논문 2007-44SP-1-13

# Damping 요소를 첨가한 매칭 퍼슈잇 정현파 모델링

## ( Matching Pursuit Sinusoidal Modeling with Damping Factor )

정 규 혁*, 김 종 학*, 임 정 우*, 주 기 호**, 이 인 성***

( Gyu-Hyeok Jeong, Jong-Hark Kim, Joung-Woo Lim, Gi-ho Joo, and In-Sung Lee )

요 약

본 논문은 정현파 모델 기반의 코덱을 위한 매칭 퍼슈잇(Matching Pursuit)의 성능을 개선시킨 새로운 정현파 모델링을 제안한다. 제안하는 damping 요소를 첨가한 매칭 퍼슈잇 정현파 모델링은 과거와 현재 프레임에서 파라미터들간의 상관성을 이용하여 damping 요소를 정의하고 현재 프레임에서 보다 정확한 정현파 파라미터를 damping 요소에 따라 매칭 퍼슈잇 방법으로 추출한 후 합성한다. 따라서 인접 프레임과의 보간 없이 현재 프레임에서의 정현파 파라미터만으로 효율적인 모델링이 가능하다. 제안한 모델링 방법은 보간법을 사용한 일반적인 정현파 모델과 달리 추가지연을 가지지 않으면서 유성음 구간 신호뿐만 아니라 모든 구간에서 개선된 음질을 보인다. 제안한 모델링 방법의 성능을 SNR, MOS값, LR(Itakura-Saito likelihood ratio), CD(cepstral distance)를 통해 보간법을 사용한 매칭 퍼슈잇과 비교 평가한다.

Abstract

In this paper, we propose the matching pursuit with damping factors, a new sinusoidal model improving the matching pursuit, for the codecs based on sinusoidal model. The proposed model defines damping factors by using a correlativity of parameters between the current and adjacent frame, and estimates sinusoidal parameters more accurately in analysis frame by using the matching pursuit according to damping factor, and synthesizes the final signal. Then it is possible to model efficiently without interpolation schemes. The proposed sinusoidal model shows a better speech quality without an additional delay than the conventional sinusoidal model with interpolation methods. Through the SNR(signal to noise ratio), the MOS(Mean Opinion Score), LR(Itakura-Saito likelihood ratio), and CD(cepstral distance), we compare the performance of our model with that of matching pursuit using interpolation methods.

**Keywords**: 정현파 모델, matching pursuit, damping 요소

## Ⅰ. 서 론

일반적인 정현파 또는 하모닉 모델은 신호를 시변하는 주파수, 진폭, 그리고 위상을 가진 정현파 성분의 선형합으로 정의한다[1]. 하지만 프레임 기반으로 처리되는 코덱에서 비트 전송률의 제약 때문에 한 프레임 구간에서 스펙트럼 크기, 주파수(또는 피치), 그리고 위상이 시간에 따라 일정하다는 가정을 전제로 하게 된다.

정현파 모델은 낮은 비트 전송율로 음성신호를 부호화하는 효율적인 기술로 알려져 왔고[2], 최근에는 음성변환[3]-[5]이나 음질 개선[6], 그리고 저전송율의 오디오 부호화[7][8]에서도 이용되고 있다. 또한 배경 잡음과 비음성 신호에 강인한 특성으로 인해 비디오 신호, 생체신호등 분석과 합성이 필요한 디지털 신호처리 분야에서 활발한 연구가 진행 중이다[9]-[12].

위에서 언급된 응용분야의 처리단계는 3단계, 즉 파라미터의 예측, 변환, 합성으로 구분할 수 있다. 첫 번째 단계는 정현파 파라미터의 예측으로 방법에 따라 크게 세 부류로 나눌 수 있다: 스펙트럼 피크 검출방법[1][3], 최소 자승법(Least Square)[7], 분석 및 합성 방법(Matching Pursuit)[8]-[11]. 오디오 코덱에서는 청각 인지 특성을 이용하는 방법을 결합하기도 한다[12]-[14]. 두 번째

* 학생회원, *** 정회원, 충북대학교 전파공학과
(Dept. of Radio Science & Engineering, Chungbuk National University)
** 정회원, 배재대학교 정보통신공학과
(Dept. of Informations and Communications Engineering, PaiChai University)
※ 본 연구는 2003년도 한국 과학재단의 특정 기초연구 사업(과제번호 R01-2003-000-11620-0)의 지원으로 수행되었습니다.
접수일자: 2006년7월13일, 수정완료일: 2006년12월29일

단계는 파라미터의 변환이다. 예측된 정현파 파라미터는 양자화하거나 수정된다. 마지막 단계에서는 두 번째 단계에서의 최종 파라미터로 신호를 합성하고, 필요에 따라 합성신호가 파라미터의 보간법이나 파형의 overlap-add를 통해 수정된다. 일반적으로 스펙트럼 크기의 선형 보간과 위상의 3차(cubic) 보간이 합성 시 함께 사용된다[1]. 이 이외에도 파라미터 분석과정에 따라 스펙트럼 크기의 지수 함수적 보간[13]이나, 위상 보간에서 2차(quadratic) 함수 보간[13][15][16]이 서로 조합되어 사용되고, 저전송률의 음성 코덱에서는 스펙트럼 크기만을 보간하여 이용하거나 피치성분에 따라 적응적으로 보간을 하기도 한다[17]-[19].

정현파 모델은 한 프레임 안에서 정현파 성분이 일정한 값을 가진다는 가정 때문에 프레임간의 불연속이 생기게 된다. 이 같은 문제를 보완하기 위해 파라미터 보간법이나 파형 보간(overlap-add)을 사용하게 되는데 이는 음성파형의 변형을 가져오게 되어 non-stationary 구간에서 파형의 왜곡이 발생한다. 특히 onset이나 offset같은 전이구간 신호에서 파형의 왜곡으로 음질저하가 뚜렷하게 나타난다. 이처럼 파라미터 보간법이 근본적인 해결책이 되지 못하기 때문에 non-stationary 구간에서는 CELP 모델 사용하는 멀티 모드 음성 코덱이나 ESM(Exponential Sinusoidal Model)과 같은 확장된 정현파 모델들이 제안되었다[20]-[22].

본 논문에서는 음성 신호의 효율적인 정현파 모델링을 위해 [23]에서 제안된 주파수 해상도가 뛰어난 매칭 퍼슈잇 방법을 개선 발전시켜 확장된 정현파 모델을 제안한다. 제안한 모델은 지수 함수적으로 증가(또는 감소)하는 정현파 성분들을 표현하기 위해 새로운 파라미터를 추가한다. 이는 시변하는 신호를 효율적으로 표현할 수 있으며 ESM과 비슷한 접근을 가진다. 하지만 보간법을 사용하는 기존의 정현파 모델과는 달리 파라미터 예측을 위해 과거 파라미터와 연관된 정보를 이용함으로 합성 시 과거 프레임과의 파라미터 보간법이나 파형의 overlap-add 방법이 필요 없게 된다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 매칭 퍼슈잇과 정현파 모델의 성능을 개선하기 위한 대표적인 파라미터 보간법들에 대해 설명하고, Ⅲ장에서는 제안하는 정현파 모델을 설명한다. Ⅳ장에서는 Ⅱ장에서 소개된 매칭 퍼슈잇과 각 파라미터 보간법과 제안한 모델과의 시뮬레이션 결과를 보이고, Ⅴ장에서 결론을 맺는다.

## Ⅱ. 매칭 퍼슈잇 방법과 파라미터 보간법

### 1. 정현파 사전을 이용한 매칭 퍼슈잇 방법[23]

정현파 모델에서 신호는 정현파 성분들의 선형합으로 다음과 같이 정의된다.

$$s[n] \approx \hat{s}[n] = \sum_{m=0}^{M} A_m^k \cdot \cos(w_m^k \cdot n + \phi_m^k) \tag{1}$$

여기서 $A_m^k$, $w_m^k$, $\phi_m^k$는 k번째 프레임에서 $m$번째 진폭, 주파수, 위상을 나타낸다.

정현파 파라미터의 예측을 위한 매칭 퍼슈잇 방법은 오류상쇄(error concealment) 원리에 바탕을 둔다. 식 (2)는 매칭 퍼슈잇의 반복과정을 위한 왜곡 측정 함수이다.

$$MSE(A_m^k, \theta_m^k) = \sum_{m=1}^{L} \left(s_{m-1}^k(n) - A_m^k \cos(w_m^k n + \phi_m^k)\right)^2 \tag{2}$$

여기서 $s_{m-1}^k(n)$ 은 $k$번째 프레임에서 $m-1$번째 반복단계의 목적신호이고, 가중치 함수 $w(n)$은 해밍윈도우를 사용한다.

$$\begin{aligned} E_m^k &= \sum_{n=0}^{N-1} [w(n)\{s_{m-1}^k(n) - A_m^k \cos(w_m^k n + \phi_m^k)\}]^2 \\ &= \sum_{n=0}^{N-1} [w(n)\{s_{m-1}^k(n) - A_m^k \cos(w_m^k n)\cos(\phi_m^k) \\ &\quad + A_m^k \sin(w_m^k n)\sin(\phi_m^k)\}]^2 \\ &\text{where} \quad s_m^k(n) = s_{m-1}^k(n) - A_m^k \cos(w_m^k n + \phi_m^k) \end{aligned} \tag{3}$$

식 (3)에서 $w_m^k$를 기본 주파수의 배수로 고정시키면, $w_m^k$은 $m \cdot w_0^k$이다. $m$은 정수값이고, 최대값은 전체 하모닉의 개수이고 피치주기를 2로 나누어 계산된다. 즉 파라미터 예측과정은 하모닉 개수만큼 반복되게 된다. 왜곡 측정 함수를 최소화하는 해를 구하면 스펙트럼 크기와 초기 위상을 아래 식과 같이 구할 수 있다[23].

$$A_m^k = \sqrt{(a_m^k)^2 + (b_m^k)^2}, \quad \phi_{ml}^k = -\tan^{-1}\left(\frac{b_m^k}{a_m^k}\right) \tag{4}$$

$$\text{where} \quad a_m^k = A_m^k \cos(\phi_m^k), \quad b_m^k = -A_m^k \sin(\phi_m^k)$$

일반적으로 파라미터 예측에 사용되는 스펙트럼 피크 검출방법보다 주파수 해상도가 높다. 매칭 퍼슈잇 방법은 윈도윙 후에 512 포인트 FFT를 취하는 방법에 비해 뛰어난 성능을 나타내고, 2048 포인트 FFT를 취하는 방법과 비슷한 해상도를 갖는다[23].

### 2. 파라미터 보간 방법

프레임간의 연결성을 유지하기 위해 정현파 파라미터들의 연결은 필수적이다. 일반적으로 음성 코딩을 위한 정현파 모델에서는 STFT(Short Time Fourier Transform) 기반의 스펙트럼 피크 검출 방법과 스펙트럼 크기의 선형 보간(linear interpolation)과 위상의 3차 보간(cubic interpolation)이 일반적으로 사용된다[1].

그림 1은 보간법을 사용할 경우 프레임의 구조를 나타낸다. 즉 보간법이 사용될 경우 분석 프레임에 비해 프레임 길이의 반이 지연된다.

스펙트럼 크기의 보간법(amplitude interpolation)은 시간 변화에 따라 피치의 변화가 크지 않다는 가정 하에 선형(linear) 함수를 이용한다. 선형 보간식은 식 (5)와 같다.

$$\widetilde{A_l^k}(n) = (1-\frac{n}{N})A_l^k + \frac{n}{N}\cdot A_l^{k+1} \qquad (5)$$

$N$은 프레임의 크기이고, $A_l^k$은 $k$번째 프레임의 $l$번째 스펙트럼 크기를 나타낸다. $\widetilde{A_l^k}$는 $k$와 $k+1$번째 프레임의 경계면을 중심으로 한 프레임의 구간에서의 보간된 스펙트럼 크기가 된다.

위상 보간(phase interpolation)은 인간의 청각 특성에 덜 민감하기 때문에 프레임의 경계 부분에서만 고려한다. 프레임 경계에서의 위상만 연결한다면 프레임내의 위상 왜곡은 들리지 않게 된다.

위상 보간은 두가지 가정을 전제로 한다[1]. 첫 번째는 프레임 경계에서 앞 뒤 프레임의 위상과 주파수가 같아야 한다는 것이다. 즉 $k$번째 프레임의 위상 $\theta_l^k(N)$과 $k+1$번째 위상 $\Theta_l^{k+1}(0)$이 같아야 하고, 주파수는 위상의 미분값과 같으므로 $w_l^k(N) = w_l^{k+1}(0)$이어야 한다. 이와 같이 함으로써 프레임 경계면에서 위상과 주파수의 불연속이 없어지게 한다. 두 번째는 합성 프레임 내에서 위상함수의 기울기가 최소가 되도록 저차 다항식을 가져야 한다. 위상 함수는 식 (6)과 같이 정의되고, 위상 모델을 위해 1차, 2차, 3차함수로 정의하고 위의 가정을 통해 위상 보간식을 유도하게 된다[24].

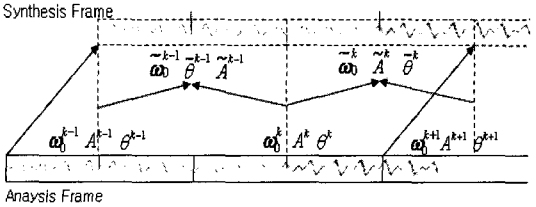



그림 1. 보간법 적용시 합성 프레임 구조
Fig. 1. Synthesis frame structure in interpolation.

$$\theta_l^k(n) = \theta_l^k(0) + \sum_{l=1}^{L} w_l^k, \quad L = harmonic\,total\,number \qquad (6)$$

● 1차(linear) 보간[1][24] :

1차 위상함수 : $\theta_l^k(n) = \zeta + \gamma \cdot n$

$$\hat{\theta}_l(n) = (1-\frac{n}{N})\theta_l^k + \frac{n}{N}\cdot\theta_{l+1}^k + \frac{2\pi M n}{N} \qquad (7)$$

● 2차(quadratic) 보간[13][15][16] :

2차 위상함수 : $\theta_l^k(n) = \zeta + \gamma \cdot n + \alpha \cdot n^2$

$$\hat{\theta}_l(n) = \theta_l^k + w_l^k \cdot n + \alpha \cdot n^2 \qquad (8)$$

$$\text{where} \qquad \alpha = \frac{w_l^{k+1} - w_l^k}{2N}$$

● 3차(cubic) 보간[1][24] :

3차 위상함수 : $\theta_l^k(n) = \zeta + \gamma \cdot n + \alpha \cdot n^2 + \beta \cdot n^3$

$$\hat{\theta}_l(n) = \theta_l^k + w_l^k \cdot n + \alpha(M) \cdot n^2 + \beta(M) \cdot n^3 \qquad (9)$$

where

$$\begin{bmatrix}\alpha(M)\\ \beta(M)\end{bmatrix} = \begin{bmatrix} 3/N^2 & -1/N \\ -2/N^3 & 1/N^2 \end{bmatrix} \cdot \begin{bmatrix} \theta_l^{k+1} - \theta_l^k - w_l^k N + 2\pi M \\ w_l^{k+1} - w_l^k \end{bmatrix}$$

$$M = e\left[\frac{1}{2\pi}\left((\theta_l^k - \theta_l^{k+1} + N w_l^k) + (w_l^{k+1} + w_l^k)\frac{N}{2}\right)\right]$$

$\theta_l^k(n)$는 $k$번째 분석 프레임에서의 위상 함수이고, $\hat{\theta}_l(n)$은 $\hat{A}_l$과 같이 $k$와 $k+1$번째 프레임의 경계면을 중심으로 한 합성프레임 구간에서의 보간된 위상이다.

식 (7)-(9)에서 $N$은 프레임의 크기이다. $\theta_l^k$, $w_l^k$는 $k$번째 프레임에서 $l$번째 위상과 주파수를 나타내고, 프레임 내에서 일정하다. $M$은 phase unwrapping integer factor이고, $e[x]$은 $x$와 가장 가까운 정수를 나타낸다.

## III. 제안하는 damping 요소를 첨가한 매칭퍼슈잇

damping 요소를 첨가한 매칭 퍼슈잇 정현파 모델은 파라미터의 예측과 합성 시의 보간법을 통합한 모델이다. 그림 2는 전체 블록도이다. damping 요소라 명명하는 2개의 전송 파라미터가 과거 정보를 이용하여 추가로 예측된다. 예측된 damping 요소들은 보간시의 오차

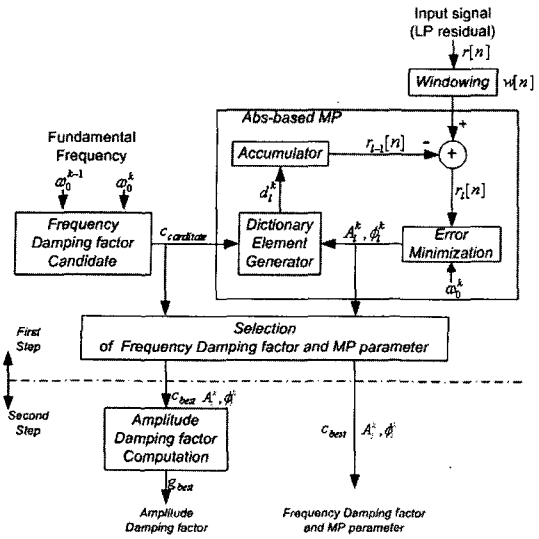



그림 2. 제안하는 정현파 모델의 블록도

Fig. 2. Block diagram of Matching pursuit with damping factor.

가 최소가 되도록 파라미터의 예측 반복과정에 응용된다. 반복되는 분석 및 합성 과정을 통해 최적의 정현파 파라미터와 damping 요소가 예측된다. damping 요소는 정현파 모델에서 합성 시 음질 개선을 위해 불가피한 보간 시의 에러를 줄이는 역할과, 기존의 정현파 모델에서 과거 파라미터와의 보간에서 생기는 지연을 없애는 역할을 한다.

현재 프레임의 파라미터에 대한 과거 프레임의 파라미터와의 비를 damping 요소라고 정의하고, 프레임간의 스펙트럼의 크기와 주파수를 다음과 같이 표현한다.

$$A_l^{k+1} = g_l^k \cdot A_l^k,\ w_l^{k+1} = c_l^k \cdot w_l^k \tag{10}$$

여기서 $A_l^k$,와 $w_l^k$는 $k$번째 프레임의 $l$번째 스펙트럼 크기와 주파수를 나타낸다. 즉, 스펙트럼 크기와 주파수에 대한 현재 프레임의 damping 요소인 적절한 $g_l^k$, $c_l^k$로 정해진다.

먼저 합성 시 사용되는 보간식을 damping 요소로 표현한다. 스펙트럼 크기의 선형 보간식은 식 (5)와 달리 합성 시 지연이 없으므로 식 (11)과 같이 정리된다.

$$\begin{aligned}\widetilde{A_l^k}(n) &= (1-\frac{n}{N})A_l^k + \frac{n}{N}\cdot A_l^{k+1} \\ &= [1+(1-g_l^k)\cdot\frac{n}{N}]\cdot A_l^k\end{aligned} \tag{11}$$

2차 위상 보간을 나타내는 식 (8)은 식 (12)와 같이 정리 된다.

$$\widetilde{\theta_l^k}(n) = \theta_l^k + w_l^k \cdot n + \alpha \cdot n^2,\ \alpha = \frac{(c_l^k-1)w_l^k}{2N} \tag{12}$$

여기서 $N$은 프레임 길이를 나타낸다.

그리고 식(2)의 오류함수를 식 (13)과 같이 수정한다.

$$\begin{aligned}&MSE(A_l^k,\theta_l^k,w_l^k,c_l^k,g_l^k) \\ &= \sum_{l=1}^{L}\sum_{n=1}^{N}\left(s_{l-1}^k(n) - A_l^k(n)\cos(w_l^k n + \phi_l^k(n))\right)^2\end{aligned} \tag{13}$$

$s^k(n)$은 $k$번째 프레임에서의 목적신호가 된다.

식 (13)에서 $c_l^k$는 현재와 과거 프레임의 기본주파수의 차와 비례하고, 현재 프레임에서의 기본주파수 성분을 조정한다. 피치를 기본으로 하모닉을 찾기 때문에 $l$개의 $c_l^k$를 $c_0^k$로 고정할 수 있다. $c_0^k$을 위해 $g_l^k$를 1로 고정하고 $c_0^k$의 후보값을 과거와 현재 프레임의 기본 주파수의 차를 기준으로 $n/2,(n=0,\pm1,\pm2)$을 곱하여 5개를 선택한다. 후보값들 중 식 (14)이 최소가 되는 $c_0^k$를 분석 및 합성과정, 즉 매칭 퍼슈잇 방법을 통해 선택한다.

$$\begin{aligned}&MSE(A_l^k,\theta_l^k,w_l^k,c_0^k,g_l^k=1) \\ &= \sum_{l=1}^{L}\sum_{n=1}^{N}\left(s_{l-1}^k(n) - A_l^k(n)\cos(w_l^k n + \phi_l^k(n))\right)^2\end{aligned} \tag{14}$$

전송률의 제약을 고려하여 $g_l^k$를 $g_0^k$로 가정하고 $g_0^k$값을 추정하면 다음과 같다. damping 요소 $g_0^k$를 하모닉과 관련되지 않도록 고정하였으므로 식 (11)의 스펙트럼 크기와 $g_l^k$와의 관련식 (11)을 식 (14)에 대입해 정리하면 식 (15)와 같이 정리된다.

$$MSE(g_0^k) \equiv \left(\sum_{n=1}^{N}\left(s^k(n) - \frac{(1-(1-g_0^k))n}{N}\hat{v}(n,c_0^k)\right)^2\right) \tag{15}$$

$$\text{where, } \hat{v}(n,c_0^k) = \sum_{l=1}^{L^k} A_l^k \cdot Re[e^{j\theta_l^k(n,c_0^k)}]$$

$g_0^k$의 최적 해는 식 (15)를 $g_0^k$에 대해 미분한 값을 0으로 놓음으로써 식 (16)을 얻는다.

$$\begin{aligned}g_0^k &= \frac{\sum_{n=1}^{N}\left(\frac{N-n}{N}\{\hat{v}(n,c_0^k)\}^2 - \frac{n}{N}s^k(n)\hat{v}(n,c_0^k))\right)}{\sum_{n=1}^{N}\left(\left(\frac{n}{N}\right)^2\{\hat{v}(n,c_0^k)\}^2\right)} \\ &= N\left(\frac{\sum_{n=1}^{N} n\cdot s^k(n)\hat{v}(n,c_0^k)}{\sum_{n=1}^{N}(n\cdot\hat{v}(n,c_0^k))^2} - \frac{\sum_{n=1}^{N} n\cdot(\hat{v}(n,c_0^k))^2}{\sum_{n=1}^{N}(n\cdot\hat{v}(n,c_0^k))^2}\right)+1\end{aligned} \tag{16}$$

최종적으로 예측된 파라미터는 스펙트럼 크기와 위상 그리고 damping 요소인 $g_0^k$, $c_0^k$가 정현파 합성식에 사용된다. 만약 $g_0^k$=1, $c_0^k$=1 일 경우 매칭 퍼슈잇을 이용한 정현파 모델로 축소된다.

## Ⅳ. 시뮬레이션 및 성능 평가

시뮬레이션을 위해 20ms 단위로 LPC 분석 후 잔여 신호를 목적 신호로 하여 정현파 파라미터를 분석한다. 분석 시 사용되는 피치주기는 정규화 된 자기 상관계수 값을 사용한 개구간(open-loop) 피치 검색 방법과 단편(fractional) 피치 검색 방법을 이용하여 구하였다.

실험에 사용된 음성은 KIST 음성 DB 중 잡음이 없는 한국어 음성 60개(남자 30개, 여자 30개)를 사용하였고, Segmental SNR과 MOS 값을 통해 비교한다. Segmental SNR은 보간 시 지연을 고려하면서 무성음 구간이 아닌 구간, 즉 유성음과 전이구간에서만을 측정하였고, MOS값은 전 구간 파형에서 ITU-T 표준 음질 소프트웨어인 "PESQ - ITU-T Recommendation P.862 Version 1.2 - 2 August 2002"로 측정하였다[25].

표 1은 분석 및 합성 방법으로 정현파 파라미터를 분석한 후 합성한 파형들의 Segmental SNR과 MOS값의 평균을 나타낸 것이다. 시뮬레이션 결과를 통해 보간법을 사용하지 않은 경우보다 보간법 사용 시 현저한 성능 개선을 확인 할 수 있다. 음성에 따라 어느 정도 편차는 보이지만 평균치를 살펴 볼 때 1차 스펙트럼 크기 보간과 3차 위상 보간을 사용시 16.125 dB의 SNR과 3.557정도의 MOS 값으로 가장 좋은 성능을 보였다. 또한 SNR과 MOS값이 절대적으로 비례하지 않는다는 사실과 SNR 증가에 비해 MOS값의 증가정도가 차이가 있다는 사실을 확인 할 수 있었다. 이는 SNR 측정 구간을 무성음이 아닌 구간, 즉 유성음과 전이구간에서 측정된 것이라 분석된다. 실험 결과 가장 좋은 성능을 보이는 보간은 1차 스펙트럼 크기 보간과 3차 위상 보간이며, 1차 스펙트럼 크기 보간과 3차 위상 보간을 이용한 합성 결과를 제안한 모델의 결과는 제안하는 damping 요소를 첨가한 매칭 퍼슈잇이 12.016 dB 개선된 SNR과 약 0.15 높은 MOS 값을 보였다.

제안한 모델의 성능을 확인하기 위해 객관적인 음질 평가 방법 인 LR(Itakura-Saito likelihood ratio)과 CD(cepstral distance)를 측정하였다[26].

LR은 식 (17)과 같이 정의되고, CD는 식 (18)과 같이 정의된다.

$$LR=\frac{\bar{a}_r R_o \bar{a}_r^T}{a_o R_o a_o^T} \tag{17}$$

$$CD=10/10\cdot\log\left[\sqrt{(c_o-c'_o)+2\sum_{i=1}^{\infty}(c_i-c'_i)^2}\right] \tag{18}$$

식에서 $\bar{a}_o$와 $\bar{a}_r$는 원 신호와 합성 신호에서의 선형 예측계수이고, $c_o$와 $c'_o$는 원신호와 합성 신호에서의 캡스트럼 계수를 나타낸다. $R_o$는 원본신호에서의 상관 행렬(correlation matrix)을 나타낸다.

표 1에서는 보듯이 제안한 모델의 LR과 CD 값들이 매칭 퍼슈잇으로 분석 후 1차 스펙트럼 크기 보간과 3차 위상 보간을 이용해 합성할 때 보다 평균적으로 낮은 값을 가지는 것을 알 수 있다.

그림 3에서 1차 스펙트럼 크기 보간과 3차 위상 보간을 이용한 합성 신호와 제안한 모델의 합성 신호를 비교한다. 사용된 신호는 무성음, 전이구간, 그리고 유성음을 포함하는 640 샘플의 LP(Linear Prediction) 잔여 신호이다. 160 샘플 단위의 프레임을 구별하였고, 각 프

표 1. 보간법에 따른 Segmental SNR과 MOS값
Table 1. Results of Segmental SNR and MOS value according to interpolation methods.

| Amplitude interpolation scheme | Phase interpolation scheme | Segmental SNR | MOS | LR | CD |
|---|---|---|---|---|---|
| constant | constant | 9.891 | 2.967 | . | . |
| | 1 order (Linear) | 9.953 | 3.231 | | |
| | 2 order (Quadratic) | 14.452 | 3.331 | | |
| | 3 order (Cubic) | 15.114 | 3.326 | | |
| 1 order (Linear) | 1 order (Linear) | 10.321 | 3.320 | | |
| | 2 order (Quadratic) | 15.125 | 3.370 | | |
| | 3 order (Cubic) | 16.125 | 3.557 | 1.63 | 3.50 |
| 제안하는 모델 (Matching Pursuit with damping factor) | | 28.089 | 3.702 | 1.31 | 3.01 |

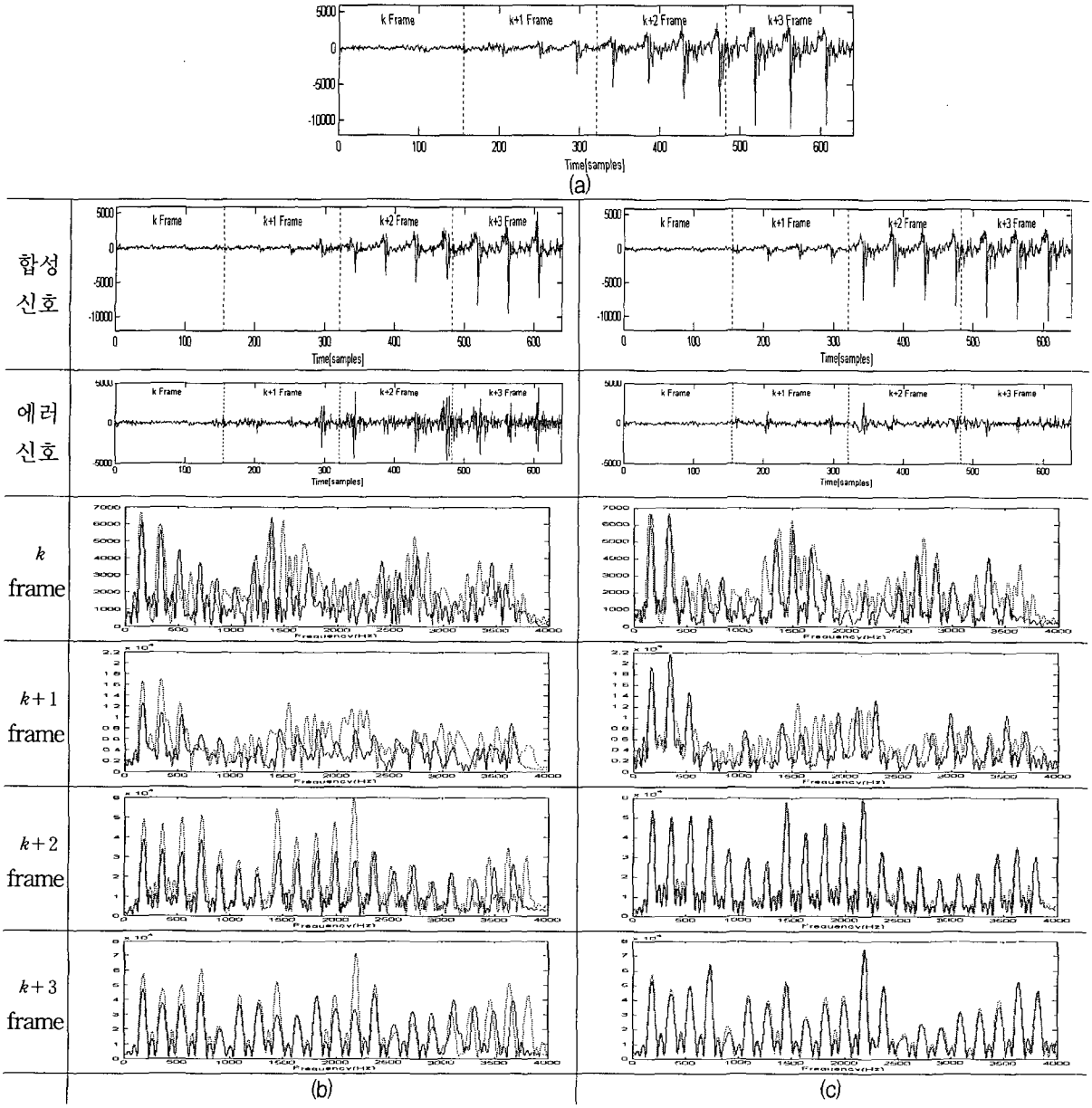



그림 3. 합성신호와 프레임에 따른 스펙트럼 결과 ; 목적 신호(점선), 합성 신호(실선), (a) 시간축에서의 목적 신호 (b) 매칭 퍼슈잇 + 3차 위상 보간 + 선형 스펙트럼 크기 보간 ($s[n+80]$), (c) damping 정현파 매칭 퍼슈잇 ($s[n]$)

Fig. 3. Synthesized signal and spectrum results over the frame ; target signal (dashed line), synthesized signal (solid line), (a) target signal in time domain, (b) matching pursuit + cubic phase interpolation + linear amplitude interpolation ($s[n+80]$), (c) matching pursuit with damping factor ($s[n]$).

레임에서 시간축의 신호와 스펙트럼의 모양을 비교해 수 있다. 그림 3의 (a)의 원본 신호과 그림 3의 (b)와 (c)의 합성 신호의 비교를 위해 시간축 오차신호를 나타내었다. 오차신호의 스케일 범위 만으로도 개선된 성능을 확인할 수 있고, 각 프레임별 스펙트럼의 비교에서도 눈에 띄게 개선된 성능을 확인 할 수 있다. 특히 그림 3의 (b)에서 유성음에 해당하는 $k+2$, $k+3$ 번째 프레임을 보면 스펙트럼 크기의 포락선은 비슷하나 스케일이 다른 결과를 보인다. 그림 3의 (c)에서 보듯이 제안한 모델의 damping 요소($g_0^k$)는 이런 스케일 보상함으로써 매칭 퍼슈잇 방법을 개선시킨다.

결과적으로 매칭 퍼슈잇으로 분석한 파라미터들은 스펙트럼 크기와 위상의 보간으로 인해 음질의 향상을 가져오기는 하지만 시간축 신호나 주파수축 신호에서의 왜곡이 발생한다. 전이구간 같은 에너지 폭이 크거나 기본 주파수 외의 하모닉들이 다수 존재하는 구간에서

는 그 왜곡이 심하게 나타난다. 본 논문에서 제안한 damping factor의 사용은 보간이 필요하지 않도록 분석할 때 보간에 대한 영향을 damping factor로 따로 추출하고, 합성 시에 과거 파라미터와의 보간을 사용하지 않고 damping factor로 보간시의 영향분을 보완해줌으로써 시간축에서 파형과 주파수축의 스펙트럼을 원신호에 맞게 유지 시킨다. 이는 음질의 향상 뿐 아니라 보간에 따른 지연도 발생하지 않게 된다.

## V. 결 론

본 논문에서는 정현파 파라미터 분석과 파라미터 보간 방법을 통합한 damping 요소를 가진 매칭 퍼슈잇 정현파 모델을 제안하였다. 제안하는 모델은 과거 정현파 파라미터와의 상관성을 damping 요소로 정의하고 현재 프레임에 최적의 값을 분석과 합성 구조로 예측해 냄으로서 합성 시 추가적인 지연이 전혀 발생하지 않으면서 보간 시 생기는 파형 왜곡을 최소화 하여 음질을 개선하는 모델이다.

제안한 모델의 성능 비교를 위해 매칭 퍼슈잇 방법으로 파라미터를 분석하고 합성 시 보간(스펙트럼 크기 선형 보간, 3차 위상보간)하는 정현파 모델과 비교를 하였다. 본 논문에서 제안된 현재 파라미터와의 상관성을 이용한 damping 요소를 첨가한 매칭 퍼슈잇은 1차 선형 보간과 2차 위상을 바탕으로 했지만, 결과적으로 1차 스펙트럼 크기 보간과 3차 위상 보간을 사용한 매칭 퍼슈잇보다 약 12 dB정도의 SNR 증가와 0.15정도의 MOS값 증가로 개선된 음질을 보인다. 현재의 모델에서는 두개의 damping 요소를 단계적으로 최적의 파라미터로 예측해낸다. 하지만 복잡도의 감소나 damping 요소의 효율적인 검출방법이 보완된다면 향후 정현파 모델링을 기반으로 한 음성/오디오 코덱이나 디지털 신호의 모델링이 필요한 분야에서 응용이 가능할 것이다.

## 참 고 문 헌


[1] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," IEEE Trans. on ASSP, vol. 34, no. 4, pp. 744 - 754, Aug. 1986.

[2] W. B. Kleijin and K. K. Paliwal, Speech coding and synthesis, Elevier Science Publishers, Amsterdam, 1995.

[3] T. F. Quatieri and R. J. McAulay, "Speech transformations based on a sinusoidal representation," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 1449-1464, 1986.

[4] E. B. George and M. J. T. Smith, "Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model," IEEE Trans. Speech Audio Processing, vol. 5, no. 5, pp. 389-406, 1997.

[5] Y. Stylianou, "Applying the harmonic plus noise model in concatenative speech synthesis," IEEE Trans. Speech Audio Processing, vol. 9, pp. 232-239, Mar. 2001.

[6] J. Jensen and J. H. L. Hansen, "Speech enhancement using a constrained iterative sinusoidal model," IEEE Trans. Speech Audio Processing, vol. 9, pp. 731-740, Oct. 2001.

[7] J. Nieuwenhuijse, R. Heusdens, and E.F. Deprettere, "Robust exponential modeling of audio signals," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '98, Seattle, Washington, USA, vol. 6, pp. 3581 - 3584, May 1998.

[8] T. S. Verma and T. H. Y. Meng, "Sinusoidal modeling using frame-based perceptually weighted matching pursuits," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '99, Phoenix, Arizona, USA, vol. 2, pp. 981 - 984, May 1999.

[9] Yuan Yuan and D. M. Monro, "Improved Matching Pursuits Image Coding," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '05, vol. 2, pp. 201-204, Mar. 2005.

[10] K. Skretting, K. Engan and J.H. Husoy, "ECG compression using signal dependent frames and matching pursuit," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '05, vol. 4, pp. 585-588, Mar. 2005.

[11] P. Vera-Candeas and N. Ruiz-Reyes, "New matching pursuit based sinusoidal modelling method for audio coding," IEE Proceedings on Vision, Image and Signal Processing, vol. 151, pp. 21-28, Feb. 2004.

[12] T. Painter and A. Spanias, "Perceptual segmentation and component selection in compact sinusoidal representations of audio," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '01, vol. 5, pp. 3289 - 3292, May 2001.

[13] X. Serra and J. Smith, "Spectral modeling synthesis: A sound analysis/synthesis system

based on a deterministic plus stochastic decomposition," Computer Music journal, vol. 14, pp. 12-24, Dec. 1990.

[14] T. S Verma and T. H. Y. Meng, " Sinusoidal Modeling Using Frame-Based Perceptually Weighted Matching Pursuit," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '99, vol. 2, pp. 981-984, 1999.

[15] Y. Ding and X. Qian, "Estimating sinusoidal parameters of musical tones based on global waveform fitting," Multimedia Signal Processing, pp. 95 - 100, Jun. 1997.

[16] I. Atkinson, S. Yeldner and A. Kondoz, "High quality split band LPC vocoder operating at low bit rates," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '97, vol. 2, pp. 1559 - 1562, Apr. 1997.

[17] T. F. Quatieri and R. J. McAulay, "Phase modelling and its application to sinusoidal transform coding," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '86, vol. 3, pp. 1713-1715, Apr. 1986.

[18] R. J. McAulay and T. F. Quatieri, "Magnitude-only reconstruction using a sinusoidal speech model," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '84, vol. 2, pp. 27.6.1-27.6.4, Mar. 1984.

[19] R. J. McAulay and T. F. Quatieri, "Computationally ecient sine-wave synthesis and its application to sinusoidal transform coding," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '88, vol. 1, pp. 370-373, Apr. 1988.

[20] J. Nieuwenhuijse, R. Heusdens, and E. F. Deprettere, "Robust exponential modeling of audio signals," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '98, vol. 2, pp. 3581-3584, Mar. 1998

[21] J. Jensen, S. H. Jensen, and E. Hansen, "Exponential sinusoidal modeling of transitional speech segments," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '99, pp. 473-476, 1999.

[22] K. Hermus, W. Verhelst, and P. Wambacq, "Psycho-acoustic modeling of audio with exponentially damped sinusoids," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '02, pp. 1821-1824, 2002.

[23] 안 영욱, 정 규혁, 김 종학, 양 용호, 이 인성, "정현파 모델 부호화기를 위한 MP(Matching Pursuit)알고리즘과 파라미터 양자화기," 음향학회지 제 24권 제 7호, pp. 402~409, 2005.

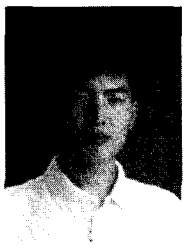
[24] L. Girin, S. Marchand, J. di Martino, A. Robel and G. Peeters, "Comparing the order of a polynomial phase model for the synthesis of quasi-harmonic audio signals," Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on. 19-22 pp. 193-196, Oct. 2003.

[25] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ) : An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codec", Feb. 2001.

[26] S. Wang, A. Sekey and A. Gersho, "An objective measure for predicting subjective quality of speech coders," Selected Areas in Communications, IEEE Journal on vol. 10, pp. 819 - 823, Jun. 1992.

## 저 자 소 개

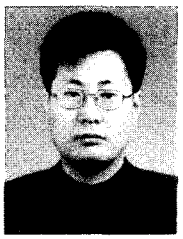

정 규 혁(학생회원)
2004년 2월 충북대학교 전기전자공학 (공학사)
2006년 2월 충북대학교 전파공학과 (공학석사)
2006년 3월 ~현재 충북대학교 전파공학과 (공학박사)
<주관심분야 : 음성/오디오 부호화, 통신신호처리, VoIP>



임 정 우(학생회원)
2005년 2월 충북대학교 전기전자공학 (공학사)
2005년 3월~현재 충북대학교 전파공학과 (공학석사)
<주관심분야 :암묵신호처리. 음성/오디오 부호화, 적응필터>

이 인 성(정회원)
1983년 2월 연세대학교 전자공학과 (공학사)
1985년 2월 연세대학교 전자공학과 (공학석사)
1992년 2월 Texas A&M University 전기공학과 (공학박사)
1993년 2월~1995년 9월 한국전자 통신연구원 이동통신 기술연구단 선임연구원
1995년10월~현재 충북대학교 전기전자공학부 정교수
<주관심분야 : 음성/영상 신호 압축, 이동통신, 적응필터>

김 종 학(학생회원)
1999년 2월 충북대학교 전자공학과 (공학사)
2000년 2월 충북대학교 전파공학과 (공학석사)
2000년 3월 ~현재 충북대학교 전파공학과 (공학박사)
<주관심분야 : 음성/오디오 부호화, 영상압축, 적응필터>



주 기 호(정회원)
1984년 2월 고려대학교 전기공학과 (공학사)
1985년 2월 고려대학교 전기공학과 (공학석사)
1992년 2월 Texas A&M University 전기공학과 (공학박사)
1995년~현재 배재대학교 정보통신공학과 부교수
<주관심분야: 임베디드통신시스템, 멀티미디어통신>