제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

음원변수 추출에서 선택적 저역통과필터링

*엄 기완, *김 진영, **최 승호

*전남대학교 전자공학과, **동신대학교 정보통신공학과

Selective Low-Pass Filtering Method on Estimation of Voice Source Parameters

*Ki-Wan Eom, *Jin-Young Kim, **Seung-Ho Choi

*Dept. of Electronic Eng., Chonnam National Univ.,

(eom@dsp.chonnam.ac.kr, kimjin@dsp.chonnam.ac.kr)

**Dept. of Information and Communication Eng. Dongshin Univ.

요 약

본 논문에서는 성문파 신호로부터 음원변수들을 추출하는 방법과 그 전 단계에서 역 필터링(inverse filtering)방법에 의해 구한 미분성문파 신호로부터 고주파 잡음을 제거하기 위해 음원구간에 따라 필터의 대역폭을 달리함으로서 음원변수 추출과정에서 저역통과 필터에 의해 발생할 수 있는 오차를 최소화하기 위한 선택적 저역통과 필터링(Selective Low-Pass Filtering) 방법을 제안한다. 이 방법은 음원모델중 하나인 LF-model 펄스를 합성하여 필터링 함으로서 그 성능을 비교, 평가하였다.

1. 서 론

현재 음성분석이나 음성합성분야에서 가장 널리 사용되고 있는 음성생성 모델은 음원(voice source)이라는 입력신호가 성도(vocal tract)필터를 통과한 출력으로 간주하는 선형생성원리에 기초를 두고 있다.

그리고 성도필터는 주로 선형예측방법(LPC)에 의한 음성의 성도를 나타내는 전극(all pole)필터로 나타내며, 이 필터의 계수들을 구하는데에는 과거의 신호들의 선형결합에 의해 현재 신호를 예측한 값과, 실제의 현재 신호값과의 차이인 잔차신호를 최소화하는 방법을 사용하며, 여기에는 자기상관법, 공분산법, 격자방법등이 있다.[1]

그리고, 무엇보다 정확한 성도의 필터계수를 구하기 위해서는 성문이 닫힌구간(glottal closed region)에 해당하는 음성 데이타들을 선택하여 분석을 해야하며, 성문이 닫힌구간을 찾는 방법에는 음성신호만을 이용하는 방법이 있으나, 보다 정확한 결과를 얻기 위해 주로 Laryngograph 신호를 이용한다.[2]

지금까지 음원변수를 추출하는 여러 가지 방법들이 제안되어 왔으며, 이러한 음원신호를 모델링하고, 추출하는 방법은 음성합성, 음성코딩, 그리고 음성인식 등 여러 분야에 매우 필수적인 연구라 할 수 있다.

본 연구에서는 역 필터링(inverse filtering)방법에 의해 구한 미분성문파 신호(differential glottal signal)로부터 음원변수들을 추출하는 방법과, 추정한 미분성문파 신호로부터 고주파잡음을 제거하기 위한, 필터링 방법에서 대해 제안하고자 한다. 특히 음원신호의 저역통과 필터링에 대한 영향을 조사하기 위해 음원신호는 음성신호로부터 직접 구한 것을 사용하지 않고 LF-model함수에 의해 음원신호를 합성한 것을 사용하였다. 음원신호로서 LF-model은 음성합성 및 음성분석에 널리 사용되고 있다. 지금까지의 연구들을 보면, 음원모델에 대한 연구는 여러 방향으로 활발히 진행되어 왔으나, 이러한 잡음처리에 대한 연구는 정확한 음원변수를 추출하는데 있어 매우 중요한 부분임에도 불구하고, 몇몇 연구자들을 제외하고는 그다지 주목을 받지 못하였다.

논문의 구성은 다음과 같다. 먼저 다음 2장에서는 음성신호로부터 구한 미분성문파 신호에서 고주파잡음을 제거하기 위해 사용되어온 기존의 방법과 함께 본 논문

에서 제안하고자 하는 선택적 저역통과 필터링(Selective Low-Pass Filtering)에 대해 설명하고 3장에서는 미분성분파 신호로부터 음원변수들을 추출하기 위한 방법에 대해 설명한다. 그리고 다음 4장에서는 제안한 방법의 성능을 평가, 검토한다.

2. 선택적 저역통과 필터링

먼저 음성신호로부터 음원신호를 구하는데에는 주로 역 필터링 방법이 사용된다. 역 필터링 방법에 의해 구한 미분성분파 신호는 일반적으로 고주파 잡음을 포함하게된다. 이러한 잡음은 시간변수들인 음원변수를 추출하는데 있어 에러를 유발하게된다. 이들 잡음 처리 방법으로 저역통과 필터링을 하게 되는데, 지금까지 제안되어온 방법은 적당한 창 함수를 사용한 필터링 기법이다.[3,4] 특히 일반적인 선형 FIR필터중 Blackman 창 함수를 사용한 것과의 성능을 비교한 논문을 보면, 후자의 방법이 더 우수하다 할 수 있다. 후자의 방법에서는 크기가 19 포인트인 창 함수를 이용하였으며, 이때의 창 함수의 크기는 경험에 의해서 얻어진 값이다. 또한 창 함수의 크기가 증가할수록 즉, 필터의 대역폭이 작을수록 음원변수 추정에러는 점차 증가하게 된다.[5]

본 연구에서도 역시 미분성분파 신호에서 고주파잡음을 제거하기 위해 Blackman 창 함수와의 컨볼류션을 이용하였나, 위에서처럼 창함수의 크기로 고정된 값을 사용하지는 않았다. 즉, 미분성분파 신호에서 잡음의 영향을 많이 받는 부분에서는 필터의 대역폭이 좁게 창 함수의 크기를 조절하였고, 반대로 잡음의 영향에 덜 민감한 부분에서는 필터링을 하지 않거나, 대역폭이 매우 넓은 창 함수를 사용하여, 미분성분파 신호를 필터링함으로서 발생할 수 있는 음원변수 추정에러를 최소화 하고자 하였다.

다음 그림 1은 실제 음성신호로부터 구한 미분성분파 신호를 나타내며, 미분성분파 신호에서 성문이 열리고 닫히는 구간에 따라 잡음에 영향을 받는 정도가 다름을 알 수 있다.

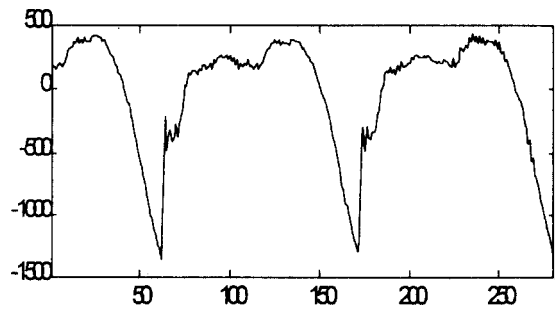

그림 1. 미분성분파 신호

위 그림에서 알 수 있듯이 각 구간(성문 열리는 구간, 열린 구간, 성문 닫히는 구간, 닫힘 구간)에 따라 잡음에 영향을 받는 정도가 다름을 알 수 있다. 이에 따라 창 함수를 사용함에 있어 고정된 크기를 사용하는 것 보다 각 구간에 따라 가변적으로 사용하는 것이 더 타당하리라 본다.

다음 그림 2의 합성한 LF-model pulse와 그의 미분신호에서도 볼 수 있듯이, 신호의 변화폭이 큰 구간 즉, 그의 미분신호가 크게 나타나는 부분에서는 잡음이 영향에 덜 민감하고, 그 반대의 경우에는 잡음의 영향에 민감하게 된다.

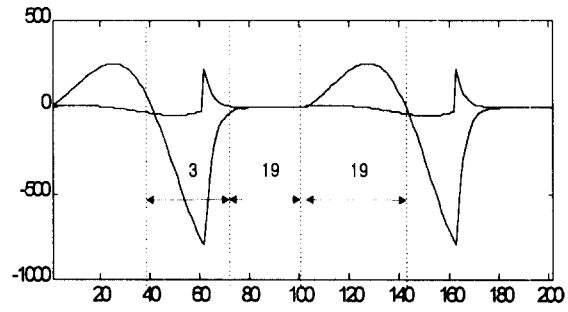

그림 2. 합성된 LF-model 펄스와 그의 미분신호

그리고 위 그림에 나타낸 바와 같이 본 연구에서는 미분성분파 신호를 3구간으로 나누어 각각 크기가 다른 Blackman창을 사용하여 저역통과 필터링을 하였다. 여기에 해당하는 크기는 경험적인 결과에 의한 것이다.

여기에서 제안한 선택적 저역통과 필터링방법과 크기가 19포인트로 고정된 Blackman 창을 사용하여 합성된 LF-model 펄스를 저역통과 필터링한 결과는 다음 그림 3에 나타내었다.

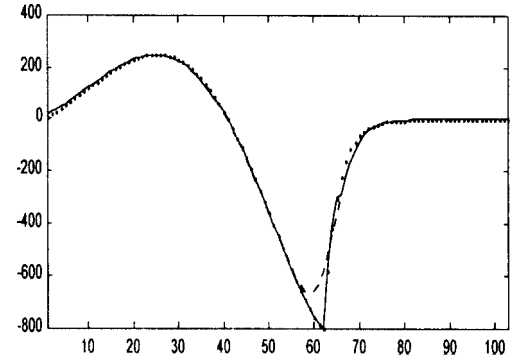

그림 3. 고정된 크기의 창함수(19포인트 Blackman 창)를 사용한 경우와 선택적 저역통과 필터링에 의한 결과.[".": 합성한 LF-model 펄스, "-": 선택적 저역통과 필터링, " - - ": 19포인트 창 함수]

3. 음원변수 추출

역 필터링 방법에 의해 구한 성문파 신호 즉 음원은 복잡한 성문벽 특성과 성도의 모델 오차를 포함하고 있으므로, 추정된 음원신호로부터 음원의 특징을 분석, 표현하기가 어렵다. 따라서 성문파의 특성을 간략화 시킨 모델을 세우고, 제한된 범위 내에서의 모델 변수들을 이용해서 음원의 특성을 제어하게 된다. 이러한 음원모델로는 Fant, Fujisaki, LF(Liljencrants- Fant)모델 등이 있으며, 본 연구에서는 음성합성분야에서 주로 이용되고 있는 LF-model을 사용하였다.[6]

LF-model 펄스와 그 파라메터들을 다음 그림 4에 나타내었다.

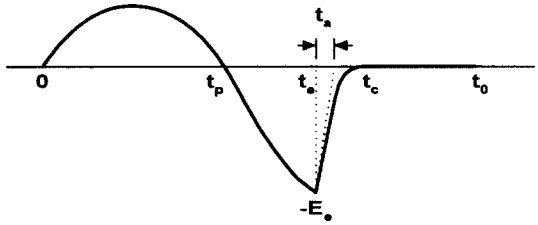

그림 4. LF-모델의 파형과 그 변수

$$E(t) = E_0 e^{\alpha t}\sin\omega_g t\ ,\ \omega_g = \frac{\pi}{t_p}\ ,\quad 0 \le t \le t_e \tag{1}$$

$$E(t) = \frac{-E_e}{\varepsilon t_a}\left(e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)}\right),\quad t_e \le t \le t_c \tag{2}$$

미분성문파 신호로부터 음원모델변수(LF-model 변수)를 구하는 방법에는 그 신호에서 직접 구하는 방법과 미분성문파 신호를 LF-model 펄스로 Fitting하는 방법이다.

먼저 전자의 방법에서는 구한 오차가 매우 크며, 위의 LF-model 변수들 중 t_a를 구할 수 없다. 그리고 두 번째 Fitting에 의한 방법에는 Simplex search 알고리즘, Levenberg-Marquardt 알고리즘등을 사용하는데 이러한 최적화(optimization)방법은 많은 계산량을 요구하게된다.

그러므로 본 연구에서는 성문 열림구간($0 \le t \le t_e$)에는 로그영역에서 미분성문파 신호와 LF-model펄스간의 차의 제곱의 합(squared error)을 최소가 되도록하는 음원변수들을 구하였으며, 성문 닫힘구간($t_e \le t \le t_c$)에서는 특이치분해(singular value decomposition)방법을 사용하였다.

성문 열림구간에 해당하는 LF-model 변수(t_p, t_e, E_e)를 구하는 과정은 다음과 같다. 변수 t_e는 역 필터링 방법에 의해 구한 미분성문파 신호에서 그 값이 최소가 되는 시간변수이므로 그 신호로부터 직접 구할 수 있다.

위의 식(1)에서 LF-모델의 펄스는 t_e 시간에 그 값이 E_e가 되므로 E_0는 다음식과 같이 쓸 수 있다.

$$\begin{aligned} -E_e &= E_0 e^{\alpha t_e}\sin\omega_g t_e \\ E_0 &= -E_e e^{-\alpha t_e}/\sin\omega_g t_e \end{aligned} \tag{3}$$

LF-model 펄스의 함수는 모델링 과정에서, 이산시간(discrete time)으로 바꿔 사용하였으며, 위 식에서 미지의 변수는 α와 E_e가 되므로, 추정된 음원신호가 $x(n)$라 한다면 다음 식을 최소하도록 하는 α와 E_e값을 구하면 된다.

$$Min_{(E_e,\alpha)}\sum_n\left[x(n) - \left(-E_e e^{\alpha(n-n_e)}\sin\omega_g n/\sin\omega_g n_e\right)\right]^2 \tag{4}$$

그러나 위 식에서 α가 지수함수의 승수로 되었으므로, 다음과 같이 두 신호에 각각 자연로그를 취한 신호에 대해서 계산하였다.

$$Min_{(E_e,\alpha)}\sum_n\left[\ln|x(n)| - \ln\left|-E_e e^{\alpha(n-n_e)}\sin\omega_g n/\sin\omega_g n_e\right|\right]^2 \tag{5}$$

여기에서 간략하게 하기 위해

$$x'(n) = \ln|x(n)| - \ln|\sin(\omega_g n)| + \ln|\sin(\omega_g n_e)|\ ,$$
$$E_e' = \ln(-E_e)$$

으로 각각 치환한다.

위 식(5)에서 E_e'에 대해 minimize하면 다음과 같다.

$$\sum_n x'(n) = N\cdot E_e' + \left(\sum_n - N\cdot n_e\right)\alpha \tag{6}$$

그리고 식(5)에서 α에 대해 minimize하면 다음 식과 같이 나타낼 수 있다.

$$\begin{aligned} &\sum_n\left(nx'(n)\right) - n_e\sum_n x'(n) = \\ &\left(\sum_n Nn_e\right)E_e' + \left(\sum_n n^2 - 2n_e\sum_n n + Nn_e^2\right)\alpha \end{aligned} \tag{7}$$

위 식(6),(7)은 행렬식으로 다음과 같이 쓸 수 있다.

$$\begin{bmatrix} E_e' \\ \alpha \end{bmatrix} = \begin{bmatrix} N & \sum_n n - Nn_e \\ \sum_n n & Nn_e\ \ \sum_n n^2 - 2n_e\sum_n n + Nn_e^2 \end{bmatrix}^{-1} \times \begin{bmatrix} \sum_n x'(n) \\ \sum_n\left(nx'(n) - n_e\sum_n x'(n)\right) \end{bmatrix} \tag{8}$$

단, $E_e = e^{E_e'}$.

위의 행렬식을 풀면 성문 열림구간에서 추정한 미분성문파 신호와 LF-model 펄스간의 에러를 최소화하는 모델 변수들을 얻을 수 있다. 그리고 식(1)을 보면 변수 ω_g는 $\omega_g = \pi/t_p$이다. 여기에서 t_p는 LF-모델 펄스에서 보면 0이 되는 지점이다. 그러나 음원신호로부터 t_p를 직접 구하기 어렵고, 직접 구했을 때 많은 오차가 발생한다. 그러므로 추정된 음원신호로부터 대략적인 t_p을 찾아 그 지점에서 좌우 1 point 정도씩 변화 시켜가면서 미분성문파 신호와 LF-model 펄스간의 에러가 최소가 되는 지점을 t_p로 정하였다.

또한 성문 닫힘구간에 해당하는 LF-model 변수들 (t_a, t_c)은 특이치분해 방법에 의해 쉽게 구할 수 있다. 다음 그림 5에는 합성된 LF-model 펄스를 선택적 저역통과 필터링을 수행한 후 이를 본 절에서 기술한 방법에 의해 LF-model 펄스로 fitting한 결과를 나타내었다.

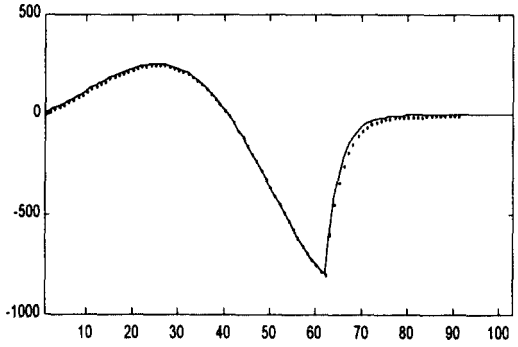

그림 5. LF-model 펄스와 선택적 저역통과 필터링방법에 의해 그 펄스를 필터링한 후 Fitting한 결과.["-": LF-model 펄스, " · ": Fitting 결과]

그리고 아래 그림은 실제 음성신호에서 미분성문파 신호를 구하고 이를 선택적 저역통과 필터링을 거친 다음 LF-model 펄스로 fitting한 결과이다.

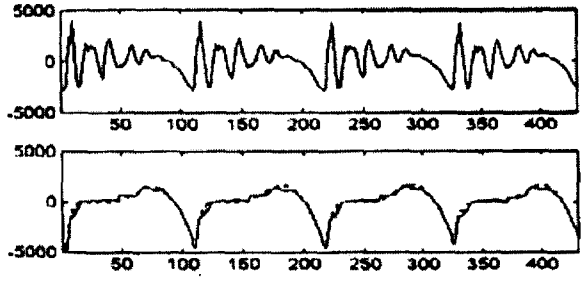

그림 6. LF-modelling 결과.["-": 미분성문파 신호, " - - ": LF-model펄스로 Fitting된 신호]

4. 결 론

이상과 같이 본 논문에서는 역 필터링방법에 의해 구한 미분성문파 신호로부터 음원변수를 추출하는 방법에 대해 기술하였다. 특히 미분성문파 신호에서 고주파 잡음을 제거하기 위한 방법으로, 성문의 열리고 닫히는 구간에 따라 필터의 대역폭을 가변적으로 사용하여 음원변수를 구할 때 발생할 수 있는 에러를 최소화 할 수 있는 선택적 저역통과 필터링(S-LPF)방법을 제안하였다. 이러한 방법은 현재 음색에 따른 음원의 특성을 연구하는 음성분석분야에 매우 유용하게 사용될 수 있으리라 기대된다.

참고문헌

[1] J. D. Markel & A. H. Gray. 1976. *Linear Prediction of Speech.* Springer-Verlag.

[2] D. G. Childers & Wong, C. F. 1994. "Measuring and modeling vocal source-tract interaction." *IEEE Trans. Biomde. Eng. 41*, 663-671.

[3] P. Alku. 1995. "Effects of bandwidth on glottal airflow waveforms estimated by inverse filtering." *J. Acoust. Soc. Am. 98*, 763-767.

[4] Helmer Strik & Boves, L. 1994. "Automatic estimation of voice source parameters." *Proc. ICSLP 94(1)*, 155-158.

[5] Helmer Strik. 1996. "Comments on "Effects of bandwidth on glottal airflow waveforms estimated by inverse filtering"[J. Acoust. Soc. Am. 98, 763-767 (1995)]." *J. Acoust. Soc. Am. 100*, 1246-1249.

[6] G. Fant et al. 1989. "Voice source rules for text-to-speech synthesis." *Proc. ICASSP 89(1)*, 223-226.

※ 본 논문은 '97대학기초연구지원사업'의 연구결과중 일부입니다.