

## G.729.1 코더에서 프레임 간의 상호상관 관계를 이용한 개선된 스펙트럼 포락 코딩 방법

### Enhanced Spectral Envelope Coding Scheme Using Inter-frame Correlation for G.729.1

조근석<sup>1)</sup> · 성종모<sup>2)</sup> · 한민수<sup>3)</sup> · 김영일<sup>4)</sup> · 정상배<sup>5)</sup>

Cho, Keunseok · Sung, Jongmo · Hahn, Minsoo · Kim, Young-Il, Jung, Sangbae

#### ABSTRACT

This paper describes a new algorithm for encoding spectral envelope in the time domain alias cancellation (TDAC) part of G.729.1. The spectral envelope and modified discrete cosine transform (MDCT) coefficients of the weighted code-excited linear predictive (CELP) coding error in lower-band and the higher-band input signal are encoded in the TDAC part. In order to reduce allocation bits for spectral envelope coding, a new algorithm using sub-band correlation between adjacent frames is proposed. In addition, to improve the quality of decoded signals, two bit allocation strategies using reduced bits from the proposed algorithm are proposed. The performance of the proposed algorithm is evaluated in terms of objective quality and bit reduction rates. Experimental results show that the proposed algorithm increases the quality of sounds significantly.

**Keywords:** spectral envelope, TDAC, MDCT coefficient, G.729.1

#### 1. 서 론

최근 음성 코덱 기술은 협대역 코덱(300~3400Hz)에서 부족한 자연성과 명료성을 살리기 위해 광대역 코덱(50~7000Hz)에 대한 표준화가 이루어 졌고, 꾸준히 연구되고 있다. 또한 나아가, 음성 과 오디오 통합 코덱에 대한 기술 연구도 많이 이루어 지고 있다. 광대역 코덱(Wideband codec)은 50Hz~7000Hz의 대역폭을 갖는 광대역 신호를 압축하고 복원하는 기술이다. 현재 유선전화망이나 이동통신망에서는 300Hz~3400Hz의 대역폭을 갖는 협대역 (또는 전화선 대역)신호를 코딩하는 협대역 코덱을 주로 사용하고 있다. 광대역 신호는 협대역 신호에 비해 50Hz~300Hz의 대역이 추가되면서 자연성이 높아지고, 3400Hz ~7000Hz의 대역이 추가

되면서 명료성이 증가하여, 장시간 통화를 할 경우에도 피곤함이 덜한 것으로 알려져 있다. 광대역 코덱은 1988년부터 표준화되기 시작하였지만, 협대역 특성을 가지고 있는 기존 전화망과의 인터 페이스 때문에 널리 사용되지는 않았다. 그러나 최근 들어서 네트 워크 기술의 발달과 더불어 좀 더 넓은 음성 대역폭의 사용이 가능해지면서 협대역 코덱에 비해 음질이 우수한 광대역 코덱에 많은 관심이 모아지고 있다. 그 동안 광대역 코덱은 많은 국제 표준 화가 이루어져 왔는데, 2006년에는 ITU-T에서 G.729에 기반 하는 임베디드 가변 비트율 형태의 G.729.1을 표준화 하였다. G.729.1 은 인터넷 전화 서비스에서 가장 널리 사용되는 G.729 비트열과 호환성을 제공하고, 임베디드 형태로 비트열 계층을 쌓아가는 구조로 협대역 신호부터 광대역 신호를 압축, 복원한다[1].

기존의 embedded 형태의 비트스트림을 전송하는 scalable codec에서 transform 코딩에서 주로 MDCT를 사용하고 있다. G.729.1 코덱에서의 경우 CELP, time-domain bandwidth expansion (TDBWE), TDAC로 세 개의 모듈로 구성되어 있는데, 저대역 과 고대역 신호의 음질을 개선하기 위해 16~32kbps (2kbps 단위)에서 TDAC 모듈을 사용한다. 이 모듈은 50~4000Hz의 입력 신호와 CELP 모듈에서 재합성한 신호 사이의 차이, 4000~7000Hz의 입력신호를 MDCT 영역에서 부호화 한다. 오디오 코덱에서

- 1) 한국과학기술원 kscho83@kaist.ac.kr
- 2) 한국전자통신연구원 jmseong@etri.re.kr
- 3) 한국과학기술원 mshahn@ee.kaist.ac.kr
- 4) 경상대학교 공학연구원 yi@gnu.ac.kr
- 5) 경상대학교 공학연구원 jeongsb@gnu.ac.kr, 교신저자

접수일자: 2009년 10월 30일  
수정일자: 2009년 12월 7일  
게재결정: 2009년 12월 7일

도 MDCT 방법이 transform 방법으로 많이 사용 되어 왔다.

본 논문에서 G.729.1의 음질 개선을 위해 새로운 spectral envelope 정보를 인코딩 알고리즘을 제안하였다. 이 새로운 방법은 TDAC 인코더에서 이전 프레임과 현재 프레임의 같은 주파수 밴드의 spectral envelope 상호상관계수가 높은 것을 이용하여 할당되는 비트를 줄이고, 줄인 비트를 음질 개선을 위해 적절히 할당하는 것이다. 새로운 방법을 통해 얻은 비트는 두 가지 제안된 방안을 통해 비트 할당이 이루어진다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 G.729.1의 개요에 대해 간략히 소개되어 있고, 3장에는 제안된 새로운 알고리즘을 적용시킬 수 있는 기존의 TDAC 인코더에 대해 설명되어 있다. 4장에는 제안된 새로운 알고리즘이 상세히 설명되어 있고, 5장에서 새로운 알고리즘을 적용한 실험결과를 보이고, 6장에서 결론을 맺는다.

2. ITU-T G.729.1 CODER 개요

ITU-T에서 표준화된 G.729.1 코덱은 인터넷 전화 서비스에서 가장 널리 사용되는 G.729를 기반으로 8~32kbps 범위 내에서 임베디드 형태로 비트열 계층을 쌓아가는 구조로 협대역 신호부터 광대역 신호를 압축, 복원한다. VoIP에서 가장 널리 사용되는 협대역 코덱인 G.729의 비트열과 호환되도록 G.729.1은 G.729 비트열을 기본 계층(layer 1)으로 두고, 그 위에 비트열 계층을 쌓아가면서 고품질의 음질을 제공하는 구조이다. G.729.1 비트열 형식은 그림 1과 같다.

G.729.1 코덱은 20ms 단위로 16kHz 16bit PCM신호를 입력으로 받아 처리한다. 8kHz 샘플링 주파수인 입력 신호도 지원한다. 마찬가지로 출력은 8, 16 kHz의 샘플링 주파수를 갖는 16bit PCM 신호이다. G.729.1 코덱은 CELP, TDBWE, TDAC로 세 개의 모듈로 구성된다. CELP 모듈은 8, 12kbps에서 저대역 신호를 생성하고, TDBWE 모듈은 14kbps에서 고대역 신호를 생성한다. 저대역

과 고대역 신호의 음질을 개선하기 위해 16~32kbps (2kbps 단위)에서 TDAC 모듈을 사용한다[3]. 이 모듈은 50~ 4000Hz의 입력 신호와 CELP 모듈에서 재합성한 신호 사이의 차이, 4000~7000Hz의 입력신호를 MDCT 영역에서 부호화 한다. CELP 모듈은 G.729와 마찬가지로 10ms마다 동작한다[2]. G.729.1의 프레임 길이는 20ms이므로 CELP 모듈은 한 프레임에 두 번 동작한다. 가변대역 광대역 음성 코덱인 G.729.1 부호화기는 그림 2와 같은 구조를 가지며, 복호화기는 그림 3과 같은 구조를 가진다.

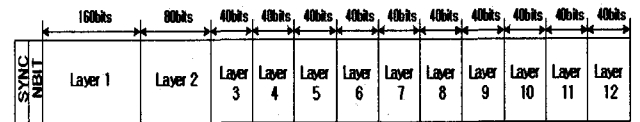


그림 1. G.729.1 비트열 형식  
Figure 1. Bitstream format of G.729.1

Quadrature mirror filtering 필터를 사용해서 입력 신호를 고대역과 저대역 신호로 분리한다. 저대역 신호는 2만쯤 decimation을 한 후 50Hz 이하 주파수 성분을 제거하기 위해 고대역 통과 필터를 통과시킨다. 이 신호는 CELP 모듈에 의해 8, 12kbps로 부호화 된다. 전처리된 저대역 신호와 CELP 모듈에 의해 재합성된 신호의 차이는 인지가중필터를 거친다. 이 때 차이 신호와 고대역 입력 신호 사이에 스펙트럼 연속성을 보장하기 위해서 이득 보상 과정을 거친다. 이 신호는 MDCT를 사용하여 주파수 영역으로 변환한다. 한편 고대역 입력 신호는 2만쯤 decimation을하고 주파수 대칭을 시킨다. 이 신호의 3000Hz 이상 성분을 저대역 통과 필터로 제거한다. 전처리된 신호는 TDBWE 모듈로 부호화되고, MDCT에 의해 주파수 영역으로 변환된다[4]. 저대역과 고대역의 MDCT 계수들은 TDAC 모듈로 부호화 된다[5]. 이 외에 프레임 손실 발생 시 음질 저하를 막기 위해 몇 개의 파라미터가 전송된다.

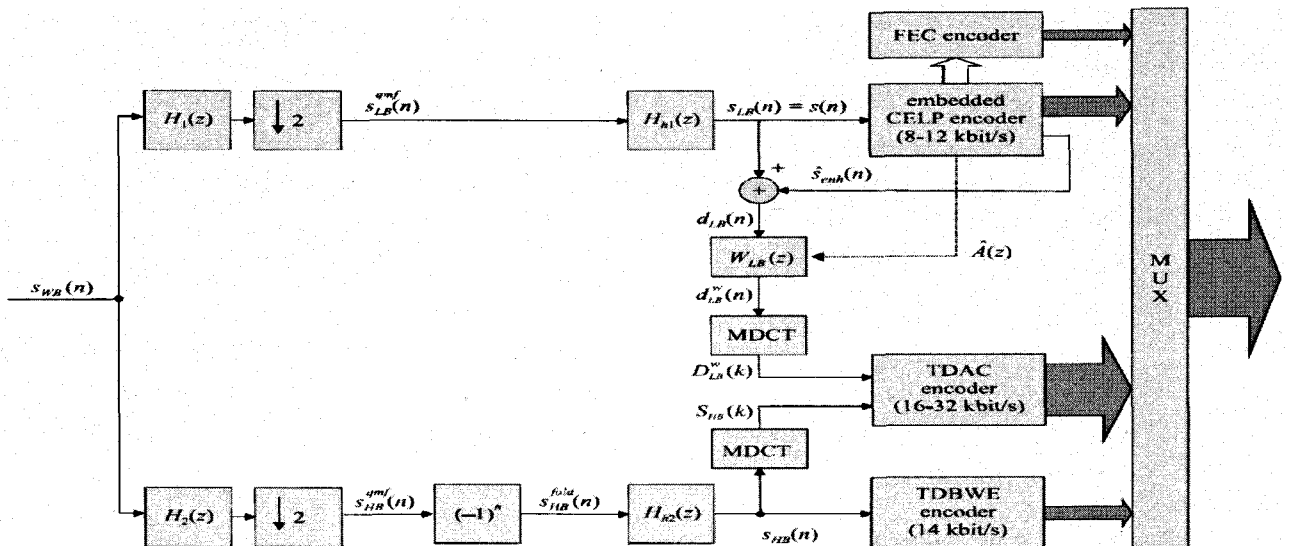


그림 2. G.729.1 부호화기 구조  
Figure 2. Block diagram of G.729.1 encoder

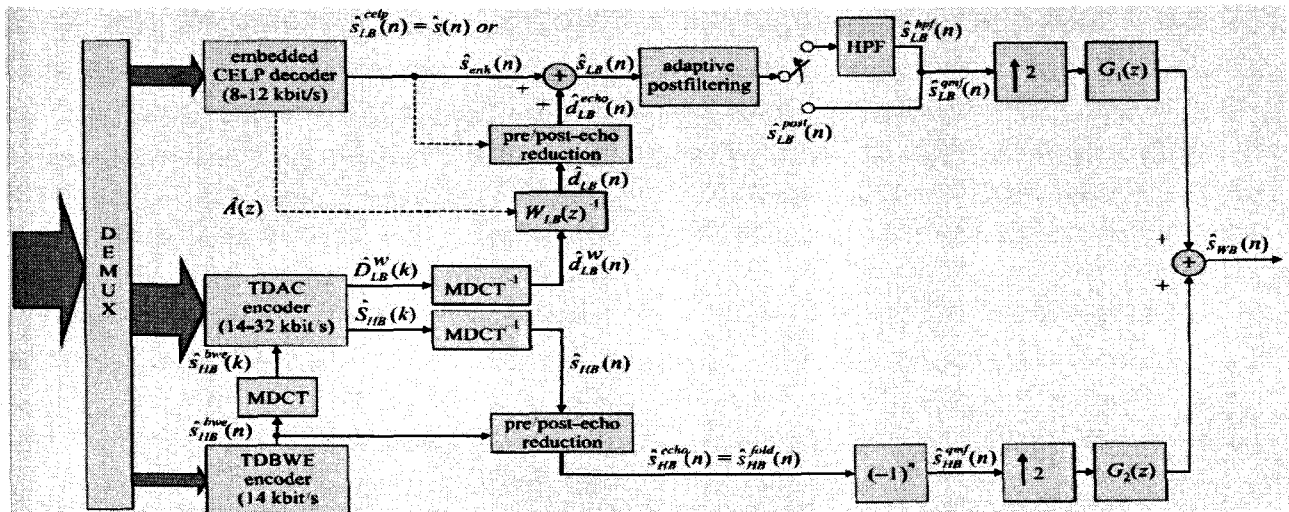


그림 3. G.729.1 복호화기 구조  
Figure 3. Block diagram of G.729.1 decoder

### 3. TDAC ENCODER

TDAC 인코더의 구성은 그림 4과 같다. TDAC에서 인코딩된 신호는 낮은 주파수 영역의 밴드에서의 인지적 가중치를 부여한 CELP 코딩의 에러 신호와 높은 주파수 영역의 밴드에서의 입력 음성 신호로 구성된다. 각 주파수 밴드들의 MDCT 계수들은 TDAC 인코더에서 gain-shape 벡터 양자화에 의해 얻어진다[1]. 0~7000Hz 밴드를 나타내는 MDCT 계수는 18개의 서브 밴드들로 나뉜다. 앞에서부터 17개의 서브 밴드는 16개의 MDCT 계수들로 구성되어 있고, 각 서브 밴드는 400 Hz 단위로 나뉜다. 그리고 마지막 18번째 밴드는 8개의 MDCT 계수들로 구성되어 있고, 200Hz 단위를 가진다. 각 서브 밴드에서 spectral envelope는 unif수들m 양자화 방식과 허프만 코딩기법을 이용하여 양자화 및 인코딩되며, MDCT 계수, 즉, spectral shape 정보는 spherical 벡터 양자화 방식을 통해 양자화 및 인코딩 된다. Spectral envelope은 root mean square (RMS)에 의해 log 도메인으로 나타내어진다. 식 (1)에 의해 spectral envelope이 구해진다.

$$\log_{rms}(j) = \frac{1}{2} \log_2 \left[ \frac{1}{N(j)} \sum_{k=b(j)}^{b(j+1)} Y^2(k) + \epsilon_{rms} \right], j=0, \dots, 17 \quad (1)$$

식 (1)에서  $\epsilon_{rms} = 2^{-24}$  를 나타내고, j와 k는 각각 서브밴드와 MDCT 인덱스를 나타낸다. 또한, N(j)와 b(j)는 각각 MDCT 계수 수와 j번째 서브밴드의 경계를 나타낸다. Y(k)는 전체 spectrum 밴드의 k번째 MDCT 계수이다. 식 (1)에서 구해진 Spectral envelope은 uniform scalar 양자화 방식을 통해 5비트로 양자화 된다. 5비트로 양자화 된 RMS 인덱스 값은 다시 3dB steps로 근사화 되게 된다. 근사화 된 인덱스들의 범위는

-11에서 +20사이의 값이다. 즉, 32가지 값으로 근사화 된다. 근사화 된 전체 밴드 envelope정보는 낮은 밴드의 spectral envelope ( $r(0) \sim r(9)$ )과 높은 밴드의 spectral envelope ( $r(10) \sim r(17)$ )로 두 가지의 서브벡터들로 나뉜다. 이런 과정을 거친 RMS 인덱스는 direction natural binary coding과 differential Huffman coding의 두 가지 방식 중 더 효율적인 방법이 계산되어 선택되어 양자화 된다. 위의 두 가지 방식은 낮은 밴드와 높은 밴드의 두 서브벡터에 각각 사용된다. 위에서 differential Huffman coding 기술은 평균 전송 비트율을 줄이기 위하여 사용된다. 반면에, direction natural binary coding 기술은 신호의 envelope정보를 differential Huffman coding 기술에 적용했을 때 오히려 비트 할당량이 더 커질 때를 방지하기 위하여 제한을 두기 위해 사용되는 방법이다. 디코더 측면에서 두 가지 방식 중 어느 방식이 이용되었는지 알기 위해서 하나의 flag 비트가 사용된다.

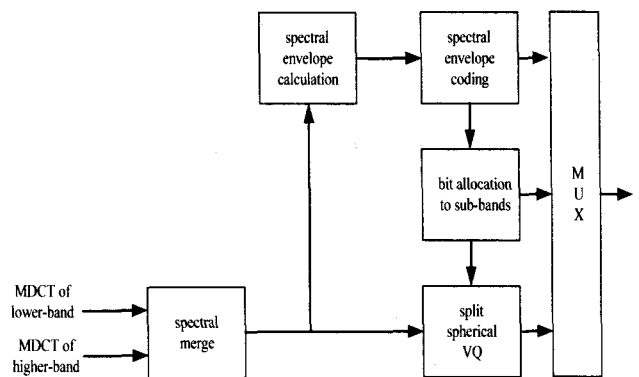


그림 4. TDAC encoder의 블록 다이어그램  
Figure 4. Block diagram of the TDAC encoder

Different Huffman coding은 다음의 세 가지 단계로 구성된다 [1]. 먼저, 첫 번째 인덱스인  $r(0)$ 이 5비트 할당을 통해 natural binary coding 방식을 사용하여 인코딩 된다. 이는 밴드 별로 인

텍스 차이 값을 양자화 할 때, 디코더 단에서 원래 값을 복원하기 위하여 기준이 되는 값은 원래의 값 그대로 전송해야 하기 때문이다. 두 번째로는 RMS 인덱스의 이전 밴드 값과의 차를 구하고, Huffman coding 방식을 이용할 경우 비트 이득이 있는지를 계산하여, flag 값을 얻는다. 식 (2)는 낮은 밴드의 경우에서 Huffman coding을 적용하기 위해 인덱스 차이 값을 구하는 방법을 나타낸다.

$$d(j) = r(j) - r(j-1), \quad j = 1, \dots, 9$$

$$mode = \begin{cases} 0 & \text{if } |d(j)| \leq 12 \text{ for } j = 1, \dots, 9 \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

여기서,  $d(j)$  는 인덱스 차이 값이며,  $j$ 는 서브밴드 값을 나타낸다. 마지막으로 비트 이득이 발생하는지의 여부에 따라 방식 flag값을 결정된다. flag값이 0일 경우 비트 이득이 발생한 것으로 Huffman coding 방식이 적용된 것이고, flag값이 1일 경우 비트 이득이 발생하지 않은 것으로 direct natural binary coding 이 적용된다. 높은 밴드에서도 이와 동일한 방식이 적용된다.

RMS 인덱스가 인코딩 된 후 식 (1)을 통해 구한 로그 RMS 를 이용하여 각 서브밴드 별로 인지적 중요도를 계산하게 된다. 인지적 중요도는 각 서브밴드 별 MDCT 계수를 인코딩 할 때 필요한 비트를 가변적으로 할당하기 위해 사용된다. 한정된 비트를 가지고 인지적 중요도에 따라 각 서브밴드 별로 비트가 할당된다[6]. 식 (3)으로부터 MDCT 계수를 인코딩 하기 위해 사용 가능한 전체 비트 수( $nbits\_VQ$ )를 구할 수 있다.

$$nbits\_VQ = 351 - nbits\_HB - nbits\_LB \quad (3)$$

$nbits\_HB$  과  $nbits\_LB$  는 각각 spectral envelope 의 낮은 주파수 밴드와 높은 주파수 밴드를 인코딩 할 때 사용된 비트 수를 나타낸다.

**4. 제안된 SPECTRAL ENVELOPE CODING 알고리즘**

실제 음성 오디오 데이터에서 하나의 일정한 프레임 내에서의 밴드 사이의 스펙트럼 상호상관 관계보다 인접한 프레임들 사이의 같은 밴드끼리의 상호상관 관계가 큰 경우가 많다. 이런 점을 이용하여 spectral envelope coding 에서 사용되는 비트 수를 좀 더 줄이기 위하여 기존 G.729.1에서 사용되는 방법 대신 인접한 프레임 사이의 상호상관 관계를 이용하는 방법을 제안하였다. 기존의 알고리즘은 한 프레임 내에서 서브 밴드와 서브 밴드의 차이 값을 Huffman coding 을 이용하였기 때문에 항상 첫 서브밴드는 기준 값이 되어 direct natural binary coding 을 통해 항상 5비트 할당을 해서 보내야 한다는 약점을 가지고 있

었다. 그러나 제안된 알고리즘에서는 이전 프레임 값을 참조하기 때문에 전체 서브 밴드들에 대해 Huffman coding 을 적용하여 비트 사용량을 감소시킬 수가 있다. 제안된 알고리즘을 사용했을 때, 비록 전체 프레임 중에 첫 번째 프레임의 서브 밴드 값들을 모두 direct natural binary coding 을 통해 인코딩 하여 전송해야 하지만, 전체 입력 음성 오디오 데이터 측면에서 첫 번째 서브 프레임 이후 모든 프레임의 서브 밴드 값들을 Huffman coding 을 적용할 수 있기 때문에 평균 비트 전송률을 줄일 수 있다. 제안된 알고리즘은 세 가지 단계로 구성된다. 첫째로, 인접한 프레임 사이, 즉 현재 프레임과 이전 프레임 사이에서 같은 서브 밴드의 RMS 인덱스 값의 차이를 구한다. 식 (4)를 통해 계산될 수 있다.

$$d^{(prop)}(j) = r(t, j) - r(t-1, j) \quad (4)$$

여기서  $t$  와  $j$  는 각각 프레임 인덱스와 서브 밴드 인덱스를 나타낸다. 두 번째 단계로, 제안된 알고리즘을 통해 spectral envelope의 전체 서브 밴드에 할당되는 비트 값이 계산된다. 마지막 단계로, 제안된 알고리즘에 의한 비트 감소율과 기존 알고리즘에 의한 비트 감소율을 비교하여 더 많은 비트가 감소하는 방법이 선택된다. 즉, 비트 감소를 위해 기존 알고리즘과 제안된 알고리즘이 병행되어 사용된다. 여기서 제안된 알고리즘과 기존 알고리즘 중에 해당 프레임에 사용된 방식이 무엇인가를 디코더 단에서 알 수 있도록 1 비트를 추가적으로 사용하여 frag를 전송한다. 물론, 프레임 마다 1 비트가 추가되는 단점이 생기지만, 실제 할당되는 평균 비트 율 측면에서 보면 많은 비트를 감소시킬 수가 있다. 그림 5는 제안된 알고리즘에 의해 spectral envelope coding 이 수행되는 절차를 블록 다이어그램으로 나타낸 것이다.

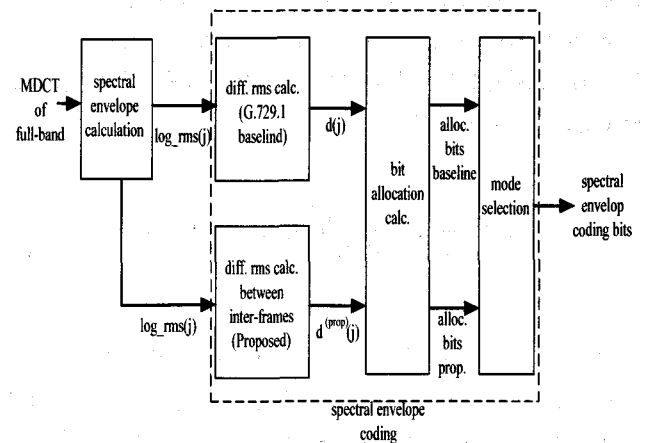


그림 5. 제안된 알고리즘을 적용한 spectral envelope coding 의 블록 다이어그램

Figure 5. Block diagram of spectral envelope coding by the proposed algorithm in the TDAC encoder

식 (3)에서 나타났듯이, 사용할 수 있는 비트가 전송률에 따라 고정되어 있기 때문에 spectral envelope을 인코딩 할 때 사용되는 비트가 적으면, 대신 MDCT 계수를 인코딩 할 때 사용될 수 있는 비트 수가 증가 하게 된다. 따라서 본 연구에서 제안된 알고리즘을 이용하여 비트 할당 감소를 통해 얻은 비트들을 이용하는 두 가지 방안을 제안하였다. 첫 번째 방안은 비트 감소를 통해 얻은 비트를 기존의 G.729.1에서의 방식대로 더 많은 비트를 모두 MDCT 계수를 인코딩 하는데 할당하는 것이다. 두 번째 방안은 비트 감소를 통해 얻은 비트를 spectral envelope 값의 resolution을 증가시켜 인코딩 하는 데에 사용을 하는 것이다. 즉, 제안된 방법을 통해 비트 감소를 통해 얻은 비트들을 spectral envelope을 좀 더 자세히 나타내는 데에 사용한다는 것이다. 이렇게 되면, 기존의 G.729.1에서 MDCT 계수를 인코딩 하는데 사용된 비트 수는 거의 변동 없이 제안된 알고리즘을 통해 spectral envelope을 더 세분화 하여 양자화 하여 인코딩 할 수가 있게 되므로 음질 개선을 이룰 수가 있다. Spectral envelope의 resolution은 식 (1)에서 나타낸 값의 두 배로 나타낸다. 예를 들어, 만약 원래 spectral envelope 인덱스 값이 1.7이면 기존의 비트 할당 전략에서 근사화 함수에 의해 실제 spectral envelope 인덱스 값은 1로 양자화 된다. 그러나 제안한 방안을 통해 resolution을 증가 시키도록 세분화를 적용시키면, spectral envelope 인덱스 값은 1.5의 값으로 양자화 시킬 수가 있다. 이런 경우 오차가 적기 때문에 디코더 단에서 spectral envelope을 좀 더 잘 복원할 수가 있다. 이 방안에서, 직접적으로 입력 신호를 처리할 수 있는 spectral envelope의 높은 밴드 부분에 대해서만 고려하여 적용하기로 하였다. 그 이유는 spectral envelope의 낮은 밴드의 경우에는 G.729.1 코더의 코어 코더인 G.729를 통과시켜 얻은 코딩 에러 신호를 가지고 계산되어 지기 때문이다. 따라서, 낮은 밴드 경우 높은 밴드의 경우보다 더 적은 음질 개선이 발생될 것이 예상됨에 따라 비트의 효율적인 할당을 위해 높은 밴드 부분에 대해서 spectral envelope의 resolution을 높이는 데에만 비트 감소를 통해 얻은 비트를 할당 하였다.

## 5. 실험 및 결과

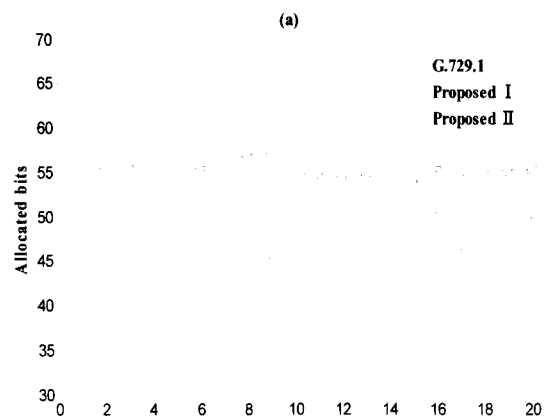
제안된 spectral envelope coding 알고리즘의 성능을 평가하기 위해 비트 감소율과 spectral distortion을 측정하였다. 실험 데이터로 한국어인 발성 음성과 오디오 신호가 사용되었다. 전체 음성과 오디오 데이터는 각각 20개씩 사용되었으며, 데이터의 길이는 대략 20초 정도이다. 각 데이터는 16 kHz로 샘플링 되었으며, 데이터의 resolution은 16비트이다.

먼저 비트 감소율 측면의 평가를 실시하였다. 그림 6은 프레임 별 비트 할당 평균 값을 나타낸다. 그림 6을 통해 제안된 알고리즘을 적용했을 때의 비트 할당량이 기존의 알고리즘을 적용했을 때 보다 더 작은 것을 알 수가 있다. 이것은 제안된 인

접 프레임을 적용한 방법의 경우 Huffman coding을 이용했을 때 인코딩에 필요한 비트를 줄이는 데에 더 효과적이라는 것을 의미한다. 즉 비트 감소율이 기존 알고리즘에 비해 제안된 알고리즘을 적용했을 때 더 크다는 것이다. 제안된 알고리즘을 통해 같은 spectral envelope을 인코딩함에 있어서 기존의 알고리즘에 비해 더 적은 양의 비트 할당으로도 가능하게 되었다. 실험 전체 데이터에 대하여 실험한 결과, 음성신호의 경우, 기존 알고리즘을 사용할 경우 프레임 당 약 55.449비트가 사용되었으며, 제안된 알고리즘을 사용할 경우 프레임 당 약 49.841비트가 사용되었다. 결과적으로 프레임 당 약 5.608 비트를 줄일 수가 있었다. 실험 전체 데이터에 대하여 오디오 신호의 경우, 기존 알고리즘을 사용할 경우 약 56.054비트가 사용되었으며, 제안된 알고리즘을 사용할 경우 약 48.286비트가 사용되었다. 프레임 당 약 7.768 비트를 줄일 수가 있었다. 실험 결과에서 오디오 데이터에 대해 더 많은 비트를 줄일 수가 있었던 것은 인접프레임이 유사한 경우가 많을 경우 비트를 많이 줄일 수 있다는 측면에서, 오디오 데이터가 음성 데이터에 비해 상대적으로 변화가 크지 않는 프레임들의 수가 많기 때문이다. 비트 감소율에 이어 음질 평가를 실시하였다. 음질 실험 평가를 위하여 객관적 수치를 나타낼 수 있는 Bark's scale distortion (BSD) 을 사용하였다[7]. 원음과 디코딩을 통해 나온 신호 사이의 BSD를 측정하였다. BSD를 계산하기 위해, 512-FFT (Fast Fourier Transform) 이 30 ms 분석 프레임으로부터 매 10 ms간격으로 수행되었으며, Bark's scale 필터 बैं크를 통해 계산된 33 energy 값이 예측 값으로 구해졌다. 식 (5)는 BSD를 나타낸다.

$$BSD = \frac{\sum_{t=0}^{T-1} (BSD^{(t)})}{\sum_{t=0}^{T-1} \left( \sum_{k=0}^{N-1} (L_x^{(t)}(k))^2 \right)} \quad (5)$$

여기서,  $t$ ,  $k$ 는 각각 프레임 인덱스와 필터 बैं크 인덱스를 나타내며,  $T$ 와  $N$ 은 각각 전체 프레임의 수와 전체 필터 बैं크의 수를 나타낸다.



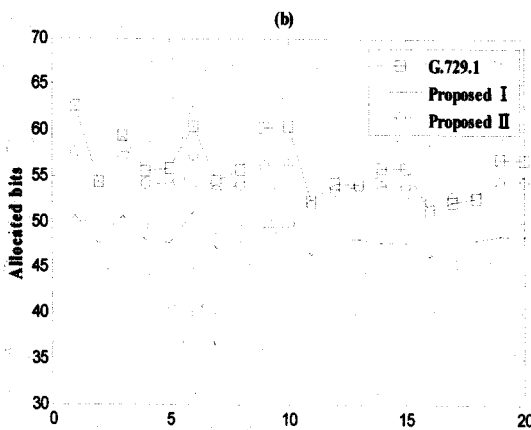


그림 6. Spectral envelope을 위해 할당된 평균 비트 수 ((a) 음성 DB, (b) 오디오 DB, proposed I: 첫 번째 비트 할당 방안, proposed II: 두 번째 비트 할당 방안)

Figure 6. Average number of bit allocation per frame for spectral envelope ((a) speech DB, (b) audio DB, proposed I: first bit allocation strategy, proposed II: bit allocation strategy)

$$BSD^{(i)} = \sum_{k=0}^{N-1} (L_x^{(i)}(k) - L_y^{(i)}(k))^2$$

는 표현될 수 있으며,

$L_x^{(i)}(k)$ ,  $L_y^{(i)}(k)$  는 각각 참고 될 신호의 필터 बैं크 에너지와 테스트 될 신호의 필터 बैं크 에너지를 나타낸다. 이전에도 언급했듯이 음질을 높이기 위한 첫 번째 방안은 줄인 비트들을 모두 MDCT 계수를 인코딩 하는 데에 할당하는 것이다. 비록 음질 개선의 정도가 다소 적긴 하지만, 음성 데이터에 대한 BSD결과가 오디오 데이터에 대한 결과보다 낮음을 확인할 수가 있다. 두 번째 방안은 줄인 비트들을 통해 얻은 비트들을 spectral envelope coding 의 resolution 을 늘리는데 할당하는 것이다. 이전에 언급했듯이 이는 spectral envelope 의 높은 주파수 밴드의 resolution을 두 배로 늘렸고, 이를 통해 얻은 결과는 그림 7을 통해 확인할 수가 있다. 전체 실험 데이터에 대해서 음성 신호의 경우 원래 알고리즘을 사용했을 시에 약 10.132의 BSD값을 얻었고, 첫 번째 방안을 적용 했을 시에는 약 10.009, 두 번째 방안을 적용 했을 시에는 약 9.441의 BSD값을 얻었다. 그리고 전체 실험 데이터에 대하여, 오디오 신호의 경우에는 원래 알고리즘을 사용했을 시에 약 11.544의 BSD값을 얻었고, 첫 번째 방안을 적용 했을 시에는 약 11.435, 두 번째 방안을 적용 했을 시에는 약 10.257의 BSD값을 얻었다. 그림 7에서 볼 수 있듯이 첫 번째 방안에 비해 spectral envelope 의 resolution을 높이는 두 번째 방안을 적용했을 시에 성능이 더 많이 향상됨을 확인할 수가 있다. 두 번째 방안을 적용한 것이 원래의 알고리즘을 통해 얻은 신호의 BSD값보다도 그리고 첫 번째 방안을 적용했을 때의 BSD 값보다도 작은 것을 알 수가 있다. 이는 BSD가 작을수록 원래의 신호에 가깝다는 것을 의미한다. 첫 번째 방안과 비교해 보았을 때, 두 번째 방안이 오디오 데이터의 음질을 개선하는 데에 더욱 효과적임을 확인할 수가 있다. 전

체 실험 결과로부터 spectral distortion 은 MDCT 계수보다 spectral envelope 에 의해 영향이 더 크다는 것을 확인할 수가 있었다.

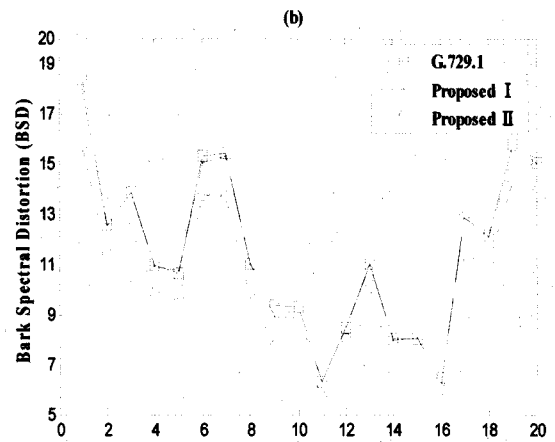
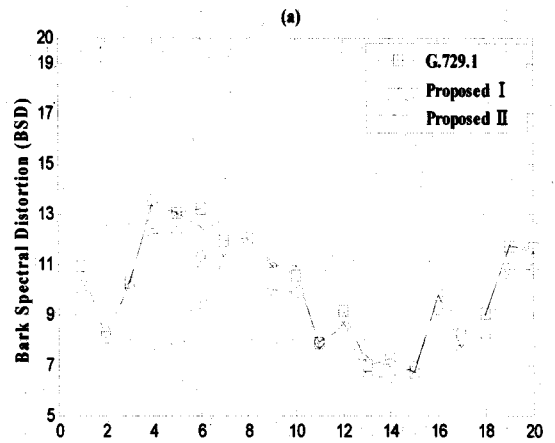


그림 7. 제안된 알고리즘에 의한 spectral distortion 결과 ((a) 음성 DB, (b) 오디오 DB, proposed I: 첫 번째 비트 할당 방안, proposed II: 두 번째 비트 할당 방안)

Figure 7. Spectral distortion result by proposed algorithm ((a) speech DB, (b) audio DB, proposed I: first bit allocation strategy, proposed II: bit allocation strategy)

### 6. 결 론

본 논문에서는 프레임 사이의 상호상관 관계를 이용하는 spectral envelope 정보 인코딩을 위한 새로운 알고리즘을 제안하고 실험을 통해 수행하였다. 비록 제안된 알고리즘은 전송 중에 프레임이 손실되었을 경우 이전 프레임을 참고 할 수 없게 되는 단점을 지니고 있지만, 실험 결과를 통해 제안된 알고리즘은 인코딩 하는 데 사용되는 비트를 줄이고 디코딩 된 음성 및 오디오 신호의 음질을 향상시키는데 효과적임을 확인할 수가 있다. 만약, 프레임이 손실되었을 경우, 기존 코더에서 FEC라는 패킷 손실에 대한 처리를 해주는 부분이 있으며, 본 제안된 방

식에서는 원래의 허프만 코딩 방식과 제안된 인접프레임의 상호관계를 이용한 방식 중 더 나은 비트이득을 얻는 방식을 프레임마다 결정하는 방식이므로, 실험 결과 두 가지 방식이 서로 번갈아 나오기 때문에 에러가 축적되지 않게 된다. 그리고 또한, 비트이득의 차이가 별로 없을 시에 주기적으로 기존의 방식을 어느 정도의 프레임 간격으로 고정 선택시키게 되면 프레임 손실에 의한 에러가 쌓이는 것을 방지할 수가 있다. 앞으로 제안된 알고리즘이 프레임 손실에서도 강인한 결과를 얻을 수 있는 방안에 관하여 연구를 진행 할 것이며, spectral envelope 정보를 추가적인 비트 할당 없이 좀 더 높은 resolution으로 양자화 하여 전송하는 방법에 관해서도 역시 연구를 진행 할 것이다.

### 감사의 글

본 연구는 지식경제부 및 정보통신진흥원의 IT신성장동력 핵심기술개발사업의 일환으로 수행된 결과의 일부입니다.  
[2008-S011-02,FMC 어커스틱 융합코덱 및 제어기술 연구]

### 참 고 문 헌

- [1] ITU-T Rec. G. 729.1, "G.729 based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729," 2006.
- [2] Ragot, S. et al, "ITU-T G.729.1: An 8-32 kbit/s scalable coder interoperable with G.729 for wideband telephony and Voice over IP," *Proc. ICASSP*, Vol. 4, pp. 529-532, May 2007.
- [3] Massaloux, D. et al, "An 8-12 kbit/s embedded CELP coder interoperable with ITU-T G.729 coder: First stage of the new G.729.1 standard," *Proc. ICASSP*, Vol. 4, pp. 1105-1108, May 2007.
- [4] Jung, S. et al, "An embedded variable bit-rate coder based on GSM EFR: EFR-EV," *Proc. ICASSP*, pp. 4765-4768, Mar 2008.
- [5] Geiser, B. et al, "Candidate proposal for ITU-T super-wideband speech and audio coding," *Proc. ICASSP*, pp. 4121-4124, April 2009.
- [6] Geiser, B., Jax, P., Vary, P., Taddei, H., Gartner, M., Schandl, S., "A qualified ITU-T G.729EV codec candidate for hierarchical speech and audio coding," *Proc. IEEE 8th Workshop on Multimedia Signal Processing*, pp. 114-118, Oct. 2006.
- [7] Wang, S., Sekey, A., Gersho, A., "An objective measure for predicting subjective quality of speech coders," *IEEE Trans. Selected Areas in Comm.*, Vol. 10, No. 5, pp. 819-829, June 1992.

#### • 조근석 (Cho, Keunseok)

한국과학기술원 전기 및 전자공학과  
대전광역시 유성구 문지동 103-6  
Tel: 042-350-6206

Email: kscho83@kaist.ac.kr

관심분야: 음성/오디오 코딩, 음성 합성

현재: 한국과학기술원 전기 및 전자공학과 박사과정

#### • 성종모 (Sung, Jongmo)

한국전자통신연구원 광대역통신망연구단

대전광역시 유성구 가정로 138

Tel: 042-860-1243

Email: jmseong@etri.re.kr

관심분야: 음성/오디오 코딩

현재: 한국전자통신연구원 선임연구원

#### • 한민수 (Hahn, Minsoo)

한국과학기술원 전기및전자공학과

대전광역시 유성구 문지동 103-6

Tel: 042-350-6206

Email: mshahn@ee.kaist.ac.kr

관심분야: 음성신호처리

현재: 한국과학기술원 전기 및 전자공학과 교수

#### • 김영일 (Kim, Young-Il)

경상대학교 공과대학(공학연구원)

경남 진주시 가좌동 900번지

Tel: 055-751-5352 Fax: 055-759-7814

Email: yi@gnu.ac.kr

관심분야: 음성신호처리

현재: 경상대학교 전자공학과 교수

#### • 정상배 (Jeong, Sangbae) 교신저자

경상대학교 공과대학(공학연구원)

경남 진주시 가좌동 900번지

Tel: 055-751-5357 Fax: 055-759-7814

Email: jeongsb@gnu.ac.kr

관심분야: 음성신호처리

현재: 경상대학교 전자공학과 조교수