

논문 2008-45SP-2-9

# ITU-T G.729/G.729E와 호환성을 갖는 광대역 음성/오디오 부호화기

( A New Wideband Speech/Audio Coder Interoperable with ITU-T  
G.729/G.729E )

김 경 태\*, 이 민 기\*, 윤 대 희\*\*

( Kyung Tae Kim, Min Ki Lee, and Dae Hee Youn )

## 요 약

광대역 신호는 16 kHz로 표본화 되어 50-7000 Hz로 밴드 제한된 신호를 말하며, 전화대역 음성 신호에 비해서 높은 자연성(naturalness)과 명료성(intelligibility)을 가진다. 이런 특징으로 광대역 부호화기는 화상회의, 디지털 AM 방송 및 고음질 음성통신 등에 사용될 수 있다. 본 논문에서는 가변대역 특징을 갖는 광대역 음성·오디오 부호화기를 제안하였다. 제안된 부호화기는 대역분할 구조를 가진다. 저주파 대역은 전화대역 음성 부호화기로 많이 사용되고 있는 8 kbit/s ITU-T G.729나 보다 높은 전송률로 오디오 신호까지 처리할 수 있는 11.8 kbit/s ITU-T G.729 Annex E로 부호화한다. 고주파 대역은 청각 모델을 기반으로 한 파라미터 부호화 방법으로 부호화한다. 제안된 고주파 대역 부호화는 감마톤 필터뱅크(gammatone filterbank)를 이용하여 입력신호를 임계대역으로 분할한 후, 각각의 임계대역 신호를 양자화한다. 저주파 대역 부호화기와 고주파 대역 부호화기는 서로 독립되어 있으므로, 복호화기에서는 채널 조건에 따라 전화대역 합성신호와 광대역 합성신호를 선택할 수 있는 특징이 있다. 성능 평가 결과, 제안된 부호화기는 낮은 전송률과 짧은 지연 시간으로 음성과 오디오 신호 모두에 대해 ITU-T G.722.1 24 kbit/s와 동등한 음질을 제공한다는 것을 확인하였다.

## Abstract

Wideband speech, characterized by a bandwidth of about 7 kHz (50-7000 Hz), provides a substantial quality improvement in terms of naturalness and intelligibility. Although higher data rates are required, it has extended its application to audio and video conferencing, high-quality multimedia communications in mobile links or packet-switched transmissions, and digital AM broadcasting. In this paper, we present a new bandwidth-scalable coder for wideband speech and audio signals. The proposed coder splits 8kHz signal bandwidth into two narrow bands, and different coding schemes are applied to each band. The lower-band signal is coded using the ITU-T G.729/G.729E coder, and the higher-band signal is compressed using a new algorithm based on the gammatone filter bank with an invertible auditory model. Due to the split-band architecture and completely independent coding schemes for each band, the output speech of the decoder can be selected to be a narrowband or wideband according to the channel condition. Subjective tests showed that, for wideband speech and audio signals, the proposed coder at 14.2/18 kbit/s produces superior quality to ITU-T 8 kbit/s G.729.1 with the shorter algorithmic delay.

**Keywords:** 광대역 음성, 음성 부호화기, 가변 전송률, 가변 대역, G.729, 감마톤 필터뱅크, 인지 모델

## I. 서 론

통신이나 네트워크 환경의 비약적인 발전으로 인해 데이터 전송률은 높아졌으나 고음질을 제공하는 저전송

률 부호화기의 응용분야는 더욱 증대되고 있다. 이는 전송 비용이나 저장 비용을 최소화하고 무선 통신 채널처럼 제한된 전송용량을 갖는 환경에서 효율적으로 데이터를 전송하기 위해서 이다. 저전송률 부호화기에 대한 연구는 전화대역 신호(200-3400 Hz)에 대해서 활발히 진행되었으나, 최근 몇 년간 광대역 신호(50-7000 Hz) 부호화의 연구가 활발히 진행되면서, 저전송률을

학생회원, \* 정희원, 연세대학교 전기전자공학과  
(School of Electrical & Electronic Engineering,  
Yonsei University)  
접수일자: 2007년11월16일, 수정완료일: 2008년2월28일

갖는 광대역 부호화에 대한 관심이 높아가고 있다.

광대역 신호는 16 kHz로 표본화 되어 50-7000 Hz로 대역 제한된 신호를 말하며 전화대역 음성 신호에 비해서 높은 자연성(naturalness)과 명료도(intelligibility)를 가진다. 그리고 음성 신호에 비해 넓은 대역을 갖는 오디오 신호의 전송도 가능하게 된다. 이런 특징으로 인해 광대역 신호는 화상회의, 디지털 AM 방송 및 고품질 음성통신 등의 용도로 사용이 가능하다. 최근 들어 광대역 부호화기에 대한 관심의 증대로 많은 표준화 작업이 이루어지고 있다. ITU-T G.722<sup>[1]</sup>, G.722.1<sup>[2]</sup>, G.729.1<sup>[3]</sup>, MPEG 4 version 1<sup>[4]</sup>, AMR-WB<sup>[5]</sup>, EVRC-WB<sup>[6]</sup>는 대표적인 광대역 음성부호화 알고리즘이다.

이러한 표준화 작업과 더불어 여러 가지 광대역 부호화기 알고리즘들이 제안되었다. 하지만, 대부분의 연구가 음성만을 목적으로 한 부호화기에 집중되어 왔기 때문에 오디오 신호까지 고려한 부호화 알고리즘은 많지 않다. 일반적으로 음성 부호화 알고리즘과 오디오 부호화 알고리즘은 서로 전혀 다른 형태를 갖고 있기 때문에 저전송률을 갖는 알고리즘으로 음성과 오디오 신호 모두를 부호화하기에는 많은 어려움이 있다. 그리고 기존 오디오 부호화기 알고리즘의 경우, 음성 부호화기와 달리 긴 지연시간을 가지고 있어서 실시간 통신에는 부적합한 면이 있다.

광대역 부호화기 최근 경향 중 하나는 대역 분할 구조와 기존의 전화 대역 음성부호화기를 사용하여 현재 시스템과 호환성을 유지하는 것이다. ITU-T G.729<sup>[7]</sup>/G.729E<sup>[8]</sup>는 전화대역 음성에 많이 사용되는 부호화기 중 하나이다. 이 부호화기를 핵심 부호화기나 저주파 대역 부호화기로 사용할 경우, 광대역 음성부호화기는 현재 상용 시스템과 호환성을 유지할 수 있다.

본 논문에서는 14.2-18 kbit/s 전송률을 지원하는 광대역 음성/오디오 부호화기를 제안한다. 제안된 부호화기는 음성/오디오 신호를 전화 대역인 저주파 대역과 3.7 kHz 이상인 고주파 대역으로 분할하여 부호화한다. 저주파 대역의 음성/오디오 신호는 8 kbit/s G.729나 11.8 kbit/s G.729E를 사용하여 부호화하며, 고주파 대역 신호는 감마톤 필터<sup>[9]</sup>를 이용하여 임계 대역으로 분할하여 부호화한다. 고주파 대역 부호화를 위해 감마톤 필터의 여기신호를 주파수 영역에서 벡터 양자화하여 낮은 전송률을 유지한다. 제안된 부호화기는 전화대역 음성부호화기와 호환이 가능하며, 감마톤 필터의 인지 모델의 효과로 낮은 전송률에서 음성과 오디오 신호를 효율적으로 양자화할 수 있다.

### II. 제안된 광대역 음성/오디오 부호화기 개요

제안된 알고리즘은 음성과 오디오 신호를 동일한 알고리즘을 사용하여 부호화하기 위해서 분할대역, 시간 영역 부호화, 서브밴드 부호화가 결합된 구조를 가진다. 전체의 구조는 크게 분할대역 방식으로 이루어져 있어서 저주파 대역은 시간 영역 부호화 방법으로, 고주파 대역은 서브밴드 부호화 방법으로 부호화된다.

제안된 광대역 음성/오디오 부호화기와 복호화기의 전체 구조는 각각 그림 1, 그림 2와 같다. 그림 1에서 볼 수 있듯이 전체적으로 입력신호를 저주파 대역 신호(0-4 kHz)와 고주파 대역 신호(4-8 kHz)로 나누고 각각을 서로 독립적으로 부호화 한다. 먼저 저주파 대역 신호는 8 kbit/s 부호화기인 표준안인 ITU-T G.729나 11.8 kbit/s G.729E로 부호화하게 된다. 저주파 대역 신호를 부호화하기 위해 입력신호를 저주파 대역 통과 필터를 통과 시킨 후, 2:1 간축 과정을 거친다.

고주파 대역 입력 신호는 먼저 감마톤 필터뱅크를 이용하여 3700 Hz에서 7770 Hz에 해당하는 네 개의 대역으로 나뉜다. 그림 3은 대역통과 필터와 고주파 대역에 적용된 감마톤 필터뱅크의 주파수 응답을 보여주고 있다. 사용된 감마톤 필터는 0-7770 Hz의 대역을 21개의 임계대역으로 나눈 것 중에서 3700 Hz에서 7770 Hz에 해당하는 18 번째부터 21 번째 필터뱅크(4개의 필터뱅크)가 사용된다. 필터뱅크의 주파수특성을 모두 합한 전체 이득 곡선에는 2 dB 정도의 ripple이 존재하지만 음질에 큰 영향을 주지 않는다<sup>[10-11]</sup>.

감마톤 필터뱅크를 통과한 4 개의 신호는 각각 다음의 과정을 거친다. 먼저 반과정류를 한 후, 지역 최대치

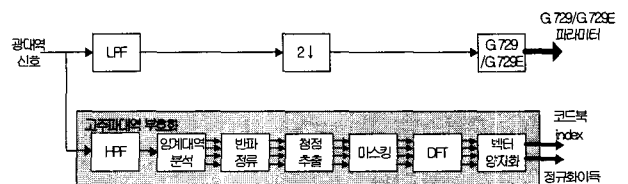


그림 1. 제안된 광대역 음성/오디오 부호화기  
Fig. 1. Proposed wideband speech/audio encoder.

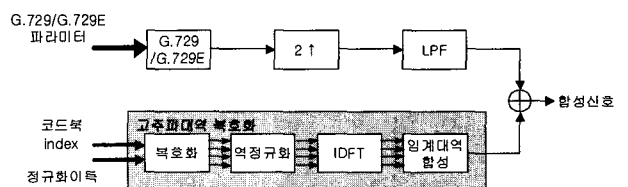


그림 2. 제안된 광대역 음성/오디오 복호화기  
Fig. 2. Proposed wideband speech/audio decoder.

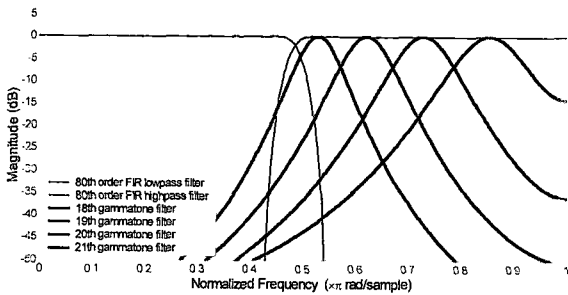


그림 3. 제안된 구조에서 사용한 필터뱅크  
 Fig. 3. Frequency responses of the filterbanks used in the proposed wideband coder.

만을 추출한다. 이렇게 얻어진 신호는 펄스가 조합된 형태로 나타나는데, 이 신호에 대해 순시 마스킹(temporal masking)과 동시 마스킹(simultaneous masking)을 적용한다. 이렇게 얻어진 신호는 몇 개의 펄스로만 이루어지는데 이것을 주파수 영역에서 벡터 양자화로 부호화한다. 이 때 주파수 계수를 얻기 위해 DFT(Discrete Fourier Transform)<sup>[12]</sup>을 사용한다.

복호화 과정은 부호화 과정의 역과정이다. 저주파 대역 신호를 얻기 위해 G.729나 G.729E로 부호화한 파라미터들은 각각에 해당하는 복호화기에 통과 시킨다. 이 때 후처리 필터는 사용하지 않는다. 고주파 대역 신호를 얻기 위해서 벡터 양자화한 변수와 정규화를 위해 사용한 이득값으로 감마톤 필터뱅크의 여기신호를 만든다. 이 때 벡터 양자화는 주파수 영역에서 이루어졌기 때문에 여기신호를 만들기 위해서는 IDFT(Inverse DFT) 과정이 필요하다. 이렇게 만들어진 여기신호를 감마톤 필터뱅크에 통과시킨 후, 저주파 대역에서 부/복호화된 신호와 합하면 제안된 부호화기의 부/복호화 과정이 완료된다.

### III 고주파 대역 신호 부호화

감마톤 필터뱅크를 이용한 광대역 음성/오디오 부호화 알고리즘은 이미 연구된 바 있다<sup>[10-11]</sup>. 이 알고리즘들은 전 대역을 감마톤 필터뱅크를 적용해서 부호화하여 감마톤 필터뱅크에 의한 많은 지연시간을 가지며 69.7 kbit/s의 높은 전송률을 요구하므로 유/무선 음성 통신과 같은 응용분야에 적합하지 않은 알고리즘이라 할 수 있다. 본 논문에서는 위의 알고리즘을 제안된 구조의 고주파 대역 신호 부호화에 적용하였다. 새로 제안된 방법은 기존에 방법에서 문제가 된 지연시간과 전송률 문제에 대한 해결 방안을 제시한다.

#### 1. 고주파 대역 신호 부호화 알고리즘 개요

제안된 구조에서는 고주파 대역에서만 감마톤 필터뱅크를 사용한다. 하지만 일반적으로 신호의 에너지 대부분은 저주파 대역에 집중되어 있으므로 원 신호를 고주파 대역 감마톤 필터뱅크에 통과시킬 경우 여전히 저주파 대역 성분이 많이 남아있을 수가 있다. 이러한 저주파 대역 성분은 우리가 고주파 대역 신호를 저주파 대역과 독립적으로 부호화하는데 방해 성분이 될 수 있다. 이 문제를 해결하기 위해 그림 1에서 볼 수 있듯이 고주파 대역 통과 필터를 사용한다.

제안된 구조에서는 기존의 구조<sup>[3]</sup>와 달리 펄스 신호를 부호화하기 위해 벡터 양자화를 이용한다. 우선 마스킹까지 거친 펄스 신호를 DFT를 한 후, 얻어진 DFT 계수의 절대값을 미리 만들어진 코드북을 바탕으로 벡터 양자화한다.

복호화단에서는 전송되어온 코드북 인덱스와 위상 진화 식을 사용하여 합성 필터뱅크의 여기신호를 만든다. 이 때, 위상 정보는 첫 프레임인 경우에는 난수를 발생시켜 사용하고, 그 이후로는 위상 진화 식으로 만들어진 값을 이용한다. 위상 진화식은 아래와 같다.

$$\theta_k^{(n)} = \theta_k^{(n-1)} + 2\pi f_k N \quad L_i \leq k \leq U_i \quad (1)$$

여기서  $\theta_k^{(n)}$ 는 현재 프레임의 위상을,  $\theta_k^{(n-1)}$ 는 과거 프레임의 위상을  $N$ 은 프레임 길이를 나타내고  $f_k$ 는 주파수 성분을 나타낸다.  $k$ 는 각 감마톤 필터뱅크의 주파수 대역 범위,  $(L_i, U_i)$ 에 속한 DFT 주파수 인덱스이다. 위상 진화 식을 이용한 경우와 원래의 위상 정보를 사용한 경우에 대한 실험 결과, 음질의 차이가 없음을 확인하였다. 이는 현재 부호화하고자 하는 부분이 음질에 있어 위상의 영향을 적게 받는 고주파 대역 신호이기 때문이다. 이렇게 만들어진 여기신호를 합성 필터뱅크에 통과시키고 각 대역 신호를 합하면 복호화 과정이 완료된다.

#### 2. 감마톤 필터뱅크

각 임계대역은 주파수가 증가할수록 사람의 인지 분해능이 떨어진다는 사실을 이용해서 인지하는 특성에 따라 주파수 대역을 나눈 것이다. 감마톤 필터뱅크는 임계대역을 기준으로 만들어진 필터 중 하나이며, FIR 필터로 구성되기도 하고 IIR 필터로도 구성되기도 한다<sup>[9]</sup>. 감마톤 필터의 충격응답은 식으로 표현된다.

$$g(t) = at^{(M-1)}e^{-2\pi h ERBf_c t} \cos(2\pi f_c t + \phi) \quad (2)$$

여기서  $M$ 은 필터의 차수를,  $f_c$ 는 중심주파수를,  $ERB(f_c)$ 는 지각 필터 (auditory filter)의 Equivalent Rectangular Bandwidth 을 나타내며,  $a$ 와  $b$ 는 상수이다<sup>[9]</sup>.  $ERB(f_c)$ 는 다음의 식으로 정해지는 함수이다.

$$ERB(f_c) = 24.7 + 0.108f_c \quad (3)$$

감마톤 필터뱅크의 Overall 이득값은 1 dB 가량의 ripple이 있지만 음질에 큰 영향을 주지 않는다.

기존의 알고리즘에서 사용한 감마톤 필터뱅크는 선형 위상 FIR 필터로 구현되었다. 충격 응답 함수  $g(t)$ 를 표본화해서 얻어진  $g(n)$ 과 이것의 시간축 역전된 신호를 convolution하여 얻어진 계수를 이용한다. 선형 위상 FIR 필터는 필터뱅크의 전체 이득 ripple을 더욱 크게 한다. 합성시 사용되는 필터뱅크의 영향을 고려할 때, 이러한 방법은 음질의 왜곡을 가져오게 된다. 또한 저대역부터 고대역까지 적절한 주파수 응답을 얻기 위해서는 40-50 msec 정도의 충격응답길이가 필요하므로 통신용 부호화기에 적합하지 않은 지연시간을 요구한다.

지각 필터로서 부/복호화 하고자 할 때 일반적으로 부호화 시 사용되는 필터와 복호화시 사용되는 필터는 서로 달라야한다. 만약 부호화 시에 사용된 필터의 전달함수가  $H_k(z)$  ( $k=1, \dots, 21$ ) 라고 한다면, 복호화 시에 사용되는 합성필터,  $G_k(z)$ 는

$$G_k(z) = \frac{H_k^*(z)}{\sum H_k(z)H_k^*(z)} \quad (4)$$

와 같이 되어야 한다. 만약  $\sum H_k(z)H_k^*(z) = 1$  이라면, 합성 필터뱅크는 시간역전 (time reverse)된 충격응답을 가지는 분석 필터뱅크와 같게 된다. 합성필터,  $G_k(z)$ 을  $H_k^*(z)$ 으로 사용할 경우 결과적으로 선형 위상 특성을 얻을 수 있다<sup>[12]</sup>. 이러한 특징을 이용하여 제안된 부호화기에서는 감마톤 필터의 충격응답,  $g(n)$ 의 적절한 길이만을 취해, 이것을 감마톤 필터의 계수로 사용하였다.

그림 4는 18번째 감마톤 필터의 충격응답을 보여주고 있다. 이 그림에서 알 수 있듯이 18번째 감마톤 필터는 60차 FIR 필터로도 충분히 표현할 수 있다. 감마톤 필터뱅크는 고주파 대역으로 갈수록 적은 필터 차수로도 정확한 스펙트럼 정보를 얻을 수 있다는 점을 고려할 때, 제안된 부호화기에서 사용되는 FIR 감마톤 필터뱅크의 차수는 60차면 충분하다.

분석 필터로  $g(n)$ 을 충격응답으로 가지는  $M$ 차 FIR 필터를 사용하고, 합성 필터로  $g(M-n)$ 을 충격응답으로 가지는  $M$  차 FIR 필터를 사용할 경우, 결과적으로  $M$  샘플 선형 지연되는 효과를 얻을 수 있다. 이를 고려할 때 제안된 구조에서는 60 샘플 지연으로도 구현이 가능하다. 하지만 저대역 합성 신호와의 지연시간을 맞추기 위해서 제안된 알고리즘에서는 120차 FIR 필터를 사용했다.

### 3. 반파정류와 첨점 추출

임계대역으로 나누어 각각의 대역 신호를 반파정류한 후 첨점 추출한 신호를 부호화하는 것은 인지생리학을 기반으로 한 것이다. 필터뱅크의 각 대역을 섬모의 한 묶음이라 생각하고 반파정류를 하는 것은 내이섬모 (inner hair cell)의 행동 방식을 모델링하는 것이다<sup>[10]</sup>.

각 섬모는 자극된 후, 자극 대기 상태 수준(reset state)까지 정해진 감퇴 시상수  $\tau$ 에 따라 지수적으로 자극 정도가 감소한다. 자극 대기 상태 수준은 자극 신호의 크기에 따라 변하는 값이다. 단위시간동안 섬모에 자극이 생길 확률은 자극 신호의 크기와 섬모의 상태 수준 크기의 차이에 비례하는 함수이다.  $\tau$ 보다 작은 주기를 갖는 정현파가 입력으로 들어 왔을 때는 정현파의 지역 최대치에서는 청각 자극이 빠르게 일어나고, 그

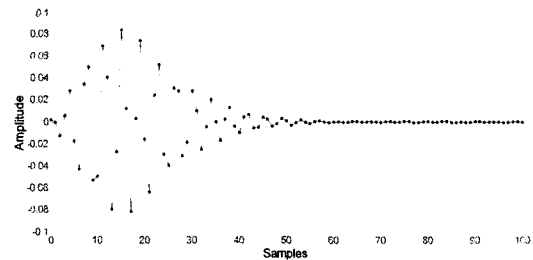


그림 4. 18번째 감마톤 필터뱅크의 충격응답  
Fig. 4. Impulse response of the 18th gammatone filterbank.

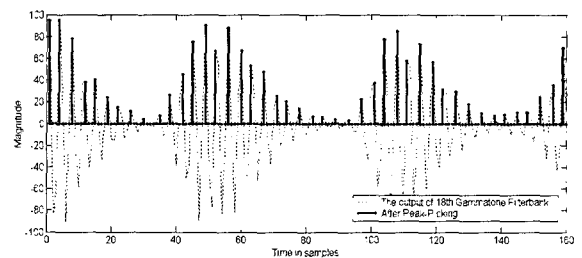


그림 5. 반파정류와 첨점 추출 과정  
Fig. 5. Example of halfwave rectification and peak-picking.

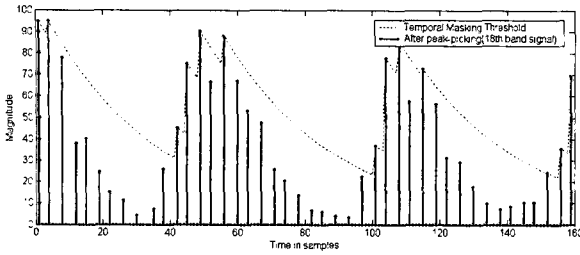


그림 6. 순시 마스킹 효과  
Fig. 6. Example of temporal masking effect.

외의 곳에서는 청각 자극이 느리게 일어난다고 가정한다. 즉, 자극률이 심해지는 곳이 신호의 주기에 따라 반복된다는 것이다<sup>[10]</sup>.

이 가정을 통해서, 각 필터뱅크의 출력신호는 섬모 자극 주기를 나타냄으로써 표현 가능하다. 즉, 각 필터뱅크의 출력신호의 지역 최대값만을 추출한 후 이를 다시 합성 필터뱅크에 통과시키면 우리가 원하는 소리를 얻을 수 있다는 것이다. 그러므로 부호화해야 하는 정보는 펄스의 위치와 크기가 된다.

위의 생리학적 모델을 신호적으로 처리하기 위해서 반파정류와 침점 추출의 작업이 필요하다. 반파정류는 다음과 같이 간단한 수식으로 표현된다.

$$w(n) = \begin{cases} y(n), & y(n) > y(n-1) \text{ and } y(n) > y(n+1) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

여기서  $w(n)$ 은 반파정류 결과 신호이고  $y(n)$ 은 침점 추출 결과 신호이다. 이 연산을 그림 5에 표현했다.

#### 4. 마스킹

반파정류/침점 추출 결과로 얻은 신호는 많은 펄스를 가지고 있으므로 이를 모두 부호화하기 위해서는 많은 비트를 필요로 한다. 기존의 부호화기는 펄스 수를 줄이고 보다 효율적인 부호화를 위해서 마스킹 효과를 이용한다.

제안된 부호화기는 고려할 수 있는 두 가지 마스킹, 동시 마스킹과 순시 마스킹 모두를 사용한다<sup>[13]</sup>. 본 논문에서는 동시 마스킹을 위해서 MPEG<sup>[14]</sup>에서 사용한 마스킹 기법을 이용했다. 각 대역에 해당하는 마스킹 임계치를 계산한 후 이보다 작은 값을 가지는 펄스는 제거했다. 하지만 동시 마스킹에 의해 제거되는 펄스의 수는 그다지 많지 않다. 이에 반해, 순시 마스킹은 많은 펄스들을 제거할 수 있다. 순시 마스킹은 어떤 강한 신호가 발생했을 때, 시간적으로 뒤에 발생하는 약한 신

호는 들리지 않는다는 특성을 반영한 것으로써 마스킹 임계치  $Y_i(n)$ 는 다음과 같은 수식으로 표현된다<sup>[11]</sup>.

$$Y_i(n) = \begin{cases} x'_i(n), & x'_i(n) > c_0 Y_i(n-1) \\ c_0 Y_i(n-1), & \text{otherwise} \end{cases} \quad (6)$$

여기서  $x'_i(n)$ 는 21개 필터뱅크 중  $i$  번째 펄스 신호를 동시 마스킹한 결과이고,  $c_0$ 는 시상수  $\tau_i$ 에 의해 결정되는 값으로  $c_0 = \exp(-\tau_i)$ 이다. 순시 마스킹 과정을 그림 6에 나타내었다.

#### 5. 여기신호 부호화와 양자화

마스킹까지 거친 펄스 신호는 몇 개의 펄스로 이루어진다. 마스킹을 하기 전에 비해서 펄스의 수는 현저히 줄어들지만, 프레임마다 다른 개수의 펄스 수와 무작위적으로 나타나는 펄스의 위치로 인해 적은 비트로 양자화하기는 어렵다. 이러한 문제를 해결하기 위해서 제안한 부호화기에서는 펄스 신호를 크기와 위치 정보로 부호화하지 않고, 주파수 영역에서 부호화하였다. 이 과정을 위해 DFT를 사용하여 주파수영역으로 변환한다. 주파수 영역의 스펙트럼 크기 정보만이 벡터 양자화의 목적 신호가 된다.

각 임계대역에 해당하는 주파수 계수의 절대값만을 추출하여 정규화 과정을 거친 후 이것과 가장 가까운 신호를 미리 만들어진 코드북에서 검색하게 된다. 여기서 각 대역에 해당하는 DFT 계수만을 양자화하는 것은 복호화단에서 여기신호가 합성필터뱅크를 통과할 때, 결국 대역에 해당하는 신호만 얻게 되기 때문이다. 제안된 부호화기는 프레임의 길이가 160 샘플이므로 각 임계대역에 해당하는 양자화를 위한 DFT 주파수 영역은 표 1와 같다.

정규화는 양자화 목적신호를 만들기 위해, DFT 계수를 최대값으로 나누는 과정이다.  $X_i(k)$ 을 마스킹된 펄스신호의 DFT 계수라고 할 때, 양자화 목적 신호  $T_i(k)$ 는

표 1. 양자화할 DFT 주파수 영역  
Table 1. Frequency ranges in DFT domain for quantization.

임계대역	DFT 주파수 영역 ( $L_i - U_i$ )
18	38-44(3700 Hz-4400 Hz)
19	45-54(4400 Hz-5300 Hz)
20	55-64(5300 Hz-6400 Hz)
21	65-71(6400 Hz-7000 Hz)

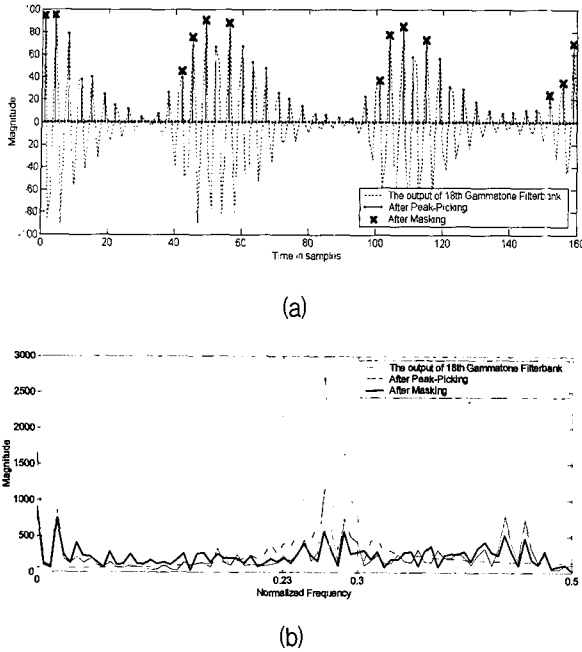


그림 7. 반파정류/첨점 추출과 마스크의 영향 (a) 시간 영역, (b) 주파수 영역)

Fig. 7. Effect of halfwave rectification / peak picking. (a) time domain, (b) frequency domain)

$$T_i(k) = \frac{|X_i(k)|}{\max(|X_i(k)|)}, \quad L_i \leq k \leq U_i \quad (7)$$

이 된다. 이 때,  $L_i$ 와  $U_i$ 는 표 1에 따른다. 최대 크기 값,  $\max(|X_i(k)|)$ 는 양자화되어 복호화단으로 전송된다. 정규화 후, MSE(Mean Square Error)을 기준으로 코드북 검색을 한다. MSE,

$$\epsilon = \frac{1}{U_i - L_i + 1} \sum_{k=L_i}^{U_i} [T_i(k) - C_i^m(k)]^2 \quad 0 \leq m < 2^M \quad (8)$$

을 최소로 하는  $m$ 을 선택하여 복호화단으로 전송한다. 이 때,  $M_i$ 는 각 대역에 할당된 비트이다. 이러한 코드북을 이용한 벡터 양자화는 부호화기의 구조를 크게 변화시키지 않고 전송률을 쉽게 조절할 수 있는 특징이 있다.

이와 같은 방식과는 달리 필터뱅크, 반파정류/첨점 추출과 마스크를 거치지 않고 원 신호의 고주파 대역 주파수 영역을 양자화하는 방법도 가능하다. 그러나 같은 길이의 코드북을 사용했을 때, 원신호를 양자화하는 방법이 제안한 방법보다 좋은 음질을 제공하지 못함을 실험적으로 확인하였다. 필터뱅크는 신호를 지각적 특징에 따라 구분하기 위한 것이고, 반파정류/첨점 추출과 마스크는 부호화를 위해 펄스 수를 줄여주는 것뿐만

표 2. 마스크의 벡터 양자화 효과 (10 비트 할당)  
Table 2. SegSNRs for vector quantization domains.

벡터 양자화 목적신호	SegSNR
임계대역 통과 신호	12.97
첨점추출 과정을 거친 신호	13.03
마스크된 신호	14.03

아니라 주파수 영역에서 양자화하기 좋은 조건으로 만들어주는 역할을 하기 때문이다.

그림 7은 반파정류/첨점 추출과 마스크의 영향을 보여주고 있다. 그림 7(a)에서는 시간 영역에서 18 번째 감마톤 필터를 통과한 신호, 반파정류/첨점 추출 후 신호와 마스크 후 펄스 신호를 각각 점선, 실선, ×로 표시했다. 그림 7(b)에서는 주파수 영역에서 같은 신호를 각각 점선, 실선, 굵은 실선으로 나타냈다. 그림 7(a)에서 볼 수 있듯이 마스크 후 신호는 몇 개의 펄스로만 이루어져 있다. 이 펄스의 크기는 순시 마스크 효과에 의해 비슷한 크기를 가진 펄스들로 이루어져 있다. 이러한 형태의 신호는 주파수 영역에서 비슷한 크기를 가진 하모닉의 조합으로 나타난다. 이러한 현상은 그림 7(b)의 점선 곡선과 굵은 실선 곡선을 비교함으로써 확인할 수 있다. 이것은 마스크 효과가 스펙트럼을 비슷한 크기를 가지는 하모닉의 조합으로 만들어 주는 역할을 한다는 것을 보여준다. 이렇게 비슷한 크기를 가지는 신호는 양자화하기가 좋다. 표 2는 감마톤 필터를 통과한 신호, 반파정류/첨점 추출 후 신호와 마스크 후 신호를 각각 10 비트로 벡터 양자화 했을 때의 segSNR을 보여준다. 표 2에서 볼 수 있듯이 마스크 후 신호를 벡터 양자화 할 때가 다른 경우보다 높은 seg-SNR을 가진다. 이는 마스크 후 신호가 벡터 양자화에 가장 적합하다는 것을 나타낸다. 즉, 제안된 부호화기에서 사용된 벡터 양자화의 효율을 높이기 위해서 마스크는 반드시 필요한 과정이다.

#### IV. 실험 결과 및 성능 분석

##### 1. 프레임과 시간정렬

저주파 대역 부호화를 위한 G.729/G.729E는 부/복호화 결과 예견구간에 의해 80 샘플 지연이 생기고 두 번의 저주파 대역 통과 필터에 의해 80 샘플 지연이 생기므로, 저주파 대역 부호화에서는 총 160 샘플의 지연이 발생한다. 반면, 고주파 대역 부호화에서는 고주파 대역 통과 필터에 의한 지연과 분석/합성을 위한 두 번의 감

마톤 필터뱅크에 의한 지연이 있다. 고주파 대역 통과 필터에 의해 40 샘플 지연이 발생하고 120차 FIR 감마톤 필터뱅크에 의해 120 샘플 지연이 발생한다. 고주파 대역 부호화 과정에서도 총 160 샘플 지연이 발생하므로, 저주파 대역과 시간정렬을 할 필요가 없다.

프레임의 길이는 G.729/G.729E의 프레임 길이에 맞추어 10 ms(160 샘플)이다. 그러므로 제안된 부호화기의 알고리즘 지연시간은 20 ms(160 샘플 + 160 샘플= 320 샘플)이 된다.

2. 비트 할당

저주파 대역 부호화에 G.729를 사용할 경우 낮은 전송률로 음성을 잘 부호화할 수 있고, G.729E는 부가적인 비트를 할당하여 오디오까지 부호화할 수 있다. 이를 이용하여, 본 논문에서는 낮은 전송률에서 음성 부호화만을 목적으로 하는 경우와 높은 전송률로 오디오 신호까지 부호화할 수 있는 경우에 대해 실험을 수행하였다.

자세한 비트할당은 표 3에 정리했다. 결과적으로 전송률은 저주파 대역 부호화에 G.729를 사용했을 경우 14.2 kbit/s, G.729E를 사용했을 경우 18 kbit/s가 된다. 여기 신호 양자화에 사용된 비트를 살펴보면, 정규화 크기는 비균일 스칼라 양자화되고, 각각 7 비트를 할당한다.

표 3. 제안된 부호화기의 비트할당  
Table 3. Bit-allocation of the proposed wideband coder.

변수		할당된 비트(bits)
저주파 대역 부호화	G.729/G.729E	80/118
고주파 대역 부호화	정규화 크기	28 (7 + 7 + 7 + 7)
	벡터 양자화	34(7 + 10 + 10 + 7)
총 계		142/180

3. 음질 평가

제안된 부호화기의 성능을 평가하기 위해서 주관적 음질평가를 했다. 주관적 음질평가는 음성 신호와 오디오 신호에 대해 수행했고, 음성 신호는 여성화자와 남성화자의 경우로 나누어서, 오디오 신호는 서로 다른 시간축 주파수축 특성을 갖는 다섯 가지 음악 샘플의 경우에 대해 평가했다. 2가지 부/복호화된 신호로 20명의 청취자에게 선호도 평가를 수행했다.

음성 신호는 16 kHz 표본화 된 8초 길이의 남녀 화자 각각 2명의 음성 샘플을 사용했고, 비교 대상은 저

표 4. 음성 신호 선호도 평가 결과  
Table 4. Results of preference tests for speech signals.

	Proposed (14.2 kbit/s)	No preference	G.722.1 (24 kbit/s)
Female 1	55 %	35 %	10 %
Female 2	40 %	45 %	15 %
Male 1	65 %	35 %	0 %
Male 2	55 %	40 %	5 %

표 5. 오디오 신호 선호도 평가 결과  
Table 5. Results of preference tests for audio signals.

	Proposed (18 kbit/s)	No preference	G.722.1 (24 kbit/s)
Sample 1	55 %	0 %	45 %
Sample 2	45 %	0 %	55 %
Sample 3	45 %	0 %	55 %
Sample 4	25 %	5 %	70 %
Sample 5	60 %	0 %	40 %

주파 대역 신호를 G.729로 부호화한 14.2 kbit/s 버전과 ITU-T G.722.1 24 kbit/s이다. 표 4는 음성 신호에 대한 선호도 평가 결과를 보여준다.

오디오 신호는 16 kHz 표본화 된 5가지 다른 음악 신호를 선택했다. 비교 대상은 저주파 대역 신호를 G.729E로 부호화한 18 kbit/s 버전과 ITU-T G.722.1 24 kbit/s 이다. 표 5는 오디오 신호에 대한 선호도 평가 결과를 보여준다. 음성의 경우 제안된 방법이 G.722.1 24 kbit/s보다 좋은 성능을 보이지만, 오디오 신호의 경우 선호도가 갈리는 경향을 보였다. 시간축 에너지 변화가 심한 신호에 대해서는 G.722.1가 좀 더 나은 성능을 보이는 것으로 관찰되었다.

V. 결 론

본 논문에서는 14.2/18 kbit/s 광대역 음성/오디오 신호를 부호화할 수 있는 알고리즘을 개발했다. 분할대역 구조를 바탕으로 한 제안된 부호화기는 저주파 대역 신호와 고주파 대역 신호를 독립적인 알고리즘으로 부호화한다.

저주파 대역 신호는 전화대역 음성 부호화기로 많이 사용하는 8 kbit/s ITU-T G.729나 11.8 kbit/s G.729E로 부호화한다. 고주파 대역 신호는 저주파 대역 합성 신호를 제거한 원 신호를 임계대역으로 나누어 주는 감마톤 필터뱅크에 통과시켜 양자화한다. 감마톤 필터뱅크 통과한 신호는 반과정류/첨점 추출과 마스킹 과정을

통과해서 펄스 신호가 되고 이 펄스 신호는 주파수 영역에서 벡터 양자화된다.

제안된 부호화기는 분할대역 구조를 갖추고 있기 때문에 저주파 대역 신호를 부호화하는 사용된 G.729나 G.729E와 호환이 가능하며 제안된 감마톤 필터뱅크의 적용으로 20 ms의 적은 알고리즘 지연시간을 갖는다. 효과적인 주파수대역 양자화의 도움으로 음성 부호화의 경우 14.2 kbit/s, 오디오까지 부호화할 경우 18 kbit/s의 낮은 전송률을 가진다.

주관적 음질평가 결과, 오디오와 음성신호 모두에 대해서 제안된 18 kbit/s 부호화기는 G.722.1 24 kbit/s와 비슷한 음질을 제공하고, 음성 신호의 경우 제안된 14.2 kbit/s 부호화기로도 G.722.1 24 kbit/s보다 우수한 음질을 제공할 수 있음을 확인할 수 있었다.

## 참 고 문 헌

- [1] ITU-T Rec. G.722 "7 kHz Audio-coding within 64 kbit/s," 1988.
- [2] ITU-T Rec. G.722.1 "Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss," 1999.
- [3] 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.
- [4] ISO/IEC, "Coding of Audiovisual Objects, Part 3: Audio, Subpart 3: CELP, Technical Report ISO/JTC 1/SC 29/N2203CELP," 1998
- [5] A. McCree, T. Unno, A. Anandakumar, A. Bernard, E. Paksoy, "An embedded adaptive multi-rate wideband speech coder," *Proc. of ICASSP*, pp.761-764, Utah, 2001.
- [6] V. Krishnan, V. Rajendran, A. Kandhadai and S. Manjunath, "EVRC-WIDEBAND: The new 3GPP2 wideband vocoder standard," *Proc. of the ICASSP*, pp.333-336, 2007.
- [7] ITU-T Rec. G.729 "Coding of Speech at 8 kbit/s CS-ACELP Speech Coder," 1996.
- [8] ITU-T Rec. G.729 Annex E "11.8 kbit/s CS-ACELP speech coding algorithm," 1998.
- [9] L. Lin, W. H. Holmes, E. Ambikairajah, "Auditory filter bank inversion," *Proc. of ISCAS*, pp.537-540, 2001.
- [10] E. Ambikairajah and J. Epps, L. Lin, "Wideband speech and audio coding using Gammatone filter banks," *Proc. of ICASSP*, pp.773-776, 2001.
- [11] G. Kubin and W. B. Kleijn, "On speech coding in a perceptual domain," *Proc. of ICASSP*, pp. 205-208, 1999.
- [12] A. V. Oppenheim and R. W. Schaffer, "Discrete-time signal processing: second edition" Prentice hall, 1998.
- [13] E. Zwicker and H. Fastl, "Psychoacoustics, Facts and Models, second updated edition," Springer, 1998.
- [14] K. B. Brandenburg, G. Stoll, "ISO-MPEG-1 audio: A generic standard for coding of high-quality digital audio," *J. Audio Eng. Soc.*, vol. 42, no.10, Oct. 1994.



— 저 자 소 개 —



김 경 태(학생회원)  
 2000년 연세대학교 기계·전자  
 공학과 학사 졸업.  
 2002년 연세대학교 전기전자  
 공학과 석사 졸업.  
 2008년 연세대학교 전기전자  
 공학과 박사 졸업.

<주관심분야 : 음성/오디오 신호처리>



이 민 기(학생회원)  
 2004년 연세대학교 기계·전자  
 공학과 학사 졸업.  
 2006년 연세대학교 전기전자  
 공학과 석사 졸업.  
 2008년 현재 연세대학교 전기전자  
 공학과 박사과정.

<주관심분야 : 음성/오디오 신호처리>



윤 대 희(정회원)  
 1982년 Kansas State University  
 전자공학과 박사 졸업  
 1982년~1985년 The University  
 of Iowa assistant  
 professor  
 1985년~현재 연세대학교  
 전자공학과 교수

2001년~2005년 연세대학교 공대 학장  
 2006년~2007년 연세대학교 교학부 부총장  
 <주관심분야 : 음성/오디오 신호처리>