

# HE-AAC 부호화의 저비트율에서 음질향상 기법

## Quality Improvement of Low-Bitrate HE-AAC Encoder

김 정 근\*, 이 재 성\*, 박 영 철\*\*, 이 태 진\*\*\*, 강 경 옥\*\*\*

(Jeong-Geun Kim\*, Jae-Seong Lee\*, Young-Cheol Park\*\*, Taejin Lee\*\*\*, Kyeongok Kang\*\*\*)

\*연세대학교 전기전자공학부 디지털신호처리연구실, \*\*연세대학교 컴퓨터정보통신공학부, \*\*\*전지통신연구원 방송미디어연구그룹

(접수일자: 2007년 12월 18일; 채택일자: 2008년 2월 18일)

본 논문에서는 HE-AAC (High Efficiency Advanced Audio Coding) 오디오 부호화기의 저주파 대역과 고주파 대역을 담당하고 있는 AAC 부호화기와 SBR (Spectral Band Replication) 부호화기에 대해 낮은 비트율에서 효과적인 개선 방법을 제안한다. AAC 부호화기가 담당하는 저주파 대역에서 과도신호가 발생하는 부분의 프리에코를 줄이기 위하여 적용 주파수범위가 저주파 대역 방향으로 선택적으로 확장되는 sTNS (extended Temporal Noise Shaping) 방법을 고안하였다. 또한 SBR에 의해 부호화되는 고주파 대역에서 톤 성분 복원 시에 잡음층 (Noise floor)이 추가 발생되지 않도록 정현파 모델을 통해 톤을 사전 인지하고 인지된 톤들의 주파수를 QMF 밴드의 중앙으로 재배치하여 성능 향상을 이루었다. 제안한 방법들을 사용하여 부호화한 샘플 음원들에 대해 주/객관적 음질평가를 실시한 결과, 표준 HE-AAC에 비해 향상된 결과를 보여주었다.

**핵심용어:** HE-AAC, TNS, 정현파 모델, AAC, SBR, 프리에코

**투고분야:** 음향신호처리 분야 (1,2)

In this paper, we propose new techniques that can improve the quality of AAC and SBR encoders comprised in low bitrate HE-AAC. To reduce the pre-echo artifacts often occurring for transient blocks in AAC, we propose an extended Temporal Noise Shaping (sTNS) in which the frequency range is selectively extended down to the low-frequency region. Also, for the high-frequency region being coded by SBR encoder, tones are identified through a sinusoidal modeling and their frequencies are adjusted within the QMF band in order to reduce the noise floor due to aliasing. Spectrograms of the decoded signals were compared and listening tests were conducted to evaluate the proposed algorithm. Results confirmed the effectiveness of the proposed algorithm.

**Keywords:** HE-AAC, TNS, Sinusoidal model, AAC, SBR, Pre-echo, Musical noise

**ASK subject classification:** Acoustic Signal Processing (1,2)

### 1. 서론

최근 10여년 동안 발달된 디지털 오디오 압축 기술은 음악을 즐기는 환경과 매체의 급격한 변화를 가져왔다. 오디오 저장 매체는 CD (Compact Disc)와 MD (Mini Disc)를 거쳐, 플래시 메모리가 대중화 되기에 이르렀으며, 더 나아가 무선 환경의 구현과 통신 환경의 발달을 위해, 비트율이 낮은 환경에서도 더 나은 음질이 보장될 수 있도록 지속적인 연구가 진행되고 있다. 낮은 비트율에서 동일한 음질을 보장하기 위한 노력은 HE-AAC (High Efficiency Advanced Audio Codec) [1] 라는 새로운 개념의 오디오 압축기술을 탄생시켰다.

HE-AAC는 일종의 BWE (Bandwidth Extension) 기법인 SBR (Spectral Band Replication) [2,3]와 MPEG-4 AAC (Advanced Audio Coding) [4]를 결합하여 만들어진 기술로서 기존 AAC 부호화기 대비 30% 이상의 부호화 효율을 향상시켰다. 현재 HE-AAC는 3G mobile 음악, A/V 서비스, 인터넷 오디오, 디지털 방송, IPTV 등 다양한 분야에 응용되고 있다.

HE-AAC가 AAC에 비해 높은 효율을 보이고 있지만 극도로 낮은 비트율 (<24 kbps)에서는 성능의 한계를 보인다 [5]. 낮은 비트율에서 동작하는 AAC 부호화기에서 발생하는 음질의 열화는 원도우 스윕칭을 제 배에 해주지 못해 발생하는 프리에코와 비트 할당이 제대로 이뤄지지 못해 발생하는 스펙트럼 홀에 의한 뮤지컬 잡음이다. 이러한 문제는 AAC와 같은 변환부호화기의 고유한 특성이나 비트율이 낮아지면 특히 왜곡이 심해져

책임저자: 박 영 철 (young00@younsei.ac.kr)  
220-710 강원도 원주시 흥업면 매지리 234  
연세대학교 정보기술학부  
(전화: 033-760-2744; 팩스: 033-763-4323)

음질을 심각하게 손상시킨다. 뮤지컬 잡음은 Spectral-Hole Avoidance 기법을 사용함으로써 상당 부분을 해결할 수 있다 [10]. 그러나 원활하지 않은 윈도우 스위칭에 의한 프리에코 문제는 TNS (Temporal Noise Shaping) [4] 기법을 적용하더라도 해결하지 못하는 경우가 자주 발생하는데, 특히 저주파 대역에서 발생하는 프리에코가 제거되지 못함으로써 심한 가청 왜곡이 발생한다 [6]. 본 연구에서는 낮은 비트율의 HE-AAC에서 발생하는 프리에코를 개선하기 위해 저주파 대역까지 확장된 eTNS (extended TNS)를 제안한다. 제안된 eTNS는 기존 TNS가 적용되는 주파수 범위를 저주파 신호 포락선의 에너지 변화에 따라서 선택적으로 확장함으로써 저주파 영역의 프리에코에 의한 왜곡 현상을 최소화하였다.

SBR은 신호의 고주파와 저주파의 연관성을 이용한 새로운 개념의 압축이다. 따라서 축력 신호의 왜곡 현상 또한 그 성향이 기존과는 상이하다 [7]. SBR 과정에서 발생하는 혼한 왜곡 중 하나가 유실 하모닉 성분을 복원해 주는 과정에서 대역간의 스펙트럼 에일리어징에 의한 잡음이다. 본 논문에서는 정현파 모델 분석을 통하여 왜곡이 발생할 수 있는 톤 성분을 조정함으로써 이러한 왜곡을 제한하는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 개선 방법을 설명하기 전에 필요한 배경지식을 미리 설명하고, 문제가 발생하는 실제 예를 통해 원인과 현상을 파악한다. 3장에서는 이를 개선하기 위한 각각의 방법을 제안한다. 4장에서는 제안한 기법을 적용하여, 주/객관적인 실험을 수행하고 결과를 분석한다. 마지막으로 5장에서는 전체적인 내용을 정리하며 결론을 맺는다.

## II. HE-AAC

그림 1은 HE-AAC 부호화기 블록도를 나타낸 것이다. 전체적인 부호화의 순서를 살펴보면 HE-AAC는

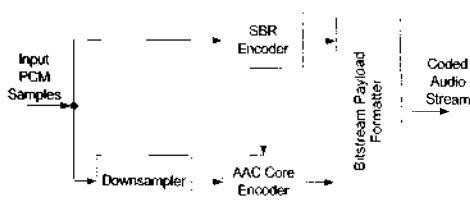


그림 1. HE-AAC 부호화기 블록도  
fig. 1. HE-AAC Encoder block diagram.

SBR 부호화를 먼저 실행하는데, SBR 부호화 시에 사용된 입력 샘플은 2:1 다운샘플링을 거쳐 AAC 블록으로 입력된다. SBR과 AAC 부호화기를 거친 각각의 주파수 대역 별 신호는 마지막으로 비트스트림으로 만들어져 전송된다 [8,9].

### 2.1. TNS (Temporal Noise Shaping)

TNS는 입력신호의 과도 (transient) 현상이 있는 부분을 다루기 위해 AAC에 포함된 툴이다 [4].

TNS 알고리즘은 시간 영역 상에 넓게 퍼져 있는 양자화 잡음의 모양을 조절하여, 시간영역에서의 마스킹 효과를 만족시키도록 만드는 것이다.

3GPP HE-AAC 부호화기에 적용된 TNS를 확인해보면 [10], 첫 단계로 TNS를 적용할지 결정하는 TNS 감지 블록이 실행된다. 다음으로, 정규화된 스펙트럼에 스무딩 필터를 적용하는 것으로 LPC (Linear predictive coding) 분석을 위한 준비를 마친다. 스무딩 필터링은 LPC 중단 주파수부터 시행 주파수까지 아래의 과정으로 필터링 다운을 실행하고 반대로 LPC 시행 주파수부터 중단 주파수까지 필터링 업을 실행한다. 다음 단계로는 Levinson -Durbin 알고리즘을 사용하여 자기 상관 함수와 LPC를 계산하고, 그 결과로 PARCOR 계수를 구하고 예측 이득을 산출해 낸다. 산출된 예측 이득을 통하여 TNS를 적용여부를 결정한다. 마지막 단계인 TNS 부호화 블록에서는, PARCOR 계수의 높은 차수부터 양자화를 시험적으로 진행하여 임계치인 0.1보다 작은 값을 가지지 않게 되는 최초 계수의 차수로 TNS 차수를 결정하게 된다. 이렇게 결정된 차수와 계수를 가진 TNS 필터를 거쳐 MDCT (Modified Discrete Cosine Transform) 스펙트럼 계수에 적용시키고, 적용된 MDCT 스펙트럼 계수를 기준으로 이후의 AAC 부호화 작업을 진행하게 된다.

### 2.2. SBR을 이용한 고주파 대역의 톤 신호 복원

SBR 부호화기의 블록도는 그림 2와 같다. 가장 먼저, 입력 신호를 64채널 복소수 QMF로 분해하는데, 여기서 1 프레임 당 2048개의 샘플은 32 (시간축) x 64 (주파수축) 구조의 T/F (시간/주파수) 단위로 분해되며, 과도 위치에 따라 주파수와 시간 간격을 가변하며 특정한 영역을 형성하게 된다. 이 구간들은 부호화 과정에서 원본신호와 복원될 신호의 에너지와 조성을 비교하는 영역의 기준이 되며, 에너지 보정을 위한 주파수 대역과 조성 보정을 위한 잡음층 대역 (Noisefloor Band)의 추가적인 조정과

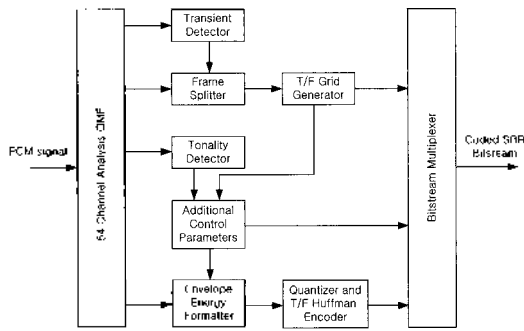


그림 2. HE-AAC의 SBR 부호화기 블록도  
fig. 2. SBR Encoder of HE-AAC block diagram.

라메터를 적용하는 기본 단위를 정하는 것에도 이용된다.

영역이 정해졌으면 에너지 보정을 위한 스케일 팩터 값은 포락선 에너지라는 위에서 정한 영역 내의 QMF 평균 에너지를 산출하고, 조성은 톤알리티 (tonality) 인식 블록을 통해 추가될 톤/노이즈량을 판단하게 된다. 이렇게 산출된 정보들은 비트스트림 형식으로 변형되고, 이 정보량을 AAC 부호화기 블록에 전달해주어 정해진 비트율의 나머지 비트량에 맞추어 AAC 코어의 비트량을 조절할 수 있도록 한다.

### III. 낮은 비트율 HE-AAC 음질향상 기법

#### 3.1. 주파수 대역이 확장된 eTNS (extended TNS)

저주파 대역을 담당하는 AAC 부호화기는 주파수 계수를 정해진 비트율에 따라 손실 압축하기 때문에, 낮은 비트율일수록 양자화 잡음이 크게 나타나게 되며, 이는 곧 음질 저하의 결과로 이어진다.

양자화 잡음은 프리에코를 발생시키게 되는데, 이를 줄이기 위해, 오디오 부호화기들은 블록 전환 (block switching) 기법을 사용한다. 하지만, 블록 전환을 사용하는 것이 안정된 신호와 낮은 비트율에서는 오히려 역효과를 가져오기도 한다. 즉, 낮은 비트율에서는 단구간 블록 대신 장구간 블록이 사용되면, 프리에코가, 단구간 블록이 과도하게 사용되면 뮤지컬노이즈가 생긴다. 결국 낮은 비트율에서 과도 신호를 효과적으로 나타내기 위해서는 블록 전환 외에 다른 방법이 필요하며, 이를 해결하기 위해, 본 논문에서는 장구간 블록만을 사용하면서, 기존의 TNS의 사용을 확장하도록 한다.

TNS가 적용되고 있는 HE-AAC에서도 프리에코는 발생하는데, 이는 TNS의 적용 주파수 대역이 장구간 블록의 경우, 1800 Hz에서부터, 단구간 블록의 경우에는

2750 Hz부터 SBR이 시작되는 경계 주파수까지 적용되고 있으므로, 이 주파수 대역보다 낮은 대역에서 프리에코가 발생하는 것이다.

만약 TNS 적용 주파수를 낮은 주파수까지 확장하여 적용하게 되면, 프리에코는 향상될 수 있지만 확장된 주파수 대역의 톤 성분을 왜곡시키게 된다. 때문에, 제안하는 방법에서는 상황에 따라 일반적인 TNS와 확장된 TNS를 적절히 선택하여 사용하며, 이를 판단하는 기준으로는 확장되는 대역에서 저주파 대역만의 과도 정보를 참고하여 결정한다. 이 때, 확장되는 저주파 대역의 과도 정보를 파악하기 위해, AAC 코어블록의 과도 검출 (Transient detection) 알고리즘을 그대로 사용한다.

AAC 코어블록에서 사용되는 과도 검출은 해당 프레임의 장구간 블록 처리할 것인지, 단구간 블록 처리할 것인지를 결정하는 것으로 신호를 입력하면, 에너지의 급격한 상승 (Attack)이 있는지의 여부와 상승 위치 (Attack index)를 판단해 준다. 이는 해당 프레임을 8개의 구간으로 나누어 각 구간별 에너지가 이전 구간까지의 누적 에너지에 비해 급격히 증가하는지를 확인하는 것으로 시간영역에서 이루어진다.

원래 HE-AAC 부호화의 과도 검출 알고리즘에서는 1차 버터워즈 고주파 통과 필터를 통과한 신호를 대상으로 판단하지만, 제안된 방법에서는 확장 대역만의 주파수 성분만을 판단하기 위해, 고주파 통과 필터와는 별도로 저주파 통과 필터도 미리 사용하였다. 다음의 식 (1)은 HE-AAC AAC 코어의 블록 전환을 위해 사용되는 고역 통과 필터의 수식이다.

$$H(z) = \frac{0.7548 \cdot (z-1)}{z-0.5095} \quad (1)$$

본 논문에서 제안하는 방법은 확장 TNS 적용 여부를 판단하기 위해, 1 kHz 미만의 대역에서 과도현상이 있는지의 정보가 필요하다. 그러므로 1차 버터워즈 저역 통과 필터인 식 (2)을 사용했다.

$$H(z) = \frac{0.2452 \cdot (z+1)}{z-0.5095} \quad (2)$$

위의 저역통과필터가 적용되고 난 뒤, 원본 신호의 해당 프레임이 TNS를 적용하기에 적합하다고 판단되며, 저주파 필터를 거친 신호가 과도하다고 판단되면 확장 TNS를 적용시킨다.

프리에코의 영향이 미치는 시간은 해당 윈도우 길이와 같은데, 에너지가 증가하는 과도한 부분이 해당 프레

입의 앞부분에 위치할수록 프리에코가 발생하는 범위는 좁아지고, 그 반대로 해당 프레임의 끝부분에 위치할수록 프리에코의 발생 범위는 넓어지게 된다. 이를 그림 3에 나타내었다. 그러므로 과도한 부분이 프레임의 뒤쪽에 위치한 경우, 프리에코가 미치게 되는 영향이 크다고 판단하여 확장 TNS를 적용하는 조건을 추가하였다.

과도한 부분의 위치는 블록 전환에서 결정되는 과도 위치 (Attack index)를 참고한다. 과도 위치는 0에서 7까지로 판단되며, 각 구간은 해당 프레임을 8개로 나눈 지점을 나타낸다. 이 과도 위치는 HE-AAC에서 효

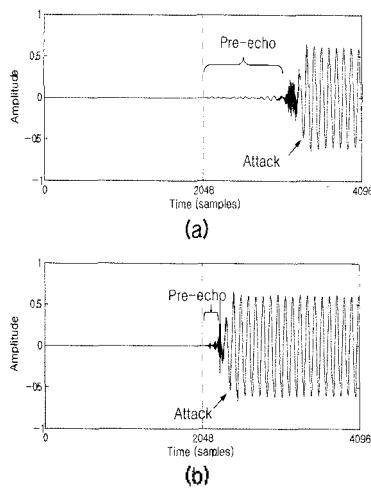


그림 3. 과도 위치에 따른 프리에코의 정도 차이. (a) 과도 부분이 프레임의 뒷부분에 있을 경우 (Attack index=3), (b) 과도 부분이 프레임의 앞부분에 있을 경우 (Attack index=1)  
 fig. 3. Pre-echo level difference for different transient position (a) Case of back transient position of frame (Attack index 3) (b) Case of front transient position of frame(Attack index 1).

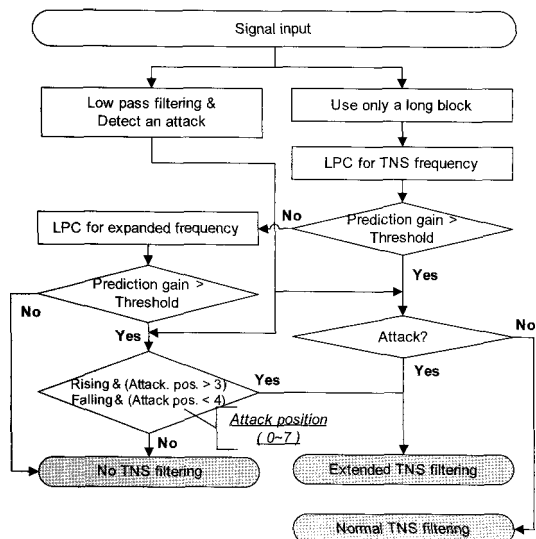


그림 4. 확장 TNS의 순서도  
 fig. 4. Flow chart of Extended TNS.

율적인 코딩을 위해 사용되는 값으로, 단구간 블록 적용 시에 8개의 단구간 블록을 4개의 그룹으로 묶을 때 참조하는 값이다.

확장 TNS에 참고할 값으로는 저주파 통과 필터를 거친 신호의 과도 위치를 참고로 하였다. 즉, 100 Hz까지 확장된 대역의 예측 이득 값이 임계치를 넘고, 에너지가 증가되는 과도 위치가 4에서 7사이의 값을 가질 경우에 확장 TNS를 적용하도록 한다. 반대로 에너지가 감소하는 과도 신호의 경우에는 Index 값이 0에서 3사이의 값을 가질 때, 확장 TNS를 적용하도록 한다.

그림 4에 지금까지 설명한 확장 TNS의 순서도를 나타내었다. 여기에서 확장 TNS 과정 중간에 마스킹 임계치를 재조정하지 않는 과정이 추가되었다. 마스킹 임계치 재조정 과정이란 TNS가 적용되지 않는 주파수 구간에 대해서는 마스킹 임계치를 낮추어 주어진 비트를 효율적으로 사용될 수 있도록 하는 것인데 [10], 확장 TNS를 사용하게 되면 TNS가 적용되지 않는 주파수 대역은 거의 없기 때문에, 마스킹 임계치를 낮추는 과정을 생략한 것이다.

### 3.2. 정현파 모델을 사용한 고주파 대역 톤 이동 기법

SBR 압고리즘은 저주파 대역을 복사하여 고주파 대역을 형성하는 것이다. 만일 고주파 대역의 톤은 유실되는 하모닉 (Missing Harmonic)이 발생하면, SBR 부호화 과정에서는 이렇게 유실되는 톤 성분을 미리 감지하여, SBR 복호화 과정에서 해당 주파수 대역에 톤신호를 삽입하는 톤신호의 주파수는 표 1과 같다.

여기서  $F_s$ 는 원본 신호의 샘플링 주파수,  $k$ 는 유실되는 하모닉이 감지된 주파수 대역 번호이며,  $f(k)$ 는  $k$ 번째 주파수 대역 내에서 가장 낮은 QMF 대역을 나타낸 것이다. 즉, 현재의 주파수 대역이 홀수 개의 QMF 대역으로 구성되어 있다면, 주파수밴드의 중앙 주파수에 톤 성분을 추가하며, 짝수 개로 구성되어 있다면 중앙에서 0.5 QMF 대역만큼 벗어난 위치에 톤 성분이 추가되는 것이다.

표 1. 유실되는 하모닉의 생성 주파수  
 Table 1. The generation frequency for loss harmonic.

$f(k+1) - f(k)$ (QMF 서브밴드 개수)	유실되는 하모닉의 생성 위치 (Hz)
Odd	$F_s \cdot \frac{(f(k+1)+f(k))}{256}$
Even	$F_s \cdot \frac{(f(k+1)+f(k)+1)}{256}$

하지만, 이런 유실되는 하모닉이 발생될 경우는 실제 신호에서 자주 일어나지 않는다. 입력 신호가 고주파 대역과 저주파 대역의 연관성이 적은 경우가 이에 해당하며, HE-AAC의 취약한 부분이기도 하다. 본 논문은 이러한 취약한 부분에 대해 개선의 여지를 살펴보고자 하는 것에 목적이 있다.

SBR 알고리즘은 고주파 대역을 포락선 스케일팩터 (Envelope scalefactor) 단위로 에너지를 산출하여, 부호화와 복호화시에 이 영역단위로 에너지와 조성을 고려하여 복원하게 된다. 그러므로 신호를 구성하는 한 성분의 에너지가 이웃하는 주파수 밴드간에 나뉘어져 있다면, 복원과정에서 심각한 왜곡이 생기게 된다.

그림 5는 디코더에서 SBR 알고리즘을 이용하여 복원된 신호의 스펙트로그램이다. 그림 5의 (a)에서 볼 수 있듯이, 48 kHz로 샘플링된 신호에서  $k=17$ 인 18번째 QMF 서브밴드의 주파수 대역은 6375 Hz부터 6749 Hz까지 해당하는데, 그 중앙에 해당하는 6562 Hz의 주파수를 가지고 있는 정현파는 이웃하는 QMF 서브밴드에 에너지가 나뉘지 않지만, 그림 5의 (b)와 같이 중앙에서 벗어난 6700 Hz의 정현파는 19번째 QMF 서브밴드에 에너지가 나뉘게 된다. 그러므로 이웃한 19번째 밴드에선 잡음층 스펙트럼을 형성하여 심한 잡음을 야기시킨다. 그림 6은 이러한 문제가 발생하는 과정을 나타낸

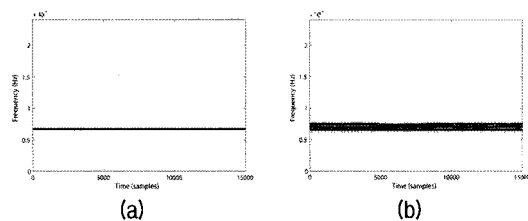
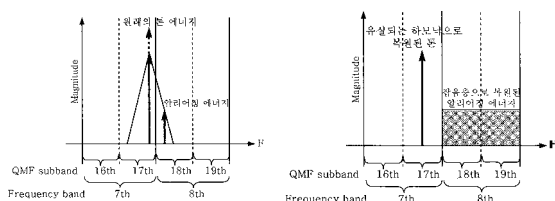


그림 5. (a) 6700 Hz 정현파 신호의 스펙트로그램  
(b) (a)를 HE-AAC로 복원한 신호의 스펙트로그램  
fig. 5. (a) 6700 Hz sine wave spectrogram (b) Decoded signal spectrogram of (a) using HE-AAC decoder.



(a) SBR 부호화 과정에서의 톤 성분 (b) SBR 복호화 과정에서의 잡음층 형성  
(a) Tone component at SBR (b) Noise layer generation at encoding process SBR decoding process

그림 6. 잡음층이 발생하는 과정의 개념도  
fig. 6. Noise layer generation Principle.

것이다.

이러한 왜곡을 개선하기 위해 본 논문에서는 문제가 되는 톤 성분의 주파수를 SBR 부호화 이전에 왜곡이 일어나지 않는 주파수로 사전에 변환시키는 방법을 제안한다. 제안하는 방법에서는 이동시킬 톤 성분을 구분해내기 위해, 음성신호 분석에 주로 사용되는 정현파 모델 (Sinusoidal model) [11]을 SBR 부호화기의 전단에 추가 적용하였다.

그림 7은 SBR 블록 전단에 추가되는 정현파 모델 블록도를 나타낸 것이며, SBR이 적용되는 고주파 대역에만 정현파 모델이 분석하도록 하기 위해, 입력 신호는 먼저 대역 통과 필터를 거치게 된다. 대역 통과 필터의 통과 대역은 SBR 알고리즘에서 결정되는 주파수 대역 테이블을 참고하고, 통과대역 주파수와 저지대역 주파수는 각각  $k_0$ ,  $k_0 + M$  번째의 QMF 서브밴드의 주파수가 된다. 여기서  $k_0$ 는 SBR 주파수 밴드 테이블의 첫 번째 QMF 서브밴드이며, M은 SBR로 복원되는 주파수 영역의 QMF 서브밴드 개수이다.

대역통과 필터를 거친 신호는 정현파 모델을 통해 정현파의 주파수 및 위상 변화에 대한 정보를 얻어내게 된다. SBR 복호화 과정에서는 유실되는 하모닉 성분을 1 프레임 동안 고정된 주파수를 가진 톤으로 대체하여 생성해내기 때문에, 정현파 모델을 통해 톤을 분석할 때, 1 프레임 동안 일정한 주파수 범위를 유지하는 톤을 감지하도록 한다. 이는 정현파 모델의 시간당 주파수 변화율의 임계치를 낮추어 줌으로써 가능하다.

다음 단계에서는 이 성분이 SBR 복호화 과정에서 왜곡이 얼마나 발생할지를 판단한다. 이는 QMF 서브밴드

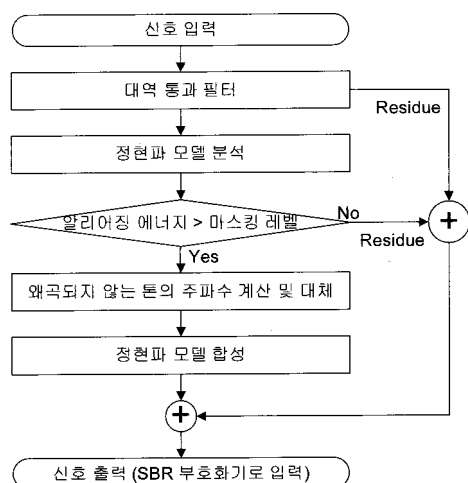


그림 7. 제안된 정현파 모델 블록도  
fig. 7. Proposed sinusoidal model block diagram.

의 에일리징 에너지를 계산하는 것으로 가능하며, 이웃한 서브 밴드로 누출되는 에너지는 QMF 분석시 사용되는 윈도우의 주파수 응답으로 산출할 수 있다. 3GPP HE-AAC의 QMF 서브밴드 간 누출되는 에너지를 그림 8에 나타내었다.

그림 8에서 가로축은 분석된 성분의 주파수가 해당 QMF 서브밴드의 중앙으로부터 얼마나 떨어져있는지를 나타낸다. 0일 경우는 QMF 서브밴드의 중앙에 위치하는 것이고, 0.5의 경우는 해당 QMF 서브밴드와 이웃한 QMF 서브밴드의 경계 주파수에 위치하는 것이다. 즉, 해당 QMF 서브밴드의 중앙 주파수를 가지고 있는 톤 성분은 이웃한 QMF 서브밴드로 에너지가 거의 나뉘지 않지만, 경계 부근에 위치할수록 절반에 가까운 에너지가 이웃한 QMF 서브밴드로 누출되게 된다.

정현파 모델에 의해 분석된 톤 성분들은 다음의 과정으로 처리한다. 먼저 각각의 톤에 대해 누출 에너지  $E_L$  은

$$E_L = E_T \cdot r(d) \quad (3)$$

로 구하며, 여기서  $E_T$ 은 분석된 톤의 평균 에너지이며,  $r(d)$ 는 누출 에너지 비율이다. 그리고,  $d$ 는 분석된 해당 톤의 평균 주파수로부터 QMF 서브밴드의 중앙 주파수까지의 거리이다. 만약 해당 주파수 밴드가 1개 이상의 QMF 서브밴드로 구성되어 있다면, 양 끝의 QMF 서브밴드에 한해서만 누출 에너지를 확인한다.

다음으로, 유실되는 하모닉이 생성될 주파수 밴드를 결정하도록 한다. 원래 톤의 주파수를 기준으로, 해당 주파수 밴드와 이웃하는 주파수 밴드의 유실되는 하모닉이 생성되는 주파수를 비교하여 근접한 유실되는 하모닉이 속하는 주파수 밴드로 결정한다.

다음 단계에서는 누출 에너지  $E_L$ 가 마스킹 레벨을 넘는지를 판단한다.

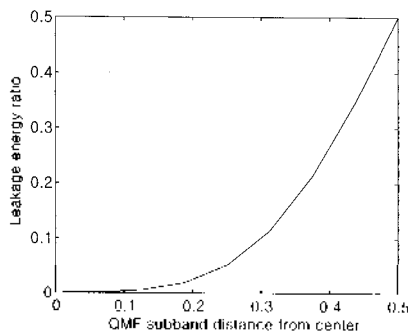


그림 8. QMF 누출 에너지 비율 그래프  
fig. 8. QMF Leakage energy ratio.

$$E_H = E_T - E_L \quad (4)$$

여기서  $E_H$ 은 이웃한 QMF 서브밴드로 나뉘고 남은 잔여 에너지 (residual energy)이다. 유실되는 하모닉의 톤 성분의 주파수와 잔여에너지가 만들어 내는 주파수 마스킹 곡선은 심리음향 모델의 확산비율 (spreading ratio)에 의해 구해지게 되며, 유실되는 하모닉의 주파수는 표 1에서 구할 수 있다.

확산 비율을 적용하기 위해, 해당 주파수 밴드를 포함하는 크리티컬 대역을 산출한다 [12].

$$Z_k = 13 \arctan(0.76 \cdot f_k / 1\text{kHz}) + 3.5 \arctan((f_k / 7.5\text{kHz})^2) \quad (5)$$

여기서  $Z_k$ 는  $k$ 번째 주파수밴드에 해당하는 크리티컬 밴드이다.  $f_k$ 는 복호화 과정에서  $k$ 번째 주파수밴드에 생성되는 유실되는 하모닉의 주파수이다. 스펙트럼 에너지를 컨벌루션하는 대신, 확산 비율을 적용하는 것으로 간단하게 산출한다. 이 때, 확산 비율은 아래와 같이 구해진다.

$$S_L[k] = -((z_{k+1} - z_k) \cdot 25) - 14.5 \quad (6)$$

$$S_H[k] = -((z_k - z_{k-1}) \cdot 15) - 14.5 \quad (7)$$

여기서  $S_L(k)$ 과  $S_H(k)$ 는 바크 (Bark)단위에서 이웃한 주파수 밴드에 의해 계산된 확산 에너지이며, 일정 감소 비율로 25 dB/Bark 와 15 dB/Bark 를 적용한다 [12]. 14.5 dB는 유사 톤 신호 (tone-like signal)이 유사 잡음 신호 (noise-like signal)을 마스킹하는 경우의 옵션값 [13]이다.

$$E_{thr}[k-1] = E_H[k] + S_L[k-1] \quad (8)$$

$$E_{thr}[k+1] = E_H[k] + S_H[k+1] \quad (9)$$

여기서,  $E_{thr}[k]$ 는  $k$ 번째 주파수 밴드의 마스킹 임계치가 된다. 누출 에너지가 이웃한 QMF 서브밴드의 마스킹 임계치보다 높다면, 식 (4)와 (5)를 이용하여 누출 에너지가 마스킹 임계치보다 낮게 될 때의 톤의 주파수를 역으로 산출한다. 이제 톤을 이동시킨 주파수가 결정되었으면, 정현파 모델에서 분석된 톤의 주파수를 수정하고, SBR 복호화기로 전달될 해당 주파수 밴드의 유실되는 하모닉 플래그 (missing harmonics flag)를 설정한다.

제안하는 톤 이동 방법을 그림 9에 예시하였다. 그림 9(a)는 QMF 분석을 통한 결과이며, 에일리징 에너

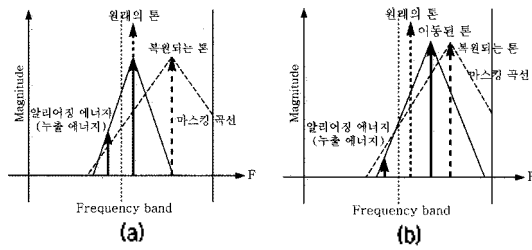


그림 9. 톤 이동 방법의 예시도 (a) 일반적인 QMF 분석 후 (b) 알리어징 성분이 들리지 않도록 사전에 톤을 이동하고 QMF 분석 후

fig. 9. Example of moving tone (a) Result of reference QMF analysis (b) Result of QMF analysis using moving tone for reducing aliasing.

지가 복원될 톤 성분의 마스크 임계치 곡선을 초과하는 것을 보여준다. 그림 9(b)는 톤을 옮겨서 복원될 톤 성분의 마스크 임계치 곡선보다 알리어징 에너지가 낮게 되도록 조정시키는 것으로 위에서 설명된 과정을 보여준다.

#### IV. 실험 및 결과 고찰

##### 4.1. 확장된 TNS (Extended TNS)의 결과

다음 그림은 이 결과를 나타낸 것이다. 그림 10(a)는 중국여성화자 신호인 "chinese\_female\_1.wav" 원본 신호이며, 그림 10(b)는 HE-AAC를 적용한 신호이다. 마지막으로, 그림 10(c)는 확장된 TNS를 적용한 신호이다. 살펴보면 그림 10(b)의 표기한 부분의 프리에코가 사라진 것을 볼 수 있다.

여러 가지 실험을 통하여 확장된 TNS는 단구간 블록을 사용하는 것 만큼 프리에코를 효과적으로 해결함을 관찰할 수 있었다. 장구간 블록만을 사용해서 과도한 신호를 충분히 처리할 수 있다면, 단구간 블록을 사용해서 문제되었던 뮤지컬노이즈 또한 향상되며, 물론 이는 낮은 비트율에서 더욱 효과를 볼 수 있다.

##### 4.2. 정현파 모델을 사용한 톤 이동 기법의 결과

제안된 방법에서 중요한 것은 표준 SBR 알고리즘이 유실되는 톤 성분을 복원할 때, 지정된 주파수로 복원된다는 점이다. 이는 해당 주파수 밴드내에서 톤 성분을 이동시켜 놓아도, 결국 복원되는 톤은 주파수의 변화가 없다는 점을 이용한 결과이다. 게다가 제안된 방법에서는 주파수 밴드 간의 유실되는 하모닉 들을 고려하여 가장 유사하게 복원하기 때문에 좀 더 자연스러운 복원이

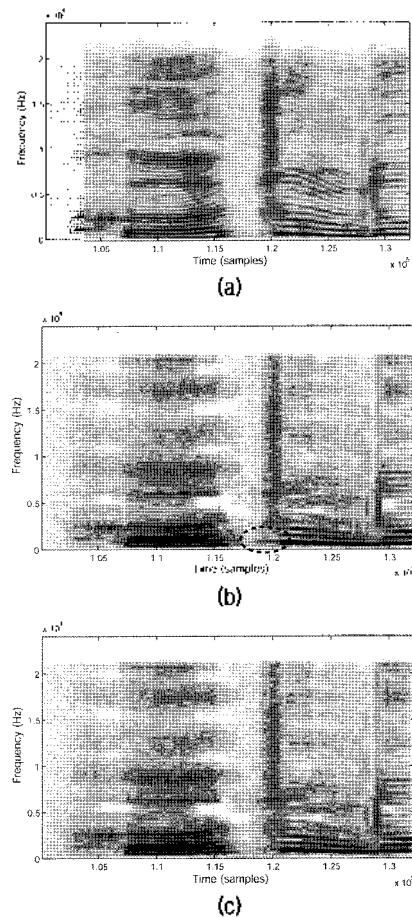


그림 10. "Chinese\_female\_1.wav" 신호 실험 결과. (a) 원본 신호, (b) HE-AAC로 복원된 신호, (c) eTNS 적용결과  
fig. 10. Result of encoded "Chinese\_female\_1.wav" (a) Original signal (b) Decoded signal by HE-AAC decoder (C) Decoded signal using eTNS.

가능하다.

제안된 방법을 사용하여, 실제 음원에 대하여 표준 HE-AAC와의 스펙트로그램 비교 결과를 그림 11에 나타내었다. 제안된 정현파 이동 방법을 사용함으로써 고주파 대역의 톤 성분이 좀 더 충실히 재생됨을 확인할 수 있다.

##### 4.3. 주관적인 실험 및 결과

확장된 TNS 기법과 정현파 모델을 사용한 톤 이동 방법에 대한 각각의 주관적인 음질 평가와 이 두가지 방법을 모두 적용하여 복원한 신호의 음질 평가를 실시하여 모두 3번의 실험을 하였다. 그리고, 톤 이동 방법만을 적용한 실험에서는 저주파 대역을 원래의 입력 신호로 대체하여 순수 SBR 영역의 개선효과만을 파악할 수 있도록 하였다. 실험자로는 경험이 풍부한 7명에 대해 실시하였으며, 음성과 음악, 또는 보컬이 가미된 음악

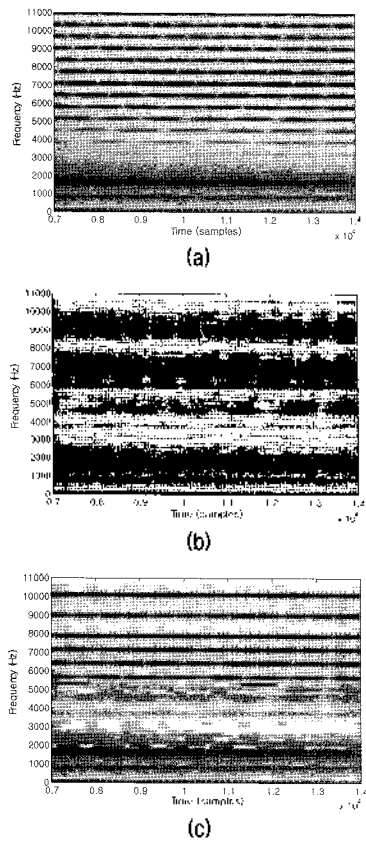


그림 11. (a) 원본 신호 (b) HE-AAC로 복원된 신호 (c) 제안된 HE-AAC로 복원된 신호  
 fig. 11. (a) Original signal (b) Decoded signal by HE-AAC decoder (c) Decoded signal using proposed algorithm.

표 2. 주관적 음질 평가에 사용된 음원  
 Table 2. Test materials.

번호	입력파일	시간 (초)	분류
1	Castanets	6.28	악기 (연속과도)
2	Chinese_female_1	19.2	음성 (여성)
3	Dongwoo	6.13	보컬 & 음악
4	Es02_s	7.24	음악 (전자악기)
5	Jazz	6.02	음악 (전자악기)
6	Moulin	3.23	음악 (전자악기)
7	velvets	5.01	악기 (짧은과도)

등의 7개의 입력 신호들로 12 kbps의 낮은 비트율에서 기존 HE-AAC와 제안된 방법을 적용한 HE-AAC의 결과를 비교 평가하였다. 평가 기준은 MUSHRA score [14]를 사용했으며 표 2에 입력 신호를 나타내었고, 그림 12에는 각 실험의 음질 평가 결과를 나타내었다.

실험에 사용된 음원 중 과도 성분이 매우 연속적인 신호 1번과 음성신호인 2, 4번은 eTNS 방법을 적용하였을 때 향상이 있고, 고주파에 톤 성분이 많은 전자 악기로 연주

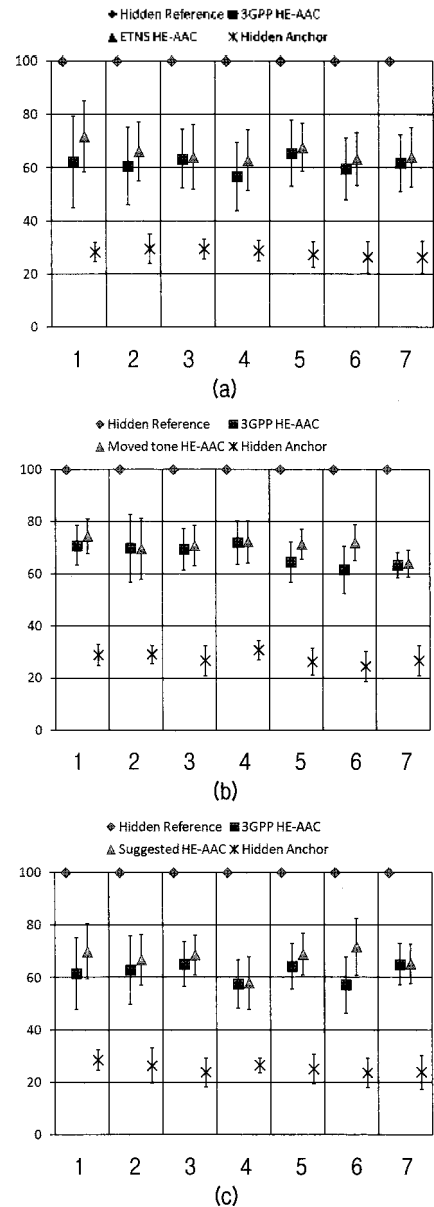


그림 12. MUSHRA 테스트 결과 확장된 TNS (a) (Extended TNS)의 음질평가 결과 (b) 톤 이동 기법의 음질평가 결과 (c) (a),(b) 모두 적용된 신호의 음질평가 결과  
 fig. 12. MUSHRA test results of eTNS (a) Subjective listening test result of eTNS (b) The result of moving tone algorithm (c) The results using both (a) and (b).

된 5, 6번은 톤 이동 기법에서 효과가 있었다. 그리고, 순간적인 과도 성분이 주기적인 7번과 보컬이 섞인 음악인 4번의 경우는 별다른 향상은 없었지만, 제안된 방법을 사용하였을 때, 기존 대비 음질이 저하되지는 않았다.

그림 12의 결과에서 볼 수 있듯이, 각각의 제안된 방법이 서로 다른 주파수 영역과 음원에서 독립적으로 효과를 보여주는 것을 확인 할 수 있었으며, 특히, 과도한



성분이 많은 음원이었던 Castanets와 고주파 대역에  
 톤 성분이 강한 Moulin의 경우에는 기존 대비 많은 향  
 상이 이루어졌다.

## V. 결론

본 논문에서는 고효율 오디오 부호화 방법인 HE-AAC  
 부호화기를 구성하고 있는 SBR 부호화기와 AAC 부호  
 화기에 대해 각각의 개선방법을 제안하였다. 저주파 대  
 역을 부호화하는 AAC 부호화기에서는 TNS 적용 주파  
 수를 확장시켜 프리에코와 유지킬노이즈를 줄였으며,  
 고주파 대역을 부호화하는 SBR 부호화기에서는 정현파  
 모델로 신호를 분석하여 잡음층을 유발시킬 수 있는 톤  
 성분을 부호화 이전에 이동시키는 방법으로 부호화 과  
 정에서 노이즈 성분이 발생하는 것을 최소화하였다. 제  
 안된 방법은 비트 사용량이 증가하지 않는다는 것과 복  
 호화기에서 추가적인 연산량이 필요없다는 점 등의 장  
 점이 있다. 각각의 방법을 사용하여 스펙트로그램을 비  
 교하여 보고, 주관적 음질평가를 수행하여 지각적으로  
 향상이 있는지를 평가하였다. 표준 HE-AAC에 비해 음  
 질이 개선 됨을 확인하였으며, 인지적으로도 향상된 결  
 과를 보여주었다.

## 감사의 글

본 연구는 정보통신부 및 정보통신연구진흥원의 IT신  
 성장동력 핵심기술개발사업 [2007-S005-01, AV코  
 덱 고도화를 통한 리치미디어 방송기술 개발]의 연구비  
 지원으로 이루어졌습니다.

## 참고 문헌

1. M. Wolters, K. Kj rling, D. Homm and H. Purnhagen, "A  
 Closer Look into MPEG-4 High Efficiency AAC," AES 115th  
 Convention, New York, October 2003.
2. ISO/IEC, "Text of ISO/IEC 14496-3:2001 / FPDAM 1, Band-  
 width extensions," ISO/IEC JTC1/SC29/WG11/N5203, October  
 2002.
3. M. Dietz, L. Liljeryd, K. Kj rling and O. Kunz, "Spectral  
 Band Replication, A Novel Approach in Audio Coding," AES  
 112nd Convention, Munich, 2002 May 10 - 13.
4. ISO/IEC, "International Standard ISO/IEC IS 13818-7, Infor-  
 mation technology - GenericCoding of Moving Pictures and

Associated Audio:, Part 7: Advanced Audio Coding (AAC)",  
 ISO/IEC JTC1/SC29/WG11, 1997.

5. Makinen, J. and Bessette, B. and Bruhn, S. and Ojala, P.  
 and Salami, R. and Taleb, A., "AMR-WB+: a new audio  
 coding standard for 3rd generation mobile audio services",  
 Proc IEEE International Conference on Acoustics, Speech,  
 and Signal Processing (ICASSP '05).
6. 박호중, 박명철, 홍진우, "음성 및 오디오 통합 부호화 기술",  
 Telecommunications Review, 17(5), 841-854, 2007, 10.
7. Chang, Chia-Ming, Hsu, Han-Wen, "Compression Artifacts  
 in Perceptual Audio Coding", AES 121th Convention, San  
 Francisco, October 2006.
8. 3rd Generation Partnership Project, "Enhanced aacPlus en-  
 coder SBR part 3GPP TS 26.404," 3rd Generation Partner-  
 ship Project Technical Specification Group Services and  
 System Aspects, September 2004.
9. <http://www.3gpp.org/ftp/Specs/html-info/26410.htm>
10. 3rd Generation Partnership Project, "Advanced Audio Co-  
 ding (AAC) part 3GPP TS 26.403," 3rd Generation Partner-  
 ship Project Technical Specification Group Services and  
 System Aspects, June 2006
11. Robert J. McAulay and Thomas F. Quatieri, "Speech Analysis  
 /Synthesis Based on a Sinusoidal Representation", IEEE  
 transactions on acoustics, speech and signal processing,  
 34(4), 744-754, august 1986.
12. E. Zwicker and H. Fasel, Springer-Verlag, Berlin Heidelberg  
 1990.
13. N. Jayant, J. Johnston and R. Safranek, "Signal Compression  
 Based on Method of Human Perception", Proc. Of IEEE,  
 81(10), 1385-1422, October 1993.
14. ITU-R, "Method for the subjective assessment of interme-  
 diate quality level of coding systems (MUSHRA)", ITU-R  
 Recommend, BS, 1534, 2001.

## 저자 약력

### •김 정 근 (Jeong-Geun Kim)

2001년 2월 : 홍익대학교 전자공학과(공학사)  
 2006년 3월~현재 : 연세대학교 전자공학과 석사과정  
 \* 주관심 분야 : 음성/오디오 신호처리

### •이 재 성 (Jae-Seong Lee)

2003년 2월 : 건국대학교 전자공학과 (공학사)  
 2005년 2월 : 연세대학교 전자공학과 (공학석사)  
 2005년~현재 : 연세대학교 전자공학과 박사과정  
 \* 주관심 분야 : 음성/오디오 신호처리

### •박 영 철 (Young-Cheol Park)

한국음향학회지 제24권 3호 참조

### •이 태 진 (Taejin Lee)

한국음향학회지 제24권 2호 참조

### •강 경 옥 (Kyeongok Kang)

한국음향학회지 제24권 2호 참조