

## 기본 주파수와 고조파 성분을 고려한 악기음의 모델링

오복환, 이동규, 이두수

한양대학교 전자공학과

전화 : (02) 2290-0358 / 팩스 : (02) 2298-1796

### Modeling of Instrumental Tone considering Main Frequency and Harmonics

Bok Hwan Oh, Dong Gyu Lee, Doo Soo Lee

Dept. of Electronic Engineering, Hanyang University

E-mail : audiosp@hyamil.hanyang.ac.kr

#### Abstract

In this paper, using one method of Additive Synthesis, Analysis-by-synthesis/Overlap-Add (ABS/OLA) method, analysis and synthesis of musical tones is processed. But peak detection of frequency domain is processed by proposed method considering the view of acoustics. It is that harmonics frequency is times of main frequency. Using this fact, peak detection of frequency domain is useful for detection of tonal component identified musical note. It is possible to realize high-quality low bit rate audio

#### I. 서론

과학문명이 발달함에 따라서 인간의 기술영역이 산업 분야에 많은 발전을 가져오게 되었다. 더 나아가서 수요가 고급화되면서 예술분야에도 과학문명의 힘이 필요하게 되었고 과학과 예술의 접목이라는 새로운 분야가 부각되기 시작하였다.

본 논문에서는 전자음향을 어쿠스틱 음향에 좀 더 가깝게 표현하기 위한 방법 중 악기음의 모델링에 관하여 논하고자 한다.

1860년에 Hermann von Helmholtz에 의해 처음으로 음악표현에 전기적인 신호개념을 도입하게 되어 오늘

날의 음향 합성 방법으로는 가산합성, 감산합성, 벡터 합성, FM합성, 샘플링, 선형산술합성 등이 이용된다[1].

본 논문에서는 모든 사운드는 정현파들의 합으로 표현된다는 원리에 기반을 둔 가산 합성방식을 사용하였다. 정현파 모델을 이용한 가산합성 방식은 신호를 분석하여 분석된 각각의 정현파 성분들의 크기, 주파수, 위상을 추출하여 합성하는 방법이다. 정현파의 추출은 주파수 스펙트럼 상에서의 피크를 통해 이루어진다. 본 논문에서는 고조파가 기본 주파수의 정수배가 된다는 사실에 중점을 두어 음향학적으로 의미 있는 순음에 해당하는 피크를 추출하는 방법을 제안한다. 모의 실험에 사용되는 악기는 음의 신호 포락선이 중간 부분에서 수평 방향으로 계속 진행되지 않는 형태인 피아노를 사용한다.

#### II. 가산 합성

초기에 제안된 음향 합성 방법은 (준)주기적인 음향 신호의 정현파 모델을 이용한 가산합성 방식으로 식(1)과 같이 표현된다.

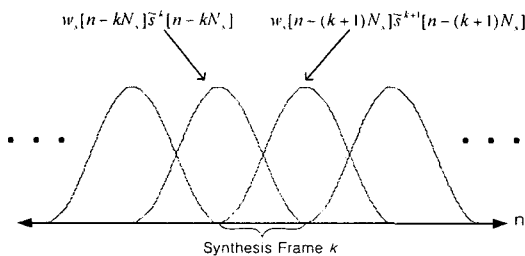
$$\hat{s}(n) = \sum_{l=1}^L A_l \cos(\omega_0 n + \phi_l) \quad (1)$$

여기서  $A_l$ ,  $\omega_0$ ,  $\phi_l$ 는 각각 분해된  $l$ 번째 정현파의 크기, 주파수, 위상에 해당한다. 식(1)과 같은 방법을

바로 적용하려면 정확한 신호의 피치 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.

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

여기서  $\alpha[n]$ 는 신호의 에너지 포락선이고,  $w_s[n]$ 은 시간 분해능을 높여주기 위해 씌우는 창함수로 각각의 프레임간의 연결을 부드럽게 해주기 위해서 complementary window를 사용하는 데 <그림 1>과 같이  $N_s$  간격으로 중첩하여 창을 씌운다.



<그림1> 창을 이용한 Overlap-add synthesis의 구조

창의 길이는 신호의 피치 주기의 약 2.5배 이상이 되어야 하고 사용한 창은 축엽 억제 특성이 좋은 Hamming window를 사용한다. 각각 창들의 정규화 형태는 식(3)과 같다.

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

각각의 프레임에서 창과 곱하게 되는  $\hat{s}^k[n]$ 은 식(4)과 같이 표현된다.

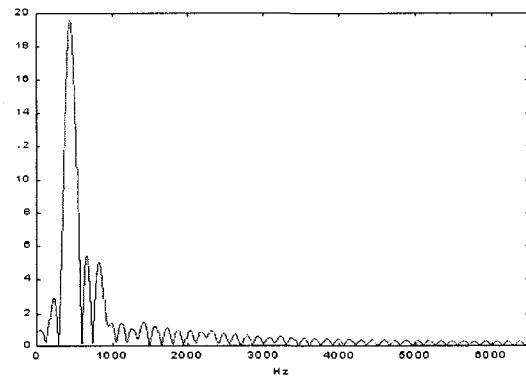
$$\hat{s}^k[n] = \sum_{l=1}^{L[k]} A_l^k \cos(\omega_l^k n + \phi_l^k) \quad (4)$$

식 (4)에서  $L[k]$ 는 각 프레임에서의 정현파의 수이다. 즉 창을 씌운 각각의 프레임들에서 신호를 분석하여 분석된 정현파의 합으로 재 합성해주는데 각각의 프레임에서 합성된 값들을 중첩하여 더해주는 방법이다. E. Bryan George와 Mark J. T. Smith에 의해 제안된 ABS/OLA 방법은 창을 씌워 단시간 고속 푸리에

변환을 통해 피크 부분의 크기, 주파수, 위상을 추출하여 합성하는 과정에서 원 신호와의 차의 자승오차를 반복 연산을 통해 최소화시키는 방법이다[3]. 이때 추출된 정현파는 음향 신호의 순음 성분에 해당한다. 그런데 이 방법은 반복 연산 과정에서 남아있는 오차 부분에서 이전에 추출한 피크의 주변성분도 다음 연산 과정에서 피크로 인식할 수 있는 단점이 있다. 본 논문에서는 신호의 분해 합성과정은 ABS/OLA 방법을 쓰되 순음 검출에 해당하는 피크 추출 과정은 제안된 방법을 이용한다.

### III. 피크 추출

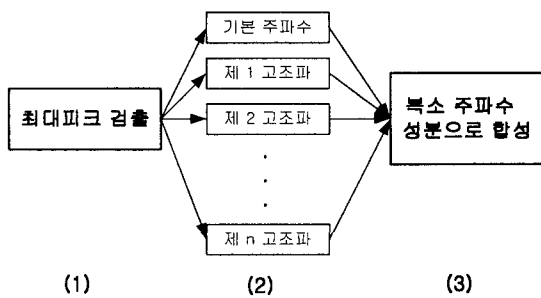
피크 추출 과정으로 심리 음향적 분석 방법이 거론될 수 있다. 이 방법은 마스킹 레벨을 산출하고 그 레벨을 문턱 값으로 설정하여 피크를 추출하는 방법이다[4]. 심리 음향적 기준에 바탕을 둔 MPEG 오디오의 경우 최대 약 12 : 1의 압축비를 제공할 수 있을 정도로 심리 음향적 기준은 매우 유용하다. 하지만 기존의 심리 음향모델은 자체적으로 정확한 순음, 잡음 판별을 수행하지 못하고 분석 과정이 복잡하다는 단점이 있다. 본 논문에서는 음향학적인 측면으로 접근하여 그 음의 특색을 잘 나타내 줄 수 있는 성분들을 효율적으로 추출한다.



<그림2> 피아노 음(A4)의 주파수 스펙트럼

<그림 2>와 같이 악기음은 기본 주파수 성분과 그것의 정수배에 해당하는 고조파 성분이 나타나게 된다. 기본 주파수 성분에는 가장 많은 에너지가 집중하게 되는데, 대부분 악기음의 경우 기본 주파수가 그 음의 음계를 나타내게 된다. <그림 2>의 경우에는 A4음의 주파수가 약 440Hz이므로 440Hz의 크기가 가장 크게 나타난다. 심리 음향적 분석의 경우 마스킹 레벨에

만 의존하여 피크 주파수들간의 관계에 대해서는 알지 못하기 때문에 고조파인지 아닌지 판별할 수 없다. 그리고 대부분 악기의 경우 크기가 가장 큰 주파수가 그 음계에 해당하지만, 어떠한 경우는 가장 큰 주파수가 그 음계와 다른 옥타브의 같은 음일 경우도 있다. 그 경우 음계에 해당하는 주파수의 크기가 마스킹 임계치보다 작으면 피크 추출과정에 빠져버릴 수도 있다. 본문에서는 우선 최대 피크치를 추출한 다음 그 주파수의 정수배에 해당하는 주파수들이 피크에 해당하는가를 판별하여 피크를 적응적으로 추출하는 방법을 이용한다.



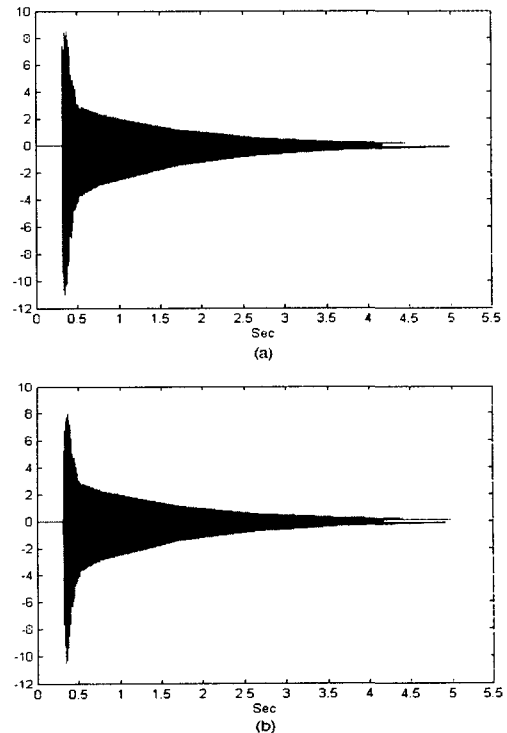
<그림 3> 제안된 피크 추출 방법

<그림 3>의 피크 추출 과정에서처럼 먼저 최대 피크를 추출한다(1). 대부분의 경우는 최대 피크 하나만으로도 어떤 음계인가는 구별해 낼 수 있다. 각각의 프레임에서 피크 추출을 위한 고속 푸리에 변환의 길이는 시간 성분의 길이와 같도록 한다. 시간 성분의 길이보다 긴 고속 푸리에 변환은 주파수의 해상도를 높여주는 하지만, 입력 신호 자체가 이산적인 신호이므로 높아진 해상도가 피크 추출과정을 더 어렵게 만들기 때문이다. 최대 피크 추출 후 그 최대 피크의 정수배 주파수 성분들을 차례로 판별하여 적절하게 피크의 수를 조정할 수 있다(2). 음의 발생 초기에는 그 음계의 주파수 특성과는 다른 잡음 성분이 나타난다. 이 경우 최대 피크와 다른 피크가 서로 정수배가 되지 않고, 정수배의 주파수 판별 과정이 얼마 가지 않아 샘플링 주파수의 반을 넘게 되므로 최대 피크가 순음에 해당하지 않는다는 것을 알 수 있게 된다. 추출된 피크에 해당하는 정현파 들에 대해서는 각각의 크기, 주파수, 위상 정보를 가지고 다시 재 합성하게 된다(3). 합성하는 과정에서 위상 정보는 매우 중요한 역할을 한다. 피크 추출 과정에서는 주파수 성분의 절대값을 이용한다. 그래서 위상 정보를 무시하기 쉬운데 위상 정보가 잘 표현되지 않을 경우 소리의 잔향 현상이 나

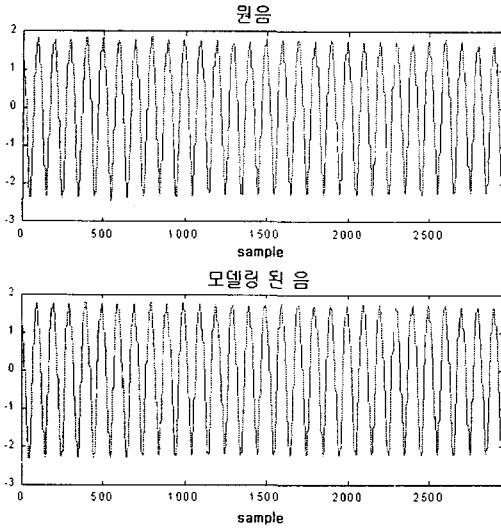
타나게 된다. 따라서 피크에 해당하는 주파수의 위치를 알아서, 합성 과정에서는 같은 위치의 복소 주파수 성분으로 합성을 한다.

#### IV 실험 및 결과

모의 실험에 사용하는 악기는 신호 포락선이 음의 진행 중간 부분에서 수평 방향으로 계속 진행되지 않는 형태인 피아노를 사용하기로 한다. 그리고 대부분 악기의 조율음으로 사용되는 A4(라4)음을 쓴다. 입력 시에는 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. 청취 실험 결과 제안된 방법에 의해 구현된 음이 원음에 비하여 attack부분의 소리는 미세한 차이로 약하게 들리지만, 음의 진행 중간 부분은 거의 차이가 없다.



<그림 4> 두 음의 신호 포락선 비교



<그림 5> 두음의 중간부분 비교

<그림 4>는 전체적인 신호의 포락선 형태를 나타낸 것인데 (a)는 원음이고 (b)는 제안된 방법에 의해서 구현된 음이다. <그림 5>는 음의 진행 중간 부분에서 몇 개의 샘플에 관해서 비교를 한 것이다. 음의 진행 중간 부분은 청취 실험 뿐 만 아니라 실제 파형 상으로도 매우 유사한 형태를 나타내고 있다.

## V. 결론

본 논문에서는 가산 합성방식으로 ABS/OLA 방법을 도입하여 음향 신호의 분석과 합성에 이용하였다. 추출과정에서는 음향학적인 측면으로 접근하여, 악기음은 기본 주파수와 그것의 정수배에 해당하는 고조파 성분이 주요한 요소라는 것을 착안하여 피크를 추출하였다. 그 결과 원음과 거의 가까운 음이 합성되었다. 피크 추출 과정에서 제외되는 피크들은 연속적인 신호를 이산적인 신호로 만들기 위해 샘플링 하였기 때문에 나타나게 되는 현상이다. 그러므로 불필요한 피크들을 제외시키고 음향학적으로 의미가 있는 피크들만 추출해냄으로써 실제 어쿠스틱 사운드에 가깝게 모델링 할 수 있었다. 피크 성분의 주파수를 미리 예측할 수 있어서 연산량도 줄게 되고, 불필요한 피크들이 제외되므로 고음질 저 비트율의 사운드 실현도 가능하게 할 수 있다.

## 참고문헌

- [1] Cochran, Connor Freff and Rychner, Lorenz. "The Truth About Synthesizers", Electronic Musician, Feb 1991, 7(2), pp. 102-105.
- [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] E. Zwicker and H. Fastl, Psychoacoustics : Facts and Models Heidelberg, Germany, Springer-Verlag, 1990
- [5] Mark Kahrs and Karlheinz Brandenburg, Applications of digital signal processing to audio and acoustics, Kluwer Academic Publishers, 1998
- [6] 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