

저 전송률 환경에서 선형예측 전처리를 사용한 HE-AAC의 성능 향상

정희원 이재성*, 이건우***, 박영철**, 종신회원 윤대희*

Quality Improvement of Low Bitrate HE-AAC using Linear Prediction Pre-processor

Jae-seong Lee*, Gun-Woo Lee***, Young-chul Park** *Regular Members*,
Dae-hee Youn* *Lifelong Member*

요약

본 논문은 선형예측 전처리를 이용하여 저 전송률 환경에 적합한 HE-AAC의 구조를 제안한다. 저 전송률 환경에서는 HE-AAC의 적절하지 못한 비트 할당 알고리즘 때문에 많은 스펙트럴 홀(스펙트럼 홀)들이 발생을 하고 있으며, 그로 인해서 심각한 음질의 열화가 발생하고 있다. 이를 해결하기 위해서 선형예측 전처리를 사용하여 저 전송률에서 비트가 적절하게 할당되도록 하였다. HE-AAC로 들어오는 입력신호는 선형예측 전처리에 의해서 LP 계수와 레지듀얼 신호로 나뉘지게 되며, AAC 부분은 분리된 레지듀얼 신호를 부호화하게 된다. 제안된 방법의 성능 평가를 위해서 지각적 잡음(Perceptual noise)의 측정을 통한 객관적인 실험과 MUSHRA 테스트를 통한 주관적인 실험을 하였고, 그 결과 저 전송률 환경에서 제안된 방법을 사용함으로써 성능을 향상시킬 수 있음을 확인하였다.

Key Words : Linear Prediction Pre-processor, Low Bitrate, HE-AAC, Spectral Hole

ABSTRACT

This paper proposes a new method of improving the quality of High Efficiency Advanced Audio Coding (HE-AAC). HE-AAC encodes input source by allocating bits for each scalefactor bands appropriately according to human ear's psychoacoustic property. As a result, insufficient bits are assigned to the bands which have relatively low energy. This imbalance between different energy bands can cause decreasing of sound quality like musical noise. In the proposed system, a Linear Prediction (LP) module is combined with HE-AAC as a pre-processor to improve sound quality by even bits distribution. To apply accurate human being's psychoacoustic property, the psychoacoustic model uses Fast Fourier Transform (FFT) spectrum of original input signal to make masking threshold. In its implementation, masking threshold of psychoacoustic model is normalized using the LP spectral envelope in prior to quantization of the LP residual. Experimental result shows that, the proposed algorithm allocates bits appropriately for insufficient bits condition and improves the performance of HE-AAC.

※ 본 연구는 방송통신위원회, 지식경제부 및 한국산업기술평가관리원의 IT 원천기술개발사업의 일환으로 수행한 연구로부터 도출된 것입니다. [과제관리번호: 2008-F-011, 과제명: 차세대 DTV 핵심기술 개발]

* 연세대학교 전기전자공학부 디지털신호처리 연구실(dream7070@dsp.yonsei.ac.kr), ** 연세대학교 컴퓨터통신공학부, ***삼성전자 논문번호: KICS2009-06-244, 접수일자: 2009년 6월 5일, 최종논문접수일자: 2009년 7월 15일

I. 서론

소리를 효율적으로 압축을 하는 방법은 크게 오디오 부호화와 음성 부호화로 나뉜다. 두 방법 모두 더 좁은 대역폭과 더 적은 메모리를 사용하면서 인지적으로 높은 음질을 보장할 수 있도록 소리 정보의 중복성이나 불연관성을 제거하는 데에 그 목적이 있다^[1].

ITU-T의 주도하에 연구가 되어온 음성 부호화 방법은 휴대단말간의 통신환경의 주로 낮은 전송률의 양방향 통신 환경에서 좁은 대역폭을 갖는 음성을 실시간으로 전송을 하기 위해 낮은 복잡도를 갖도록 연구가 진행되어 왔다^[2]. CELP로 대표되는 이러한 방법은 4.8~16kbps의 환경에서 전화 통화음질을 보장한다. 이에 반해, MPEG 주도로 연구되어 온 오디오 부호화 방법은 저장 또는 단방향 통신을 위하여 높은 전송률 환경에서 넓은 대역폭을 갖는 소리에 대해 좋은 음질을 보장할 수 있도록 연구되어 왔다^{[3],[4]}.

하지만 각 코더의 뚜렷한 특성에도 불구하고 통신의 발달로 인하여 실시간 스트리밍 서비스와 같이 양방향 통신에 대한 요구와 높은 음질에 대한 요구가 동시에 이루어지고 있는 것이 요즘의 추세이다^[4]. 따라서 음성 부호화는 더 높은 전송률과 넓은 대역폭에서 더 좋은 음질을 보장하는 방향으로, 이와는 반대로 오디오 부호화는 더 낮은 전송률과 더 좁은 대역폭을 사용하면서 좋은 음질을 유지하는 방향으로 연구가 진행되고 있다. 이에 대표적인 부호화기로 각각 AMR-WB+와 HE-AAC가 있다. 그 중 HE-AAC는 오디오 부호화기 중 낮은 전송률에서 현재 가장 좋은 성능을 나타내는 코더로서, 32kbps 이하의 전송률에서도 기존 AAC 대비 30% 이상의 부호화 효율을 향상시켜서 뛰어난 음질을 보장한다^[5].

하지만 HE-AAC가 뛰어난 성능을 보임에도 불구하고 매우 낮은 전송률에서는 AMR-WB+ 등에 비해서 낮은 성능을 나타내고 있다. 그것은 높은 전송률에서는 효과적이었던 비트 할당 알고리즘이 낮은 전송률의 환경에서는 더 이상 효율적으로 비트를 할당하지 못하기 때문이다. 따라서 본 논문은 낮은 전송률에서 효율적인 부호화를 하기 위하여 HE-AAC에 선형예측 전처리기를 이용하는 방법을 제안하였다.

본 논문은 2장에서 기존 HE-AAC의 양자화 과정에 대하여 간단히 설명을 하고, 그것으로 인한 낮은 전송률 환경에서의 문제점을 살펴본다. 3장은 본

논문에서 제안하는 선형예측 전처리에 대해서 설명을 하고, 그것에 따른 심리음향모델에 대해서 설명을 한다. 4장에서는 제안된 알고리즘의 성능을 비교 평가하기 위하여 실험한 결과를 보여준다. 마지막으로 5장에서 간단하게 정리를 하고 결론을 맺는다.

II. HE-AAC의 양자화 과정

HE-AAC는 양자화 과정에서 적은 연산을 통하여 효율적인 코딩을 하기 위해서 임계치와 스펙트럼의 정보를 이용하여 스케일팩터를 산출한다. 이 때 적용되는 임계치는 사용가능한 정보량에 의해서 수정된 값이다. 따라서 정확한 청각적 특성을 반영하기 보다는 효율적인 코딩을 위하여 계산된 값이다. 특히 이렇게 계산된 최종 임계치는 매우 낮은 전송률에서는 지나치게 높은 레벨을 갖는다. 최종 임계치를 구하는 과정에서 사용되는 감쇠값은 전체 대역에 대해서 허용 가능한 비트를 만족시키기 위한 값으로 결정이 되고, 따라서 임계치의 상승은 전체 대역에 대해서 하나의 감쇠값에 의해서 결정된다. 그런데 실제 임계치의 상승은 loudness domain에서 이루어지게 되므로 작은 에너지 레벨에 대해서는 상대적으로 많은 양의 상승이 이루어지게 되고, 반대로 높은 에너지 레벨에 대해서는 상대적으로 적은 양의 상승이 이루어지게 된다.

그림 1은 PAM으로부터 얻어진 초기 임계치와 감쇠값에 의해서 상승된 후의 임계치를 나타낸 것이다. 그림 1에 표시된 부분에서 비교를 할 수 있듯이, 모든 밴드에 대해서 같은 감쇠값이 적용되었지만, 낮은 에너지를 갖는 밴드에 대해서 더 많은 임계치의 상승이 이루어 졌음을 알 수 있다. 이것은 상대적으로 높은 에너지 레벨에 대해서는 임계치의 상승을 줄임으로써 더 높은 Signal to Mask Ratio (SMR)을 갖도록 하기 위함이다. 하지만 16kbps이

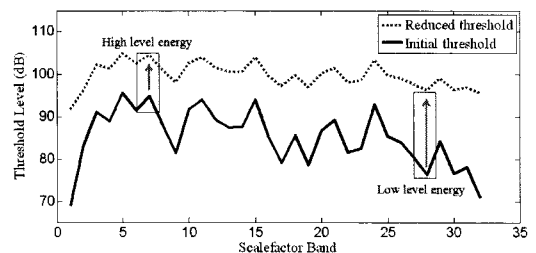


그림 1. 마스크 임계치와 사용가능한 정보량의 부족으로 인해 미리 올려진 최종 임계치

하의 매우 낮은 전송률에서는 사용가능한 정보량이 너무 적기 때문에 감쇠값의 값이 매우 커지게 되어, 상대적으로 적은 에너지 레벨이 분포한 구간에서는 지나친 임계치의 상승이 이루어지게 된다. 이렇게 만들어진 최종 임계치가 스펙트럼 에너지 레벨 보다 큰 레벨을 갖게 되면 그러한 스케일팩터 밴드는 코딩이 되지 않고 스펙트럼 홀(hole)을 발생시킨다. 즉, 낮은 레벨의 에너지가 분포한 구간에 대해서는 지나친 최종 임계치의 상승으로 인하여 스펙트럼 홀이 발생하게 된다.

즉, HE-AAC의 양자화 전략은 높은 에너지를 갖는 부분에 대해서는 더 정확한 표현을 위해서 SMR을 높이고, 반대로 낮은 에너지를 갖는 부분에 대해서는 SMR을 낮춤으로써 정보량의 할당을 덜 되게 하는 것이다. 따라서 매우 낮은 전송률의 환경에서는 상대적 에너지 크기의 차이가 심한 음원에 대해서는 적은 에너지를 갖는 구간에 대해서 지나치게 임계치를 상승시킴으로써 스펙트럼 홀들을 많이 발생시키고 있다. 인위적으로 마스킹 임계치를 낮은 전송률에 대하여 수정을 할 경우 심리음향 모델에 기반한 수정이 어렵고 절대적으로 낮은 전송률 환경에서는 사람의 청각적 특성을 고려한 심리음향 모델이 제대로 구현되기 어렵다. 즉 이미 모든 대역의 양자화 잡음이 마스킹 임계치를 넘게되므로 마스킹 임계치에 대한 수정보다는 스펙트럼 홀을 상쇄하여 양자화의 효율성을 높이는 방안이 필요하다.

III. 제안된 선형예측 전처리기

선형 예측 방법은 과거의 인접한 샘플들로부터 현재의 샘플을 예측하는 방법으로서 다음의 식과 같다.

$$\tilde{x}(n) = \sum_{k=1}^L a_k x(n-k) \quad (1)$$

수식(1)에 나타난 예측된 신호 $\tilde{x}(n)$ 은 과거 L 개의 인접한 샘플들과 예측 계수 a_k 로부터 현재의 샘플을 예측한 것이다. 레지듀얼 신호는 실제 신호와 예측된 신호 사이의 에러를 나타내며 다음의 식과 같다.

$$r(n) = x(n) - \sum_{k=1}^L a_k x(n-k) \quad (2)$$

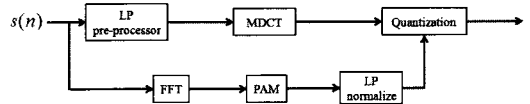


그림 2. 제안된 선형예측 전처리기를 사용한 AAC의 구조

이렇게 나온 레지듀얼 신호는 원 신호에 있던 인접 샘플간의 중복성이 제거된 신호이고, 원 신호에 비해서 평평한 모양의 스펙트럼을 갖는 특성이 있다. 따라서 이러한 특성을 이용하여 HE-AAC의 AAC 부분으로 들어가는 입력신호의 스펙트럼 에너지 분포를 더 균등하게 만들어줌으로써 불균등한 에너지의 분포로 인해 부적절하게 정보량이 할당되던 낮은 전송률 환경에서의 문제점을 해결할 수 있다.

선형예측 전처리기를 적용한 블록다이어그램은 다음과 같다.

불균등한 에너지 분포를 갖는 입력신호는 선형예측 전처리기를 통해서 선형 분석(analysis)되고, 스펙트럼 envelope 정보를 나타내는 선형예측(Linear Prediction : LP)계수는 Immitance spectral pair (ISP)로 변환되어 전달된다⁶⁾. HE-AAC는 이 과정에서 나온 레지듀얼 신호를 부호화하게 된다. 또한 새롭게 변한 입력신호의 특성에 맞도록 심리음향 모델과 양자화 모듈을 최적화 시켜주었다.

3.1 선형예측 전처리기

선형예측 전처리기의 구조는 다음과 같다.

AAC는 매 프레임마다 1024개의 새로운 입력신호를 받아서 앞 프레임의 1024입력신호와 함께 2048 MDCT를 수행한다. 여기에 선형 예측을 적용하기 위해서 새로 들어오는 1024개의 입력신호에 대해서 선형예측 전처리기를 통과시킨다. 하지만 48kHz로 샘플링 된 입력신호에 대해서 AAC의 한 프레임의 길이는 약 42.6ms가 되므로 하나의 주파수 응답으로 스펙트럼 포락선을 표현하기에는 긴 시간이다. 따라서 스펙트럼 포락선 모양의 변화를 반영하기 위해서 20ms의 길이를 갖는 두 개의 서브 프레임

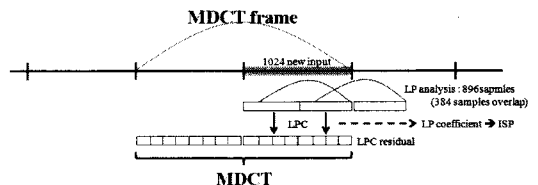


그림 3. 선형예측 전처리기의 구조

(sub frame)으로 나누어서 각각에 대해서 LP를 수행하였다. 각 서브프레임에 대해서는 896 샘플 길이의 콰드러칭의 sine 윈도우가 사용되었고, 그 수식은 다음과 같다.

$$w(n) = \sin\left(\frac{\pi}{L}(i+0.5)\right)$$

$$i = 0, \dots, L-1 \quad L = 896 \quad (3)$$

윈도우가 취해진 입력 신호에 대해서 자기상관 (autocorrelation)값이 계산되고, Levinson-Durbin 알고리즘을 이용하여 LP 계수를 산출한다. LP의 차수는 입력 신호에 상관없이 고정 10차를 사용하였다. 스펙트럼 포락선 정보를 나타내는 LP계수는 ISP로 변환을 하여 표현하였다. 그렇게 함으로써 정확한 filter의 주파수 응답을 전달할 수 있었다. 10개의 ISP 값은 Immittance spectral Frequencies(ISF)으로 옮겨진 후 Split Multistage Vector Quantization (S-MSVQ)을 이용하여 양자화된다. 이 양자화 방법은 AMR-WB+[6]의 ISP 양자화 과정과 동일한 방법을 사용하였다. 10개의 계수를 양자화 하는 데에는 약 1.2kbps가 소요된다고 가정을 하였고, 그에 따라서 HE-AAC의 AAC부분은 1.2kbps가 적은 정보량으로 레지듀얼을 코딩하게 된다. 이렇게 만들어진 LP계수를 이용하여 각 서브프레임에 대해서 레지듀얼을 만들고, AAC는 레지듀얼을 부호화하게 된다.

그림 4와 그림 5는 각각 제안된 선형 예측 전처리기를 사용함으로써 인해서 발생한 스펙트럼의 변화와 이에 따른 스케일팩터 밴드별 정보량 할당의 변화를 그린 것이다. HE-AAC는 제한된 정보량으로 신호를 표현하기 위해서 큰 에너지가 분포한 부분에 많은 정보량을 할당하고, 그렇지 않은 부분에는 적게 할당하려고 한다. 따라서 전체적으로 할당되는 정보량의 분포는 에너지의 분포와 관련이 깊다. 따

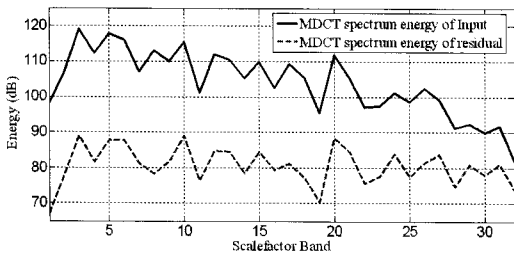


그림 4. 선형예측 전처리기로 인한 입력 신호의 스펙트럼 변화

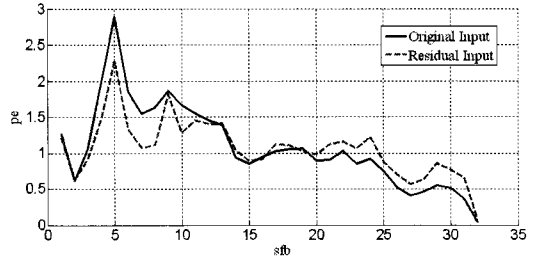


그림 5. 선형예측 전처리기로 인한 스케일팩터 밴드별 정보량 할당의 변화

라서 제안된 선형예측 전처리기를 사용함으로써 인해서 스펙트럼의 분포가 그림 4와 같이 균등해짐으로써 인해서 이제까지 대체적으로 정보량의 할당이 덜 되던 고주파부분에 기존보다 더 많은 정보량이 할당됨을 그림 5를 통해서 확인할 수 있다.

3.2 FFT에 기반한 심리음향모델

심리음향모델은 사람 귀의 인지적인 특성을 반영하여 입력신호의 스펙트럼 에너지로부터 마스킹 임계치를 만들어낸다. 하지만 제안된 선형예측 전처리기를 사용할 경우에 AAC로 들어가는 신호는 원래의 입력신호가 아닌 입력신호의 레지듀얼신호가 된다. 따라서 선형예측 전처리기로 인해서 특성이 변화된 새로운 레지듀얼 입력신호를 이용해서 기존의 심리음향모델을 그대로 사용하게 되면 온전한 사람의 인지적인 특성을 반영할 수 없게 된다. 따라서 레지듀얼 신호를 이용하여 마스킹 임계치를 구하더라도 정확한 인지적 특성을 반영할 수 있도록 후처리과정이 요구된다. 그것은 그림 6에 나타나있는 LP 정규화 과정으로써 구체적인 과정은 다음의 블록다이어그램과 같다.

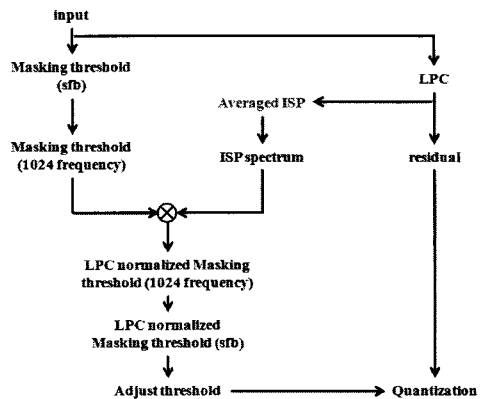


그림 6. LP 계수를 이용한 마스킹 임계치 normalize의 과정

우선 정확한 마스크 임계치를 만들기 위해서 선형예측 전처리를 통과한 레지듀얼 신호가 아닌 원래의 입력신호의 FFT 스펙트럼으로부터 마스크 임계치를 구한다. 그러므로 이렇게 얻어진 마스크 임계치는 실제 부호화하게 될 레지듀얼 신호의 MDCT 스펙트럼과는 레벨이 맞지 않게 된다. 즉, 원래의 입력신호로부터 얻은 마스크 임계치를 스펙트럼 인벤텔 정보 제거된 레지듀얼 신호에 적용할 수 있도록 LP 계수를 이용하여 정규화시켜주게 된다. 그렇게 함으로써 마스크 임계치가 가지고 있는 가장 중요한 정보중의 하나인 SMR을 레지듀얼 도메인에서도 그대로 유지시킬 수 있게 된다.

SMR이 유지되는 과정은 다음과 같이 설명된다. 원 신호의 스펙트럼 에너지와 마스크 임계치의 에너지를 각각 S_M 와 M_M 라 할 때, 원 신호로부터 얻을 수 있는 정확한 SMR은 다음의 식과 같다.

$$SMR = \frac{|S_M(e^{j\omega})|^2}{M_M(e^{j\omega})} \quad (4)$$

여기서 선형예측 전처리가 적용된 경우, 레지듀얼 신호의 스펙트럼 에너지는 다음과 같이 계산된다.

$$|E_M(e^{j\omega})|^2 = |S_M(e^{j\omega})|^2 |A_M(e^{j\omega})|^2 \quad (5)$$

A_M 은 LP 분석 과정에서 얻어진 원 신호의 스펙트럼 인벤텔 정보를 나타내는 LP 스펙트럼으로서 LP 계수를 2048 MDCT변환을 함으로써 얻을 수 있다. 이렇게 신호의 스펙트럼 에너지가 S_M 에서 E_M 으로 바뀌었으므로 원래의 SMR을 유지시켜주기 위해서는 원신호로부터 얻어진 마스크 임계치도 다음과 같이 정규화시켜 준다.

$$M'_M(e^{j\omega}) = M_M(e^{j\omega}) |A_M(e^{j\omega})|^2 \quad (6)$$

여기서 M'_M 은 정규화된 마스크 임계치를 나타낸다. 결국 이렇게 정규화된 마스크 임계치를 적용시키게 되면 선형예측 전처리를 사용하지 않았을 경우의 SMR을 다음의 수식과 같이 레지듀얼 도메인에서도 그대로 적용을 시킬 수 있게 된다.

$$SMR = \frac{|S'_M(e^{j\omega})|^2}{M'_M(e^{j\omega})} = \frac{|S_M(e^{j\omega})|^2 |A_M(e^{j\omega})|^2}{M_M(e^{j\omega}) |A_M(e^{j\omega})|^2} \quad (7)$$

$$= \frac{|S_M(e^{j\omega})|^2}{M_M(e^{j\omega})}$$

IV. 실험 결과

코어 밴드에 대하여 선형예측 전처리를 사용하여 스펙트럼 인벤텔 정보를 추출하고, 나머지 레지듀얼 신호는 HE-AAC로 코딩하게 된다. 선형예측 전처리를 사용함으로써 달라지는 비트 할당 전략에 의해 얻을 수 있는 이득 및 효과를 알아보기 위해서 다음과 같은 실험이 진행되었다. 먼저 각 전송률에 따라서 선형예측 전처리를 적용한 경우에 발생하는 심리음향적 잡음(perceptual noise)의 정도를 측정하여 원래의 경우와 비교를 해 봄으로써 선형예측 전처리의 사용이 근거가 있음을 관찰하였다. 그리고 각 전송률에 따라서 스펙트로그램의 변화를 관찰하고, 그에 따른 왜곡의 변화 양상을 관찰하였다. 그리고 MUSHRA 테스트를 이용한 주관적인 음질 평가를 통하여 제안된 선형예측 전처리가 실제 음질의 향상에 미치는 영향을 측정하였다.

실험은 48kHz로 샘플링된 모노신호에 대해서 진행되었으며, 실험에 사용된 선형예측 전처리는 10차를 사용하였다. 또한 모든 실험에 있어서 LP 계수는 1.2kbps의 정보량을 사용하여 ISP로 변환되어 부호화하였으며, 모든 실험에서 사용된 선형예측 전처리는 HE-AAC 중 AAC부분의 성능 개선을 위해서 제안된 것이기에, 모든 비교는 SBR을 제외한 AAC대역에 대해서만 이루어졌다.

4.1 객관적인 음질 평가

각 전송률 환경에서 다양한 음원들에 대하여 심리음향적 잡음을 NMR을 통하여 측정하여 보았다. 사용된 음원의 종류는 다음과 같이 일반 오디오 신호와 음성을 포함한 다양한 음원을 사용하였으며, 각 경우에 사용되는 양자화기는 심리음향적 잡음을 최소화하는 최적의 compressor를 사용한 값들을 사용하였다. 또한 심리음향적 잡음은 각 음원에 대해서 NMR을 측정하여 시간에 따른 평균을 구하여 사용하였다.

그림에서 나타난 바와 같이, 대부분의 음원들에 대해서 약 24kbps이하의 낮은 전송률에 대해서는 LP 모듈을 사용함으로써 전체적인 심리음향적 잡음을 줄일 수 있음을 확인하였다. 이것을 통해서 낮은 전송률에 대해서는 기존의 비트 할당 전략이 아닌 선형예측 전처리를 사용한 새로운 비트 할당 전략이 더 효과적일 수 있음을 알 수 있었다. 하지만 전반적인 경우에서 살펴보면 선형예측 전처리를 사용하지 않는 원래의 경우에는 약32kbps이상의 전

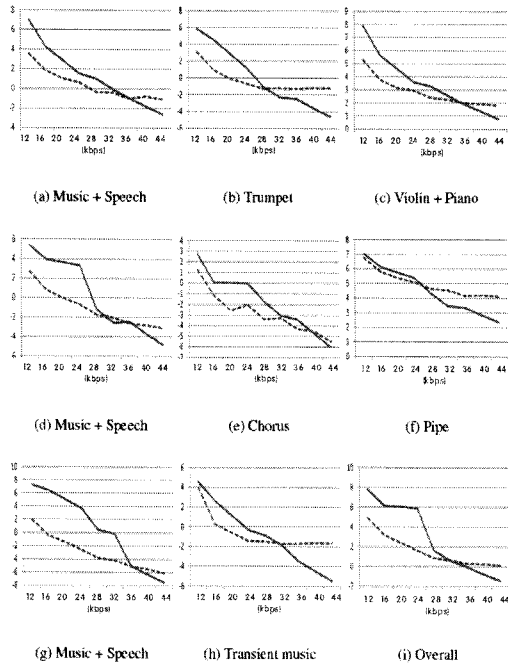


그림 7. 여러 종류의 음원에 따른 다양한 전송률 별 심리음향적 잡음추정 결과

송률에서는 심리음향적 잡음이 0dB아래로 내려가지만, 선형예측 전처리기를 사용하는 경우에는 전송률이 높아져도 0dB이하로 내려가지 않았다. 결국 전송률이 높은 환경에서는 선형예측 전처리기의 사용이 오히려 효과가 없음을 알 수 있었다.

4.2 주관적 음질 평가

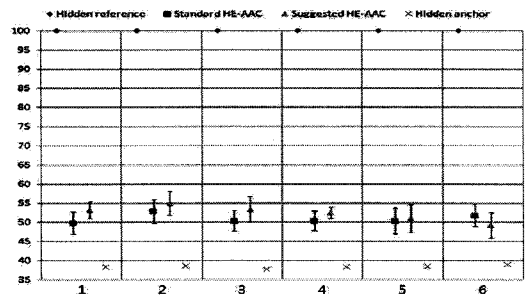
MUSHRA 테스트를 통하여 제안된 선형예측 전처리기가 실제 음질향상에 미치는 영향을 측정하였다. 실험은 헤드폰 환경에서 총 10명의 실험자들에게 대해서 진행하였으며, 실험에 사용된 음원들의 종류는 아래의 표에서 보이는 바와 같이 4개의 뮤직과 2개의 음성 등 다양한 종류로 선택되었고 모두 모노(Mono) 신호이다. 실험은 12와 16kbps의 전송률에 대해서 진행되었다.

표 3. MUSHRA 테스트에 사용된 음원의 종류

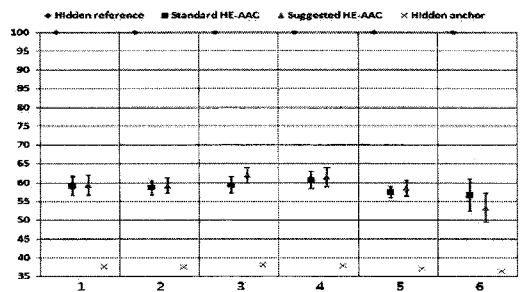
Index	Name	Description
1	rock	Base Guitar + drum
2	music 2	Symphony
3	fools	Instrument + female
4	chorus	Chorus
5	chinese female	Female (Chinese)
6	te19	Male (English)

그림 8에서 보이듯이 12kbps의 낮은 전송률 환경에서 제안된 알고리즘이 오디오 샘플에 대해서 음질 향상에 효과가 있었음을 확인할 수 있다. 기존에 12kbps의 환경에서 발생하던 많은 스펙트럼 홀들에 의한 뮤지컬 노이즈가 많이 없어짐으로써 전체적인 음질 향상이 이루어졌음을 알 수 있다. 제안된 선형예측 전처리기를 적용함으로써 낮은 주파수 대역에서 커지는 양자화 잡음에 의한 영향보다 높은 주파수 대역에서 줄어드는 뮤지컬 노이즈에 의한 음질 향상이 전체적인 음질에 더 큰 영향을 미쳤다. 반면 음성 신호들에서는 효과가 미미하거나 오히려 음질이 나빠지는 현상이 나타났다. 그것은 대부분의 음성 샘플들은 낮은 대역의 주파수 성분들에 의해서 전체적인 음질이 결정되는데, 본 논문에서 제안한 알고리즘은 높은 대역에 발생하는 문제점을 해결하기 위해서 낮은 주파수 대역의 성능이 나빠지기 때문이다. 본 알고리즘을 적용함으로써 높은 대역에 발생하는 스펙트럼 홀등은 제거가 되어 음질향상이 이루어졌지만, 정작 음성에 있어서 음질을 결정하는 중요한 부분인 낮은 주파수 대역에 있어서는 오히려 양자화 잡음이 증가하였기 때문이다.

다음으로 16kbps의 전송률에서의 결과를 보면,



(a) 12kbps



(b) 16kbps

그림 8. MUSHRA 테스트 결과

제안된 선형예측 전처리기에 의한 효과가 크게 나타나지 않는 것을 볼 수 있다. 각 음원들에 대해서 음질의 향상이 12kbps에 비하면 많은 차이를 보인 않았다. 역시 음성들에 대해서는 오히려 안 좋아지는 경우도 발생을 하고 있다. 이것은 기존의 HE-AAC 알고리즘도 충분한 전송률에 있어서는 스펙트럼 홀과 같이 높은 주파수에 발생하는 양자화 잡음에 의한 음질의 저하가 전체적인 음질에 미치는 영향이 적기 때문이다.

따라서 선형예측 전처리기를 적용함으로써 스펙트럼 홀이 줄어드는 것과 같이 높은 주파수 대역의 양자화 잡음은 줄어들었지만, 전체적인 음질에 미치는 영향은 아주 적었다. 이것으로부터 선형예측 전처리기를 통한 비트 할당 전략의 수정은 매우 낮은 전송률에서 적합하다는 것을 알 수 있다.

V. 결 론

본 논문에서는 낮은 전송률에서 발생하는 HE-AAC의 문제점을 분석하고, 선형예측 전처리기를 사용하여 문제점을 해결하는 방법을 제안하였다.

HE-AAC는 낮은 전송률에서 기존의 코더에 비해 매우 효과적인 코더임에도 불구하고 낮은 전송률에서는 비효율적으로 작동하는 HE-AAC의 비트 할당 전략으로 인하여 음질열화가 발생한다. HE-AAC의 전략은 높은 전송률에서는 매우 효과적으로 작동을 하지만 낮은 전송률에서는 상대적으로 에너지가 적은 구간에 너무 적은 양의 정보량을 할당한다. 따라서 주위에 비해 상대적으로 적은 에너지가 분포하는 구간에서 적당한 코딩이 이루어지지 못하게 되고, 결국 전체적으로 심각한 음질의 저하를 불러오게 된다.

이러한 문제점을 해결하기 위해서 제안된 선형예측 전처리기는 입력신호의 스펙트럼을 평탄하게 해 줌으로써 상대적인 크기의 차이를 줄여주었다. 따라서 HE-AAC는 전체적으로 기존보다 균일한 bit의 할당을 할 수 있게 되었다. 또한, 보다 정확한 심리음향적인 특성의 적용을 위해서 입력 신호의 FFT 스펙트럼을 이용하여 얻은 정확한 마스킹 임계치를 적용하였다. 그리고 이 과정에서 LP 계수를 이용하여 마스킹 임계치를 정규화 시켜줌으로써 레지듀얼 도메인에서도 정확한 마스킹 임계치의 적용이 이루어지도록 하였다.

제안된 방법의 성능 평가를 위하여 주, 객관적인 방법을 사용하였다. 객관적인 방법에서는 출력 신호

의 NMR을 비교하여 낮은 전송률에서 제안된 방법이 기존보다 더 효과적으로 비트를 할당하여 사용함을 보였다. 그리고 주관적 방법으로는 MUSHRA 테스트를 수행하여 지각적으로 음질이 향상되었는지를 확인하였다. 이를 통하여 제안된 방법이 아주 낮은 전송률에서 음질을 향상시키고 있음을 확인할 수 있었다.

참 고 문 헌

- [1] E. Peter Noll, "Wideband Speech and Audio Coding," in *IEEE Communications Magazine*, vol.31, pp.34-44, Nov. 1993.
- [2] Ming Yang, "Low bit rate speech coding," in *IEEE Potentials*, vol.23, pp.32-36, Oct-Nov. 2004.
- [3] M.Wolters, K. Kjorling, D. Homm, and H. Purnhagen, "A closer look into MPEG-4 high efficiency AAC," in *115th AES Convention*, Preprint 5871, Oct. 2003.
- [4] K. Brandenburg, "Low bitrate audio coding-state-of-the-art, challenges and future directions," in *5th International Conference on Signal Processing Proceedings*, 2000, pp.1-4.
- [5] M. Wolters, K. Kjorling, D. Homm, and H. Purnhagen, "A closer look into MPEG-4 high efficiency AAC," in *Proc. 115th AES Convention*, Los Angeles, USA, Oct. 2003, paper 5871.
- [6] Redwan Salami et al, "Extended AMR-WB for High-Quality Audio on Mobile Devices," in *IEEE Communications Magazine*, May 2006.

이 재 성 (Jae-seong Lee)
한국통신학회 논문지 제31권 6C호 참고

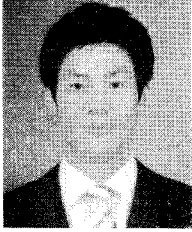
정회원

박 영 철 (Young-chul Lee)
한국통신학회 논문지 제29권 6C호 참고

정회원

이 건 우 (Gun-Woo Lee)

정회원



2007년 2월 연세대학교 전자
공학과 졸업
2009년 2월 연세대학교 전자
공학과 석사
2009년 3월~현재 삼성전자

<관심분야> 전자공학, 오디오신
호처리

윤 대 희 (Dae-hee Youn)
한국통신학회 논문지 제29권 6C호 참고

종신회원