

논문 2007-44SP-4-14

정현과 음성-오디오 모델의 빠른 하모닉 합성 방법

(Fast Harmonic Synthesis Method for Sinusoidal Speech-Audio Model)

김 규 진*, 김 종 학*, 정 규 혁*, 이 인 성**

(Gyu-Jin Kim, Jong-Hark Kim, Gyu-Hyeok Jung, and In-Sung Lee)

요 약

대부분의 2차 및 3차 위상 보간을 사용하는 하모닉 합성 방법은 각각의 정현과 성분에 대해 샘플단위로 합성되기 때문에 구현하는데 있어 많은 연산량이 요구된다. 본 논문에서는 이러한 문제를 해결하기 위해 2차 및 3차 위상 항을 가지는 정현과 음성 및 오디오 모델을 위한 빠른 하모닉 합성 방법을 제안한다. 제안하는 빠른 하모닉 합성 방법은 2차 및 3차 위상함수의 계수를 하모닉과 독립적으로 강요함으로써 오버 샘플링 함수와 위상 변조 함수를 정의하고, Inverse Fast Fourier Transform(IFFT)을 이용한 합성식을 유도한다. 제안한 빠른 하모닉 합성 방법은 연산량과 Segment SNR(Segment Signal-to-Noise Ratio)을 코사인 함수를 이용한 합성 방법과의 비교를 통해 음질의 저하없이 연산량이 현저히 줄어드는 것을 확인할 수 있었다.

Abstract

Most harmonic synthesis methods using phase information employ a quadratic or cubic phase interpolation. The methods are computationally expensive to implement because every component sinewave must be synthesized on a per sample basis. In this paper, we propose a fast harmonic synthesis method for sinusoidal speech/audio coding based on the quadratic and cubic phase function to overcome the complexity problem. To derive the fast harmonic synthesis method, we define the over-sampling function and phase modulation function by constraining the parameter of phase function to be independent for harmonic index and derive the fast synthesis method using IFFT. Experimental results show that the proposed method significantly reduce the complexity of conventional cosine synthesis method while maintaining the performance.

Keywords: 정현과 모델, cubic phase, interpolation, inverse fast fourier transform(FFT)

I. 서 론

STC(Sinusoidal Transform Coding)^[1~3], MBE (Multi-Band Excitation)^[4~5], HVXC(Harmonic Vector Excitation Coding)^[6], MELP(Mixed Excitation Linear Prediction)^[7~8]과 같은 하모닉 음성 부호화기들은 4 kbps 이하에서도 좋은 음질을 만들고 있다. 또한, 이러

한 정현과 모델은 MPEG-4 HILN(Harmonic Individual Line)^[6]과 같은 저 전송률을 갖는 파라메트릭 오디오 부호화기에서도 쓰이고 있고, 음성 및 오디오의 통합 부호화를 위한 유니버설 코더 연구 그룹인 SiCAS (Sinusoidal Coding for Audio and Speech), ARDOR (Adaptive Rate-Distortion Optimized sound codeR)에서도 활발히 연구되고 있다.

대부분의 하모닉 부호화기에서는 합성방법으로서 코사인 함수를 사용한 중첩 및 합산 방법을 사용하고 있다. 이러한 합성방법은 구현에 있어 간단하지만, 각각의 정현과 성분에 대해 샘플단위로 합성되기 때문에 많은 연산량을 요구하고 있다. 이러한 문제를 해결하기 위해 MPEG-4 HVXC는 IFFT를 이용한 합성법을 사용하고

* 학생회원, ** 정회원, 충북대학교 전자공학과
(Department of Radio Engineering,
Chungbuk National University)

※ “이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임” (지방연구중심대학육성사업/충북BIT연구중심대학육성사업단)

접수일자: 2007년3월13일, 수정완료일: 2007년6월8일

있다^{[6][11]}. 또한, 최근의 정현파 코더에 대한 많은 연구들이 분석 합성 방법에 기초하기 때문에 원본 신호의 위상 정보가 유지 되어야 한다^{[3][9-10]}. 하지만, HVXC에서는 기본 주파수가 선형보간 되었다는 전제로 위상정보를 이전 프레임에서 구해서 사용하고, 선형위상을 사용하고 있다^{[6][11]}.

본 논문에서는 2차 및 3차 위상 항을 가지는 정현파 코더에서의 IFFT를 이용한 빠른 하모닉 합성 방법을 제안하고, 원본 정현파 합성모델로부터 IFFT를 이용한 빠른 하모닉 합성 수식을 유도하고, 연산량과 Segment SNR을 비교함으로써 그 성능을 검증한다.

II. 2차 및 3차 위상 항을 가지는 정현파 모델

정현파 모델은 낮은 비트 전송률로 음성신호를 부호화하는 효율적인 기술로 알려져 왔고, 최근에는 음성 변환이나 음질 개선, 그리고 저 전송률의 오디오 부호화에서도 이용되고 있다. 일반적인 정현파 또는 하모닉 모델은 신호를 시변하는 주파수, 진폭, 그리고 위상을 가진 정현파 성분의 선형합으로 다음과 같이 정의된다^[1-3].

$$\hat{s}^k(n) = \sum_{l=1}^{L_s(n)} A_l^k \cdot \cos(\theta_l^k(n)), n=0, \dots, N-1 \quad (1)$$

여기서 A_l^k , θ_l^k 는 k 번째 프레임에서 l 번째 하모닉 성분의 스펙트럼 크기와 위상을 나타낸다.

프레임간의 연속성을 유지하기 위해 정현파 파라미터들의 연결은 필수적이다. 일반적으로 음성 코딩을 위한 정현파 모델에서는 스펙트럼 크기의 선형 보간(linear interpolation)과 3차 위상 보간(cubic interpolation)이 일반적으로 사용된다^[1-3]. 스펙트럼 크기의 선형 보간은 다음과 같다.

$$\hat{A}_l(n) = (1-\lambda(n)) \cdot A_l^k + \lambda(n) \cdot A_l^{k+1} \quad (2)$$

where, $\lambda(n) = n/N$

여기서 N 은 프레임의 크기이고, A_l^k 는 k 번째 프레임에서 l 번째 하모닉의 스펙트럼 크기를 나타낸다. \hat{A}_l 는 k 와 $k+1$ 번째 프레임 경계면을 중심으로 한 구간에서의 보간된 스펙트럼 크기이다.

합성 시 위상 보간(phase interpolation)은 두 가지 가정을 전제로 한다^[1-3]. 첫 번째는 프레임의 경계에서 위

상과 주파수가 같아야 한다는 것이다. 즉, k 번째 프레임의 위상 $\theta_l^k(N)$ 과 $k+1$ 번째 프레임의 위상 $\theta_l^{k+1}(0)$ 이 같아야 하고, 위상의 미분 값인 주파수 $\omega_l^k(N)$ 과 $\omega_l^{k+1}(0)$ 도 같아야 한다는 것을 의미하므로 프레임 경계면에서 위상과 주파수의 불연속이 없도록 한다. 두 번째는 3차 위상에서 합성 프레임 내에서 위상함수의 기울기가 최소가 되도록 최대 평탄함(maximally smooth)을 가져야 한다는 것이다. 위의 가정을 통해 2차 및 3차 위상 보간식은 다음과 같이 유도된다.

2차 위상 보간(quadratic phase interpolation) :

$$\hat{\theta}_l(n) = \omega_l^k \cdot n + \alpha_l \cdot n^2 + \phi_l^k \quad (3)$$

$$\text{where, } \alpha_l = \frac{\omega_l^{k+1} - \omega_l^k}{2 \cdot N}$$

3차 위상 보간(cubic phase interpolation) :

$$\hat{\theta}_l(n) = \omega_l^k \cdot n + \alpha_l \cdot n^2 + \beta_l \cdot n^3 + \phi_l^k \quad (4)$$

where,

$$\begin{bmatrix} \alpha_l \\ \beta_l \end{bmatrix} = \begin{bmatrix} 3/N^2 & -1/N \\ -2/N^3 & 1/N^2 \end{bmatrix} \begin{bmatrix} \phi_l^{k+1} - \phi_l^k - \omega_l^k \cdot N + 2\pi M \\ \omega_l^{k+1} - \omega_l^k \end{bmatrix}$$

$$m = \frac{-1}{2\pi} (\phi_l^{k+1} - \phi_l^k - \omega_l^k \cdot N) + \frac{N}{2} (\omega_l^{k+1} - \omega_l^k)$$

$$M = \text{round}(m)$$

여기서 ϕ_l^k , ω_l^k 는 k 번째 프레임에서의 l 번째 위상과 주파수를 나타내고, 프레임 내에서 일정하다. 위상은 $-\pi$ 와 π 사이로 제한되므로 프레임 사이의 연속성을 최대화하기 위해서는 시간에 따른 위상 항의 변화율이 작도록 정수 M 값이 조정되어야한다. $\hat{\theta}_l$ 는 \hat{A}_l 와 마찬가지로 k 와 $k+1$ 번째 프레임 경계면을 중심으로 한 구간에서의 보간된 위상이다.

III. IFFT를 이용한 빠른 하모닉 합성

1. 2차 위상 항을 가지는 하모닉 합성

제한된 하모닉 합성 방법은 기존의 하모닉 합성 방법인 코사인 함수를 사용한 중첩 및 합산 방법에서 요구되는 많은 연산량을 IFFT를 이용하여 합성함으로써 줄이는 방법이다. 2차 위상 항을 가지는 정현파 모델의 경우 식 (1)을 중첩 및 합산 방법으로 나타내면, 다음과 같다.

$$\begin{aligned} \hat{s}(n) &= \lambda_1(n) \cdot \sum_{l=1}^{L^k} A_l^k \cdot \cos(\theta_l^k(n)) \\ &+ \lambda_2(n) \cdot \sum_{l=1}^{L^k} A_l^{k+1} \cdot \cos(\theta_l^{k+1}(n)) \\ &= \text{Re} \left\{ \lambda_1(n) \cdot \sum_{l=1}^{L^k} A_l^k \cdot e^{j\theta_l^k(n)} \right\} \\ &+ \text{Re} \left\{ \lambda_2(n) \cdot \sum_{l=1}^{L^k} A_l^{k+1} \cdot e^{j\theta_l^{k+1}(n)} \right\} \end{aligned} \quad (5)$$

where, $\lambda_1(n) = 1 - n/N$, $\lambda_2(n) = n/N$

$$\theta_l^k(n) = \omega_l^k \cdot n + \alpha_l \cdot n^2 + \phi_l^k$$

$$\theta_l^{k+1}(n) = \omega_l^{k+1} \cdot n + \alpha_l \cdot n^2 + \phi_l^{k+1}$$

식 (5)에서 IDFT를 이용한 합성식을 유도하기 위해서는 ω_l 과 α_l 이 하모닉에 독립이 되어야 한다. 하모닉이 기본 주파수의 배수에 위치한다고 가정하면, 2차 위상 함수에서 주파수 ω_l^k 와 2차 항의 계수 α_l 는 식 (6)과 같이 표현될 수 있고, 이를 통해 새로운 오버 샘플링 함수 $f^k(n)$ 과 $f^{k+1}(n)$ 을 식 (7)과 같이 정의함으로써 IDFT를 이용한 방법으로 표현이 가능해진다.

$$\omega_l^k = \omega_0^k \cdot l, \quad \alpha_l = \frac{(\omega_0^{k+1} - \omega_0^k) \cdot l}{2 \cdot N} \quad (6)$$

$$f^k(n) = \frac{(\omega_0^k \cdot n + \alpha \cdot n^2) \cdot P}{2\pi} \quad (7)$$

$$f^{k+1}(n) = \frac{(\omega_0^{k+1} \cdot n + \alpha \cdot n^2) \cdot P}{2\pi}$$

여기서 P 는 IDFT 크기이다.

식 (6), (7)을 식 (5)에 대입하여 정리하면, 식 (8)과 같이 스케일이 조정된 IDFT를 이용한 합성식을 유도할 수 있다.

$$\hat{s}(n) = \text{Re} \left(\begin{aligned} &\lambda_1(n) \sum_{l=0}^{P-1} A_l^k e^{j\phi_l^k} e^{j \frac{2\pi f^k(n)l}{P}} \\ &+ \lambda_2(n) \sum_{l=0}^{P-1} A_l^{k+1} e^{j\phi_l^{k+1}} e^{j \frac{2\pi f^{k+1}(n)l}{P}} \end{aligned} \right) \quad (8)$$

$$= \text{Re} \left(\begin{aligned} &\lambda_1(n) F_{IDFT} \left(A_l^k e^{j\phi_l^k}, f^k(n) \right) \\ &+ \lambda_2(n) F_{IDFT} \left(A_l^{k+1} e^{j\phi_l^{k+1}}, f^{k+1}(n) \right) \end{aligned} \right)$$

$$\begin{aligned} \text{where, } F_{IDFT}(X, n) &= \sum_{l=0}^{P-1} X \cdot e^{j \frac{2\pi nl}{P}} \\ &= P \left(\frac{1}{P} \sum_{l=0}^{P-1} X e^{j \frac{2\pi nl}{P}} \right) \end{aligned}$$

여기서 IDFT를 IFFT로 대체함으로써 연산량을 감소시킬 수 있다.

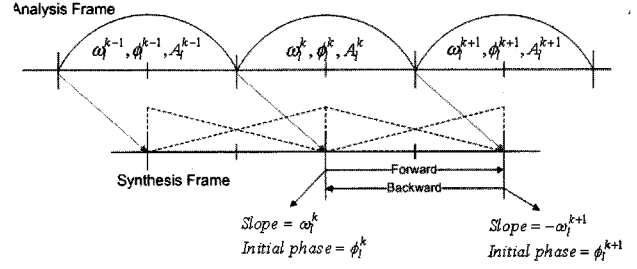


그림 1. 보간법 적용시 합성 프레임 구조
Fig. 1. Synthesis frame structure in interpolation.

2. 3차 위상 항을 가지는 하모닉 합성

3차 위상 항의 경우 3차 위상함수의 3차 항의 파라미터 β_l 을 l 에 독립적으로 강요함으로써 IDFT를 이용한 합성 수식을 유도할 수 있다. 여기서 β_l 을 강요할 때 어디를 중심으로 하는가는 그림 1과 같이 보간법이 적용된 합성 프레임 구조에서 보듯이 프레임의 시작점과 끝점을 중심으로 강요하여 보간하는 방법을 생각할 수 있다. 먼저, 합성 프레임의 시작점을 중심으로 강요된 3차 항을 갖는 3차 위상함수를 강요된 전향 3차 위상 함수(forward constrained cubic phase; FCCP)라 정의하면, 다음과 같다.

$$\theta_l^{fcp}(n) = \omega_l^k \cdot n + \alpha_l^{fcp} \cdot n^2 + \beta_l^{fcp} \cdot n^3 + \phi_l^k \quad (9)$$

여기서 β_l^{fcp} 는 하모닉과 독립적으로 강요된 파라미터이고, α_l^{fcp} 은 β_l^{fcp} 값이 적용된 경계조건에 따라 새로이 구해진 파라미터이다. 강요된 3차 항의 파라미터 β_l^{fcp} 는 3차 위상함수를 이용해 합성한 신호와 강요된 β_l^{fcp} 값을 이용해 합성한 신호의 오류함수로 정의할 수 있고, 위상함수의 오차가 작다면, 1차 테일러 급수에 의해 식 (10)과 같이 위상함수의 오류함수로 근사화하여 구할 수 있다^[12].

$$\begin{aligned} E_l(n) &= \sum_{n=0}^{N-1} |\hat{s}(n) - s_{fcp}(n)|^2 \\ &\cong \sum_{n=0}^{N-1} \sum_{l=1}^{L^k} W_l^2 (\theta_l(n) - \theta_l^{fcp}(n)) \end{aligned} \quad (10)$$

여기서 W_l 은 가중치 값으로 $W_l = (A_l^k + A_l^{k+1})/2$ 와 같이 k 번째와 $k+1$ 번째 프레임 사이의 하모닉 크기값으로 구해지고, $E_l(n)$ 를 최소화시키기 위한 β_l^{fcp} 의 해는 β_l^{fcp} 에 대한 편미분 값이 0이 되는 방정식의 해가 된다. 여기서 β_l^{fcp} 의 계산을 위해 α_l 에 대한 프레임 경계조건을 다음과 같이 정의할 수 있다.

$$\alpha_l = \frac{(\omega_l^{k+1} - \omega_l^k) \cdot l}{2N} - \frac{3\beta_l \cdot N}{2} \quad (11)$$

식 (10)의 오류함수를 3차 위상함수와 강요된 전향 3차 위상함수 (9)를 대입하여 β^{fcp} 에 대한 편미분 방정식으로 나타내고, α_i 에 대한 프레임 경계조건을 적용하여 β^{fcp} 에 대하여 정리하면 식 (12)와 같다.

$$\beta^{fcp} = \frac{\sum_{l=1}^{L^k} (W_l^2 \cdot \beta_l)}{\sum_{l=1}^{L^k} W_l^2} \quad (12)$$

위상함수의 3차 항의 계수가 하모닉에 독립적으로 강요됨에 따라 2차 항의 계수 α_i 도 경계조건에 맞도록 다시 계산되어야 한다. 여기서 α_i^{fcp} 의 계산은 경계조건에서 β_i 이 β^{fcp} 로 대체되어 구할 수도 있지만, 추가 연산량의 감소를 위해 식 (13)과 같이 이미 구해진 값인 α_i 로부터 구해질 수 있다.

$$\alpha_i^{fcp} \equiv \alpha_i + \frac{3N(\beta_i - \beta^{fcp})}{2} \quad (13)$$

IFFT를 이용한 빠른 합성법을 유도하기 위해 2차 위상함수에서와 마찬가지로 하모닉이 기본 주파수의 배수에 위치한다고 가정하면, IFFT 크기 P 를 갖는 오버 샘플링 함수 $f_1(n)$ 과 위상 변조함수 $f_2(n)$ 을 식 (14), (15)와 같이 정의한다.

$$f_1^{fcp}(n) = \frac{P}{2\pi} \left(\omega_0^k n + \frac{(\omega_0^{k+1} - \omega_0^k)n^2}{2N} \right) \quad (14)$$

$$f_2^{fcp}(n) = e^{\frac{-3\beta^{fcp}N}{2}n^2 + \beta^{fcp}n^3} \quad (15)$$

앞에서 구해진 오버 샘플링 함수와 위상 변조함수를 이용해 중첩 및 합산 방법의 구조를 가지는 IDFT를 이용한 합성 방법으로 나타내면, 다음과 같다.

$$\begin{aligned} s_{fcp}(n) &= \text{Re} \left(f_2^{fcp}(n) \left(\lambda_1(n) \sum_{l=0}^{P-1} A_l^k e^{j\phi_l^k} e^{j \frac{2\pi f_1^{fcp}(n)l}{P}} \right. \right. \\ &\quad \left. \left. + \lambda_2(n) \sum_{l=0}^{P-1} A_l^{k+1} e^{j\phi_l^{k+1}} e^{j \frac{2\pi f_1^{fcp}(n)l}{P}} \right) \right) \\ &= \text{Re} \left(f_2^{fcp}(n) \left(\lambda_1(n) F_{IDFT} \left(A_l^k e^{j\phi_l^k}, f_1^{fcp}(n) \right) \right. \right. \\ &\quad \left. \left. + \lambda_2(n) F_{IDFT} \left(A_l^{k+1} e^{j\phi_l^{k+1}}, f_1^{fcp}(n) \right) \right) \right) \quad (16) \end{aligned}$$

where, $\lambda_1(n) = 1 - n/N$, $\lambda_2(n) = n/N$

$$\begin{aligned} F_{IDFT}(X, n) &= \sum_{l=0}^{P-1} X e^{j \frac{2\pi nl}{P}} \\ &= P \left(\frac{1}{P} \sum_{l=0}^{P-1} X e^{j \frac{2\pi nl}{P}} \right) \end{aligned}$$

여기서 IDFT를 IFFT로 대체함으로써 연산량을 현저

히 줄일 수 있지만, 강요된 전향 3차 위상함수는 합성 프레임의 시작점을 중심으로 구해진 것이기 때문에 그림 2와 같이 합성 프레임의 끝부분으로 갈수록 왜곡이 증가하는 현상을 보인다. 이러한 왜곡을 보상하기 위해 강요된 전향 3차 위상함수와 반대의 왜곡 특성을 갖는 합성 프레임의 끝을 중심으로 구해진 강요된 후향 3차 위상함수(backward constrained cubic phase; BCCP)를 다음과 같이 정의할 수 있다.

$$\theta_i^{bcp}(n) = -\omega_i^{k+1}n + \alpha_i^{bcp}n^2 + \beta_i^{bcp}n^3 + \phi_i^{k+1} \quad (17)$$

먼저, 강요된 후향 3차 위상함수의 계산을 위해 프레임의 끝을 중심으로 하는 3차 위상함수 $\theta_i^{end}(n)$ 를 정의하면 다음과 같다.

$$\theta_i^{end}(n) = -\omega_i^{k+1}n + \alpha_i^{end}n^2 + \beta_i^{end}n^3 + \phi_i^{k+1} \quad (18)$$

where,

$$\begin{bmatrix} \alpha_i^{end} \\ \beta_i^{end} \end{bmatrix} = \begin{bmatrix} 3/N^2 & -1/N \\ -2/N^3 & 1/N^2 \end{bmatrix} \begin{bmatrix} \phi_i^k - \phi_i^{k+1} + \omega_i^{k+1}N + 2\pi\hat{M} \\ -\omega_i^k + \omega_i^{k+1} \end{bmatrix}$$

$$\hat{m} = \frac{-1}{2\pi} (\phi_i^k - \phi_i^{k+1} + \omega_i^{k+1}N) + \frac{N}{2} (-\omega_i^k + \omega_i^{k+1})$$

$$\hat{M} = \text{round}(\hat{m})$$

여기서 2차 및 3차 항의 파라미터 α_i^{end} 과 β_i^{end} 의 계산은 식 (4)의 3차 위상함수에서 주파수와 위상을 프레

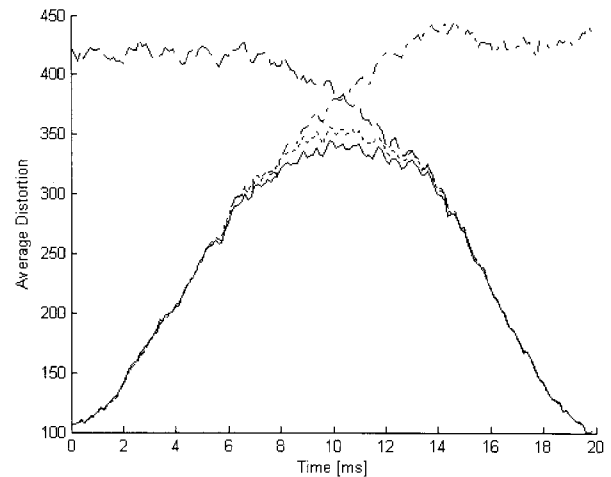


그림 2. 합성 프레임에서의 평균 왜곡; 3차 위상함수(점선), 강요된 전향 3차 위상함수(1점 쇄선), 강요된 후향 3차 위상함수(파선), 강요된 전-후향 3차 위상함수(실선)

Fig. 2. Average distortion at the span of synthesis frames; cubic phase(dotted), forward constrained cubic phase(dashed dot), backward constrained cubic phase(dashed), forward-backward constrained cubic phase(solid).

임의 끝을 중심으로 한 주파수와 위상으로 대체됨으로써 구해질 수 있다($\omega_i^k = -\omega_i^{k+1}$, $\omega_i^{k+1} = -\omega_i^k$, $\phi_i^k = \phi_i^{k+1}$, $\phi_i^{k+1} = \phi_i^k$). α_i^{end} , β_i^{end} 과 \widehat{M} 의 계산은 상당히 많은 추가 연산량을 발생시키지만, 식 (19)과 같이 α_i , β_i 과 M 으로부터 구함으로써 추가 연산량을 감소시킬 수 있다.

$$\alpha_i^{end} = -\alpha_i + \frac{\omega_i^{k+1} - \omega_i^k}{N} \quad (19)$$

$$\beta_i^{end} = -\beta_i, \widehat{M} = -M$$

합성 프레임의 끝을 중심으로 강요된 3차 항의 파라미터 β^{bccp} 는 전향 파라미터와 같이 프레임의 끝을 중심으로 구해진 위상함수 $\theta_i^{end}(n)$ 과 강요된 후향 3차 위상함수 $\theta_i^{bccp}(n)$ 의 오류함수로부터 구해지고, α_i^{bccp} 는 강요된 3차 항의 파라미터 β^{bccp} 과 프레임 경계조건으로부터 구할 수 있지만, α_i 로부터 구함으로써 적은 연산량으로 구해진다.

$$\beta^{bccp} = \frac{\sum_{l=1}^L (W_l^2 \beta_l^{end})}{\sum_{l=1}^L W_l^2} \quad (20)$$

$$\alpha_i^{bccp} = \alpha_i + \frac{3N(\beta_i^{end} - \beta^{bccp})}{2}$$

강요된 후향 3차 위상함수를 IDFT를 이용한 합성법을 유도하기 위해 IDFT 크기 P 를 갖는 오버 샘플링 함수와 위상 변조 함수를 정의하면 다음과 같다.

$$f_1^{bccp}(n) = \frac{P}{2\pi} \left(-\omega_0^{k+1}n + \frac{(\omega_0^{k+1} - \omega_0^k)n^2}{2N} \right) \quad (21)$$

$$f_2^{bccp}(n) = e^{-\frac{3\beta^{bccp}N}{2}n^2 + \beta^{bccp}n^3}$$

여기서 f_1^{bccp} 와 f_2^{bccp} 는 강요된 전향 3차 위상함수의 오버 샘플링 함수와 위상 변조 함수로부터 구함으로써 연산량을 감소시킬 수 있다.

$$f_1^{bccp}(n) = f_1^{fccp}(n) - \frac{P}{2\pi} (\omega_0^k + \omega_0^{k+1})n \quad (22)$$

$$f_2^{bccp}(n) = \frac{1}{f_2^{fccp}(n)}$$

강요된 후향 3차 위상함수에 대한 중첩 및 합산 구조의 IDFT를 이용한 합성 방법으로 나타내면, 다음과 같다.

$$s_{bccp}(n) = \text{Re} \left(f_2^{bccp}(n) \left(\lambda_1(n) F_{IDFT} \left(A_i^{k+1} e^{j\phi_i^{k+1}}, f_1^{bccp}(n) \right) + \lambda_2(n) F_{IDFT} \left(A_i^k e^{j\phi_i^k}, f_1^{bccp}(n) \right) \right) \right) \quad (23)$$

중첩 및 합산 구조의 강요된 3차 위상 항을 가지는 하모닉의 합성은 강요된 전향 3차 위상함수와 강요된 후향 3차 위상함수를 이용해 합성한 파형의 선형 보간을 통해 이루어진다.

$$s_{fbccp}(n) = \lambda_1(n) s_{fccp}(n) + \lambda_2(n) s_{bccp}(n) \quad (24)$$

IV. 시뮬레이션 및 성능 평가

표 1은 DSP에 사용되는 기본 연산인 덧셈, 곱셈, 곱셈-덧셈 누산기(multiply and add)의 연산 복잡도 가중치를 1로 하여 구한 연산량이다. 여기서 L 은 하모닉 개수, N 은 프레임 길이, P 은 IDFT 크기이고, Exp 연산은 코사인파와 사인 Table로 처리하였고 가중치는 1로 두었다. IDFT를 이용한 합성법의 성능 평가를 위해 기존의 코사인 합성법과 제안된 방법의 연산량을 비교하여 측정하였다. 그림 3은 제안된 방법이 기존의 코사인을 이용한 합성법에 비해 현저히 적은 연산량을 갖는다는 것을 보여준다. 그림 3의 (a)는 제안된 방법이 IDFT 크기에 크게 영향을 받지 않는다는 것을 나타낸다. 그림 3의 (b), (c)와 같이 하모닉 개수와 프레임 크기가 증가하면, 코사인 합성법과 제안된 방법 모두 연산량이 증가하지만, 제안된 방법이 코사인 합성법에 비해 영향을 적게 받는다. 또한, 하모닉 개수나 프레임 크기가 증가할수록 연산량의 감소 효과가 크다는 것을 알 수 있다.

제안된 모델의 성능을 확인하기 위해 주관적 음질 테스트와 Segment SNR을 측정하였다. 실험은 음성의 경우 NTT 음성 DB 중 한국어 남·여 각각 32문장과 영어 남·여 각각 32문장에 대해 HVXC(Harmonic Vector eXcitation Coding)에서 측정된 피치와 유무성 음 정보를 사용하여, 전이구간을 포함한 유성음 구간에

표 1. 합성위상에 따른 추정 연산량
Table 1. Approximated complexity for synthetic phase.

Synthesis method		Operations
Quadratic phase	Conventional	$L(12N+2) + 3N$
	Proposed (fast)	$4L + P \log_2 P + 14N$
Cubic phase	Conventional	$L(13N+16)$
	Proposed (fast)	$43L + 2P \log_2 P + 31N$

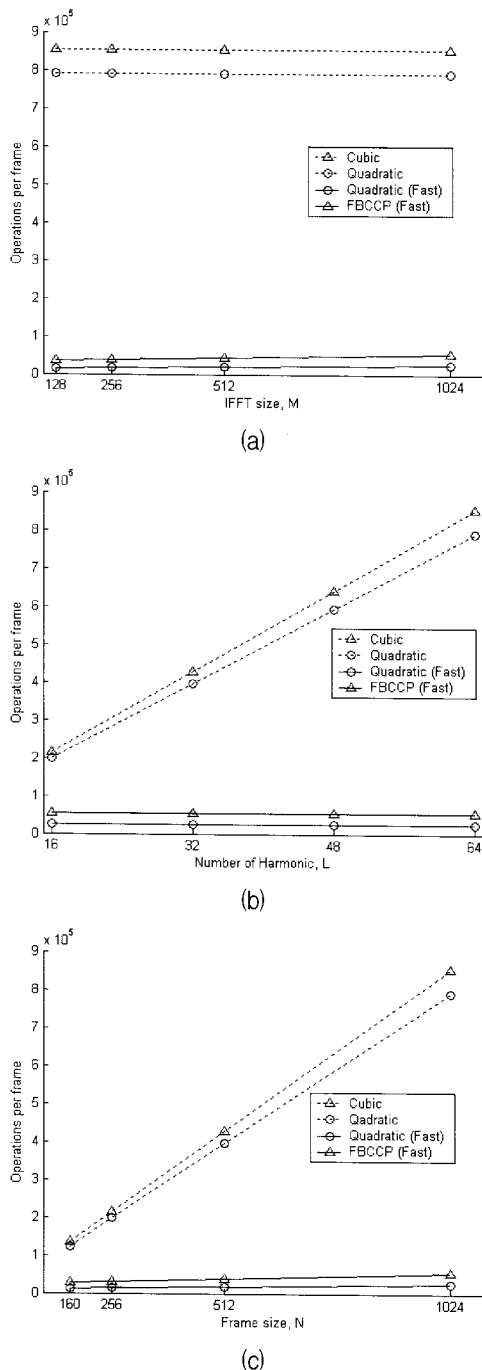


그림 3. 음성 및 오디오 신호에서의 계산량.
 (a) IFFT 크기에 따른 계산량 ($L=64, N=1024$),
 (b) 하모닉 개수에 따른 계산량 ($N=1024, P=1024$), (c) 프레임 크기에 따른 계산량 ($L=64, P=1024$)
 Fig. 3. Complexity of speech/audio signals.
 (a) Complexity for IFFT size ($L=64, N=1024$),
 (b) Complexity for number of harmonics ($N=1024, P=1024$), (c) Complexity for frame size ($L=64, P=1024$).

서의 Segment SNR을 계산하였고, 오디오는 MPEG-4 SQAM(Sound Quality Assessment Material) 16개에

표 2. 음성 및 오디오 신호에 대한 Segment SNR
 Table 2. Segment SNR for Speech and Audio signal.

		Segment SNR (dB)			
		Speech (English : Man)			
IFFT size		128	256	512	1024
quadratic phase		13.90			
cubic phase		14.28			
quadratic phase(fast)		13.67	13.73	13.74	13.75
FBCCP(fast)		14.28	14.34	14.35	14.36
		Speech (English : Woman)			
IFFT size		128	256	512	1024
quadratic phase		15.61			
cubic phase		16.05			
quadratic phase(fast)		15.43	15.45	15.46	15.46
FBCCP(fast)		16.15	16.17	16.18	16.18
		Speech (Korea : Man)			
IFFT size		128	256	512	1024
quadratic phase		14.82			
cubic phase		15.25			
quadratic phase(fast)		14.60	14.67	14.69	14.69
FBCCP(fast)		15.11	15.20	15.21	15.22
		Speech (Korea : Woman)			
IFFT size		128	256	512	1024
quadratic phase		17.12			
cubic phase		17.40			
quadratic phase(fast)		16.98	17.00	17.01	17.01
FBCCP(fast)		17.40	17.42	17.42	17.42
		Audio			
IFFT size		128	256	512	1024
quadratic phase		11.09			
cubic phase		11.30			
quadratic phase(fast)		11.04	11.08	11.08	11.09
FBCCP(fast)		11.31	11.36	11.36	11.37

대해 HILN(Harmonic Individual Line plus Noise)에서 측정된 피치와 하모닉 사용 정보를 이용하여, 하모닉 성분이 강한 부분에서의 Segment SNR을 계산하였다. 2차 위상의 경우 IFFT 크기가 증가할수록 제안된 합성법의 SNR이 코사인 합성법의 SNR에 수렴하는데, 이것은 제안된 방법이 위상함수로부터 정의된 오버샘플링함수로 IFFT를 통해 만들어진 한 피치주기의 시간영역 파형에 대해 오버샘플링하는 방법이기 때문에 IFFT 크기가 커질수록 SNR 측면에서 좋은 성능을 보인다. 3차 위상은 IFFT 크기가 증가할수록 코사인 합성법보다 약간 좋은 성능을 보였다. 이것은 제안된 방법이 기존의 중첩 및 합산 방법과 달리 프레임의 시작과 끝을 중심으로 합성된 파형을 중첩 및 합산 연산을 수행함으로써 그림 2와 같이 합성 프레임 내에서 왜곡을 줄였기 때문이다. 이 경우 연산량과 성능을 고려한다면, IFFT 크기는 256-point가 적당하다는 것을 알 수 있다. 256-point IFFT로 합성한 그림 4를 보면 제안된 2차 및 3차 합성

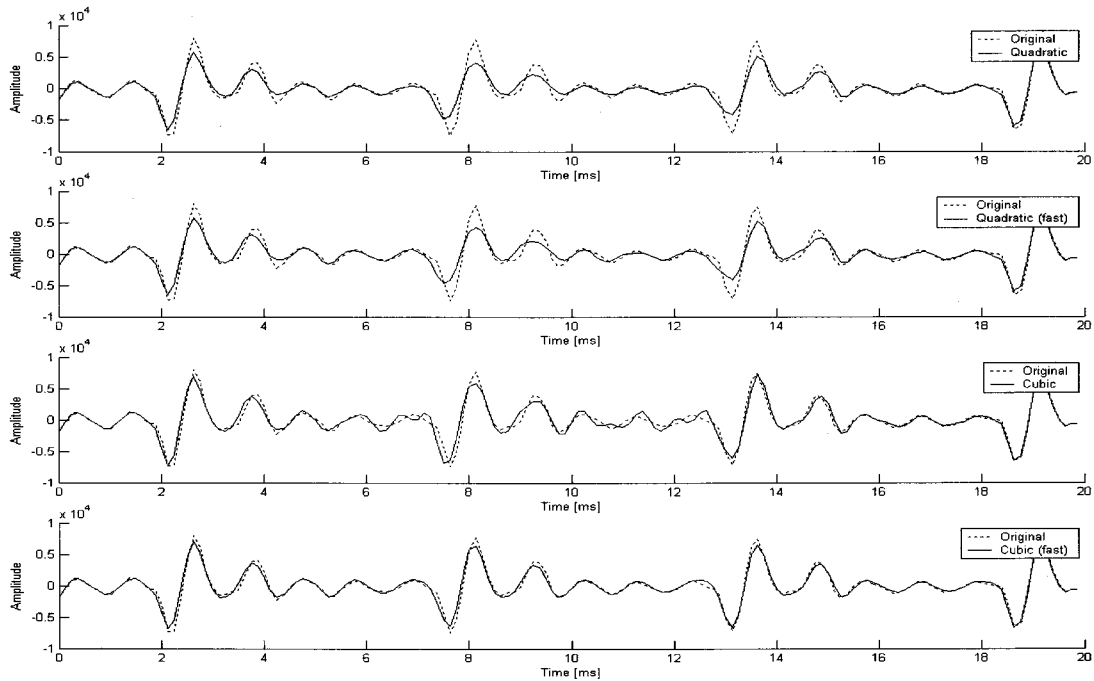


그림 4. 음성 신호에 대한 합성 파형
Fig. 4. Synthesis signal for speech.

표 3. 제안된 합성법에 대한 선호도
Table 3. Preference of proposed synthesis method.

	quadratic		cubic	
	conventional	fast	conventional	fast
man	49.43 %	50.57 %	52.65 %	47.35 %
woman	49.81 %	50.19 %	51.71 %	48.29 %
Total	49.62 %	50.38 %	52.18 %	47.82 %

법 모두 코사인을 이용해 합성한 신호와 유사하다.

주관적 음질테스트는 NTT 음성 여자 24문장과 남자 24문장을 사용하였고, 테스트에 참여한 인원은 12명이다. 테스트 방법은 잡음이 없는 동일한 음성을 각각의 합성법으로 합성하여 약 2초간의 간격을 두고 들려준 후 더 듣기 좋은 음성을 고르는 A/B 테스트 방법을 사용하였다. 음질 테스트 결과에서 보여주듯이 제안된 합성법은 2차 위상에서 기존의 코사인 합성법과 유사한 선호도를 보였지만, 3차 위상에서는 약간 낮은 선호도를 보였다. 이것은 제안된 합성법의 유도과정에서 위상 함수의 3차 항의 파라미터 β_1 이 제한되었기 때문에 근사화 오차에 의한 것으로 보인다.

결과적으로 제안된 방법은 코사인을 이용한 합성법에 비해 1/16~1/5의 적은 연산량으로 코사인 합성법과 유사한 성능을 보인다.

V. 결 론

본 논문에서는 2차 및 3차 위상 항을 가지는 음성 및 오디오 정현파 모델에서의 빠른 하모닉 합성 방법을 제안하였다. 제안하는 모델은 2차 위상함수에서 오버 샘플링 함수를 정의하고, 3차 위상함수에서는 3차 항의 파라미터 β_1 을 하모닉과 독립적으로 강요함으로써 오버 샘플링 함수와 위상 변조 함수를 정의하여 크기 P 를 갖는 IFFT를 이용한 빠른 하모닉 합성법이다.

제안된 방법의 성능을 비교하기 위해 DSP에 사용되는 기본 연산인 덧셈, 곱셈, 곱셈-덧셈 누산기(multiply and add)의 연산 복잡도 가중치를 1로 하여 구한 연산량과 Segment SNR을 측정하였다. 제안된 IFFT를 이용한 빠른 하모닉 합성법은 기존의 코사인 합성법과 유사한 음질을 유지하면서 1/16~1/5의 연산량을 가진다. 3차 위상의 정확도 개선이 더 이루어진다면 정현파 모델링을 기반으로 한 음성 및 오디오 부호화기 및 구현에 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. Kleijn and K.K Paliwal, "Speech coding and synthesis", ELSEVIER, chapter 4, 1995.

[3] 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

[4] D.W.Griffm, "Multiband Excitation Vocoder", Ph.D. dissertation, M.I.T., Cambridge, MA, 1987.

[5] D.W.Griffm and J.S.Lim, "Multiband Excitation Vocoder", IEEE Trans. on Acoustics, Speech, and Signal Processing, pp1223-1235, 1988.

[6] ISO/IEC 14496-3, "Information Technology - Coding of Audio Visual Object, Part 3 : Audio, Subpart 2 : Parametric Coding", ISO/IEC International Standard, 2000.

[7] A. V. McCree and T. P. Barnwell 111, "Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding," IEEE Trans. on Speech and Audio Processing, vol. 3, pp. 242-250, July 1995.

[8] A. McCree, K. Truong, E. B. George, T. P. Barnwell 111, and V. Viswanathan, "A 2.4 kbit/s MELP Coder Candidate for the New U.S. Federal Standard," in Proc. IEEE Int. Con\$ ASSP, (Atlanta), pp. 200-203, May 1996.

[9] 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.

[10] 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.

[11] Masayuki Nishiguchi, "Harmonic vector excitation coding of speech", Acoustical Science and Technology, Vol. 27, No.6 pp.375-383, 2006.

[12] David L. Thomson, "Parametric Models of the Magnitude/Phase Spectrum for Harmonic Speech Coding," ICASSP 1988, pp.378-381, 1988

저 자 소 개



김 규 진(학생회원)
 2006년 2월 충북대학교
 전자공학과 (공학사).
 2006년 3월~현재 충북대학교
 전파공학과 (공학석사)
 <주관심분야 : 음성/오디오 부호
 화, 통신신호처리, 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, 적
 응필터>



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



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, 적
 응필터>