

G.718 초광대역 코덱의 음질 향상을 위한 개선된 Generic Mode Coding 방법 Modified Generic Mode Coding Scheme for Enhanced Sound Quality of G.718 SWB

조근석¹⁾ · 정상배²⁾
Cho, Keunseok · Jeong, Sangbae

ABSTRACT

This paper describes a new algorithm for encoding spectral shape and envelope in the generic mode of G.718 super-wide band (SWB). In the G.718 SWB coder, generic mode coding and sinusoidal enhancement are used for the quantization of modified discrete cosine transform (MDCT)-based parameters in the high frequency band. In the generic mode, the high frequency band is divided into sub-bands and for every sub-band the most similar match with the selected similarity criteria is searched from the coded and envelope normalized wideband content. In order to improve the quantization scheme in high frequency region of speech/audio signals, the modified generic mode by the improvement of the generic mode in G.718 SWB is proposed. In the proposed generic mode, perceptual vector quantization of spectral envelopes and the resolution increase for spectral copy are used. The performance of the proposed algorithm is evaluated in terms of objective quality. Experimental results show that the proposed algorithm increases the quality of sounds significantly.

Keywords: G.718 SWB, generic mode, MDCT, cepstrum, spectral envelope

1. 서론

최근 음성 코덱 기술은 협대역에서 광대역, 그리고 초광대역 코덱으로 발전해 나가고 있다. 기존에는 유선전화망이나 이동통신망에서는 300 Hz - 3400 Hz의 대역폭을 갖는 협대역 코덱을 주로 사용해 왔지만 50 Hz - 7 kHz의 대역폭을 갖는 광대역 코덱과 50 Hz - 14 kHz의 대역폭을 갖는 초광대역 코덱이 표준화가 되고 현재 광대역 코덱이 이동통신망에서 사용되는 추세이다[1][2]. 광대역 코덱은 1988년부터 표준화되기 시작하였지만, 협대역 특성을 가지고 있는 기존 전화망과의 인터페이스 때문에 널리 사용되지는 않다가, 네트워크 기술의 발전으로 좀 더 나은 통화품질이 요구되면서부터 음질이 우수한 광대역, 초광대역 코덱에 많은 관심이 모아지고 있다. 나아

가 기술적으로 코덱간의 호환성 문제를 근원적으로 해결하고 음성서비스와 음악서비스를 대역폭 확장을 통해 동시에 제공할 수 있는 초광대역 음성기반 오디오 융합코덱 표준화가 이루어지고 있다.

임베디드 가변비트율 코덱기술을 대표하는 광대역 코덱인 G.718, G.729.1 코덱과 호환성을 제공하여 기존의 통신 시스템에 대한 영향을 최소화 하면서 초광대역 음성 및 오디오 품질을 필요로 하는 응용분야에 사용하기 위해 지난 2007년부터 표준화가 이루어졌다. 현재 3GPP에서는 SA4 산하에 EVS (enhanced voice service) 그룹을 구성하여 AMR-WB (adaptive multirate-wideband), AMR-WB+ 코덱의 후속인 LTE (long term evolution) 서비스를 위한 EVS의 표준화를 위해 기술적 요구 사항을 도출하고 있으며 음성정보와 오디오 대역(20 - 20,000 Hz)의 정보를 모두 처리할 수 있는 융합코덱 형태의 기술개발과 표준화가 진행되고 있다.

G.718 SWB 코덱에서는 low band (LB) 영역(0 - 7 kHz)과 high band (HB) 영역(7 - 14 kHz)을 신호 영역에서 독립적으로 양자화를 수행한다. LB 영역에서는 기존의 G.718 양자화가 그대로 사용되며 HB 영역에서는 신호의 특성 (tonality

1) 한국과학기술원, 전기 및 전자공학과, kscho83@kaist.ac.kr

2) 경상대학교, 전자공학과(공학연구원), jeongsb@gnu.ac.kr, 교신저자

접수일자: 2012년 8월 3일

수정일자: 2012년 9월 20일

게재결정: 2012년 9월 20일

detection)에 따라서 양자화를 달리하고 있다. HB 영역에서는 기본적으로 신호의 MDCT 변환이 이루어지며 검출된 신호의 특성, 즉, 톤 성분 존재 여부에 따라서 generic mode coding 과 sinusoidal mode coding 기법이 선택적으로 사용된다. 본 연구에서는 HB 영역에서 인지적 관점에서 MDCT 계층의 인지적 캡스트럼 벡터 양자화를 통해 기존의 generic mode coding을 개선한 modified generic mode coding 방식을 제안하였다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 G.718 SWB의 개요에 대해 간략히 소개되어 있고, 3장에는 제안된 새로운 알고리즘을 적용시킬 수 있는 기존 G.718 SWB 코덱의 generic mode 부분에 대해 설명되어 있다. 4장에는 제안된 새로운 알고리즘이 상세히 설명되어 있고, 5장에서 새로운 알고리즘을 적용한 실험결과를 보이고, 6장에서 결론을 맺는다.

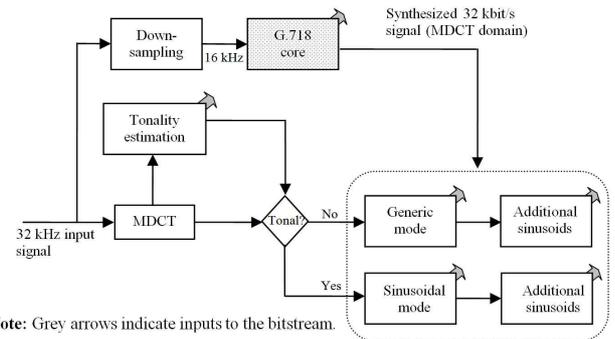
2. G.718 SWB 개요

G.718 SWB 코덱은 G.718 코덱 (광대역 코덱)과 호환성을 제공하여 기존의 통신 시스템에 대한 영향을 최소화 하면서 초광대역 음성 및 오디오 품질을 필요로 하는 응용분야에 사용하기 위해 표준화 되었다. 기존의 G.718에 SWB 대역폭 확장을 적용하여 협대역, 광대역, SWB 대역 신호를 점진적으로 제공하여 고품질의 응용서비스의 다양화를 가능하게 한다. 또한, 스테레오 기능을 통해 음악품질을 향상시킬 수 있다. SWB 부호화기는 tonality 검출 기법, 파라미터 세부구조 부호화 기법, gain-shape 기반의 MDCT 계수의 부호화를 사용하며, 특기사항으로 G.718 코덱이 잘 처리하지 못하는 톤과 같은 정보처리를 위해 추가적인 TDAC (time-domain alias cancellation) 정보를 송신하는 방법으로 광대역 향상 기능을 추가하였다. 이러한 음성기반 오디오 융합코덱 기술을 통해 와이브로, 무선랜, 4G이동통신 등에서 인터넷전화(VoIP) 서비스, 오디오 컨퍼런싱, 원격교육, 실시간 인터넷 게임, 대화형 IPTV와 같은 서비스에서 융합멀티미디어 서비스를 제공하는 데 널리 사용될 수 있다[3].

기존 G.718 코덱에서 지원되는 비트율은 8, 12, 16, 24, 32 kbps이다. ITU-T에서 표준화된 G.718 코덱은 8 - 32 kbps 범위 내에서 임베디드 형태로 비트율 계층을 쌓아가는 구조로 협대역 신호부터 광대역 신호를 압축, 복원한다. G.718 SWB 코덱은 G.718 코덱의 확장 코덱으로 기존 32 kbps에서 더해져, 36 kbps, 40 kbps, 48 kbps가 지원된다. AMR-WB로 널리 알려진 G.722.2 부호화기의 12.65 kbps 모드와 호환성을 제공하며, 구조는 비트율이 높아짐에 따라 비트를 쌓아 올리는 임베디드 가변 비트율 부호화 기술로 이루어져 있다[2]. <표 1>은 G.718 SWB 코덱으로 확장될 때 각 전송 계층 구조를 설명한다. 각 전송 계층마다 다른 기술이 사용되어 있는데, 특히 SWB로 확장된 첫 번째 전송 계층인 Layer 6mono에서는 generic mode 혹은 sinusoidal mode가 사용된다[4]. 이는 광대역에서 초광대역으로 확장될 때, 적은 비트율로도 최적의 음질을 얻기 위해 사용되는 기술이다.

표 1. G.718 SWB의 각 전송 계층 별 구조 및 기술
Table 1. Structure and technology for each transmission layer of G.718 SWB

전송 계층	비트율	포함 기술
G.718 codec	32 kbps	Multi-mode ACELP (algebraic code-excited linear prediction core), embedded enhancement layers
Layer 6mono	+4 kbps	Generic mode / Sinusoidal mode HB coding
Layer 7mono	+4 kbps	Additional HB sinusoids
Layer 8mono	+8 kbps	Additional HB sinusoids



Note: Grey arrows indicate inputs to the bitstream.

그림 1. G.718 SWB 코덱의 부호화기 블록다이어그램
Figure 1. Block diagram of G.718 SWB Encoder

<그림 1>은 G.718 SWB 코덱의 부호화기 블록다이어그램을 나타내고 있다[4]. <그림 1>에서 알 수 있듯이 G.718 코어 코덱에서는 표본화율이 32 kHz인 원 신호를 16 kHz로 다운샘플링하여 수행되고, SWB 확장에서는 원 신호가 입력으로 사용된다. SWB 인코더에서 먼저 입력 신호를 MDCT 영역으로 변환하고, 변환된 MDCT를 가지고 톤 성분을 측정하게 된다. 톤 성분을 측정하여 톤이 존재하면 sinusoidal mode로 들어가서 부호화가 수행되고, 톤이 존재하지 않으면 generic mode를 이용하여 부호화가 수행된다. 이는 전체 SWB 확장 코덱에서 첫 번째 전송 계층에서 수행이 이루어진다. 그 이후 전송 계층에서는 추가적인 sinusoidal mode를 통해 sinusoidal coding을 통해 원음에 가깝게 부호화와 양자화가 수행되어 비트율에 따라 음질이 향상된다.

3. G.718 SWB Generic Mode

G.718 SWB 코덱에서의 generic mode는 Layer 6mono에서 사용된다. SWB 인코더에서 먼저 입력 신호를 MDCT 영역으로 변환하고, 변환된 MDCT를 가지고 톤 성분을 측정하게 된다. 만약, 톤이 존재하면 sinusoidal mode를 통해 부호화가 수행되고, 톤이 존재하지 않으면 generic mode를 통해 부호화가 수행된다. 톤 성분에 대한 플레그 비트로 1 비트가 사용되며, 톤 성분이 존재하지 않을 경우, tonality = 0 으로, 톤이 존재할 경우, tonality = 1 로 나타낸다. 이는 전체 SWB extension 코덱에서 첫 번째 전송 계층에서 수행이 이루어진

다. Generic mode를 요약 설명하면, 기본적으로 LB 영역의 MDCT를 복사하여 HB 영역의 신호를 생성하는 것이다[4].

앞에서 설명하였듯이, generic mode는 토널 성분이 존재하지 않을 경우에만 사용된다. Generic mode에서는 G.718 광대역코덱 (core codec)에서 부호화된 MDCT (LB의 MDCT)를 HB에 이용한다. HB의 MDCT를 4 개의 밴드(40 MDCT, 70 MDCT, 70 MDCT, 100 MDCT)로 나누고, 양자화된 LB MDCT와의 correlation을 통해 가장 잘 정합되는 MDCT 시작 인덱스를 전송한다. 이를 위해서, 먼저, 양자화된 LB MDCT를 35개 밴드로 나눈 후 두 개의 스케일 팩터로 이득 조절을 한다. 합성되는 HB MDCT를 이용해 선형 영역의 이득과 로그 영역의 이득을 구해 최종적인 정규화를 한다. 또한, 추가적인 sinusoidal coding을 통해 양자화 오차를 개선을 한다. 복호화 기에서 LB MDCT로부터 전송된 인덱스를 이용하여 HB MDCT를 생성하고, 양자화하여 전송된 선형 영역의 이득과 로그 영역의 이득을 곱하여 HB MDCT 신호를 최종적으로 생성한다. 양자화를 위한 HB MDCT는 수식 (1)과 같이 나누어진다[4].

$$\begin{aligned} M_{32}^0 &= M_{32}(k+280), k=0, \dots, 39 \\ M_{32}^1 &= M_{32}(k+320), k=0, \dots, 69 \\ M_{32}^2 &= M_{32}(k+390), k=0, \dots, 69 \\ M_{32}^3 &= M_{32}(k+460), k=0, \dots, 99 \end{aligned} \quad (1)$$

여기서, $M_{32}^j(k)$ 는 각 서브밴드를 나타내고, 각 서브밴드의 길이는 40, 70, 70 그리고 100 개의 MDCT 영역 응답을 갖는다.

4. 제안된 Modified Generic Mode

Modified generic mode의 기본 개념은 기존의 generic mode에서 MDCT shape을 나타내는 밴드의 개수를 확장하고, 기존의 전체 이득 대신에 밴드마다 이득 특성을 반영할 수 있도록 스펙트럼 궤적 (spectrum envelope)을 양자화하여 필터뱅크 기반의 MDCT 이득을 보정하는 방법이다. G.718 SWB 코덱에서는 <표 1>에서 알 수 있듯이 32 kbps에서부터 HB 영역에 비트할당이 주어진다. 기존 베이스라인에서는 전체 이득을 사용하는 generic mode를 사용했으나 본 연구에서는 이러한 방식의 개선이 주요 목적이다. 본 연구에서는 SWB확장의 첫 번째 전송 계층에서 HB 쪽 양자화에 주어진 비트는 80 비트를 최적 활용하여 MDCT shape과 필터뱅크 기반의 MDCT 이득 전송에 적절히 사용할 수 있도록 조절하였다.

기존의 generic mode방식에서는 HB의 밴드 분할이 4 개 밖에 안 되므로 LB 영역에서 만족할만한 MDCT를 복사해오기가 어렵다. 우선 밴드의 길이가 길기 때문에 근사한 밴드를

찾는데 한계가 있다. 밴드의 길이가 짧아지고 밴드 수가 많아질수록 좀 더 잘 정합되는 MDCT 값을 찾아올 수가 있다. 따라서 제안한 modified generic mode에서는 기존의 밴드의 길이를 줄이고 밴드 수를 늘리는 데에 비트를 좀 더 할당하였다. 두 번째로는 전체 이득을 보내기 때문에 서브밴드의 에너지 변화를 나타내기 어렵다는 단점이 있다. 기존의 generic mode에서는 단지 4 개의 서브밴드 분할을 통해 각 밴드를 대표하는 전체 이득을 보내게 되는데 이 역시 밴드 길이가 길기 때문에 밴드 내에서의 에너지 변화, 즉 궤적을 나타내는 데에 한계가 있다. 이를 개선하기 위하여 제안한 방식에서는 필터뱅크 기반의 MDCT 이득 보정을 위한 캡스트럼 벡터 양자화를 제안하였다. G.718 SWB의 첫 번째 전송 계층에 주어지는 80비트를 활용하여 6개의 밴드분할 및 LB 영역 복사에 총 48 비트 (8, 8, 8, 8, 8, 8 비트), 캡스트럼 양자화에 총 26 비트 (캡스트럼 0 차(Energy 성분)에 8 비트, 1 - 5 차에 9 비트, 6 - 13 차에 9 비트)를 할당하였다. 나머지 6비트는 여분으로 남겨 두었다. 기존 generic mode와 modified generic mode의 MDCT 밴드 분할 폭을 비교하면, 기존 generic mode의 경우 4 밴드 (40, 70, 70, 100 MDCTs)인 반면에, 제안한 방식에선 6 밴드 (30, 30, 40, 40, 60, 80 MDCTs)이다.

제안된 modified generic mode 방식은 다음과 같다. 먼저 LB쪽의 MDCT 양자화 신호를 저장해 둔다. 기존 G.718과는 다르게 전체 이득을 전송하지는 않기 때문에 LB쪽의 MDCT 양자화 신호 값에서 가장 근사한 MDCT값의 첫 인덱스를 전송하게 된다. 6 개 밴드를 각 34, 40, 40, 40, 70, 80 의 간격을 갖는 서브밴드로 분할한다. HB의 각 서브밴드 값과 가장 상관도가 큰 LB 쪽의 MDCT를 찾게 된다. 수식 (2)는 가장 근사한 LB MDCT 인덱스를 찾는 상관도 식이다.

$$C(k) = \frac{\sum_{k=0}^{sbWidth(i)} (M_{low}(k) \times M_{high}(k))}{\sqrt{\sum_{k=0}^{sbWidth(i)} (M_{low}(k))^2 \times \sum_{k=0}^{sbWidth(i)} (M_{high}(k))^2}}, j=0, \dots, 280 - sbWidth(i)$$

$$\begin{aligned} M_{low}(k) &= m(j+k) \times Spec(sbStart(i) + k) \\ M_{high}(k) &= m(280 + sbStart(i) + k) \end{aligned} \quad (2)$$

수식 (2)에서 $Spec(\cdot)$ 은 복호화 단에서 합성을 고려하여 좀 더 합성될 때의 MDCT 값과 잘 정합시키기 위하여 이용되는 HB의 스펙트럼 궤적 값이다. $m(k)$ 는 MDCT의 k 번째 인덱스를 나타내며, $sbStart(i)$ 는 서브밴드의 시작 인덱스 값이며, $sbWidth(i)$ 는 서브밴드의 간격을 나타낸다. 복호화 단에서의 합성을 고려하여 LB MDCT에 HB의 스펙트럼 에너지 궤적을 씌워서 상관도를 계산하여 인덱스를 찾게 된다. 이는 전체 이

득 대신에 14 개의 서브밴드로 나눠 구한 스펙트럼 궤적의 캡스트럼을 양자화 하여 전송 후 다시 복원한 스펙트럼 궤적을 가지고 전송된 MDCT shape과의 곱으로 MDCT를 복호화 하기 때문이다. 복호화 단에서 전송된 MDCT 인덱스를 통해 LB의 MDCT을 복사하여 HB의 각 서브밴드에 대한 MDCT 값을 생성하게 된다.

필터뱅크 기반의 MDCT 이득 보정 방식은 기존의 간소화된 이득 보정을 개선한 것이다. 기존 방식은 4 밴드의 이득을 벡터 양자화하여 전송한다. 본 제안된 방식에서는 7 - 14 kHz 구간을 500 Hz 단위(20 MDCT bins)로 분할하여 14 개의 필터뱅크를 생성하게 된다. 그 후, 각 주파수 밴드의 평균 에너지를 계산하고 이를 이용하여 캡스트럼을 추출한다. <그림 2>는 MDCT를 이용하여 필터뱅크를 분할하는 방식을 나타낸다. 여기서 $m_i(k)$ 는 i 번째 필터뱅크의 k 번째 MDCT 계수를 나타낸다. 수식 (3)은 로그 영역에서의 필터뱅크를 구하는 식이다. 수식 (3)을 통해 필터뱅크 에너지 값을 구하며, 이는 스펙트럼 궤적 값이 된다. 수식 (4)는 필터뱅크 에너지를 가지고 캡스트럼을 추출하는 식이다. 수식 (4)를 통해 캡스트럼 0 - 13 차 값을 구하게 된다[6].

$$LFB(i) = \log\left(\frac{1}{20} \sum_{k=1}^{20} |m_i(k)|\right), \quad i = 0, 1, \dots, 13 \quad (3)$$

$$c(n) = \frac{1}{14} \left(LFB(0) + 2 \sum_{i=0}^{13} LFB(i) \cos\left(\frac{2\pi n i}{13}\right) \right), \quad n = 0, 1, \dots, 13 \quad (4)$$

이중에서 0 차 값은 7 - 14 kHz의 로그에너지 정보를 담고 있기 때문에 중요한 정보로서 0 차 값만 따로 8비트 양자화를 수행하게 된다[8]. 이 때, 훈련 DB를 통해 얻은 벡터값들을 modified K-means (MKM) clustering을 사용하여 양자화 하여 256 코드값을 생성한다[7]. 1 차에서 13 차는 스펙트럼의 세부 구조(스펙트럼 라인)를 구성한다[8]. 1 차에서 5 차까지, 6 차에서 13 차까지 각각 두 개의 벡터집합으로 두고 MKM clustering 양자화를 수행한다. 각 두 개의 벡터 집합을 각 9 비트씩 양자화 하여 각 512 개의 코드값을 생성한다. 복호화 단에서는 캡스트럼 계수로부터 필터뱅크 이득 (spectrum envelope)을 직접 계산하게 된다. 수식 (5)는 캡스트럼 계수로부터 필터뱅크 이득을 구하는 식이다.

$$FB_q(i) = 2 \sum_{n=1}^{13} C_q(n) \cos\left(\frac{2\pi n i}{14}\right) \quad i = 0, 1, \dots, 13 \quad (5)$$

여기서, $C_q(n)$ 은 양자화 된 캡스트럼 코드 벡터를 나타내고, $FB_q(i)$ 는 $C_q(n)$ 으로부터 구한 양자화된 필터뱅크 에너지를 나타낸다. 복호화 단에서 필터뱅크 에너지를 역양자화하는 방법

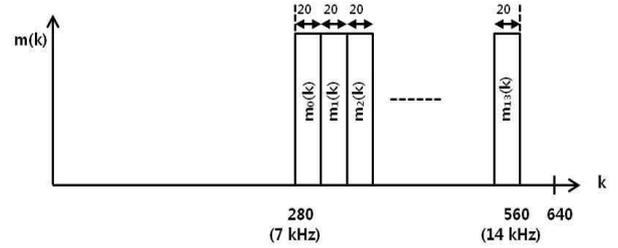


그림 2. 필터뱅크 분할 방식
Figure 2. Filterbank partitioning method

은 수식 (6)과 같다.

$$\hat{m}_i(k) = \frac{\sqrt{FB_q(i)}}{\sqrt{G_g(i)}} m_{g,i}(k) \quad (6)$$

수식 (6)에서 i 와 k 는 각각 필터뱅크 인덱스와 MDCT 계수 인덱스를 나타내고, $m_{g,i}(k)$ 는 Generic mode에 의해 복사 된 MDCT 계수를 나타내며, $FB_{q,i}$ 는 전송된 캡스트럼 VQ 를 통해 구한 필터뱅크 에너지를 나타낸다. 그리고 $G_g(i)$ 는 $m_{g,i}(k)$ 의 에너지를 나타내는데 이는 수식 (7)을 통해 구할 수가 있다.

$$G_g(i) = \frac{1}{20} \sum_{k=0}^{19} |m_{g,i}(k)| \quad (7)$$

수식 (7)을 통해 modified generic mode에 의해 생성된 MDCT와 캡스트럼 양자화를 통해 구한 필터뱅크 에너지를 생성된 MDCT의 밴드 에너지로 정규화한 값으로 HB의 MDCT를 생성할 수가 있다.

5. 실험 및 결과

캡스트럼 양자화 훈련을 위한 DB의 사양을 <표 3>에 나타내었다. 음악은 143 개 파일이 사용되었으며 음성은 169개의 파일이 이용되었고, 음악의 경우 각각 1 분 이상의 지속시간을 갖는다. 음성은 대체적으로 1 분보다 짧은 약 50 초 음성 10 개, 약 6 초 음성 159 개가 사용되었다. 음악의 경우는 다양한 장르를 포함하고 있으며, 음성의 경우 30명의 화자가 발성한 DB를 사용하였다. 음성과 음악 DB를 통해 총 343,072 개의 벡터를 추출하였으며, MKM 기반 양자화를 수행하였다.

<그림 4>에 코드북 크기별 intra-distnace 측정 결과를 나타내었다. <그림 4>의 결과 및 시행착오 과정을 거쳐서 HB MDCT 에너지 궤적의 캡스트럼 벡터 양자화에 9 비트, 전체 이득 (캡스트럼 0차)에 8 비트를 할당하였다. 캡스트럼 전체를 한 개의 벡터로 묶어 512 (9 비트) 양자화 한 경우 어느 정도 스펙트럼 궤적을 따라가지만 양자화 에러가 많이 존재하므로 실제로 적용할 때는 캡스트럼을 부분적으로 분할해서 여러 개

표 3. 켈스트럼 양자화를 위한 훈련 DB
Table 3. Training DB for cepstrum quantization

음 악 (143 파일)	POP (Korean POP 포함) 8 세트
	Classic 4 세트
음 성 (169 파일)	약 50 초 음성 10 파일
	약 6 초 음성 159 파일
총 343,072 벡터	

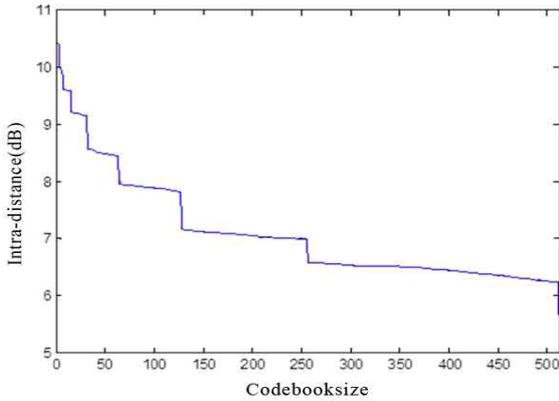


그림 4. 코드북 크기 별 intra-distance 측정 결과
Figure 4. Intra-distance result for each codebook size

의 벡터 집합으로 묶어 각각 양자화를 수행하였다. 이득에 대한 켈스트럼 양자화의 경우 1 개의 에너지 벡터 (켈스트럼 0 차)를 8 비트 양자화 하였고, 나머지 차수들은 총 2 개의 벡터 집합으로 묶어 각각 9 비트 양자화를 수행하였다

음질 평가를 위해 <표 4>와 같이 실험 환경을 구성하였다. G.718 SWB의 전송 비트율은 generic mode를 지원하는 36 kbps 로 설정하였다. 실험 결과를 측정하기 위하여 객관적 평가 방법인 Bark's scale distortion (BSD)를 사용하였다[5]. BSD를 계산하기 위해 512-FFT (fast Fourier transform) 이 30 ms 분석 프레임으로부터 매 10 ms간격으로 수행되었으며, Bark's scale 필터 뱅크를 통해 계산된 33 에너지 값이 예측 값으로 구해졌다. 수식 (8)에 BSD를 구하기 위한 수식을 나타내었다.

$$BSD = \frac{\sum_{t=0}^{T-1} (BSD^{(t)})}{\sum_{t=0}^{T-1} \left(\sum_{k=0}^{N-1} (L_x^{(t)}(k))^2 \right)} \quad (8)$$

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

$BSD^{(t)} = \sum_{k=0}^{N-1} (L_x^{(t)} - L_y^{(t)}(k))^2$ 로 계산되며,

$L_x^{(t)}(k), L_y^{(t)}(k)$ 는 각각 원신호의 필터 뱅크 에너지와 테스트 신호의 필터뱅크 에너지를 나타낸다.

<그림 5>와 <그림 6>에서 음성 및 음악에 대한 성능 비교

표 4. 음질 평가를 위한 실험 환경

Table 4. Experimental environments for sound quality evaluation

테스트 DB	32 kHz 표본화율, 모노, 16 비트
	약 6 초 음성 14 파일
	약 20 초 음악 7 파일
	36 kbps의 비트율 환경
성능 평가 방법	BSD 측정

를 나타내었다. 음성 파일에 대한 실험 결과 G.718 SWB 36 kbps 보다 제안한 방법을 사용하였을 경우 성능이 나은 것을 확인할 수가 있으며, 음악 파일에 대해 실험한 것 역시 제안한 방법을 사용하였을 경우 성능이 나은 것을 확인할 수가 있다. <표 4>는 테스트 데이터의 평균 BSD 결과를 나타낸 것이다. 음성 데이터 평균 BSD 결과를 보면 기존 G.718 SWB보다 제안한 modified generic mode를 사용할 경우 약 2.5 dB 이상 개선되는 것을 확인할 수가 있다. 또한 음악 데이터 평균 BSD 결과를 보면 기존방식에 비해 제안된 방식을 사용하였을 경우, 약 1.7 dB 이상 개선되는 것을 확인할 수가 있다. 이는 기존의 generic mode에서 획일적인 전체 이득보다는 밴드 별 에너지 변화를 나타내는 스펙트럼 궤적을 근사화하여 전송 및 복원하는 것이 음질을 높는데 더 효과적임을 나타내고 있으며, HB의 MDCT 서브밴드를 확장함으로써 음질을 높일 수 있음을 보인다. 또한, 기존 generic mode의 경우 할당된 80 비트를 모두 사용하는 반면, 제안된 방법에서는 추후 개선부분을 위해 6 비트의 여분의 비트를 남겨 놓았기 때문에 남은 비트를 새로운 음질 개선 방법에 사용하면 더 나은 결과를 얻을 수 있을 것으로 기대된다.

6. 결론

본 논문에서는 G.718 SWB에서 HB의 신호 정보를 위해 임베디드 가변 비트 구조에서 첫 번째 초광대역 layer에서 사용되는 기술인 generic mode를 개선하는 알고리즘을 제안하였다. 제안된 알고리즘에서는 기존 generic mode에서 사용되는 넓은 서브 밴드 길이를 줄였으며 밴드 수를 늘림으로써 LB에서 좀 더 잘 정합되는 부분을 선택하도록 하였다. 또한, 기존의 전체 이득을 좀 더 세분화 하여 나타낼 수 있도록 서브밴드의 스펙트럼 궤적에 효율적인 비트 할당을 위해 켈스트럼으로 변환하여 양자화를 수행하여 할당 비트를 줄이면서도 음질을 높이는 방법을 사용하였다. 제안된 방식을 적용하여 실험한 결과, 기존의 generic mode 보다 음질이 향상됨을 알 수 있었다. 제안된 방법 중에 이득을 필터뱅크 기반의 MDCT 이득을 보정하는 방법, 즉, 켈스트럼 벡터 양자화를 통해 스펙트럼 궤적을 양자화 하여 전송하는 방법은 generic mode coding 뿐만 아니라 다른 양자화 방식인 algebraic vector quantization(AVQ)나 factorial pulse coding (FPC)에서도 전체 이득을 대체하여 더 나은 결과를 얻을 수 있을 것이라고 판단

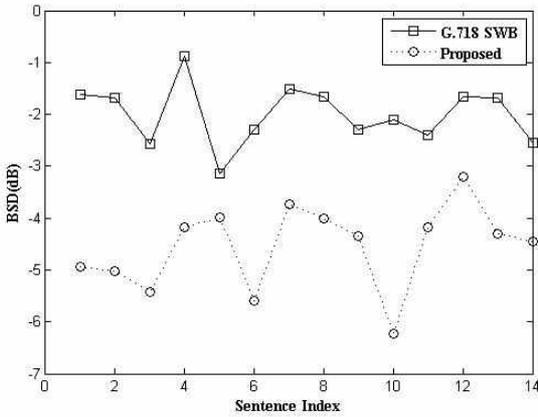


그림 5. 제안된 알고리즘에 의한 spectral distortion 결과 (음성 파일)

Figure 5. Spectral distortion result by proposed algorithm (Speech files)

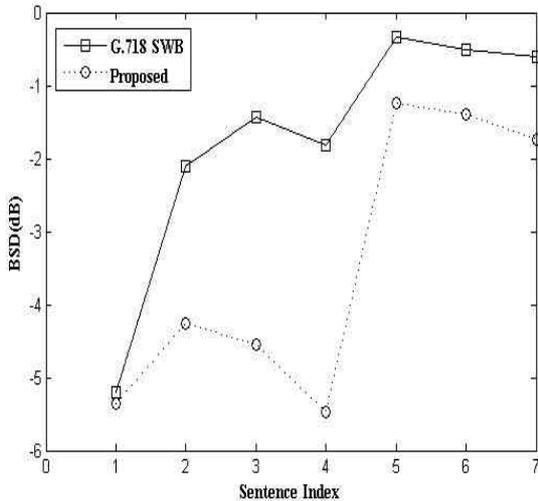


그림 6. 제안된 알고리즘에 의한 spectral distortion 결과(음악 파일)

Figure 6. Spectral distortion result by proposed algorithm (Music files)

표 4. 테스트 데이터의 평균 BSD 측정 결과 (dB)
Table 4. Average BSDs for test DB (dB)

	G.718 SWB	Proposed
음성	-2.004	-4.543
음악	-1.716	-3.431

된다. G.718 SWB 측면에서 현재 제안한 방식은 할당된 비트 측면에서 기존 방식에 비해 더 적은 비트를 사용하기 때문에 여분의 비트를 이용하여 좀 더 나은 성능을 얻을 수 있을 것으로 기대된다. 향후, 제안된 알고리즘을 G.718 SWB 뿐만 아니라 다른 초광대역 코덱에도 적용할 예정이며 AVQ나 FPC 등의 스펙트럼 궤적 정보를 양자화 하는 방식을 대체하여 음질을 개선할 예정이다. 추가적으로 G.718 SWB에서는 제안된

modified generic mode를 사용할 때의 잉여 비트를 파라미터 양자화 오차 보정 방법에 적용하여 음질을 개선하는 연구가 필요하다.

감사의 글

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업지원금을 받아 수행된 것입니다 (2012R1A1A1014222).

참고문헌

[1] Cho, K., Sung J., Hahn, M., Kim, Y. & Jeong, S. (2009). Enhanced Spectral Envelope Coding Scheme Using Inter-Frame Correlation for G.729.1, *Journal of Korean Society of Speech Sciences*, Vol. 1, No. 4. 97-103.
(조근석, 성종모, 한민수, 김영일, 정상배 (2009). G.729.1 코더에서 프레임 간의 상호상관 관계를 이용한 개선된 스펙트럼 포락 코딩 방법, 말소리와 음성과학, 제1권 제4호, 97-103.)

[2] Kim, D., Sung, J., Lee, M., Bae, H. & Lee, B. (2009). Trends of Speech-Based Audio Convergence Codec Technology. *Electronics and Telecommunications Trends*. Vol. 24, No 5, 10-19.
(김도영, 성종모, 이미숙, 배현주, 이병선 (2009). 음성기반 오디오 융합코덱 기술동향, 전자통신동향분석. 제24권 제5호, 10-19.)

[3] Lee, M., Kim, D. & Lee, B. (2010). Trends of Codec Technology for 4G Mobile Enhanced Voice Service. *Electronics and Telecommunications Trends*. Vol. 25, No 6. 29-37.
(이미숙, 김도영, 이병선 (2009). 4G 모바일 증감음성 서비스를 위한 코덱 기술 동향, 전자통신동향분석. 제25권 제6호, 29-37.)

[4] ITU-T Rec. (2010). Superwideband scalable extension for G.718, New G. 718 Annex B.

[5] Wang, S., Sekey, A. & Gersho, A. (1992). An objective measure for predicting subjective quality of speech coders, *IEEE Trans. Selected Areas in Comm.* Vol. 10. No. 5, 819-829.

[6] Linde, Y., Buzo, A. & Gray, R. (1980). An Algorithm for Vector Quantizer Design. *IEEE Transaction on Communications*, 28 (1), 84-94.

[7] Wim, D., & Xavier, R. (2003). Discrete Cepstrum Coefficients as Perceptual Features, *International Computer Music Conference (ICMC)*, 1-4.

- [8] Oh, Y. (1998). *Speech language information processing*.
Seoul: Hongleung Science Publisher.
(오영환 (1998). 음성언어정보처리, 홍릉과학출판사.)

• **조근석 (Cho, Keunseok)**

한국과학기술원 전기 및 전자공학과
대전광역시 유성구 구성동 373-1
Tel: 042-350-5474
Email: kscho83@kaist.ac.kr
관심분야: 음성/오디오 코딩, 음성 합성
현재: 한국과학기술원 전기 및 전자공학과 박사과정

• **정상배 (Jeong, Sangbae)** 교신저자

경상대학교 공과대학
경남 진주시 가좌동 900번지
Tel: 055-751-5357 Fax: 055-759-7814
Email: jcongsb@gnu.ac.kr
관심분야: 음성신호처리
현재: 경상대학교 전자공학과(공학연구원) 조교수