

# 저전송률 오디오 부호화에서 음성 신호의 성능 개선을 위한 마스킹 임계값 적응기법 향상

## Enhanced Adjustment Strategy of Masking Threshold for Speech Signals in Low Bit-Rate Audio Coding

이 창 현\*, 강 흥 구\*  
(Chang-Heon Lee\*, Hong-Goo Kang\*)

\*연세대학교 전기전자공학과

(접수일자: 2009년 11월 9일; 수정일자: 2009년 12월 7일; 채택일자: 2009년 12월 21일)

본 논문에서는 기존 마스킹 임계값 적응 방식을 개선하여 저전송률 오디오 부호화에서 음성 신호에 대한 성능을 향상시킨다. 포먼트 영역 검색 이후, 각 포먼트 영역의 평균 에너지와 해당 서브밴드의 에너지 비율을 이용하여 마스킹 임계값을 변화시킨다. 상대적으로 에너지가 큰 밴드에 대해서는 더 많은 양자화 노이즈가 허용되는 반면, 청각적으로 민감한 스펙트럴 밸리에서는 비트 할당을 높여 양자화 에러를 좀 더 줄인다. 이는 음성 부호화에서 널리 사용되는 지각 가중(perceptual weighting) 개념을 반영한 것이다. 객관적 음질 평가 결과, 제안한 알고리즘이 기존 방식에 비해 음성 신호에 대한 성능을 향상시킨다는 것을 확인하였다.

**핵심용어:** Enhanced aacPlus, HE-AAC, 심리음향 모델, 마스킹 임계값, 지각 가중, 오디오 부호화, 음성 부호화

**투고분야:** 음성처리분야 (2.2), 뉴미디어분야 (13.3)

This paper proposes a new masking threshold adjustment strategy to improve the performance for speech signals in low bit-rate audio coding. After determining formant regions, the masking threshold is adjusted by using the energy ratio of each sub-band to the average energy of each formant. More quantization noises are added to the bands that have relatively large energy, but less distortion is allowed in spectral valley regions by allocating more bits, which reflects the concept of perceptual weighting widely used in speech coding. From the results of objective speech quality measure, we verified that the proposed method improves quality for the speech input signals compared to the conventional one.

**Keywords:** Enhanced AacPlus, HE-AAC, Psychoacoustic Model, Masking Threshold, Perceptual Weighting, Audio Coding, Speech Coding

**ASK subject classification:** Speech Signal Processing (2.2), New Media (13.3)

### I. 서론

인지 오디오 부호화 (perceptual audio coding) 방식에서는 심리음향 모델 (psychoacoustic model) [1]에 의해 정의되는 마스킹 임계값 (masking threshold)을 기준으로 허용 가능한 양자화 노이즈 레벨을 결정하고, 이에 따라 각 서브밴드에 비트를 할당함으로써 복호화된 신호의 음질을 원음과 동등하게 만드는 것을 목표로 한다 [2].

그러나 저전송률 모드에서는 비트 부족으로 인하여 모든 서브밴드에서의 양자화 노이즈 레벨을 마스킹 임계값 이하로 낮추는 것이 불가능하다. 이 경우, 양자화 노이즈에 의한 음질 저하를 최소화하면서 정해진 비트율 (bit-rate)을 만족시킬 수 있는 효율적인 마스킹 임계값 적응기법이 필요하게 된다.

MPEG HE-AAC (High-Efficiency Advanced Audio Coding) [3]의 한 종류로서 3GPP (3<sup>rd</sup> Generation Partnership Project)에 의해 표준화된 Enhanced aacPlus 오디오 부호화기에서는, 정해진 비트율을 만족할 때까지 모든 서브밴드의 마스킹 임계값에 동일한 오프셋

(offset)을 더하여 증가시키는 방법을 사용한다 [4]. 이 과정에서, 모든 서브밴드의 마스킹 임계값이 동일한 육셋만큼 증가한다는 것은 결국 각 서브밴드에 대해 같은 양의 양자화 노이즈 레벨을 허용한다는 것을 의미한다. 그러나 이는 스펙트럼의 특성을 전혀 고려하지 않는 방식이라는 측면에서, 특히 스펙트럼의 변화가 다양한 음성 신호에 대해서는 개선의 여지가 충분히 있다고 할 수 있다.

우리는 기존 연구에서 실제 코딩이 되는 전체 서브밴드의 평균 에너지와 각 서브밴드의 에너지 비율을 이용하여 가중치를 결정한 후, 이를 각 마스킹 임계값에 적용함으로써 서브밴드의 에너지 분포에 따라 다르게 마스킹 임계값을 적용시키는 방법을 제안하였다 [5]. 포먼트(formant) 주파수 영역과 같이 에너지가 큰 밴드들에 대해서는 마스킹 임계값을 기존 방식보다 더 높여 할당되는 비트수를 줄임으로써 양자화 노이즈를 더 허용하는 반면, 스펙트럴 밸리(spectral valley) 영역에서는 기존 방식보다 마스킹 임계값을 낮추어 보다 많은 비트할당을 통해 양자화 잡음을 더 줄일 수 있게 하였다. 이는 기존의 많은 저전송률 표준 음성 부호화기에서 널리 사용되고 있는 지각 가중(perceptual weighting) 개념을 반영한 것이다 [6]. 하지만 이 방법은 첫 번째 또는 두 번째 포먼트 영역의 에너지가 상대적으로 매우 큰 경우, 전체 서브밴드의 평균 에너지가 지나치게 높아짐으로써 나머지 포먼트들에 대한 특성이 잘 반영되지 않는 문제점을 지니고 있다.

이를 해결하기 위해, 본 논문에서는 개선된 가중치 결정 방법을 제안하고 이를 이용하여 마스킹 임계값을 정해진 비트율에 맞게 적응시킨다. 가중치를 계산하는 과정에서 코딩이 되는 전체 서브밴드의 평균 에너지 대신에 추정된 각 포먼트 영역에서의 로컬 평균 에너지와 그 영역에 포함되는 각 서브밴드의 에너지 비율을 사용함으로써, 기존 연구에서 제기된 문제점을 해결할 수 있다. 개선된 방식으로 가중치를 결정한 후, 한정 최소 사승(constrained least square) 방법을 이용하여 마스킹 임계값의 최종 변화량을 계산한다 [7].

객관적 음질 평가를 통하여 저전송률 오디오 부호화에서 추가적인 비트 할당 없이 음성 신호의 음질을 향상시키기 위한 제안한 알고리즘의 효율성을 검증한다.

## II. 동일 세기(loudness)의 양자화 노이즈 허용 방법

3GPP Enhanced aacPlus 오디오 부호화기에서는 정해진 비트율을 맞추기 위해, 모든 서브밴드에 대하여 동일 세기의 양자화 노이즈를 허용하면서 다음과 같이 마스킹 임계값을 증가시킨다 [4].

$$T_r(n) = (T(n)^{0.25} + r)^4, \quad (1)$$

이 식에서  $T(n)$ 은 심리음향 모델에 의해 얻어진  $n$  번째 서브밴드의 마스킹 임계값을 나타내고, 0.25 지수 값은 loudness 단위로 근사화하기 위해 사용된다.  $r$ 는 각 서브밴드에 동일하게 더해지는 loudness 값을 의미한다. 증가된 마스킹 임계값  $T_r(n)$ 으로부터 전체 지각 엔트로피(perceptual entropy)는 아래와 같이 근사화되고, 이 식에서  $pe$ 를 정해진 비트율에 맞는 엔트로피  $pe_c$ 로 대체함으로써  $r$ 에 대응하는  $r$  값을 구할 수 있다.

$$pe = \sum_n pe(n) = \sum_n l_q(n) \log_2 \left( \frac{E(n)}{(T(n)^{0.25} + r)^4} \right) \quad (2)$$

$$\approx A - 4B \log_2 (T_{av}^{0.25} + r),$$

이 식에서  $pe(n)$ 과  $E(n)$ 은 각각  $n$  번째 서브밴드의 지각 엔트로피와 에너지를 의미하고,  $l_q(n)$ 은 0값으로 양자화되지 않을 주파수 빈 개수의 추정치를 나타낸다.  $A$ 값은  $\sum_n l_q(n) \log_2 E(n)$ 와 같고,  $B$ 값은  $\sum_n l_q(n)$ 에 의해 결정된다. 또한, 식 (2)에서  $T_{av}$  값은 전체 마스킹 임계값들의 추정 평균값을 의미한다. 결국,  $r$  값은 다음과 같이 계산되고, 이에 따르는 지각 엔트로피 값과  $pe_c$ 과의 차이가 기준 임계치보다 크면 조건을 만족할 때까지 위의 과정을 반복하게 된다.

$$r = 2^{(A-pe_c)/4B} - T_{av}^{0.25}. \quad (3)$$

## III. 지각 가중식 마스킹 임계값 적응 방식

우리의 이전 연구에서는, 식 (1)의 마스킹 임계값 변형 방식에서 상수값  $r$ 에 가중치를 적용하여 아래와 같은 적응 방법을 제안하였다 [5].

$$T_r(n) = (T(n)^{0.25} + w(n)r)^4, \quad (4)$$

### 3.1. 기존 가중치결정 방법

이전 연구에서  $w(n)$  값은 식 (5)와 같이 코딩 될 전체 서브밴드의 평균 에너지와 각 서브밴드의 에너지 비율로 결정되는데, 이는 음성 신호에 있어서 포먼트 영역과 같이 상대적으로 에너지가 큰 밴드들에 대해서는 1보다 큰 값을 갖고 스펙트럴 벨리 영역에서는 1보다 작은 값을 갖게 된다. 다시 말해, 포먼트 영역에서는 마스킹 임계값을 기존 방식보다 더 높여 할당되는 비트 수를 줄임으로써 양자화 노이즈를 더 허용하고, 벨리 영역에서는 더 많은 비트를 할당하여 노이즈를 좀 더 줄일 수 있게 된다. 이는 높은 에너지를 갖는 주파수 영역보다 낮은 에너지 영역에서의 노이즈 성분이 사람 귀에 더 민감하다는 사실을 반영한 것으로, 음성 부호화 과정에 널리 사용되고 있는 각각 가중 (perceptual weighting) 개념과도 맥락을 같이 한다. 클래식이나 팝 음악과 같이 에너지가 전 주파수 영역에 고르게 분포되어 있는 경우에는 평균에너지와 각 서브밴드 에너지의 차이가 크지 않으므로, 이 경우에 있어서는  $w(n)$  값이 거의 1에 가까워 기존 식 (1)과 같은 마스킹 임계값 적용 방식을 따를 것이다. 다시 말해, 많은 음악 신호에 대해서는 기존의 성능을 유지하면서 음성 신호 그 중에서도 특히 유성음 구간에 대하여 가중치 적용 방식이 영향을 미치게 된다.

$$w(n) = \frac{Es(n)}{\frac{1}{N} \sum_{n=1}^N Es(n)} \quad (5)$$

이 식에서  $N$ 은 코딩 될 전체 서브밴드의 수를 의미하고,  $Es(n)$ 은 에너지 spreading 함수에 의해 얻어진  $n$  번째 밴드의 spreaded 에너지 값을 나타낸다. 각 밴드의 실제 에너지  $E(n)$ 보다  $Es(n)$ 이 스펙트럴 포락선 (envelope)에 더 가깝기 때문에,  $Es(n)$ 을 이용하는 것이 각각 가중 효과를 적용하는데 있어 더 효율적이다.

그런데 지나친 가중치들의 변화는 오히려 심각한 음질 저하를 유발할 수 있기 때문에, 이를 방지하기 위해 다음과 같은 매핑(mapping) 과정을 통해  $w(n)$ 의 변화를 완화하였다.

$$\tilde{w}(n) = \frac{1}{1 + e^{\alpha(1 - w(n))}} + 0.5, \quad (6)$$

이 때,  $\alpha$ 값은 실험에 의해 1로 고정하였다.  $w(n)$ 이 1이면  $\tilde{w}(n)$  역시 1의 값을 갖고,  $w(n)$ 이 음수가 될 수 없으므로 최소  $1/(1+e)+0.5$ , 최대 1.5의 값을 갖게 된다. 최종적으로 얻어진  $\tilde{w}(n)$  값이 식 (4)의  $w(n)$ 으로 사용된다.

그러나 이 방법은 첫 번째나 두 번째 포먼트 영역의 에너지가 상대적으로 매우 큰 경우 평균 에너지 값이 지나치게 높아짐으로써 다른 포먼트 영역의 특성을 제대로 반영하지 못할 뿐만 아니라, 첫 번째나 두 번째 포먼트 영역의 마스킹 임계값만 지나치게 높여 중요 하모닉 성분들에 대한 왜곡 현상을 심화시킬 수 있는 문제점을 안고 있다.

### 3.2. 개선된 가중치 결정 방법

이러한 문제점들을 해결하기 위해, 가중치 계산 방법에서 전체 서브밴드의 평균 에너지를 이용하지 않고, 각 포먼트 영역의 로컬 평균 에너지와 해당 서브밴드의 에너지 비율을 통해 가중치를 결정하였다. 아래 식은 제안한 가중치 결정 방법을 나타낸다.

$$w_i(n) = \frac{Es(n)}{\sum_{k=n_i}^{n_{i+1}-1} Es(k) / (n_{i+1} - n_i)}, \quad \text{for } 0 \leq n_i \leq n < n_{i+1} \leq N \quad (7)$$

위 식에서  $n_i$ 는  $Es(n)$ 의  $i$ 번째 극소점에 해당하는 서브밴드의 인덱스 (index)를 의미하고, 이 극소점들은 포먼트들을 구분짓는 역할을 한다. 따라서 가중치 값들은 각 포먼트 영역 내에서 해당 영역의 평균 에너지를 바탕으로 계산된다.  $n_i$ 가 적절히 포먼트 영역들을 구분 질 수 있다면, 기존의 방식과 달리 높은 주파수 영역에 있는 포먼트들의 특성도 잘 반영할 수 있는 장점을 지니고 있다.  $n_i$  결정 방식은 다음과 같다.

우선  $d(n) = Es(n) - Es(n-1)$ 을 모든 서브밴드들에 대하여 계산한 후,  $n$ 번째 밴드가  $d(n-1)$ 과  $d(n)$  모두 음수,  $d(n+1)$ 과  $d(n+2)$ 가 모두 양수인 조건을 만족하면 그 밴드를 극소점에 해당하는 서브밴드로 결정한다. 위의 과정을 통해  $M$ 개의 극소 밴드들이 결정되었다면  $(n_1, \dots, n_M < N)$ ,  $n_0$ 와  $n_{M+1}$ 이 각각 0과  $N$ 으로 정해지게 되는데, 이는 양 끝 쪽에서 가중치 값들의 계산을 위해 필요하기 때문이다. 그림 1.(a)는 음성 신호의 유성음에 해당하는 한 프레임에 대하여 선택된 극소 밴드들의 예시를 보여준다. 선택 과정에서  $N$ 은 약 4 kHz 주파수에 해

당하는 밴드 인덱스로 결정하였는데, 이는 그 이상의 주파수에서는 음성 신호의 포먼트 성분들이 뚜렷하지 않기 때문이다. 그림에 나타난 바와 같이, 각 극소 밴드들에 의해 주요 포먼트 영역들이 잘 구분되고 있는 것을 알 수 있다. 만약 조건을 만족하는 극소 밴드가 전혀 결정되지 않았다면, 모든 가중치 값들은 1로 정해져 불필요한 변형으로 인한 음질 저하를 방지하게 된다.  $N$  이상의 서브밴드들에 대한 가중치는 불필요한 변형을 방지하기 위해 아래와 같이 1로 정하였다.

$$w(n) = \begin{cases} w_r(n), & \text{for } n < N \\ 1.0, & \text{for } n \geq N \end{cases} \quad (8)$$

아울러 기존 방식과 마찬가지로 가중치들의 지나친 변화를 방지하기 위해 식 (6)의 매핑 함수를 사용하였는데, 변화를 좀 더 완화시키기 위해  $\alpha$  값은 0.6으로 결정하였

다. 그림 2는 이러한 매핑 함수를 나타낸 것이다. 최대 한계 값에는 변화가 없지만, 최소 값은  $1/(1+e^{\alpha n})+0.5$ 으로 약간 커져 전체 변화 폭을 좀 더 완화시키게 된다. 이와 같이 완화된  $\tilde{w}(n)$  값이 식 (4)의  $w(n)$ 으로 사용된다.

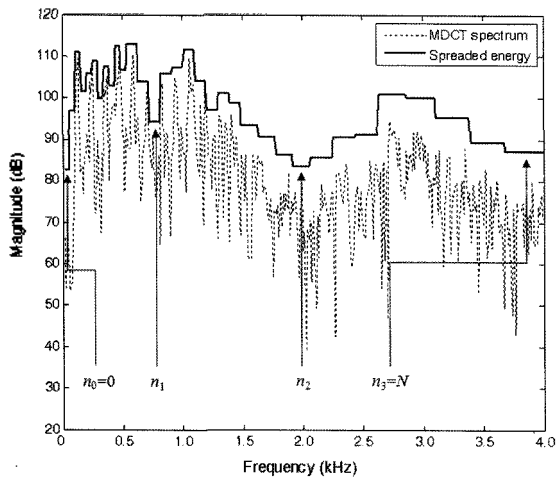
그림 1.(b)에서는 앞서 설명한 기존 (Conventional) 방식과 제안한 (Proposed) 방식에 의해 결정된 최종 가중치들을 비교하여 나타내었다. 가중치가 1을 기준으로 마스크 임계값의 증가와 감소 여부를 결정한다고 볼 때, 제안한 방식은 기존 방식에 비해 세 번째 포먼트 영역에 대한 특성을 잘 반영하는 것을 알 수 있다. 뿐만 아니라, 첫 번째 포먼트 영역에서 지나치게 마스크 임계값을 높인 기존 방식에 비해 증가 정도를 완화함으로써 중요 포먼트 성분들에 대한 왜곡 심화 현상을 방지할 수 있음을 확인하였다. 그러나 이 또한 최적의 파라미터는 아니기 때문에, 이에 대한 연구가 더 필요하다.

### 3.3. $r$ 값 결정

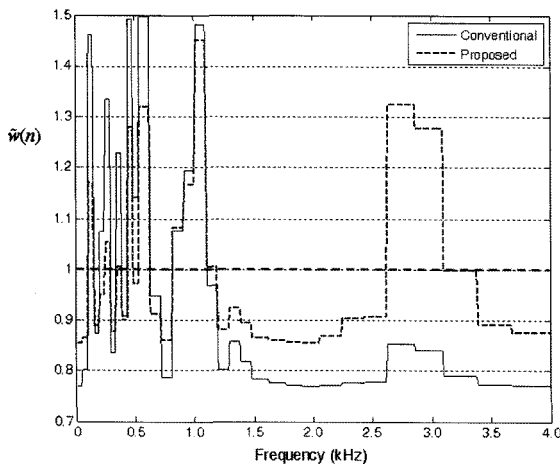
식 (1)에 나타낸 바와 같이 가중치  $w(n)$ 을 구한 후, 최종  $r$ 값을 구해야 제안한 마스크 임계값 적응 방식을 적용할 수 있다.  $r$ 값은 우리의 이전 연구에서 제안했던 한정 최소 자승(constrained least square) 방법을 통하여 계산한다. 우선 아래와 같이 두 근인  $r_1$ 과  $r_2$ 를 결정한다.

$$r_1 = \max\left(\frac{c_3}{c_1\lambda_1 - c_2}, 0\right), \quad r_2 = \max\left(\frac{c_3}{c_1\lambda_2 - c_2}, 0\right),$$

$$(\lambda_1, \lambda_2) = \text{Re}\left\{\frac{(2c_2c_4 - c_1^2) \pm c_1\sqrt{c_3^2 + 2c_1c_4}}{2c_1c_4}\right\}. \quad (9)$$



(a) 극소 밴드 선택 예시



(b) 가중치 비교

그림 1. (a) 극소 밴드 선택 예시, (b) 가중치 비교  
Fig. 1. (a) An example of selected local minimum bands, (b) comparison of weighting values.

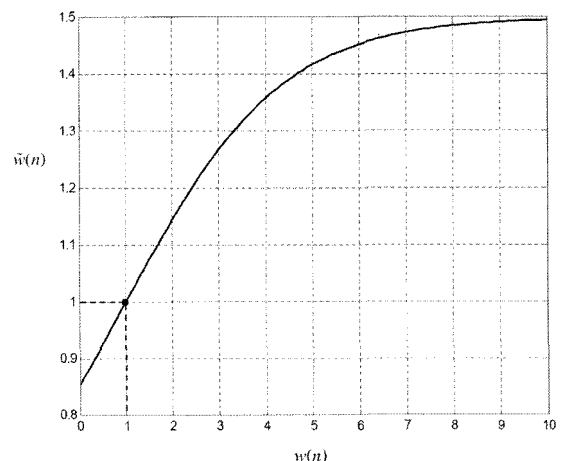


그림 2.  $w(n)$ 에서  $\tilde{w}(n)$ 으로의 매핑 함수  
Fig. 2. Mapping function from  $w(n)$  to  $\tilde{w}(n)$ .

식 (9)에서 사용된 각 계수들은 다음과 같이 계산되어 적용된다.

$$c_1 = \frac{1}{\ln 2} \sum_{n=1}^{N_c} \left( \frac{l_q(n) \tilde{w}^2(n)}{T(n)^{0.5}} \right), \quad c_2 = \sum_{n=1}^{N_c} 2 \tilde{w}^2(n),$$

$$c_3 = \frac{1}{\ln 2} \sum_{n=1}^{N_c} \left( \frac{l_q(n) \tilde{w}(n)}{T(n)^{0.25}} \right),$$

$$c_4 = \sum_{n=1}^{N_c} l_q(n) \log_2(T(n)^{0.25}) - \left( \sum_{n=1}^{N_c} l_q(n) \log_2(E(n)) - pe_r \right) / 4. \quad (10)$$

이 식에서,  $l_q(n)$  과  $E(n)$ 은 식 (2)에서 설명한 것과 같은 의미이고,  $N_c$ 는 실제 코딩이 이루어지는 서브밴드 수를 나타낸다. 그리고  $pe_r$ 은 주어진 비트율에 맞는 시각 엔트로피 (perceptual entropy)를 나타낸다. 최종  $r$ 값은 두 근 중에서 식 (11)과 같이 결정된다. 두 근 모두 양수인 경우에는 더 작은 양자화 노이즈를 허용하기 위해 둘 중 작은 값을 선택하고, 그렇지 않은 경우에는 큰 값을 선택한다. 식 (9)에 나타난 것과 같이  $r_1$ 과  $r_2$ 값들은 음의 값에 대해서 강제로 0이 되기 때문에, 근이 0인 경우에는 정확한 값이라 할 수 없다. 따라서 이 경우에는 둘 중 더 큰 값을 선택하게 된다.

$$\hat{r} = \begin{cases} \min(r_1, r_2), & \text{if } r_1 > 0 \text{ and } r_2 > 0 \\ \max(r_1, r_2), & \text{otherwise} \end{cases} \quad (11)$$

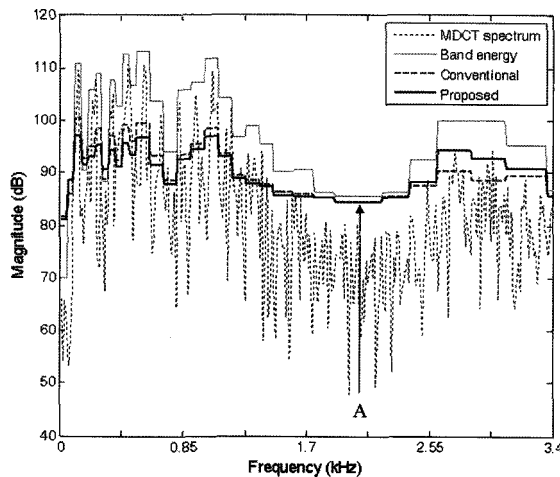


그림 3. 14 kbps mono에서의 유성음 구간에 대한 마스킹 임계값 비교

Fig. 3. Masking threshold comparison for speech vowel segment at 14 kbps mono.

## IV. 실험 결과 및 토의

본 논문에서 좀 더 개선한 마스킹 임계값 적응 방식을 3GPP Enhanced aacPlus 오디오 부호화기에 구현하여 실험하였다. Enhanced aacPlus 오디오 부호화기는 코딩 효율을 높이기 위해 코어 레이어 (core layer)에서 저주파 성분들을 부호화하고, 저주파 정보를 이용한 SBR (Spectral Band Replication) 방식을 통해 적은 비트로 고주파 성분들을 부호화한다 [8]. 이 과정에서 코어 레이어에서 코딩하는 주파수 대역이 비트율에 따라 다르게 정해지는데, 예를 들어 16 kbps 이하의 모노 (mono) 모드인 경우 약 3.4 kHz 까지, 18 kbps 모노 모드에서는 약 4.2 kHz 까지 코어 레이어에서 코딩한다. 마스킹 임계값은 코어 레이어 코딩에만 사용되기 때문에, 제안한 적응 방식에 대한 성능 평가 역시 SBR 기술을 사용하지 않고 코어 레이어에 의해서만 부/복호화된 신호에 대해 시행하였다.

### 4.1. 마스킹 임계값(masking threshold) 비교

그림 3과 그림 4는 음성 유성음 구간에서 각각 14 kbps 모노, 18 kbps 모노에 대하여 기존 가중치 결정 방식에 의해 얻어진 마스킹 임계값을 제안한 방식과 비교하여 나타낸 것이다. 그림 1.(b)의 가중치 비교 결과에서 나타난 바와 같이, 기존 방식에 비하여 제안한 방식으로 얻어진 마스킹 임계값은 세 번째 포먼트 영역의 특성을 반영하여 보다 증가한 것을 알 수 있다. 반면, 첫 번째 포먼트 영역에 대해서는 기존 방식에 비해 마스킹 임계값 증가

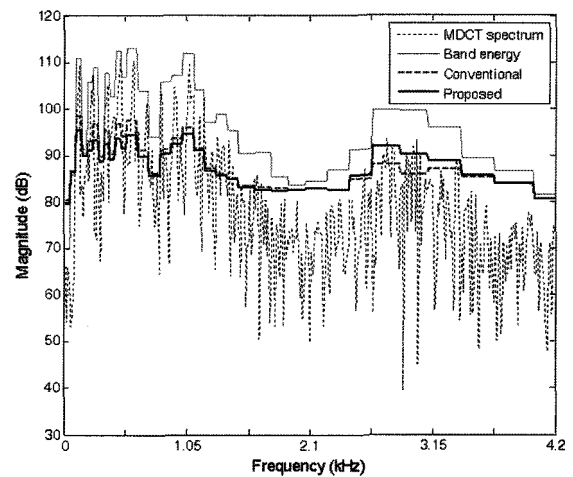


그림 4. 18 kbps mono에서의 유성음 구간에 대한 마스킹 임계값 비교

Fig. 4. Masking threshold comparison for speech vowel segment at 18 kbps mono.

표 1. PESQ 결과 비교

Table 1. Comparison of PESQ results.

	14 kbps mono		18 kbps mono	
	기존 방식	개선된 방식	기존 방식	개선된 방식
남성 샘플	3.61	3.62	3.78	3.79
여성 샘플	3.56	3.57	3.73	3.76
합계	3.59	3.60	3.76	3.78

정도를 완화하여 비트를 보다 많이 할당함으로써 중요 포먼트 성분들에 대한 왜곡 심화 현상을 개선하였다. 하지만 그림 3에 나타난 것과 같이, A와 같은 밴드에서 그림 1에 나타난 가중치의 효과가 적용되지 않고 기존 방식과 동일한 마스킹 임계값을 보이고 있는데, 이는 Enhanced aacPlus 부호화기에서 사용하고 있는 SHA (Spectral Hole Avoidance) 기술에 의한 것이다 [4]. 증가된 마스킹 임계값이 비트율에 따라 부호화기 내에서 정해지는 최소 SMR (Signal to Mask Ratio) 요구보다 커지게 되면, 임계값은 강제적으로 최소 SMR을 만족하도록 정해지게 된다. 이러한 현상은 특히 스펙트럴 밸리 영역에서 자주 발생하게 되는데, 그림 4에 보이는 것처럼 18 kbps 모노의 경우에도 이러한 현상은 동일하게 나타나고 있다.

#### 4.2. 객관적 음질 평가 결과

기존 가중치 결정 방식과 제안한 방식의 성능 평가를 위해 객관적 음질 평가를 수행하였다. 마스킹 임계값 적용 기법은 코어 레이어에서 부/복호화된 음성 신호에 대하여 음질 향상을 목적으로 하기 때문에, 남성 화자의 48개 음성 샘플과 여성 화자의 48개 음성 샘플로 이루어진 NIT-AT 한국어 데이터베이스를 이용하여 ITU-T P.862 PESQ 테스트를 비교 실시하였다. 실험을 위해 14 kbps 모노와 18 kbps 모노의 두 가지 모드를 사용하였다.

표 1은 기존 방식과 제안한 방식에 대하여 PESQ 결과를 비교하여 나타난 것이다. PESQ 수치로부터 제안한 알고리즘이 기존 가중치 결정 방법에 비해 음성 신호의 음질을 보다 개선한다는 것을 확인할 수 있다.

## V. 결론

본 논문에서는 개선된 마스킹 임계값 적용 알고리즘을 제안함으로써, 저전송률 오디오 부호화에서 음성 신호의 음질을 향상시킬 수 있었다. 제안한 방식은 추가적인 비트 할당과 디코더 구조의 변화를 요구하지 않는 면에서 또한 장점을 지니고 있다. 음성 신호에 대하여 포먼트 영역을 검색한 후, 각 포먼트 영역의 평균 에너지와 해당 서브밴드의 에너지 비율을 이용하여 마스킹 임계값을 변화시켰다. 제안한 방식은 기존 방식에서 반영하지 못했던 포먼트 영역에 대한 특성을 잘 반영할 수 있을 뿐 아니라, 상대적으로 에너지가 큰 중요 포먼트 성분들에 대한 왜곡 심화 현상 또한 방지할 수 있었다. 각 서브밴드에 대한 마스킹 임계값 증가 정도를 한정 최소 사승 방법을 이용하여 최종적으로 결정하였다. 객관적 음질 평가를 통해, 제안한 알고리즘이 기존 방식에 비해 음성 신호에 대한 성능을 보다 향상시킨다는 것을 확인하였다.

## 참고 문헌

1. E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Models*, 2nd Updated ed., New York: Springer, 1999.
2. J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Select. Areas Commun.*, vol. 6, pp. 314-323, 1988.
3. M. Wollers, K. Kjørling, D. Hømm, and H. Purnhagen, "A closer look into MPEG-4 High Efficiency AAC," *115th AES Convention*, New York, USA, Oct. 2003, preprint 5871.
4. 3GPP TS 26.403 v7.0.0, *Enhanced aacPlus general audio codec; Encoder specification; Advanced audio coding (AAC) part*, June, 2006.
5. C. H. Lee, H. O. Oh and H. G. Kang, "On the study of noise allocation for speech signal in low bit-rate audio coding," *IEEE Signal Processing Letters*, vol. 16, no. 10, pp. 849-852, 2009.
6. M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *J. Acoust. Soc. Amer.*, vol. 66, pp. 1647 - 1652, 1979.
7. E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*, Second ed., New York: Wiley, 2001.
8. 3GPP TS 26.401 v6.2.0, *Enhanced aacPlus general audio codec; General description*, Mar., 2005.

---

## 저자 약력

---

•이 창 현 (Chang-Heon Lee)



2003년 2월 : 연세대학교 공과대학 전기전자공학과  
(공학사)  
2005년 2월 : 연세대학교 공과대학 전기전자공학과  
(공학석사)  
2005년 3월~현재 : 연세대학교 공과대학  
전기전자공학과 박사과정

•강 홍 구 (Hong-Goo Kang)



연세대학교 전기전자공학부 교수  
연세대학교 디지털 신호처리 연구실  
(DSP lab, Yonsei University)