

적응 필터뱅크를 이용한 오디오 부호화

Audio Coding Using Adaptive Filter Bank

신 유 철*, 강 현 철**, 변 윤 식*

(Yoo-Cheol Shin*, Hyunchul Kang**, Youn-Shik Byun*)

*본 논문은 인천대학교 '97년도 교내연구비 지원에 의함.

요 약

본 논문은 두 부류의 오디오 원에 대해 각각 다른 구조를 가지는 필터뱅크를 설계하고 스위칭 기준을 제안한다. 균일한 필터뱅크로는 MDCT 필터뱅크를 사용하고 비균일한 필터뱅크로는 웨이블릿 패킷 필터뱅크를 사용하였으며 오디오 신호의 시변 특성에 기초하여 두 필터뱅크를 스위칭한다. MDCT 필터뱅크는 정상신호 표현에 적절하지만 급격한 변화를 포함하는 오디오 신호를 표현하는데는 적절하지 못하다. 따라서, 본 연구에서 사용한 웨이블릿 패킷 필터뱅크는 인간의 청각 특성을 고려한 임계대역(critical band)과 유사하게 설계하였으며 스위칭 기준에는 에너지-엔트로피(energy-entropy), 영교차(zero-crossing)법 그리고 차분(difference)기준을 사용하였다. 입력되는 오디오 신호의 통계적 특성에 기안하여 두 필터뱅크를 스위칭하는 방식의 오디오 부호화기에 대해서 새로운 스위칭 기준을 제안하였다. 여러가지 오디오 신호에 대한 주관적 평가(MOS)를 수행한 결과, 기존의 부호화기보다 좋은 성능을 보였다.

ABSTRACT

This paper designs a filter bank which has different structure for two classes of audio source and proposes a new switching criterion. In this filter design, the MDCT(modified discrete cosine transform) filter bank is used as uniform filter bank and wavelet packet filter bank as nonuniform filter bank. Two filter banks are switched based on the time-varying characteristic of audio signal. Modified discrete cosine transform filter bank is adequate for stationary signal representation, but not for audio signal including attack. Wavelet packet filter bank is designed similarly to the critical band of human auditory system. Switching criterions are calculated from energy-entropy, zero-crossing and difference method. The enhanced switching criterions based on the statistical characteristics of the input signals are proposed. This audio coder has received higher score than existing audio coder from the result of MOS(mean opinion score) test.

I. 서 론

고음질 오디오 신호는 매우 큰 대역폭을 가지는 것이 일반적인 특징이다. 즉, 신호의 대역폭은 가청 주파수를 44.1kHz로 표본화하고 각 표본치를 16비트 선형 PCM을 이용하여 양자화 했을 경우, 결과적인 전송율은 약 1.5Mbits/sec(2채널)이다. 고음질 오디오 압축 알고리즘의 목적은 산업표준인 CD음질을 유지하면서 비트율을 가능한 적게 유지하는 것이다. 그예로서 MPEG(Moving Pictures Experts Group) 오디오 표준이 가장 널리 보편화 되어 있다. 그러나, MPEG오디오는 64kbps이하의 전송율에서는 음질이 크게 떨어지는 단점이 있다. 이러한 MPEG 오디오는 MDCT(modified discrete cosine transform)를 사용하여 32개의 균일한 필터뱅크로 부대역(subband) 분

할을 한다[2, 3, 4].

최근, 유용한 수학적 도구로서 알려져 있는 웨이블릿 변환(wavelet transform)에 대한 연구가 활발히 진행되어 많은 분야에 적용되고 있다. 웨이블릿 변환은 비정상(non-stationary) 신호에 대해서 다양한 분해능력을 가지고 있기 때문에, 부대역 부호화(subband coding), 다중해상도 분석(multiresolution analysis) 과도신호의 검출, 음성/영상 신호의 압축 코딩, 잔향 제거기 등과 같은 응용에 적용된다[5, 6, 7]. 특히, 웨이블릿 변환은 비정상신호에 대해서 탁월한 분해능을 가지기 때문에 오디오와 같이 비정상신호가 빈번히 발생하는 응용에는 필수적이라 할 수 있다. 그러나 웨이블릿 변환은 정상신호에 대해서는 균일한 주파수 해상도를 가지는 MDCT 필터뱅크보다 변환 이득(transform gain)의 관점에서 불합리하다[8, 9].

본 논문은 입력신호의 시변 특성에 기초를 두고 각각의 필터뱅크를 스위칭하는 구조를 제시하고 더 나아가 스위칭 기준을 제안한다. 균일한 필터뱅크는 MDCT 필터

*시립 인천대학교 전자공학과

**시립 인천대학교 정보통신공학과

접수일자: 1997년 12월 19일

뱅크를 비균일한 필터뱅크에는 웨이브렛 패킷 필터뱅크를 사용한다. MCDT 필터뱅크는 정상신호(대부분의 음악, 악기, 음성)에 대해서 사용되고 웨이브렛 필터뱅크는 비정상신호(드럼, 캐스터네츠)에 대해서 사용된다. MDCT 필터뱅크의 윈도우 스위칭을 이용하여 비정상신호를 처리할 수 있지만 모든 주파수에서 동일한 시간해상도를 가지게 된다. 그리고 웨이브렛 필터뱅크는 고유적으로 합성 과정 중 양자화 잡음의 피침현상(preecho)을 제거할 수 있다. MDCT 필터뱅크와 웨이브렛 패킷 필터뱅크와의 스위칭 기준은 매우 중요하다. 만약, MDCT 필터뱅크에서 비정상 신호를 처리한다면 저전송율에서 잡음을 만들어 낼 것이고, 웨이브렛 필터뱅크에서 정상신호를 처리한다면 저전송율에서 잡음을 만들어 낼 것이고, 웨이브렛 필터뱅크에서 정상신호를 처리한다면 처리전력의 낭비를 가져올 것이다. 전형적인 비정상 신호인 캐스터네츠를 DCT보다도 WT로 변환했을 때 변환 이득(transform gain)이 10-20dB 정도 더 크다[8, 9, 10].

II. 부대역 오디오 부호화

2.1 균일한 필터뱅크

균일한 필터뱅크를 구성하기 위해서 MPEG 오디오에서 사용한 폴리페이즈(Polyphase) 필터뱅크를 이용하여 오디오 입력신호를 균일한 32개의 부대역으로 나누었다. 부대역, $h_{sub}(n)$ 의 임펄스 응답은 원형저역필터 $h(n)$ 의 임펄스 응답의 변조에 의해서 구해진다[2, 4, 7]. 여기서 사용되는 분석행렬과 합성행렬은 다음과 같다.

$$M_{ik} = \cos \left[(2i + 1)(k - 16) \frac{\pi}{64} \right] \quad i = 0, 1, \dots, 31 \quad (2-1)$$

$$N_{ik} = \cos \left[(i + 16)(2k + 1) \frac{\pi}{64} \right] \quad k = 0, 1, \dots, 63$$

합성과정은 분석과정의 역순으로 수행되어진다. 물론 완전 복원을 만족한다.

전체 필터뱅크는 식(2-2)와 같이 표현된다.

$$s_i[i] = \sum_{k=0}^{63} \sum_{j=0}^7 M[i][k] \times (C[k + 64j] \times x[k + 64j]) \quad (2-2)$$

여기서, i 는 범위가 0에서 31를 가지는 부대역 인덱스이다. $s_i[i]$ 는 시간 t 에서의 부대역 i 에 대한 출력 표본값이다. $C[m]$ 은 표준에 정의된 분석 윈도우 512개의 계수 중 하나를 나타내고, $x[m]$ 은 512개의 표본 버퍼에서 읽혀 들어오는 입력 오디오 표본값이다. 그리고, 분석행렬과 합성행렬은 식(2-1)에서 구해진다.

32개의 입력 표본값에 대해서 분석 필터뱅크는 32개의 출력값을 가진다. 32개의 필터출력들은 $512 + 32 \times 64 = 2,560$ 번의 곱셈과 $64 \times 7 + 63 = 2,464$ 번의 덧셈이 요구된다. 즉, 한 개의 출력당 80번의 덧셈과 곱셈이 필요하다.

그러나 FFT 알고리즘으로 상당히 많은 곱셈과 덧셈 연산의 감소를 기대할 수 있다. 부대역 필터의 각각은 입력 오디오 신호를 32로 다운샘플링(downsampling-subsampling)하여 대역별로 한 개의 표본값만 출력한다.

식(2-2)를 중첩(convolution)으로 다시 표시하면 다음 식과 같다.

$$s_i[i] = \sum_{n=0}^{511} x[t-n] H_i[n] \quad (2-3)$$

식(2-3)에서 $x[t]$ 는 시간 t 에서 오디오 표본값이고, $H_i[n]$ 은 다음 식에서 구해진다.

$$H_i[n] = h[n] \times \cos \left[\frac{(2 \times i + 1) \times (n - 16) \times \pi}{64} \right], \quad 0 \leq n \leq 511 \quad (2-4)$$

$$h[n] = \begin{cases} -C[n], & (n/64) \text{의 정수부가 홀수} \\ C[n], & \text{otherwise} \end{cases} \quad (2-5)$$

여기서, 필터뱅크의 각 부대역은 고유의 대역통과필터(band pass filter) 응답 $H_i[n]$ 을 가진다. 식(2-4)는 신호해석을 위해서는 편리하지만 충분한 해를 제공하지 못한다. 즉, 이 식을 직접 적용한 경우에는 32개의 필터뱅크 출력을 계산하기 위해서 $32 \times 512 = 16,384$ 개의 곱셈과 $32 \times 511 = 16,325$ 번의 덧셈이 요구된다.

2.2 비균일한 필터뱅크

본 연구에서 사용한 비균일한 필터뱅크는 웨이브렛 패킷 변환(Wavelet Packet Transform)을 이용하여 설계하였다. 먼저, 웨이브렛 변환에 대한 기본적인 이론을 기술한 다음에 본 연구에서 사용된 웨이브렛 패킷 필터뱅크에 대해서 기술한다[5, 6, 8, 16, 17, 18, 19].

2.2.1 웨이브렛 변환(Wavelet Transform)

STFT의 시간/주파수 분해능의 제한을 극복하기 위한 방법으로 웨이브렛 변환을 이용할 수 있다. 웨이브렛 변환 이론은 신호나 프로세스를 특정한 함수들의 집합(wavelet set)을 가지고 모델링하는 수학적 도구이다. 웨이브렛 함수는 진동하며 평균값이 0인 유한 에너지 함수라고 할 수 있으며, 특정한 하나의 웨이브렛(mother wavelet)으로부터 발생하는 웨이브렛 함수들의 집합은 임의의 함수를 근사화 하는데 사용될 수 있다. 즉, 임의의 함수는 적당한 가중치를 준 웨이브렛 집합의 각 원소에 신호를 투사(projection)시켜 얻을 수 있으며, 이를 웨이브렛 변환으로 정의한다. 웨이브렛 변환식은 식(2-6)와 같다[5, 6, 17].

$$W_g f(a, \tau) = |a|^{-\frac{1}{2}} \int f(t) g^* \left(\frac{t-\tau}{a} \right) dt \quad (2-6)$$

a : 스케일링 인자(scaling factor)

τ : 병진 인자(translation facton)
 $g(t)$: mother 웨이브렛 함수

웨이브렛 평면에 투사된 함수는 웨이브렛 역변환을 거쳐 다시 복원될 수 있다. 역변환식은 식(2-7)과 같다.

$$f(t) = \frac{1}{C_x} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_x f(a, \tau) = |a|^{-1/2} g\left(\frac{t-\tau}{a}\right) \frac{d\tau da}{a^2} \tag{2-7}$$

웨이브렛 변환은 관측 윈도우(observation window)가 고정되어 있는 STFT와 달리, 분석창이 신호의 주파수 대역에 따라 상대적인 대역폭을 가지게 되어 시간/대역폭 곱 범위 내에서 임의로 시간/주파수 분해능을 조절할 수 있다. 해석 방법을 필터뱅크의 관점에서 볼 때, 시간 분해능은 해석필터의 중심 주파수에 따라서 증가해야 한다. 따라서, $\Delta\omega$ 는 ω 에 비례한다는 것을 이용한다[5, 20].

$$\frac{\Delta\omega}{\omega} = c, \quad c: \text{임의의 상수} \tag{2-8}$$

저주파 부분에서는 긴 윈도우를 적용하고 고주파 부분에서는 짧은 윈도우를 사용함으로써, 저주파 부분에서는 시간 분해능보다 주파수 분해능을 좋게 하며, 고주파 부분에서는 주파수 분해능은 나쁘지만 시간 분해능을 좋게 하여 신호 분석을 유리하게 해준다. 이들은 웨이브렛 변환의 기저함수(basis function)로서 mother 웨이브렛을 스케일링(scaling)하거나 병진(translating)함으로써 구해지며, 상호 직교성을 가진다. 이런 직교 함수들은 mother 웨이브렛의 선택에 따라 그 특성이 결정된다.

시간/스케일 인자 a , τ 를 이산화함으로써 식(3-11)로부터 식(3-14)과 같이 이산 웨이브렛 변환을 구할 수 있다.

$$C_{m,n} = \sum_{k=-\infty}^{\infty} f(k) g_{m,n}^*(k) \tag{2-9}$$

$$g_{m,n}^*(k) = a_0^{-m/2} g(a_0^{-m}k - nT) \\ a = a_0^m, \quad \tau = na_0^m T$$

식(2-9)에서 이산 웨이브렛 패킷, $g_{j,l}(n)$ 는 각각의 다른 값을 갖는 스케일 변수 a 와 병진 변수 τ 에 의해 서로 다른 분해능을 제공함으로써, 웨이브렛 변환은 신호를 각각의 분해능에 따른 성분들로 분해함으로써 다분해능 해석(multiresolution)이라 할 수 있다. 실제로 이산 웨이브렛 변환에서 $a_0 = 2$, $\tau = 1$ 로 놓는다.

$g(n)$ 이 이상적인 반대역 저역통과 필터이고 $h(n)$ 이 이상적인 반대역 고역 통과 필터이면 두 개의 서브 샘플링 형태로 원신호를 완전하게 표현할 수 있다.

대역분할 신호처리 기법에서 발생할 수 있는 대역 간섭 현상을 제거한 웨이브렛 변환의 장점은 이상적인 필터를 사용하지 않아도 되며, $x(n)$ 을 두개의 필터링되고 다운샘

플링된 신호 $y_0(n)$ 과 $y_1(n)$ 으로 복원할 수 있다는 것이다. 원신호를 복원하기 위해서는 두 신호를 오버 샘플링(over sampling)한 뒤 각각 $g'(n)$ 과 $h'(n)$ 으로 필터링하고 두 신호를 더한다. 복원된 신호 $\hat{x}(n)$ 은 필터들이 특정한 조건을 만족하지 않으면 $x(n)$ 과 다르다. 이들 조건을 만족하는 필터들은 완전 재구성 특성(PR-Perfect Reconstruction property)을 갖는다[5, 6, 16, 20, 21].

여기서 사용하는 FIR(finite impulse response)필터를 가정하면 고역통과 필터와 저역통과 필터사이의 관계는 식(2-10)과 같이 나타낼 수 있다.

$$h(n) = (-1)^n g(L-1-n) \\ g'(n) = g(L-1-n) \tag{2-10}$$

여기서 L 은 필터의 길이이다.

2로 다운샘플링한 뒤의 중첩(convolution)을 계산하는 이 필터뱅크는 $x(n)$ 과 $\{g(2k-n), h(2l-n)\}$ 의 내적을 구하는 것이므로 식(2-11)과 같이 각 필터 단의 출력을 나타낼 수 있다.

$$y_0(k) = \sum_n x(n) g(2k-n) \\ y_1(k) = \sum_n x(n) h(2k-n) \tag{2-11}$$

필터의 임펄스 응답은 정규화된 직교 집합으로 구성하므로, 식(2-12)과 같이 $x(n)$ 을 간단히 복원할 수 있다.

$$\hat{x}(n) = \sum_{k=-\infty}^{\infty} \{y_0(k) g(2k-n) + y_1(k) h(2k-n)\} \tag{2-12}$$

식(3-10)과 (3-12)으로부터 합성필터가 시간에 역으로 분석필터와 동일하다는 것을 알 수 있다. 본 연구에서는 가장 일반적으로 사용하는 Daubechies 필터를 사용하였다. Daubechies 필터는 식(2-13)의 완전복원 조건인 Power complementary 조건을 만족한다[16, 21].

$$|G(\omega)|^2 + |H(\omega)|^2 = c \tag{2-13}$$

여기서, c 는 상수이다.

위에서 설명한 QMF 뱅크를 이용하여 다해상도 분석(Multiresolution Analysis) 즉, 고속 웨이브렛 변환(FWT)이 가능해진다. 본 연구에서는 인간의 청취각각을 고려한 임계대역(critical band)과 유사하게 비균일한 필터뱅크를 설계하였다[8, 18, 22].

III. 하이브리드(Hybrid) 필터뱅크를 이용한 오디오 부호화

본 연구에서는 위에서 기술한 균일한 필터뱅크인 MDCT 필터뱅크와 비균일한 필터뱅크인 웨이브렛 패킷 필터뱅크

크를 스위칭하는 하이브리드 필터뱅크를 사용한다. 오디오 신호의 갑작스런 변화 즉, 비정상 신호를 부호화하는 것은 오디오 부호화기의 설계에 있어서 문제점(동일한 블록에서 낮은 레벨의 신호에 대한 양자화 잡음-preecho)을 가지고 있어서 많은 연구가 진행되어져 왔다. 또한 입력 오디오 신호의 특성에 따라 변환 이득의 관점에서 상당한 불합리성을 가지고 있다. 균일한 필터뱅크는 정상 신호를 표현하는데 적절한 반면에 비균일한 필터뱅크는 비정상신호를 표현하는데 적절하다는 특성을 가진다. 따라서, 하이브리드 필터뱅크는 두 종류의 필터뱅크의 장점을 취함으로써 부호화기의 성능을 향상시킬 수 있다. 비정상 신호를 부호화하는데 있어서 이러한 필터뱅크는 저 전송율에서 향상된 음질을 유지할 수 있다. 또한, 시간/주파수 해상도의 적절한 임계점을 찾는 것은 상당히 중요한 요소로서 작용한다. 이러한 시간/주파수 해상도는 인간의 청각구조에 따라 설계되어진다. 본 장에서는 스위칭 구조의 이점에 대해서 기술하고 본 연구에서 제안된 스위칭 기준에 대해서 설명한다.

3.1 스위칭 구조의 이점

MDCT 필터뱅크(폴리페이즈 필터뱅크)와 웨이브렛 패킷 필터뱅크를 스위칭 하는 기준은 시스템 전체의 성능을 좌우하는 매우 중요한 요소이다. 입력되는 오디오 신호의 갑작스런 변화(attack)를 적절히 추출해 내는 것이 스위칭 기준을 설정하는 목적이다. 기본적인 정의는 입력되는 신호의 통계적 특성, 즉 변화량(평균, 분산, 엔트로피, 주파수)을 적절히 추출해 내야한다. 만약, attack이 MDCT 필터뱅크의 입력으로 사용된다면 특히, 저전송율(low bit rate)에서 상당한 왜곡을 가져올 것이다. 반면에, 정상신호가 웨이브렛 패킷 필터뱅크로 사용된다면 처리 전력에 있어서 낭비를 가져올 것이다[22].

이러한 스위칭 기법을 이용한 오디오 부호화기는 두 가지 관점에서 명확히 설명될 수 있다. 입력되는 신호의 특성에 따라 웨이브렛 변환과 변형된 이산 역변환에 대한 변환이득의 차이가 있다. 또 하나는 각 변환에 대한 지각적 마스킹 특성에 의해 설명된다. 여기서 각 관점에 대해서 간단히 설명하도록 한다.

◎ 변환이득

스위칭 기법의 사용여부에 대한 한가지 관점은 변환 이득(transform gain)의 차이이다. 즉, 어느 한 필터뱅크만을 사용한다면 변환 이득의 관점에서 상당한 불합리성을 내포하고 있다. 여기서 변환이득은 식(3-1)로서 정의된다[21, 22].

$$G_{TC} = \frac{\frac{1}{N} \sum_{i=0}^{N-1} \sigma_i^2}{\left(\prod_{i=0}^{N-1} \sigma_i^2\right)^{\frac{1}{N}}} \quad (3-1)$$

여기서, σ_i^2 는 i 번째 변환된 계수 θ_i 의 분산(variance)이다. 식(3-1)이 내포하고 있는 의미는, 변환의 효율성은 그 변환에 의해서 얼마나 많은 에너지가 compaction하는가가 보여준다. 즉, 변환된 계수의 분산에 대한 산술적 평균(arithmetic mean)과 기하학적 평균(geometric mean)의 비(ratio)로서 나타낸다. 일반적으로 정상신호에 대해서는 비균일한 필터뱅크인 웨이브렛 패킷 변환보다는 균일한 필터뱅크인 MDCT 필터뱅크가 변환 이득이 10~20dB 정도 높고, 반대로 비정상 신호에 대해서는 MDCT 필터뱅크보다 웨이브렛 패킷 필터뱅크가 더 높은 변환 이득을 가진다. 그러므로, 이러한 스위칭 구조를 가지는 필터뱅크가 오디오 신호와 같이 비정상 신호가 빈번히 발생하는 응용에는 바람직하다[8, 22].

◎ 지각적 마스킹 특성(Perceptual masking property)

지각적 마스킹 특성의 관점에서 스위칭 기법에 대한 타당성을 설명하도록 한다. 정상 신호에 대한 마스킹 효과에 대한 결과를 고려한다면, 변형된 이산 역변환을 이용한 부호화기가 웨이브렛 변환을 이용한 부호화기보다 더 많은 이점이 있다. 매우 낮은 상호 주파수 마스킹(inter-frequency masking) 특성을 가지는 순음과 같은 임의의 신호에 대해서, 웨이브렛 변환 또는 어느 한 부대역의 마스킹 특성은 임의로 정해진 제한된 주파수 해상도 때문에 문제점을 가질 수 있다. 즉, 설계한 필터뱅크의 주파수 해상도가 임계대역과 일치하지 않을 경우 근접한 대역에 대한 왜곡을 발생시킨다. 그러나, 이러한 문제점의 웨이브렛 패킷 필터뱅크를 임계대역 스케일(bark)과 일치하도록 한다면 감소시킬 수 있다. 반면에 MDCT 필터뱅크를 이용한 부호화기에 있어서, 고주파에서 제한된 시간 해상도는 프리에코(preccho)를 발생시킬 것이다. 즉, 어느 한 블록(프레임)에서 attack이 발생할 경우 낮은 레벨의 신호에 대해서 양자화의 합성과정에서 잡음이 발생하게 된다. 고유적으로 웨이브렛 패킷 변환은 양자화 스텝 사이즈를 연속적으로 가변하기 때문에 프리에코를 제거할 수 있고 고주파에서 높은 시간 해상도를 제공한다. 물론, MDCT 필터뱅크를 이용한 부호화기에서 이러한 프리에코의 문제는 윈도우 스위칭(window switching) 기법을 이용해서 제거할 수 있으나, 이것은 프리에코를 완전히 제거하지는 못한다. 따라서, 웨이브렛 패킷 변환과 MDCT 필터뱅크를 스위칭 하는 구조가 이러한 문제를 해결할 수 있다.

3.2 스위칭 기준

3.1절에서 언급했듯이 스위칭 기준(switching criterion)은 매우 중요한 요소로서 사용된다. 본 연구에서는 3가지 스위칭 기준을 가지고 모의 실험을 해보았다. 실제 구현상, 입력되는 오디오 신호의 표본값에 대한 정상성, 비정상성을 판별하기 위해서 블록(프레임) 단위로 검사하게

된다. 이 절에서는 각 스위칭 기준에 대한 특성을 설명하고 실험결과는 4장에 보인다.

1. 에너지-엔트로피 기준(Energy-entropy criterion)

본 기준은 식(4-2)에서 보여진 것처럼 확률적 엔트로피와 유사한 형태를 가지고 있다[8].

$$I = - \sum_{i=0}^{N-1} \sigma_i^2 \log_2 \sigma_i^2 \text{ bits} \quad (3-2)$$

여기서, N 은 프레임당 세그먼트의 수이다. 에너지-엔트로피는 각 프레임을 N 개의 세그먼트로 나눔으로써 계산된다. 44.1kHz로 표본화된 오디오 표본치에 대해서는 $N=16$ 으로 설정하는 것이 적당한것으로 알려져 있다[8]. 입력되는 오디오 신호의 에너지는 이러한 각 세그먼트상에서 계산되고 전체 프레임의 에너지에 대해서 정규화된다. σ_i^2 는 i 번째 세그먼트의 정규화 에너지(<1)이다. 일반적으로 상수(constant)의 에너지를 가지는 프레임에서 에너지-엔트로피가 가장 크게 나타난다. 반면에, 갑작스런 에너지 천이를 가지는 프레임에서 에너지-엔트로피는 떨어진다. 즉, 현재 프레임과 이전 프레임 사이의 에너지-엔트로피 변화율에 따라 오디오 부호화기는 스위칭 작용을 하게 된다[8].

2. 영교차 기준(Zero-crossing criterion)

본 기준은 입력되는 오디오 신호의 영교차 수를 카운트(count)하여 그 수의 많고 적음에 따라 스위칭하는 기준이다. 즉, 시간 축을 기준으로 교차되는 수를 카운트하여 그 값을 저장한다. 정확한 스위칭 판별을 위하여 입력되는 오디오 신호의 저주파 부분의 신호를 높은 주파수 성분만을 가지고 카운트하였다. 저역부분을 제거하는데도 웨이블릿 변환을 2번 수행하여 저역신호를 추출한 다음, 원신호에서 감산하였다. 일반적으로 비정상신호는 높은 주파수를 포함한다는 특성을 이용하였다. 오디오 신호의 갑작스런 변화에서 높은 수의 영교차수를 가졌다. 에너지-엔트로피 기준과는 달리, 현재 프레임의 영교차 수가 정해진 임계값을 넘을 경우 스위칭하도록 설계하였다.

3. 차분 기준(Difference criterion)

본 스위칭 기준은 이전 프레임의 제곱 평균과 현재 프레임의 제곱 평균을 빼어 그 차이가 큰 것을 비정상 신호를 포함한 것을 간주하였다. 비정상 신호일수록 높은 제곱 평균값을 가진다는 특성을 이용하였다. 마찬가지로 대부분 비정상 신호는 이전 프레임의 평균과 현재 프레임의 평균과의 차이가 크다.

IV. 모의 실험 및 결과 검토

4.1 전체구조

본 연구에서 사용된 전체 시스템의 구성도는 그림(4-1)과 같다. 먼저, 3장에서 기술된 스위칭 기준을 이용하여 입력신호의 통계적 특성에 따라 스위칭 정보를 추출했다. 이러한 스위칭 기준에 따라 입력 오디오 신호는 어느 한 필터뱅크의 입력으로 선택되어지고 필터링된 후, 양자화 과정을 거쳐 부호화 하였다. 저전송율을 실현하기 위한 동적 비트할당(dynamic bit allocation)정보는 MPEG 오디오의 심리음향 모델을 사용하여 추출해 내었다. 모든 입력 데이터는 프레임 단위로 하였으며, 한 프레임은 MPEG 오디오와 같은 1152개의 오디오 샘플을 포함시켰다. 오디오 샘플은 저음, 중음, 고음 그리고 변화량이 큰 오디오 신호까지 포함하여 어느 한 부분에 편중되지 않도록 선택되어졌으며 5가지(Piano, Violin, Drum, Crash, Saxophon)를 사용하였다. 양자화 과정에서 스케일 팩터와 스위칭 정보는 부정보(Side information)로 프레임마다 패킹(packing)되어 비트스트림(bitstream)을 구성한다.

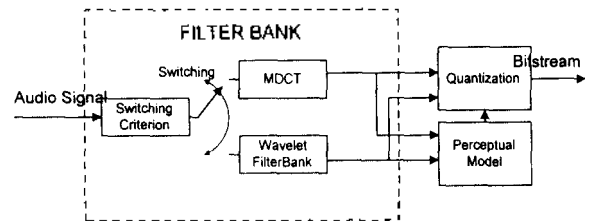


그림 4-1. 사용된 오디오 부호화기의 구조

4.2 각 스위칭 기준에 대한 성능 평가

44.1kHz로 표본화된 오디오 신호, 즉 Drum, Piano, Violin, Crash 그리고 Saxophon을 가지고 모의 실험을 하였다. 이 오디오 신호 중, Drum에 대해서 스위칭한 결과를 그림(4-2), (4-3), (4-4)에 나타내었다. 세가지 기준 중, 일반적으로 에너지-엔트로피 기준과 영교차 기준이 좋은 성능을 가짐을 알 수 있다. 그림(4-2)의 에너지 엔트로피 기준은 에너지 엔트로피가 3.8bits 이하로 떨어지면 스위칭을 하도록 하였으며 그림(4-3)의 영교차 기준은 수

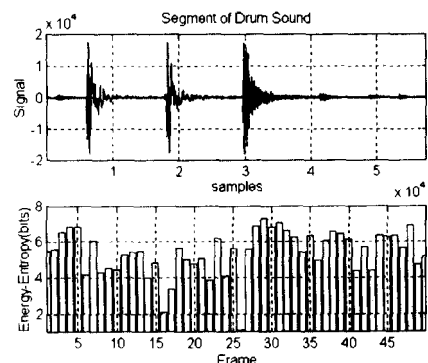


그림 4-2. 에너지-엔트로피 기준

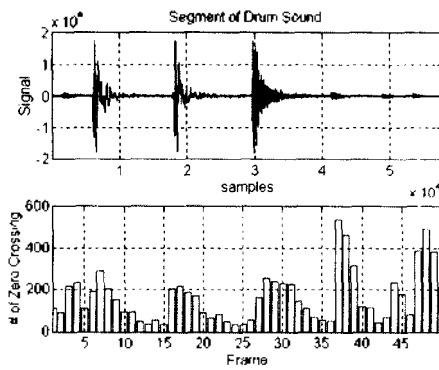


그림 4-3. 영교차 기준

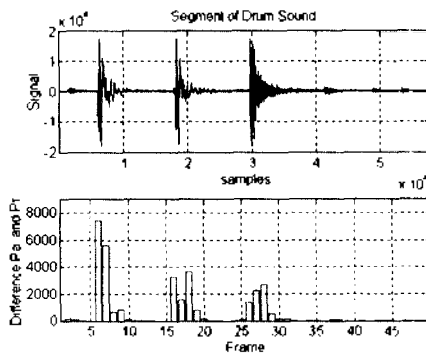


그림 4-4. 차분 기준

를 계상하여 120 이상인 경우에 스위칭하도록 하였다. 마지막으로 그림(4-4)의 차분 기준은 이전 프레임의 평균과 750이상이 차이가 나면 스위칭 하도록 하였다. 그 결과 에너지 엔트로피 기준과 제로 크로싱 기준이 좋은 결과를 얻었다. 이후의 모의 실험에서는 영교차 기준에 의하여 스위칭을 하였다.

4.3 필터뱅크의 설계

• 균일한 필터뱅크

그림(4-5)은 균일한 필터뱅크에서 사용된 필터 $h[n]$ 과 $C[n]$ 을 비교하여 그림으로 나타내었다. 부분적으로 최적화된 식(2-7)에서 사용된 $C[n]$ 은 $h[n]$ 을 64개의 계수들로 묶어서 매 홀수번째 묶음을 음수로 바꾸어준 것이다. 식(2-4)에서의 $H_1[n]$ 에 대한 방정식은 cosine함을 가지고 적절한 주파수 대역으로 저역통과필터의 응답을 변조한 것으로 볼 수 있다. 이러한 필터들은 $\pi/(64T)$ 의 홀수 곱에서 중심 주파수를 가지고 있다. 여기서, T는 오디오 표본화 주파수이고 각각 $\pi/(32T)$ 의 대역폭(nominal bandwidth)을 가지고 있다.

그림(4-6)에서 보는 것과 같이 원형 필터의 응답은 자신의 대역폭에서 날카로운 차단 주파수를 가지지 않는다. 결국, 32로 다운샘플링할 경우 상당히 많은 양의 aliasing이 발생할 수 있다. 그러나 적당한 cosine을 사용하여 원

형필터가 위상 천이(phase shift)를 가지도록 한다면 복호화의 합성 필터뱅크에서 완벽한 alias cancellation을 이룰 수 있다.

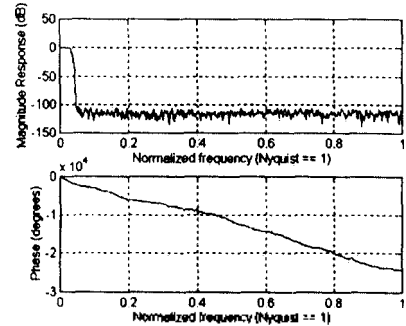


그림 4-6. 원형필터의 진폭, 위상 응답, $h[n]$

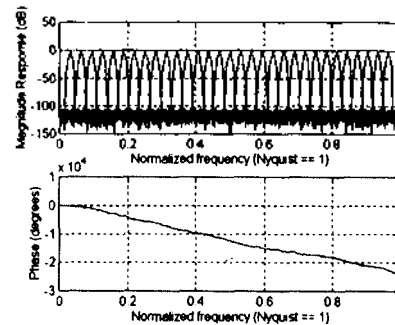


그림 4-7. 원형필터의 변조에 의한 대역통과필터

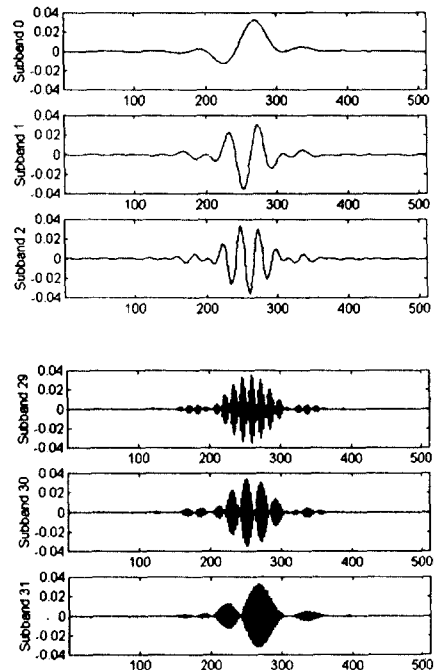


그림 4-8. 대역통과필터의 임펄스 응답(1, 2, 3, 30, 31, 32 대역)

그림(4-7)은 그림(4-6)의 원형필터의 변조를 이용한 필터뱅크의 응답을 보여 주고 있다. 그림(4-8)은 각 부대역 필터의 임펄스 응답을 보여주고 있다. 즉, 원형필터에 분석행렬을 곱하여 얻어진다.

• 비균일한 필터뱅크

그림(4-9)과 같이 비균일한 필터뱅크로서 웨이브렛 패킷 필터뱅크를 설계하였다. 사용된 필터는 6차와 20차 Daubechies 필터를 사용하여 모의 실험을 해보았고 이후의 모든 모의 실험은 20차 Daubechies 필터를 사용하였다. 본 연구는 시스템의 성능에 초점을 맞추었기 때문에 계산량에 대한 연구는 추후 과제로 남겨 놓았다. 웨이브렛 패킷 필터뱅크는 인간의 청각구조를 고려한 임계대역과 유사하게 설계하였다.

그림(4-11)에서 실제 오디오 신호(Drum)에 대해서 웨이브렛 패킷 변환을 한 후, 다시 역변환 한뒤의 파형을 보여주고 있다. (a)는 원신호이고, (b)는 복원된 신호이다. 그리고, (c)는 원신호와 복원된 신호와의 차, 에러를 보여주고 있다. (c)에서 보여진 것처럼 에러는 -1~1 사이에 존재하고 인간의 청각으로는 구별이 불가능하기 때문에 완전복원이 되었다고 가정한다. 이러한 에러는 컴퓨터의 유한어장(finite word length)에 의한 에러와 필터길이가 유한이기 때문인 것으로 사료된다. 이 결과로서 비균일한 필터인 웨이브렛 패킷 필터뱅크는 완전복원 조건을 만족함을 확인하였다.

4.4 양자화

본 연구에서는 균일한 양자화기를 가지고 모의 실험을 하였다. 오디오 부호화에서는 입력이 0일 때(silent), 출력도 없어야 하기 때문에 midtread 양자화기를 사용하는 것이 바람직하다. 또한, 양자화 과정에서 양자화 레벨의 전

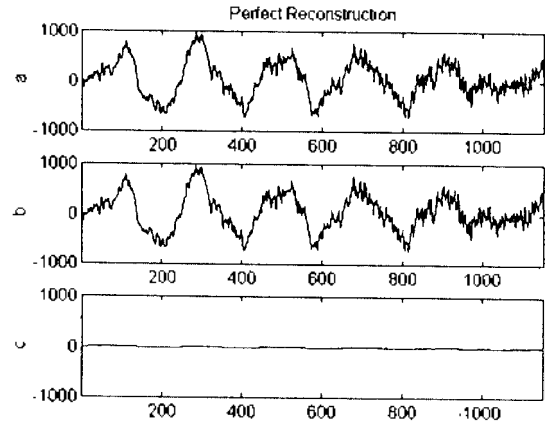


그림 4-10. 완전복원

체를 이용하기 위해서 스케일링(6-bits)을 하였다.

4.5 오디오 부호화기의 성능평가

본 연구의 모의 실험 결과는 주관적 평가(MOS: mean opinion score)를 통해서 이루어졌다. 일반적으로 심리음향 모델이 사용되어지면 MOS 테스트로서 음질을 평가하는 것이 일반적이다. 이유는, 심리음향 모델을 사용한 오디오 부호화기의 경우 인간의 청각이 청취할 수 없는 부분은 부호화 하지 않는다는 기본 개념을 포함하고 있기 때문이다. 그러나 객관적 평가인 SNR과 Segmental SNR은 이러한 부분까지도 잡음으로 간주하기 때문에 전체 오디오 부호화기의 성능이 떨어지는 것으로 판단하기 때문이다.

5종류의 오디오 신호에 대해서 실험을 하였고, 총 4가지의 전송율(192kbps, 128kbps, 64kbps, 32kbps: 단채널)에서 실험을 하였다. 주관적 평가 방식에 의해 부여받은 점

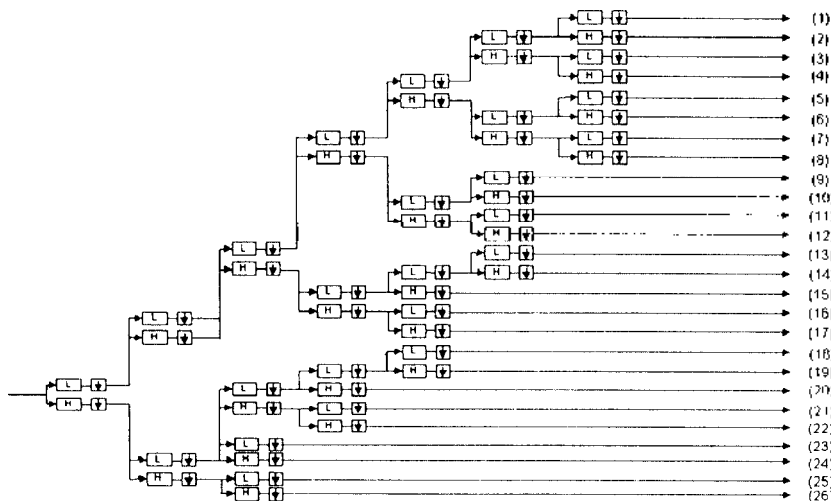


그림 4-9. 웨이브렛 패킷 필터뱅크

수의 통계가 표 1에 나와있다. 성능평가에 참가한 인원은 비전문 청취자 12명이었다. 192kbps와 128kbps에서는 균일한 필터뱅크만을 사용했을 경우와 비슷한 결과를 보였으나 저전송율, 즉 64kbps, 32kbps에서는 스위칭 필터뱅크가 더 향상된 음질을 가짐을 확인하였다. 64kbps 이상의 전송율에서 균일한 필터뱅크만을 사용한 부호화기보다 하이브리드 필터뱅크를 사용한 부호화기가 낮은 점수를 받았는데, 이유는 두 필터뱅크가 서로 다른 표현력을 가지기 때문으로 사표된다. 그림(4-11)에 주관적 평가에 대해서 부여 받은 점수와 평균을 그래프로 표시하였다.

표 1. 청취테스트 결과(단체별)

스위칭 필터뱅크	192Kbps	128Kbps	64Kbps	32Kbps
Drum	4.34	4.12	3.3	2.68
Piano	4.54	4.23	3.54	3.06
Violin	4.76	4.36	3.62	3.1
Crash	4.32	4.1	3.21	2.83
Saxophon	4.53	4.21	3.01	2.89
평균	4.498	4.204	3.336	2.912
편차	0.1344	0.0752	0.1952	0.1344
균일한 필터뱅크	192Kbps	128Kbps	64Kbps	32Kbps
Drum	4.53	4.79	3.06	2.03
Piano	4.4	4.43	3.21	2.11
Violin	4.87	4.64	3.42	2.32
Crash	4.45	4.32	3.5	1.76
Saxophon	4.65	4.28	3.44	2.01
평균	4.58	4.492	3.326	2.046
편차	0.144	0.1784	0.1528	0.1352

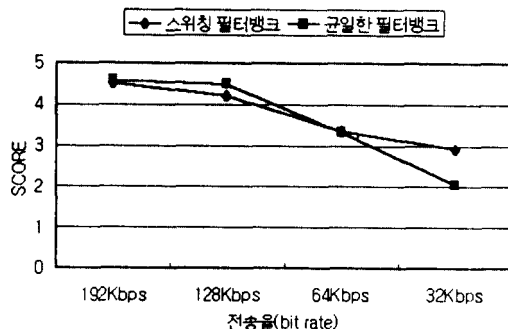


그림 4-11. MOS 테스트

그림(4-11)에서 사각형으로 표시된 선이 균일한 필터뱅크만을 이용했을 경우의 전송율과 부여받은 점수를 나타내고 세모로 표시된 선이 스위칭 필터뱅크로 부호화했을 경우의 점수를 나타낸다.

V. 결 론

본 논문은 기존의 폴리페이즈 필터뱅크를 이용한 MPEG 오디오가 64kbps이하에서 상당히 음질이 떨어지는 단점을 보완하기 위해서 입력신호의 특성에 따라 비균일한 필터뱅크인 웨이브렛 패킷 필터뱅크를 적용적으로 적용하였다. 스위칭 필터뱅크의 필요성은 에너지 Compaction의 기준이 되는 변환이득의 관점에서 바람직한 시도라고 여겨진다. 본 연구에서 사용된 오디오 부호화기와 MPEG 오디오의 전체 시스템과는 비교하지 않고 필터뱅크와 심리음향 모델만을 사용하였지만 충분한 비교와 개선의 여지가 있었다. 사용된 오디오 신호는 저음, 중음, 고음 그리고 급격한 변화를 가지는 신호를 선택하여 어느 음원에 편중되지 않도록 하였다. 그 결과, 스위칭 필터뱅크를 이용한 부호화기는 오디오 신호의 급격한 변화에도 적절히 대처하였다. MDCT보다 웨이브렛 변환이 계산량이 상당히 적기 때문에 스위칭 필터뱅크를 이용한 오디오 부호화기는 기존의 MPEG 오디오에 약간의 계산량만이 추가되어 상당한 효과를 얻을 것으로 여겨진다.

스위칭 기준으로는 3가지 즉, 에너지-엔트로피 기준, 영교차 기준, 그리고 차분 기준을 사용하였다. 그 중에서 본 연구에서 제안된 영교차 기준과 차분 기준중 영교차 기준이 더 적절히 스위칭 하였다.

향후 연구 과제로서 제시된 문제점은 본 연구에서 설계된 웨이브렛 패킷 필터뱅크의 동적비트할당은 MPEG 오디오의 심리음향 모델을 사용하였기 때문에 최적의 비트할당이 이루어지지 않았다. 그러므로 임계대역과 유사한 웨이브렛 패킷 필터뱅크에 대한 정확한 심리음향에 대한 연구가 이루어진다면 더 나은 성능을 유지할 것으로 기대된다. 또한, 양자화에 있어서 균일한 양자화기는 한계성을 가지고 있는 것으로 사료되며, 나아가 더 낮은 전송율에서 향상된 음질을 유지하기 위해서는 비균일한 양자화기와 벡터 양자화를 적용되어야 할 것이다. 이와 같은 문제점을 보완한다면 더 향상된 음질의 오디오 부호화기를 설계할 수 있을 것이다.

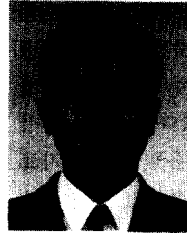
참 고 문 헌

1. Noll, P., "Wideband Speech and Audio Coding", IEEE Communication Magazine, Nov., 1993.
2. Shlien, S., "Guide to MPEG-1 Audio Standard", IEEE Transactions on Broadcasting, Vol., 40, No. 4, DEC. 1994.
3. 정계창 번역, 洋 著, 그림으로 보는 최신 MPEG, 교보문고, 1995.
4. Pan, D., "A Tutorial on MPEG/Audio Compression", IEEE Multimedia, 1995.
5. Riouland, M. O. Vetterli, "Wavelets and Signal Processing", IEEE SP Magazine, pp. 14-38, Oct. 1991.
6. Fliege, N.J., Multirate Digital Signal Processing, John Wiley & Son, 1994.

7. Sinha, D. and Tewfik A. H., "Low Bit Rate Transparent Audio Compression Using Adapted Wavelets", IEEE Trans. on Sign. Proc., Vol. 41, No. 12, Dec. 1993.
8. Jayant, N.S., P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, 1984.
9. Sayood, K., *Introduction to Data Compression*, Morgan Kaufmann Pub., 1996.
10. 安世瑛, 디지털오디오의 이론과 응용, 탐구원, 1996.
11. Noll, P., "MPEG Digital Audio Coding", IEEE Signal Processing Magazine, Sept., 1997.
12. RAO, K.R., HWANG, J.J., *Techniques and Standards for Image, Video, and Audio Coding*, Prentice Hall, 1996.
13. Zwicker, E., Fastl, H., *Psychoacoustics Facts and Models*, Springer-Verlag, 1990.
14. Strang, G. and Nguyens, T., *Wavelet and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA, 1996.
15. Chan, Y.T., *Wavelet Basics*, Kluwer Academic Publishers, 1995.
16. Ali, Murtaza, Khaled N. Hamdy and Ahmed H. Tewfik, "High Quality Audio Coding Using Adaptive Signal Representation", from Internet.
17. Vaidyanathan, P.P., *Multirate System and Filter Bank*, Prentice Hall, 1993.
18. Sinha, D. and Johnston, J., "Audio Compression at Low Bit Rates Using A Signal Adaptive Switched Filter Bank", Proc. ICASSP, pp. 1053-1056, Atlanta, GA, MAY 1996.
19. Khaled N. Hamdy, Murtaza Ali and Ahmed H. Tewfik, "Low Bit Rate High Quality Audio Coding With Combined Harmonic And Wavelet Representations", IEEE ICASSP, pp. 1045-1048, 1996.
20. 배석모, 김도형, 정재호, "웨이브렛 변환을 이용한 오디오 코딩", 한국음향학회지 제16권 4호, 1997.

▲신 유 철(Yoo-Cheol Shin)

1997년 1월 15일생



1996년 2월: 인천대학교 전자공학과 졸업(공학사)

1998년 2월: 인천대학교 대학원 전자공학과 졸업예정(공학석사)

※주관심분야: 신호처리, 신경망, 데이터 압축

▲강 현 철(Hyunchul Kang)

현재: 시립인천대학교 정보통신공학과 부교수
한국음향학회지 제14권 제5호 참조

※주관심분야: 영상처리, 패턴인식, 화상통신

▲변 윤 식(Youn-Shik Byun)

현재: 시립인천대학교 전자공학과 부교수
한국음향학회지 제16권 제3호 참조

※주관심분야: 신호처리 및 응용, 디지털 통신