

16kbps 광대역 음성 압축기 개발

16kbps Wideband Speech Codec

송재종*, 박호종*
(Jai-Chong Song*, Hochong Park*)

* 광운대학교 전자공학과

(접수일자: 2001년 8월 16일; 채택일자: 2001년 10월 30일)

본 논문에서는 7 kHz 대역폭을 가지는 새로운 16 kbps 광대역 음성 압축기를 제안한다. 제안하는 압축기는 QMF (Quadrature Mirror Filter)를 사용하여 입력 신호를 하위 밴드와 상위 밴드로 분리하고, 하위 밴드는 AMR (Adaptive Multi Rate) 음성 압축기를 이용하여 압축하고 상위밴드는 G.722.1 광대역 음성 압축기를 기반으로 새로운 압축기를 개발하여 압축한다. 특히 입력신호의 성질에 따라 상위 밴드와 하위 밴드에 할당되는 비트 수를 다르게 함으로써 고정적으로 비트를 할당하는 음성 압축기보다 향상된 압축 성능을 제공하고, 웨이블릿 변환을 이용하여 상위 밴드 신호의 압축 성능을 향상시켰다. 주관적 방법으로 제안된 음성 압축기의 성능을 측정하였으며, 다양한 입력신호에 대하여 제안한 16 kbps 음성 압축기가 G.722 48 kbps SB-ADPCM보다 우수한 성능을 가지는 것을 확인하였다.

핵심용어: 음성 압축기, 광대역 신호, 웨이블릿 변환, 가변적 비트 할당

투고분야: 음성처리 분야 (2,2)

This paper proposes new 16 kbps wideband speech codec with bandwidth of 7 kHz. The proposed codec decomposes the input speech signal into low-band and high-band signals using QMF (Quadrature Mirror Filter), then AMR (Adaptive Multi Rate) speech codec processes the low-band signal and new transform-domain codec based on G.722.1 wideband codec compresses the high-band signal. The proposed codec allocates different number of bits to each band in an adaptive way according to the property of input signal, which provides better performance than the codec with the fixed bit allocation scheme. In addition, the proposed codec processes high-band signal using wavelet transform for better performance. The performance of proposed codec is measured in a subjective method, and the simulations with various speech data show that the proposed codec has better performance than G.722 48 kbps SB-ADPCM.

Keywords: Speech codec, Wideband signal, Wavelet transform, Adaptive bit allocation

ASK subject classification: Speech signal processing (2,2)

I. 개요

디지털 통신에서 음성 압축기는 아날로그 음성 신호를 적은 양의 디지털 데이터로 변환하는 역할을 하며, 통신 시스템의 용량과 품질을 결정하는 매우 중요한 요소이다.

책임저자: 박호종 (hcpark@mail.gwu.ac.kr)
139-701 서울시 노원구 월계동 447-1
광운대학교 전자공학과
(전화: 02-940-5104; 팩스: 02-912-7305)

그 동안 음성 압축기에 대한 많은 연구를 통하여 현재 여러 종류의 음성 압축기가 표준으로 지정되어 다양한 응용 분야에 사용되고 있고, 대표적으로 디지털 이동 통신에서 IS-96 QCELP8k[1], IS-733 QCELP13k[2], IS-127 EVRC[3]가 표준으로 정해져 사용 중에 있고, 차세대 이동 통신인 비동기 IMT-2000에서는 AMR 음성 압축기가 표준으로 지정되었다[4]. 이상의 음성 압축기들은 음성 신호를 8 kHz로 표본화하여 4 kHz 이하의 대역폭을 가

지는 협대역 음성 압축기이다.

그러나, 최근에 디지털 이동통신, 인터넷 통신 (VoIP), 원격회의 등의 새로운 통신 서비스가 개발되고 있고, 이들은 서비스의 경쟁력을 높이기 위하여 기존의 음성 통신에 비하여 향상된 음성 품질을 요구하고 있다. 이와 같은 고품질 음성 통신을 위하여 음성 신호를 16 kHz 표본화하여 전달하는 것이 필요하며, 이러한 요구에 따라 ITU와 3GPP 등의 표준화 기구에서 광대역 (Wideband) 음성 압축기의 표준화 작업이 진행 중이다. ITU는 광대역 음성 압축기를 위하여 G.722를 제정하였고[5], 최근에 원격회의를 위하여 24/32 kbps 전송률을 가지는 광대역 음성 압축기로 G.722.1을 지정하였으며[6], 3GPP에서는 제3세대 이동통신을 위하여 Wideband AMR을 표준으로 정하였다[7]. 이와 같이 다양한 응용 분야에서 고품질의 음성 서비스를 제공하기 위하여 광대역 음성 압축기의 개발이 요구됨에 따라 본 논문에서는 저전송률을 가지며 고품질의 음성 통신을 할 수 있는 새로운 광대역 음성 압축기를 개발한다.

대표적인 광대역 음성 압축기의 구조에는 CELP (Code Excited Linear Prediction) 방식[8], 주파수 영역에서 압축하는 TC (Transform Coding) 방식[9], 그리고 신호를 두 밴드로 나누어 각각 압축하는 Hybrid 방식이 있다. CELP 방식은 현재 사용되고 있는 13 kbps 이하의 대부분의 협대역 음성 압축기에서 사용되는 방식으로서 광대역 음성 압축기 중에는 6.60 kbps ~ 23.85 kbps 사이의 전송률을 지원하는 Wideband AMR이 CELP 방식을 사용한다 [7]. TC 방식은 음성 신호를 주파수 영역으로 변환하여 주파수 영역의 계수를 양자화하는 방법을 사용하며, 간단한 형태로 모델링이 불가능한 오디오 신호를 압축하는 MPEG 계열에서 많이 볼 수 있다. TC 방식을 이용한 광대역 음성 압축기에는 24 kbps와 32 kbps 전송률을 지원하는 G.722.1이 있다. 마지막으로 Hybrid 방식은 QMF를 이용하여 입력신호를 상위밴드와 하위밴드로 나누고 일반적으로 하위밴드는 CELP 방식을 이용하고 상위밴드는 TC 방식을 이용하여 압축한다.

본 논문에서는 Hybrid 방식의 새로운 광대역 음성 압축기를 제안하며, 16 kHz로 표본화된 신호를 QMF를 이용하여 상위밴드와 하위밴드로 나누고 하위밴드는 AMR의 고전송률 4개 모드를 이용하여 압축하고, 상위 밴드는 G.722.1을 기반으로 새로운 형태의 TC 압축기를 개발하여 압축한다. 특히 각 밴드의 비트 할당, 상위 밴드 압축을 위한 새로운 압축기 개발에 대하여 집중적으로 연구하며, 주관적 성능을 G.722 48 kbps SB-ADPCM과 비교하

여 측정한다. 2장에서는 제안하는 음성 압축기의 구조와 동작에 대하여 상세하게 설명하고, 3장에서는 압축기의 성능을 측정하고 4장에서 결론을 맺는다.

II. 제안하는 음성 압축기의 구조와 동작

2.1. 전체 구조

본 논문에서 제안하는 광대역 음성 압축기의 전체적인 구조는 그림 1과 같이 크게 4 블록으로 구성된다: (i) 입력 신호를 상위 밴드와 하위 밴드로 분리하는 QMF 과정, (ii) 입력 신호의 특성에 따라 상위와 하위 밴드에 비트를 할당하는 과정, (iii) 상위 밴드와 하위 밴드를 각각 압축하는 과정, (iv) 상위 밴드와 하위 밴드의 파라미터들을 직렬화하는 과정.

입력신호의 형태는 16-비트 선형 PCM이고, 프레임의 길이 및 Look-Ahead는 각각 20 ms이다. 먼저 입력 신호를 4 kHz를 기준으로 하위 밴드와 상위 밴드로 분해하며, 이는 G.722에서 사용하는 기본적인 QMF를 사용하여 구현한다[5]. 두 밴드로 분해된 신호는 이 이후에 각각 완전 독립적으로 압축된다. 다음, 4 kHz 이하의 하위 밴드는 AMR의 12.2 kbps ~ 7.40 kbps 전송률을 가지는 4 가지 모드를 사용하여 압축하며, 각 프레임에서 실제 사용하는 모드는 2.2절에서 설명하는 방법으로 입력 신호의 특성을 고려하여 정한다. 상위 밴드의 압축은 G.722.1을 기반 새로운 TC 구조 압축기를 개발하여 사용하며 2.3절 및 2.4절에서 상세히 설명한다. 파라미터를 비트열로 만드는 과정은 음성 압축기의 동작과 직접적인 관련이 없고 전송 및 표준화와 관련되는 내용이므로 본 논문에서는 다루지 않는다.

이상과 같이 제안한 압축기의 하위 밴드는 기존의 AMR을 사용하여 압축하고 단지 전송률을 정하는 방법만 새롭게 개발되므로 자세한 압축 동작은 AMR 표준안을 참조하고 본 논문에서는 다루지 않는다[4]. 반면, 본 논문

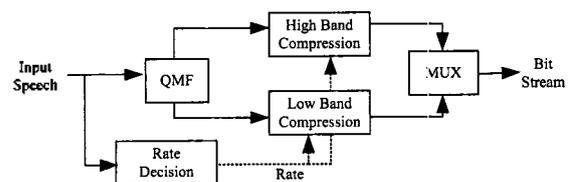


그림 1. 제안한 광대역 음성 압축기의 구조
Fig. 1. Structure of proposed wideband speech codec.

에서는 하위 밴드와 상위 밴드에 적절한 비트를 할당하여 AMR의 모드를 정하는 방법과 상위 밴드를 압축하는 방법에 대한 연구를 중점적으로 진행하며, 이 부분에 대한 새로운 방법의 개발을 통하여 음성 압축기의 성능을 향상시키도록 한다.

2.2. 상위밴드와 하위밴드의 비트 결정

기존의 많은 광대역 음성 압축기에서는 입력 신호를 4 kHz 기준으로 하위 밴드와 상위 밴드로 분해하여 상위 밴드와 하위밴드에 고정된 수의 비트를 할당한다. 대표적인 예로 G.722 48 kbps 압축기에서는 하위 밴드에 32 kbps, 상위 밴드에 16 kbps를 할당하여 각 밴드를 독립적으로 압축한다[5]. 그러나, 본 논문에서는 다양한 특성의 입력 신호에 대하여 우수한 성능을 가지도록 입력 신호의 특성에 따라 상위밴드와 하위밴드의 비트 할당을 가변적으로 한다. 즉, 한 프레임에 할당된 총 320 비트를 상위 밴드 및 하위 밴드에 할당하는 방법을 표 1과 같이 4가지 정의하고, 입력신호의 특성을 분석하여 각 프레임에 대한 적정 프레임 모드를 결정한다.

프레임 모드를 결정하는 과정은 입력신호에 해밍 윈도우를 씌운 다음, 256-Point DFT를 이용하여 파워 스펙트럼을 구하고 16개의 500 Hz 단위 밴드로 나누어 각 밴드에 대한 RMS 값을 구하여 식 (1)과 같이 상위 밴드에 대한 하위 밴드의 RMS 비율을 각 500 Hz 밴드 단위로 구한다.

$$Rate_RMS[k] = \frac{RMS[k]}{RMS[k] + RMS[8+k]}, \quad 0 \leq k \leq 8 \quad (1)$$

다음, 8개의 Rate_RMS[k]에 대한 평균을 구하는데, 이때 앞 4개 (2 kHz 이하)의 값이 더 중요하기 때문에 뒤

표 1. 하위밴드와 상위밴드에 비트를 할당하는 4가지 프레임 모드
Table 1. Four modes of bit allocation for low band and high band.

Frame Mode	3	2	1	0
Low Band (Bits/Frame)	244	204	159	148
High Band (Bits/Frame)	76	116	161	172

4개 (2 kHz 이상)보다 두 배 만큼 가중치를 주어 평균을 구한다. 마지막으로 가중치가 적용된 RMS 비율 평균값이 0.5 보다 작으면 프레임 모드는 0, 0.55보다 작으면 1, 0.6보다 작으면 2, 0.6 이상이면 3으로 프레임 모드를 결정한다. 또한, 급격한 프레임 모드의 변화를 막기 위한 프레임 전의 프레임 모드와 현재 프레임 모드가 2 이상 차이가 있을 때는 전 프레임 모드보다 한 단계 높거나 낮은 프레임 모드를 최종적으로 결정한다.

제안한 비트 할당 방법의 성능을 검증하기 위하여 전체 음성 압축기에서 상위/하위 밴드 비트 할당 블록만 기존의 고정 방법과 제안한 가변 방법으로 변경하여 전체 음성 압축기의 성능을 각각 측정하였으며, 그 결과 제안한 가변적 방법으로 비트 할당할 때의 복원 음성 신호의 Segmental SNR이 고정 방법을 사용할 때보다 평균 0.27 dB정도 향상되는 것을 확인하였다.

그림 2는 특정 음성 신호에 대하여 프레임 모드를 결정하는 예를 보여준다. 신호의 특성에 따라 적절하게 최종 프레임 모드가 결정되는 것을 볼 수 있다.

2.3. 상위 밴드 압축

상위밴드는 G.722.1의 기본 구조를 기반으로 새로운 형태의 TC 압축기를 개발하여 사용한다. 특히 320 샘플 단

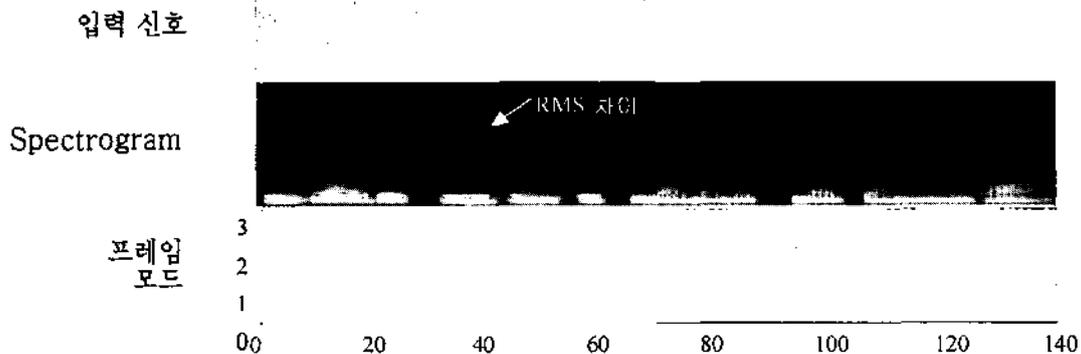


그림 2. 입력 신호의 특성에 따른 하위/상위 밴드의 가변적 비트 할당 예
Fig. 2. Example of adaptive bit allocation for low/high band according to the input property.

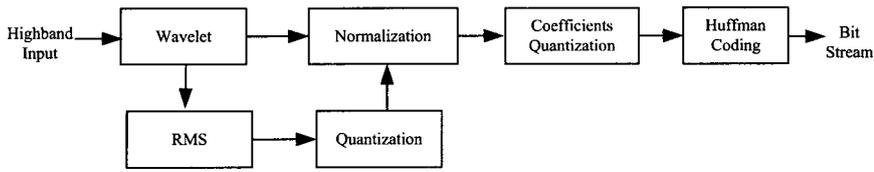


그림 3. 제안한 압축기의 상위 밴드 압축기 구조

Fig. 3. Structure of high-band encoder in the proposed wideband codec.

위로 압축하던 G.722.1 방식을 160 샘플 단위로 압축하는 방식으로 변환하였고, 입력신호의 변환은 MLT (Modified Lapped Transform) 대신 웨이블릿 변환을 이용하였고, RMS 값의 양자화 방법을 새로 개발하였다. 상위 밴드 압축기의 전체 구조는 그림 3에 있다.

먼저, QMF에 의하여 구하여진 상위 밴드 신호에 웨이블릿 변환을 적용하여 Low와 High로 나누고, 레벨1에서 Low를 다시 Low/High로, High를 Low/High로 나누고, 이러한 과정을 레벨3까지 수행하여 8개의 밴드로 분리하는 3-Level 웨이블릿 패킷을 구한다. 단, 웨이블릿 패킷을 구하는 과정에서 High 밴드에서의 시간축 Decimation 영향으로 아래 레벨의 Low와 High 밴드의 배열 순서가 바뀌게 되며, 이를 해결하여 정상적인 주파수 배열이 되도록 웨이블릿 패킷의 순서를 재조정하는 추가 과정이 필요하다. 본 논문에서는 웨이블릿 기저함수로 12-Tap Daubechies 직교 필터를 사용하였으며, 기저 함수의 Tap 수가 많아지면 Aliasing 현상이 줄어드는 장점은 있으나, Look-Ahead가 증가하는 단점이 있다.

G.722.1에서는 주파수 변환을 위하여 MLT를 사용하는 데, 일반적으로 MLT는 정적 신호에 대하여 우수한 분석 성능을 가지는 반면, 웨이블릿 변환은 천이 신호에 대하여 더 좋은 성능을 가진다[10]. 제안된 압축기에서 사용되는 변환 방식은 상위 밴드 신호를 압축하기 위한 것이고, 이 신호는 주로 잡음성 신호이며 하모닉 구조를 가지지 않는 경우가 많으므로 MLT 보다 웨이블릿 변환이 더 효과적인 변환이 되며, 따라서 본 논문에서 개발하는 압축기는 웨이블릿 변환을 사용하여 상위 밴드를 압축하는 방법을 사용한다.

웨이블릿 변환을 통하여 3-레벨 웨이블릿 패킷을 완성한 후 각 밴드의 RMS 값을 구하고 양자화하며, 양자화된 RMS 값으로 각 웨이블릿 밴드의 계수를 정규화하고 최종적으로 양자화한다. 웨이블릿 계수의 양자화 과정은 G.722.1 방법을 사용한다. 즉, Categorization 방법을 사용하여 각 500 Hz 밴드에 대한 비트 분포를 정하고 전달하며, 각 밴드의 웨이블릿 계수를 정해진 스텝 크기로

스칼라 양자화하고 양자화 인덱스를 벡터 Huffman 코딩하여 최종 비트열을 만든다. 자세한 웨이블릿 계수 양자화 방법은 [6]을 참조한다.

2.4. 상위 밴드의 RMS 양자화 방법

상위 밴드에서는 RMS 양자화와 웨이블릿 계수 양자화가 수행되며, 각각의 양자화를 위한 비트 수가 미리 결정되어 있는 것이 아니라 정해진 방법으로 RMS 양자화를 먼저 실행하고 그 다음 상위 밴드에 할당된 총 비트 중에서 남은 비트를 이용하여 웨이블릿 계수를 양자화한다. G.722.1에서는 모든 밴드의 RMS 값을 로그 영역에서 스칼라 양자화하며, 그 결과 RMS 양자화를 위하여 매우 많은 비트를 사용한다. 그러나 제안한 압축기에서는 상위 밴드에 상대적으로 적은 수의 비트가 할당되므로 G.722.1과 동일한 방법으로 상위 밴드의 RMS 값을 양자화하면 웨이블릿 계수의 양자화를 위한 비트 수가 너무 적게 되어 웨이블릿 계수의 양자화 오차 증가에 따른 성능의 저하가 매우 심해진다. 따라서, 제안하는 압축기에서는 새로운 구조에 적합하도록 RMS 양자화를 위한 새로운 방법을 개발한다.

개발된 상위 밴드 RMS 양자화 과정의 구조는 그림 4와 같다. 제안한 압축기의 밴드 폭이 7 kHz이므로 상위 밴드는 6개의 500 Hz 밴드로 구성되며, 따라서 6 개의 RMS 값에 대한 양자화가 필요하다. 각 밴드의 RMS 중에서 최대값을 구하여 6-비트 예측 스칼라 양자화하고, 양자화된 최대 RMS 값으로 각 밴드의 RMS 값을 정규화하여 6차 5-비트 벡터 양자화한다. 예측기로 4차 MA 구조 필터를 사용하고 스칼라 양자화는 식 (2)와 (3)의 방법을 사용한다.

$$q[n] = 1.13^n, \quad -2 \leq n \leq 29 \quad (2)$$

$$q_bound[n] = 1.13^{n+0.5}, \quad -2 \leq n \leq 29 \quad (3)$$

여기서, $q[n]$ 은 양자화값, $q_bound[n]$ 은 양자화기의 경계값을 나타낸다.

IV. 결론

본 논문에서 7 kHz 대역폭을 가지는 광대역 신호를 16 kbps로 압축하는 광대역 음성 압축기를 제안하였다. 입력신호를 하위와 상의 밴드 신호로 분리하여 각 밴드를 독립적으로 압축하는 기존 구조를 가지며, 하위 밴드는 AMR의 4가지 모드로 압축하고 상위 밴드는 G.722.1에 기반을 둔 새로운 압축기를 개발하여 사용하였다. 입력 신호의 특성에 따라 상위밴드와 하위밴드에 비트 수를 가변적으로 할당하고, 상위 밴드의 RMS 양자화 방법을 제안한 압축기의 구조에 맞도록 새로 개발하고, 상위 밴드의 특성에 적합하도록 웨이블릿 변환을 적용하여 상위 밴드 신호를 압축하였다. 이와 같은 새로운 방법을 적용하여 기존의 방법에 비하여 향상된 성능을 얻었으며, 주관적인 방법으로 압축기 성능을 측정할 결과 제안하는 16kbps 음성 압축기가 G.722 48 kbps SB-ADPCM보다 우수한 성능을 가지는 것을 확인하였다.

성능 측정 결과에 의하면 제안한 압축기가 아직 만족할 만한 성능을 가지지 못하며, 이를 해결하기 위하여 여러 블록에서의 성능 향상이 요구된다. 특히, 현재의 RMS 값을 이용한 단순한 하위/상위 밴드 비트 할당 방법의 문제점을 해결하고, 음향 심리학을 기반으로 각 밴드에서 필요 비트 수와 신호의 특성과의 상관 관계를 분석하고 이를 이용하여 성능을 향상시키는 연구가 필요하다. 또한, 웨이블릿 변환 및 웨이블릿 계수의 양자화, 상위 밴드 RMS 값의 양자화 방법에 대한 추가 연구가 필요하다.

감사의 글

본 연구는 2000학년도 광운대학교 교내 학술연구비의 지원으로 이루어졌습니다.

참고 문헌

1. TIA/EIA/IS-96, "Speech Service Option Standard for Wideband Spread Spectrum Digital Cellular System," 1994.
2. TIA/EIA/IS-733, "High Rate Speech Service Option for Wideband Spread Spectrum Digital Cellular System," 1996.
3. TIA/EIA/IS-127, "Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital System," 1997.
4. GSM 06,90, "Digital cellular telecommunications system (Phase 2+); Adaptive multi-rate (AMR) speech transcoding," 1998.
5. ITU Recommendation G.722, "7 kHz Audio-coding within 64 kBit/s," 1988.
6. ITU Recommendation G.722.1, "7 kHz Audio-Coding at 24 and 32 kbps for Hands-Free operation in system with low frame loss," 2000.
7. 3GPP TS 26.190, "AMR Wideband Speech Codec," 2001.
8. A. M. Kondoz, *Digital Speech: Coding for Low Bit Rate communication Systems*, England: Wiley, 1994.
9. N. S. Jayant and P. Noll, *Digital Coding of Waveform Principles and Applications to Speech and Video*, Prentice Hall, 1984.
10. R. M. Rao and A. S. Bopardikar, *Wavelet Transforms: Introduction to Theory and Application*, Addison Wesley, 1998.

저자 약력

● 송재중 (Jai-Chong Song)

1999년: 원광대학교 전자공학과 (공학사)
2001년: 광운대학교 전자공학과 (공학석사)
2001년~현재: 전자부품연구원 연구원
* 주관심분야: 음성신호처리, 오디오 신호처리

● 박호중 (Hochong Park)

1986년: 서울대학교 전자공학과 (공학사)
1987년: Univ. of Wisconsin-Madison, 전자공학과 (M.S.)
1993년: Univ. of Wisconsin-Madison, 전자공학과 (Ph.D.)
1993년~1997년: 삼성전자 선임연구원
1997년~현재: 광운대학교 전자공학과 부교수
* 주관심분야: 음성/오디오 신호처리, 통신 신호처리, 영상 신호처리