DBpia

논문 99-24-11B-22 한국통신학회논문지 '99-11 Vol.24 No.11B

웨이블렛을 이용한 64 kb/s 이하의 저 전송률 오디오 신호 압축

정회원 박 관 우*, 정 재 호**

Transparent Low Bit Rate Below 64 kb/s Audio Compression Using Wavelets

Kwanwoo Park*, Jae Ho Chung** *Regular Members*

요 약

본 논문에서는 신호처리 분야에서 주목을 받고 있는 웨이블렛 변환과 이에 적합한 심리음향 모델을 설계함으로써 저 전송률에서 고품질의 오디오 부호화를 위한 부호화기/복호화기를 설계하였다.

웨이블렛 변환할 경우, 오디오 신호를 임계대역에 근사한 29개의 밴드로 이루어진 tree구조의 웨이블렛 패킷(wavelet packet)을 이용하여 분해하였으며, 심리음향모델에서 얻어진 정보를 효과적으로 적용할 수 있도록 오디오 신호를 분해하는 웨이블렛을 선정하여 사용하였다. 웨이블렛 계수를 양자화 할 때, 웨이블렛 변환된 신호의 대역별 특성을 이용하여 scale factor를 점차적으로 넓은 범위에서 찾고, scale factor를 양자화하기 위하여 매 frame마다 웨이블렛 패킷의 각 stage별 dynamic range를 구하여 sacle factor를 양자화 하였다. 실험에서는 제안된 codec의 성능 평가를 위해 blind test와 MPEG-1 Layer II와 비교하였다. 실험결과는 제안된 방법으로 채널 당 52-64(kb/s)에서 CD 수준의 양호한 오디오 신호 코딩을 할 수 있었으며, MPEG-1 Layer II와의 비교에서는 48 kb/s에서 월등히 우수함을 보였다.

ABSTRACT

In this paper, we propose a low bit rate, high fidelity audio codec using wavelet, which is well-matched to the psychoacoustic model. We decompose the audio signal into 29 non-uniform subbands approximated to the critical bands using the tree structured wavelet packet decomposition. When this transform is performed, we select a wavelet which can be best applied to the information obtained from the psychoacoustic model. In the quantization procedure of wavelet coefficients, we select the scale factors of each frame from gradually broader range as moving to the higher frequency region, and quantize these scale factors according to their pre-searched dynamic ranges to minimize the number of bits needed in expression of each audio frame under certain distortion level. Using the proposed method, we get the compact disc(CD) quality of audio at 52-64 kbps(kilo bits per second) per channel and in comparison with MPEG-1 Layer II, the coding quality of the proposed codec is greatly superior to that of MPEG-1 Layer II at 48 kbps.

I. 서 론

오디오 신호의 경우, 전송률(bit rate)을 낮추는 것은 오디오 시스템을 디자인하는데 있어서 중요하다. 멀티미디어 워크스테이션이나 고품질 오디오 신호의 전송 또는 저장과 같은 다양한 적용분야에서는 낮은 전송률에서 고 충실도의 오디오 신호 코딩은 중요한 문제이다. 기준이 되고 있는 CD(44.1 kHz 표본화율)의 경우 16 bit/sample의 높은 해상

* LG정보통신(kwanwoo@lgic.co.kr), ** 인하대학교 전자공학과(jhchung@inha.ac.kr)
논문번호 : 98534-1216, 접수일자 : 1998년도 12월 16일

www.dbpia.co.kr

도로 양자화 하므로 한 channel 당 705 kb/s의 전송률을 필요로 한다. 원거리 방송라인, 고충실 오디오 신호의 위성전송이나 멀티미디어 제품등과 같은 적용분야에서는 한정된 채널 용량을 효율적으로 사용하기 위해, 오디오 신호는 50~60 kb/s이하의 낮은 전송률에서도 성능이 우수한 오디오 부호화 알고리즘으로 압축되어야 한다.

오디오 신호 압축의 방법은 크게 두 가지로 나눌 수 있다. 하나는 오디오 부호화에서 주목을 받고 있는, 인간 청각계의 인지 특성인 매스킹(masking)현상을 이용하여 지각적으로(perceptually) 투명(transparent)하게 부호화 하는 것이고, 다른 하나는 오디오 신호자체의 통계적인 중복성을 제거하여 전송률을 줄이는 것이다. 그러나 매우 낮은 전송률(64 kb/s이하)에서 투명한 오디오 부호화를 위해서는 인간 청각계의 인지 특성을 이용한 지각 부호화와 신호 자체의 통계적 중복성을 제거하는 것이 모두 고려되어져야 한다.

오디오 신호 압축은 MPEG을 주축으로 지난 10여년 동안 눈부시게 발전하였다. 이것은 오디오 신호압축의 가장 핵심적인 인간 청각계의 지각특성을 이용하여 잡음을 은폐시키는 심리음향모델에 기인한다. MPEG의 경우 MPEG-1,-2의 Layer I과 II에서는 주로 MUSICAM의 개발 과정에서 나온 심리음향 모델 1을 사용하였으며, Layer III에서는 ATT사에서 PAC(Perceptual Audio Coder)의 개발 과정에서 만들어진 심리음향 모델 2를 사용하였으며, 후에 심리음향 모델 2는 MPEG-2 AAC와 MPEG-4에서도 사용되었다[1][2][3][4].

심리음향의 매스킹 현상은 인간 청각계의 대역별 특성에 따른 임계대역 내에 있는 신호의 순음(tonal) 성분과 잡음(noiselike) 성분에 따라 달라진다. 이와 같은 이유로 심리음향 모델의 해석은 주로 각 임계대역 내에서 이루어진다. 그러나 MPEG의 경우 MDCT(Modified Discrete Cosine Transform)를 이용하여 입력 오디오 신호를 32개의 일정한 대역으로 서브밴드화 함으로써 임계대역과 불일치하는 서브밴드를 갖게 된다. 심리음향에서 임계대역 별로 얻어진 정보를 임계대역과 불일치하는 불완전한 서브밴드에 bit할당 정보로 사용하기 위하여 MPEG에서는 어쩔 수 없이 심리음향에서 bit할당 정보(SMR값)를 얻는 과정 즉, 각 서브밴드에 해당하는 임계대역의 SPL(Sound Pressure Level) 중에 가장 큰 값과 매스킹 레벨 중에 가장 작은 값을 사용하여 worst case를 적용하였다. 이것은 부정확한 bit 할당 정보에 의한 많은 bit의 손실을 의미하며 전송률이 낮아질수록 심각한 음의 왜곡을 야기한다. 또한 MPEG에서는 낮은 전송률에서 안정적인 코딩을 하기 위하여, 각 밴드에 bit를 할당하는 과정에서 고주파대역을 bit할당 대상에서 제외하는 band limiting 기법을 사용하였으나 이것은 정확한 심리음향에 의한 bit할당보다 낮은 전송률에서는 성능이 현저히 떨어지는 원인이 되고 있다.

본 논문에서는 이와 같은 임계대역과 불일치하는 대역분할에 의한 문제점을 해결하기 위하여 tree 구조의 웨이블렛 패킷을 이용하여 인간 청각계의 임계대역과 일치하는 29개의 대역으로 신호를 분해하였다. 또한, 웨이블렛 변환된 신호에 적합하고 각각의 임계대역에 일치하는 대역에 대한 bit할당 정보를 얻기 위하여 MPEG의 심리음향 모델1을 개선/변형하였다. Bit할당 정보를 얻어내는 과정에서는 개선된 심리음향 모델에서 얻어진 각 임계대역의 정확한 SMR값으로부터 정확한 MNR(SNR/SMR)값을 산출함으로써 웨이블렛 변환된 각 서브밴드 샘플들을 양자화할 때, 정확하게 산출된 bit할당 정보에 의하여 양자화 하였다.

웨이블렛 변환할 때 웨이블렛의 다양성 때문에 주어진 신호와 심리음향 모델에서 해석된 정보를 잘 반영할 수 있는 웨이블렛을 선정하는 문제는 중요하다. 웨이블렛의 선정에서는 최대 vanishing moment수를 갖는 웨이블렛이 가장 적합하다는 결론을 내렸다.

낮은 전송률에서 투명한 오디오 부호화를 위해서는 잘 설계된 양자화기는 필수적이다. Tree 구조의 웨이블렛 패킷 분해된 신호는 각 stage별로 다른 성질을 갖는 것과 일반적인 오디오 신호의 에너지가 저주파 대역에 몰려있는 것을 감안하여 양자화기를 설계하였다. 임계대역에 따라 웨이블렛 변환된 서브밴드가 비등간격이므로 scale factor를 찾는 구간도 비등간격으로 구하였으며, 각 stage별로 다른 dynamic range를 가지므로 매 프레임(frame)마다 scale factor의 stage별 dynamic range를 구하여 scale factor를 양자화 하였다. 한 프레임은 약 23msec에 해당하는 1024샘플을 사용하였다.

제안된 codec의 성능을 평가하기 위하여 MPEG-1 Layer II와 비교하였다. 제안된 codec은 객관적인 평가에서뿐만 아니라 주관적인 청취 평가에서도 우수한 성능을 얻을 수 있었다.

본 논문의 전체적인 구성은 다음과 같다. II 장에서는 웨이블렛의 선정과 웨이블렛 패킷 변환에 대

www.dbpia.co.kr

하여 제안하고, III 장에서는 웨이블렛 변환된 샘플에 적용될 심리음향 모델에 대하여 제안하며, IV 장에서는 양자화기에 대하여 언급하였으며, V 장에서는 실험 및 주/객관적 음질 평가에 대하여 언급한다. 마지막으로 VI장에서 결론을 맺는다.

Ⅱ . 웨이블렛

2.1 웨이블렛 패킷 (Wavelet packet)

인간의 청각계는 각 주파수 대역에 따라 그 반응 특성이 달라진다. 이것을 임계대역이라 하며 양자화 잡음은 임계대역 내에서 청각계에 의해 더해진다. 그러므로 같은 임계대역 내에 있는 잡음 성분들은 같은 가중치를 가지며 더해지게 된다[5][6]. 본 연구에서 사용한 임계대역은 표 1에 나타내었다. 임계대역 내의 이러한 성질을 이용하여 심리음향모델에서 얻어진 매스킹 정보를 각 밴드의 bit 할당에 이용한다.

표 1. Tree 구조의 웨이블렛 패킷을 이용하여 분해된 29 대역

밴드넘버	아래쪽 경계 (Hz)	중심 주파수 (Hz)	위쪽 경계 (Hz)
1	0	45	90
2	90	130	170
3	170	215	260
4	260	300	340
5	340	385	430
6	430	475	520
7	520	605	690
8	690	775	860
9	860	945	1030
10	1030	1115	1200
11	1200	1300	1400
12	1400	1550	1700
13	1700	1900	2100
14	2100	2250	2400
15	2400	2600	2800
16	2800	2950	3100
17	3100	3250	3400
18	3400	3750	4100
19	4100	4450	4800
20	4800	5150	5500
21	5500	5850	6200
22	6200	6550	6900
23	6900	7600	8300
24	8300	8950	9600
25	9600	10300	11000
26	11000	12400	13800
27	13800	15150	16500
28	16500	17900	19300
29	19300	20900	22500

Bit를 할당 시 각 임계대역내의 인지되는 잡음을 일정수준 이하로 유지하도록 반영하기 위하여, 각 임계대역의 MNR값에 따라 bit를 할당하게 된다. 이와 같은 이유로 입력 오디오 신호는 임계대역과 일치하는 밴드로 분해되어야 한다. 본 연구에서는 29 밴드의 트리 구조의 웨이블렛 패킷을 이용하여 입력 오디오 신호를 분해하였다[7][8]. 그림 1에 분해된 구조를 나타내었다. 그림에서 위로 갈수록 저주파 대역이며 아래로 갈수록 고주파 대역이다. 그림 1의 각 노드에서의 웨이블렛 필터링에 의한 다음 단 노드로의 진행은 그림 2에 나타낸 것과 같다. 간단한 두개 채널의 웨이블렛 필터뱅크를 보여준다. $h_0[n]$과 $h_1[n]$은 분석 필터이고, $g_0[n]$과 $g_1[n]$은 합성 필터를 나타낸다.

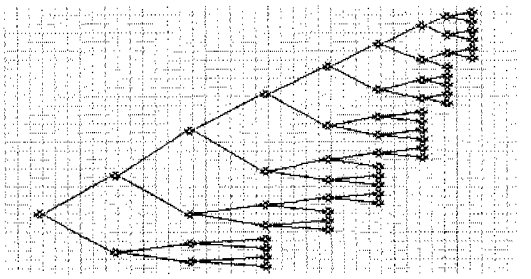
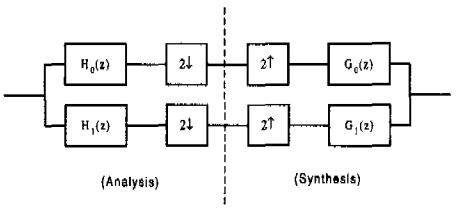
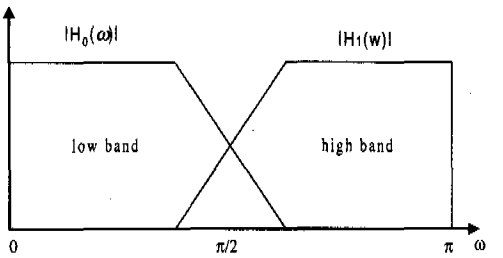


그림 1. 29밴드의 웨이블렛 패킷으로 분해된 트리 구조



(a) Block Diagram



(b) Spectrum Splitting Performed by the Filter Bank

그림 2. 2-채널 필터 뱅크

그림 2(a)는 신호를 분석하고, 합성하는 과정을 보여주고, 그림 2(b)는 분석 필터에 의하여 분할된

www.dbpia.co.kr

대역을 보여준다. 그림 2(a)에서 입력된 신호는 저대역 통과 필터($h_0[n]$)와 고대역 통과 필터($h_1[n]$)를 통하여 분석되고, down-sampling된다. 다시 합성부에서는 down-sampling된 샘플을 up- sampling하고, 합성 필터를 통해서 원래의 신호를 복원하게 된다[9][10][11][12][13].

2.2 웨이블렛의 선택

오디오 신호를 웨이블렛 변환하는 것은 심리음향 모델에서 얻어진 매스킹 정보를 정확하게 적용하기 위함이다. 매스킹 현상은 바크 주파수 영역에서의 오디오 신호 특성이 지배적으로 작용한다. 이러한 성질은 웨이블렛 변환된 계수들이 바크 주파수 대역(임계대역)에 충실하게 대역분할 되어야 하는 것을 의미하며, 웨이블렛 변환 시 충분한 레귤레러티(regularity)를 보장하여야 한다. 본 논문에서는, 충분한 레귤레러티를 유지하기 위하여 최대의 vanishing moment수를 갖는 웨이블렛을 이용하였다. 최대의 vanishing moment수를 가질 때 웨이블렛 필터는 통과대역에서 거의 일정한 값을 가지며 천이대역은 좁아져 좋은 저대역 통과 필터처럼 작용하게 된다[10][14]. 그림 3에 vanishing moment에 따른 웨이블렛 필터의 주파수응답을 나타내었다. 또한 웨이블렛 필터의 주파수 분해능은 그림 4에서 보는 바와 같이 필터의 길이가 증가할수록 증가한다. 이러한 성질 때문에 vanishing moment의 수가 많고 필터의 길이가 길수록 코딩성능이 향상되는 것을 보인다. 그러나 필터의 길이가 증가할수록 연산량과 시스템 지연이 길어지므로 필터의 길이는 적정수준에서 결정되어야 한다. 본 연구에서는 필터 길이가 60을 넘어서면서부터 코딩성능 향상이 급격히 포화되는 현상을 보였다.

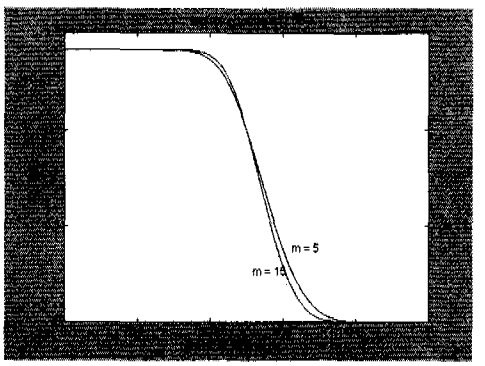


그림 3. 필터 길이가 30이고, vanishing moment의 개수가 5, 15인 경우 각각의 크기응답

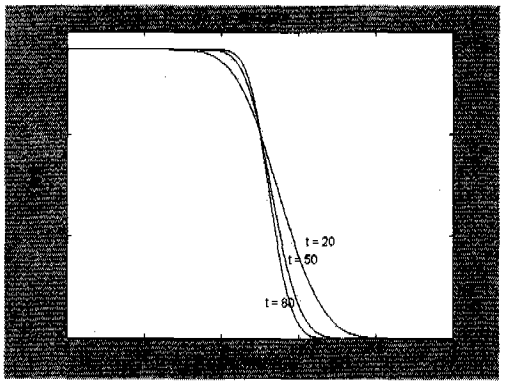


그림 4. Maximum vanishing moment를 갖고, 필터길이가 20, 50, 80일 때의 크기응답

Ⅲ. 심리음향모델에서의 SMR(signal to masking ratio) 산출

심리음향을 이용한 오디오 신호 압축은 오디오 신호로부터 음향학적으로 무관한 부분을 제거함으로써 압축을 시도한다. 즉, 인간의 청각계가 청각 매스킹의 조건하에서 양자화 잡음을 감지할 수 없음을 이용한다. 이 매스킹은 인간 청각계의 지각 특성이며, 강한 오디오 신호가 시간적 또는 주파수적으로 근처의 약한 오디오 신호를 감지할 수 없게 할 때 발생한다[15].

인간의 청각계는 제한된 주파수 의존적 해상도(frequency dependent resolution)를 가지고 있다. 이러한 의존(dependency)은 제 II장에서 제시한 표 1에 주어진 임계대역 폭의 관점에서 표현될 수 있는데, 임계대역의 폭은 저주파에서는 100Hz, 고주파에서는 4kHz로, 낮은 주파수에서 높은 주파수로 갈수록 넓어진다. 인간의 청각계는 비록 주파수 선택도가 한 개의 임계대역 보다도 훨씬 해상도가 좋더라도 한 개의 임계 대역 내의 다양한 신호 성분들을 완벽하게 감지하지 못한다. 이러한 인간 청각계의 주파수 의존적 해상력(frequency-dependent resolving power) 때문에, 주어진 임의의 주파수에서의 잡음 매스킹 임계값은 그림 5와 같이 오직 해당 주파수 근처의 제한된 대역폭 내에 있는 신호 에너지에만 의존한다. 오디오 신호 압축은 이러한 인간 청각계의 매스킹 현상을 이용하여 이루어진다. 즉, 2장에서 제시한 29밴드의 웨이블렛 서브밴드 내의 양자화 잡음의 가청성(audibility)에 따라서 bit 할당을 한다. 그러므로, 가장 효과적인 오디오 압축을 위해서는 매스킹 현상을 이용하여 각 band의 양

www.dbpia.co.kr

자화 잡음이 감지되지 않는 범위의 레벨 내에서 양자화 되어야 한다.

본 연구에서는 MPEG에서 사용한 심리음향모델-1을 웨이블렛 변환된 샘플들에 효과적으로 적용될 수 있도록 개선/변형하여 새로운 모델을 사용함으로써 청각계에 의하여 감지되는 양자화 잡음을 최소화하였다.

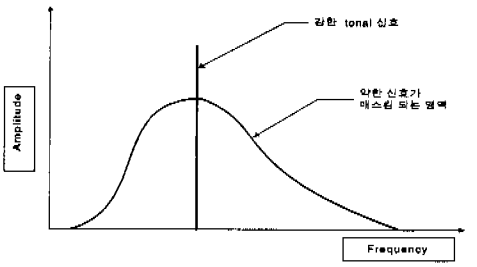


그림 5. 오디오 신호의 마스킹 현상

3.1 심리음향 모델의 전체 구조

본 연구에서는 MPEG에서 사용한 심리음향 모델 1을 기반으로 하여 심리음향 모델을 구현하였다. 그림 6에서는 심리음향 모델의 전체적인 블록 다이아그램을 나타내었다. 매 프레임(frame)의 오디오 입력 신호에 대하여 Hanning window를 취한 후에 1024-point FFT(Fast Fourier Transform)를 수행한다. 웨이블렛으로 변환된 영역에서 직접 심리음향 해석을 하지 않은 것은 심리음향에서 정확한 마스킹 한계(masking threshold)를 계산해 내기 위해서는 더 정밀한 주파수 해상도를 요구하기 때문이다. Fourier Transform을 사용하여 변환한 후에 각 spectrum line을 Bark 주파수로 mapping한다. Mapping된 각 임계대역에 해당하는 spectrum line들로부터 SPL(Sound Pressure Level)을 구한다. 이때 심리음향에서 구하여진 마스킹 정보를 서브밴드의 bit 할당 정보로 사용하는 과정에서 MPEG에서는 임계대역과 서브밴드의 불일치로 인하여 worst case를 고려한 부정확한 SPL을 구하게 되지만, 임계대역에 근사하게 웨이블렛 변환된 서브밴드에서는 정확하게 SPL을 구하게 된다. 또한 심리음향 모델에서 얻어진 SPL값을 바로 사용하는 것보다 웨이블렛 변환된 신호에 안정적인 bit할당을 위하여 weighting factor를 곱해줌으로써 프레임간의 불연속성에 기인하는 pre-echo현상을 상당부분 제거할 수 있다. 자세한 내용은 다음 절에서 설명한다.

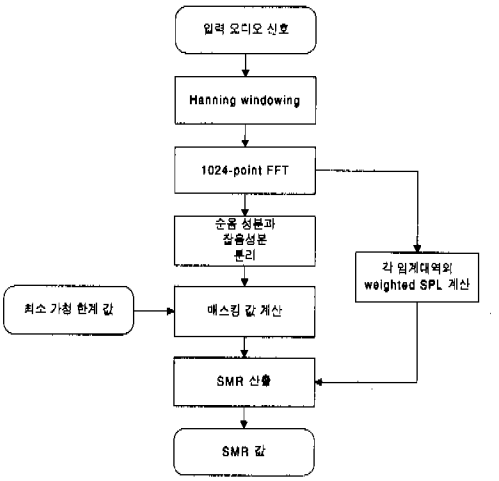


그림 6. 심리음향 모델의 블록 다이아그램

FFT된 오디오 신호의 spectrum line에서 tonal 성분과 non-tonal 성분을 가려내는데 그 이유는 두 가지 신호 형태의 매스킹 능력이 서로 다르기 때문이다. Tonal 성분을 결정하는 기준은 오디오의 power spectrum의 local peaks에 기초하여 결정한다. Tone 성분에 대한 처리를 한 후에, 남아 있는 spectral value를 하나의 임계대역에 하나의 non-tonal 성분으로 합산하여 집중시킨다. Non-tonal 성분으로 집중된 각각의 성분들의 주파수 index는 감싸고 있는 임계대역의 기하평균에 가장 가까운 값으로 정해진다. 다음으로 tonal 성분과 non-tonal 성분에 의한 매스킹 값을 구하게 된다. 구하는 식은 다음과 같다[16].

$$LT_{tm}[z(j),z(i)] = X_{tm}[z(j)] + av_{tm}[z(j)] + vf[z(j),z(i)] \text{ dB} \quad (1)$$

$$LT_{nm}[z(j),z(i)] = X_{nm}[z(j)] + av_{nm}[z(j)] + vf[z(j),z(i)] \text{ dB} \quad (2)$$

이때 j는 마스커(masker)의 인덱스이며 i는 마스키의 인덱스이다. $z(j)$는 인덱스 j를 바크 주파수로 환산한 값이다. $X_{tm}[z(j)]$는 인덱스 넘버 j를 갖는 바크 주파수 $z(j)$에서의 SPL이다. 매스킹 인덱스에 대한 식은 다음과 같이 주어진다.

$$av_{tm} = -1.525 - 0.275 * z(j) - 4.5 \text{ dB} \quad (3)$$

$$av_{nm} = -1.525 - 0.175 * z(j) - 0.5 \text{ dB} \quad (4)$$

www.dbpia.co.kr

또한 매스킹 함수는 마스커와 마스키(maskee)의 위치에 따라 다음과 같이 주어진다.

$$vf = 17*(dz+1) - (0.4*X[z(j)]+6)\ \text{dB} \quad \text{for } -3 \le dz < -1 \quad \text{Bark} \tag{5}$$

$$vf = (0.4*X[z(j)]+6)*dz\ \text{dB} \quad \text{for } -1 \le dz < 0 \quad \text{Bark} \tag{6}$$

$$vf = -17*dz\ \text{dB for } 0 \le dz < 1 \quad \text{Bark} \tag{7}$$

$$vf = (dz-1)*(17-0.15*X[z(j)]) - 17\ \text{dB} \quad \text{for } 1 \le dz < 8 \quad \text{Bark} \tag{8}$$

$$dz = z(i) - z(j) \tag{9}$$

식 (1)과 (2)에 의하여 구하여진 매스킹 한계값에 가청한계를 더하여 최종적인 매스킹 한계값이 얻어진다. MPEG에서는 매스킹 한계를 구하는 과정에서도 서브밴드와의 불일치로 인한 오차를 보상하기 위하여 최소 매스킹 한계를 선택함으로써 부정확한 해석을 하게 되나 웨이블렛을 이용한 본 연구에서는 정확한 매스킹 한계를 산출하였다.

앞서 구한 WSPL(Weighted Sound Pressure Level)과 매스킹 한계로부터 최종적인 SMR을 구하게 된다.

3.2 Pre-echo현상을 제거하기 위한 WSPL의 산출

본 논문에서 제안한 코덱은 프레임 단위로 입력 오디오 신호를 처리하며, 각 프레임은 1024샘플로 이루어져 있다. 이것은 44.1kHz의 표본화율에서 23msec에 해당한다. 연속하는 각 프레임의 경계에서는 overlap and add method를 사용하였다. Overlap되는 샘플 수는 128sample을 Hanning window의 제곱을 곱하여 overlap시켰다. 이때 시간 영역에서 변화가 적은 신호일 때는 overlap에 의한 pre-echo현상이 temporal masking효과에 의하여 들리지 않으나 변이가 심한 즉, nonstationary한 오디오 신호인 경우에는 각 프레임의 경계지점에서 블록화 현상이 일어나 pre-echo현상을 야기한다. 이것은 급격히 변하는 오디오 신호에 대하여 심리음향에서 해석된 bit할당 정보가 프레임 내에서는 잡음이 들리지 않도록 bit 할당이 되었다 하여도 상대적으로 에너지가 큰 저주파 대역의 신호가 인접한 프레임에서 다르게 해석될 경우 즉, 인접한 프레임에서 상이하게 bit 할당이 되었을 때 프레임간의 양자화된 신호의 에너지 차에 의하여 블록의 경계지점에서 블록현상이 일어나 pre-echo현상을 야기하게 된다. 이러한 현상을 제거하기 위하여 프레임 크기를 조절하여 sudden burst가 존재하는 프레임 즉, 천이가 심한 연속한 프레임에서 sudden burst가 프레임의 앞쪽에 위치하도록 함으로써 temporal masking효과를 이용하는 방법이 있으나 pre-echo현상을 만족할 만큼 제거하기는 힘들다. 본 논문에서 제안된 코덱에서는 이러한 블록간의 변이가 심한 경계에서 생기는 pre-echo현상을 제거하기 위하여 심리음향모델에서 인접한 프레임과의 에너지 차가 일정수준 이상인 경우, SPL을 구하는 과정에서 에너지가 많은 저주파 대역에 대하여 가중치를 줌으로써 이러한 현상을 제거하였다. 이와 같이 weighted SPL을 사용함으로써 변이가 심한 부분에서 인접한 프레임에서 해석되는 심리음향모델에 의한 bit 할당정보의 상이함에 기인하는 pre-echo현상을 보상하여 이전 프레임과의 연관성을 갖게 하는 효과가 있다. 천이구간에서의 WSPL에 의한 bit할당은 일반적인 오디오 신호가 저주파 대역에 많은 에너지가 존재하며, 에너지가 급격히 변화하는 부분에서만 이루어지므로 전체적인 오디오 신호에서는 매우 짧은 시간이며, 또한 매우 적은 bit를 사용하여 심리음향에서 인접한 프레임간의 경계에서 급격히 변하는 오디오 신호에 대해 잘못 해석된 bit 할당 정보를 바로잡아 pre-echo현상을 제거하는 장점이 있다.

Ⅳ. 양자화

앞서 언급한 바와 같이 낮은 전송률(64kb/s이하)에서 효과적인 오디오 부호화를 위해서는 인간 청각계의 인지 특성을 이용한 지각 부호화와 신호 자체의 통계적 중복성을 제거하는 것이 모두 적용되어야 한다.

이번 장에서는 심리음향 모델에서 얻어진 SMR값을 이용하여 bit를 할당하는 과정과 이렇게 얻어진 bit할당 정보로부터 웨이블렛 변환된 샘플을 양자화하는 방법에 대하여 설명하고, 오디오 신호의 넓은 다이나믹 레인지(dynamic range)와 중복성(redundancy)을 줄이기 위하여 오디오 신호 자체의 특성과 웨이블렛 변환된 샘플들의 특성을 이용하는 기법에

www.dbpia.co.kr

대하여 설명한다.

4.1 웨이블렛 서브밴드에 대한 bit 할당

Bit 할당 과정은 심리음향 모델로부터 얻은 매스킹 정보인 SMR 값에 기초하여 각 웨이블렛 서브밴드에 할당될 bit의 수를 결정한다. Bit 할당 과정은 매스킹 대 잡음비(MNR)의 계산으로부터 시작한다.

$$MNR_{dB} = SNR_{dB} - SMR_{dB} \tag{10}$$

$$SMR_{dB} = WSPL_{dB} - LTMIN_{dB} \tag{11}$$

식 (10)에서 MNR_{dB}는 매스크 대 잡음비, SNR_{dB}는 신호 대 잡음비, 그리고 SMR_{dB}는 신호 대 매스크 비이다. SMR_{dB}는 식 (11)과 같이 구해지며, $WSPL_{dB}$는 각 웨이블렛 서브밴드의 $WSPL$의 데시벨 값이고 $LTMIN_{dB}$는 웨이블렛 서브밴드 내의 식 (1)과 식 (2)와 같이 구한 매스킹값과 가청한계값을 더한 값들 중에 가장 작은 값이다. 각각의 웨이블렛 서브밴드에 대하여 식 (10)과 같이 구한 MNR값이 가장 작은 웨이블렛 서브밴드에 bit를 할당하며, SNR 값은 0부터 시작하여 각 웨이블렛 서브밴드에 bit가 할당될 때마다 7dB씩 증가시킨다. 이와 같은 방법으로 더 이상 할당할 bit가 없을 때까지 가장 작은 MNR값을 가지는 웨이블렛 서브밴드에 bit를 할당하게 된다.

4.2 양자화기의 설계

적절한 양자화기의 설계는 오디오 신호 압축뿐만 아니라 모든 신호 압축에서 매우 중요한 부분이다. 본 연구에서는 오디오 신호의 특성과 웨이블렛 변환된 웨이블렛 계수들의 특성을 이용하여 오디오 신호의 넓은 다이나믹 레인지와 신호의 중복성을 줄일 수 있는 양자화기를 설계하였다.

4.2.1 각 프레임의 스테이지별 다이나믹 레인지의 결정

오디오 신호는 다른 신호에 비하여 넓은 다이나믹 레인지를 갖는다. 이러한 특징 때문에 오디오 신호를 부호화 하는데 상대적으로 많은 bit를 필요로 하게 된다. 이러한 오디오 신호를 웨이블렛 변환하는 과정에서 정규화(normalization)하여 변환한다. 웨이블렛 변환은 비균일(non-uniform)하게, 즉, 비균일한 대역폭과 웨이블렛 변환 스테이지(stage)를 가지기 때문에 각 대역의 웨이블렛 계수의 수와 다이나믹 레인지도 다르다. 이러한 성질 때문에 스케일 인자(scale factor)를 양자화하기 위한 각 밴드별 다이나믹 레인지가 결정되어야 한다.

본 연구에서는 스테이지가 다른 각 웨이블렛 서브밴드에 녹음실에서 녹음할 수 있는 최대 레벨의 순음을 인가하여 각 스테이지에서 가질 수 있는 최대값을 결정하였다. 표 2는 웨이블렛 계수들의 각 스테이지별 최대값을 보인다. 표 2에서 보는 바와 같이 스케일 인자를 양자화하기 위한 다이나믹 레인지가 저주파 대역인 8번째와 7번째 스테이지와 같은 경우 고주파 대역에 비하여 상당히 큰 것을 볼 수 있다. 오디오 신호 자체가 상당히 역동적인 성질을 가지고 있기 때문에 표 2의 값들을 바로 스케일 인자의 다이나믹 레인지로 사용하는 것보다 매 프레임마다 웨이블렛 변환된 오디오 신호의 대역별 다이나믹 레인지를 찾는 것이 효과적이다. 본 연구에서는 표 2와 같이 얻어진 웨이블렛 계수의 스테이지별 최대값으로부터 매 프레임마다 step size 2dB를 갖는 logarithmic uniform quantizer를 이용하여, 처리되고 있는 프레임의 스케일 인자의 다이나믹 레인지를 결정하였다. 매 프레임의 스케일 인자의 양자화를 위한 다이나믹 레인지를 전송하는데 6개의 스테이지에 대하여 6bit씩 모두 32bit를 사용하였다.

표 2. 각 웨이블렛 변환 스테이지의 최대값

웨이블렛의 스테이지	최대값
3^{rd} 스테이지	2.8
4^{th} 스테이지	3.4
5^{th} 스테이지	4.7
6^{th} 스테이지	6.6
7^{th} 스테이지	9.6
8^{th} 스테이지	15.5

4.2.2 스케일 인자의 선택

웨이블렛 계수들을 양자화하기 위하여 매 L샘플마다 변환된 계수들의 스케일 인자를 전송해 주어야 한다. 이때 L값을 어떻게 결정해야 하는가는 매우 중요하면서도 민감한 부분이다. L값을 너무 작게 할 경우 샘플들을 양자화 하기 위한 부가 정보가 증가하여 전송률이 증가하며, 너무 크게 하였을 경우 양자화기의 성능을 떨어뜨려 이를 보상하기

www.dbpia.co.kr

위한 전송률의 증가를 가져올 것이다. 본 연구에서는, 스케일 인자를 양자화 하기 위하여 6bit씩 할당하였다. 일정한 구간에서 스케일 인자를 전송하는 경우, 8샘플마다 전송하는 것이 가장 좋은 성능을 나타내는 것을 보였다. 그러나, 8샘플마다 스케일 인자를 전송하는 경우 한 frame에 128개의 스케일 인자를 전송해야 하며, 이때 사용되는 부가정보의 양은 CD와 같은 44.1kHz의 표본화 율에서 33kb/s (128 × 6bits × 44100 sample per second / 1024 sample)로 많은 부가정보가 필요하게 된다. 이러한 문제를 해결하기 위하여 본 연구에서는 최소 8샘플마다 스케일 인자를 전송하며 최대 128샘플마다 스케일 인자를 전송하였으며, 8샘플마다 일정하게 스케일 인자를 전송한 것과 비교하였을 때 성능의 차이는 거의 없었다. 이때 스케일 인자를 전송할 간격을 결정하는 방법으로는 웨이블렛 변환된 신호의 밴드별 신호특성에 따라 저주파 대역에서 고주파 대역으로 갈수록 넓은 간격마다 전송하는 것이 스케일 인자를 양자화 하는데 필요한 bit수를 줄일 수 있다. 이러한 방법으로 모든 밴드에 bit가 할당될 경우 전송되는 스케일 인자의 수는 한 frame(1024 sample)당 26개가 되며, 정보량은 6.7kb/s로 일정간격(8 sample)마다 스케일 인자를 전송하였을 때 보다 약 26kb/s의 전송률을 줄일 수 있다.

웨이블렛 변환된 샘플들을 양자화하기 위한 양자화기는 동 간격 step size를 갖는 midtread 양자화기를 사용하였다. 사용한 양자화기의 특성으로 인하여 성능은 향상시키며 반면에 코덱의 구조는 매우 간단하게 유지할 수 있었다.

웨이블렛을 이용한 오디오 코덱의 전체 구조는 그림 7, 8과 같다. 그림 7은 임계대역에 근사한 웨이블렛을 이용한 오디오 부호화기를 나타내며 그림 8은 복호화기를 나타낸다. 또한 그림 9는 부호화된 오디오 신호의 bit stream 구조를 보이고 있다. 그림 9에서 보는 바와 같이 부가정보에는 스케일 인자의 다이나믹 레인지, 즉, 프레임내의 각 스테이지의 최대값을 양자화하기 위하여 36 bits가 사용되며, 스케일 인자를 양자화하기 위해서는 스케일 인자가 속한 밴드에 bit가 할당되지 않았을 때는 스케일 인자도 전송하지 않으므로 모든 밴드에 bit가 할당되었을 경우 최대 156 bits가 할당되고, 각 웨이블렛 서브밴드의 bit 할당 정보를 전송하기 위하여 각 밴드에 4 bits씩 29밴드에 대하여 116 bits가 할당된다. 부가 정보에 사용되는 총 bit 수는 최대 304 bits가 된다. 웨이블렛 계수를 양자화하기 위한 bit 수는 전체 사용가능 bit 수에서 부가정보로 사용된 bit 수를 제외함으로써 얻어진다.

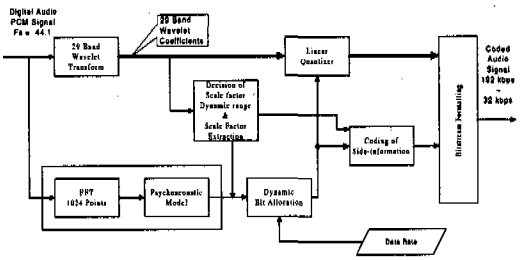


그림 7. 웨이블렛 오디오 부호화기의 구조

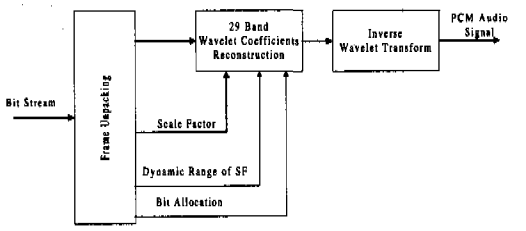


그림 8. 웨이블렛 오디오 복호화기의 구조

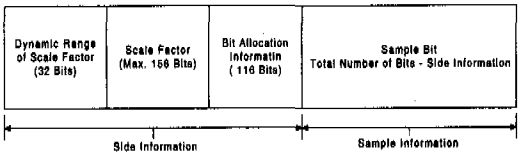


그림 9. Bit stream의 구조

V . 실험결과

본 연구에서 제안한 오디오 코덱의 성능 평가에 사용한 실험 샘플은 오디오 신호의 다양성과 다이나믹한 성질을 반영하기 위하여, 표 3에 있는 다양한 종류의 실험 샘플을 사용하였으며 연주 시간은 60초이다. 주로 ALANPARSONS와 STEPHEN COURT가 제작한 CD인 SOUND CHECK와 STEREOPLAY/AUTOHIFI사에서 제작한 HIFI CHECK를 이용하였다.

원음과 비교해 코딩된 음의 누화가 인지되는지 안 되는지를 판별하기 위하여 블라인드 테스트(blind test)를 실시하였다. 이 실험은 원음을 들려준 후에 코딩된 음과 원음을 순서 없이 들려준다. 이때 청취자는 어느 것이 코딩된 것인지 모르는 상태에서 각각에 10점 만점으로 점수를 주도록 하였다. 위와 같은 실험방법을 선택한 이유는 청취자의 선입

www.dbpia.co.kr

견과 서로 다른 판단기준을 배제하고 코딩된 음과 원음과의 차이를 구별할 수 있는지 없는지를 객관적으로 테스트하기 위하여 블라인드 테스트를 사용하였다. 실험에 참가한 인원은 20명이며 연령은 23세에서 31세 사이의 자원자들로 구성되었다. 실험결과는 표 4에 나타내었다. 표 4에서 보는 바와 같이 64kb/s에서 원음(CD음질)과 비교하였을 때 투명한 오디오 코딩 성능을 보였다. HC와 PF를 제외하고는 제안된 codec으로 부호화된 샘플이 원음보다 오히려 선호되는 것을 볼 수 있다. 이것은 부호화 된 음과 원음을 분별하지 못하는 것을 명백하게 보이고 있다.

또한 MPEG-1 layer II와 비교하기 위하여 64kb/s의 전송률과 48kb/s의 전송률에서 원음을 들려준 후에 두 개의 코덱으로 코딩된 음을 모르는 상태에서 각각 들려주고 어느 것이 더 좋은지를 선택하도록 하였다. 실험 샘플은 저주파 대역에 신호가 많은 첼로, 고주파 성분이 많은 하프시코드와 어택이 강한 봉고를 실험 샘플로 사용하였다. 실험 결과를 표 5에 나타내었다. 표 5에서는 48kb/s와 64kb/s에서 각각 MPEG_1 Layer II에 대한 제안된 코덱의 선호도를 확률로 나타내고 있다. 표 5에서 보는 바와 같이 64kb/s에서 첼로와 같이 저주파 대역에 에너지가 많은 신호인 경우에는 MPEG이 band limiting 방법을 사용하므로 조금 선호도가 높았으나, harpsichord와 같이 고주파 대역으로 널리 퍼져 있는 신호에 대해서는 제안된 코덱이 월등히 선호되는 것을 보인다. 봉고와 같이 어택이 강한 신호에 대해서도 제안된 코덱이 우수함을 보인다. 또한 48kb/s의 전송률에서는 제안된 코덱이 월등히 성능이 우수한 것을 나타내고 있다. 이와 같은 결과는 MPEG으로 코딩된 신호와 본 연구에서 제안한 코덱으로 코딩된 신호의 파형분석으로도 명확하게 증명된다. 그림 10에서는 attack이 강한 봉고의 48kb/s와 64kb/s의 전송률에서의 MPEG과 제안된 코덱의 스펙트럼 분석 결과를 보였으며, 그림 11에는 고주파 성분이 많은 harpsichord의 64kb/s 전송률에서의 MPEG과의 비교를 보인다. 그림 10(b)와 (c)에서 보는 바와 같이 MPEG의 경우 band limiting하는 방법을 사용한 결과 낮은 전송률에서 심각한 파형의 왜곡을 가져온다. 그러나 제안된 코덱의 경우 낮은 전송률에서도 정확한 심리음향 정보에 의한 bit 할당을 하므로 파형의 왜곡이 적게 나타나고 있음을 볼 수 있다. 그림 10의 (d)와 (e)에서는 전송률이 64kb/s에서 band limiting에 의한 왜곡이 적어져 제안된 코덱과 비슷한 것을 볼 수 있다. 그러나 그림 11의 (b)와 (c)에서 보는 바와 같이 고주파 성분이 많은 harpsichord와 같은 경우, 역시 부정확한 bit 할당을 보상하기 위한 band limiting에 의하여 파형의 왜곡이 64kb/s에서도 뚜렷하게 나타나는 것을 볼 수 있다. 그러나 본 연구에서 제안된 코덱은 고주파 성분이 많은 신호에 대하여서도 충실하게 음을 표현하고 있는 것을 볼 수 있다.

표 3. 실험에 사용된 오디오 샘플

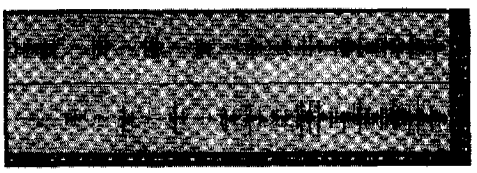
CL	CELLO
CB	CONTRABASS
HC	HARPSICHORD
PS	PIANO_SLOW
PF	PIANO_FAST
BO	BAROQUE ORGAN
BG	BONGO
DS	DRUM SET

표 4. 원음과 코딩된 음의 분별도

오디오 샘플	원음 평균점수	코딩된 신호 점수	분별도
CB	5.57	6.14	분별못함
HC	7.42	7.35	분별못함
PS	9.07	8.07	분별못함
PF	7.93	7.79	분별못함
BO	7.86	7.86	분별못함
BG	7.21	7.29	분별못함
DS	7.57	7.64	분별못함

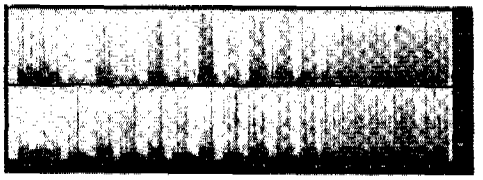
표 5. 제안된 코덱과 MPEG_1 Layer II의 48kb/s와 64kb/s에서 비교시 제안된 코덱이 선호되는 평균 확률

오디오 샘플	48kb/s	64kb/s
CL	1	0.46
HC	0.92	0.83
BO	1	0.54

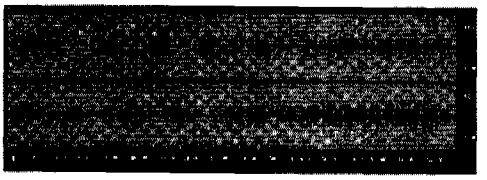


(a) Bongo의 original waveform

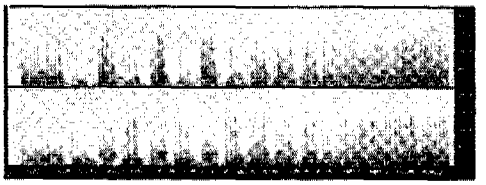
www.dbpia.co.kr



(b) Bongo의 original spectrogram

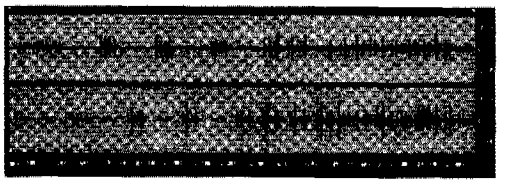


(c) Bongo를 MPEG-1 layer II를 이용하여 64kb/s로 압축한 신호의 waveform

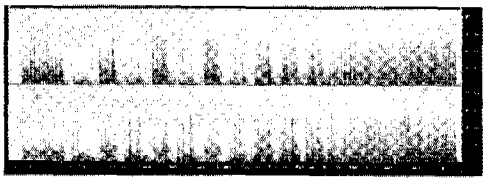


(d) Bongo를 MPEG-1 layer II를 이용하여 64kb/s로 압축한 신호의 spectrogram

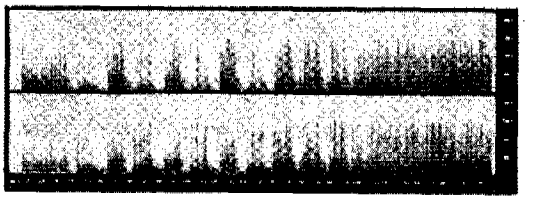
(e) Bongo를 MPEG-1 layer II를 이용하여 48kb/s로 압축한 신호의 spectrogram



(f) Bongo를 제안된 코덱을 이용하여 64kb/s로 압축한 신호의 waveform

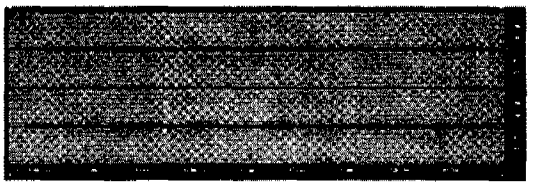


(g) Bongo를 제안된 코덱을 이용하여 64kb/s로 압축한 신호의 spectrogram

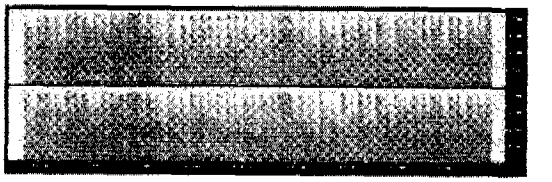


(h) Bongo를 제안된 코덱을 이용하여 48kb/s로 압축한 신호의 spectrogram

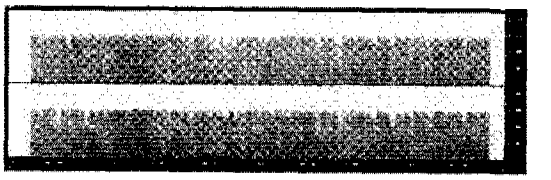
그림 10. Bongo를 64kb/s와 48kb/s의 전송률로 MPEG과 제안된 코덱을 이용하여 각각 압축하였을 때의 wave 파형과 spectrogram



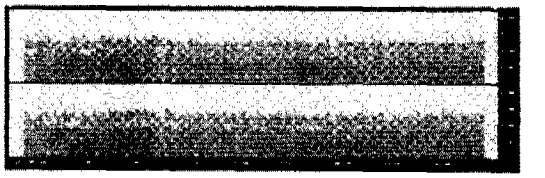
(a) Harpsichord의 original waveform



(b) Harpsichord의 original spectrogram



(c) Harpsichord를 제안된 코덱을 이용하여 64kb/s로 압축한 신호의 spectrogram



(d) Harpsichord를 MPEG-1 layer II를 이용하여 64kb/s로 압축한 신호의 spectrogram

그림 11. 고주파 성분이 많은harpsichord를 64kb/s의 전송률로 MPEG과 제안된 코덱을 이용하여 각각 압축하였을 때의 wave 파형과 spectrogram

www.dbpia.co.kr

Ⅳ. 결 론

제안된 웨이블렛을 이용한 코덱은, 웨이블렛의 신호해석에 있어 시간영역에서의 로컬라이제이션과, 주파수 영역에서의 해상도의 융통성이 인간 청각계의 반응특성과 다이나믹한 오디오 신호의 해석에 적합하다는 결론을 내렸다. 오디오 신호 압축의 핵심적인 심리음향에 적절한 웨이블렛 변환을 수행하기 위해서는 다양한 웨이블렛 필터들 중에 최대 vanishing moment를 갖는 웨이블렛이 적절하며, 웨이블렛 필터길이가 증가할수록 이러한 제약이 줄어든다는 결론을 내렸다. 그러나, 필터길이는 무한정 길어질 수 없으며 위의 moment 조건과 함께 적절히 결정되어져야 한다. 또한 심리음향모델에서 해석된 정보는 웨이블렛 계수들의 밴드별 특징을 고려하여 적용되어야 하며 이를 위하여 적절한 심리음향 모델의 설계는 중요한 문제이다.

전송률을 낮추기 위해서는 심리음향모델에 의한 방법만이 아니라, 매우 중복성(redundancy)이 많은 오디오 신호의 통계적인 특징을 이용하여 전송률을 낮추도록 시도하였다. 웨이블렛 변환된 신호는 오디오 신호의 이러한 중복성을 어느 정도 제거하는데 효율적이라는 것이 본 연구에서 입증되었으며, 이러한 신호 자체의 통계적인 특징은 여러 연구에서 VQ(vector quantizer)와 같은 방법으로 상당한 전송률을 줄일 수 있는 것이 입증되었다. 또한 MPEG-1 layerIII에서는 Huffman 코딩과 같은 엔트로피 코딩 방법으로 이러한 신호의 중복성을 상당부분 제거하였다. 그러나, 본 연구에서는 이러한 복잡한 알고리즘을 사용하지 않고도 오디오 신호 자체의 특성과 웨이블렛의 특성을 이용하여 상대적으로 간단한 코덱으로도 64 kb/s이하의 낮은 전송률에서 좋은 성능의 오디오 신호 코딩을 할 수 있었으며, MPEG오디오 코덱과의 성능비교에서도 48 kb/s에서는 월등히 우수한 성능을 보였다.

연구의 향후 과제로, 중복성이 많은 오디오 신호를 이에 적합한 VQ나 Huffman 코딩과 같은 방법을 사용하여 40kb/s이하의 전송률에서 양질의 오디오 코딩을 수행 하고자 한다.

참 고 문 헌

[1] Mark Kahrs "The Past, Present, and Future of Audio Signal Processing", *IEEE Signal Processing magazine*, Vol.14, No.5, pp.30-81, Sep. 1997.

[2] D. Pan, "A Tutorial on MPEG/Audio Compression", *IEEE Multimedia*, pp. 60-74, 1995.

[3] S. Shilen, "Guide to MPEG-1 Audio Standard", *IEEE Trans. On Broadcasting*, Vol. 40, No. 4, December, 1994.

[4] K. R. Rao and J.J. Hwang, *Techniques & Standards For ImageVideo & Audio Coding*, Prentice Hall, 1996.

[5] Thomas D. Rossing, *The Science of Sound*, Addison-Wesley, 1990.

[6] E. Zwicker and H. Fastl, *Psychoacoustics*, Springer-Verlag, 1990.

[7] D. Johnston, "Transform Coding of Audio Signal Using Perceptual Noise Criteria", *IEEE J. on Selected Areas in Comm.*, Vol.6, No.2, pp.314-323, Feb. 1988.

[8] D. Sinha and A. H. Tewfik, "Low Bit Rate Transparent Audio Compression Using Adapted Wavelets", *IEEE Trans. On Signal Processing*, Vol. 41,No. 12, pp. 3463-3479, Dec. 1993.

[9] Mladen Victor Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, IEEE Press, 1994.

[10] M. Vetterli, *Wavelets and Subband Coding*, Prentice Hall, 1995.

[11] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, 1993.

[12] P.P. Vaidyanathan, "Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial", *Proc. of The IEEE*, Vol.78, No.1, Jan. 1990.

[13] P.P. Vaidyanathan, "Orthonormal and Biorthonormal Filter Banks as Convolvers, and Convolutional Coding Gain", *IEEE Trans. On Signal Processing*, Vol. 41, No. 6, June 1993.

[14] Ail N. Akansu and Richard A. Haddad, *Multiresolution Signal Decomposition*, Academic Press, Inc., 1992.

[15] S. J. Elliot and P. A. Nelson, "Active Noise Control", *IEEE Signal Processing Magazine*,

www.dbpia.co.kr

vol. 10, no. 4, pp. 12-35, Oct. 1993.

[16] ISO/IEC 11172-3 "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to 1.5Mbit/s-Audio Part", *International Standard*, Nov. 1992.

박 관 우(Kwanwoo Park) 정회원

1997년 2월 인하대학교 전자공학과 (학사)

1999년 2월 인하대학교 전자공학과 대학원졸업(석사, 디지털신호처리 전공)

1999년 3월~현재 LG 정보통신

<주관심 분야> 디지털 신호처리, 오디오 신호처리, 심리음향

정 재 호(Jae Ho Chung) 정회원

1982년 Univ. of Maryland (학사)

1984년 Univ. of Maryland (석사)

1990년 Georgia Institute of Technology (박사)

1984년~1985년 미국 국방성 산하 해군 연구소, 신호처리 연구실, 연구원

1991년~1992년 AT&T Bell Laboratories, 음성신호처리 연구실, 연구원 (MTS)

1992년~현재 인하대학교 공과대학 전자공학과 부교수

<주관심 분야> 디지털 신호처리, 음성신호처리, 영상신호처리

이 논문은 1997년 한국 학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

www.dbpia.co.kr