

위상 변환 인자가 적용된 음성의 중첩합산 정현파 합성 방법

*박종배, 김종학, 김규진, 양용호, 이인성
충북대학교 전자공학과

e-mail : tmplover@nate.com, inslee@chungbuk.ac.kr

Overlap and Add Sinusoidal Synthesis Method of Speech Signal Using Phase Shaping Factor

*Jong-Bae Park, Jong-Hark Kim, Kyu-Jin Kim, Yang Yong Ho, In-Sung Lee
Dept. of Radio Engineering, Chungbuk National University

Abstract

In this paper, we propose a new method for overlap and add synthesis using phase shaping factor in a sinusoidal synthesis method of speech signal, which improves continuity and SNR(Signal Noise Ratio) efficiency of synthesized speech.

I. 서론

일반적인 정현파 모델은 시간에 따라 변화하는 주파수, 진폭, 그리고 위상을 가진 정현파 성분의 선형 합으로 정의한다[1]. 정현파 모델은 한 프레임 안에서 정현파 파라미터가 일정한 값을 가진다는 기본 가정 때문에 프레임간의 불연속성이 생기게 된다. 일반적으로 이러한 프레임 불연속을 줄이기 위해 스펙트럼 크기의 선형 보간과 위상 보간이 합성 시 함께 사용된다[1]. 본 논문에서는 기존의 합성신호들 간의 오차를 최소화하는 방법으로 정현파 파라미터를 추정하는 합성법에서 위상변환 인자를 적용하여 합성성능을 개선한다.

II. 정현파 크기로 가중치 된 선형 위상 중첩합산 정현파 합성법

프레임간의 연속성을 유지하기 위해 정현파 파라미터들의 연결은 필수적이다. 3차 위상 보간[1]에 비해 적

은 복잡도를 가지면서 프레임간의 연속성을 유지하기 위한 방법으로 정현파 크기로 가중치 된 선형 위상 중첩합산(WE-OLA-LP: weighted - overlap and add - linear phase) 정현파 합성법이 제안되었다[2].

WE-OLA-LP 정현파 합성법에서는 중첩되는 경계면에서의 최종 합성된 신호들 간의 차이를 최소화 하는 방법을 사용하여 보간 하려는 지점의 정현파 파라미터를 구함으로써 프레임간의 불연속성을 최소화한다.

WE-OLA-LP의 파라미터는 정현파 크기가 가중치된 위상오차 함수식으로부터 유도된다[2]. 그림 1과 같이 N/2지점의 $\bar{\theta}$ 는 두 지점 N/4지점과 3N/4지점에서의 MSE(mean-square error)를 최소화 하는 값으로 구해진다. 여기서, N/4, 3N/4 지점에 대한 에러 $\epsilon_{N/4}(\bar{\theta})$, $\epsilon_{3N/4}(\bar{\theta}, M)$, $l(\bar{\theta}, M)$ 값은 다음과 같다[2].

$$\begin{aligned} \epsilon_{N/4}(\bar{\theta}) &= W_{1/4} \left(\theta_k + w_k \cdot \frac{N}{4} \right) + \bar{w} \cdot \frac{N}{4} - \bar{\theta} \\ \epsilon_{3N/4}(\bar{\theta}, M) &= W_{3/4} \left(\theta_{k+1} + 2\pi M - w_{k+1} \cdot \frac{N}{4} \right) - \bar{w} \cdot \frac{N}{4} - \bar{\theta} \\ l(\bar{\theta}, M) &= \epsilon_{N/4}^2(\bar{\theta}) + \epsilon_{3N/4}^2(\bar{\theta}, M) \end{aligned} \quad (1)$$

여기서 $W_{1/4}$ 와 $W_{3/4}$ 는 각 지점에서의 크기의 평균값이다. 위상 $\bar{\theta}$ 와 위상 비구속 정수 인자 M은 식 (1)을 최소화 하도록 다음과 같이 구해질 수 있다[2].

$$\begin{aligned} \bar{\theta} &= \frac{W_{1/4}A + W_{3/4}B}{W_{1/4} + W_{3/4}} \quad \text{where } A = \theta_k + \frac{3w_k + w_{k+1}}{2} \cdot \frac{N}{4} \\ B &= \theta_{k+1} + 2\pi M - \frac{w_k + 3w_{k+1}}{2} \cdot \frac{N}{4} \end{aligned} \quad (2)$$

V. 결론

본 논문에서는 WE-OLA-LP 정현파 합성법에서 얻은 보간 지점의 정현파의 위상 파라미터 값에서 위상 변환 인자를 적용하여 기존의 WE-OLA-LP보다 합성 성능을 개선하였다.

“이 논문 또는 저서는 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임”(지방연구중심대학육성사업/충북BIT연구중심대학육성사업단)

참고문헌

- [1] R. J. McAulay and T. F. Quatieri, "Computationally efficient sine-wave synthesis and its application to sinusoidal Transform coding," .
- [2] 박종배, 김규진, 김종학, 정규혁, 이인성 “가중치 된 위상 오류함수를 사용한 음성의 중첩합산 정현파 합성방법”, 제 17회 통신정보 합동학술대회 2007.
- [3] Goodwin, M.: "Matching Pursuit with Damped Sinusoids", Proc. *IEEE ICASSP* 1997, vol.3, pp.2037-204

$$M = \text{round} \left(\frac{\frac{W_{1/4}A + W_{3/4}C}{W_{1/4} + W_{3/4}} - \theta_{k+1} + w_{k+1} \cdot \frac{N}{4} + \bar{w} \cdot \frac{N}{4}}{\left(2\pi - \frac{W_{3/4} \cdot 2\pi}{W_{1/4} + W_{3/4}} \right)} \right) \quad (3)$$

최종 합성과형은 위의 $\bar{\theta}$ 와 M 파라미터를 사용하여 N/2 WE-OLA-LP정현파 합성법으로 합성된다[2].

III. 위상변환 인자가 적용된 음성의 중첩합산 정현파 합성 방법

WE-OLA-LP 정현파 합성법에서 구한 보간 지점의 정현파 파라미터 중 위상 변환 인자를 적용하여 위상 값을 변화시켜 원 신호의 위상 값과 유사하게 함으로써 음성신호의 합성성능을 개선한다. MP(matching pursuit)[3]를 통해서 얻은 중간 지점의 위상 파라미터 θ_{opt} 와 WE-OLA-LP를 통해 얻은 N/2지점의 위상 $\bar{\theta}$ 를 이용하여 위상 변환 인자를 β 를 구한다. 다음과 같이 에러 E가 β 값을 구하기 위해 사용된다..

$$E = [\theta_{opt,\ell} - (\bar{\theta}_\ell + \beta_\ell * 2\pi)]^2$$

$$\left(\begin{array}{l} \theta_{opt,\ell} : \ell\text{th optimal phase} \\ \bar{\theta}_\ell : N/2 \text{ rate } \ell\text{th phase (WE-OLA-LP)} \\ \beta_\ell : \ell\text{th phase shapping factor} \end{array} \right) \quad (4)$$

식(4)를 최소화 하는 β 값으로부터 다음과 같이 $\theta_{sf,\ell}$ 를 새로 정의한다.

$$\theta_{sf,\ell} = \bar{\theta}_\ell + \beta_\ell * 2\pi \quad (5)$$

WE-OLA-LP 정현파 합성과정에서 ℓ 번째 하모닉 N/2지점 위상 값을 $\theta_{sf,\ell}$ 으로 사용함으로써 원본 위상에 더욱 근접한 합성을 얻을 수 있다. 본 논문에 사용된 β 값의 적정 양자화 레벨은 실험적으로 16단계가 적절하여 이를 이용하였다.

IV. 시뮬레이션

식(5)에서 얻은 변화된 위상 값으로 N/2지점에서의 새롭게 합성된 것과 기존의 WE-OLA-LP의 합성과형으로 원본 파형과의 SNR을 구한 시뮬레이션은 그림 2와 같다. NTT 남, 여 각각 32문장씩이 사용되었으며, 8kHz데이터에 160샘플 분석 프레임에 갖는다. 그림에서 보논바와 같이 프레임 N/2지점에서의 SNR 성능이 약 3dB정도 향상되었음을 볼 수 있다. 전체적으로는 평균 SNR값이 15.51dB에서 16.62dB으로 향상됨을 확인할 수 있었다.

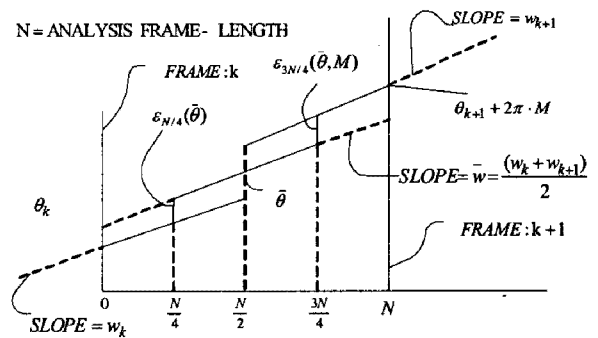


그림 1. N/2지점의 프레임이 첨가된 위상 궤적

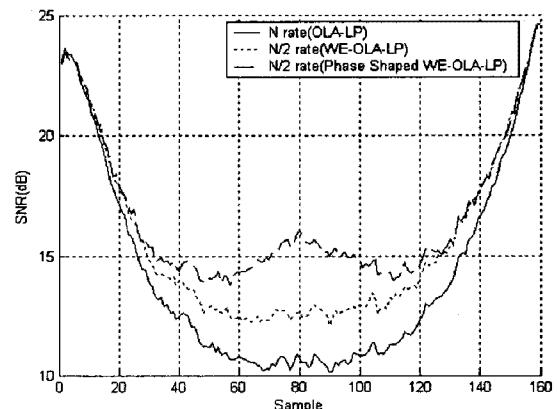


그림 2. 합성법에 따른 각 샘플위치의 평균 SNR 값