

하모닉 보상방법에 기반한 파라메트릭 코덱 구현에 관한 연구

정종훈, 이남숙, 이권형
삼성전자 디지털 미디어 연구소

Novel harmonic coding method for parametric audio codec

Jonghoon Jeong, Namsuk Lee, Geonhyoung Lee
Digital Media R&D Center, Samsung Electronics

Abstract - 본 논문은 오디오 압축시 하모닉의 특성을 적용함으로써 신호의 압축률을 향상시킬 수 있도록 하는 내용을 기술하고 있다. 하모닉 코딩은 오디오 신호가 가지는 특징인 복합음(complex tone)의 특성을 이용하여, 주파수 공간에서 정수배의 주파수가 존재하며, 정현파의 특성상 시간적으로 인접 신호들간의 유사성이 매우 높은 특징을 이용하여 압축효율을 향상시키는 방법이다. 하지만 실질적인 오디오 신호의 경우, 악기들의 harmonic stretch, 전달과정에서 발생하는 신호의 왜곡, 외부 잡음등의 특성으로 인하여 수집된 오디오 신호를 분석하는 과정에서 부정확한 하모닉의 판단이 이루어질 가능성이 높으며, 이는 압축과정에서 심각한 음질의 열화를 가져오게 된다. 따라서 본 논문에서는 프레임간의 변화 추이의 판단을 통하여 하모닉의 변화를 예측하고, 예측 오류에 대한 보상값을 전달함으로써 오디오 신호의 안정적인 압축/복원이 가능하도록 하는 신호처리 방법에 대한 내용을 기술하고 있다.

1. 서 론

무어의 법칙(Moore's law)로 대변되는 디지털 기술의 발전은 상상 속에서만 존재하던 기술들을 하나둘씩 현실화 시켜가고 있다. 특히 초고속 통신망, 저장기술의 발달은 기존 정보들의 효율적인 소통을 가능케 하였을 뿐만 아니라, 데이터의 표현방식을 현실감 있는 멀티미디어 형태로 변화시켜가고 있다. 이러한 변화는 데이터량의 급속한 증가와 함께, 이를 효율적으로 표현하기 위한 신호의 압축이라는 기술개발의 필요성을 지속적으로 요구하게 되었으며, 수많은 연구개발 결과 오늘날 신호압축 기술은 우리의 실 생활속에서 없어서는 안될 중요한 분야로서 자리잡게 되었다.

지금까지의 압축기술의 발달은 전달된 신호의 왜곡을 최소화하며 가능한 원래의 신호를 그대로 표현하는데 중점을 두어왔다. 하지만 최근에는 신호의 발생과정의 분석을 통하여 이를 파라미터화 하여 압축효율을 증가시키려는 노력이 가시화 되고 있다.

오디오 압축에서는 이러한 파라메트릭 방식을 이용하여 낮은 bitrate하에서도 만족할만한 음질성능을 나타내기 위한 다양한 노력이 경주되고 있다. 그중 MPEG-4 Parametric audio coding은 대표적인 압축 방식의 하나로서, 오디오 신호의 분석을 통하여, 각 신호의 특성에 해당하는 파라미터를 추출/저장함으로써 고효율 압축 성능을 보여주고 있는 오디오 압축 방식이다.

본 논문에서는 오디오 신호의 특징중의 하나인 하모닉 특성을 추출하고, 이를 효율적으로 표시할 수 있는 신호 모델을 제시함으로써, 오디오 압축 성능의 향상을 가져올 수 있는 파라메트릭 모델을 제안하고자 한다.

2. 본 론

2.1 MPEG-4 Parametric audio coding

MPEG-4 Parametric audio coding은 저비트율 압축 코덱의 하나로서 오디오 또는 음성 신호를 3가지 파라미터(transient, sinusoid, noise)로 분리, 각각에 대하여 최적의 파라미터를 추출함으로써, 고효율 압축을 가능하도록 하는 신호처리 방법이다. 각 파라미터는 신호의 패턴에 따른 특성을 나타내기 위한 것들로서, transient는 급격한 변화를 포함하는 신호, sinusoid는 시간적으로 안정된 정현파의 표현시, 그리고 noise는 특정한 주파수를 가지지 않는 랜덤신호의 표현시 사용된다 [1][2]. 이들 파라미터를 이용하여 신호를 표현하는 방법은 식1에서 나타나고 있다.

$$x[n] = \sum transients + \sum sinusoids + \sum noise \quad \text{식1}$$

이들 파라미터중 가장 많은 비중을 차지하는 요소는 sinusoid 성분이다. Sinusoid는 정적이며, 많은 변화가 발생하지 않는 정현파를 파라미터화 하여 나타내며, 신호의 시작을 나타내는 위상성분, 신호의 높/낮이를 나타내는 주파수 성분, 크기를 나타내는 진폭으로 표현되며, 식2에서 이를 수식으로 나타내고 있다.

$$s[n] = \sum_p A_p(n) \cos(\omega_p(n)n + \phi_p) \quad \text{식2}$$

2.2 Harmonic coding

파라메트릭 코딩의 장점은 입력 신호의 특성을 적절히 반영할 수 있는 패턴을 적용할 시, 소수의 파라미터만으로도 신호의 효율적인 표현이 가능하다는데 있다. 이를 바꾸어 말하면 기존의 waveform 코딩과는 달리, 입력 신호를 효율적으로 표현할 수 있는 모델의 선정이 매우 중요함을 의미한다. 본 절에서 언급하고자 하는 하모닉 신호는 일반적인 자연음에서 흔히 발견되는 신호의 패턴으로서, 기본주파수를 가지는 정현파와 함께 기본 주파수의 정수배 주파수를 가지는 고조파로 구성된다. 특히 악기의 경우 본 하모닉 성분은 악기 고유의 음색을 결정짓는 매우 중요한 요소로서 취급된다. 따라서 본 논문에서는 종래의 파라메트릭 코덱에서 제공하는 3가지 신호모델과 함께 하모닉의 특성을 파라미터화 하여 코딩시 적용시킴으로써 성능 향상을 이루고자 한다. 이를 위하여 하모닉 신호의 특성을 주파수, 시간적인 측면에서의 분석을 통하여 신호모델로서 활용할 수 있는 방안을 강구하고자 한다.

우선 주파수 측면에서의 특성을 살펴보면, 앞서 언급한바와 같이 하모닉 신호는 기본주파수와 고조파의 조합으로서 구성된다. 따라서 정확한 기본 주파수의 추출이 이루어질 경우 나머지 고조파들이 가지는 주파수는 예측이 가능함을 의미한다. 이는 정현파 모델을 이용하여 고조파를 코딩할 시 포함되어야 하는 주파수 성분에 대한 파라미터를 생략할 수 있음으로 인하여 코딩 효율의 증가를 가져올 수 있게 됨을 의미한다.

두 번째 시간 측면에서의 특성을 살펴보도록 하자. 하모닉은 정현파들의 조합으로 구성된다. 이는 정현파의 특성을 그대로 수용함을 의미하며, 일정 시간동안 신호의 특성이 안정적임을 의미한다.

따라서 본 논문에서는 정현파의 시/공간적인 반복성 및 유사성을 이용하여 예측을 수행함으로써 압축효율을 향상시키고자 한다.

2.2.1 Harmonic encoding

하모닉 예측을 위해서는 먼저 프레임 데이터를 구성하고 있는 정현파들의 추출과, 하모닉의 기본 주파수의 판단이 선행되어야 한다. 만일 적절한 정보들이 추출되었을 경우, 다음 네 단계의 연산을 수행하게 된다.

첫 번째 단계는 추출된 정현파들 중 고조파에 해당하는 성분들을 선정하는 단계이다. 이때 신호에 포함된 잡음 및 정현파의 추출시 발생가능한 오차등을 고려하여 일정 범위 이내 존재하는 신호들을 정현파로서 판단한다.

$$f_{n,prev_est} = n \cdot f_{0,prev_real} \quad \text{식3}$$

$$\text{if} \left(\left| f_{n,prev_est} - f_{n,prev_real} \right| < \alpha \right) \text{flag}_n = \text{harmonic} \\ \text{else} \quad \text{flag}_n = \text{sinusoid} \quad \text{식4}$$

두 번째 단계는 기존 주파수의 변화를 판별하는 단계이다. 정현파 성분의 경우 일정시간동안 안정적임 특징이 존재하나, 악기의 연주 방법 등에 의하여 어느 정도의 주파수 변화가 존재하게 된다. 따라서 인접 frame간의 기본 주파수의 변화를 판별한 뒤 이를 시간 영역에서의 예측에 반영하도록 한다.

$$\Delta f_0 = f_{0,curr_real} - f_{0,prev_real} \quad \text{식5}$$

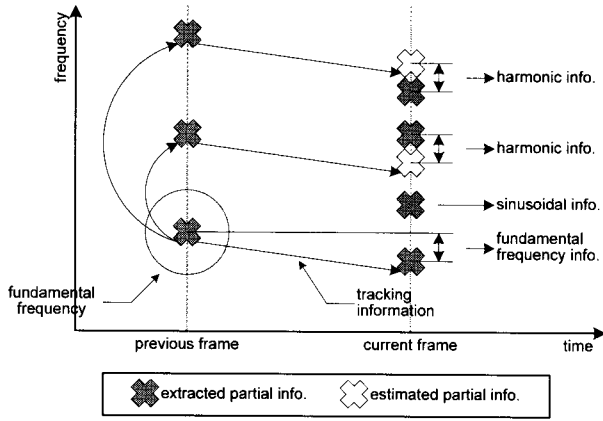
세 번째 단계는 첫 번째 단계에서 하모닉 성분으로 판별된 신호들에 대하여 적용되는 단계로서, 시간 영역에서의 하모닉 예측 과정이다. 하모닉의 예측은 측정된 기본 주파수의 변화량을 하모닉으로 선정된 정현파에 적용함으로써 수행된다.

$$f_{n,curr_est} = f_{n,prev_real} + n \cdot \Delta f_0 \quad \text{식6}$$

마지막 단계는 실제 측정된 정현파의 주파수와, 예측에 의하여 산출된 주파수와의 차분 신호를 구하는 단계로서, 본 과정에서는 기존의 파라메트릭 코딩에서 지원하는 tracking과정이 필요하게 되며 해당 과정은 MPEG-4 Parametric audio coding 문서를 참조하기 바란다.

$$\text{if} (\text{flag}_n == \text{harmonic}) f_{n,diff_code} = f_{n,curr_real} - f_{n,curr_est} \\ \text{else} \quad f_{n,diff_code} = f_{n,curr_real} \quad \text{식7}$$

상기 네 단계를 거쳐 생성된 잔차 성분은 기존의 sinusoid 모델만을 이용하여 추출된 방법에 비하여 변위에 대한 절대값의 감소를 가져올 수 있으며, 이는 0에 근접한 데이터의 확률분포를 증가시킴으로써 압축률의 향상을 가져올 수 있게 된다 [3]. 그림 1에서는 이러한 인코딩 과정을 도식화하여 나타내고 있다. 회색으로 표시된 마크는 정현파로부터 추출된 sinusoid 성분을 나타내며, 현재 프레임의 성분중 흰색으로 표시된 마크는 시간 영역의 예측을 통하여 하모닉으로 판단된 정현파를 의미한다.



〈그림 1〉 주파수/시간 영역에서의 하모닉 예측과정

2.2.1 Harmonic decoding

하모닉 디코딩 과정은 인코딩 과정의 역으로서, 프레임 신호들간의 연결 관계 및 유사성의 판단에 의하여 각 신호의 종류를 판별하고, 단일 판단된 신호가 하모닉의 경우 예측된 신호성분과 전달된 잔차성분간의 합에 의하여 신호를 복원하게 되며, 단일 사인파로 판별하게 될 경우 기존의 파라메트릭 디코딩에서 사용하던 sinusoid 복원을 수행하게 된다.

디코딩 과정 또한 인코더와 마찬가지로 크게 네 단계의 연산을 수행하게 된다.

그중 첫 번째 단계는 기본 주파수의 복원 단계이다. 전달된 기본 주파수의 잔차 성분을 이전 프레임의 기본주파수에 적용시킴으로써 원래의 기본 주파수를 획득하게 된다.

$$f_{0,curr_real} = f_{0,prev_real} + \Delta f_0 \quad \text{식8}$$

두 번째 단계는 이미 복원된 이전 프레임들의 성분들을 대상으로 각 정현파 성분이 고조파에 해당하하는지를 판별하는 단계이다.

$$f_{n,prev_est} = n \cdot f_{0,prev_real} \quad \text{식9}$$

$$\begin{aligned} & \text{if } (|f_{n,prev_est} - f_{n,prev_real}| < \alpha) \quad \text{flag}_n = \text{harmonic} \\ & \text{else} \quad \text{flag}_n = \text{sinusoid} \end{aligned} \quad \text{식10}$$

세 번째 단계는 기본 주파수의 변화량을 기반으로 현재 프레임상의 하모닉을 예측하는 단계로서, 그 연산과정은 인코더와 동일하다.

$$f_{n,curr_est} = f_{n,prev_real} + n \cdot \Delta f_0 \quad \text{식11}$$

네 번째 단계는 상기 연산과정으로부터 추출된 각 신호들의 성분 및 현재 프레임에 의하여 전달된 신호성분들을 이용하여 원래의 신호를 복원하는 단계이다.

$$\begin{aligned} & \text{if } (\text{flag}_n == \text{harmonic}) \quad f_{n,curr_real} = f_{n,curr_est} + f_{n,diff_code} \\ & \text{else} \quad f_{n,curr_real} = f_{n,diff_code} \end{aligned} \quad \text{식12}$$

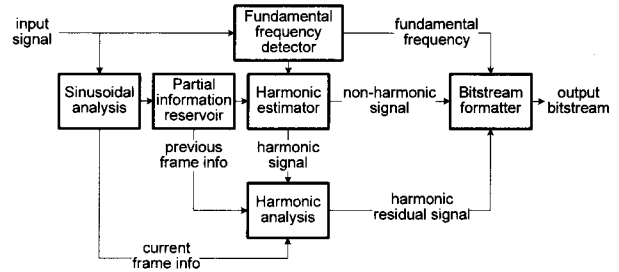
2.2.3 시스템 구성

본 절에서는 하모닉 신호 모델링을 위한 시스템의 전반적인 구성에 대하여 알아보려 한다. 하모닉 코딩 과정은 파라메트릭 코덱의 다른 신호 모델(transient, noise)과 달리, sinusoid 모델과 유기적으로 동작을 수행한다. 이는 하모닉 신호가 정현파의 부분집합인 관계로, sinusoid 모델과 독립적으로 존재할 수 없기 때문이다.

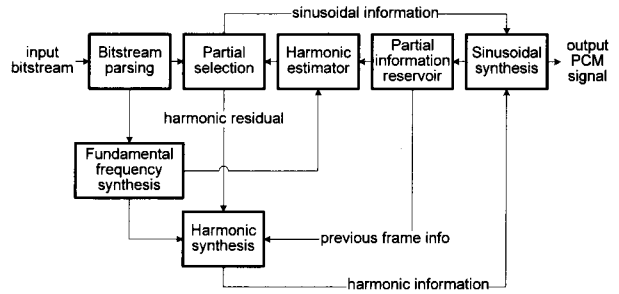
본 논문에서 제시하는 알고리즘의 특징은 sinusoid 성분과 하모닉의 성분의 판단을 위한 별도의 부가정보가 존재하지 않는다는 점이다. 저 비트율 압축 코덱의 경우, 신호의 표현을 위하여 가능한 bit이 매우 제한적이다. 따라서 알고리즘의 우수성에도 불구하고 각 경우에 최적화된 알고리즘을 선

택을 위하여 부가정보가 포함될 시, 최적화된 알고리즘에 의한 이득을 부가정보의 저장을 위한 공간을 위하여 상당부분 손실하는 경우가 발생하게 된다. 따라서 부가정보의 저장을 위한 bit을 할당하지 않으며, 연산 과정에서 해당 신호의 특성을 판단할 수 있는 알고리즘은 그 효과가 매우 크다고 할 수 있다.

연산과정은 대부분 간단한 산술 연산 및 비교에 의한 신호의 특성 판단으로서 구성된다. 또한 앞서 인코더 및 디코더의 연산 과정의 설명과정에서 나타내는바와 같이 두 모듈의 상당부분의 알고리즘이 동일함을 알 수 있다. 이는 실제 구현과정에서의 복잡도가 매우 낮음을 의미하며, 동일한 연산 블록을 공유함으로써 전체적인 시스템의 복잡도를 크게 낮출 수 있는 장점이 존재한다. 그림 2와 그림 3은 이러한 인코더와 디코더의 블록 다이어그램을 나타내고 있다.



〈그림 2〉 Harmonic encoder의 구성도



〈그림 3〉 Harmonic decoder의 구성도

3. 결 론

본 논문에서는 기존의 MPEG-4 Parametric audio coding을 기반으로, 하모닉 성분에 대한 신호모델을 추가시킴으로서 연산 효율을 증가시킬 수 있는 방안에 대하여 기술하였다. 파라메트릭 연산은 종래 음성코덱에 사용하던 알고리즘의 일종으로서, 성능의 우수성으로 인하여 대부분의 음성 코덱의 주요 알고리즘으로 적용되고 있다. 또한 오디오 코덱 분야에서도 HE-AAC의 고주파 코딩방법(spectral band replication), 스테레오 신호의 부호화 방법(parametric stereo)등에서 획기적으로 비트율을 감소시킴은 물론, 음질적인 측면에서도 매우 우수한 성능을 나타냄으로 인하여 그 적용범위가 점점 확대되어 가고 있다.

본 논문에서는 오디오 주요한 특성중의 하나인 하모닉 성분에 대한 신호 모델을 제시함으로써 기존의 파라메트릭 코딩 성능 향상을 목적으로 하고 있다. 또한 연산의 효율성 및 각 모듈간의 유기적인 동작을 수행하도록 함으로써 부가정보의 필요성을 배제하여 연산의 효율성 및 비트스트림의 효율성을 향상시킬 수 있도록 하였다.

[참 고 문 헌]

- [1] M. G. Muzzi, "Improvement of the audio quality of a parametric audio coder," Master's thesis, IRCAM, Paris, 2003
- [2] ISO/IEC JTC1/SC29/WG11/N6130 "Text of ISO/IEC 14496-3:2001/FDAM2 (Parametric coding for high quality audio), ISO/IEC, 2003
- [3] Richard W. Hamming, "Coding and information theory", Prentice-hall, pp 103-137, 1986