정현파 모델을 이용한 오디오 신호의 심리음향적 분석 및 합성

Analysis and Synthesis of Audio Signals using a Sinusoidal Model with Psychoacoustic Criteria

남 승 현*, 강 경 옥**, 홍 진 우**
(Seung Hyon Nam*, Kyeong Ok Kang**, Jin Woo Hong**)

요 약

정현파 모델은 음성과 오디오 신호의 분석과 합성에 많이 활용되어 왔으며 최근 고음질 저비트율 오디오 부호화에 효율적인 방법의 하나로 대두되고 있다. 정현파 모델을 이용한 오디오 신호의 분석과 합성에서 중요한 단계 중의 하나는 순음의 검출이다. 본 논문은 정현파를 이용한 오디오 신호의 분석과 합성에 매스킹 효과와 매스킹 인덱스 그리고 JNDf(Just Noticeable Difference in Frequency) 등의 심리음향적 기준들을 활용하는 효율적인 방안을 제안하였다. 모의실험 결과, 심리음향적 기준을 사용하면 합성된 음질에 거의 영향을 주지 않으면서 합성에 사용되는 정현파의 개수를 현저하게 줄일 수 있었음을 알 수 있었다.

ABSTRACT

A sinusoidal model has been widely used in the analysis and synthesis of speech and audio signals, and becomes one of the efficient candidates for high quality low bit rate audio coders. One of the crucial steps in the analysis and synthesis using a sinusoidal model is the detection of tonal components. This paper proposes an efficient method for the analysis and synthesis of audio signals using a sinusoidal model, which uses psychoacoustic criteria such as masking effect, masking index, and JNDf(Just Noticeable Difference in Frequency). Simulation results show that the proposed method reduces the number of sinusoids significantly without degrading the quality of the synthesized audio signals.

I. 개 요

지난 10여년 간 고음질 디지털 오디오 부호화는 사람의 청각특성을 이용한 심리음향모델의 활용에 힘입어 급격한 발전을 이루었다[1]. MPEG 오디오 표준은 이러한 연구의 산물로서 최대 약 12 : 1의 압축비를 제공하고 있다. 고음질을 유지하면서 압축비를 높이는 시도는 앞으로 지속될 것이며 많은 진보가 기대된다. 기존의 오디오 부호화기에서 사용하고 있는 스펙트럼의 단순한 양자/부호화 방식은 근본적으로 오디오 신호의 중복성(redundancy)을 효과적으로 제거하지 못하므로 더 높은 압축비를 얻는데 한계를 지닐 수 밖에 없다. 따라서 보다 효율적인 부호화를 위해서 신호의 중복성을 제거할 수 있는 새로운 오디오 신호 모델이 필요하다. 이런 측면에서 정현파 모델은 최근 많은 관심을 끌고 있는 것 중의 하나이다[2].

정현파 모델은 오디오 신호를 주기적인 순음 성분들의 합으로 표현하는 방법으로 초기에 음성 신호용과 컴퓨터 음악등에 활용되어 왔으나, 배경 잡음과 비음성 신호에 강인한 특성으로 인해 오디오 신호의 분석, 합성, 부호화에도 활용되기 시작했다[3]. 정현파 모델이 음성과 오디오 신호의 부호화에 효율적이기는 하지만 순음 만을 이용하는 경우 오디오 신호를 효과적으로 모델링할 수 없다. 예를 들면, 현악기음에 포함된 마찰음과 같은 잡음 성분을 표현하기 위해서는 비주기적인 잡음 성분이 필요하다. 이와 같이, 오디오 신호를 주기적인 순음 성분과 잡음성분으로 분류하거나[4], 순음 성분과 잡음 성분 그리고 신호의 attack에서 나타나는 일시적인 (transient) 성분으로 분류하는 방안이 제안되었다[6].

본 논문에서는 오디오 신호의 성분 중 순음 성분의 검출/분류/추적/합성하는 문제에 초점을 맞추어 논의하고자 한다. 먼저 정현파 모델을 이용한 오디오 신호의 분석 및 합성 모델을 설명하고 심리음향적 기준을 이용한 순음 성분의 검출, 추적, 합성 방식을 제안하고 모의실험 결과를 설명한다.

II. 정현파 모델

McAulay와 Quatieri에 의해 제안된 정현파 모델은[2]

* 배재대학교 전자공학과
** 한국전자통신연구원 무선방송기술연구소 방송기술연구부
접수일자 : 1998년 11월 9일

$$\hat{s}[n] = \sum_{l=1}^{L} A_l[n]\cos(\omega_l[n]\,n+\theta_l) \qquad (1)$$

로 주어진다. 여기서 $A_l[n]$, $\omega_l[n]$, $\theta_l[n]$은 각각 l번째 순음 성분의 진폭, 순간 주파수, 그리고 위상이다. 그림 1은 정현파 모델을 활용한 오디오 신호의 분석 및 합성 과정이다.

분석의 첫번째 과정은 입력 오디오 신호에 분석 윈도우를 씌워 STFT(Short-Time Fourier Transform)을 취하여 주파수 영역으로 변환하는 것이다. 이때 분석 윈도우 $w_a[n]$은 홀수 길이로 정해지며

$$\sum_{n=-N}^{N} w_a[n] = 1 \qquad (2)$$

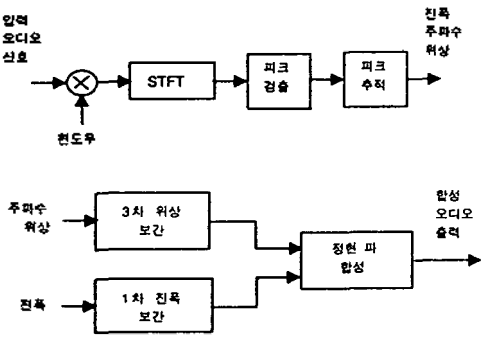

그림 1. 정현파 모델을 이용한 오디오 신호 분석 및 합성
Fig. 1. Analysis and synthesis of audio signals using a sinusoidal model.

와 같이 정규화된다. 순음 검출을 효과적으로 하기 위해서 분석 윈도우는 충분한 측엽 억제(sidelobe suppression) 특성을 갖아야 한다. 일반적으로 해밍 윈도우나 Blackman-Harris 윈도우 등이 많이 사용된다.

윈도우된 오디오 신호는 순음의 파라미터를 예측하기 위해 STFT를 사용하여 주파수 영역으로 변환된다. 이때, 윈도우의 선형 위상 특성의 영향을 배제하기 위해서 윈도우된 데이터가 STFT 버퍼의 중앙에 위치하도록 조정된다[4].

일반적으로 오디오 신호 중의 순음 성분은 STFT 스펙트럼의 피크로 해석된다[2]. 이것은 순음의 스펙트럼이 주파수 영역에서 임펄스와 윈도우 스펙트럼의 컨볼류션으로 이루어지기 때문이다. 따라서 순음 성분의 검출은 STFT 스펙트럼의 피크 중 피크를 검출함으로써 이루어지며 검출된 피크로부터 진폭, 순간주파수, 위상의 정보가 산출된다. 이때, STFT 스펙트럼으로부터 순음의 파라미터를 제대로 산출하기 위해서는 분석 윈도우의 길이가 피치 주기의 약 2.5배 이상 되어야 한다. 또한 STFT 스펙트럼으로부터 산출되는 주파수의 해상도는 분석 윈도우의 길이에 반비례하므로 요구되는 주파수 해상도를 결정한 다음 분석 윈도우의 길이를 결정하는 것이 필요하다. 실제적으로 요구되는 주파수 해상도는 사람의 청각 특성을 고려하여 결정되어져야 한다. 사람의 청각 특성 중 JNDf (Just Noticeable Difference in Frequency)는 사람이 인식할 수 있는 가장 작은 주파수의 차이를 의미하며 500Hz 미만에서는 3.6 Hz이고 500Hz 이상에서는 중심 주파수의 약 0.7%로 정의되어진다[5]. 따라서, 3.6Hz의 해상도를 얻기 위해서는 매우 긴 STFT를 사용해야만 한다. 그러나 이것은 시간 해상도를 떨어뜨리며 계산량을 급증시키기 때문에 비현실적이다. 따라서 STFT 크기를 적절한 수준으로 유지하면서 주파수 해상도를 높이는 방법들이 활용되고 있다[4][7]. 이들은 보간법을 사용하거나 regression을 사용하여 스펙트럼의 피크가 FFT 선상에서 벗어나 있는 경우 정확한 피크에서의 주파수를 예측할 수 있기 때문에 매우 효과적이다.

일반적인 분석에서는 피치 검출이 생략되지만, 경우에 따라 피치 검출 과정이 포함될 수 있다. 검출된 피치 정보는 앞에서 언급한 것과 같이 분석 윈도우의 길이를 적응적으로 변화 시키는데 활용될 수 있으며, 다음 단계에서 전개될 피크 추적 과정을 단순하게 하거나, 또한 검출된 피크들로부터 하모닉스를 인식하고 도출하는 과정에서도 활용될 수 있다. 일반적으로 오디오 신호에서의 피치 검출은 음성 신호와는 달리 매우 까다롭지만 two-way mismatch 방안을[8] 사용하면 비교적 안정적인 좋은 결과를 얻을 수 있다.

각 프레임에서 검출된 피크들로부터 산출된 각 프레임의 진폭, 순간 주파수, 위상 정보로부터 오디오 신호를 합성하기 위해서는 프레임 간의 부드러운 연결이 필요하다. 이러한 연결을 위한 첫번째 과정은 각 프레임에서 검출된 피크들을 연결시키는 작업이다. 이 작업은 그림 2와 같이 현재 프레임에서 검출된 피크들을 과거 프레임에서 검출된 피크들과 비교하는 생성-소멸 매칭 방식의 피크 추적 과정으로 이루어진다[2]. 피크 추적 과정을 거쳐 프레임 간의 연속성이 찾아진 피크들에 대해서는 프레임 간의 보간법이 사용되는데 진폭에 대해서는 1차 보간이 위상에 대해서는 3차 보간이 실행된다[2].

보간 후 k번째 프레임에서의 오디오 신호 합성은

$$\hat{s}^k[n] = \sum_{l=1}^{L^k} A_l^k[n]\cos(\psi_l^k[n]) \qquad (3)$$

과 같이 이루어진다. 여기서 L^k는 k번째 프레임에서의 순음의 개수이며, $A_l^k[n]$과 $\psi_l^k[n]$은 k번째 프레임에서 보간 후 얻어진 l번째 순음 성분의 진폭과 위상이다.

일반적으로 위상 정보의 정확도는 음질에 크게 영향을 미치지 않는 것으로 알려져 있으나, 위상 정보가 생략될 경우 울림(reverberance)이 발생하기 때문에 이를 방지하기 위해서는 위상 정보의 사용이 필수적이다. 또한 오디오 신호를 순음 성분과 잡음 등의 잔여 성분으로 분리하여 합성할 경우 순음 성분의 위상 일치성이 오디오 신호의 효과적인 분리에 매우 중요하므로 상당한 정확도의 위상 정보가 요구된다[4]. 식 (3)을 이용한 시간 영역의 합성은

시간 영역의 파형을 매우 정확하게 복원할 수 있는 장점이 있지만 계산량이 많다는 단점도 있다. 따라서 계산량을 줄이기 위해 IFFT를 활용하는 주파수 영역의 합성이 활용되기도 한다[4].

III. 심리음향적 기준을 이용한 순음 성분의 검출 및 추적

정현파 모델을 이용한 오디오 신호의 분석과 합성에 가장 중요한 과정은 순음 성분의 검출과 추적이다. 일반적으로 오디오 신호에서 순음 성분은 스펙트럼 상의 피크로 인식되어진다. 그러나 STFT 스펙트럼에 피크가 무수히 많기 때문에 이중 어느 피크가 순음 성분에 해당하는가가 문제가 남는다. 순음 검출 방법 중 전통적으로 많이 사용되어 온 파라메트릭 방법들은 순음의 개수가 알려진 상황에서 사용되어지므로 여기에 적용하기는 어렵다. 일반적으로 많이 사용되었던 방법은 앞에서 언급한 바와 같이 단순히 일정한 크기의 문턱값을 초과하는 모든 스펙트럼 피크들을 검출한 다음 피크 추적 과정에서 순음이라고 판별되는 피크들을 추출하는 것이다[4]. 이때 피크 추정 과정에서 사용하는 방법은 피크가 순음일 경우 일반적으로 일정 시간 이상 지속된다는 성질을 활용하는 것이다. 그러나, 이 경우 매우 많은 수의 피크들이 검출되기 때문에 신호의 압축을 목적으로 하는 오디오 신호의 부호화에 그대로 적용하기는 어렵다. 다른 순음 검출 방법으로 least square 최적화 방법이 제안되었다[3]. 그러나, 이 방법은 원래의 스펙트럼에서 순음에 해당하는 스펙트럼을 뺀 후 남게 되는 잔여 스펙트럼이 다시 피크로 인식되기 때문에 하나의 순음 성분을 여러 개의 순음으로 모델링하는 결과를 가져오는 단점이 있다. 이러한 순음 검출 방식의 단점을 피하기 위해, 최근에는 Thomson의 하모닉 분석 알고리즘을 채택하는 것이 제안되었다[6]. 이 방법은 5개의 prolated spheroidal window를 사용하여 각각을 이용한 스펙트럼을 구하고 이들의 평균값과 오차로부터 순음을 검출하는 것으로 매우 효과적인 것으로 보고되었다[6]. 그러나 각 프레임에서 FFT를 5번이나 반복적으로 계산해야 하는 계산량의 증가가 큰 단점이다.

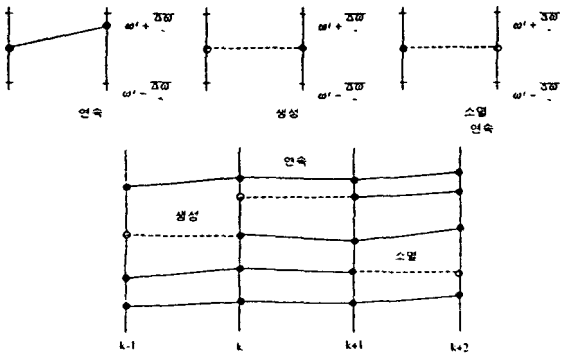

그림 2. 생성-소멸 매칭 과정을 이용한 피크 추적
Fig. 2. Peak tracking using a birth-death matching process.

본 논문에서 제안하는 방법은 입력 오디오 신호로부터 심리음향모델을 이용하여 문턱값을 산출하고, 이 문턱값을 넘는 스펙트럼 피크들을 검출하는 방법으로 비교적 정확한 순음 판별이 가능하다. 문턱값의 산출은 기본적으로는 매스킹 레벨의 산출 과정과 거의 유사하다. 매스킹 레벨의 산출은 일반적인 지각 오디오 부호화기에 필수불가결의 요소이기 때문에 계산량의 추가적인 부담은 없다고 볼 수 있다. 다만 기존의 심리음향모델이 자체적으로 정확한 순음/잡음 판별을 수행하지 못하기 때문에 심리음향모델을 이용하여 순음과 잡음을 구분하려는 것이 일견 모순된 것과 같이 여겨질 수 있다. 그럼에도 불구하고, 심리음향모델의 활용은 다음과 같은 이유에서 충분히 그 의미가 있다.

먼저 심리음향모델을 사용함으로써 매스킹 레벨보다 크기가 작은 피크들이 순음으로 검출되는 것을 방지할 수 있다는 장점을 들 수 있다. 이들 작은 피크들의 대부분은 분석 윈도우의 측엽에 의한 것으로 당연히 합성에서 제외되어져야만 한다. 측엽에 의한 피크들을 순음으로 판단하여 합성에 사용하면 경우에 따라 불필요한 위상 일치성(phase coherency)이 증가하게 되며, 이로부터 울림 현상이 일어나는 등의 음질 저하를 가져올 수 도 있다. 따라서, 심리음향모델을 사용함으로써 음질의 저하를 초래하지 않으면서도 순음의 개수를 현저하게 줄일 수 있는 것이다.

이러한 방법은 오디오 신호의 스펙트럼에 존재하는 분석 윈도우의 측엽에 의한 피크들을 효과적으로 제거함으로써 하모닉 성분들이 잘 나타나는 오디오 신호에 대해서 매우 효과적이다. 그러나, 만일 오디오 신호에 순음과 잡음이 함께 뒤섞여 있다면 잡음 성분도 매스킹 레벨 보다 클 수 있으므로 피크로 선택될 수 있다. 이러한 문제를 해결하기 위해 본 연구에서는 매스킹 인덱스를 도입하였다.

매스킹 레벨은 신호의 자극 레벨과 매스킹 레벨의 차이로서 신호의 성분이 순음인가 잡음인가에 따라 다르며 bark 주파수 z의 함수이다. MPEG-1 오디오의 심리음향모델 1에서 정의된 순음 매스킹 인덱스 av_t와 잡음 매스킹 인덱스 av_n는 각각 다음과 같다[9].

$$av_t = 6.025 + 0.25z$$

$$av_n = 2.025 + 0.175z$$

식 (4)에서 알 수 있듯, 순음의 매스킹 인덱스는 잡음의 매스킹 인덱스 보다 크다. 이것은 동일한 음압의 순음과 잡음 성분이 주어졌을 때 순음에 대한 매스킹 레벨이 잡음에 대한 매스킹 레벨보다 더 낮다는 것을 의미한다. 따라서 이미 산출된 매스킹 레벨에 잡음의 매스킹 인덱스를 더한 값을 문턱값으로 설정하고 이 값보다 큰 피크들을 검출한다면 순음에 해당되는 피크들만이 검출될 것이다. 따라서, 오디오 신호에 순음과 잡음이 섞여 있는 경우 잡음의 피크들을 제외한 순음의 피크들 만이 효과적으로 검출될 것이다.

앞에서 언급한 바와 같이 이 방법은 이미 순음과 잡음

이 완벽하게 검출된 것을 전제로 하기 때문에 논리적으로 모순과도 같으나 모의 실험 결과 실제적으로는 매우 효과적인 것으로 드러났다. 사용 가능한 심리음향모델은 기존의 MPEG 오디오의 모델 1과 2가 있으며[9], 모델2가 더 정교하고 정확한 매스킹 레벨을 산출하는 것으로 알려져 있으나 어느 방법이나 사용 가능하다. 계산량을 줄이기 위해 매스킹 레벨의 산출 과정을 위한 별도의 STFT 과정을 거치지 않고 이미 신호 분석 과정에서 얻어진 STFT 스펙트럼을 사용한다. 그런데 일반적으로 피크의 주파수 해상도를 높이기 위해서는 STFT에서 사용하는 FFT 길이가 윈도우 길이 보다 길어야만 한다. 특별히, 보간법을 이용하여 높은 주파수 해상도를 얻으려면 STFT에서 요구되는 FFT 길이는 윈도우 길이의 약 4배 정도가 적당하다. 이 값은 일반적으로 매스킹 레벨의 산출을 위해서 요구되는 FFT 길이 보다 훨씬 큰 값이 된다. 따라서 피크 검출에서 사용하는 STFT 스펙트럼과 심리음향모델에서 사용하는 FFT 스펙트럼 사이의 적절한 조절이 필요하다.

정현파 모델의 분석과 합성 과정에 중요한 또 하나의 단계는 피크의 생성-소멸 매칭 과정이다. 이 과정은 프레임 간 순음들을 적절하게 매칭시키고 연결시키는 과정으로 합성된 신호의 음질에 지대한 영향을 끼친다. 심리음향모델을 이용한 피크 검출은 불필요한 피크를 제거함으로써 피크 추적 과정을 단순화시켜준다. 피크의 수가 많아지면 생성-소멸 매칭 과정을 통한 피크 추적 과정은 잘못된 매칭을 결론지을 수 있으며, 결국 위상의 불연속으로 이어지며 경우에 따라서는 음질이 오히려 저하되는 결과를 야기할 수도 있기 때문이다. 그러나 피크 수의 감소만으로 피크의 생성-소멸 매칭 과정이 다 이루어지는 것은 아니다. 그림 2에서 볼 수 있듯 피크 생성-소멸 매칭 과정에서 중요한 요소는 생성, 소멸, 연속 여부를 판별하는 기준이 되는 주파수 범위를 설정하는 것이다. 앞에서 언급한 바와 같이, 사람은 신호의 중심 주파수가 500Hz를 넘게 되면 중심 주파수의 절대 변위에 따라 음의 변화를 감지하는 것이 아니라 주파수의 상대적 변위에 따라 반응한다. 따라서, 본 연구에서는 피크의 생성-소멸 매칭 과정에 절대적인 주파수 범위값을 사용하는 대신 JNDf를 활용하였다.

IV. 모의실험 결과

모의 실험을 통해 심리음향모델을 사용한 순음 성분 검출의 효과를 조사하였다. 심리음향모델로는 MPEG 오디오에서 사용되는 모델 1을 활용하였다. 입력 오디오 신호는 44.1 kHz로 샘플링 된 것이며 분석 윈도우의 길이는 2047 샘플, 합성 프레임의 길이는 512 샘플로 고정하였다. 그림 3은 클라리넷 소리에 대한 피크 검출 결과를 보여준다. 심리음향모델을 사용하지 않은 경우, 최대 피크 검출 수는 300개이며 이 중 합성에 사용된 정현파의 수는 80개로 제한되었다. 심리음향모델을 사용한 경우, 순음은 매스킹 레벨에 잡음의 매스킹 인덱스를 더한 문턱값 보다 큰 피크 만을 검출함으로써 추출되었다. 이 경우, 그림에서 볼 수 있듯 단지 12개의 피크만이 검출되었다. 본 논문에서 제안한 방법의 성능을 검토하기 위해, Thomson의 하모닉 분석 방법을 활용하여 순음을 검출한 결과와 비교하여 보았다. 그림에서 볼 수 있는 바와 같이 Thomson의 하모닉 분석은 작은 피크들까지 순음으로 검출하는 것을 보여준다. 사실 이들 피크들은 순음이 아닐 가능성이 높은 거짓 피크들이거나 또는 심리음향적으로 의미가 없는 순음에 속한다. 따라서 Thomson의 하모닉 분석 결과를 그대로 활용하는 것 보다는 매스킹 레벨을 적용하여 실제로 의미있는 순음에 해당하는 피크를 검출하는 것이 바람직하다. 매스킹 레벨을 적용하면 그림 3에서 볼 수 있듯 약 14개의 순음 만이 검출되는 것을 알 수 있다. 이 결과는 심리음향모델만을 사용한 결과와 거의 유사하다. 따라서, 5번의 FFT 계산을 필요로 하는 Thomson의 하모닉 분석에 비해 심리음향모델만을 고려한 방안이 훨씬 경제적임을 알 수 있다.

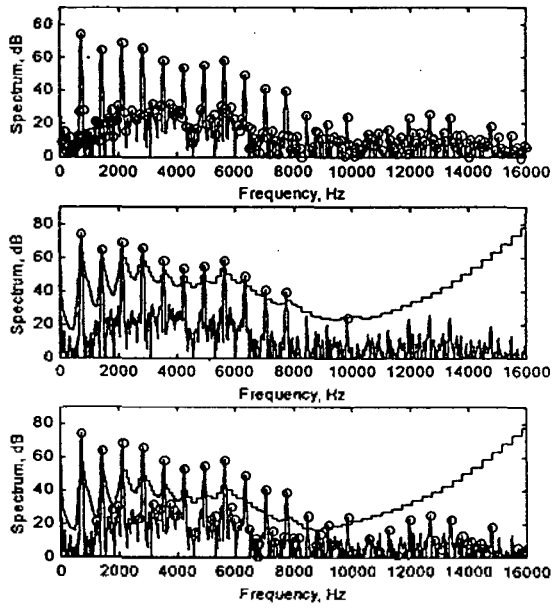

그림 3. 피크 검출 : 심리음향모델을 사용하지 않은 경우(위), 사용한 경우(중간), Thomson의 하모닉 분석 방식을 사용한 경우(아래)

Fig. 3. Peak detection : without (top) and with (middle) the psychoacoustic criterion, Thomson's harmonic analysis (bottom).

그림 4는 합성된 오디오 신호의 스펙트럼을 보여준다. 심리음향모델을 사용하지 않은 경우와 사용한 경우 합성 스펙트럼 성분 상의 차이는 심리음향적으로 거의 의미없는 영역의 것이다. 이 사실은 그림 5에 보여진 합성음의 파형을 비교함으로써 확인될 수 있다. 실제 청음으로도 두 합성음의 음질 차이를 거의 느낄 수 없었다.

그림 6은 앞에서 사용한 클라리넷 소리의 피크 주파수 추적 결과를 보여준다. 그림에서 알 수 있듯, 일반적인 오디오 신호에서는 순음의 주파수가 급격하게 변하지 않

는다. 따라서 차등부호화(differential coding)를 활용하여 양자화하면 아주 낮은 비트율에서도 고음질의 합성음을 얻을 수 있다. 이 점이 바로 정현파 모델이 오디오 부호화에 효과적인 이유 중 하나이다.

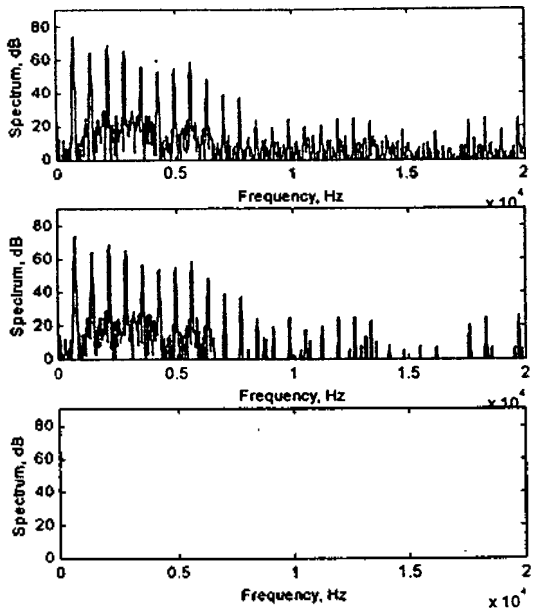

그림 4. 합성 오디오 신호의 스펙트럼: 원음(위), 심리음향모델을 사용하지 않은 경우(중간), 사용한 경우(아래)

Fig. 4. Spectrum of synthesized audio signals: original(top), synthesized without(top) and with(middle) the psychoacoustic criterion.

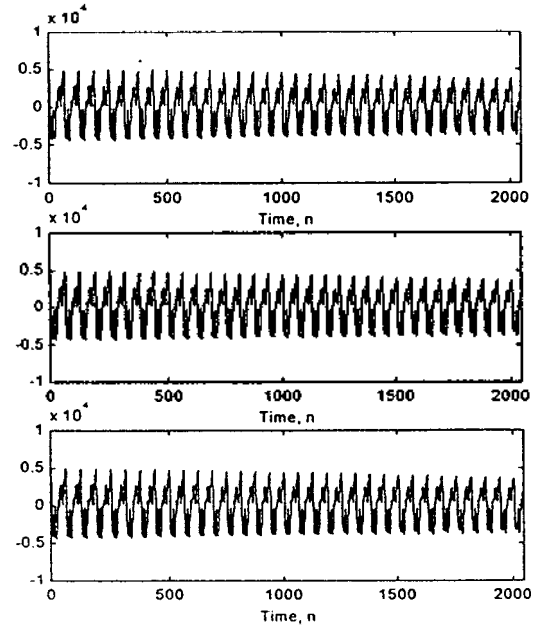

그림 5. 합성 오디오 신호의 파형 (피치가 일정한 부분): 원음(위), 심리음향모델을 사용하지 않은 경우(중간), 사용한 경우(아래)

Fig. 5. Waveforms of synthesized audio signals(stable pitch): original(top), synthesized without(top) and with(middle) the psychoacoustic criterion.

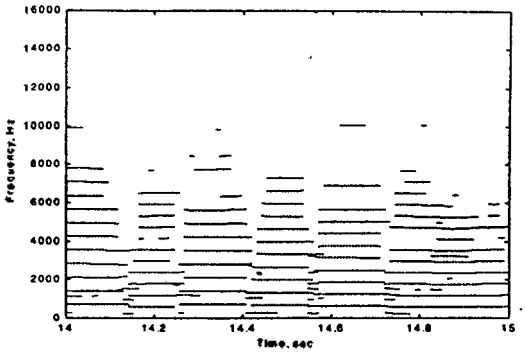

그림 6. 피크 주파수 추적(클라리넷)

Fig. 6. Trajectories of peak frequencies(Clarinets).

악기 음들은 일반적으로 많은 부분에서 정적인 성질을 나타내지만 한 음에서 다른 음으로 이동하는 기간에는 상당한 피치의 변화가 있다. 또 사람의 음성 신호의 경우 이러한 비 정적인 현상은 더욱 심하다. 그림 7은 음성 신호에서 피치가 변하는 부분에 대해 심리음향모델을 사용하지 않고 피크를 검출한 경우와 사용하여 검출한 결과 경우 각각의 합성음을 보여준다. 이 경우, 심리음향모델을 사용하지 않고 많은 수의 피크를 검출하여 합성한 것이 오히려 더 나쁜 결과를 얻는다는 사실을 알 수 있다. 이것은 앞에서 지적한 바와 같이 심리음향모델을 사용하지 않으므로 순음이 아닌 잡음이나 또는 분석 윈도우의 측엽(sidelobe)등이 순음으로 잘못 사용되어져 위상의 예측이 제대로 이루어지지 않았기 때문이다. 또한, 본 논문

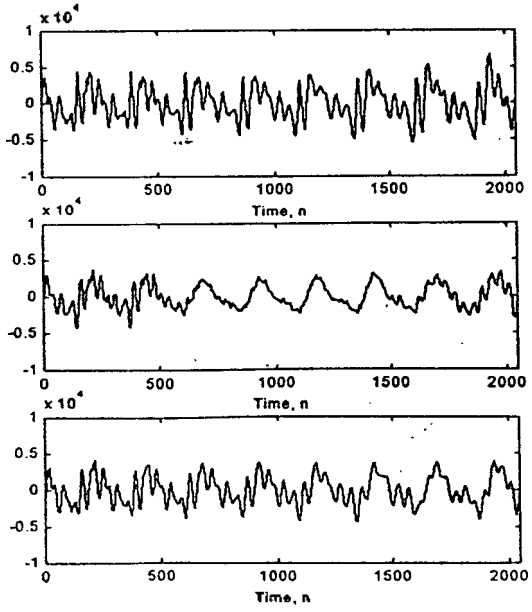

그림 7. 합성음의 파형(피치가 변하는 부분): 원음(위), 심리음향모델을 사용하지 않은 경우(중간), 사용한 경우(아래)

Fig. 7. Waveforms of synthesized audio signals(changing): original(top), synthesized without(top) and with(middle) the psychoacoustic criterion.

에서는 실험 결과가 생략되었으나 앞에서 언급한 바와 같이 피크의 생성-소멸 매칭 과정에 JNDf를 사용한 경우 피치가 변화하는 부분에서의 음질이 상당히 개선됨을 알 수 있었다. 따라서 심리음향적 기준들을 사용함으로써 잘못된 순음/잡음 판별로 인한 오차와 그로부터 발생하는 위상 예측으로 인한 오차를 줄일 수 있으며 효율적인 피크의 생성-소멸 매칭을 이룰 수 있다.

V. 결 론

본 논문에서는 정현파 모델을 이용한 오디오 신호 분석과 합성에 대해 살펴보고 심리음향모델을 활용한 순음 검출 방법과 JNDf를 이용한 피크의 생성-소멸 매칭 과정에 대해 살펴보았다. 모의실험 결과, 심리음향모델을 사용하면 합성음의 음질을 저하시키지 않으면서도 합성에 사용되는 순음의 개수를 현저하게 줄일 수 있었음을 알 수 있었다. 또한 피치가 변하는 부분에서 심리음향적 기준들을 사용하여 피크를 검출하고 추적할 경우 상당한 음질의 개선이 있었음을 확인할 수 있었다. 이것은 매스킹 레벨과 매스킹 인덱스를 이용한 심리음향적 피크 검출과 JNDf를 활용한 피크들의 생성-소멸 매칭이 효과적임을 확인시켜주는 것이다.

참 고 문 헌

1. K. Brandenburg and M. Bosi, "Overview of MPEG Audio: Current and Future Standards for Low-Bit Rate Audio Coding," *J. Audio Eng. Soc.*, Vol. 45, No. 12, Jan/Feb 1997.
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, pp. 744-754, Aug. 1986.
3. E. B. George and M. J. T. Smith, "Analysis/Synthesis Overlap-Add Sinusoidal Modeling Applied to the Analysis and Synthesis of Musical Tones," *J. of Audio Eng. Soc.*, Vol. 40, No. 6m pp. 497-516, June, 1992.
4. X. Serra, "Musical Sound Modeling with Sinusoids plus Noise," *Musical Signal Processing*, Swets & Zeitlinger Publisher, 1997.
5. E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Heidelberg, Germany, Springer-Verlag, 1990.
6. K. Hamdy, M. Ali, and A. Tewfik, "High Qaulity Audio Coding of Audio Signals with a Combined Harmonics and Wavelet Representation," *ICASSP-96*, Atlanta, GA.
7. ISO/IEC JTC1/SC29/WG11 MPEG, CD 14496-3 Subpart 2: Parametric Coding, 1997.
8. R. C. Maher and J. W. Beauchamp, "Fundamental frequency estimation of musical signals using a two-way mismatch procedure," Jour. of Acousti. Soc. of America 95(4), pp. 2254-2263, April 1994.
9. ISO/IEC JTC1/SC29/WG11 MPEG, International Standard IS-11172-3, Part 3: Audio, 1992.
10. Digital Voice Systems, "Inmarsat-M Voice Codec-Version 2," Inmarsat-M specs, Inmarsat, Feb. 1991.

▲남 승 현(Seung Hyon Nam)

1980년 2월 : 서강대학교 전자공학과 (학사)
1987년 8월 : The Univ. of Alabama, Huntsville, 전기 및 컴퓨터공학과(석사)
1992년 12월 : Texas A&M University, 전기공학과(박사)
1979년 12월~1985년 6월 : 국방과학연구소 연구원
1993년 3월 : 배재대학교
현재 : 전자공학과 부교수
※주관심분야 : 음성 및 오디오 부호화, 적응필터, 이동통신

▲강 경 옥(Kyeong Ok Kang) : 한국음향학회지 제16권 3E호

▲홍 진 우(Jin Woo Hong) : 한국음향학회지 제16권 3E호