멀티미디어를 위한 디지털 오디오 알고리즘

□ 차형태*, 김헌중*, 김수현*/ *숭실대학교 멀티미디어 시스템 연구실

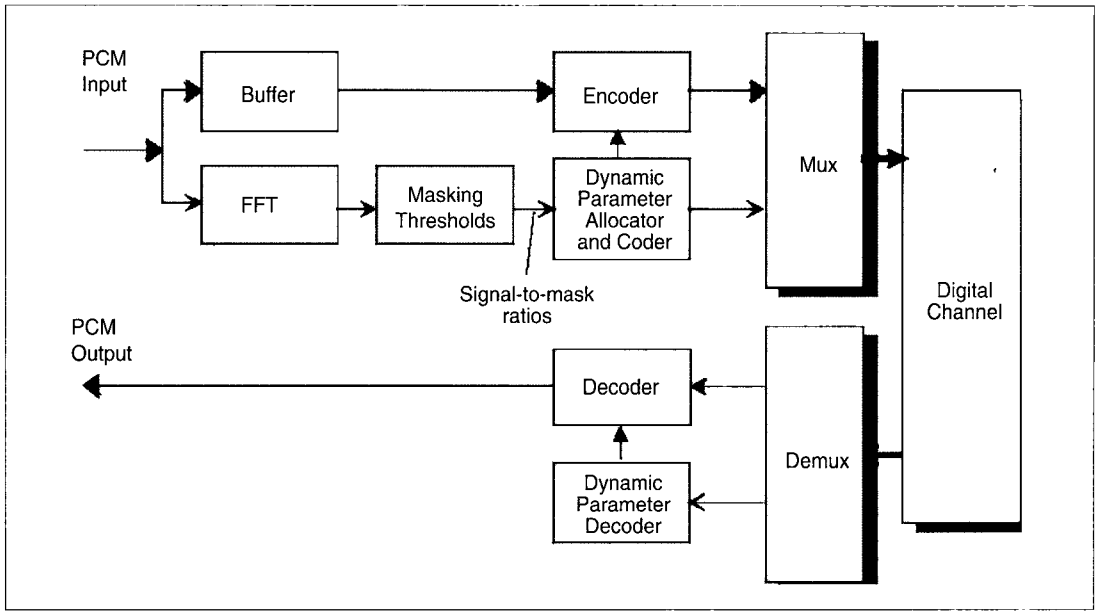
Ⅰ. 서 론

일반적으로 오디오 신호는 telephone speech, wideband speech와 wide band audio로 분류할 수 있으며, 이러한 것들은 각기 다른 대역폭, dynamic range와 제공되는 음질에 대한 청자(聽子)의 기대치가 각기 다르다. 이 중 wide band audio (즉, high quality) 표현 방식이라 함은 일반적으로 최소한 20kHz 이상의 대역폭과 다 채널 오디오를 포함한 오디오 표현 방식을 의미한다.

이미 오늘날의 디지털 오디오 표현 방식의 de facto standard로 자리 잡은 CD(Compact Disc)의 도입은 다양한 사용자의 요구 조건을 충족시키고 디지털 표현 방식이 주는 이점으로 인해 사용자들에게 선공 적으로 수용되었다. 또한 새로운 인터넷 기반의 네트워크, 무선망, 멀티미디어 시스템들은 재생 음질(reproduction quality)의 저하 없이 고압축률을 갖는 새로운 오디오 포멧의 도래를 요구하게 되었다. 이와 같은 요구 조건을 만족 시키기 위해 지난 10여년간 지각적으로 인위적인 잡음의 영향을 받지 않고 투명성을 유지하는 코딩기법(perceptually transparent coding) 을 통한 압축 알고리즘의 개발을 위한 연구가 수행되어 지게 되었다. 그 결과 몇몇 알고리즘들은 국제 또는 산업 규격들로 채택되어 지게 되었고, 뿐만 아니라 방대해진 통신 환경에서 다양한 목적에 부합하는 알고리즘의 개발이 이루어 지게 되었다.

이와 같은 지각적으로 투명성을 갖는 코딩을 위한 고음질의 디지털 코딩 알고리즘들로는 ISO/IEC MPEG family (-1,-2,-2 AAC,-4), Lucent Technologies PAC/EPAC/MPAC, Dolby AC-2/AC-3와 Sony ATRAC/SDDS 알고리즘들이 있다.

이 중 MPEG 오디오의 경우, 1992년에 4년여



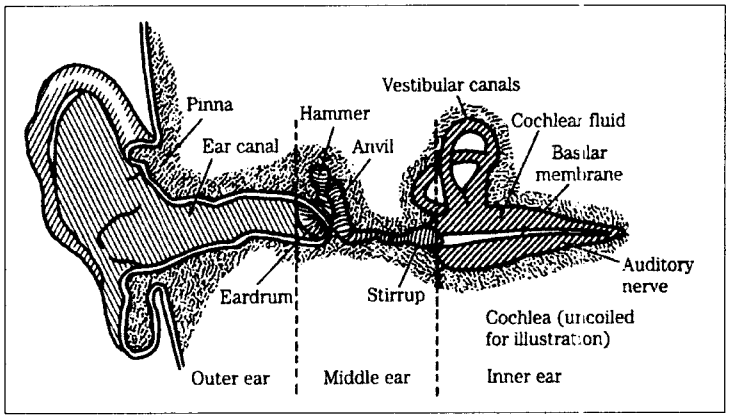
〈그림 1〉 지각적으로 인위적인 잡음의 영향을 받지 않고 투명성을 유지하는 코딩기법(perceptually transparent coding)

간의 세계 각 국의 오디오 신호처리 전문가들의 광범위한 연구를 통해 스테레오 채널의 CD음질 수준의 오디오를 위한 ISO/MPEG(International Standard Organization/Moving Picture Experts Group) 오디오 코딩 규격(ISO/IEC 11172 (MPEG-1))을 채택을 시작으로, 현재 MPEG-2, MPEG-2 AAC(Advanced Audio Coding), MPEG 4에 이르기까지 다양한 적용환경과 확장성을 가지고 발전해 가고 있다. 이러한 MPEG 오디오기술은 subband decomposition, filter bank analysis, transform coding, entropy coding, dynamic bit allocation, nonuniform quantization, adaptive segmentation와 psychoacoustic analysis등의 기술을 결합시킨 하이브리드 코딩 기술로써 적용 환경에 따라 몇 가지 다양한 방법들로 구성할 수 있도록 되어 있다. 이러한 MPEG Audio Family중 MPEG-1의 경우 현재 다양한 어플리케이션에서 성공을 거두었다. 예를 들어, MPEG-1 Layer III (.MP3)의 경우 압축 오디오의 전송과 저장에 있어 WWW과 포터블 멀티미디어 시스템(e.g. MP3 Player, MP3CDP등)분야에서 많은 사용자들에 의해 표준화되었다. 이와 같이 MPEG-1 오디오 코딩기법은 점진적으로 끊임없이 다양한 적용 분야들에 수용되어지면서 현재에는 DBA(European digital radio) 또는 Eureka, 위성 방송과 디지털 컴팩트 카세트(digital compact cassette) 같은 대규모 시스템에까지 수용되기에 이르렀다. 더욱이 최근에는 ACTS(the collaborative European Advanced Communications Technologies and Service)에서 MPEG 오디오와 비디오를 텔레비전 프로그램 제작과 배급에 다양한 기능성을 제공하기 위해 핵심 압축기술로 채택하기 이르렀다 [1][3][5][6][7][8].

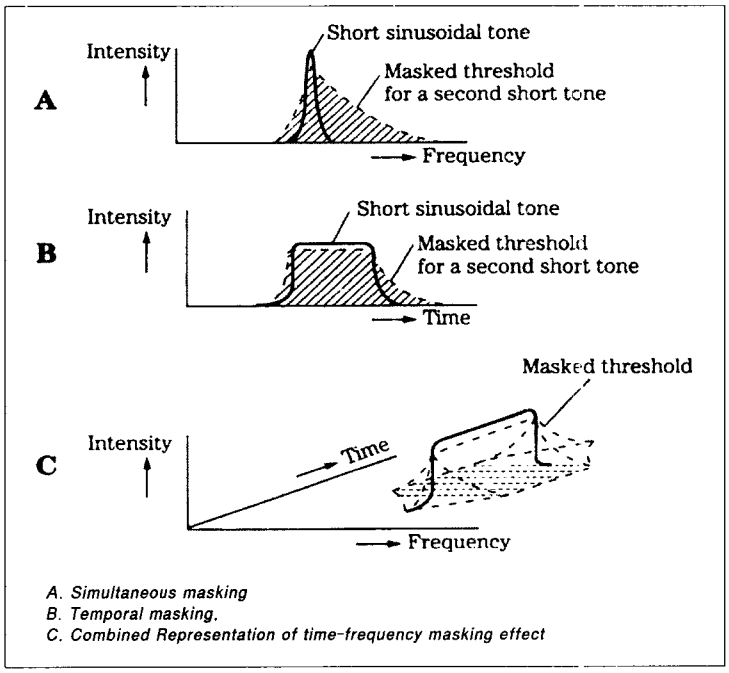
Ⅱ. 오디오 코딩을 위한 기본 개념

1. Perceptual Coding

일반적으로 오디오 코딩 알고리즘들은 코딩 효율을 최적화하기 위해 오디오 신호에 대한 일반적인 수신기 모델에 의존하게 된다. 오디오 신호에 있어서 이러한 수신기는 사람의 귀가 되고, 사운드에 대한 사람의 지각(知覺)은 이러한 귀의 마스킹 현상 등의 영향을 받게 된다. 심리음향 분야는 이러한 청각적 지각특성 특히 내이(內耳:inner ear)에서의 시간 또는 주파수 성분들에 대한 해석 능력 등을 이해하는 분야로, 오디오 코딩에서는 이러한 청각적 지각 특성을 신호의 부호화에 적용하여, 청각적으로 무관한 신호 정보가 청자(聽者)에게 감지되지 않는 다는 것을 이용하여 신호의 압축을 꾀하고 있다. 이러한 지각적으로 무관한 정보는 신호 분석 시 심리음향적 특징들, 즉, 절대 가청 한계(absolute hearing thresholds), 임계대역 주파수 분석(critical band frequency analysis), 순시 마스킹(simultaneous masking), 시간영역에서의 마스킹(temporal masking)등을 이용하여 추정하게 된다[1][3][10][11][12][13].



〈그림 2〉 귀의 구조



〈그림 3〉 Masking effect

2. Subband Coding

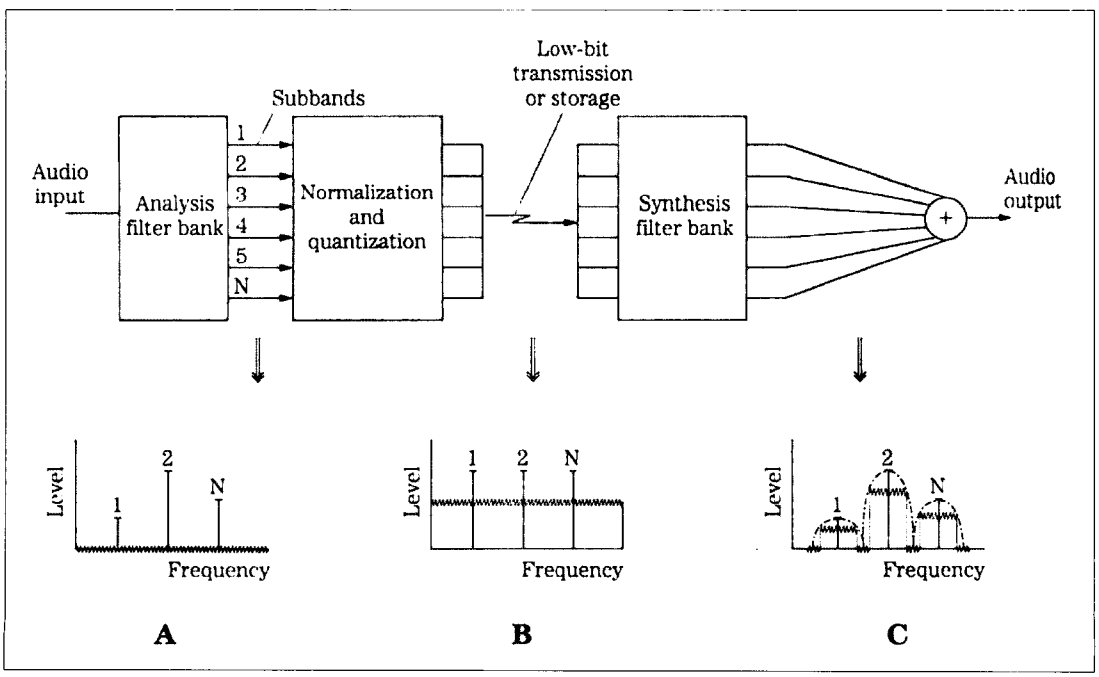
서브밴드 코딩 기법은 1980년대 초반 Bell Labs에서 처음 개발된 이래, 계속해서 발전적인 연구가 수행 되어져 왔다. 이것은 broadband signal을 나타내는 시간 영역의 오디오 데이터 블록을 필터 뱅크에 적용하여 여러 개의 다중 대역으로 분리하게 된다. 오디오 코딩에 있어 이것은 사람의 귀의 기저막에서의 주파수 분해능을 표현하는 임계대역(Critical Band)에 근사화 시키기 위한 것이다.

이러한 필터 뱅크는 임계대역에서의 주파수 응답 특성을 표현하고, 각 대역내에 양자화 노이즈를 제한하기 위해 매우 가파른 cutoff특성을 가져야만 한다(대략 100dB/Octave). 이러한 조건을 만족시키기 위해서는 디지털 필터를 통한 필터 뱅크 구성만이 가능하며, 또한 양자화 노이즈가 사람의 귀의 temporal limit일 넘지 않도록 하기 위해 처리시 지연속도가 약 3~4ms보다 낮아지도록 구현을 해야만 한다.

지각적인 투명성을 이용한 오디오 코딩 기법에서는 이러한 각각의 서브밴드에 있는 오디오 신호들을 분석하고, 심리음향 모델을 통해 계산되어진 그러한 서브밴드의 마스킹 임계치에 근거해서 각각의 서브밴드내에 있는 오디오 샘플들을 적응적으로 양자화를 하게 된다.

이 경우 양자화 노이즈는 각각의 서브밴드 내에서 증가하게 되는데, 신호의 복원 시 이러한 각각의 서브밴드 내의 양자화 노이즈는 그 밴드에 대해서만 제한이 되며, 오디오 신호 성분들에 의해 마스킹 되어지게 된다.

이때, 각 서브밴드 내의 오디오 샘플들에 대한 bit 할당은 매 입력 샘플 블록에 대해서 각 서브밴드



〈그림 4〉 subband coding with quantization noise masking

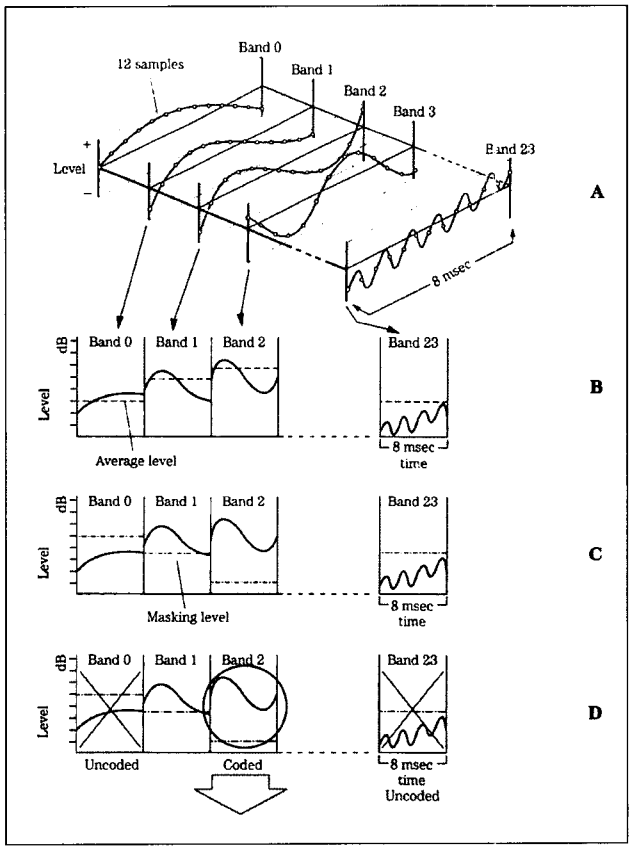
내의 신호 자체에 대한 분석과 심리음향 모델을 통해서 결정 되어 지는데, 즉, 오디오 샘플들은 오디오 신호 자체와 양자화에 의한 양자화 노이즈의 가청도(audibility)에 따라 동적으로 양자화가 이루어지게 되는 것이다.

이렇게 양자화 된 각각의 서브밴드 샘플들은 합성 필터 뱅크에 의해서 broadband 오디오 신호로 합성되어 지게 되는 것이다.

〈그림 4〉는 임의의 서브밴드를 이용한 이러한 과정을 나타내고 있다. 부호화기는 낮은 광대역 노이즈 에너지 층을 갖고 있는 광대역 오디오 샘플들을 여러 개의 서브밴드로 분리 시키고(A), 이러한 각 서브밴드 내에 있는 오디오 샘플들을 정규화 시킨 후에 심리음향 모델에 의한 정보를 이용하여 적은 bit수에 의한 양자화를 통해 broadband noise floor를 적정 수준까지 증가시키게 된다(B). 그러나 합성 필터를 통해 오디오 샘플들이 복원이 되었을 때, 이러한 각 서브밴드들은 양자화 노이즈 에너지 floor를 신호 에너지에 의해 마스킹되는 곳에 제한 하게 된다.

이와 같이 서브밴드를 이용한 지각적 투명성을 이용한 부호화기는 입력 오디오 신호들을 여러 개의 밴드로 분할 하기 위해 디지털 필터 뱅크를 사용하게 되는데, 일부 부호화기에서는 각 서브밴드내의 신호 에너지를 분석하기 위해 FFT와 같은 side processor가 사용되기도 한다. 이와 같이 계산되어진 각 서브밴드 내의 신호 에너지 값들은 각각의 밴드 내에 존재하는 신호들에 적용하게 될 전역 마스킹 임계치를 계산하기 위해 심리음향 모델에 적용되어 진다.

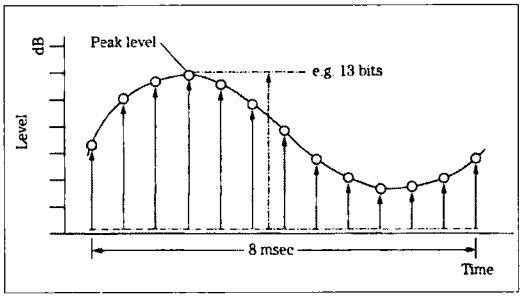
즉, 이와 같이 이러한 서브 밴드 부호화와 같은 기법들은 시간 영역의 오디오 샘플들을 부호화 하는데 있어서 좀더 선택적인 부호화 방법들을 가능하게 하는데, 특히, 부호화기에서는 각 서브밴드에 대해서 어떠한 서브밴드가 가청 가능한 오디오 신호를 포함하고 있는가를 결정하기 위해 각 서브밴드내의 평균 에너지 레벨을 분석하게 된다(A,B). 이렇게 계산되어진 각 서브밴드의 평균에너지 레벨은 각 서브밴드 내에서의 신호에너지에 의한 마스



〈그림 5〉 Subband Coding Scheme for Audio Coding

킹 레벨을 계산하는데 사용되어질 뿐만 아니라 인접 서브밴드로 부터의 마스킹 영향을 계산하는데도 사용되어 지게 된다.

그런 다음 심리음향 모델로부터 계산되어진 마스킹 레벨이 각 서브밴드에 적용이 되고(C), 이러한 마스킹 레벨과 비교하기 위해 각 서브밴드내의 peak에너지 레벨이 계산되어 진다.



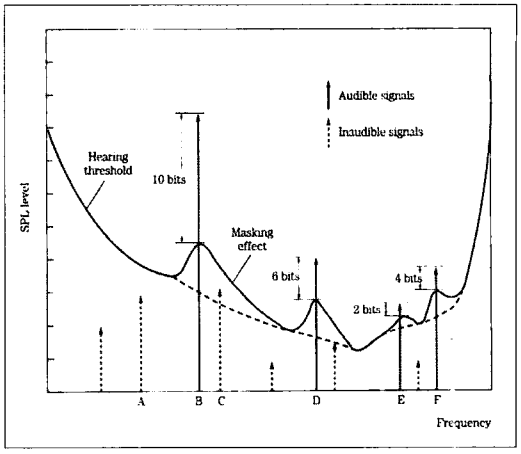
〈그림 6〉 subband sample in time domain

이렇게 계산되어진 peak에너지 레벨은 마스킹 레벨과 비교하여 부호화 되지 않을 서브밴드를 선택하고(D), 미스킹 레벨 이상의 peak레벨을 갖고 있는 서브밴드의 경우 그 레벨에 따라 bit를 할당하게 된다.

같은 개념으로 근처의 에너지가 높은 tone에 의해 마스킹 되는 tone의 경우도 부호화 되지 않으며, 인접 서브밴드의 에너지에 의해 마스킹 되어지는 서브밴드의 경우도 부호화 되지 않게 된다 [3][5][9][10][11].

3. 동적 bit 할당

양자화를 통해 bit을 할당하는 것은 가청도(audibility) 곡선을 나타내는 전역 마스킹 임계치를 넘는 신호 에너지의 세기 정도에 따른 priority schedule에 따라 각 서브밴드에 bit을 할당 시키게 된다. 예를 들어 〈그림 7〉에서 최저 가청 한계치 이하에 있는 신호 들은 부호화 되지 않게 되는데, 그림에서 신호 A의 경우 최저 가청 한계치 아래에 존재 하게 되므로 부호화 되지 않게 되고, 신호 B의 경우는 마스킹 한계치 위에 존재하게 되므로 부호화 되어지게 되나, 자신의 에너지의 영향으로 최저 가청 한계치 이상에서의 상대적인 진폭값을 낮추게 되는 마크킹 현상을 발생시키게 되어, 신호 C의 경우 이러한 신호 B의 영향에 의해 부호화 되지 않게 된다. 이때 마스킹 임계치와 신호 B 사이의 부분을 이러한 마스킹 임계치를 고려하여 요구되어지는 최소한의 bit수를 통해서 부호화를 하게 되는데, 이와 같이 양자화를 통해 bit을 할당하는 것은 SNR(Signal-to-Masking Ratio)에 의존하기 보다는 각각의 서브밴드에 대해서 계산되어지는 신호 에너지 레벨과 마스킹 임계치 사이의 차이값으로 정의되는 SMR(Signal-to-Masking Ratio)에 근거해서 각각의 서브밴드에 대해 양자화 노이즈 레벨을 마스킹 임계치 아래에 유지시킨다는 목표를 통해 bit을 할당하게 된다.



〈그림 7〉 Bit Allocation

이와 같이 bit 할당 알고리즘은 각 서브밴드 내에서의 요구되어지는 SNR을 획득하기 위해, 마스킹

영향에 잠기게 될 최적의 양자화 노이즈를 계산한다. 일반적으로 반복적인 할당을 통해 제한된 bit rate를 유지한 채 최대한 coding margin을 증가시키기 위해, 사용 가능한 여분의 bit들을 할당하게 된다. 이때 bit rate은 해당하는 신호 블록에 대해서 출력 옵션에 따라 fixed bit rate 또는 variable bit rate을 갖도록 구성 되어 진다.

이러한 bit 할당 방법으로는 크게 두 가지 방식에 기초해서 구성이 된다.

▶ **Bit-Pool approach를 이용한 방법**

주로 *fixed bit rate*을 요구할 때 사용되어지는 방식으로, 높은 SMR을 갖는 신호성분을 우선으로 bit을 할당하는 방식으로, 초기 할당 후에도 Pool에 여분의 bit가 남아 있다면 시스템의 데이터 처리용량이 허용하는 범위 내에서 가지고 있는 bit을 모두 할당할 때 까지 반복 수행하는 방식이다. 이 경우 높은 SMR을 갖는 신호가 대부분의 bit을 모두 소진하거나, 때에 따라서는, 이전에는 inaudible로 분류되었던 서브밴드가 이러한 여분의 bit들로부터 bit를 할당 받아 마스킹 임계치 밑에 있는 신호라 할지라도 실제로는 coding될 수 있어 optimal coding 보다는 다소 낮은 결과를 초래하기도 한다.

▶ **Analysis-by-Synthesis technique를 이용한 방법**

두 개의 반복 loop가 analysis-by-synthesis technique을 이용한 양자화와 부호화를 수행하게 된다. 1) Inner Loop : 신호의 주파수 계수 들에 대해서 초기에는 임의의 양자화 step size를 할당하고, 오디오 신호 샘플 블록내의 신호를 부호화 하는데 필요한 bit수를 계산하게 된다. 그 다음으로 만약 그 블록에 대해서 허용치의 bit rate을 초과할 경우 루프는 조금 더 큰 양자화 step size를 다시 할당하게 된다. 이러한 반복은 target bit rate에 도달할 때까지 수행 되어 지게 된다. 2) Outer loop : 복원된 신호에서 발생하게 될 양자화 에러를 계산하고, 임의의 band내의 양자화 에러가 마스킹 모델에 의해 허용된 에러의 범위를 초과할 경우, 그 band에 대해 양자화 step size를 줄여 bit를 할당한다. 즉, 이러한 두 개의 반복 루프를 optimal coding을 획득 할 때까지 계속 수행하게 된다[3][5][9][10][11].

4. Transform Coding & Filter Bank

서브 밴드 코딩 시스템 경우 시간 영역 기반의 샘플들을 부호화하기 위해 주파수 분석을 하는 것과는 반대로 변환 coder는 주파수 계수들을 부호화 하게 된다. 정보 이론의 관점에서 볼 때, 변환(transform)은 신호의 엔트로피를 줄임으로써 효율적인 부호화를 제공하게 된다. 그러나 변환의 블록이 커질수록 높은 주파수 분해능(spectral resolution)을 제공하지만, 그로 인해 시간 분해능(temporal resolution)을 잃게 된다. 예를 들어, long block의 경우 신호의 급격한 변화(transient)에 앞서 pre-echo를 발생시키는 결과를 초래 할 수 있다. 이것으로 인해 대부분의 부호화기는 시간 분해능(temporal resolution)을 개선 시키기 위해 시간 축에서 50%씩 연속적으로 block들을 중복시킨다. 이것은 주파수 영역에서의 블록간 변화를 줄이게 된다. 또한 일부 design에서는 이러한 블록 인파를 signal condition에 따라 적응 시킨다. 이와 같이 시간 영역의 샘플들이 주파수 영역으로 변환이 되어서 spectral 계수를 산출 하게 되는데, 대게의 경우 512, 1024, 또는 그 이상의 개수 들로 구성 되어진 주파수 계수들은 임계대역analysis을 emulation하는 대략 32밴드들로 그룹지어 진다.

이와 같이 변환 operation은 사람의 청각 시스템의 기저막에서 주파수를 분석하는 방법을 근사화 시킨 것으로, 주파수 분석된 계수들은 심리음향 모델의 결과에 따라 양자화 된다. 즉, 마스킹 되어진 성분들은 제거되고, 양자화는 신호 성분의 가청도(audibility)에 따라 이루어 지게 된다.

일반적으로 모든 오디오 압축 알고리즘이 일종의 time-frequency analysis block에 의존해서 시간

영역의 입력 샘플로부터 양자화와 지각적인 왜곡정도에 대한 정보를 얻고 있다. 이와 같은, time-to-frequency mapping에 일반적으로 사용되어 도구 중 하나가 필터 뱅크이다.

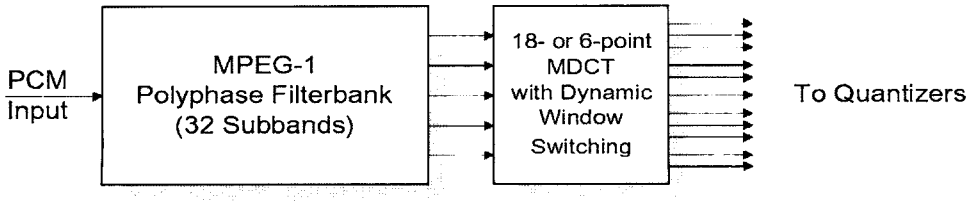
오디오 코딩을 위해 이러한 필터뱅크로부터 요구되지는 것들은

- *signal adaptive time-frequency tiling*
- *low-resolution "critical-band" mode, e.g., 32 subbands*
- *high-resolution mode, up to 4096 subbands*
- *efficient resolution switching*
- *good channel separation*
- *strong stopband attenuation*
- *perfect reconstruction*
- *availability of fast algorithm*

등이다.

그러나 대부분의 필터 뱅크는 필터의 transition band의 finite width, band overlap, 1/N subsampling등으로 인해 aliasing을 발생 시키기 때문에 이상적인 성능을 발휘 하지 못한다. QMF(Quadrature mirror filter)의 경우 N개의 overlapping 서브밴드로부터 original signal을 aliasing 없이 복원 시키는 특성을 갖고 있다. 그러나 완벽한 복원은 일반적으로 N=2인 경우, 즉, 두 개의 equal-width를 갖는 서브밴드를 생성하는 경우로 제한되므로, 이러한 QMF process를 반복 수행함으로써 더 많은 대역으로 나눌 수 있게 되는데, 이러한 과정은 tree structure를 통해 수행 될 수 있으나, processing delay를 증가 시키는 단점이 있다.

그밖에 다양한 필터 뱅크가 지각적으로 투명한 부호화 기법(perceptual coding)에 사용되어 진다. polyphase filterbank는 위상 정보의 상호관계를 통해 효율적인 구현을 가능하게 하는 equal bandwidth를 구성하게 된다. 이러한 Polyphase filter는 equal bandwidth를 갖는 서브밴드와 높은 stop-band attenuation을 갖는 좋은 주파수 분해능(frequency resolution)을 가지고 있어 aliasing을 통제하기에 좋다. MPEG-1 Layer I과 II의 encoder의 경우 이러한 32-band polyphase filter를 사용하게 된다. 그 외에 sine-taper window를 이용한 DFT(discrete Fourier transform)와 DCT(discrete cosine transform)도 동일 간격의 주파수 대역을 구성 할 수 있으니 주파수 성분의 수가 시간 샘플의 수보다 커져, critical sampling을 제공하지 못한다. MDCT (modified discrete cosine transform)의 경우는 "time domain aliasing cancellation"을 통해 높은 주파수 분해능을 가지면서도 critical sampling을 가능하게 해, 좋은 효율성을 나타낸다. Hybrid filters는 적당한 수준의 복잡도를 가지면서도, 각기



〈그림 8〉 MPEG-1 Layer 3에서의 Hybrid Filter Bank의 구성

다른 주파수에서 각기 다른 주파수 분해능을 제공한다. 예를 들어 MPEG-1 Layer III 부호화기가 polyphase filter bank와 MDCT로 구성된 hybrid filter를 사용한다[3][5][9][10][11].

Ⅲ. MPEG-1 Audio

MPEG-1 표준안(ISO/IEC 11172)은 디지털 저장 매체, 즉 CD(Compact Disc), DAT(Digital Audio Tape), magnetic hard disc 등을 대상으로 최대 1.5 Mbits/s의 전송률로 고 품질의 비디오 · 오디오 복호화 신호를 얻을 수 있도록 제정되었다.

MPEG-1 표준안은 다음과 같이 5개 부분으로 구성되어 있다.

- ▶ 11172-1 : systems
- ▶ 11172-2 : video
- ▶ 11172-3 : audio
- ▶ 11172-4 : conformance
- ▶ 11172-5 : software

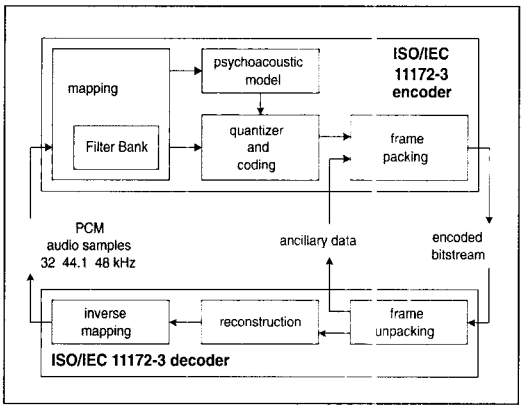
systems부에서는 동영상 재생시 오디오와 비디오의 동기를 맞추기 위해 오디오 및 비디오 신호를 조합하는 방법에 관하여, 비디오와 오디오부에서는 각기 비디오와 오디오 신호의 효율적인 부 · 복호화 방법에 관하여, conformance부에서는 응용분야에 MPEG-1 표준안에 적합한지를 검증하기 위한 방법에 관하여, 소프트웨어부에서는 검증 모델에 기초한 실험 프로그램에 관하여 기술하고 있다.

1. MPEG-1 오디오

MPEG-1 오디오 표준안에서는 32 kHz, 44.1 kHz, 48 kHz의 표본화 주파수를 사용하여 최대 2 채널을 지원하며, 다양한 응용들에 대한 유연성을 제공하기 위해 3개의 계층(계층 I, II, III)으로 구분된다. 계층에 따라 복잡도(complexity), 처리 지연 시간(delay), 압축 성능이 높아지게 된다. 3개(32 kHz, 44.1 kHz, 48 kHz)의 표본화 주파수를 사용하여 계층에 따라 고정 비트율 또는 임의 선택(free format) 비트율을 적용할 수 있다. 각 계층 I, II, III의 최대 고정 비트율은 448 kbits/s, 384 kbits/s, 320 kbits/s이며, 임의 선택(free format) 비트율 적용시 복호화기에서는 각 계층의 최대 고정 비트율까지 지원하면 된다.

계층 II의 복호화기는 계층 I, II 복호화기에서 부호화된 비트열을 복호화 할 수 있어야 하며, 계층 III의 복호화기는 계층 I, II, III 부호화기에서 부호화된 모든 비트열을 복호화 할 수 있어야 한다. MPEG-1 오디오는 프레임 단위로 처리되기 때문에 복호화시 비트열에서의 빠른 속도의 전 · 후방 및 임의의 위치로의 이동 처리가 가능한 특징을 가진다.

〈그림 9〉는 MPEG-1 오디오 표준안의 부 · 복호



〈그림 9〉 MPEG-1 오디오 구성도

화기의 구성도를 도시한다.

부호화기에 들어온 입력신호는 필터 뱅크를 통과하여 subband 샘플로 되어진다. 심리음향 모델에서는 마스킹 임계치를 얻어 양자화에 쓰이는 비트 할당 정보를 주게된다. 양자화된 데이터와 부가정보를 가지고 비트열을 생성한다. 복호화기에서는 이러한 비트열을 입력으로 받아 각 subband 샘플들을 복원하고 합성 필터를 통과시켜 복호화된 신호의 PCM 샘플(복원신호)를 얻는다.

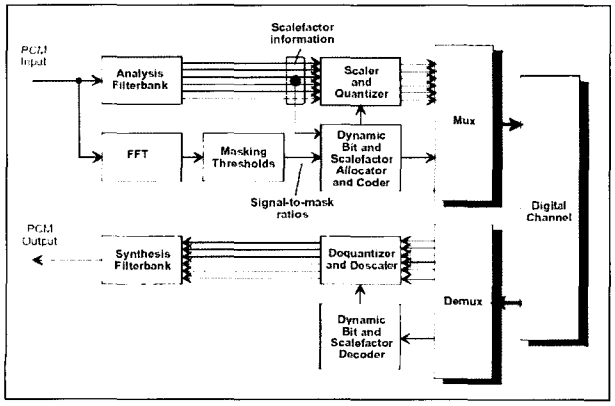
MPEG-1 오디오 표준안에서는 지각적 중복성 제거와 통계적 중복성 제거를 이용하여 효율적인 부호화를 제공한다. 지각적 중복성 제거란 이전에 설명된 심리음향 모델을 적용하여 사람의 청각적 특성에 의거 인지할 수 없는 신호를 제거하는 것을 의미하며, 통계적 중복성 제거란 32개의 등 간격 필터 뱅크(Filter Bank)를 이용한 subband 부 · 복호화 방식과 양자화기와 그룹핑(Grouping) 또는 허프만(Huffman) 부 · 복호화 방식 등을 이용하여 불필요한 성분을 제거함을 의미한다.

입력된 PCM 신호를 이용한 32 밴드의 subband 부호화를 기본으로 하여 부가적인 FFT(Fast Fourier Transform)를 통해서 청각 특성을 이용한 마스킹 임계치와 신호의 크기로부터 얻어지는 신호대 마스크 비(SMR)를 기본으로 비트 할당이 이루어짐으로써 귀가 인지할 수 없는 모든 성분을 버리게 되므로 인간의 청각 특성상 원음과 거의 구별할 수 없는 복원 신호를 얻을 수 있으면서 큰 압축율을 얻을 수 있다.

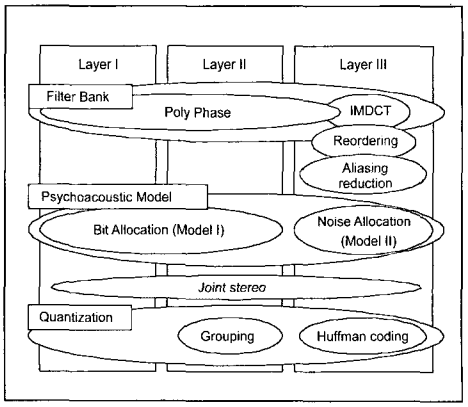
계층 I에서는 약 192 kbit/s에서 고 음질을 제공해주고 부호화기도 간단하며 처리 시간 지연도 적다. 여기서는 32개의 각 밴드 신호를 12 샘플씩 총 384 샘플을 프레임으로 비트 할당을 해준다.

계층 II에서는 계층 I을 확장한 것으로서, 좀더 복잡해진 부호화기로 더 큰 압축을 얻을 수 있다. 부가정보와 샘플들을 부호화 할 때 사용 비트를 줄이기 위하여 양자화된 샘플들에 그룹핑(Grouping)을 적용한다. 1152 샘플을 한 프레임으로 비트 할당 방식(Bit Allocation)을 적용하며 128 kbits/s 보다 작은 비트율에서 응용된다. 계층 I과 II는 비트 심리음향 모델 I이 적용되며, 비트 할당 방식으로 처리된다.

계층 III에서는 1152개의 샘플들을 한 프레임으로 보다 세밀한 주파수 해석을 위해 서브밴드(subband 또는 poly phase) 및 MDCT(Modified



〈그림 10〉 MPEG-1 Layer I, II 오디오 구성도

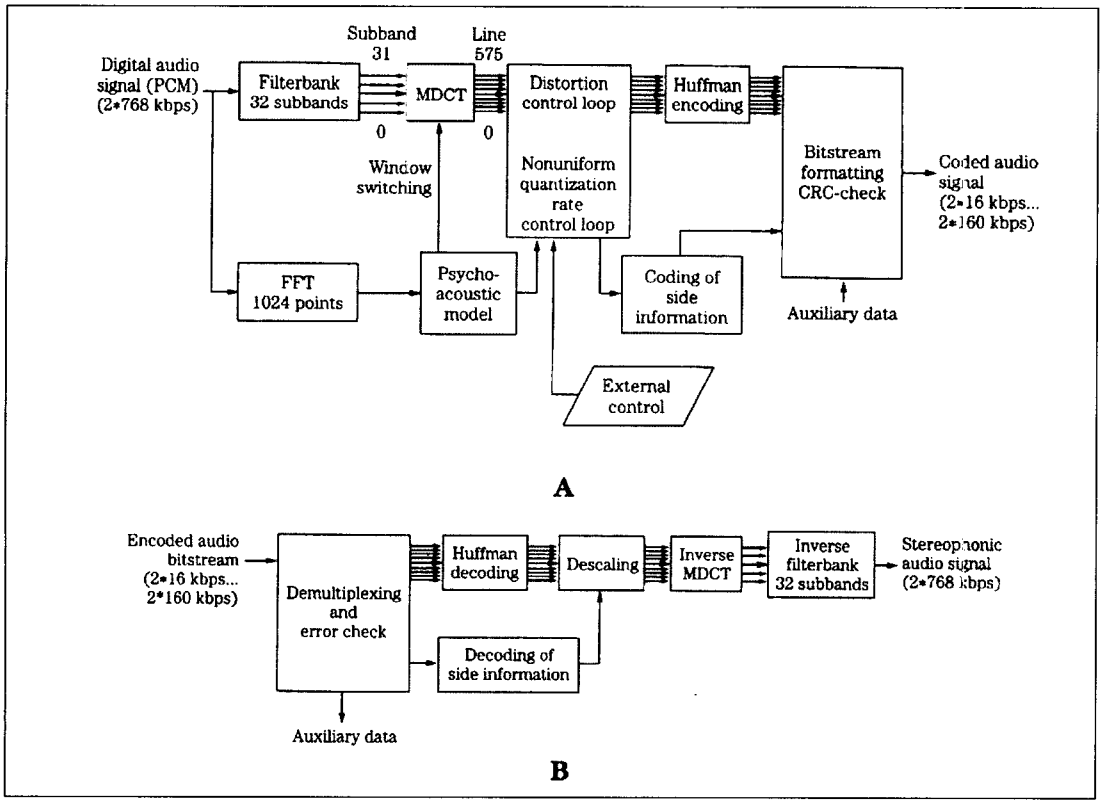


〈그림 11〉 계층별 처리 방법

Discrete Cosine Transform) 필터 뱅크와 심리음향 모델 II를 이용한 잡음 할당 방식(Noise Allocation)을 적용하여 비균일 양자화기를 사용하며, 허프만(Huffman) 부호화 방법을 사용한다. MDCT 필터 뱅크 적용으로 인해 서브밴드 필터 뱅크와의 차수를 재 정렬(reordering)하고, 발생된 엘리어싱을 감소시키는 과정이 추가된다. 허프만 부호화와 같은 가변장 부호화 방식을 사용하기 때문에 지정된 프레임 크기에 제약을 극복하기 위해서 이전 프레임의 여분의 비트를 이용 현 프레임의 정보를 저장할 수 있는 비트열 구조를 가진다. 따라서 압축 효율성은 증대하나 복잡한 연산을 필요로 하므로 복잡도(complexity) 및 처리 지연 시간(delay) 또한 증가한다. 약 64 kbps에서도 거의 CD 음질에 가까운 복원 신호를 만들 수 있다.

MPEG-1 오디오는 모노(mono), 스테레오(stereo), 이중 모노(dual mono), 채널간의 중복성을 제거한 결합 스테레오(joint stereo) 모드로 구성된다. 고주파 신호에 민감하지 않은 인간의 청각적 특성을 이용한 결합 스테레오는 계층 I, II에서 intensity 스테레오를 의미하며, 계층 III에서 intensity/MS 스테레오를 의미한다.

intensity 스테레오는 한 쌍의 두 채널중 고주파 영역에 대하여 한 채널(L)을 이용하여 다른 채널(R)의 변환정보로 구성하여 보다 효율적인 부호화



〈그림 12〉 MPEG-1 Layer III 오디오 구성도

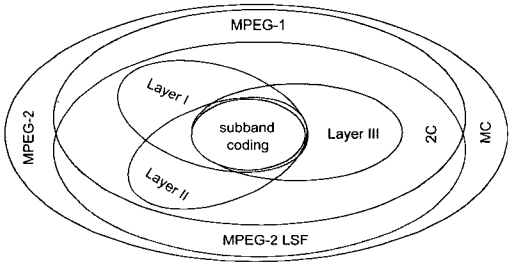
를 제공하는 방법이며, MS 스테레오는 2-point 직교 변환의 간단한 방법으로 두 신호의 합과 차로서 구성하는 방법이다[2][3][5][6][9][16].

Ⅳ. MPEG-2 Audio

MPEG-2 표준안(ISO/IEC 13818)은 방송용도의 HDTV(High Definition TV), DVD(Digital Versatile Disk) 및 극장에서 사용되는 5.1 채널의 오디오 등을 대상으로 약 4~15 Mbits/s의 전송율로 고 품질의 비디오 · 오디오 복호화 신호를 얻을 수 있도록 제정되었다.

1. MPEG-2 BC(Bcakward Compatible)

〈그림 13〉에서와 같이 MPEG-1 오디오 알고리즘을 기반으로 하여 확장 구성되며, 5.1 채널 즉, Left, Right, Center, Left Surround, Right Surround, Low Frequency Enhancement 채널을 지원하는 등의 다중 채널과 여러 언어를 추가하는 다중 언어를 제공함으로써 방송용도로 사용되고 있다. 추가되는 다중 언어 및 부가기능 채널은 최대 7개 채널까지 지원된다. 다중 채널 · 언어를 제공하는 의미에서 이를 MPEG-2 MC (MultiChannel)라 호칭한다[2 :p.44]. 오디오 채널 당 64 kbits/s 미만의 매우 낮은 비트율에서 좋은 성능을 유지하도록 MPEG-1을 확장하여 기존 샘플링 주파수(32, 44.1, 48 kHz)의 1/2(16, 22.05, 24 kHz)을 사용하는 MPEG-2 LSF(Low Sampling Frequency) 또한 제공된다. MPEG-2 오디오는 다중의 연속적인 계층들의 구성과 역방향 호환성(Backwards compatibility, 기본 2채널 : Lo, Ro)을 유지하고 하는 가변(scalable) 코딩을 지원한다. 역방향 호환성의 의미는 MPEG-1 오디오 복호화기에서도 MPEG-2 오디오 비트열을 복호화 할 수 있음을 의미하며 MPEG-1 오디오 비트열을 MPEG-2 오디오 복호화기에서 복호화할 수 있다. 반면, 이러한 역방향 호환성을 가지지 않으며 보다 좋은 성능을 구현하고자 추가된 방식이 MPEG-2 AAC(Advance Audio Coding)이다. MPEG-2 오디오는 계층적 구조를 가짐으로 인해서 복호화기의 성능에 따라 선택적으로 채널의 수를 줄여서 복호화할 수 있는 하향 호환성(Downwards compatibility)을 제공한다.



〈그림 13〉 MPEG-1 오디오와 MPEG-2와의 관계[2]

MPEG-2 오디오에서는 신호의 중복성 제거를 위해서 다음의 방법들을 사용한다.

1) Dynamic Cross Talk

음장에 영향이 적은 채널의 주파수 성분을 다른 채널에서 대체 사용하며, 전체 대역이나 일부 subband에서 사용가능하다. 스테레오 신호중 공간 지각과 관계없는 부분을 차폐되지 않더라도 복원 신호의 음질 및 음원의 위치 추정에는 무관하므로 스테레오 신호 중 어느 신호라도 스테레오 와 무관한 성분들은 음질 및 입체 음향감에 영향을 주지 않고도 임의의 스피커에 의해서 또는 여러 스피커에 적절한 배치에 의해서 재생가능하다. 서브밴드 단위로 수행되며 비트 할당 정보와

부호화된 서브밴드 샘플은 비트열에 포함되지 않고 다른 전송 채널로 전송된 서브밴드 샘플로부터 복사하여 사용된다.

2) Dynamic channel switching

채널간의 직교성을 높이기 위해 5개 채널에서 dynamic range가 작은 3개 채널을 보내고 Lo, Ro를 전송한다. 신호를 단지 행렬 변환만 시켜주고 이를 전송 시켰을 때는 복호화 단에서 역 변환시켜 원 신호를 복원시켰을 때 누적 잡음들이 서로 합쳐진 가청 잡음을 형성, 복호화 후의 이러한 가청 잡음을 줄이는 한 방법으로서 확장 채널을 유동적으로 선택하는 전송 채널 변환 방법을 도입한다. 이것은 복호화된 채널의 잡음이 주로 전송된 채널의 신호 크기에 좌우된다는 사실에 근거, 역변환 시키는 과정은 MPEG-1 호환 채널(Lo, Ro)에서 확장 채널의 성분을 제거하는 것과 같으며, 이러한 과정이 가청 잡음을 덜 형성시키기 위해서는 제거되는 확장 채널이 MPEG-1 호환 채널과 비교하여 상대적으로 신호 크기가 작아야한다.

3) 예측(Prediction)

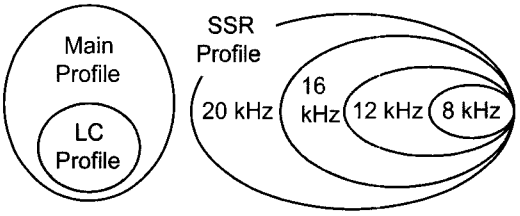
채널 신호간의 연관성을 이용하여 다른 채널의 신호를 예측하여 예측오차와 예측 계수만을 전송한다. 이때 모든 계산은 프레임 단위 수행한다.

4) Phantom Coding

Center 채널의 고주파 성분을 L, R 채널로 합성하여 대신 전송한다. 만약 비트 수가 모자란다면 Center 채널에 대해 Phantom coding을 사용하여 음질에는 큰 영향을 주지 않고서도 상당한 이득을 얻을 수 있다. Center 채널의 고주파 부분은 3dB 감소하여 L, R 채널에 합성된다[3][7][9][16].

2. MPEG-2 AAC(Advanced Audio Coding)

앞의 *1. MPEG-2 BC* 에서 언급한 바와 같이 MPEG-2 오디오의 특징 중의 하나인 역방향 호환성을 배제하여, MPEG-2 MC보다 낮은 비트율에서 좋은 음질을 얻기 위한 부호화 방식 AAC가 MPEG-2 표준안에 추가로 제정되었다. 초기에는 MPEG-1과의 역방향 호환성을 배제하였다는 의미에서 MPEG-2 NBC(Non-Backward Compatible)로 호칭되었으며, ISO/IEC 13818-7로서 차후 MPEG-4의 오디오 부분인 T/F 부호화기의 기본이 된다. MPEG-2 AAC에서는 넓은 범위의 샘플링율(8~96kHz)과 비트율을 제공하며, 1~48 오디오 채널, 최대 15개의 LFE 채널, 다중언어방식(multilanguage capability)를 지원한다. MPEG-2 BC 보다 개선된 압축율을 제공한다.



〈그림 14〉 MPEG-2 AAC 프로파일

〈그림 14〉에 도시한 바와 같이 MPEG-2 AAC에서는 다양한 용도에 적용하기 위하여 시스템의 복잡도와 지원 기능에 따라 제공하는 3가지 프로파일은 다음과 같다.

1) 메인 프로파일(Main profile)

3개의 프로파일 중에서 주어진 비트율에서 최상의 음질을 제공하는 프로파일이며, 이득제어(Gain Control)를 제외한 MPEG-2 AAC의 모든 방식이 사용된다. 많은 메모리와 연산량이 필요하며, 저 복잡도(LC) 프로파일로 부호화된 비트열을 복호화 할 수 있어야 한다.

2) 저복잡도 프로파일(LC profile, Low Complexity profile)

메인 프로파일에서 예측기와 coupling 채널을 사용하지 않으며, 시간 영역 잡은 변형(TNS, Temporal Noise Shaping) 방식의 차수을 제한하여 메모리, 연산량, 압축율이 메인 프로파일에 비해 적다.

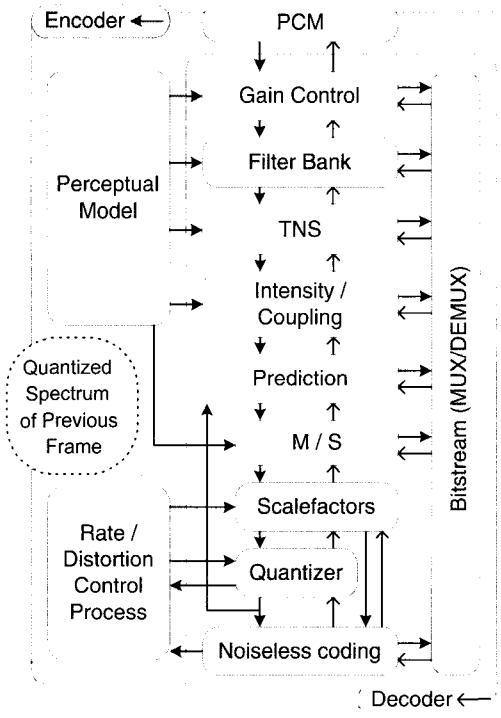
3) 가변 샘플링율 프로파일(SSR profile, Scalable Sampling Rate profile)

이득 제어와 제한된 TNS의 차수 및 대역폭을 사용하는 반면, 예측기와 coupling 채널은 사용하지 않는다. 4

PQF(Polyphase Quadrature Filter)를 사용하여 제일 낮은 서브 밴드에는 이득 제어를 사용하지 않는다. 오디오 대역폭이 감소된 경우 복잡도에 따라 주파수를 가변 시킬 수 있다.

MPEG-2 AAC에서는 비트열의 구조를 ADIF(Audio Data Interchange Format)과 ADTS(Audio Data Transport Stream)으로 구성할 수 있다. ADIF는 하나의 디스크 파일에서 복호화하는 방식 처럼, 1024 샘플 단위로 처리한 부호화 정보들을 이용하여 순차적으로 구성된 비트열에 대해 하나의 헤더를 포함한다. ADTS는 MPEG-1과 MPEG-2 오디오에서 사용되는 문법과 유사하여, ISO/IEC 11172-3 복호화기에서는 계층 Ⅳ 비트열인 것으로 인식한다.

〈그림 15〉에서는 MPEG-2 AAC의 부·복호화기의 구성도를 도시한다. 복호화기는 부호화기의 역순으로 처리되며 복호화기의 필수 구성 요소들을 명암으로 도시하였다. 〈그림 15〉에서의 화살표는 데이터 또는 제어 정보의 흐름을 표시한다. MPEG-2 AAC에서는 가장 기본이 되는 필수 요소와 보다 높은 성능을 제공하는 선택적 요소들로 구성되어 있으며, 음질과 비용의 trade-off를 고려해야 한다.



〈그림 15〉 MPEG-2 AAC 부·복호화기 구성도

각 부분에 적용되는 방식들은 다음과 같은 역할을 한다.

1) **이득 제어부**에서는 PQF(Polyphase Quadrature Filter), 이득 검출기 및 변환기가 사용된다. 입력된 시간 영역의 신호와 윈도우 시퀀스를 이용 이득 제어 정보와 MDCT 윈도우의 길이와 동일한 이득 제어된 신호들을 생성하게 된다. PQF를 이용 4개의 등간격의 주파수 밴드들로 구분하여 각 PQF 밴드의 계수를 생성한다. 이득 검출기에서는 비트열 문법에 적합하도록 이득 변화의 수와 이득 변화 위치 및 레벨의 색인을 구성하여 이득 제어 정보를 생성하며, 한 프레임의 지연 시간으로 인해 이전 프레임의 이득 제어 정보를 의미하게 된다. 이득 변환기는 각 PQF 밴드에 대하여 각 신호 밴드의 이득을 조정하게 된다. 복호화기에서는 이득 보상 제어를 통해 프리에코를 감소시켜 음질을 개선한다. 신호가 작은 부분에서 큰 부분으로 변하는 천이 부분이 존재하는 경우 DCT, TDAC 등의 블록 부·복호화 적용 시 뒤의 큰 신호가 앞의 작은 신호에 영향을 미쳐 앞의 신호가 증가되는 오차가 생기며, 앞의 신호가 작으므로 오차가 조금만 증가된다면 인간이 인지할 수 있을 정도의 왜곡이 발생하게 되는데 이를 프리 에코라 한다.

2) **필터 뱅크(Filter Bank)부**는 가장 필수적인 처리 과정으로써 시간 영역의 신호와 시간-주파수 영역의 신호로의 변환을 한다. AC-3에서 사용되는 시간 영역 엘리어싱 제거(TDAC; Time-Domain Aliasing Cancellation) 기법을 사용한 MDCT(Modified Discrete Cosine Transform)을 적용한다. 2048 또는 256 샘플들의 단위로 처리되는 블록은 50%가 overlap-add 처리되며, 시간 영역의 샘플들

을 블록 단위로 윈도우 함수에 의하여 변형시킨 후 MDCT를 수행한다. 윈도우 함수는 필터 뱅크의 주파수 응답에 중요한 영향을 주게 되므로, 입력 신호의 특성에 적합하도록 윈도우 함수를 변경할 수 있도록 설계되어 있다. 윈도우의 모양은 다양한 입력 신호에 대해 필터 뱅크가 스펙트럼 성분을 효율적으로 분석할 수 있도록 부 · 복호화기에서 변형된다. 장변환(2048 샘플)은 복잡한 스펙트럼을 가진 정상 상태 신호에 관하여 부호화 효율을 향상시키는 장점을 가지나, 블록내의 샘플의 값이 작은 값에서 큰 값으로 변하는 천이 구간에 대해서는 프리 에코등이 발생하는 문제를 피하기 위하여 단변환(256 샘플)을 한다. 그러나 단변환은 주파수 분해능이 낮기 때문에 정상 상태 신호의 부호화 효율이 좋지 않다.

3) **TNS(Temporal Noise Shaping)부**는 필터 뱅크의 특성을 변형시키는 역할, 즉 부호화된 잡음의 정교한 시간 구조로 제어를 한다. 지정된 주파수 범위에서 스펙트럼의 MDCT 계수들을 선형 예측 부호화(LPC)를 적용하여 포락선(temporal envelop)을 평탄화(flattened, 부호화) 또는 역평탄화(복호화)해주는 필터링을 한다. 이는 필터 뱅크와 함께 사용되어 입력 신호의 시간 · 주파수 특성을 적응적으로 처리함으로써 보다 세밀한 제어를 수행한다.
부호화 효율을 높이기 위한 스테레오 신호의 처리 방법으로 Intensity 스테레오와 MS 스테레오를 지원한다. MPEG-1, 2 MC에서 사용된 것과 유사 하다.

4) **Coupling 채널**은 청각이 고주파에서 두 개의 인접한 주파수에 대한 방향성을 독립적으로 인지할 수 없는 특성을 이용하여 상반된 채널 경계에서 채널 스펙트럼이 유사할 때 일반화된 intensity 스테레오 부호화를 적용하기 위해 사용되고 한 개의 음원 객체를 스테레오 음원으로의 동적 구성을 제공하기 위해 사용된다.

5) 중복성 제거를 향상시키기 위해 사용되는 **예측기부**는 연속적인 프레임들의 필터 뱅크의 출력인 스펙트럼 데이터들의 자기 상관성(auto-correlation)을 이용한 2차 역방향-적응 lattice 구조의 선형 예측 방법을 사용하여 프레임간의 채널(or 모노)에 적용한다. 부 · 복호화기에서 동이한 예측 방법을 사용함으로써 예측 계수만을 전송한다. 예측기 매개변수들은 프레임 단위를 기반으로 한 현재 신호의 통계적 특성에 적합하기 위해서 LMS 적응 알고리즘을 사용한다.

6) **스케일 인자(scalefactor)부**에서는 심리 음향 모델을 기반으로 스케일 인자를 이용하여 스펙트럼(MDCT) 계수들을 장 프레임과 단 프레임으로 구분하여 각 49, 14개의 대역 밴드로 분류하여 대역별로 정규화 시키게 된다. 이때 각 밴드마다 정규화를 위해 사용되는 계수들을 스케일 인자라 호칭하며, 장 프레임의 경우 49개의 밴드로 분할하여, 각 밴드별 스케일 인자를 조정하게 되므로 기존의 MPEG-1, 2 MC 보다 정확한 제어가 가능하다.

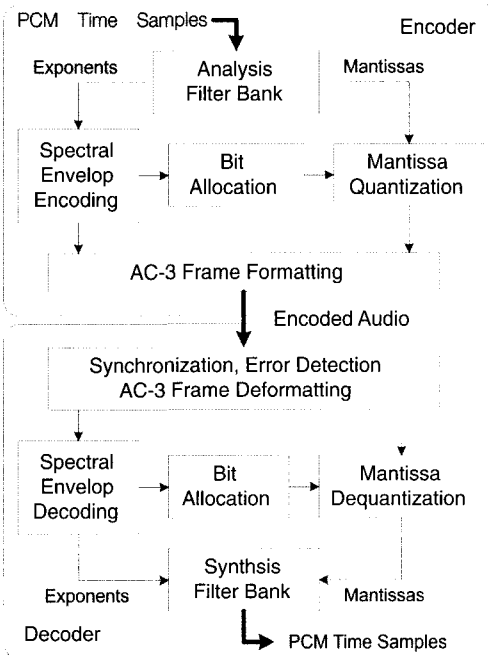
7) **양자화부**에서 심리 음향 모델을 기반한 각 밴드의 양자화 비트 수 할당에 의거 스케일 인자와 정규화된 스펙트럼 데이터를 비선형 양자화 한다.

8) **Noiseless 부호화**에서는 양자화된 출력을 최대 값을 이용한 허프만(Huffman) 부호화 방식으로 압축한다. 스펙트럼 클리핑 등의 양자화 스펙트럼 중에 진폭이 큰 부분과 해당되지 않는 부분으로 분리하여 처리함으로써 효율적인 부호화를 제공한다[8][9][16].

Ⅴ. Dolby AC-3

미국의 Dolby사의 AC-3 오디오 압축 알고리즘은 MPEG-2 Audio 부호화에 대응되는 부호화 방식으로 MPEG의 경우와 마찬가지로, AC-1(1987)와 AC-2(1989) 부호화 방식에 이어 1995년 미국 HDTV(High Definition Television) 표준 안으로 제정되었다.

이러한 Dolby AC-3부호화 방식은 32 kHz, 44.1 kHz, 48 kHz의 표본화 율 사용하여, 1~ 5.1 채널의 다중채널을 32 kbps~640 kbps의 비트 율로 부호화할 수 있으며, 최대 24비트까지의 디지털 오디오 신호를 입력으로 사용할 수 있다. 일반적으



〈그림 16〉 Dolby AC3 Encoder/Decoder

로 5.1 채널에 대하여 384 kbps의 비트 율로 사용한다.

이러한 AC-3 알고리즘은 〈그림 16〉에서와 같이 TDAC(Time Domain Aliasing Cancellation)을 통한 MDCT(Modified Discrete Cosine Transform)을 사용하여 시간영역의 신호를 주파수 영역의 신호로 변환하고, 이를 부동 소수점(floating-point) 표현 방식인 지수(exponent)와 가수(mantissa)로 분리한다. 이러한 두 가지 신호(exponent, mantissa)를 심리음향 모델을 기반으로 각각 양자화하며, 압축 효율을 높이기 위해 채널간의 중복성을 제거하는 커플링(coupling) 방법을 사용한다.

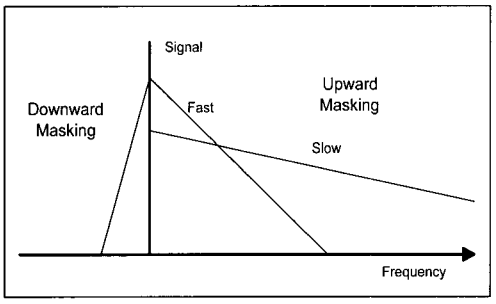
또한 AC-3는 5.1 채널로 부호화된 비트열을 복호화기에서 전송환경 및 복호화 성능 등의 제약 조건에 적응적으로 모노 또는 스테레오 등으로 복호화할 수 있도록 하향 호환성(backward compatibility)을 제공 하며, 균일하게 라우드니스(loudness)를 조정하는 대화 레벨 조정(dialog level control)기법을 사용하여 채널간의 발생하는 라우드니스의 불균형 문제를 해결하고 있다. 또한 동적 가청 영역 조정(dynamic range control)을 통하여 복호화 시, 보다 다양한 환경에 적합하게 대응할 수 있도록 하고 있다. 입력 신호로는 최대 24 bits까지의 PCM 신호를 사용할 수 있으며, 3 Hz보다 적은 DC 주파수 성분은 전처리 과정을 통해 필터링을 한다. 또한 LFE(Low Frequency Enhancement) 채널이 사용되는 경우에는 120 Hz이하의 신호들은 LFE 채널에서 처리하므로 이를 제거하기 위해 저주파수 필터링을 거쳐 MDCT의 입력으로 사용된다.

이렇게 MDCT를 통해 주파수 계수로 표현된 신호의 성분들은 지수(exponent)와 가수(mantissa)로 성분으로 분리하여 각각 양자화를 취하게 되는데, 지수의 경우는 고정 양자화 방식을, 가수의 경우는 가변 양자화 방식을 사용하여 각각의 신호 성분들에 bit를 할당하게 된다. 각각의 신호 성분에 의한 부호화 방식을 살펴 보면,

1) 지수 신호 성분은 전체 신호의 스펙트럼을 표현하는 스펙트럼 포락선(spectral envelope) 부호화 방식에 의한 매개변수를 이용하여 부호화하며, 입력 오디오 신호의 3가지 천이 구간(D15, D25, D45)에 따라 각기 다른 방식으로 부호화 하는데,

- *D15 : 천이 구간이 없는 구간에 대해 적용하며, 모든 변환 계수에 대한 지수를 부호화(2.33 bits/sample)한다.*
- *D45 : 천이 구간인 경우에 적용하며, 4개의 변환 계수 당 한 개의 지수를 부호화(0.58 bits/sample)한다.*
- *D25 : 상기 두 구간에 해당되지 않는 경우 적용되며, 2개의 변환 계수 당 한 개의 지수를 부호화(1.16 bits/sample)한다.*

2) 가수 신호 성분은 심리음향 모델에 기반한 비트 할당 방식에 의한 양자화 방식을 취하게 되는데, 참고적으로 AC-3에서는 심리음향 모델의 연산량을 감소시키기 위해 신호 에너지의 지각적인 측면에서의 에너지 확산 현상을 나타내는 마스킹 곡선을 다음과 같이 근사화 시켜서 사용한다.

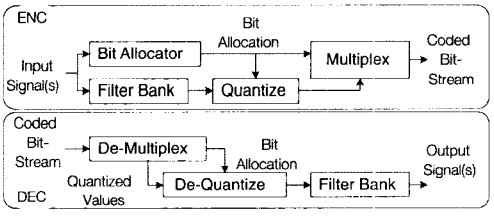


〈그림 17〉 AC-3에서 적용하고 있는 마스킹 곡선

사용하는 비트 할당(bit allocation) 방식에는

1) 순방향 적응 비트 할당 방법

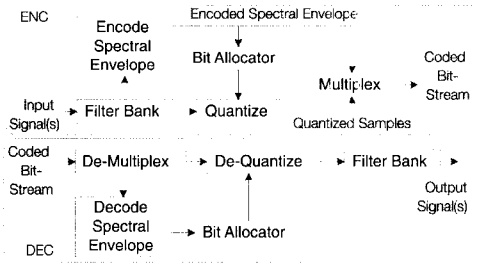
순방향 적응 비트 할당 방식은 주로 MPEG 알고리즘에서 사용하는 방식으로 부호화기에서만 비트를 할당하기 때문에 부호화기에 비해 낮은 복잡도의 복호화기를 사용할 수 있다.



〈그림 18〉 순방향 적응 비트 할당(forward adaptive bit allocation)

2) 역방향 적응 비트 할당 방법

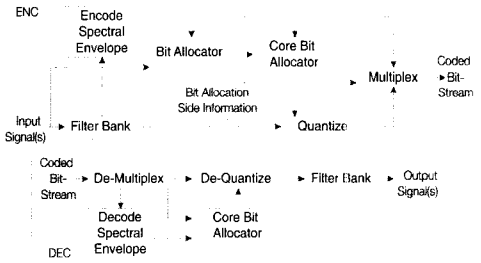
AC-2에서 사용되는 방식으로 부/복호화기에서 모두 스펙트럼 포락선(Spectral Envelope)에의한 비트 할당을 함으로써 스펙트럼 포락선에 대한 매개 변수들만을 전송하여 압축효율을 개선시킬 수 있으나 복호화기의 복잡도가 부호화기의 복잡도만큼이나 높아지게 되는 문제점을 가진다.



〈그림 19〉 역방향 적응 비트 할당(Backward adaptative bit allocation)

3) 결합 순/역방향 적응 비트 할당 방법들로 구분할 수 있다.

AC-3에서는 이러한 순방향 적응 비트 할당과 역방향 적응 비트 할당의 장단점을 보완하기 위해 이들을 조합한 결합 순/역방향 적응 비트할당을 사용하여 복호화기의 비트 할당 방식을 단순화 하였다.



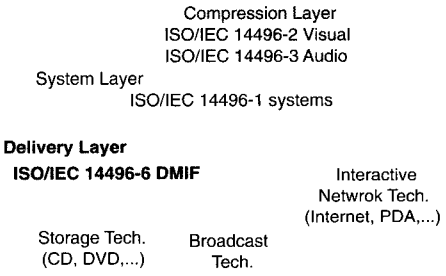
〈그림 20〉 결합 순/역방향 적응 비트 할당(hybrid forward/backward adaptive bit allocation)

그 밖에 AC-3에서는 커플링(Coupling) 기법을 사용하는데, 이 방법은 신호의 포락선(envelope)을 통해 신호의 지향성을 판단하게 되는데, 고주파 신호의 경우 인접한 두개의 신호 성분들에 대한 지향성은 사람의 청각 시스템에서는 민감하게 인지 하지 못한다. 이러한 청각적 특성을 이용하여 고주파 성분들을 하나의 커플링 채널로 감소시킴으로써 매우 낮은 비트 율을 제공한다 [14][15][16][17].

Ⅵ. MPEG-4

MPEG-4 표준안(ISO/IEC 14496)은 인터넷 동영상, 무선 동영상, 양방향 홈쇼핑, 가상 현실 게임(또는 시뮬레이션, 학습)등의 멀티미디어 정보를 처리하기 위하여 매우 낮은 비트율의 비디오 · 오디오 복호화 신호를 얻을 수 있도록 제정되었다. 자연영상과 자연오디오 신호와 더불어 컴퓨터 그래픽스 데이터와 합성 음성 및 합성 음향 등을 처리하여 이들의 합성 처리를 도모함으로써 멀티미디어 정보를 객체별로 독립적이며 유연성 있게 처리할 수 있다.

〈그림 15〉와 〈그림 21〉과 같이 객체 단위의 정보들을 계층적으로 구성(system)하여 기존에 사용되던 많은 전송 방식(DMIF; Delivery Multimedia Interface Framework)을 사용할 수 있도록 제정되었다.

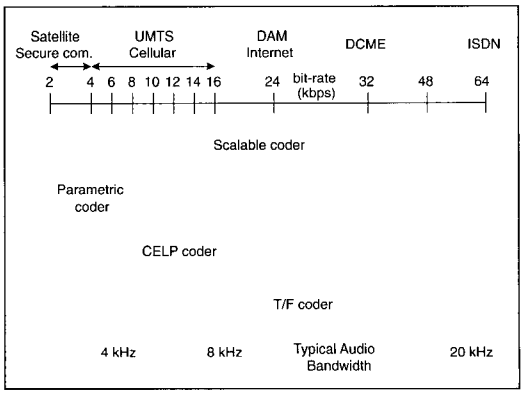


〈그림 21〉 MPEG-4 표준안의 구성[1, p104-105]

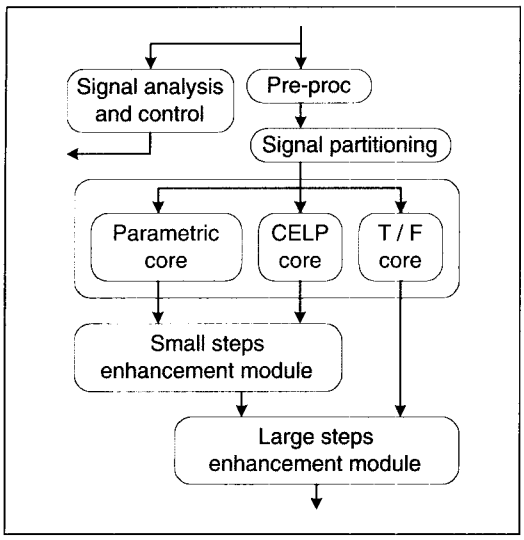
MPEG-4 표준안은 버전 1.과 버전 2.로 구분되는데, 버전 1.은 기본적인 부 · 복호화 방식과 비트열 구조와 관련한 구성과 전송 방식에 대하여 제정되었으며, 버전 2.는 이를 기반으로 확장된 표준안을 제정한다.

1. MPEG-4 오디오

MPEG-4 오디오는 크게 자연 오디오와 합성 오디오로 구분을 할 수 있으며, 각 오디오는 음성을 포함하고 있다. 자연오디오는 고품질의 오디오를 대상으로 하는 T/F(Time/Frequency) 부 · 복호화기, 저 비트율을 제공하는 음성 전용의 CELP (Code Excited Linear Prediction) 부 · 복호화기와 오디오 및 음성 대상인 Parametric 부 · 복호화기로 구성되어 있으며, 〈그림 22〉에 도시한 바와 같이 여러 응용분야에 적합한 2 ~ 64 kbps의 비트율을 지원한다. T/F 부 · 복호화기에서는 T/F 기술들을 적용하여 16 ~ 64 kbps의 비트율에서 최적화된 AAC 부 · 복호화 방식을 확장하여 8 ~ 96 kHz의 샘플링 주파수에서 고품질의 오디오를 제공한다. CELP 부 · 복호화기에서는 8 kHz 또는 16 kHz의 샘플링 주파수에서 각각 4 ~ 14 kbps, 14 ~ 24 kbps의 중간 정도의 저 비트율을 제공한다. Parametric 부 · 복호화기에서 대상에 따라 두가지 방식이 적용되는데, 오디오에 대해서는 HILN(Harmonic Individual Line with Noise) 방식을, 음성에 대해서는 HVXC



〈그림 22〉 MPEG-4 자연 오디오 관련 비트율과 대역폭



〈그림 23〉 MPEG-4 자연 오디오 부호화

(Harmonic Vector Excitation Coding) 방식을 적용한다. 일반적으로 8 kHz 샘플링 주파수에서 1.4~4 kbps의 초저 비트율을 제공할 수 있다. 〈그림 23〉에서와 같이 여러 방식을 계층적으로 결합한 확장형 부 · 복호화기(scalable coder) 또한 지원된다. 합성 오디오부는 합성 음성 합성 인터페이스(TTSI; Text-To Speech Interface)와 구조화 오디오(SA; Structured Audio)로 구성된다.

다양한 응용들에 적용을 위해 MPEG-4 오디오 표준안에서 제공되는 특징적인 기능들은 다음과 같다. 첫 번째로 복호화 중 피치(pitch)와 관계없이 time scale 변환을 제공하는 속도(speed) 변환 기능을 들 수 있다. 이 기능은 데이터 베이스를 검색하는 등의 작업에 유용한 fast forward 기능, 주어진 video sequence에 대해 audio sequence의 길이를 적응적으로 적용할 수 있는 기능을 제공한다. 두 번째로 부 · 복호화 중 time scale에 관계없이 피치(pitch) 변환을 제공하는 피치(pitch) 변환 기능을 들 수 있으며, 이 기능은 voice alteration이나 Karaoke 형태의 응용제품 등에서 사용될 수 있다. 세 번째로 부호화 비트열의 일부만을 사용해서도 비트열을 복호화할 수 있는 확장성(scalability)을 들 수 있으며, 비트율과 복잡도(complexity)에 대한 확장성(scalability)을 지원하는 특징을 가진다. 전송 및 복호화 등에서 전송망의 환경에 적응적으로 비트율을 변경시킬 수 있는 비트율 확장성(Bitrate scalability)을 제공한다. 비트율 확장성(bitrate scalabitity)의 특별한 경우인 대역폭 확장성(Bandwidth scalability)은 전송이나 복호화 과정 중 주파수 스펙트럼의 일부를 처리하지 않고 제거함으로써 무선 환경 등에서 유용한 유연성을 제공한다. 부 · 복호화기에서 복잡도를 변경시킬 수 있는 복잡도 확장성(complexity)을 지원하며, 특히 복호화기에서는 부호화 시 적용된 복잡도와는 다른 복잡도를 적용할 수 있다. 자연 오디오를 대상으로 하는 부호화를 도시한 〈그림 23〉에서와 같이 입력 신호의 특성에 따라 각 core를 선택하여 처리함과 동시에 특정 core내에서의 확장성(Small steps enhancement module)과 전체에 대한 확장성(Large steps enhancement module)을 적용하여 계층구조의 부 · 복호화(Scalable coder)를 지원함으로써 고정된 전송속도를 보장하지 못하는 경우에 적응적으로 사용할 수 있다. 네 번째로 인터넷, 통신 등을 대상으로 하였기 때문에 오류 내성(Error robustness)을 갖도록 제정되었다.

2. MPEG-4 자연 오디오

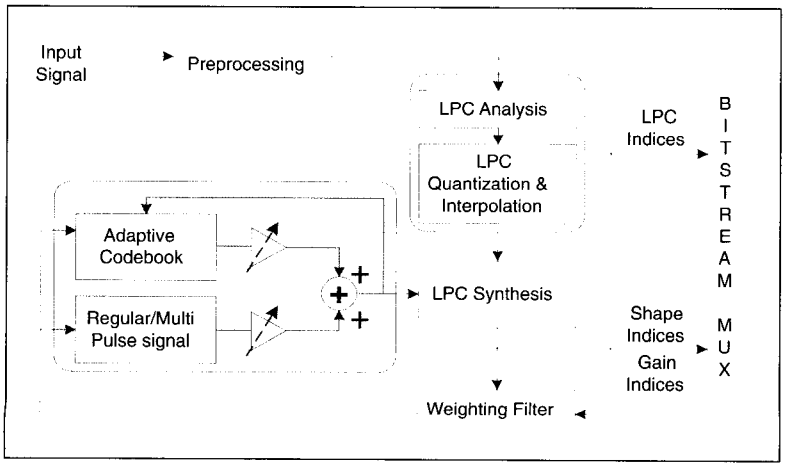
자연적인 오디오신호를 부호화하는 3가지 방식으로 이루어져 있다.

2.1 CELP 부 · 복호화기

MPEG-4 CELP 부 · 복호화기는 음성 전용 부 · 복호화기이며 선형 예측 부호화(LPC; Linear Predictive Coding)를 기반으로 한 압축 방식이다. 두 가지의 샘플링 주파수를 사용하여 고정 및 가변 비트율을 지원하며, 4~24 kbps 범위의 다양한 비트율을 제공한다.

음성은 유성음과 무성음으로 구분할 수 있으며, 특히 유성음은 특정의 주파수를 갖는 주기적인 특성을 가지고 있으며 무성음은 잡음과 유사한 비주기적인 특성을 가진다. 이때 음성의 주기적인 특성을 대표하는 주기를 피치(pitch) 주기, 해당 주파수를 기본 주파수(fundamental frequency)라 한다. 이러한 음성의 특성을 이용하여 유성음을 대상으로 LPC를 적용한 후에 LPC로 표현되는 성분을 원래의 신호에서 제거한 잡음과 유사한 잔차 신호(residual signal)를 대상으로 잡음 생성기를 이용하여 적응적으로 필터링하게 된다. MPEG-4 CELP(Code Excited Linear Prediction Coder) 부 · 복호화에서는 이와 유사한 방법을 사용하며, 부호화기를 〈그림 24〉에 도시한다. CELP 방식은 부호화기 구성에서 사용되는 모듈들을 복호화기에서도 사용한다. 또한 잡음 생성기로써 코드북을 사용하여 여기신호를 생성하게 된다.

입력된 음성 신호를 분석한 선형예측계수(LPC;Linear Prediction Coefficients)와 여기(Excitation) 코드북의 조합으로 생성된 여기신호를 이용하여 원래 신호와의 오차가 최소화되도록 적응시킴으로써 원래의 음성 신호를 선형예측 계수 및 여기 코드북 관련 정보만으로 표현할 수 있다. MPEG-4 CELP에서는 양자화 방법에 따라 스칼라 양자화(SQ : Scalar Quantization)와 벡터 양자화(VQ;Vector Quantization)로 구분되며, 8 kHz(협대역;narrowband) 또는 16 kHz(광대역;wideband)의 샘플링 주파수를 사용한다. 샘플링 주파수에 따라 200~3400 Hz의 대역폭을 갖는 8 kHz에서는 RPE(Regular Pulse Excitation) 코드북을 사용하여 4~14 kbps의 비트율을 지원하며, 50~7000 Hz의 대역폭을 갖는 16 kHz에서는 MPE(Multi Pulse Excitation) 코드북을 사용하여 14~24 kbps의 비트율을 지원한다. 전송되는 LPC의 보간 여부에 따라서 보다 정밀한 비트율 제어(Fine Rate Control)를 할 수 있다. 이러한 특성들의 선택적 적용에 의해서 8개의 모드로 구분된다. MPEG-4 CELP



〈그림 24〉 MPEG-4 CELP 부호화기의 블록도

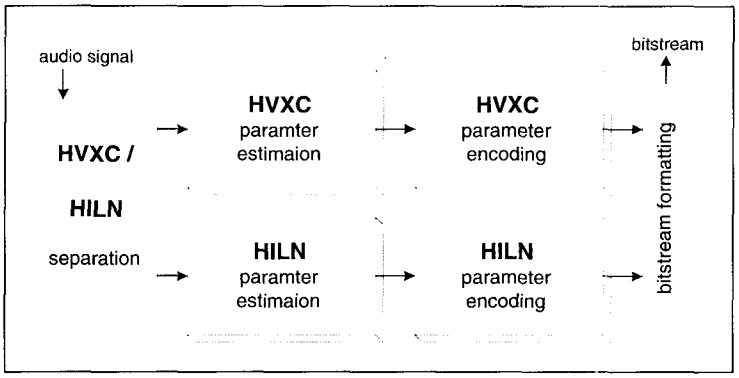
에서는 다양한 전송율을 지원하며, 음성의 자연성과 명료성을 증가시키기 위해 비트율 확장성(Birate Scalability)과 대역 확장성(Bandwidth Scalability)을 지원한다. 비트열의 구조는 기본 계층(Base Layer)과 향상 계층(Enhancement layer)으로 구성되며, 비트율 또는 대역폭 확장성 모드일 경우에 향상 계층이 첨부되게 된다. LPC에 대한 두 가지의 보간 방식과 비트열에 포함된 LPC 차수보다 적은 차수의 적용에 의하여 3개의 다른 복잡도를 지원한다. LPC는 양자화 오차에 민감하여 안정성(stability)의 단점을 가지고 있으나 이를 보완하기 위해서 LPC를 LSF(Line Spectral Frequency)라고도 호칭되는 LSP (Line Spectrum Pair)나 LAR(Log-Area-Ratio)로 변환하여 처리하며, 2가지의 보간 방식은 이들에 대한 보간을 의미한다.

2.2 Parametric 부 · 복호화기

Parametric 부 · 복호화 방식은 MPEG-4 오디오 부 · 복호화 방식, 즉 T/F와 CELP 보다 높은 압축률을 가지고 있다. MPEG-4 Parametric 부 · 복호화 방식은 음성신호를 대상으로 2.0~4.0 kbps의 비트율로 부호화 하는 HVXC(Harmonic Vector eXcitaton Coding)와 음악신호를 대상으로 4.0 ~ 8.0 kbps의 비트율로 부호화 하는 HILN (Harmonic Individual Line plus Noise)을 사용한다.

파라메트릭 부 · 복호화는 〈그림 25〉에서와 같이 부호화 하려는 대상을 부 · 복호화기에 공통적으로 사용되는 스펙트럼 영역에서 모델링한 생성모델을 이용 입력 신호에 근접한 신호를 생성하는데 필요한 파라미터를 추출하고 해당 파라미터만을 전송하여 고 압축율을 제공한다. 생성모델은 스펙트럼 영역을 기반으로 처리되며, HVXC에서는 LPC를 이용한 파라미터와 피치를 이용하여 분리된 유 · 무성음에 대해 각기 벡터 양자화, 벡터 여기 부호화 방식을 사용하고 HILN에서는 3가지 모델(harmonic lines, individual lines, noise) 중 입력 오디오의 선스펙트럼 특성에 적합한 모델에 의한 특징 파라메터들을 추출하고 이를 심리음향 모델을 기반으로 한 양자화 방식을 사용한다. 파라메트릭 부 · 복호화는 HVXC 또는 HILN 각각을 독립적으로 사용하는 모드뿐만 아니라 음성과 오디오가 혼성된 경우에 해당 프레임에 적합한 HVXC 또는 HILN을 번갈아 적용하는 스위칭 모드와 HVXC와 HILN을 같이 사용하는 혼성 모드를 제공한다.

파라메트릭 부 · 복호화기에서는 입력 신호를 모델화 할 때 비교적 간단하고 효과적인 선스펙트럼을 사용한다. 선스펙트럼은 주파수 영역에서 특정



〈그림 25〉 Parametric 부호화기의 일반 블록도

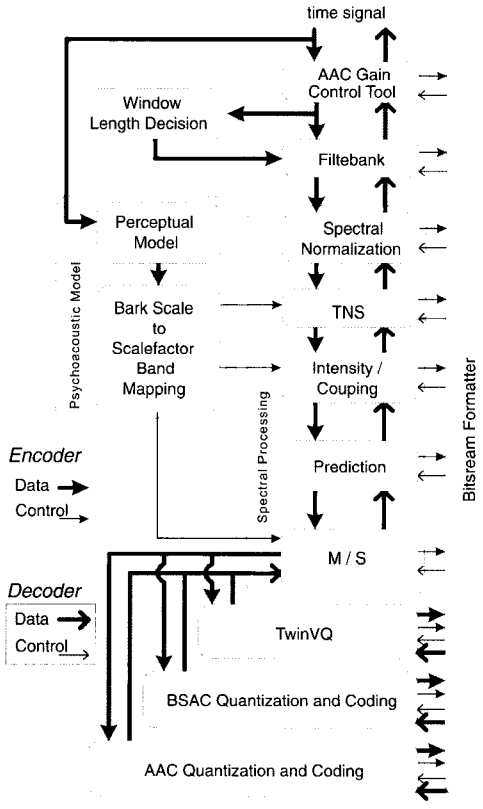
대역만을 정보로 취하는 것으로써 특히, 음성신호 같은 경우는 신호의 형태가 기본 주파수의 정수배 만큼에 해당되는 주파수 값마다 선스펙트럼이 존재하는 특징을 가지고 있어 음성신호를 모델화하기에 적합하다.

2.3 T/F 부 · 복호화기

시간/주파수(Time/Frequency) 부 · 복호화기는 광범위한 샘플링 주파수(8-96kHz)와 비트율(6-64kbps)을 가진 고음질의 오디오 부 · 복호화 방식으로서 세 개의 알고리즘인 AAC(Advanced Audio Coding), BSAC(Bit Sliced Arithmetic Coding), TwinVQ(Transform-domain Weighted Interleaved Vector Quantization)가 하나로 구성되어 있다. AAC는 MPEG-2에서 표준화되었던 알고리즘을 기본으로 하였으며, BSAC는 비트량의 제약에 구애받지 않고 가변적으로 서비스할 수 있는 기술을 말한다. TwinVQ는 입력 신호를 직접 부호화하는 기술과 달리 여러 개의 데이터를 모아 일정한 패턴으로 만든 후, 미리 준비된 표준 패턴과 비교해서 가장 유사한 패턴을 선택하여 부호화하는 방식으로 높은 음질을 유지하고 압축률을 높일 수 있다.

〈그림 26〉은 시간/주파수변환방식의 구조를 나타낸 것이다. 양자화를 제외한 각각의 모듈들은 세 개의 알고리즘에 공통으로 사용된다.

각각의 모듈을 살펴보면, AAC 이득제어도구(Gain Control Tool)는 프레임마다 각기 다른 진폭을 일정하게 한다. 윈도우 길이 결정(Window Length Decision)은 다양한 입력신호에 대해 필터뱅크(Filterbank)가 스펙트럼 성분을 효율적으로 분석할 수 있도록 윈도우의 모양과 길이를 결정한다. 필터뱅크는 시간영역의 신호를 주파수 영역을 신호로 변환하거나 주파수 영역의 신호를 시간 영역의 신호로 변환하여 스펙트럼 신호를 효율적으로 분석하는 역할을 한다. 스펙트럼 정규화(Spectral Normalization)는 스펙트럼 상에서 진폭이 일정하게 되도록 처리하는 것이고, TNS(Temporal Noise Shaping)는 각 윈도우 변환에서 양자화 잡음의 Temporal Shape를 조절하기 위해서 사용되며 필터링 과정과 스펙트럼 데이터 부분에 의해 적용된다. 지각 모델(Perceptual Model)은 마스킹 임계값을 예측하기 위해 사용되며, MPEG-1 오디오 표준안의 계층 III에서 사용된 심리음향 모델 II와 유사하다. 스테레오 처리 방식으로서 M/S는



〈그림 26〉 T/F 부 · 복호화기 블록도

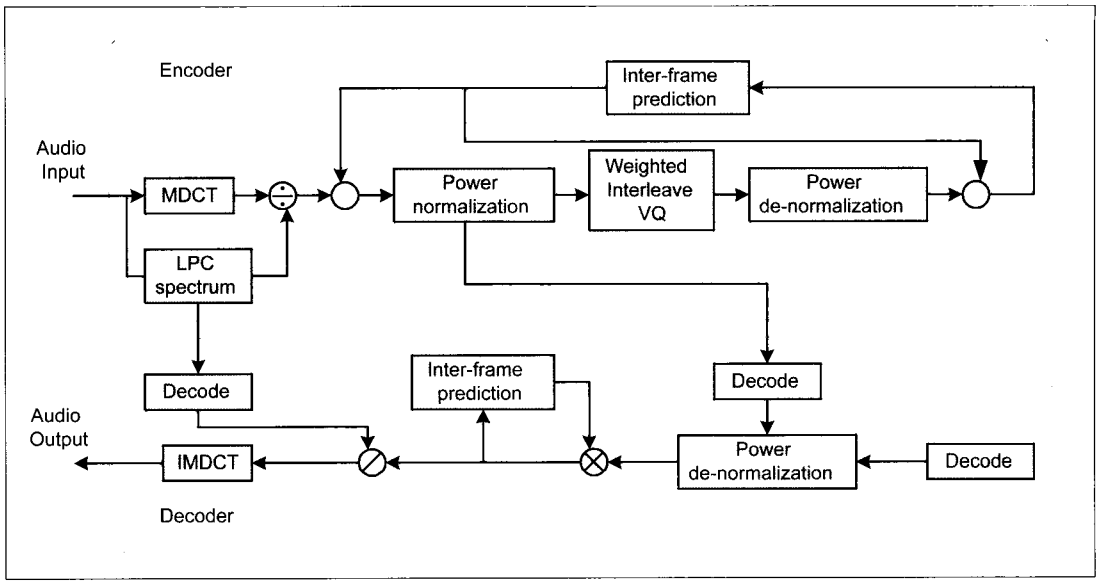
left/right 두 신호의 합과 차를 이용하고 Intensity/Coupling은 스테레오의 고주파 신호를 한 채널로 코딩하여 압축을 수행함으로써 코딩 효율을 개선시킨다. 예측(Prediction)은 프레임과 프레임사이에 신호의 유사성을 이용하는 것으로 신호의 예측신호와 잔차 신호로부터 원래의 신호를 복원해낸다.

상위 기능 블록들을 통과한 신호는 양자화 및 부호화를 수행하게 되며, 이는 AAC, BSAC, TwinVQ의 세 가지 방식이 사용될 수 있다. AAC 양자화 및 부호화(AAC Quantization and Coding)와 BSAC 양자화 및 부호화(BSAC Quantization and Coding)에서는 양자화 레벨이 큰 값에 대해서 적은 비트를 할당하기 위해 비선형 양자화를 수행하고 이 값들을 무손실(lossless) 부호화 한다. 사용되는 무손실 부호화로는 허프만(Huffman) 부호화 또는 BSAC(Quantization and Coding)가 사용된다. BSAC는 최저 주파수 계수들로부터 최대 주파수 계수의 순서에 입각한 4개의 양자화된 스펙트럼 데이터 단위로 bit-sliced 벡터를 구성하여 반복적으로 MSB에서부터 LSB의 순서로 이동하면서 이전 상태를 고려한 두 개의 부벡터로 재구성하여 산술부호화를 적용함으로써 중복성을 제거한다. 마지막으로 TwinVQ는 스펙트럼 신호를 재배열(Interleave)하여 에너지를 균일화한 후 벡터 양자화하는 방법이다.

BSAC에서 양자화된 샘플은 bit-sliced 처리되며, bit-sliced 처리란 샘플단위로의 처리가 아닌 샘플을 표현하는 비트의 위치에서 여러 샘플의 공통 비트 위치의 비트들을 묶어 데이터로 처리함을 의미한다. bit-sliced 데이터의 중복성을 감소시키기 위하여 최저 주파수 계수들에서부터 최고 주파수 계수 순서의 최상위비트(MSB) 데이터를 중복되지 않은 4개(4-tuples)의 데이터 단위로 벡터를 구성한다. 이러한 벡터들은 이전 상태에 따라 두 개의 부 벡터(sub-vector)로 분리되며, bit-sliced 시퀀스의 1 ~ 4차원의 부 벡터들은 산술 부호화(arithmetic coding) 된다. 최저 주파수에서부터 최고 주파수의 모든 MSB 데이터들이 부호화 된 후 하위 비트 위치로 순차적으로 이동하면서 LSB 데이터들을 부호화 할 때까지 반복적으로 적용한다. 이전 상태(previous state)들은 MSB로부터 LSB로의 순서로 부호화 벡터들에 따라 갱신되며, 초기화 상태는 모두 '0' 사용한다. 비트 값이 '0' 변하지 않으며, '0' 아닐 경우에 '1' 변하게 된다. 이 값들은 이전 상태가 0인지 또는 1인지에 기인하여 분리하는, 즉 '0' 이전 상태를 갖는 경우의 부 벡터와 '1' 이전 상태를 갖는 경우의 부 벡터로 분리하게 된다.

2.3.1 TwinVQ 부 · 복호화기

〈그림 27〉에 도시된 바와 같이 TwinVQ (Tranform domain Weighted INterleave Vector Quantization) 부호화기에서 MDCT를 이용하여 시간/주파수 변환되어진 스펙트럼 신호는 LPC 분석에 의해 예측된 스펙트럼 포락에 의해서 평탄화 된다. 평탄화된 스펙트럼 신호는 남아 있는 스펙트럼 피크를 평탄화하기 위하여 Inter-frame 예측(prediction)을 통과한다. 통과된 신호는 각 서브벡터의 에너지를 균일화하는 목적에서 재배열(Interleave)된 서브벡터로 분할된 후 벡터 양자화하게 된다. 복호화기에서는 비트열에서 양자화된 오디오 스펙트럼으로부터 코딩된 정보를 분석한다. 그리고, 양자화된 값과 기타 복호화하는데 필요한 부가 정보들을 찾아 양자화된 스펙트럼



〈그림 27〉 TwinVQ 오디오 부 · 복호화기의 흐름도(5)

을 재구성하고, Inter-frame 예측(prediction)과 LPC 스펙트럼(spectrum)을 통과한 후 IMDCT를 거쳐 시간 영역의 신호로 변환된다.

3. 합성(Synthetic) 오디오

MPEG-4 Audio의 마지막 분야로써 합성 오디오(Synthetic Audio)가 있다. 이 부분은 MPEG-4의 T/F, CELP, Parametric과는 달리 대상 오디오신호를 압축(compression)하는 것이 아니라 합성(synthesize)한다는 점에서 차이를 보이고 있다. 합성 오디오는 크게 2가지로 구분할 수 있으며, 이들은 입력된 문자열을 음성신호로 변화시키는 TTS(Text to Speech)와 Music language인 SAOL(Structured Audio Orchestra Language)과 SASL(Structured Audio Score Language)을 사용하여 음악을 합성하는 구조화 오디오(Structured Audio)이다.

3.1 TTS(Text To Speech)

TTS란 입력 문자를 해석하여 미리 구현된 음성 DB(Data Base)에서 해당 음원을 검색하여 문장을 만들어 출력시키는 시스템을 말한다. 음성DB 및 음성의 합성방법에 따라 TTS를 구현하는 방법은 여러 가지로 적용될 수 있기 때문에 MPEG-4 합성 오디오(Synthetic Audio) 에서는 다음과 같은 인터페이스(interface)만을 제정하고 있다.

① 음성을 합성하는데 있어서 필요한 음운학적(prosodic)인 파라미터
② 합성음성과 얼굴 에니메이션(FA; Facial Animation)간의 동기 문제
③ 영화 사운드 편집시 등장 인물간의 Lip synch(영상과 사운드와의 동기화) 문제
④ 합성 음성의 Random Access 문제

3.2 구조화(Structured) 오디오 부 · 복호화기

구조화 오디오(Structured Audio) 부 · 복호화기는 음악 신호를 아주 현저히 낮은 비트 율로 음악을

합성해 내는 시스템이다. 이 시스템에는 음악을 만들어내는 두 가지 언어(Language)가 있는데 해당 악기의 물리적 특성을 모델링 하여 음원을 만들어 내는 언어인 SAOL(Structured Audio Orchestra Language)과 SAOL(Structured Audio Score Language)에서 만든 악기의 음원 정보를 시간 영역에서 적절히 배열하는 SASL이 그것이다. 그러므로 MPEG-4 구조화 오디오에서는 두 언어로 작성된 음악 정보를 코딩하여 저장하거나 전송하는 걸로 압축을 실현하고 있다. 사용되는 악기의 수(SAOL에관한 정보)나 음악의 길이 및 복잡도(SASL에 관한 정보)에 따라 다르지만 약 0.01 ~ 10 kbps 정도로 압축을 실현하고 있다. 두 가지 언어 이외에 SAOL로 음원을 만드는 대신에 표현하기 힘든 음원은 자연 현상으로부터 녹음하여 wavetable에 저장하여 사용하거나 SASL대신 기존의 MIDI 인터페이스를 그대로 이용할 수도 있다.

하지만 매우 높은 압축률을 제공하는 반면 구조화 오디오는 모든 음악을 표현하기 위해서 현존하는 모든 악기를 SAOL로 모델링 해야하는 부담과 음악 합성시 다른 부 · 복호화기보다 훨씬 높은 성능의 하드웨어 시스템을 요구하므로 실현되어 상용화되기까지는 긴 시간이 필요할 것이다[18][19].

Ⅶ. 결론

본 논고에서는 디지털 오디오 신호를 처리, 전송, 저장하는데 사용되는 알고리즘들에 대하여 살펴보았다. 현재 활발히 사용되고 있는 mp3 파일 포맷을 규정하고있는 MPEG-1 Layer-3 방식과 앞으로 많이 사용될 것으로 기대되고 있는 AAC방식에 대하여 자세히 설명하였다. 또 HDTV의 오디오 방식으로 이미 확정되어진 MPEG-2와 Dolby AC-3 방식에 대해서도 설명하였다. 또 고압축률을 가지고 있는 MPEG-4 방식에 대하여 간단히 기술하였다. 각 방식들은 독특한 구조를 가지고 있으며, 상황에 따라 사용할 수 있도록 복잡한 기술들은 복합적으로 적용되었으며 많은 활용분야를 가지고 있다.

새로운 기술이 제안됨에 따라 차세대 압축기법에 대한 연구도 많이 진행되고 있다. 또 특허 때문에 이 기법들을 사용하기 어려워하는 사용자를 위하여 공개된 기술을 연구하는 그룹도 많아지고 있다. 현재 HDTV 규격과 DAB 규격등에 이미 규격화된 기술들이 사용되고 있으며, 새로이 개발되는 기술들은 점차적으로 활발히 쓰일 것으로 예상된다.