웨이블릿 변환을 이용한 음성신호의 AbS/OLA 정현파 모델링

김기홍*, 홍진근**, 정용익*, 이상이*
*국가보안기술연구소
**천안대학교
e-mail:hong0612@etri.re.kr

AbS/OLA Sinusoidal Modeling of Speech Signal Using Wavelet Transform

Ki-Hong Kim*, Jin-Keun Hong**, Yong-Ik Jung*, Sang-Yi Lee*
*National Security Research Institute
**Chonan University

요 약

본 논문에서는 합성에 의한 분석(Analysis-by-Synthesis) 및 가산중첩(Overlap-Add) 방식을 채택하고 있는 음성신호의 AbS/OLA 정현파 모델에 웨이블릿 변환을 적용한 새로운 모델을 제안하였다. 즉, 기존의 모델에 웨이블릿 변환을 적용하여 입력신호를 몇 개의 부대역 신호로 나눈 다음 각각 다른 길이의 분석 윈도우를 적용한다. 이는 기존 모델의 정현파 파라미터 추출 시 고정된 길이의 분석 윈도우를 이용하는 단점을 극복하여 좀 더 정확한 파라미터 추출을 가능하게 한다.

시험결과 제안된 정현파 모델이 기존 모델에 비해 합성음의 스펙트럼 및 위상 특성, 음질 등에서 성능이 개선됨을 확인할 수 있었다.

1. 서론

정현파 모델은 유·무성음의 여기신호를 각기 다른 크기, 주파수, 그리고 위상을 가지는 정현파들의 합으로 음성신호를 합성하는 방식이다. 즉, 유성음인 경우에는 주파수 영역의 스펙트럼 상에서 기본주파수를 포함하여 각 고조파 성분들을 그 주파수에 해당하는 정현파로 표현하고, 무성음인 경우에는 피크치들을 찾아 그에 해당하는 정현파로 생성한다[1-3]. 이 때 크기, 주파수, 위상 등의 파라미터를 추출함에 있어 크게 DFT(Discrete Fourier Transform) 방식[1,2]과 AbS(Analysis-by-Synthesis) 방식[3]이 있으며 AbS 방식이 DFT 방식에 비해 좀 더 나은 성능을 가지고 있다. 본 논문에서는 정현파 파라미터를 추출함에 있어 합성에 의한 분석 및 가산중첩 방식을 채택하고 있는 AbS/OLA 정현파 모델을 이용하였다. 그러나, 기존 모델은 음성신호 분석 시 피치 주기의 약 2~2.5배 정도의 고정된 길이의 분석 윈도우를 이용하는 단점을 가진다.

본 논문에서는 이러한 기존 모델의 단점을 극복하고 좀 더 정확한 정현파 파라미터 추출을 위해 기존 AbS/OLA 정현파 모델에 웨이블릿[4,5] 변환을 적용하였다. 즉, 파라미터 추출 시 입력신호를 몇 개의 부대역 신호로 나눈 다음 각기 다른 길이의 분석 윈도우를 적용하는 새로운 모델을 제안하여, 기존 모델과 합성음의 파형, 스펙트럼 및 위상, 그리고 음질을 비교·분석하고자 하였다.

2. AbS/OLA 정현파 모델

AbS/OLA 정현파 모델은 DFT를 이용하는 모델에 비해 계산량 감소 및 천이 구간, 무성음 구간에서 좀 더 정확한 파라미터를 추출할 수 있다[3]. 이는 합성에 의한 분석 과정을 통하여 입력신호와 추정된 합성신호와의 에러를 최소화하는 정현파 파라미터를 추출한다. AbS/OLA 정현파 모델에서 추정된 신호 $\tilde{s}(n)$은 식 (1)과 같이 표시된다.

$$\tilde{s}(n) = \sigma(n) \sum_{k=-\infty}^{\infty} w_s(n-kN_s)\, \tilde{s}^k(n-kN_s) \quad (1)$$

여기서, $\sigma(n)$은 입력신호의 에너지 포락선을, $w_s(n)$은 합성 윈도우 함수를 나타낸다.

입력신호의 각 분석 프레임의 합성신호 $\tilde{s}^k(n)$은 식 (2)와 같이 표시된다.

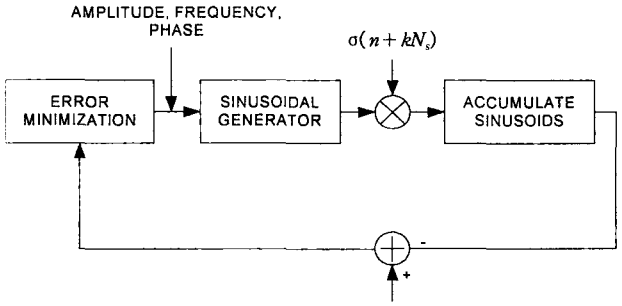
$$\tilde{s}^k(n) = \sum_{i=1}^{I^k} A_i^k \cos(\omega_i^k n + \phi_i^k) \qquad (2)$$

여기서, A_i^k, ω_i^k 그리고 ϕ_i^k 는 정현파의 크기, 주파수, 그리고 위상 정보이다. 이들 각각의 정현파 파라미터는 합성에 의한 분석 방식을 통하여 반복적으로 추출되며 식 (3)에 주어진 입력신호와 추정된 합성신호와의 에러를 최소화하는 파라미터가 최종적인 파라미터가 된다.

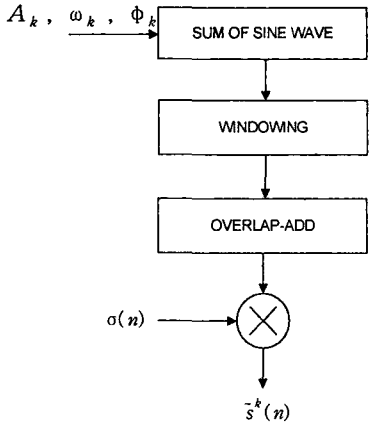
$$E = \sum_{n=-\infty}^{\infty} (s(n) - \tilde{s}(n))^2 \qquad (3)$$

$$\tilde{s}(n+kN_s) = \sigma(n+kN_s)(w_s(n)\,\tilde{s}^k(n) + w_s(n-N_s)\,\tilde{s}^{k+1}(n-N_s)),\ 0 \le n < N_s$$

그림 1은 AbS/OLA 정현파 모델의 분석·합성 과정을 나타낸다.



(a) 분석 과정



(b) 합성 과정

그림 1. AbS/OLA 정현파 모델의 분석·합성 과정

분석 과정에서는 입력신호와 합성신호와의 에러를 최소화하는 파라미터를 추출하고, 합성 과정에서는 분석 과정에서 추출한 파라미터를 이용하여 합성신호를 생성한다.

3. AbS/OLA 정현파 모델

AbS/OLA 정현파 모델에서 분석 윈도우의 길이는 피치 주기의 약 2~2.5배 정도이다. 이는 신호의 압축효율 면에서 단점을 가지며 또한, 정현파 모델링 시 pre-echo 등의 문제를 발생시킨다[6]. 이런 고정된 길이의 분석 윈도우를 이용함으로써 발생되는 문제들을 극복하고 무성음 구간이나 천이 구간에서 좀 더 정확한 파라미터 추출을 위해, 본 논문에서는 기존 모델에 웨이블릿 변환을 적용한 새로운 모델을 제안하였다. 즉, 입력신호를 웨이블릿 변환을 이용하여 몇 개의 부대역 신호로 나눈 다음 저주파 신호인 경우에는 주파수 해상도가 시간 해상도보다 높아야 하므로 길이가 긴 분석 윈도우를 적용하고, 고주파 신호인 경우에는 시간 해상도가 주파수 해상도보다 높아야 하므로 길이가 짧은 분석 윈도우를 적용하였다.

그림 2는 본 논문에서 제안한 모델의 구성도를 보여준다. 그림에서 보는 것처럼, 필터뱅크를 통과한 입력 신호의 부대역 신호를 각각 독립적으로 AbS/OLA 정현파 모델을 이용하여 분석·합성한다. 이 때 신호의 급격한 변화를 잘 반영하는 주파수가 상대적으로 높은 대역의 신호는 짧은 길이의 분석 윈도우를, 신호의 전체적인 변화를 잘 반영하는 주파수가 상대적으로 낮은 대역의 신호는 긴 길이의 분석 윈도우를 적용하여 최적의 파라미터를 한다.

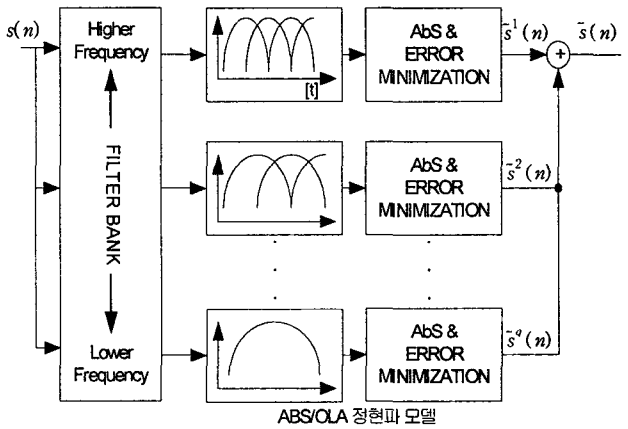
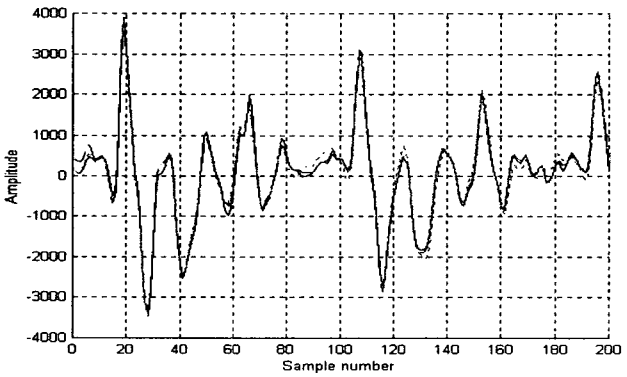


그림2. 제안한 AbS/OLA 정현파 모델의 분석·합성 과정

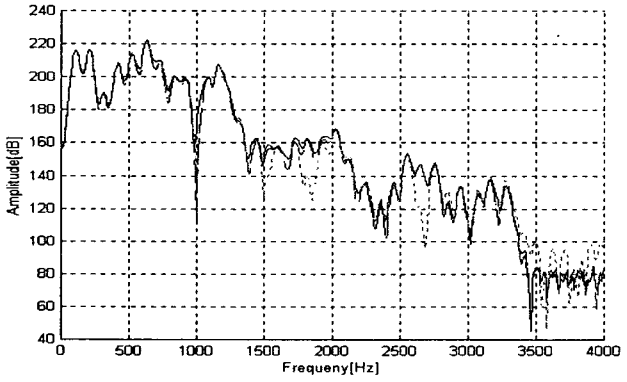
4. 실험 및 검토

본 논문에서는 8KHz로 샘플링 되고, 16비트로 양자화된 임의의 음성데이터를 이용하여 원 음성신호와 기존 모델을 이용하여 생성한 합성신호 및 제안한 모델을 이용하여 생성한 합성신호의 파형, 스펙트럼 및 위상을 비교·분석하였다. 또한 음질평가를 위해서 객관적 음질평가 테스트인 PESQ(Perceptual Evaluation of Speech Quality)[7]를 이용하여 기존 모델과 제안된 모델의 음질을 비교·평가하였다. 기존 모델의 분석 과정에서는 20ms의 고정된 길이를 가지는 해밍 윈도우 함수를 이용하였다. 한편, 입력신호의 웨이블릿 변환을 위해서 Daubechies 10-탭 필터를 이용하여 입력신호를 3개의 부대역으로 나누었다. 사용된 분석 윈도우 함수의 길이는 주파수가 가장 낮은 대역부터 높은 대역까지 각각 40ms, 20ms, 10ms이다.

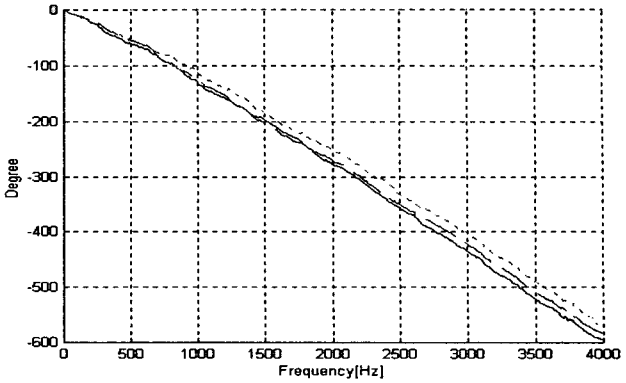
그림 3은 각각 원음성, 기존 모델을 이용하여 생성한 합성음, 그리고 본 논문에서 제안한 모델을 이용하여 생성한 합성음의 신호 파형, 스펙트럼 및 위상을 나타낸 것이다. 합성음의 파형에서는 비슷한 형태를 유지함을 볼 수 있으나, 스펙트럼 및 위상 특성에서는 제안된 방식이 원 음성신호의 특성과 좀 더 유사함을 확인할 수 있다.



(a) 원음성 및 합성음의 신호 파형



(b) 원음성 및 합성음의 스펙트럼



(c) 원음성 및 합성음의 위상

그림 3. 원음성과 합성음 비교(실선:원음성, 점선;기존 모델, 단선:제안 모델)

표 1은 기존 모델과 제안한 모델을 이용하여 생성한 10개의 합성음을 대상으로 PESQ 테스트를 수행한 결과를 보인 것이다. 기존 모델과 비교해서 대체적으로 추정 MOS 0.2 정도의 음질향상 효과가 있음을 알 수 있다.

표 1. 합성음의 PESQ 테스트

	기존 모델	제안 모델
추정 MOS	3.89	4.07

5. 결론

본 논문에서는 합성에 의한 분석 및 가산중첩 방식을 채택하고 있는 AbS/OLA 정현파 모델에 웨이블릿 변환을 적용한 새로운 모델을 제안하고 합성신호의 파형, 스펙트럼 및 위상, 그리고 음질을 비교·분석하였다. 제안한 모델은 기존 모델을 이용한 음성신호 분석 시 피치 주기의 약 2~2.5배 정도의 고정된 길이의 분석 윈도우를 이용하는 단점을 극복하고, 좀 더 정확한 파라미터 추출을 위하여 웨이블릿 변환을 이용하였다. 즉, 입력신호를 3개의 부대역 신호로 나눈 다음 각각 다른 길이의 분석 윈도우를 적용한 후 각각 독립적으로 AbS/OLA 정현파 모델을 이용하여 음성신호를 분석하고 합성한다.

실험결과 합성신호의 파형에서는 기존 모델과 비슷한 형태를 유지하지만, 스펙트럼 및 위상에서는 기존 모델보다 우수한 성능을 가짐을 확인하였다. 또한, 객관적 음질평가 테스트인 PESQ를 이용한 음질 테스트에서도 기존 모델과 비교하여 대체적으로 추정 MOS 0.2 정도의 음질향상 효과가 있음을 확인하였다.

앞으로 본 연구의 결과를 참조하여 보다 많은 음

성 데이터에 대한 분석·합성 실험과 기존 AbS/OLA 정현파 모델에서 보다 정확하고 효율적인 파라미터 추출 방식을 연구하고자 한다.

참고문헌

[1] R. J. McAulay and T. F. Quatieri, "Speech Anslysis/Synthesis Based on a Sinusoidal Representation", *IEEE Trans. on ASSP*, vol.34, pp.744-754, Aug., 1986.

[2] T. F. Quatieri and R. J. McAulay, "Speech Transformation Based on a Sinusoidal Representation", *IEEE Trans. on ASSP*, vol.34, pp.1449-1464, Aug., 1986.

[3] E. B. George and M. J. T. Smith, "Speech Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model", *IEEE Trans. on ASSP*, vol.5, pp389-406, Sep., 1997.

[4] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.

[5] O. Rioul and M. Vetterli, "Wavelet and Signals Processing", *IEEE Signal Processing Magazine*, pp.14-38, Oct., 1991.

[6] M. M. Goodwin, "Adaptive Signal Models : Theory, Algorithms, and Audio Applications", *ph. D. dissertation*, UCB, 1997.

[7] ITU-T Rec. P.862, "Perceptual Evaluation of Speech Quality(PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codes", Jan., 2002.