1999년도 한국음향학회 학술발표대회 논문집 제18권 제2(s)호

저연산 정현파 합성을 이용한 악기음의 모델링

오 복 환, 이 동 규, 송 인 호, 이 두 수
한양대학교 전자공학과
전화 : (02) 2290-0358 / 팩스 : (02) 2298-1796

Modeling of Instrumental Tone Using Low Computation Sinusoidal Synthesis

Bok Hwan Oh, Dong Gyu Lee, In Ho Song, Doo Soo Lee
Dept. of Electronic Engineering, Hanyang University
E-mail : audiosp@hyamil.hanyang.ac.kr

요 약

음향 신호의 모델링방법은 크게 분석,해석,합성의 3가지 과정으로 나눌 수 있다. 본 논문에서는 분석과 합성에 가산 합성방법의 한가지인 Analysis-by-synthesis /Overlap-Add 방법을 사용한다. 그리고 해석에 해당하는 주파수 영역에서의 피크추출은 제안한 방법에 의한다. 제안한 피크 추출 방법은 고조파 성분이 기본 주파수의 정수배가 된다는 점을 고려하여 적은 연산량으로 음향학적으로 의미있는 순음을 검출하는 방법이다. 음질보다 연산량에 더 주를 두었지만 모의 실험 결과를 통하여 음질 면에서도 원음과 거의 차이가 없음을 알 수 있었다.

I. 서 론

1860년에 Hermann von Helmholtz에 의해 처음으로 음악표현에 전기적인 신호개념을 도입하게 되어 오늘날의 음향 합성 방법으로는 가산합성,감산합성,벡터합성,FM합성,샘플링,선형산술합성 등이 있다[1].

본 논문에서는 모든 소리는 정현파들의 합으로 표현된다는 원리에 기반을 둔 가산 합성방식을 사용하여 어쿠스틱 음향을 전기적인 신호 음향으로 표현하는 효과적인 방법에 대해 제안한다. 정현파모델을 이용한 가산 합성방식은 신호를 분석하여 분석된 각각의 정현파성분들의 크기,주파수,위상을 추출하여 합성하는 방법이다. 정현파의 추출은 주파수 스펙트럼 상에서의 피크 추출을 통해 이루어진다. 본 논문에서는 악기음의 모델링에 관하여 논하고자 한다. 악기음의 경우 고조파가 기본 주파수의 정수배가 된다는 사실에 중점을 두어 음향학적으로 의미 있는 순음에 해당하는 피크를 추출하는 방법을 제안한다. 모의 실험에는 가장 표준적인 악기인 피아노를 사용하였다.

II. 가산 합성

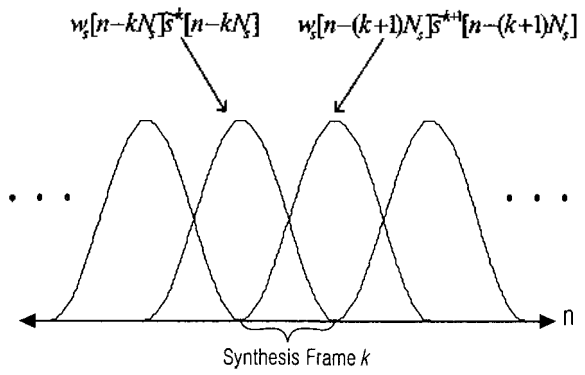
초기에 제안된 음향 합성방법은 (준)주기적인 음향신호를 정현파모델을 이용해 가산 합성하는 방식으로 식(1)과 같다.

$$\hat{s}(n)=\sum_{l=1}^{L}A_l\cos(l\omega_0 n+\phi_l) \qquad (1)$$

여기서 $A_l, l\omega_0, \phi_l$는 각각 l번째 정현파의 크기, 주파수, 위상에 해당한다. 식(1)과 같은 방법을 바로 적용하려면 정확한 신호의 피치(pitch) 주기를 알아야 하고 실시간 구현이 어렵기 때문에 Analysis- by- Synthesis /Overlap-Add(ABS/OLA)방법을 적용한다[2][3].
ABS/OLA방법의 가장 일반적인 형태는 식(2)와 같이 표현된다.

$$\hat{s}[n]=\sigma[n]\sum_{k=-\infty}^{\infty}w_s[n-kN_s]\,\hat{s}^k[n-kN_s] \quad (2)$$

여기서 $d[n]$는 신호의 에너지포락선이고, $w_s[n]$은 시간분해능을 높여주기 위해 사용하는 창함수로 각각의 프레임간의 연결을 부드럽게 해주기 위해서 complementary window를 사용하는데 <그림 1>과 같이 N_s 간격으로 창을 씌운다.



<그림1> 창을 이용한 Overlap-add synthesis의 구조

창의 길이는 신호의 피치 주기의 약 2.5배 이상이 되어야 한다[2]. 사용한 창은 측엽 억제의 특성이 좋은 Hamming window를 사용한다. 각각 창들의 정규화 형태는 식(3)과 같다.

$$\sum_{k=-\infty}^{\infty} w_s[n-kN_s]=1 \qquad (3)$$

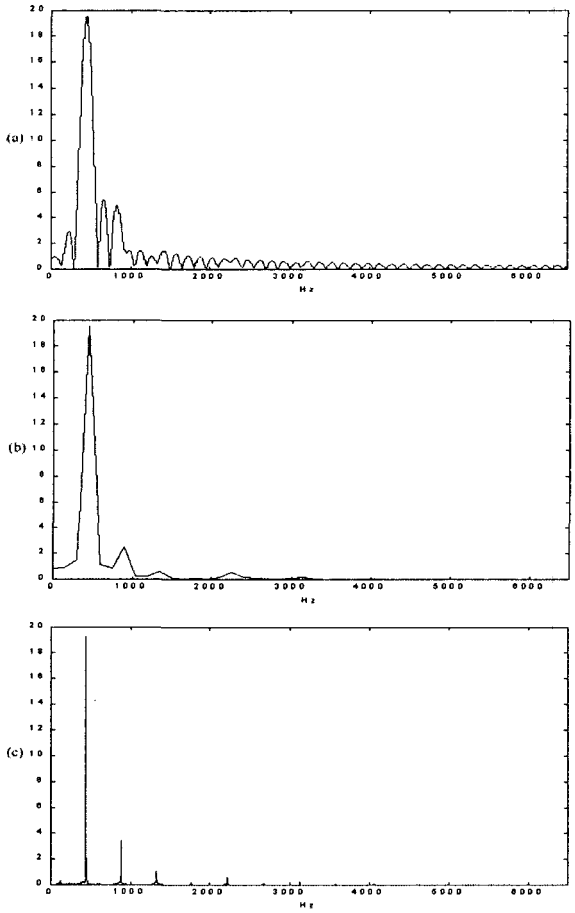
창에 의한 각각의 프레임에서 $\tilde{s}^k[n]$은 식(4)과 같이 표현된다.

$$\tilde{s}^k[n]=\sum_{l=1}^{L[k]} A_l^k \cos(\omega_l^k n+\phi_l^k) \qquad (4)$$

식(4)에서 $L[k]$는 각 프레임에서의 정현파의 수이다. 즉 창을 씌운 각각의 프레임들에서 신호를 분석하여 분석된 정현파의 합으로 재합성해주는데 각각의 프레임에서 합성된 값들을 중첩하여 더해주는 방법이다. E. Bryan George와 Mark J. T. Smith에 의해 제안된 ABS/OLA 방법은 창을 씌운 단시간 고속 퓨리에 변환을 통해 피크 부분의 크기, 주파수, 위상을 추출하여 합성하는 과정에서 원신호와의 차이를 자승하는 반복연산을 통해 오차 최소화시키는 방법이다[3]. 이때 추출된 정현파는 음향 신호의 순음성분에 해당한다. 그런데 이 방법은 반복 연산과정에 남아있는 오차부분에서 이전에 추출한 피크의 주변성분도 다음 연산과정에서 피크로 인식할 수 있는 단점이 있다. 본 논문에서는 신호의 분해 합성과정은 ABS/OLA 방법에서 순음검출에 해당하는 피크 추출 과정에 대한 효과적인 방법을 제안한다.

III. 제안한 피크 추출 방법

본 논문에서는 음향학적인 측면으로 접근하여 그 음의 특색을 잘 나타내 줄 수 있는 성분들을 효율적으로 추출한다. 악기음은 주파수 영역에서 보면 기본 주파수 성분과 그 것의 정수배에 해당하는 고조파 성분이 지배적으로 나타난다. 기본 주파수 성분에는 가장 많은 에너지가 집중하게 되는데, 대부분 악기음의 경우 기본 주파수가 음의 음계를 나타낸다.



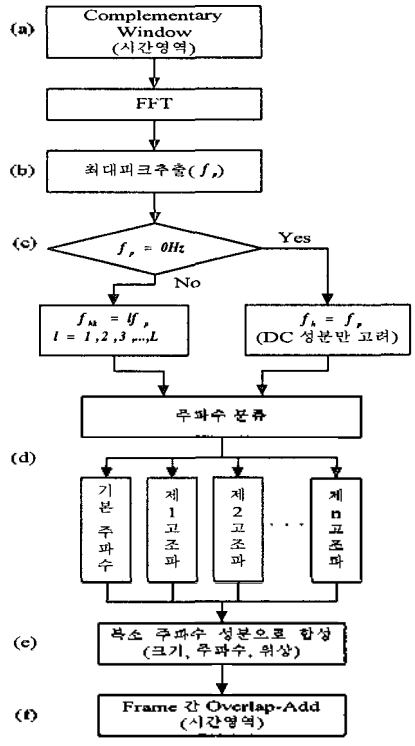
<그림2> 피아노 음(A4)의 스펙트럼

<그림 2>는 고속 퓨리에 변환의 길이 선택 방법을 각각 다르게 했을 때의 주파수 특성을 나타낸다. (a)는 한 개의 창을 초당 샘플의 개수(44100개)의 길이로 한 것이고 (b)는 한 개의 창을 그 창의 길이로 한 것이다. 그리고 (c)는 전체 샘플을 전체 샘플의 길이로 한 것이다. (a)의 경우는 시간 영역의 길이보다 긴 고속 퓨리에 변환이고 (b),(c)는 시간 영역과 같은 길이의 고속 퓨리에 변환이다. 모두 A4음에 관한 주파수 특성이

므로 A4음의 주파수에 해당하는 440Hz의 크기가 가장 크게 나타난다. (a)의 경우 주파수 해상도가 높아지기는 하지만 피크의 개수가 많아서 피크 추출이 더 어렵고 (b),(c) 경우 의미 있는 피크의 개수를 결정하기가 용이하다. 육안으로 피크의 개수를 결정하는 것이 청취 실험을 통해 상당히 유용함을 알 수 있다. 실험에 의해 <그림 2>의 경우 기본 주파수까지 포함해서 5개의 피크를 추출하면 된다. 피크 개수의 구별은 악기에 따라 다르게 되고 같은 악기의 다른 음에 관해서는 식(5)에 의해서 주파수 스케일만 변경되므로 같게 잡아주면 된다.

$$f_i = f_0 \times 2^{\frac{i}{n}} \quad (5)$$

식(5)는 평균율에 의한 음계 산출법이다. f_0는 기본 주파수이고 f_i는 산출된 주파수이다. 현대 음악의 경우 12평균율을 사용하므로 $n=12$가 된다.



<그림 3> 제안한 피크 추출 방법의 블록도

<그림 3>의 피크 추출 과정의 블록도이다. 과정을 설명하면 다음과 같다.

(a) 시간 분해능을 높여주고 실시간 처리를 위해서 Complementary Window를 씌워준다.

(b) 각각의 창에서 고속 퓨리에 변환을 하여 주파수 스펙트럼 상에서 최대 피크를 추출한다. 대부분의 경우는 최대 파크 하나만으로도 어떤 음계인가는 구별해 낼 수 있다.

(c) 최대 피크가 0*Hz*인 경우 DC 성분에 해당하는 데이 성분을 제외한 나머지 성분은 무시한다. 0*Hz*가 아닌 경우는 기본 주파수로 간주하고 추출할 피크의 수 만큼 정수배를 한다.

(d) (c)를 통해서 추출한 피크가 순음에 해당하는 기본 주파수와 고조파가 된다.

(e) 추출된 각각의 성분들을 크기, 주파수, 위상을 모두 고려하여 복소 주파수 성분으로 합성한다. 추출된 성분은 식 (6)과 같다.

$$d_l^k = \alpha_l^k + j\beta_l^k \quad (6)$$

식(4)에서의 크기, 주파수, 위상은 각각 다음과 같다.

$$A_l^k = \sqrt{\alpha_l^{k2} + \beta_l^{k2}} \quad (7)$$

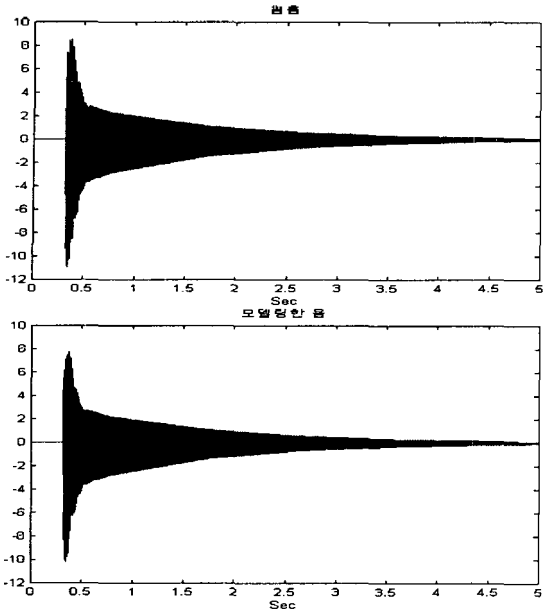
$$\omega_l^k = 2\pi l f_p^k \quad (8)$$

$$\phi_l^k = atan(\beta_l^k / \alpha_l^k) \quad (9)$$

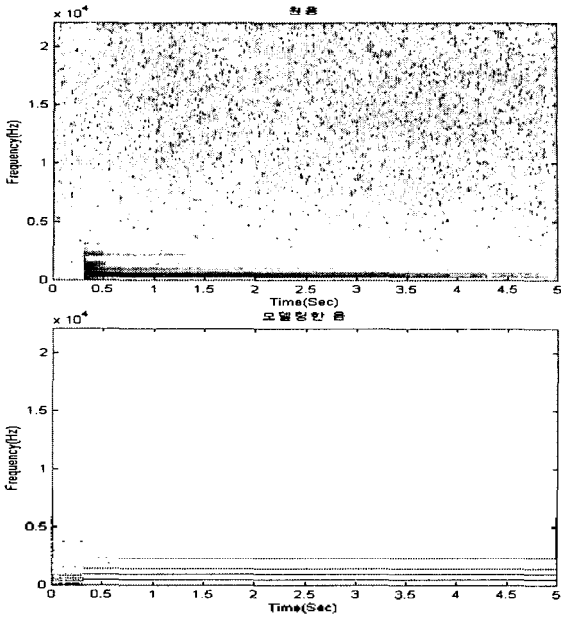
(f) 합성된 각각의 프레임들을 시간영역에서 반씩 중첩하여 더한다.

IV 실험 및 결과

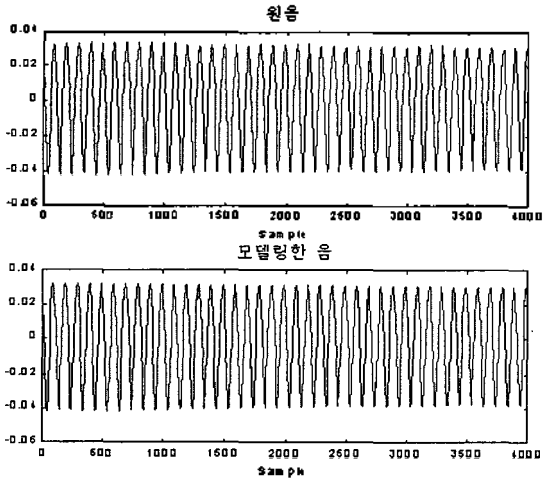
모의 실험에는 가장 표준적인 악기인 피아노를 사용한다. 그리고 대부분 악기의 조율음으로 사용되는 A4(라4)음을 쓴다. 입력 시 주변 잡음을 최소로 한 상태에서 16bit 44.1kHz의 샘플링 주파수로 5초간 녹음하여 실험 데이터를 만든다. 총 220500개의 샘플을 길이가 294인 hamming window 1499개를 이웃하는 것과 절반 길이 만큼 중첩되게 창을 씌워 분석한다. 추출하는 피크의 개수는 <그림 2>를 통해 5개로 정하고 주파수의 대칭구조를 고려하여 음의 주파수 성분까지 각 프레임 당 10개의 피크를 추출한다. <그림 4>는 원음과 모델링한 음의 전체적인 신호의 포락선 비교이다. <그림 5>는 Spectrogram 비교인데 모델링한 음의 경우 정해진 몇 개의 주파수 성분만 나타나고 원음은 주파수 성분이 여러 영역에 퍼져서 나타난다. 특히 음의 attack 부분에서는 다른 부분에 비해서 고조파 성분이 더 많이 나타난다. 제안한 방법에 의해 모델링한 음의 경우 각각의 프레임에서 추출하는 피크의 개수를 같도록 하였기 때문에 청취 실험 결과 원음과 비교하여 attack부분에서 미세한 차이를 보인다. 하지만 연산량을 줄일 수 있다는 장점이 있었다. 그리고 음의 진행 중간 부분에서는 청취 실험뿐 아니라 실제 파형 상으로도 원음과 거의 차이가 없었다. <그림 6>은 음의 중간 부분의 몇 개의 샘플에 관하여 원음과 모델링한 음의 비교이다.



<그림 4> 신호 포락선 비교



<그림 5> Spectrogram 비교



<그림 6> 음의 중간 부분 비교

V. 결 론

본 논문에서는 가산 합성방식으로 ABS/OLA 방법을 도입하여 음향 신호의 분석과 합성에 이용하였다. 추출과정에서는 음향학적인 측면으로 접근하여, 악기 음의 중요한 성분인 기본 주파수와 고조파 성분에 대한 피크를 추출하였다. 그 결과 기존의 방법보다 적은 연산량으로 원음과 거의 유사한 음이 합성되었다. 그러나, 음질의 개선을 위한 향후 연구가 계속 진행되어야 할 것이다.

참고문헌

[1] http://www.media.mit.edu/~gan/Gan/Education /NUS/Physics/MScThesis

[2] R. J. McAulay and T. F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation", IEEE trans. Acoust. Speech, Signal Processing, Vol. ASSP-34, No. 4, Aug. 1986, pp. 744-754

[3] E. B. George and M. J. T. Smith, "Analysis-by-Synthesis/Overlap-Add Sinusoidal Modeling Applied to the Analysis and Synthesis of Musical Tones", J. of Audio Eng. Soc. Vol. 40, No. 6, 1992 June, pp. 497-516

[4] Mark Kahrs and Karlheinz Brandenburg, Applications of digital signal processing to audio and acoustics, Kluwer Academic Publishers, 1998

[5] 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 Speech and Audio Processing, Vol. 5, No. 5, Sep, 1997, pp.389-406

[6] E. B. George "Practical High-Quality Speech And Voice Synthesis Using Fixed Frame Rate ABS/OLA Sinusoidal Modeling" Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing Vol 1, May, 1998, pp. 301-304