特輯

HDTV용 고음질 디지탈 오디오 데이타 압축기술

金 鍾 一, 李 炳 旭
大宇電子(株) 映像研究所

I. 서 론

효율적으로 압축된 다채널 디지탈 오디오 신호는 기존의 NTSC 방식보다 다양한 서비스를 제공할 수 있다는 점과 정보량의 증가없이 고주파 성분을 살려 양질의 음성 및 음악신호의 전송을 가능하게 함으로써 제한된 전송채널을 통한 CD(compact disk) 수준의 오디오 신호재생을 목표로 하는 고화질 텔레비젼 정보 전송 시스템에 응용되고 있다.

협대역(7KHz 이하) 음성 및 음악신호와는 달리 광대역(20Hz - 20KHz) 오디오 신호는 동작 범위와 용장도(冗長度) 면에서 데이타 압축이 어려운 관계로 협대역 신호를 위한 기존의 전화 선로 통신에 상용되는 μ-law PCM(pulse code modulation)[17] 방법이나 CCITT(International Telegraph and Telephone Consultative Committee) 권고안 (G.721, G.722)과 같이 신호의 가우시안 발생 모델에 가정을 둔 평균 자승오차를 최소화 시키려는 파형부호화 방법으로는 정보 압축의 효율성을 잃게 된다.

디지탈 라디오 방송(digital radio broadcasting), 고화질 텔레비젼의 오디오 전송 및 고음질 오디오 데이타의 저장용으로 최근에 제안된 부호화 방법[11],[13],[15]은 공통적으로 청각특성을 이용한 적응 변환 부호화기로써 청각 파라메타의 검출 과정과 추출된 파라메타를 이용하기 위한 신호의 주파수 대역으로의 변환 및 변환된 신호에 대한 적응 비트 할당으로 구성되며 압축률이 7:1 정도로써 2.5 bits/sample의 정보 전송률로 CD 수준의 음질을 목표로 한다.

Zenith 방식의 HDTV에서 오디오 데이타 압축 방법으로 Dolby에서 제안한 AC-2 방식[11]과 SA(scientific atlanta)에서 북미주지역의 디지탈 라디오 방송망에 응용할 목적으로 제안된 SEDAT(spectrum efficient digital audio technology) 방식은 공통적으로 TDAC (time domain aliasing cancellation)[14] 구조를 이용하여 신호를 주파수 대역으로 변환한 후 변환된 계수로부터 청각 파라메타를 검출하여 각 임계대역에 해당되는 변환계수를 적응적으로 부호화 하고 있다.

GI 및 MIT에서 제안한 HDTV용 오디오 부호화기는 raised-cosine 윈도우와 변형된 DCT를 이용한 TDAC 구조를 통하여 입력신호를 주파수 대역으로 변환한다. 중첩된 1,024개의 한 프레임 데이타를 변환하면 512개의 독립적인 변환 계수를 얻게 되며 각 프레임 마다 계산되는 변환 계수들은 인간의 청각 특성인 임계대역에 맞게 분류되어 부호화 된다.

MPEG motion picture experts group)에서 제안한 MUSICAM 시스템[15]은 분할대역 부호화(sub-band coding) 기법을 이용하여 16비트 균일 PCM 데이타를 입력으로 한 채널당 64Kbps에서 196Kbps까지 가변적으로 부호화할 수 있으며 인간의 청각 특성인 임계대역과 매스킹 현상을 이용하여 layer II에서는 128Kbps에서 CD 수준의 음질을 낼 수 있는 방식이다.

본 논문에서는 최근 고음질 오디오 데이타 압축 기술의 핵심인 인간의 청각특성[2,3,4,5]을 기반으로한 디지탈 오디오 신호처리 개념과 추출된 청각 파라메타를 응용한 디지탈 오디오 신호원의 인지 정보량(perceptual entropy)[5,6] 예측 및 인간의 청각 특성인 매스킹 문턱치를 이용한 객관 음질 평가 기술을 소개하고 한 채널당 128Kbps의 전송률에서 CD 수준의 음질을 갖는 고음질 오디오 데이타 압축 기술을 고찰하였다.

Ⅱ. 청각 특성을 이용한 디지탈 오디오 신호처리

본 장에서는 오디오 신호 압축에 응용하기 위하여 시간 및 주파수 영역상의 오디오 신호와 인간의 청각 특성에 대하여 고찰하고 이를 이용한 효율적인 부호화 방법을 기술한다.

1. 오디오 신호의 통계적 특성

1) 확률 밀도 함수(probability density function)

아날로그 신호원 x(t)로부터 16비트 균일 양자화기로 디지탈화한 대표적인 디지탈 오디오 신호 x(n)은 그림 1과 같다. 이와 같은 디지탈 오디오 데이타로부터 순시 진폭 분포(instantaneous amplitude distribution)를 구하면 근사적으로 확률 밀도 함수를 구할 수 있게 되는데 N개의 유한 구간 동안의 디지탈 신호원 x(n)에 대한 평균과 분산을 각각 μ_x, σ^2_x라 하면 식 (1), (2)와 같이 구한 후 48KHz로 샘플링한 10분 동안의 데이타에 대한 순시 진폭 분포를 표준편차(σ_x)로 정규화한 오디오 신호의 확률 밀도 함수는 그림 2와 같다.

$$\mu_x = \sum_{n=0}^{N-1} x(n) \cdots \quad (1)$$

$$\sigma^2_x = \sum_{n=0}^{N-1} [x(n) - \mu_x]^2 \quad (2)$$

이 분포는 Laplace 확률 밀도 함수[17]에 근사화 시킬 수 있음을 알 수 있다.

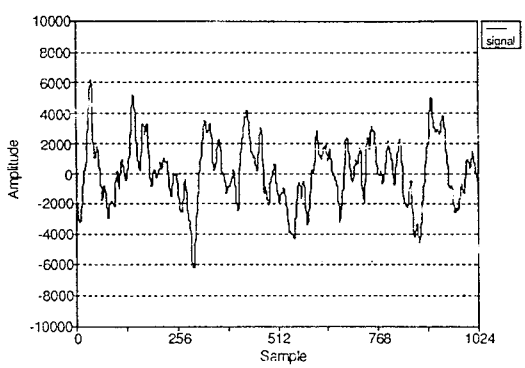


그림 1. 대표적인 디지탈 오디오 신호원

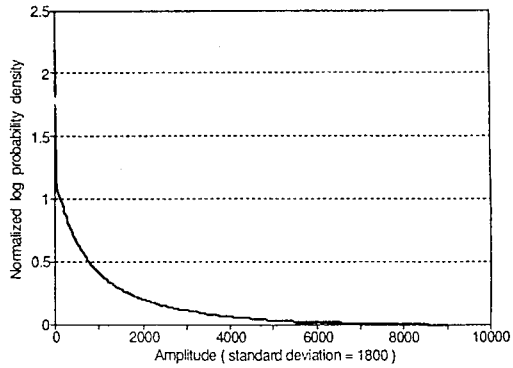


그림 2. 오디오 신호의 확률 밀도 함수

2) 전력 밀도 함수(power spectral density function)

신호원이 정상과정(stationary process)이면 전력 밀도 함수는 이 신호원의 푸리에 변환(Fourier transform)을 이용하여 구할 수 있다[19]. 오디오 신호는 약 20msec 구간에서 단구간 정상과정으로 가정할 수 있으므로 장시간 전력밀도 함수는 푸리에 변환으로부터 구한 단시간 전력밀도 함수들의 산술평균으로 근사화하여 구할 수 있다[17,20]. 디지탈 신호원 x(n)이 정상과정이고 ergodic이라 가정하면 전력밀도 함수 $P_{xx}(\omega)$는 이론적으로 식 (3)과 같이 정리할 수 있다.

$$P_{xx}(\omega) = \lim_{n\to\infty} E\left[1/(2N+1) \left| \sum_{n=-N}^{N} x(n)\cdot e^{-j\omega n} \right|^2\right]$$

$$E[x] = \int x \cdot p(x)\, dx,$$

$p(x)$: probability density function (3)

N개의 유한 신호원에 대한 근사적인 전력밀도 함수를 $\tilde{P}_{xx}(\omega)$라 하면 식 (3)으로부터 식 (4)와 같이 구할 수 있다.

$$\tilde{P}_{xx}(\omega) = 1/N \left| \sum_{n=0}^{N-1} x(n)\cdot e^{-j\omega n} \right|^2 \quad (4)$$

x(n)의 이산 푸리에 변환(discrete Fourier transform)으로부터 j번째 단구간(N개의 신호) (4)의 성분에 대한 전력 밀도를 $P_{xx}(\omega_{ij})$라 할 때 주파수 ω_i 성분에 대한 장시간 전력밀도는 식 (5)와 같이 $P_{xx}(\omega_{ij})$에 대한 산술평균으로 구할 수 있다.

$$P_{xx}(\omega_i) = 1/(2p+1) \left| \sum_{j=-p}^{j=p} P_{xx}(\omega_{ij}) \right| \quad (5)$$

10분동안의 오디오 데이타베이스에 대하여 1,024개의 이산 푸리에 변환 계수로부터 식 (5)로 구한 장시간 평균 전력밀도 스펙트럼은 그림 3과 같다.

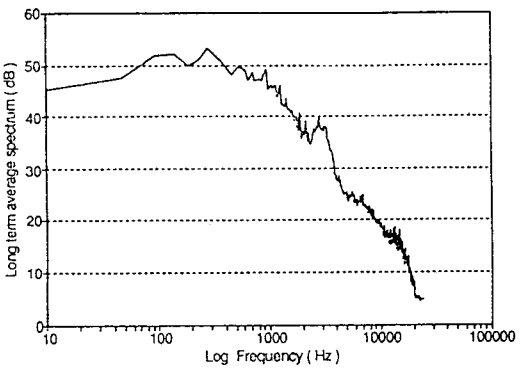


그림 3. 장시간 전력밀도 함수

그림 1의 단구간 오디오신호에 대하여 식 (4)로 구한 순시 전력밀도 스펙트럼은 그림 3의 장시간 분포와 많은 차이가 있고, 식 (2)로부터 구한 장시간 분산값과 단시간 분산값도 차이가 있게 된다[17]. 이와같은 주파수 영역과 시간영역에서 장시간 특성과 단시간 특성의 차이는 오디오 신호의 비정상 특성을 반영한다. 따라서 오디오 신호를 위한 부호화기를 설계할 때 시간영역 및 주파수 영역에서의 변화하는 통계적 특성을 충분히 고려해야 한다.

2. 인간의 청각 특성

1) 소리 인식 과정

인간의 청각 기관은 크게 나누어 외이, 중이, 내이로 구성된다[18]. 외이는 주변의 소리에 대한 음향 에너지의 압력변화를 모아 고막(중이)에 전달하게 된다. 중이는 외이로부터 전달된 압력의 변화를 증폭하여 기계적 운동으로 변환시키는 변환기(tranducer)의 역할을 수행하며 내이는 중이로부터 전달된 기계적 진동양에 대하여 청각 신경계를 자극하도록 하는 전기적 운동으로 변환시켜 뇌에 전달하게 된다. 이때, 외이의 한 부분인 와우각(cochlea)의 기저막(basilar membrane)을 자극하여 소리의 양을 감지하는 과정은 입력신호를 시간 영역 및 주파수 영역으로 분석하여 해석될 수 있다.

2) 매스킹 현상(masking effect)

외이에서 전기적 운동양으로 변환된 소리가 와우각의 기저막을 자극하여 소리의 세기가 인식되는 과정에서 상대적으로 큰 신호가 작은 신호를 매스킹하게 되는데 이 매스킹 현상에는 시간 매스킹과 주파수 매스킹이 있다[4,19].

시간 매스킹 현상은 특정시간을 기준으로 상대적으로 세기가 큰 신호는 그 시간 이전 및 이후의 방향으로 에너지가 전파되어 주변 시간 영역상의 신호를 덜 감지하도록 매스킹하는 현상이다. 그림 4에서 시간 $t=t_1$ 부근의 세기가 상대적으로 큰 신호라면 이전시간의 신호에 대하여 후방 매스킹(backward masking)의 작용을 이후의 신호에 대하여 전방 매스킹(forward masking)의 작용을 하게 되며 이때의 후방매스킹 영향은 약 4msec, 전방 매스킹의 지속시간은 약 40msec인 것으로 알려져 있다. 시간 매스킹 현상을 이용하면, 신호의 크기를 전송하는 부분에서 정보를 줄일 수 있게 된다.

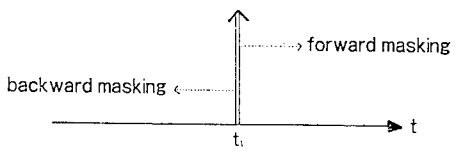


그림 4. 시간 매스킹 현상

주파수 매스킹 현상은 일정시간 동안의 주파수 성분을 인식하는 과정에서 발생하며, 다음절에 기술될 임계대역(critical band)내의 매스킹과 임계대역외의 매스킹으로 구분된다.

한 임계대역으로 입사된 신호는 주변의 임계대역으로 에너지가 전파됨으로써 다른 임계대역의 신호를 매스킹한다. 데시벨 단위의 SPL(sound pressure level)의 기준 레벨을 1KHz 순음(pure tone)의 압력 10^{-6} watt /cm^2(0.0002 dyne /cm^2)에 해당하는 크기로 정의할 때 96dB SPL 크기의 1KHz 순음이 다른 임계대역으로 에너지가 전파되어 그대역의 신경계를 자극시킴으로써 발생되는 상대적인 매스킹 양의 크기는 그림 5와 같다.

매스킹되는 크기는 매스커(매스킹하는 신호) 이전의 대역으로는 급격히 작아지고 매스커 보다 높은 대역으로는 약 8 bark 주파수 동안 넓게 지속됨을 알 수 있다.

3) 임계대역(critical band)

귀에 입사된 신호가 기저막을 거쳐 뇌 신경에 전달되는 과정에서 인간의 귀는 미세한 대역 통과 필터를 통하여 각 대역의 감각량을 분석하여 인지하게 된다. 이때의 대역을 임계대역이라하고 각 임계대역에 대한 감각능력은 입사되는 신호의 전력밀도 스펙트럼과 매스킹함

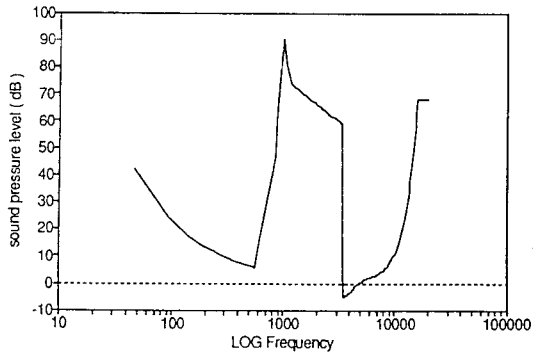


그림 5. 1KHz 순음에 의한 주파수 매스킹 현상

수 및 무 신호시의 각 임계대역에 대한 감각레벨과 밀접한 연관이 있다. 안정상태에서 주파수 대역상의 신호 감지량이 급격히 변하는 부분을 조사하면 임계대역의 각 구간을 얻을 수 있다[4],[5].

선형주파수(Hz) 단위에서 임계대역 주파수단위(critical band rate, bark)로 변환하는 관계는 식(6)과 같다.

$$Z = 13\cdot\arctan(0.76\cdot f/\mathrm{KHz}) + 3.5\cdot\arctan(f/7\cdot 5\mathrm{KHz})^2 \quad (6)$$

임계대역의 주파수 단위는 bark이며 1 bark에 해당하는 선형 주파수 대역을 임계대역 구간으로 간주할 수 있다.

3. 매스킹 문턱치(Masking Threshold) 추출

일정구간 동안의 입력 신호에 대하여 임계대역에서 인지되는 감각량은 입력신호의 스펙트럼과 청각기관의 매스킹 함수에 따라 결정된다.

본 절에서는 단시간 스펙트럼의 예측과정과 여러 매스커가 존재하는 경우의 매스킹 패턴으로부터 각 임계대역에서의 최적화된 문턱치를 추출하기로 한다.

1) 자극 패턴(excitation pattern)의 모델링[2,3,6]

N개의 오디오 신호에 대한 단시간 전력밀도 스펙트럼을 $P_{xx}(\omega)$라 할 때 M개의 임계대역에 대한 i번째 임계대역 구간이 $\omega_{i,l}$에서 $\omega_{i,u}$라 하면, i번째 임계대역의 입력파워는 식 (7)과 같이 구할 수 있다.

$$\sigma^2[i] = 1/\pi \int_{\omega_{i,l}}^{\omega_{i,u}} P_{xx}(\omega)d\omega \quad i = 1, \cdots, M \quad (7)$$

식 (7)로부터 구한 임계대역 입력파워 값으로부터 매스킹 현상을 고려하여 각각의 임계대역에서 감지되도록 자극하는 자극패턴은 식 (8)과 같이 간단하게 근사화하여 모델링 될 수 있다.

$$e_m(i,j) = \begin{cases} \sigma^2[j]\cdot\beta^{dz} & i < j \\ \sigma^2[j] & i = j \\ \sigma^2[j]\cdot\alpha^{dz} & i > \end{cases} \quad (\alpha=0.25,\ \beta=0.003) \quad (8)$$

식 (8)에서 $e_m(i,j)$는 j번째 임계대역의 매스커(매스킹하는 신호)가 i번째 임계대역을 자극하는 양에 해당하는 값이고 dz는 i번째와 j번째 임계대역간의 bark 주파수 단위로 차이의 절대값을 나타내고 $\sigma^2[j]$는 j번째 매스커에 해당하는 입사 스펙트럼 값에 해당된다. 식 (8)로서 모델링된 분해능(resolution)은 1 bark에 해당하므로 한 임계대역 내에서의 자극 패턴은 고려하지 않고 구한 것이다.

여러개의 매스커에 의하여 i번째 임계대역에서 감지되는 자극패턴을 $e_m(i)$라 할 때 무신호 상태의 자극 패턴 $e_q(i)$를 고려하여 식 (9)와 같이 정의하고 i번째 임계대역에서 잡음에 의한 자극패턴은 j번째 임계대역으로 입사되는 잡음의 파워를 $\sigma^2_n[j]$로 가정할 때 식 (8)의 $\sigma^2[j]$를 $\sigma^2_n[j]$로 대치하여 $e_n(i,j)$를 모델링하면 식 (10)과 같이 구할 수 있다.

$$e_m(i) = [e_q[i]^p + \sum_{j=1}^{N} e_m(i,j)^p]^{1/p} \quad (p=0.48) \quad (9)$$

$$e_n(i) = [\sum_{j=1}^{N} e_m(i,j)^p]^{1/p} \quad i = 1, \cdots, M \quad (10)$$

매스커에 의한 자극에 의해 잡음성분에 의한 자극이 매스킹 될 조건을 N = M으로 하여 식 (11)과 식 (12)를 얻을 수 있다[6].

$$10\cdot\log_{10}\{\frac{[e_n[i]^p+e_m[i]^p]^{1/p}}{e_m[i]}\} < 1, \quad i=1, \cdots, N \quad (11)$$

$$e_n[i]^p/e_m[i] < (10^{p/10}-1)^{1/p} \quad i = 1, \cdots, N \quad (12)$$

P=0.48일 때 i번째 임계대역에서 매스킹이 일어날 조건은 $e_n[i]/e_m[i]$의 값이 −19.4 dB이하 임을 알 수

있다. N개의 임계대역에 대하여 식 (12)의 등호 조건을 만족하는 $e_n[i]$를 이용하여 입사전력 스펙트럼으로부터 구한 $e_m[i]$ 값으로부터 식 (10)과 식 (8)을 결합하면 각각의 임계대역에 해당하는 잡음의 파워인 $\sigma^2_n[i]$를 구할 수 있게 된다. 결국 식 (12)의 등호조건을 만족하는 $\sigma^2_n[i]$가 매스킹 문턱치임을 알 수 있다.

4. 인지 정보량(Perceptual Entropy)의 예측

단구간 가우시안 정상과정의 신호를 균일 양자화기로 부호화 할 때 양자화 비트수가 크다고 가정하면 양자화 오차는 각 양자화 레벨의 구간내에서 균일 확률밀도를 갖는다[17]. 이때 최소 평균자승오차 조건을 만족하는 양자화기의 오차 분산값은 overload가 없는 경우 식 (13)과 같이 계산된다[1],[17].

$$\sigma^2_q = \Delta^2/12 \tag{13}$$

식 (13)에서 Δ는 양자화기의 계단크기(step-size)이다. 16비트의 균일 양자화기로 입력신호를 양자화할 때 계단크기는 1이므로 오차신호의 분산값은 1/12이 된다. 따라서 CD(compact disk) 수준의 음질을 갖기 위한 최소 비트 전송률, R_{min}은 식(14)와 같이 구할 수 있다.

$$R_{min} = \frac{1}{\pi}\int_0^{\pi} MAX[0, 1/2\ \log_2(12\cdot P_{xx}(\omega)]\ d\omega \tag{14}$$

식 (14)에서 $P_{xx}(\omega)$는 입력신호에 대한 전력 스펙트럼이다.

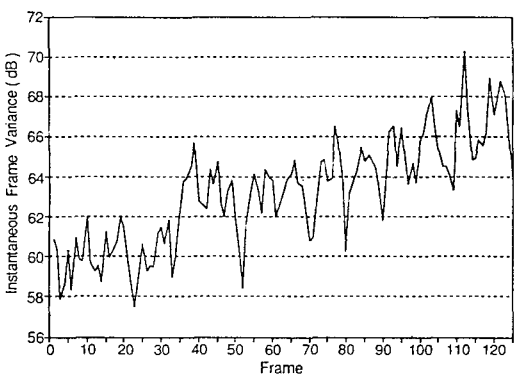


그림 6. 오디오 신호의 순간 분산값

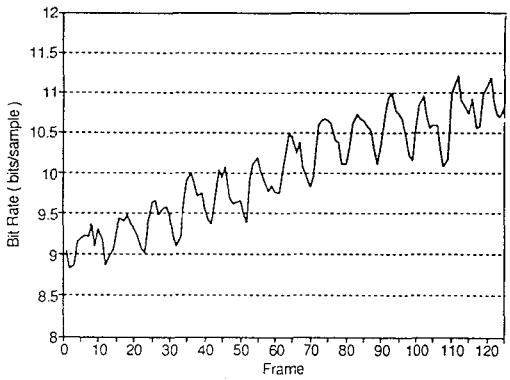


그림 7. 평균 최소 전송률

그림 6의 125프레임 동안의 오디오 신호에 대하여 CD 수준의 음질을 갖기 위한 평균 최소 전송률은 식 (14)를 적용하여 그림 7과 같이 구할 수 있다. 따라서 기존의 평균자승오차를 최소화시키는 양자화기를 평균 최소 전송률이 약 10비트임을 알 수 있다.

입력신호를 K개의 대역으로 분할하여 각 대역을 균일 양자화기로 부호화하는 경우 매스킹된 오차 신호 스펙트럼을 $P_{ee}(\omega)$라 가정하면 i번째 분할대역의 매스킹 문턱치 $\sigma^2_n[i]$는 식 (15)와 같이 구할 수 있고 i번째 분할대역의 입력신호 파워는 식(16)과 같이 구할 수 있다. 기존의 전송률 왜곡함수[1]를 이용하여 인간의 귀로서 오차를 감지할 수 없도록 하면서 최소의 전송률로 전송할 수 있는 정보량에 해당하는 인지 정보량(perceptual entroy)은 식(17)과 같이 구할 수 있다.

$$\sigma^2_n[i] = 1/K\cdot Min[P_{ee}(\omega)],$$
$$(i-1)\pi/K \le \omega < i\pi/K,\ i=1, \cdots, K \tag{15}$$

$$\sigma^2_x[i] = \int_{(i-1)/K}^{(i)/k} P_{xx}(\omega)\ d\omega \tag{16}$$

$$R_{min} = 1/K\sum_{i=1}^{N} MAX[0, 1/2\ \log_2(\sigma^2_x(i)/\sigma^2_n(i))] \tag{17}$$

$P_{xx}(\omega)$를 1,024개의 이산푸리에 변환 계수로부터 근사적으로 계산하여 32개의 분할 대역 (K=32)으로 나누어 부호화하는 경우 그림 6의 오디오 신호에 대하여 식 (17)을 만족하는 인지정보량(최소 정보비트율)은 그림 8과 같다. 약 10분 동안의 오디오 데이타에 대하여 식 (17)로서 구한 평균 인지정보량은 1.2 bit/sample

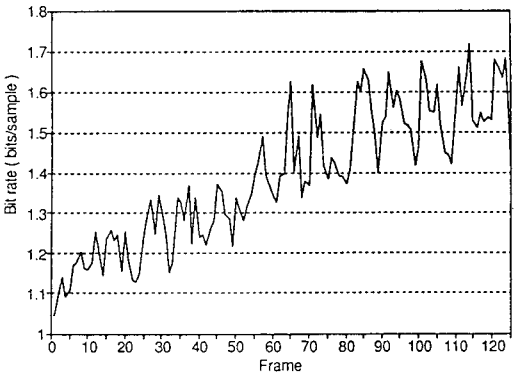


그림 8. 인지 정보량

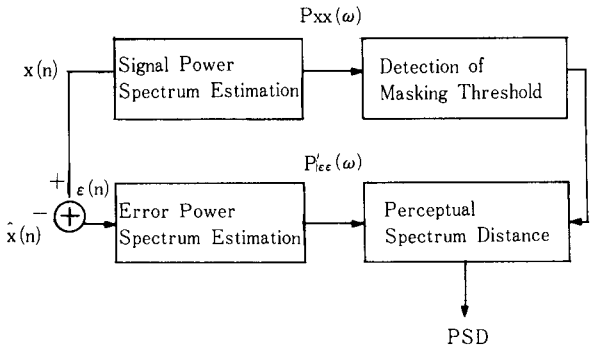


그림 9. 청각 특성을 이용한 부호화기의 객관성능 평가

을 얻을 수 있었다. 한 샘플당 16비트의 길이로 초당 48KHz로 샘플링한 오디오 신호원을 128Kbps로 데이타를 압축하는 경우(6:1 압축) 샘플당 비트 전송률이 2.67 비트에 해당한다. 이 비트율은 평균 인지 정보량(1.2 bit/sample)에 비하여 큰 값이지만 오디오 신호의 비정상 특성을 고려하면 의미있는 압축률이 된다.

부호화기의 버퍼 크기 및 복잡도를 증가시키면 이론적으로 평균 인지 정보량에 가까운 값으로 오차없이 전송할 수 있는데 무한 길이의 버퍼 및 무한 분할대역 부호화기를 이용하는 경우 16비트 균일 양자화기로 디지탈화한 오디오 신호의 압축한계는 약 13:1 임을 알 수 있다.

5. 청각특성을 이용한 부호화기의 객관 성능평가 (Objective Performance Test)

기존의 협대역(7KHz이하) 음성 부호화기의 객관성능 평가방법[10]은 광대역(20Hz−20KHz) 오디오 신호의 주관적 성능(subjective performance)을 반영할 수 없게 된다.

본절에서는 매스킹 문턱치(masking threshold)를 적용하여 광대역 오디오에 대한 복호화된 신호의 왜곡정도를 정량화 시킬 수 있는 객관 성능 평가방법을 제시하였다.

제시된 평가방법의 블럭도는 그림 9와 같다.

그림 9에서 x(n)은 부호화기의 입력신호 $\hat{x}(n)$은 복호화기의 출력신호를 각각 나타내고 $\varepsilon(n)$은 오차신호로서 x(n)과 $\hat{x}(n)$에 대한 시간 영역상의 차에 해당하며 $P_{xx}(\omega)$와 $P_{\varepsilon\varepsilon}(\omega)$는 입력신호 x(n)과 오차신호 $\varepsilon(n)$에 대한 순시전력 스펙트럼을 나타낸다.

N개의 입력신호 및 오차신호에 대하여 전력 스펙트럼 $P_{xx}(\omega)$ 및 $P_{\varepsilon\varepsilon}(\omega)$를 식 (4)로서 구하고 주파수 ω에 대한 매스킹 문턱치 함수를 $M(\omega)$라 하면 왜곡값 PSD (perceptual spectrum distance)는 $E(\omega)$를 식 (18)로 정의하면 식 (19)로 나타낼 수 있다.

$$E(\omega) = P_{\varepsilon\varepsilon}(\omega) - M(\omega) \qquad (18)$$

$$PSD = 1/\pi \int_0^{\pi} MAX\,[0, E(\omega)]\, d\omega \qquad (19)$$

식 (19)의 PSD값은 N개의 복호화된 신호에 대한 오차신호의 스펙트럼이 매스킹 문턱치보다 크게 되는 성분들의 합을 정규화한 것이므로 이론적으로는 모든 주파수 성분에 대하여 복호화된 신호가 입력신호와의 귀로 감지될 수 있는 차이를 나타내는 값이다.

식 (19)에서 MAX 연산은 $E(\omega)$값이 0보다 작아지면 이 부분의 주파수에서는 오차가 없는 경우이므로 0을 취하게 된다.

Ⅲ. 디지탈 오디오 데이타 압축 기술

1. 제안된 HDTV용 오디오 코덱

고화질 텔레비전용으로 Dolby에서 제안한 AC-2 방식[11]은 청각 특성을 이용한 변환 부호화기로써 블럭도는 그림 10과 같다.

중첩된 윈도우와 변형된 DCT(discrete cosine tramsform)로 구성된 TDAC 구조로써 16 비트로 양자화된 오디오 입력신호를 주파수 대역으로 분할하며 이 과정에서 인간의 청각특성에 맞는 비트를 각각의 임계대역에 해당하는 변환계수에 할당하여 부호화하고 각

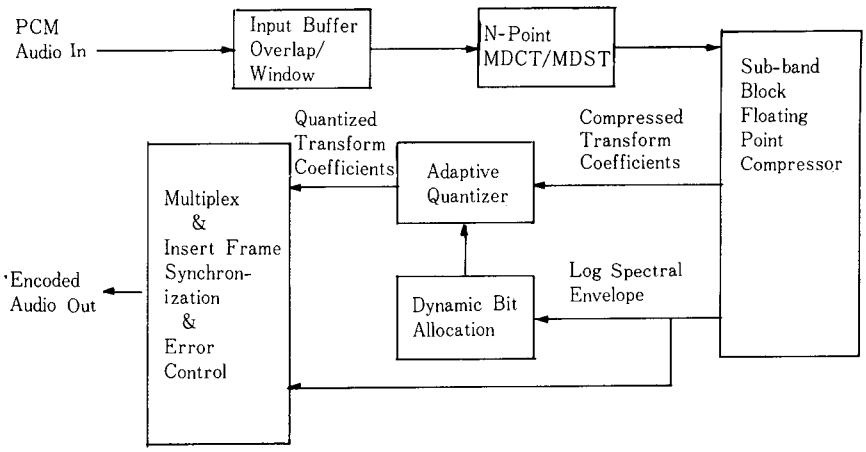


그림 10. AC-2 부호화기의 블럭도

대역의 크기 정보를 비균일 특성을 갖는 양자화기로 부호화하여 수신단에 전송하게 된다.

수신단에서는 비트할당정보와 각 대역의 크기정보(scale factor)를 이용하여 양자화된 변환계수를 재생하고 합성 TDAC 구조로써 원래의 신호를 만들게 된다. 이 방법은 FFT(fast Fourier transform)를 이용한 TDAC 구조를 이용한 경우 낮은 복잡도를 갖는 수신기를 구현할 수 있다.

MIT에서 제안한 오디오 부호화기의 블럭도는 그림 11과 같다. 48KHz로 샘플링된 디지탈 오디오 신호는 raised-cosine 윈도우와 변형된 DCT를 이용한 TDAC 구조를 통하여 주파수 대역으로 변환된다. 중첩된 윈도우를 이용하여 1,024개의 한 프레임 데이타를 변환하면 512개의 독립적인 변환계수를 얻게 되며 각 프레임마다 계산되는 변환계수들은 인간의 청각특성인 임계대역에 맞게 분류되고 각 임계대역안의 변환계수는 기준이 되는 변환계수와 나머지 변환계수로 나누어 부호화 된다. 기준이 되는 변환계수는 고정된 비트를 사용하여 상대적으로 정확하게 부호화되고 나머지 변환계수는 가변비트로써 전송된다.

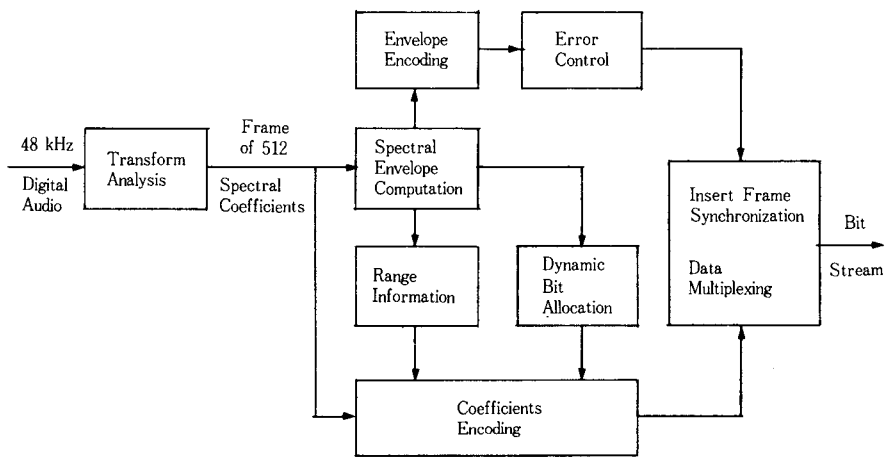


그림 11. MIT-AC 시스템의 부호화기 블럭도

각 임계대역의 변환계수 값으로부터 스펙트럼 포락 파형정보를 계산하여 총 전송비트양의 약 20% 정도를 할당하여 수신단에 전송하므로 임계대역에 해당하는 포락파형은 인간의 청각특성인 매스킹 양에 따른 대역의 상대적 중요도를 간접적으로 반영할 수 있게 된다.

MPEG에서 제안한 MUSICAM시스템[15]은 분할대역 부호화(subband coding) 기법을 이용하여 16비트 균일 PCM 데이타를 입력으로 한 채널당 64Kbps에서 196Kbps까지 가변적으로 부호화 할 수 있으며 인간의 청각특성인 임계대역과 매스킹 현상을 이용하여 layer Ⅱ에서는 112Kbps와 128Kbps에서 CD 수준의 음질을 낼 수 있는 방식이다.

MUSICAM(masking pattern adapted universal subband integrated coding and multiplexing) 방식의 블럭도는 그림 12와 같다.

입력된 신호는 분할대역 필터(subband analysis filter)로부터 32개의 균일한 주파수 대역으로 분리되며 동시에 FFT(fast Fourier transform)를 통한 입력신호의 스펙트럼으로부터 인간의 청각특성에 부합되는 파라메타가 검출되어 이 특성에 맞게 각 대역은 블럭 압신 부호화(block companding coding) 방법으로 부호화된다.

이 과정에서 전송선로 저장 매체 및 수요자의 요구에 따라 선택적으로 오디오 서비스를 제공받을 수 있도록 layer 구조로써 계층적으로 포맷화된다. 수신단에서는 전송되어진 layer 형태와 비트 할당 정보를 받아 신호를 재생하게 된다.

무신호시의 매스킹 문턱치와 순음 비순음 성분의 모든 매스커들에 의한 i번째 임계대역의 매스킹 문턱치를 $LT_g(i)$라 하면 각 분할 대역에서의 매스킹 문턱치는 임계대역 구간에 해당하는 매스킹 문턱치 값들로부터 식(20)과 같이 각 분할대역 구간에 속하는 값들 중 최소값을 선정하여 구할 수 있다.

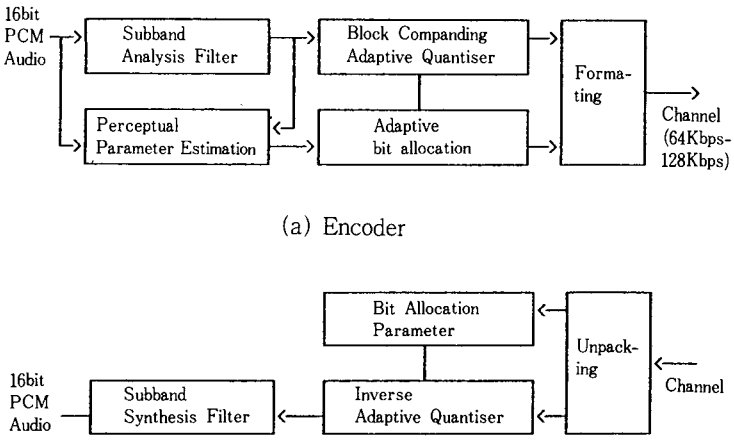
$$LT_{min}(n)=Min[LT_g(i)\ dB,\ i\in Bw(n)] \quad (20)$$

식 (20)에서 Bw(n)은 n번째 분할대역의 주파수 구간이다. n번째 분할대역에서 부호화시 필요한 신호대 잡음비(신호대 매스킹 문턱치 비) $SMR_{sb}(n)$은 식(21)과 같이 구할 수 있다.

$$SMR_{sb}(n)-L_{sb}(n)-LT_{min}(n)\ dB \quad (21)$$

신호원이 ergodic[19]하다면 j번째 프레임에서 최소화해야 할 목적함수 E(j)는 x비트를 입력으로 한 양자화기의 신호대 양자화 잡음 특성함수를 Q(x), 각 주파수 ω성분에 대한 양자화 비트수를 $R(\omega, j)$, 각 주파수 ω 성분에 대한 신호대 매스킹 문턱치 함수를 $SMR(\omega, j)$라 하면, 식 (22)와 같이 정의할 수 있다.

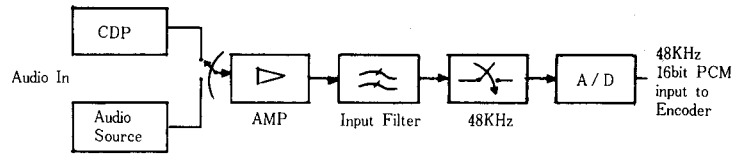
$$E(j)=1/\pi \int_0^{\pi} \{SMR(\omega, j)-Q[R(\omega, j)]\}^2\ d\omega \quad (22)$$



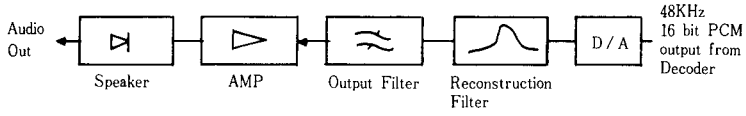
(a) Encoder

(b) Decoder

그림 12. MUSICAM 방식의 블럭도



(a) 입력장치



(b) 출력장치

그림 14. 오디오 입출력 처리장치의 블럭도

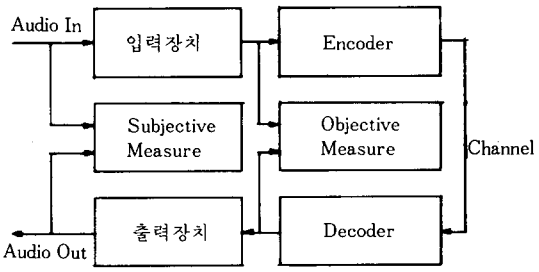


그림 15. 모의실험 과정을 나타내는 블럭도

Ⅳ. 결 론

최근의 HDTV를 고음질 디지탈 오디오 신호처리 기술은 인간의 청각특성을 적용한 주파수대역 변환부호화 방식으로 한 채널당 128Kbps 전송률에서 간단한 수신기로 CD 수준의 음질을 재생하는 기술로 집약된다.

CD(compact disk) 수준의 음질을 갖는 광대역(20Hz-20KHz) 오디오 부호화기에서 청각특성을 고려하지 않은 기존의 최소 평균 자승 오차(mimimum mean square error)의 조건을 만족시키도록 접근하는 파형부호화 방법으로는 데이타압축이 거의 불가능하며 매스킹현상과 자극패턴 모델을 전송률 왜곡 이론과 결합하면 최소의 정보 전송률로 오차신호를 감지할 수 없도록 하는 광대역 오디오 부호화기의 최적화 문제를 해결할 수 있다. 매스킹된 오차 스펙트럼으로부터 32대역으로 분할하여 부호화하는 경우 16비트 균일 PCM 데이타를 입력으로하는 광대역 오디오 부호화기의 압축한계는 약 13:1임을 알 수 있다.

자극 패턴의 모델링 과제는 인간의 청각특성을 반영하는 파라메타 추출의 핵심 부분으로 좀더 정확한 자극패턴 모델링은 곧 좋은 특성의 부호화기와 연결된다. 따라서 기존의 자극 패턴 모델링 부분과 이를 통한 청각 파라메타 추출과정은 오디오 신호가 일정구간에서 정상과정(stationary process)이라는 가정하에 전개된 것이므로 오디오 신호에 대한 시간 영역상의 비정상 특성을 고려한 새로운 자극 패턴의 모델링 부분은 앞으로의 연구과제이다.

기존 CD와의 인터페이스를 위한 샘플링 변환장치와 디지탈 신호처리기 및 VLSI 기술로서 광대역 오디오 부호화기를 구현하는 부분은 현재 활발히 연구 진행되는 분야이다.

參 考 文 獻

[1] D. J. Sakrison, "The rate distortion function of a Gaussian process with a weighted square error criterion," *IEEE Trans. Inform. Theory*, vol. 14, pp. 506-508, 1968.

[2] L. E. Humes and W. Jesteat, "Models of the activity of masking," *J. Acoust. Soc. Amer.*, vol. 85, no. 3, pp. 1285-1294, 1989.

[3] M. Florentine and S. Buus, "An Excitation-pattern model for intensity discrimination," *J. Acoust. Soc. Amer.*, vol. 70,

N개의 유한분할 대역으로 나누어 부호화하는 경우, 유한개(N개)의 양자화기 및 각 양자화기의 제한된 비트를 고려하면 새로운 목적함수 $E_m(j)$를 식 (23)과 같이 재정의하면 이론적으로 j번째 프레임에서 최적화된 비트 할당은 j번째 프레임에서 사용가능한 비트양을 B(j)라 할 때 식 (24)를 만족하는 R(i,j), i=1,… N(N은 분할 대역수)임을 알 수 있다.

$$E_m(j) = \sum_{i=1}^{N} \{SMR(i,j) - Q[R(i,j)]\}^2 \quad (23)$$

$$\text{minimize } E_m(j)$$
$$\text{under the constraint,}$$
$$\sum_{i=1}^{N} R(i,j) \le B(j) \quad (24)$$

그러나 실제 오디오 데이타를 처리하기 위한 부호화기에서 R(i,j)의 정수 제약조건, 양자화기 특성함수 Q[R(i,j)]의 비선형 특성, overflow 및 underflow 상태를 모두 고려해 볼 때 식 (24)를 만족하는 R(i,j)를 실시간으로 구하는데 제약이 따른다.

계산량 및 효율성을 고려한 적응 비트할당 과정은 그림 13과 같다.

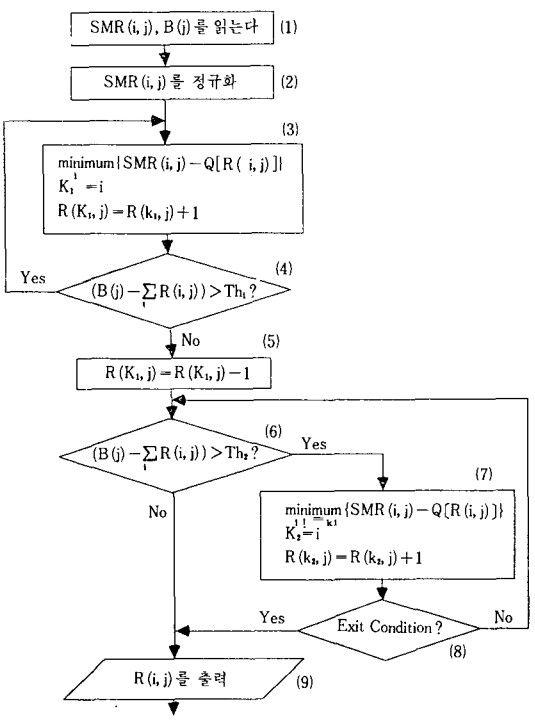


그림 13. 적응 비트 할당 알고리듬의 순서도

그림 13에서 (2)의 정규화 과정은 overflow 및 underflow 상태가 있는 경유를 대비해 SMR(i,j)를 보정하여 계산의 효율을 높이는 과정이고, Th_1은 허용 가능한 최대 비트 증가양, Th_2는 허용되는 최소 비트증가양에 해당한다. (8) 부분은 무한 loop를 방지하는 부분이다. 이와 같은 비트할당 알고리즘은 실제의 오디오 데이타를 처리하는 과정에서 간단한 비교 논리를 구현할 수 있으므로 실시간 처리를 하는 경우 많은 계산양의 감축이 예상된다.

복호화기의 구조는 그림 14의 (b)와 같다.

복호화 과정은 역적응 양자화기와 합성대역 피터 부분으로 구성된다. 역적응 양자화기는 전송되어온 각 대역 신호에 대한 부호화 정보 및 보조 정보인 블럭 압신 부호(block companding code)와 비트 할당 정보를 이용하여 각 대역의 신호를 복원하게 되고, 합성대역 필터는 복원된 각 대역 신호로부터 분할대역 필터의 역과정으로 16비트 PCM 데이타를 재생하게 된다.

2. 오디오 입출력 처리

일반적인 오디오 신호에 대한 코덱의 입출력 처리과정은 그림 14와 같이 나타낼 수 있다. CD 수준의 음질을 갖는 HDTV용 오디오 코덱에 대한 실험을 위해서는 아날로그 데이타를 20Hz−20KHz의 대역을 갖는 anti-aliasing 필터로서 대역 통과시키고 2-channl의 48KHz 샘플링를 통해 16비트 균일 양자화한 신호를 디지탈 신호원으로 간주할 수 있고 CD에 저장된 16비트 디지탈 데이타를 직접 48KHz로 디지탈 영역에서 샘플링 변환장치를 이용하여 얻을 수 있다.

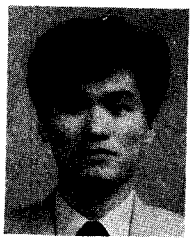
디지탈 오디오 코덱을 통하여 복호화된 출력신호는 DAC(digital to analog convertor) 및 합성필터와 주변장치로서 아날로그 신호를 재생하게 된다.

그림 14의 주변 입출력 처리장치를 포함하여 전체적인 모의 실험 및 성능평가를 수행하는 블럭도는 그림 15와 같다.

입출력 아날로그 신호로부터 주관평가(subjective measure)를 수행하고 부호화기의 입출력 신호인 16비트 균일 PCM 신호로서 객관 평가(objective measure)를 실시하게 된다.

오디오 코덱을 개발하는 과정에 있어서는 주관평가를 통한 실험은 많은 시간과 경비가 소요되므로 객관평가를 통해 정량적인 데이타분석을 하게 되지만 객관평가의 결과가 주관평가의 결과를 잘 반영해야만 한다. 따라서 객관평가의 기준은 실험 데이타의 특성에 맞게 선정해야 한다.

筆者紹介

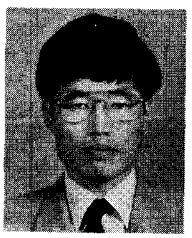


金 鍾 一

1965年 1月 5日生
1988年 2月 연세대학교 전자공학과(학사)
1990年 8月 연세대학교 대학원 전자공학과(석사)

1990年 8月～현재 대우전자 영상연구소

주관심분야 : 디지탈 신호처리, 영상 및 음성부호화



李 炳 旭

1957年 1月 14日生
1979年 2月 서울대학교 전자공학과(학사)
1981年 8月 한국과학기술원 전기 및 전자공학과(석사)
1990年 6月 Stanford Univ. Electrical Eng.(박사)

1981年 8月 ~ 1985年 8月 대우전자 중앙연구소
1990年 7月 ~ 현재 대우전자 영상연구소 책임연구원

주관심분야 : 영상 및 음성부호화, Computer Vision, Computer Graphics

no. 6, pp. 1646-1654, 1981.

[4] M.R. Schroeder, B. S. Atal,and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *J. Acoust. Soc. Amer.*, vol. 66, no. 6, pp. 1647-1651, 1979.

[5] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Select Areas Commun.*, vol. 6, pp. 314-323, 1988.

[6] R. N. J. Veldhuis, "Bit rates in audio source coding," *IEEE J. Select. Areas Commun.*, .vol. 10, 1992.

[7] P. P. Vaidyanathan, "Multirate digital filters," *Proc. IEEE*, vol. 78, no. 1, pp. 56-93, 1990.

[8] P. L. Chu, "Quadrature mirror filter design for an arbitrary number of equal bandwidth channels," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 33, no. 1, pp. 203-218. 1985.

[9] R. N. J. Veldhuis, M. Breeuwer, and R. V. Waal, "Subband Coding of Digital Audio Signals without loss of Quality," Proc. IEEE ICASSP'89, pp. 2009-2012.

[10] N. Kitawaki, K. Itoh, and M. Honda, "Speech quality accessment methods for speech coding systems," *IEEE Communication Magazine*, vol. 22, no. 10, pp. 26-33, 1984.

[11] G. Davidson, L. Fielder, and M. Antill, "High-Quality Audio Transform Coding at 128 Kbits/sec.," Proc. IEEE ICASSP '89, pp. 1117-1120.

[12] T. B. Keller, "Proposal for Advanced HDTV Audio," Proc. NAB HDTV World Conference, pp. 38-43, 1991.

[13] W. Koos, L. Koos, and K. Malinowski, "Spectrum Efficient Digital Audio Technology (SEDAT)," Proc. NAB Broadcast Engineering Conference, pp. 20-25, 1991.

[14] J. P. Princen and A. B. Bradley, "Analysis/synthesis filter bank design based on time-doman aliasing cancellation," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 34, no. 5, pp. 1153-1161, 1986.

[15] ISO/IEC JTCI/SC2/WG 11, Part 3. Audio proposal, CD-11172-3, 1991.

[16] S. Bergman, C. Grewin and T. Ryden, "The SR Report on the MPEG/AUDIO Subjective Listening Test," Stockholm, 1991.

[17] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Englewood Chiffs, Prentice-Hall, NJ, 1984.

[18] D. O'Shaughnessy, *Speech Communication*, Addison-Wesley Pub. Com., 1987.

[19] A. Papoulis, *Probability, Randon Variables, and Stochastic Process*, MacGrow-Hill, 1984.

[20] S. L. Marple, Jr., *Digital Spectral Analysis with Applications*, Englewood Cliffs, Prentice-Hall, NJ, 1984.