

MPE-LPC 음성합성에서 Maximum-Likelihood Estimation에 의한 Multi-Pulse의 크기와 위치 추정

(Multi-Pulse Amplitude and Location Estimation by Maximum-Likelihood
Estimation in MPE-LPC Speech Synthesis)

李 基 鎔*, 崔 洪 燮*, 安 秀 桔*

(Ki Yong Lee, Hong Sub Choi, and Souguil ANN)

要 約

본 논문에서는 multi-pulses을 음원으로 사용하는 MPE-LPC 음성합성에서 pulses의 크기와 위치를 구하는 방법으로 maximum-likelihood estimation를 사용하는 방법을 제시한다. 이 MLE 방법은 관측된 데이터에서 미지의 파라미터(pulse의 크기와 위치)에 대하여 likelihood function 최대화 하는 값을 구하는 것으로, pulse가 overlap된 경우 Ozawa의 cross-correlation 방법과 같으므로, MLE의 계산량과 음질은 cross-correlation 방법과 같음을 알 수 있다.

컴퓨터 시뮬레이션에 의하여 MLE 방법에 의하여 구한 multi-pulse는 (1)유성음인 경우, 피치의 준주기성을 가지고 있고, (2)무성음인 경우, 펄스가 불규칙하게 나타나며, (3)무성음에서 유성음으로 바뀌는 부분에서도 펄스가 불규칙에서 피치의 주기성으로 변화되는 것을 알 수 있었고, 그리고 스펙트럼에서 multi-pulse를 음원으로 하여 합성한 음과 원음의 포오만트 크기가 같음을 알 수 있었다.

Abstract

In this paper, We propose a maximum-likelihood estimation (MLE) method to obtain the location and the amplitude of the pulses in MPE (multi-pulse excitation)-LPC speech synthesis using multi-pulses as excitation source. This MLE method computes the value maximizing the likelihood function with respect to unknown parameters (amplitude and position of the pulses) for the observed data sequence. Thus in the case of overlapped pulses, the method is equivalent to Ozawa's crosscorrelation method, resulting in equal amount of computation and sound quality with the cross-correlation method.

We show by computer simulation : the multi-pulses obtained by MLE method are (1) pseudo-periodic in pitch in the case of voiced sound, (2) the pulses are random for unvoiced sound, (3) the pulses change from random to periodic in the interval where the original speech signal changes from unvoiced to voiced. Short time power spectra of original speech and synthesized speech obtained by using multi-pulses as excitation source are quite similar to each other at the formants.

*正會員, 서울대학교 電子工學科
(Dept. of Elec. Eng., Seoul Nat'l Univ.)
接受日字: 1989年 2月 14日

I. 서 론

지금까지 낮은 Bit-rate의 음성통신을 위한 음성의 분석 및 합성방법으로는 선형예측 부호화(LPC)

방법이 많이 연구되어져 왔다.^[4,5]

비록 LPC방법이 음성에 잘 적용되기는 하지만 이러한 모델에 의하여 합성된 합성음은 다음의 문제때문에 질이 나빠진다. 첫째는 분석하는 데이터를 voice/unvoice/silence로 판단할 때, 잘못 판단하는 경우, 해당되는 프레임에 분석하는 방법이 다르게 되므로 합성음의 질이 저하된다. 둘째는 유성음시 피치를 검출할 때 일어나는 문제로 정확하지 않는 피치를 추출하는 경우 합성단에서 피치를 음원으로 사용하므로 합성된 음의 질이 떨어진다. 그리고 올바르게 voice를 판단하고, 정확하게 피치를 검출하는 경우에도 합성된 음은 한프레임에 하나의 피치를 사용하므로 자연스럽지 못하게 들리게된다. 또한 음원을 유성음인 경우 pulse, 무성음인 경우 백색잡음만을 고려하지만, 실제 음성의 음원은 두가지로만 구분되지 않고 때로는 복합적인 경우도 있다.

따라서 자연스럽게 질 높은 합성음을 얻기 위하여 보다 정확한 음원 모델을 구하기 위하여 최근들어 많은 연구가 진행되었다.^[6,7] 이러한 방법중의 하나로 음원으로 하나의 pulse가 아닌 다수개의 pulse를 이용하는 알고리즘을 사용함으로써 v/u/s 판단과 피치 검출에서 생기는 문제를 해결하여 자연스럽게 질 높은 합성음을 얻을 수 있었다.

이러한 방법으로 최근에 Atal과 Remede는 새로운 음원(multi-pulse mode)을 제안하였다.^[8] 이 방법은 보통 multi-pulse excitation로 알려져 있고, 이 음원 모델을 사용함으로써 좋은 음질의 합성음을 얻을 수 있다. 이 multi-pulse excited(MPE) 방법에 의한 음성합성 과정은 합성 filter와 상이한 위치와 크기를 가진 다수의 pulse를 출력으로 하는 음원 발생기에 의하여 모델화할 수 있다.

합성 filter의 특성은 기존의 LPC 분석에 의하여 쉽게 얻을 수 있다. 그러나, 음원으로 사용하는 pulse들의 위치와 크기를 동시에 구하는 것은 비선형 방정식을 풀어야 함으로 매우 어렵다.

이러한 문제를 해결하기 위하여, Atal과 Remede는 한번의 과정에 하나씩 pulse의 크기와 위치를 구하는 A-b-S (analysis-by-synthesis)에 기본을 둔 순차적으로 pulse를 찾는 방법을 제안하였고,^[9] Ozawa와 Ono, Araseki는 cross-correlation 방법을 기본으로 하여 pulse의 위치와 크기를 찾는 방법을 제안하였다.^[6,7]

본 논문에서, pulse의 크기와 위치를 구하는 문제를 통계학적 통신이론(Statistical Communication Theory)에서 잘 알려진, 관측된 데이터에서 미지의 파라미터(여기서는 pulse의 크기와 위치)를 추정하기

위하여 likelihood function(또는 log-likelihood function)을 최대화하는 파라미터를 택하는 Maximum-Likelihood Estimation^[6,7]을 이용한 방법을 제안한다. 이 방법은 pulse가 overlap된 경우 Ozawa의 cross-correlation 방법과 같다.

그리고 논문의 구성은 II장에서 기본적인 MPE 방법을 간단히 소개하고, III장에서는 제안하려는 MLE 방법과 그 방법에 의하여 pulse의 크기와 위치를 구하는 과정을 소개하고, IV장에서는 제안한 방법에 대한 컴퓨터 시뮬레이션 결과를, V장에서는 결론 순으로 되어있다.

II. 음성부호화에서의 MPE

기본적인 음성합성 과정은 그림 1에서 보는 바와 같이 음원부와 합성 filter에 의하여 나타낼 수 있다. 음원에 따라서 음성은 기본적으로 유성음과 무성음 두가지로 나눌 수가 있다. 유성음인 경우 음원은 pitch 주기에 해당하는 간격을 가진 준주기적 impulse 열에 의하여 나타내고, 무성음인 경우 음원은 백색잡음으로 나타낸다. 이러한 일반적인 음원 모델은 매우 간단하여 bit-rate를 2kbit/s로 낮출 수가 있지만, pitch 검출에 의한 오차와 voice/unvoice 결정에 의한 오차로 인하여 좋은 음질의 합성음을 얻을 수가 없다.

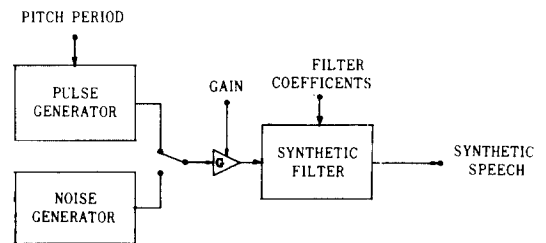


그림 1. LPC음성 합성 모델

Fig. 1. LPC speech synthesis model.

위의 이러한 문제를 해결하고 합성음의 질을 높이기 위하여 multipulse excitation(MPE) 모델이 제안되었다. MPE 모델을 사용하는 음성합성 과정은 그림 2와 같다. 그림 2에서 음원은 시간 m_1, m_2, \dots, m_M 에서 크기 g_1, g_2, \dots, g_M 을 가진 multipulse를 출력으로 한다. 이렇게 발생된 pulse열은 합성 filter의 입력으로 사용되어 음성을 합성한다.

합성 filter의 계수는 LPC 방법에 의하여 쉽게 얻을 수 있다.^[10]

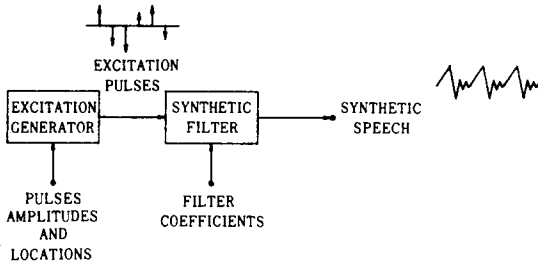


그림 2. Multi-pulse 음원 모델
Fig. 2. Multi-pulse excitation model.

MPE 방법에 의하여 합성된 음은 낮은 bit-rate에서 질 높은 합성음을 얻을 수가 있으나, 한 frame내의 모든 pulse의 크기와 위치를 동시에 알기 위해서는 비선형 방정식을 풀어야 함으로 매우 복잡하고 어렵다.

이 문제의 해결로 Atal과 Remede는 A-b-S 방법을 기본으로 하는 sequential pulse search 방법을 제안하였다. 그림 3은 pulse search과정을 나타낸 것이다. 이 과정에서 한 pulse의 크기와 위치는 원음과 합성음 사이의 mean-squared-weighted-error를 최소화하는 크기와 위치를 한번에 하나씩 원하는 수의 pulse나 최소 error의 임계값에 도달할 때까지 여러번 반복과정을 거쳐 적절한 펄스의 위치와 크기를 선택한다.

이때, 합성된 음성은 전단에서 얻어진 pulse가 합성 filter를 통과함으로써 계속해서 얻어진다.

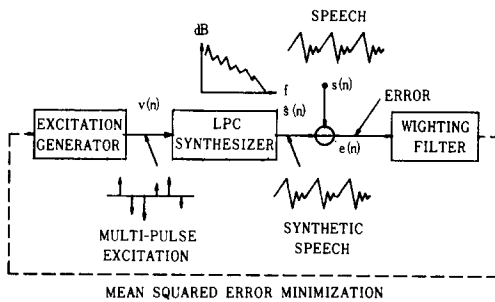


그림 3. ABS 방법의 블록도
Fig. 3. Blockdiagram of ABS method.

Weighting filter는 perceptual distortion을 줄이기 위하여 사용된다. 이 filter의 특성은 다음과 같이 주어진다.^[5]

$$W(z) = \frac{1 - \sum_{i=1}^p a_i z^i}{1 - \sum_{i=1}^p a_i r^i z^i} \quad (1)$$

여기서 a_i 는 합성 filter의 계수, p 는 filter의 차수, r 은 perceptual weighting factor. r 이 1이면 이 경우, $W(z)=1$ 로 weighting이 되지 않는다. 합성음 내에 존재하는 잡음은 거의 평탄한 spectrum 특성을 가지고 있다. 이러한 잡음의 특성으로 low-bit rate의 음성 coder에서 잡음이 고주파 대역과 formant들 사이에서 들리게 된다. 이와같은 부분에서 잡음을 억제하고, formant 부분에서 잡음을 원음에 의하여 mask시키기 위하여 weighting filter를 사용한다. r 의 값은 $0 < r < 1$ 에서 선택한다.

III. Maximum-Likelihood Estimation (MLE)에 의한 Pulse 추정

이 장에서는 MLE 방법에 의하여 pulse의 크기와 위치를 추정하는 문제를 생각하기로 한다. MPE 음성합성 모델에서 합성음은 pulse g_i 와 합성 filter의 impulse response $h(j)$ 의 convolution으로 볼 수 있다.^[11]

이 문제에서의 수학적 모델은 아래와 같이 나타낼 수 있다. pulse의 수를 M 이라 놓으면, 합성음은

$$Y'_j = \sum_{i=1}^M g_i h(j-m_i) \quad (2)$$

으로 볼 수 있고,

이때 원음은 식(3)과 같이 모델화 할 수 있다.

$$Y_j = \sum_{i=1}^M g_i h(j-m_i) + E_j \quad (3)$$

단, $j=1, 2, \dots, N$.

여기서 Y'_j 는 합성음을, Y_j 는 원래의 음성신호이다. N 은 한 프레임 내의 데이터 길이, h 는 시스템의 특성을 나타내며 알고 있다고 가정한다. g_i 는 pulse의 크기(random), m_i 는 pulse의 위치(random) 그리고 E_j 는 원음과 합성음 사이의 에러(random)로 본다. 그리고 에러 $E_j (= Y_j - Y'_j)$ 는 mean이 0이고, variance가 v^2 인 gaussian으로, g_i 와 m_i 는 서로 독립이라 가정한다.

Weighting filter의 impulse response를 $w(j)$ 라 놓고, weight된 신호를 아래와 같이 정의하기로 한다.

$$hw(j) = w(j) * h(j) \quad (4a)$$

$$S_j = w(j) * Y_j \quad (4b)$$

$$S'_j = w(j) * Y'_j \quad (4c)$$

식 (3)에서 pulse의 수가 고정되어 있는 경우만을 생각하기로 한다. Maximum-Likelihood 방법은 파라미터 추정을 위한 일반적으로 잘 알려진 방법중의 하나이다. ML 방법은 likelihood function를 최대화 하도록 미지의 파라미터를 선택하는 문제로 볼 수 있다. Likelihood function은 추정하려는 파라미터 (pulse의 크기와 위치)와 관측된 데이터(원래의 음성신호)의 함수이다. 그러므로, pulse의 크기와 위치를 추정하기 위해 사용되는 log-likelihood function은 3 가지 항으로 구성된다.¹²⁾ 1항은 pulse의 크기와 위치에 대한, 원래의 음성신호(s_1, s_2, \dots, s_n)의 log-likelihood 이다.

$$L_1 = \log \{ p_{s_1, \dots, s_n | g_1, \dots, g_M, m_1, \dots, m_M} (S | g, m) \}$$

$$= \sum_{j=1}^n \log \{ p_E (s_j - \sum_{i=1}^M g_i hw(j - m_i)) \} \quad (5)$$

여기서 $S^T = (s_1, \dots, s_n)$ 는 관측된 신호벡터이고 $g^T = (g_1, \dots, g_M)$ 와 $m^T = (m_1, \dots, m_M)$ 은 pulse의 크기와 위치 벡터, p_E 은 에러($S_j, -S'_j$)의 확률밀도 이다.

그리고 2항과 3항은 pulse의 크기와 위치에 대한 a priori log-likelihood function L_2, L_3 으로 나타난다.

$$L_2 = \sum_{i=1}^M \log \{ p_g(g_i) \} \quad (6)$$

여기서 p_g 는 pulse 크기의 확률밀도 이다.

$$L_3 = \log \{ p_{m_1, \dots, m_M} | M(m | M) \} \quad (7)$$

식 (7)은 pulse의 수가 M이므로, m의 joint probability density의 loglikelihood function이다. 이러한 확률 밀도는 식 (8)에 의하여 구할 수 있다.

$$p_{m_1, \dots, m_M} | M(m | M)$$

$$= \begin{cases} M!/N^M & 0 < m_1 < \dots < m_M < N \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

$N = -L_1 - L_2 - L_3$ 이라하면,

$$N(g, m, M) = \frac{1}{2\sigma^2} \sum_{j=1}^n [s_j - \sum_{i=1}^M g_i hw(j - m_i)]^2$$

$$+ \frac{1}{2v^2} \sum_{i=1}^M g_i^2 + \frac{1}{2} N \log(2\pi \sigma^2)$$

$$+ \frac{1}{2} M \log(2\pi v^2) - \log(M!/N^M) \quad (9)$$

pulse의 크기와 위치의 maximum-likelihood를 추정하기 위해서 g 와 m 에 관해 $N(g, m, M)$ 를 최소화 하는 값을 얻으면 된다. 식 (9)에서 g 와 m 에 관한 항은 단지 앞의 3항으므로 아래와 같이 다시 쓰면,

$$N'(g, m, M) = \frac{1}{2} \left(\frac{d}{v}\right)^2 \left[\frac{1}{\alpha} \sum_{j=1}^n s_j^2 - 2 \sum_{i=1}^M g_i z(m_i) + \sum_{i=1}^M \sum_{t=1}^M g_i g_t r(m_i, m_t) + \frac{1}{d^2} \sum_{i=1}^M g_i^2 \right] \quad (10)$$

여기서,

$$\alpha = \sum_{j=1}^n hw^2(j), \quad (11)$$

$$d = \left(\frac{\alpha v^2}{\sigma^2}\right)^{-1/2} \quad (12)$$

$$z(m_i) = \frac{1}{\alpha} \sum_{j=1}^n s_j hw(j - m_i), \quad (13)$$

$$i = 1, 2, \dots, M$$

$$r(m_i, m_t) = \frac{1}{\alpha} \sum_{j=1}^n hw(j - m_i) hw(j - m_t), \quad (14)$$

$$i, t = 1, 2, \dots, M$$

식 (14)은 다음과 같이 근사화 할 수 있다.

$$r(m_i, m_t) \approx \frac{1}{\alpha} \sum_{j=1}^n hw(j - m_i) hw(j - m_t)$$

$$= \rho(m_i - m_t)$$

$$\text{단, } \rho(\tau) = \frac{1}{\alpha} \sum_{j=1}^n hw(j) hw(j - \tau) \quad (16)$$

함수 ρ 는 시스템 impulse response의 normalize된 autocorrelation 이다. 식 (10)을 vector 형태인, R_m 로 놓으면

$$R_m = \frac{1}{2} \left(\frac{d}{v}\right)^2 (\mu - 2g^T z_m + g^T H_m g) \quad (17)$$

$$\text{단, } \mu = \frac{1}{\alpha} \sum_{j=1}^n s_j^2 \quad (18)$$

$$z_m = \text{col}(z(m_1), \dots, z(m_M)),$$

H_m 는 (i, t)의 element가 $r(m_i, m_t) + \delta_{i,t}/d^2$ 인 ($M \times M$) 행렬이다. 여기서 error가 0인 경우 $d^2 \approx \infty$ 이므로 $H_m = r(m_i, m_t)$ 이 된다.

식 (17)을 g 에 대하여 미분을 한 후 최소값을 구하면,

$$g_m = H_{m-1}^{-1} z_m \quad (19)$$

이고, 식 (19)를 식 (17)에 대입하면

$$R_m = \frac{1}{2} \left(\frac{d}{v}\right)^2 (\mu - z_m^T H_{m-1} z_m) \quad (20)$$

식 (20)의 R_m 에서 m 에 관한 최소값을 구하는 것이 어렵다.

만약 pulse가 overlap되지 않았다면, 즉 pulse의 간

격 $(m_i - m_k)$ 이 매우 커서 $g(m_i - m_k)$ 의 값이 0에 매우 가깝다. 이 경우 식 (20)은 식(21)로 근사화 할 수 있다.

$$R_m = \frac{1}{2} \left(\frac{d}{v} \right)^2 \left[\mu - \frac{d^2}{1+d^2} \sum_{i=1}^M z^2(m_i) \right] \quad (21)$$

식 (21)에서 pulse의 위치는 쉽게 구할 수가 있다. R_m 의 m_1, m_2, \dots, m_M 에 대한 최소값은 $\sum z^2(m_i)$ 의 최대값에서 존재하므로 $z^2(\tau), \tau \in [0, N-1]$ 가 최대값을 가지는 M개의 위치를 구하면된다. 그러나 위의 과정은 pulse가 overlap되었을 때는 사용할 수가 없다. 따라서 우리는 pulse가 overlap 되었을 경우에 대해 다른 방법을 생각하여야 한다. z 를 random process Z로 생각하면,

$$Z(\tau) = \frac{1}{\alpha} \sum S_j hw(j-\tau), \tau \in [0, N-1] \quad (22)$$

식 (20)에 식 (3)를 대입하면

$$\begin{aligned} Z(\tau) &= \frac{1}{\alpha} \sum_{j=1}^M \left[\sum_{i=1}^M g_i hw(j-m_i) + E_j \right] hw(j-\tau) \\ &= \sum_{i=1}^M g_i \frac{1}{\alpha} \sum_{j=1}^M hw(j-\tau) hw(j-m_i) \\ &+ \sum_{j=1}^M E_j hw(j-\tau) \\ &= \sum_{i=1}^M g_i r(\tau, m_i) + U(\tau) \\ &\approx \sum_{i=1}^M g_i \rho(\tau - m_i) + U(\tau) \end{aligned} \quad (23)$$

여기서 에러가 없다고 ($E=0$) 가정하면, $U(\tau) = 0$ 이 된다. 따라서 $d^2 \rightarrow \infty$ 이 되므로 $H_m = r(m_i, m_k)$ 이 된다. 식 (19)에 대입하면, 식 (19), (23)은 Ozawa의 cross-correlation 방법과 같으므로 pulse의 크기와 위치를 쉽게 구할 수 있다.^{6,7)}

식 (23)에서 만약 pulse의 위치 m 를 알고 있다면, pulse의 크기 $\theta_i, i=1, \dots, M$ 는 식 (19)에서 쉽게 구할 수가 있다. pulse의 위치를 구하기 위하여 pulse의 수만큼 반복하는 순차적인 방법을 생각한다.

- (1) set $i=1$.
- (2) $z'(\tau)$ 를 $\tau \in [0, N-1]$ 구한다.
- (3) pulse의 위치 m_i 를 구한다.

$$|z'(m^i)| = \max |z'(\tau)|, \tau \in [0, N-1] \quad (24)$$

- (4) m_i 를 이용하여 pulse의 크기를 구한다.

$$g_{m_i} = H_{m_i}^{-1} z_{m_i}^i \quad (19)$$

- (5) $i=i+1$

$$(6) z'(\tau) = z(\tau) - \sum_{j=1}^{i-1} g_j \rho(\tau - m_j). \quad (25)$$

- (7) 만약 $i < M$ 이면 (3)으로가고, 아니면 정지한다.

이상에서 우리는 statistical parameter(에러의 variance σ^2 , pulse 크기의 variance v^2)를 알고 있다고 가정을 했는데, 우리가 원하는 식 (8)에서 필요한 것은 d 의 값이므로, 만약 d 의 값이 매우 크다면 (에러가 전혀 없다면) $1+1/d^2 \approx 1$ 이 되므로 무시할 수 있다. 실제의 경우에서 $d \rightarrow \infty$ 로 생각하므로 pulse의 크기와 위치를 과정 (1), (2), (3), (4), (5), (6), (7)를 반복 함으로써 구할 수 있다.

IV. 컴퓨터 시뮬레이션 결과 및 검토

III장에서 제안한 MLE 방법을 테스트하기 위하여 컴퓨터 시뮬레이션을 수행하였다. 여기서 사용된 데이터는 남성화자에 의하여 발음된 음성('에', '사')을 데이터로 사용하였다.

화자에 의하여 발생된 음성을 관측된 데이터로 보고 MLE 방법에 의하여 펄스의 위치와 크기를 구하고, 유성음에서의 multi-pulse가 피치에 의한 준주기성을 가지고 있는지, 무성음에서는 펄스가 불규칙하게 나타나는지, 또 무성음에서 유성음으로 변화할때 구해진 펄스의 특성이 유성음 부분에서는 피치의 주기성이 있는가를 알아보고, 이러한 multi-pulse를 음원으로하여 합성 filter에 통과시켜 음성을 합성한 후 원래의 음성과 비교하여 보았다. 이때, 화자에 의하여 발생된 음성신호는 마이크로폰을 통하여 녹음테이프에 저장시킨 뒤, 3.4kHz의 차단주파수를 갖는 저역여파기에 통과시키고 8kHz로 샘플링하여 12bit로 양자화하여 분석, 합성을 하였다.

음성분석은 LPC 방법으로 하였고 이때, 한 프레임의 구간은 80샘플(10msec), filter의 차수는 10차로 하여 autocorrelation 방법에 의하여 LPC 계수를 구하였고, 필요한 합성 filter의 impulse response sequence를 얻어진 LPC 계수에 의하여 구하였다. 그리고 pulse의 크기와 위치는 1프레임에 8개 즉 1개/10 sample로 구하였다.

그림 4는 $M=4$ 일때 $m=0,2,3,4$ 의 증가에 의한 $z^m(\tau)$ 의 변화를 나타내고 있다. 그림에서 보는 바와 같이 가장 큰 pulse부터 제거됨을 볼 수가 있고, $d \rightarrow \infty$ 로 가정하였다.

원음과 합성음의 파형은 그림 5와 6에 나타냈다. 그림 5는 순수한 유성음인 '에'음을 나타낸 것이고, 그림 6은 무성음+유성음인 '사'음을 나타낸 것이다. '에'음의 합성음과 원음에 대한 스펙트럼은 그림 7에 나타냈다. 이 그림에서 합성음과 원음의 포오만트의 크기가 유사함을 알 수가 있다. 그림 8은 각 음에 대하여 MLE 방법으로 구한 multi-pulse열을 나타낸 것이다.

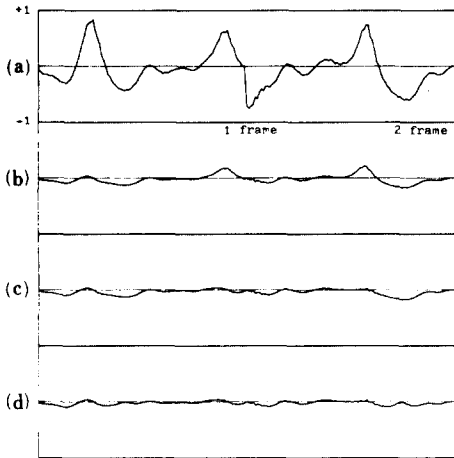
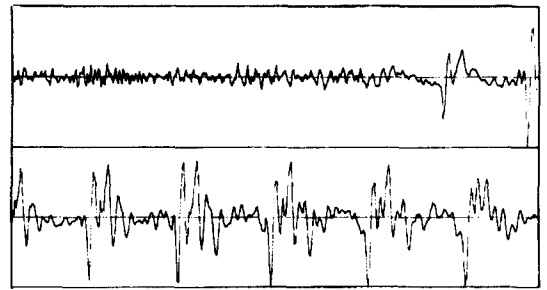
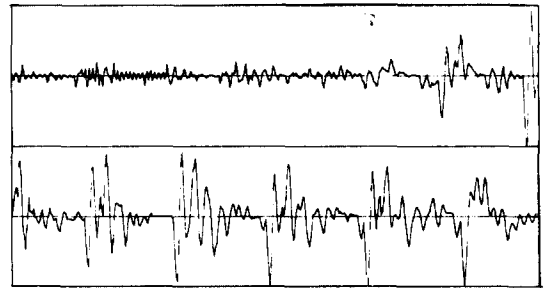


그림 4. i 에 의한 $z'(\tau)$ 의 변화
 (a) $i = 0$ (b) $i = 2$
 (c) $i = 3$ (d) $i = 4$
 Fig. 4. The change of $z'(\tau)$ by i .
 (a) $i = 0$. (b) $i = 2$.
 (c) $i = 3$. (d) $i = 4$.

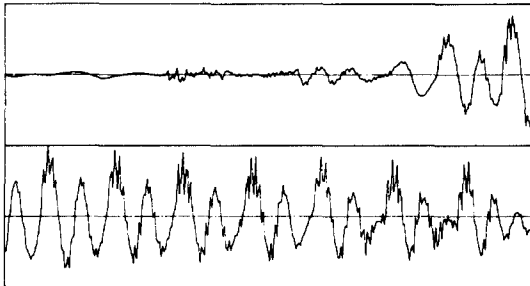


(a) 원 음

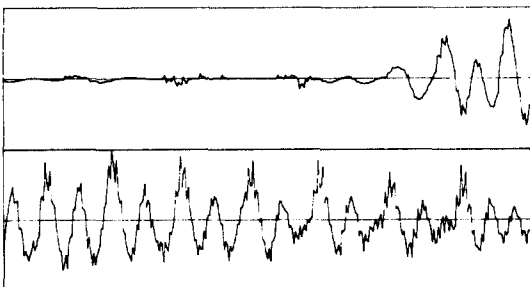


(b) 합성음

그림 6. /사/음의 원음과 합성음 파형
 Fig. 6. Waveform of original speech and synthetic speech for syllable /사/.



(a) 원 음



(b) 합성음

그림 5. /에/음의 원음과 합성음 파형
 Fig. 5. Waveform of original speech and synthetic speech for syllable /에/.

그림 8(a)에서 보는 바와 같이 유성음인 경우, 음원의 준주기적인 성격이 유성음/무성음의 판단 정보, 피치 주기의 정보 없이도 잘 나타남을 알 수가 있다. 그리고 그림 8(b)에서 파형의 앞 부분('ㅅ')인 무성음 부분에서 기존의 무성음 LPC 합성에서 백색잡음을 음원으로 사용하는 것과 같이 펄스가 불규칙하게 나타나는 것을 볼 수가 있고, 무성음에서 유성음으로 변화되는 부분(ㅅ→ㅌ)에서도 펄스가 불규칙에서 주기성으로 변화되는 것을 알 수가 있다.

V. 결 론

본 논문에서는 최근들어 많이 연구되고 있는 유성음/무성음 구별없이 multi-pulse를 음원으로 사용하는 MPE-LPC 음성합성 방법에서 pulse의 위치와 크기를 구하는 방법으로 maximum-likelihood estimation를 사용하는 방법을 제안하였다. 이 MLE 방법은 pulse가 overlap된 경우 Ozawa의 cross-correlation 방법과 같으므로 계산량과 음질은 cross-correlation 방법과 같다.

시뮬레이션 결과에서 알 수 있듯이 MLE 방법에

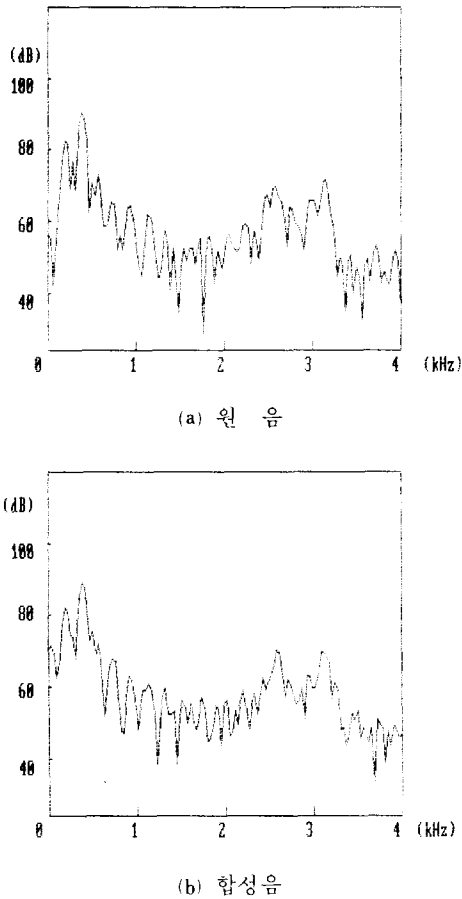


그림 7. /에/음의 원음과 합성음의 스펙트럼
Fig. 7. Spectrum of original and synthetic speech signals.

의하여 구한 multi-pulse는 유성음인 경우 피치의 준 주기성이 뚜렷이 나타나고, 무성음인 경우, 펄스가 불규칙하게 나타나며, 무성음에서 유성음으로 바뀌는 부분에서도 펄스가 불규칙에서 피치의 주기성으로 변화되는 것을 알 수 있었다. 그리고 이러한 multi-pulse를 음원으로 사용하여 얻은 합성음과 원음의 파형이 유사하고, 스펙트럼에서 보는 바와 같이 포오만트의 크기가 거의 같음을 알 수 있다.

參 考 文 獻

[1] R.E. Crochiere, R.V. Cox, and J.D. Johnston, "Real-time speech coding," *IEEE Trans. Commun.*, vol. COM-30, pp. 621-634, Apr. 1982.

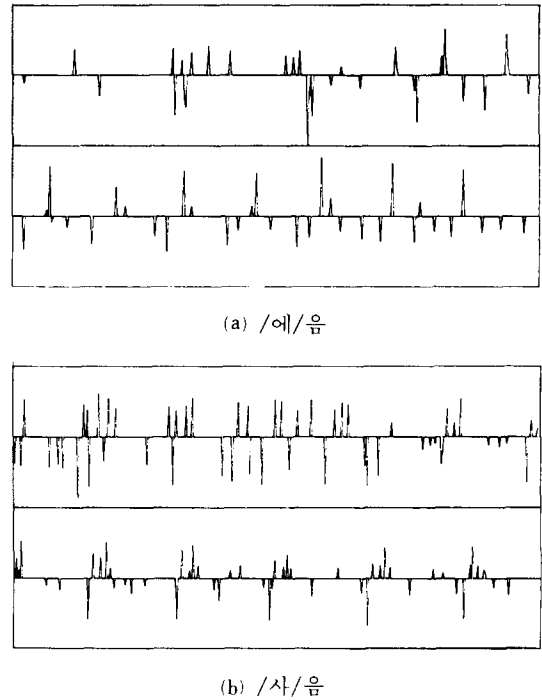


그림 8. 추출된 /에/음과 /사/음의 multi-pulse 열
Fig. 8. Extracted multi-pulse sequence for syllable /에/, /사/.

- [2] B.S. Atal, "Predictive coding of speech at low bit rates," *IEEE Trans Commun.*, vol. COM-30, pp. 600-614, Apr. 1982.
- [3] L.J. Siegel and A.C. Bessey, "Voiced/unvoiced/mixed excitation classification of speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 451-460, Jun. 1982.
- [4] S.Y. Kwon, A.J. Goldberg, "Anenhanced LPC vocoder with no voiced/unvoiced switch," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 851-858, Aug. 1984.
- [5] B.S. Atal and J.R. Remede, "A new model of LPC excitation for producing natural-sounding speech at low bit rates," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1982, pp. 614-617.
- [6] T. Araseki, K. Ozawa, S. Ono, and K. Ochiai, "Multi-pulse excited speech coder based on maximum crosscorrelation search algorithm," in *Proc. IEEE Global Telecommun. Conf.*, 1983, pp. 794-798.

[7] K. Ozawa, S. Ono and T. Araseki, "A study on pulse search algorithms for multipulse excited speech coder realization," *IEEE J. Select. Areas Commun.*, vol. SAC-4, Jan. 1986.

[8] H.L. Van Trees, *Detection, Estimation and Modulation Theory. Part I, III.* Wiley, New York, 1971.

[9] A.D. Whalen, *Detection of signals in noise.* Academic Press, 1971.

[10] J.D. Markel and A.H. Gray, Jr., *Linear Prediction of Speech*, New York: Springer-Verlag, 1976.

[11] A.V. Oppenheim, "Speech analysis-synthesis system based on homomorphic filtering," *JASA.*, vol. 45, pp. 458-465, Feb. 1969.

[12] J.J. Kormylo and J.M. Mendel, "Maximum likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 482-488, May, 1982.

著 者 紹 介



李 基 鎔 (正會員)

1960年 5月 8日生. 1983年 2月
 숭실대학교 공과대학 전자공학과
 졸업(B. S) 1985年 2月 서울대학
 교 대학원 전자공학과 졸업(M. S)
 1985年 3月~현재 서울대학교 대
 학원 전자공학과 박사과정. 주관심

분야는 통신이론, 시스템인식 및 신호처리임.



安 秀 桔 (正會員)

1930年 4月 17日生. 1950年 5月
 서울대학교 공과대학 통신공학과
 졸업(B. S) 1957年 4月 서울대학
 교 대학원 전자공학과 졸업(M. S)
 1974年 3月 서울대학교 대학원 전
 자공학과 졸업(Ph. D) 1969年 3월

~현재 서울대학교 전자공학과 교수. 1980年~현재 IEEE 원로회원 및 한국지부회장. 주관심분야는 통신이론, 디지털통신, 정보이론 및 회로이론 등임.



崔 洪 燮 (正會員)

1957年 10月 3日生. 1985年 2月
 서울대학교 공과대학 전자공학과
 졸업(B. S) 1987年 2月 서울대학
 교 대학원 전자공학과 졸업(M. S)
 1987年 3月~현재 서울대학교 대
 학원 전자공학과 박사과정. 주관심

분야는 디지털신호처리(음성신호), 통신이론임.