

Dolby AC-3: 고품질 오디오 부호화의 현재와 미래

윤 대 희, 이 근 섭
연세대학교 전자공학과

1. 서 론

오늘날 CD (Compact Disk)의 대중화로 많은 사람들이 디지털 오디오를 생활 속에서 친숙하게 접하게 되었다. LP로 대표되던 아날로그 오디오가 디지털 오디오로 점차 바뀌어가고 있는 것이다. 이와 같은 오디오 매체의 변화는 아날로그 오디오가 따라올 수 없는 디지털 오디오만의 장점으로 가능하게 되었다.

디지털 오디오의 장점은 아날로그에 비해 음질이 훨씬 깨끗하고, 음질의 변화가 거의 없다는 것이다. 따라서 이제는 원음을 손실 없이 듣고 다른 곳으로 전송할 수 있는 환경이 갖추어진 것이다. 또한 디지털 오디오는 영화나 컴퓨터와 같은 다른 매체와의 호환성이 뛰어나므로 계속해서 그 응용범위가 확장되고 있다.

그러나 디지털 오디오는 이와 같은 장점에도 불구하고 고음질의 오디오 신호를 얻는데 발생하는 비용이 크다는 단점이 있다. 아날로그 오디오 신호를 샘플링해서 얻어지는 고음질의 디지털 오디오 신호는 데이터량이 아주 많아서 막대한 용량의 저장매체를 요구하며, 전송시 매우 넓은 주파수 대역을 차지하게 된다. 따라서 오디오 신호를 음질의 손실 없이 효과적으로 압축하는 방법이 필요하게 되었다.

디지털 오디오 신호를 효과적으로 압축하는 방식은 80년대에 접어들면서부터 미국의 AT&T, Dolby Lab, 유럽의 IRT, Philips, CCETT와 일본의 Sony 등의 여러 연구소에서 개발되어 왔다. 이러한 오디오 압축기술 중에서 현재 세계적으로 널리 사용되고 있는 것은 유럽의 MUSICAM 방식에 기반을 둔 MPEG과 미국의 Dolby Lab에서 제안한 AC-3 알고리즘이 있다. 현재 MPEG은 국제 표준안으로 채택되었고, AC-3는 미국 내 표준안으로 받아들여져 있다.

이 글에서는 위에서 언급된 여러 표준들 중에서 Dolby사의 AC-3 알고리즘에 대해 자세히 살펴보고, 현재의 국제 표준인 MPEG과 비교해 볼 것이다. 또한 현재까지 상용화된 AC-3 시스템에 대해서도 알아본다.

2. Dolby AC-3의 발전 과정[17]

1) AC-1

Dolby AC-1은 1985년 호주에서 DBS (Direct Broadcast Satellite) 시스템에 처음 사용되었다. AC-1 디코더의 비용이 매우 낮았기 때문에, 이후의 다른 여러 DBS 서비스, 위성통신망, 그리고 디지털 케이블 라디오 시스템 등에 사용되었다.

AC-1은 ADM (Adaptive Delta Modulation)을 바탕으로 만들어졌다. 따라서 AC-1 비트열은 원신호 값의 정보 대신 샘플간의 차이 값에 대한 정보를 가진다. 여기에 Dolby 사의 Noise Reduction 기술을 더하여 기본적인 ADM이 갖는 성능을 향상시켰다. AC-1이 갖는 비트율은 채널 당 220~325 kbps 사이이다.

2) AC-2

Dolby AC-2 부호화 방법의 개발은 1980년대 중반에 시작되었고, 1989년 이후 상용화되었다. AC-2는 지점대지점(Point-to-Point)을 포함하는 응용분야에 사용하기 쉽도록 인코더와 디코더의 복잡도가 낮도록 비슷하게 설계되었고, 20kHz 대역폭의 오디오를 제공하기 위한 비트율은 채널 당 128~192 kbps사이이다.

AC-2는 TDAC (Time Domain Aliasing Cancellation)를 사용하여 MDCT/MDST (Modified Discrete Cosine Transform / Modified Discrete Sine Transform)가 적용된다. 또한 Evenly-Stacked TDAC 변환의 시변 특성을 사용하여 신호 특성에 따라 시간과 주파수 분해능 사이에 동적 Trade-off를 허용하는 특징을 갖는다.

AC-2는 신호의 스펙트럼 포락선을 추정하고 주파수에 대한 함수와 지정된 비트율 사이를 고려하여 워드 길이를 결정한다. 유용한 데이터 중 17~20%만 신호의 특성으로 채택하고, 사람의 청각 특성을 고려한 마스킹 효과를 이용하여 비트를 할당하는 알고리즘을 이용한 다.

3) AC-3

Dolby AC-3는 극장용 디지털 사운드 시스템을 위해 개발되었고, 1991년에 처음으로 도입되었다. 그리고 1993년 ATSC (United States Advanced Television System Committee) 에 의해 복미 HDTV 표준을 위한 오디오 부호화 시스템으로 채택되었다. AC-3는 5개의 오디오 채널과 1개의 저주파 효과 채널로 51 채널의 구성을 가지며, 32 ~640 kbps의 비트율을 갖는다. AC-2와 비교했을 때, AC-3가 약간 높은 복잡도로 더 낮은 비트율에서 좀더 우수한 음질을 들려준다.

AC-3는 Oddly-Stacked TDAC 변환과 시간과 주파수 사이의 가변 분해능이 적용된다. AC-3의 비트 할당 처리는 AC-2와는 달리, 전체 청각모델의 비트 할당 단계를 한 후 부호화 또는 복호화 처리 과정에서 간략한 파라메트릭 비트 할당 단계를 수행하는 2단계의 처리를 한다. 그리고 AC-3에 도입된 채널 커플링은 오디오의 고주파 부분에 대해서 변환 계수를 조합하여 각 채널간의 데이터를 공유하는 식으로 압축률을 높인다.

3. AC-3의 기본 기술 설명

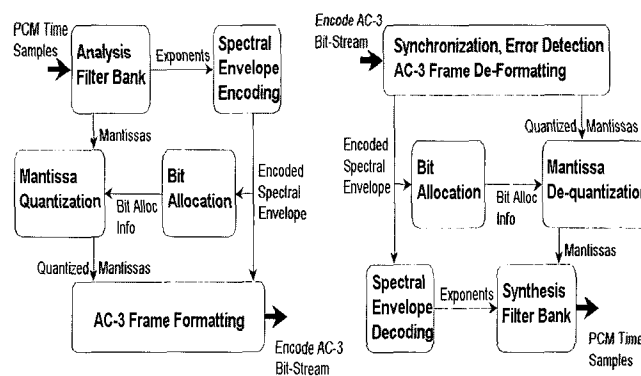
AC-3 오디오 부호화 알고리즘은 복미 지역의 HDTV 오디오 압축 기술의 표준안으로 돌비 연구소가 제안한 방식이다. SMPTE 권고안인 51 채널을 수용하고, 32~640 kbit/s의 비트율을 갖는다. 압축비를 높이기 위해서 채널간 혹은 채널 내의 마스크 특성을 이용하며, 고주파 대역의 채널 커플링을 이용하여 보다 낮은 비트율을 제공한다. 다음 그림은 Dolby AC-3 Encoder와 Decoder의 System 블록도이다.

기본적인 부호화 방법은 시간축 에일리어징 제거 방법에 기초한 MDCT 변환 부호화 방식을 사용하였다. 변환 부호화 방식의 큰 단점으로 프레임 내에서 크게 변화하는 신호를 변환하였다가 다시 역변환할 때 양자화 오차에 의해 발생하는 프리에코(pre-echo)를 들 수 있다. AC-3에서는 이를 해결하기 위해 윈도우 크기를 신호 특성에 따라 변화시키는 방법을 사용하고 있다. 즉, 안정 구간의 신호일 경우에는 윈도우 크기를 512 샘플로 하고, 천이 구간과 같이 불안정 구간일 경우에는 윈도우 크기를 256 샘플로 작게 하여 변환 부호화 시 발생하는 프리에코를 줄이는 방법이다.

입력 신호는 윈도우를 씌운 후 MDCT (Modified Discrete Cosine Transform) 를 이용하여 주파수 영역으로 변환한다. MDCT는 시간 영역 에일리어징을 제거하는 특성을 가질 뿐만 아니라 임계 샘플링을 보장하고 계산량을 크게 줄일 수 있는 고속 알고리즘을 적용할 수 있다[3]. 또한 주파수 선택성이 매우 향상된 특성을 갖는다.

MDCT 계수들은 지수부와 가수부로 나뉘어져 부호화되는데, 이는 간단한 비트 연산을 통하여 추출하므로 실제 구현 시 계산량이 많은 로그 계산을 효과적으로 줄일 수 있는 장점이 있다. 또한 지수값을 사용하여 거칠게나마 신호의 스펙트럼을 구할 수 있기 때문에 심리 음향 모델을 구현하기 위하여 따로 FFT와 같은 계산을 필요로 하지 않아 복잡도를 크게 줄일 수 있다. 부호화 시 지수부는 고정 비트로 할당하여 전송하되, 전송 시 비트의 낭비를 막기 위해서 몇 개의 지수를 묶어 부호화하는 grouping 방법을 사용한다.

가수값을 양자화하기 위한 비트 할당은 지수값을 사용하여 구성된 간략화된 스펙트럼에 대해 심리 음향



AC-3 Encoder AC-3 Decoder

그림 1. Dolby AC-3 Block Diagram

모델을 적용하여 결정한다. 이러한 방법을 사용하면, 복호화 단계에서도 전송 받은 지수값들만으로 다시 비트 할당 정보를 만들어 낼 수 있기 때문에, 비트 스트림에 비트 할당에 대한 정보를 수록하지 않아도 되는 장점이 있다.

다채널 부호화 시에는 커플링(coupling)이라는, 채널 간의 데이터 공유를 통하여 압축률을 높인다. 인간의 청각 특성상 음상 정위(localization)는 저주파 영역에서는 신호의 위상 성분, 고주파 영역에서는 크기 성분에 의존하므로, 다채널일 경우 어떤 주파수 이상의 고주파 대역에서는 채널간의 스펙트럼 정보를 공유하게 하고, 각 채널간의 크기 정보만을 전송하여도 충분한 임장감을 얻을 수 있게 된다.

1) 전처리 과정

전처리 과정은 변환 부호화 방식의 단점인 Pre-echo를 없애기 위한 것으로 먼저 부호화기에서 입력 PCM 파형을 하이패스 필터에 통과시켜 DC성분을 모두 없앤 후 천이구간을 검출한다. 이때 각 블록별로 천이가 검출되면 512샘플의 긴 블록을 256샘플의 짧은 블록으로 전환시킨다. 그리고 나서 주파수 선택성을 높이고 블록 경계 현상에 의해 나타나는 왜곡을 줄이기 위해서 적절한 윈도우 함수를 곱한다. 복호화기에서는 이에 대한 정보를 전송 받아서 블록 길이를 다시 스위칭하여 각 블록의 길이에 맞게 역변환하게 된다.

2) TDAC (Time Domain Aliasing Cancellation) 변환

일반적으로 신호 분석 시 중첩, 가산을 행하는 시스템에서는 인접한 시간축 윈도우간의 중첩에 의해서 Aliasing이 일어나게 되는데, TDAC 방법을 사용하면 분석단에서 발생한 Aliasing을 효과적으로 제거함으로써 신호를 완전히 복원해 낼 수 있다. 다음 그림 2에 TDAC Transform을 이용한 변환/역변환 과정이 잘 나타나 있다.

분석단은 입력 신호에 윈도우를 씌워 중첩하는 과정과, 그 신호를 주파수 영역으로 변환시키는 과정으로 구성된다. 이 때 변환은 주파수축에서 의미가 있는 종류이어야 하는데, 예를 들면 DFT (Discrete Fourier Transform), DCT (Discrete Cosine Transform), DST (Discrete Sine Transform) 등이 있다. 합성 시에는 역방향 변환을 행하고, 합성 윈도우와 곱한 후 미리 합성되어있는 이전 구간 신호와 중첩함으로써 원신호를 복원한다. 원신호 $x(n)$ 이 복원될 수 있는 조건은 분석 및 합성 윈도우가 시간 영역에서 중첩되고 가산된 결과가 평탄해야 한다는 것이다.

MDCT 방식을 사용했을 경우 가장 큰 장점은 윈도우

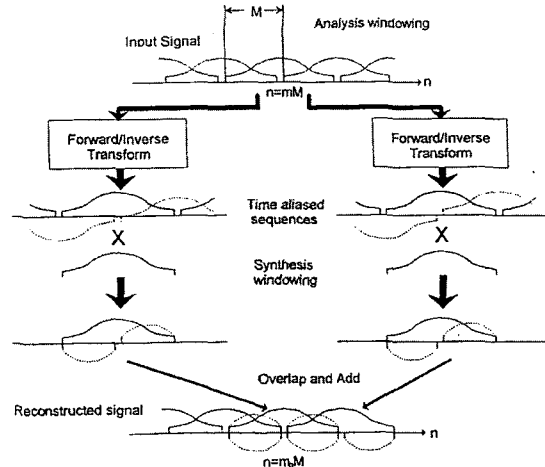


그림 2. MDCT forward/inverse transform

우의 50% 중첩(overlap)을 허용하면서도 추가되는 비트 요구량이 없어진다는 것이다. 즉, 임계 표본화를 보장하는 것으로, 프레임 크기가 N 인 윈도우를 50% 중첩하여 변환하여도 $N/2$ 개 샘플로부터 완전 복원된 신호를 얻을 수 있다는 것이다. 이외에도 다음과 같은 장점이 있다[5].

- 주파수 선택성이 높다
- 블록간의 경계에서 일어나는 왜곡 현상을 줄일 수 있다
- FFT와 같은 빠른 계산 방법이 존재한다

3) 지수 부호화

부호화 시 심리 음향 모델을 적용하기 위해서는 입력 신호의 스펙트럼을 구하여야 하는데, 이는 FFT를 통한 주파수 영역 분석을 행하는 것이 가장 일반적이지만 AC-3에서는 MDCT 변환 계수로부터 스펙트럼 정보를 얻어낸다. 변환 계수는 지수부(exponent)와 가수부(mantissa)로 나뉘어져 부호화되며, 지수부는 주로 음압의 크기, 즉 dynamic range를 나타내는 파라미터이고, 가수부는 양자화 잡음에 영향을 미치는 파라미터이다. 심리 음향 모델에서 사용하는 스펙트럼 정보는 바로 이러한 변환 계수의 지수로부터 얻어진다.

AC-3에 사용된 TDAC 필터 뱅크의 주파수 해상도는 93.75 Hz이며, 인접 대역에 대하여 12 dB의 감쇄율을 갖는다. 지수값의 변화는 ± 2 를 초과하지 않으므로, 인접 샘플간의 차를 구하여 부호화하면 -2, -1, 0, +1, +2를 나타내는 3 bit로써 지수값을 나타낼 수 있게 되고, 3개의 계수를 묶어서 grouping하게 되면 3 샘플 당 7 bit로써 부호화할 수 있게 된다. DC 값을 나타내는 처음

샘플만 5 bit를 사용하여 절대값으로 부호화하고, 나머지 샘플들은 이와 같은 방법으로 차를 구하여 부호화하면 지수값 하나를 부호화하는데 소요되는 비트수는 2.33 bit가 된다. 이렇게 부호화하는 방법을 D15 부호화라 한다.

그러나 모든 블록에 대하여 2.33 bit를 사용한다는 것은 비효율적이다. 이렇게 정밀한 해상도를 가지게 되는 경우는 대체로 안정된 구간의 음일 경우가 많고, 일반적인 경우 한번 전송되면 이후 5개 오디오 블록(32ms)은 이 값을 계속하여 사용한다. 이때의 정보량은 0.39 bits/exponent가 된다.

비안정 구간 신호의 스펙트럼은 넓은 대역에 대해 분포하는 특징이 있고, 주파수 영역 해상도보다는 시간 영역의 해상도가 중요하게 된다. 이 경우 사용하는 것이 D25 또는 D45 부호화 방법인데, D25의 경우는 D15 방법의 절반 즉, 지수값을 하나 걸러 하나씩만 전송함으로써 비트율을 낮춘다. D25 방법은 안정된 블록이 2-3개정도 나타난 후 다시 급격하게 변화하는 블록이 2-3개정도 이어져 나타날 경우에 사용한다. D45 방법은 4개 걸러 하나씩 지수를 전송하는 방법으로, 이때의 정보량은 D15에 비하여 1/4인 0.58 bit/exponent가 되며, 천이 구간의 짧은 블록을 부호화하는데 사용된다.

AC-3에서는 이와 같이 스펙트럼의 형태에 따라 적응적인 방법들을 사용하여 효과적인 부호화를 수행한다. 어떤 방법을 사용할 것인가는 부호화기에서 결정하며, 이 정보는 지수 부호화 전략(exponent coding strategy)이라 불리는 2 bit field의 정보(D15, D25, D45, REUSE)로서 복호화기에게 전달된다. 일반적으로 D15 방법인 경우는 프레임의 첫 블록에만 사용되며, 이후 5개의 블록은 그것을 재사용(REUSE)한다.

4) BIT 할당 ALGORITHM

① 심리 음향 모델

변환 계수를 부호화할 때는 비트율을 일정하게 유지하기 위해 분석 블록 단위로 비트를 할당하게 된다. 비트 할당 전략은 최소의 비트수로 최대의 신호 대 잡음비를 얻는 것을 목표로 하며, 전송을 대 왜곡 이론(rate-distortion theory)[4]에 근거하여 신호 스펙트럼의 에너지에 따라 비트를 할당한다. 그러나 사람의 청각 특성에 따라 지각적으로 느끼는 음질은 수치적인 SNR과는 다른 양상을 보이므로 이 특성을 이용하면 수치적으로 필요한 비트수보다 더 적은 비트를 사용하여도 주관적인 음질을 보장할 수 있게 된다. 이렇게 사람의 청각 특성을 모델링한 것을 심리 음향 모델이라 하며 이를 통하여 부호화 효율을 높일 수 있다[7,8].

사람의 귀는 임계 대역이라 불리는, 몇 개의 대역 별

로 음을 분석하고 인식한다. 이때 한 임계 대역의 에너지가 주위 대역의 에너지보다 훨씬 크게 되면, 주위 대역의 작은 음은 들리지 않게 된다. 이러한 마스킹 현상을 이용하여 임계 대역 내에서 마스크 대 잡음비가 0 dB 이상이 되도록(잡음이 마스킹 곡선의 아래쪽에 들어 가도록) 비트를 할당하여 부호화하면, 지각적으로 원음과 동일한 재생음을 얻을 수 있게 된다. 일반적으로 샘플 당 16 비트의 데이터를 1.5 - 3 비트로 부호화하여도 지각적으로 차이가 없는 복원음을 얻을 수 있다.

② 혼성 비트 할당

일반적으로 사용되어지는 비트 할당 방법은 크게 두 가지 부류로 나눌 수 있는데 하나는 전방 적응 할당(forward adaptive bit allocation) 방법이고, 다른 하나는 후방 적응 할당(backward adaptive bit allocation) 방법이다[9]. AC-3에서는 이 두 방법을 결합한 하이브리드 방법을 사용하는데 부호화기와 복호화기는 동일한 형태의 비트 할당 계산 모듈을 가지며, 비교적 간단한 심리 음향 모델을 사용하여 높은 정밀도를 가지도록 구성된다. 이 모듈에 대한 입력은 신호의 스펙트럼 포락선(spectrum envelope)이며, 이는 비트열 정보에 포함되어 복호화기에 전달된다.

이 모듈에서 전방 적응 할당 방법의 특성을 갖는 점은 심리 음향 모델과 차등 비트 할당(delta bit allocation)이다. 비트 할당 계산 모듈은 심리 음향 모델에 근거하여 최적의 비트 할당을 얻어내며, 개선의 여지가 있는 부분에 대해서는 그에 대한 부가 정보를 복호화기로 전송하여 준다. 즉, 부호화기에서 볼 때 약간의 파라미터를 조정함으로써 더 좋은 비트 할당이 이루어질 수 있다면, 복호화기에 그러한 정보를 전달하여 보다 향상된 재생음을 얻을 수 있도록 한다. 이러한 측면은 부호화기의 개발에 개선의 여지를 남겨둔다고도 할 수 있다. 지금 사용되어지는 심리 음향 모델보다 더 정확한 모델을 사용하더라도, 변경된 부분에 대한 정보를 전송함으로써 복호화기 시스템을 변경하지 않고도 호환성을 유지할 수 있다는 것이다.

③ AC-3 비트 할당 알고리즘

AC-3에서는 변환 계수의 각 지수와 가수에 할당되는 비트수는 미리 정해진 값을 사용하지 않고 신호의 스펙트럼에 따라 최적의 비트수를 계산해낸다. 여기서 최종 마스킹 레벨을 구하기 위하여 심리 음향 모델이 적용되며 이에 필요한 파라미터들은 신호에 따라 부호화기에서 결정한다. 각 가수에 대한 비트 할당은 최종 마스킹 레벨과 입력 신호의 전력 밀도 스펙트럼(PSD: Power Spectrum Density)과의 비교를 통해 테이블 룩업(look-up)으로 결정된다. 계산량을 줄이고 로그 함수적

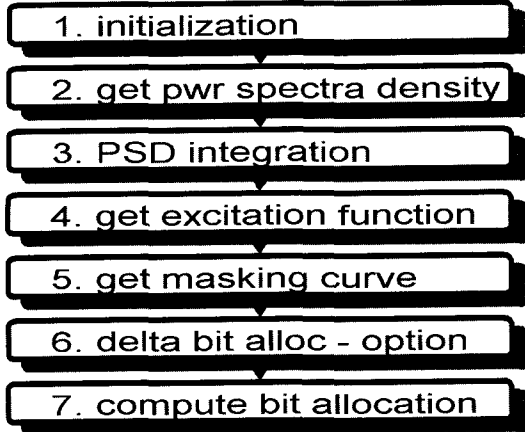


그림 3. Bit Allocation 과정

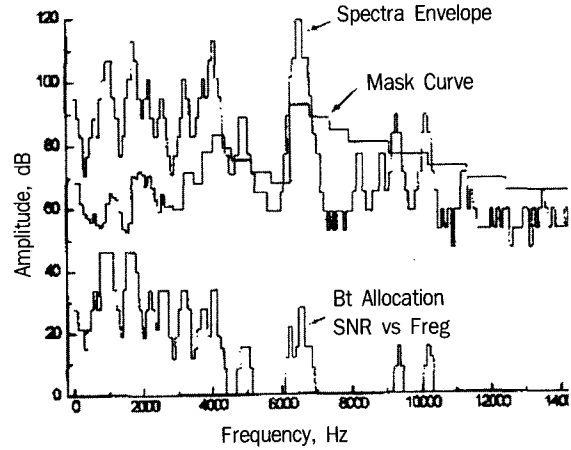


그림 4. Bit Allocation의 예

인 청각특성을 반영하기 위해 원래의 PSD를 작은 밴드들로 묶어서 처리한다.

PCM 샘플의 동적 영역이 96 dB이므로 샘플당 최대 16 bit까지 할당될 수 있다. 비트 할당 결과 사용할 수 있는 비트가 남으면 SNR Offset을 증가시키고, 모자라면 SNR Offset을 줄이는 방법으로 사용 가능한 비트를 모두 할당한다. 그러나 비트를 모두 다 사용하여도 비트가 모자라게 되는 경우에는 커플링(coupling)이라 불리는 블록간 정보 공유를 통하여 비트율을 낮춘다.

5) 가수 부호화

앞서 얻은 비트 할당 정보를 통해 가수부를 양자화하게 되는데, 1 - 5 bit까지 할당된 경우에는 대칭형 양자화(symmetric type quantization)를 사용하고, 6 - 15 bit의 경우에는 비대칭형 양자화(asymmetric type quantization)를 사용한다. 비대칭형 양자화는 실수값을 갖는 가수값을 할당된 비트만큼을 사용하여 양자화하는 방법이다. 부동 소수점 형식으로 표기된 실수값은 할당된 비트수 만큼을 추출하여 전송한다. 대칭형 양자화는 할당된 비트에 따라 정의된 표를 사용하여 양자화하는 방법이다.

지수부 부호화에서와 마찬가지로, 가수부를 부호화할 때에도 몇 개의 샘플들을 묶어서 grouping함으로써 비트를 줄이는 방법을 사용한다. 즉, 3, 5, 11 레벨을 갖는 부호들은 2개 또는 3개씩 묶이어 각각 5비트, 7비트, 11비트 코드워드로 부호화된다. 이때 grouped된 가수값은 그 묶음 내에서 맨 처음 위치에 놓인다.

복호화기에서는 복호한 지수를 입력으로 다시 부호화기에서와 동일한 bit 할당 알고리즘을 통과시켜서 가수에 할당된 비트수를 알아내고 이를 기준으로 대칭형

양자화, 혹은 비대칭형 양자화에 따라 가수를 복원하며 grouping도 풀어낸다.

6) CHANNEL간 부호화

① CHANNEL COUPLING

돌비 알고리즘에서 사용하는 방법은 채널간의 상관관계를 이용한 채널 커플링 방법으로서, 이 방법 역시 심리 음향 모델을 이론적인 배경으로 삼는다. 인간이 소리의 발생 위치를 추정하는데 있어서 저주파 대역에서는 시간 지연에 따른 위상차 정보를 이용하지만 약 2kHz 이상되는 고주파대역에 대해서는 신호의 크기 정보만을 얻어내어 이용한다. 이것은 인간의 귀가 주파수 상으로 매우 가깝게 위치해 있는 두 개의 고주파 신호의 방향성을 구별하지 못한다는 점을 이용한 것이다. 따라서 저주파 대역 신호는 기본 부호화 방법을 사용하여 전송하고, 특정 주파수 이상의 신호에 대해서는 두개 이상 채널에 대한 정보를 공유하여 전송하는 방법을 사용하면 주관적인 음질을 손상시키지 않으면서 압축률을 높일 수 있는 것이다.

커플링이 시작되는 주파수를 커플링 주파수라 하는데 이보다 높은 주파수에서부터 채널들은 커플링 채널로 묶이게 된다. 부호화기는 채널 커플링된 각 채널들의 변환 계수들의 평균값과 함께, 원래 신호의 에너지와 커플링된 신호의 에너지의 비에 대한 정보(이를 coupling coordinate라 한다)를 전송한다. 복호화기는 각 채널과 밴드에 대하여 변환 계수에 coupling coordinate를 곱함으로써 커플링된 채널로부터 원래의 채널들을 복원해낸다.

② REMATRIXING

이 방법은 2/0 모드일 때만 사용되는 방법으로서 원래의 신호(L, R)과, 다음과 같이 정의된 그들의 합/차 신호(L', R')에 대한 에너지를 조사하여 그 중 큰 에너지를 갖는 쌍을 부호화하는 방법이다.

$$L' = 0.5*(L + R)$$

$$R' = 0.5*(L - R)$$

모노 신호와 같이 상관 관계가 높은 신호일 경우 L' 신호는 원래의 L 및 R 신호와 동일한 형태를 갖게 되고, R' 신호는 이론적으로 0이 되므로, R'에는 적은 비트를 그리고 L' 신호에 대해서는 보다 많은 비트를 할당하여 줄 수 있어 좋은 음질을 보장할 수 있게 된다.

이 방법이 중요한 의미를 갖는 것은 기존 돌비 서라운드 복호화기와의 호환성을 유지할 수 있기 때문이다. 현재 보급된 돌비 프로 로직 복호화기에서는 in-phase 성분은 센터 채널로 보내고, out-of-phase 성분은 서라운드 채널로 보내는 방법을 사용한다. 이 때 rematrixing 방법이 사용되지 않을 경우에 서라운드 채널에는 양자화 잡음이 신호성분보다 커져서 신호를 가리게 된다. 그러나 rematrixing이 사용된 경우에는 처음부터 합과 차가 각기 양자화되기 때문에 차에 해당하는 서라운드 채널의 경우에도 신호성분에 비해 양자화 잡음이 작아져서 신호를 가리는 일은 발생하지 않게 된다.

7) DOWNMIXING

일반적으로 수신자 측에서 5.1 전 채널을 재생할 수 있는 복호화기를 보유하고 있지 못한 경우가 많을 것이며, 이 경우 신호의 왜곡을 줄이기 위해서는 채널 간 matrixing 과정을 통해 재생 채널 수를 조정할 필요가 있다. 이 때의 중요 관건은 양자화 잡음의 감소와 재생음의 자연스러움이 된다. 돌비 알고리즘에서는 아날로그 믹서를 포함한 기존 다채널 부호화기들이 사용하였던 intermediate matrix를 사용하지 않고, 각 채널들을 개별적으로 부호화하여 복호화기에서 상황에 맞도록 복호화하는 방법을 택하였다. 이로써 2 채널 뿐 아니라 mono, stereo, matrix surround 등과도 호환성을 가지게 하였다. mixdown은 시간 영역이 아닌 주파수 영역 상에서 행해지게 되며, downmix된 채널만큼의 inverse filter bank만을 필요로 하므로 시스템의 복잡성도 줄어들게 된다.

4. AC-3와 MPEG-2의 비교

1) MPEG-2의 개요

MPEG-2 오디오 부호화 알고리즘은 국제 표준화 기구(ISO : International Organization for Standardization) 산

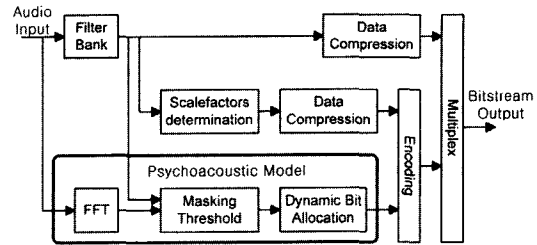


그림 5. MPEG-2 오디오 부호화 과정

하의 동영상 전문가 그룹(MPEG : Motion Picture Expert Group)에서 제안한 디지털 오디오 압축 기술로서 현재 고품질 오디오 부호화 방식에 대한 국제 표준안으로 채택되었다[15]. MPEG-2 오디오 부호화 알고리즘은 유럽에서 제안한 MUSICAM 방식[14]을 기반으로 하며 기본적인 부호화 방식은 서브밴드 부호화 방식(subband coding method)[10~12]이다. 제공하는 비트율은 최고 1011 kbps까지의 비트율을 제공하고 128 kbps에서도 CD 수준의 고품질을 얻을 수 있다. 또한 압축률을 높이기 위해서 마스킹 특성을 반영한 심리음향 모델을 사용하며 채널이 증가할 경우 채널간의 중복성을 이용하여 압축률을 더욱 향상시킬 수 있다. 그림 5에 MPEG-2 오디오 부호화 과정을 나타내었다.

그림 5에서 입력된 오디오 신호는 분석 필터뱅크를 통해서 서브밴드 영역으로 변환된다. MPEG-2 알고리즘에서 사용하는 필터뱅크는 32개의 동일 대역폭(uniform bandwidth)을 갖는 가중 중첩 가산(weighted overlap-add) 방법의 SSB(Single-Side Band) 필터뱅크이다. 이 필터뱅크는 완전 복원(perfect reconstruction)과 임계적 표본화(critical sampling)를 보장하며 각 서브밴드의 대역폭은 입력 신호에 대한 표본화 주파수의 1/32에 해당한다. 이때 필터뱅크에 사용되는 각 필터는 512-탭(tap) 저역통과 필터(low pass filter)로서 분석행렬에 의해 주파수 천이되어 32개의 동일 크기 서브밴드를 형성한다.

서브밴드 부호화 방식은 각 밴드(band)의 양자화 잡음(quantization noise) 수준을 신호의 특성과 청각적 특성에 따라 다르게 조절할 수 있는 장점을 가진다. 즉 마스킹 효과를 이용한 심리음향 모델을 사용하여 각 서브밴드에서 원음에 의해 마스킹되는 최대 잡음레벨을 결정할 수 있으며, 이를 이용해서 각 서브밴드에 할당되는 비트수를 조절함으로써 적은 비트를 가지고 주관적으로 양자화 잡음이 들리지 않도록 부호화할 수 있다.

그림 5의 심리음향 모델은 MPEG에서 제안하는 모델 1로서 신호의 스펙트럼으로부터 마스킹 임계값을 구하여 신호 대 마스크 비(SMR : Signal-to-Mask

Ratio)를 계산하는 과정이다. 먼저 입력신호에 대해서 해닝(Hanning) 윈도우를 취하고 1024-point 고속 푸리에 변환(FFT : Fast Fourier Transform)을 수행하여 전력 밀도 스펙트럼(power spectral density)을 구한다. 계산한 스펙트럼으로부터 각 서브밴드에서의 음압레벨(sound pressure level)을 구하고 순음(tonal)과 잡음(non-tonal) 성분을 찾는다. 다음으로 순음성분과 잡음성분을 간축하고 개별 마스크 임계값을 계산한 후 절대 가청 한계와 비교하여 전체 마스크 임계값을 구해낸다. 마지막으로 음압레벨로부터 최종 마스크 임계값을 빼주면 신호 대 마스크 비를 구할 수 있다.

그림 5의 크기인자 결정과정(scalefactor determination)은 서브밴드 샘플값을 부호화하는데 APCM(Adaptive Pulse Coded Modulation)을 적용하기 위해서 정규화에 필요한 크기인자(scalefactor) 정보를 결정하는 과정이다. 먼저 각 서브밴드의 매 12개의 샘플에 대한 절대값 중 최대값을 찾아 MPEG에서 제안하는 63개의 크기인자와 비교하여 크기인자 정보를 결정한다. 이렇게 결정된 크기인자는 해당 인덱스를 부호화하여 전송하며 중복 전송을 막아 전송 비트수를 줄이기 위해 크기인자 선택정보(scalefactor selection information)를 사용한다.

비트할당 과정은 프레임 전체의 잡음 대 마스크 비(NMR : Noise-to-Mask Ratio)를 최소화하도록 수행된다. 먼저 각 서브밴드마다 SNR(Signal-to-Noise Ratio)에서 SMR을 뺀 MNR(Mast-to-Noise Ratio)을 구한 후, 최소 MNR을 갖는 서브밴드에 비트를 할당하는 과정을 프레임내 사용 가능한 비트를 모두 할당할 때까지 반복해서 수행한다.

비트할당 정보와 크기인자 정보를 가지고 샘플을 양자화하는 과정은 대칭형(mid tread) 선형 양자화를 사용한다. 먼저 크기인자를 사용하여 샘플을 정규화시키고 양자화한 후 동기신호(syncword)와의 혼동을 막기 위해 부호를 반전한다. 이렇게 양자화된 샘플은 필요에 따라 비트수를 더욱 줄이기 위해 그룹화 기법을 적용한다.

MPEG-2 부호화 알고리즘이 사용하는 다채널 오디오 압축방식에는 동적 전송채널 전환 방법, 다이내믹 크로스토크, 적응 다채널 예측, 중앙 채널의 가상 부호화 방법 등이 있다. 동적 전송채널 전환방법은 MPEG-1과의 호환성을 위해 사용하는 채널 매트릭싱에 의해 발생하는 잡음을 줄이는 방법이고, 다이내믹 크로스토크는 고주파 대역에서의 귀의 청각특성을 이용하여 비트율을 낮추는 방법이다. 적응 다채널 예측은 채널 상호간의 중복성을 제거하기 위한 부호화 방법이며, 중앙 채널의 가상 부호화 방법은 중앙 채널을 좌, 우 채널에 섞어 보냄으로써 중앙 채널을 전송하는데 필요한 비트율을 줄이는 방법이다.

2) AC-3와 MPEG-2의 알고리즘 비교

앞서 언급한 바와 같이 AC-3와 MPEG-2는 각각 다른 응용분야에서 표준으로 공존하고 있다. 두 알고리즘 모두 고품질의 오디오 신호를 얻는 것을 목적으로 비슷한 구조와 연산을 갖지만, 응용분야가 다르므로 요구되는 비트율도 다르고, 시스템의 복잡도 등도 달라지게 되었다. 여기에서는 AC-3와 MPEG-2에서 나타나는 알고리즘 상의 특징에 대해 알아본다.

AC-3와 MPEG-2의 알고리즘에서 공통적으로 등장하는 중요한 특징은 필터뱅크와 심리음향 모델을 사용한다는 것이다.

① 필터뱅크

필터뱅크는 기본적으로 두가지로 분류할 수 있다. 하나는 MPEG 계층-1과 계층-2 등에서 사용되는 서브밴드(Sub-band) 코딩이고, 다른 하나는 Dolby AC-3, MPEG 계층-3, MPEG NBC 등에서 사용되는 변환(Transform) 코딩이다. 서브밴드 코딩은 일반적으로 32개의 밴드를 갖는데 비해, 변환 코딩은 256~1024개 정도의 밴드 수를 갖는다. 이러한 밴드 수의 차이는 바로 주파수축 해상도 차이로 나타나는데, 밴드 수가 많은 후자 쪽이 당연히 주파수축 해상도가 뛰어나게 된다. 주파수축 해상도가 높아질수록 밴드 폭이 좁아지고 밴드 수가 많아지므로, 이렇게 많은 밴드들 중에서 신호가 들어있는 것만을 코딩할 수 있다. 따라서 코딩의 이득(Coding Gain)이 높아지는 장점을 가진다. 그러나 주파수축 해상도가 높아지면서 시간축 해상도가 낮아지므로, 프레임 내에서 천이가 발생하는 신호(Transient signal)의 경우 별도의 특별한 처리가 필요한 단점이 있다.

Dolby AC-3는 TDAC에 기초한 변환방식을 사용한다. AC-3는 MPEG 계층-1, 계층-2 보다 높은 주파수축 해상도를 가지며, 시간축에 대해 가변적인 변환을 하므로 필요한 경우 높은 시간축 해상도를 가질 수 있다.

MPEG 계층-1과 계층-2는 서브밴드 코딩을 사용한다. 여기서는 폭이 일정한 32개의 밴드를 갖는 PQMF 필터뱅크를 사용한다. 밴드 폭이 일정하기 때문에, 사람의 청각특성을 고려한 임계 대역(Critical Band)과 비교하면 저주파(100Hz 이하)에서의 해상도가 떨어진다. MPEG 계층-3은 더 좋은 주파수축 해상도를 얻기 위해 혼합형 필터뱅크를 사용한다. 계층-3은 가변적인 주파수/시간축 분해능을 갖는 MDCT와 병행한 서브밴드 필터방식을 사용하므로 서브밴드 필터와 TDAC 변환의 장점을 동시에 가질 수 있다. 그러나 경계면 사이의 Aliasing을 고려해야할 필요가 있으며 복잡도가 증가하는 단점이 있다.

표 4-1 . Dolby AC-3 와 MPEG 의 필터뱅크 특성 비교[17]

| | 필터 뱅크 | 주파수 해상도 | 시간 해상도 |
|------------|---------------|----------|---------|
| MPEG 계층 1 | PQMF | 750 Hz | 0.66 ms |
| MPEG 계층 2 | PQMF | 750 Hz | 0.66 ms |
| MPEG 계층 3 | 혼합형 PQMF/MDCT | 41.66 Hz | 4 ms |
| Dolby AC-3 | MDCT | 93.75 Hz | 2.66 ms |

② 심리음향모델 [13]

Dolby AC-3에서는 시간/주파수 표현으로부터 얻은 신호의 전력 스펙트럼 밀도를 사용한다. 첫 단계에서 선형 주파수 단위로 표현되는 전력 스펙트럼 밀도가 임계대역 단위로 변환된다. 다음 단계에서는 전력 스펙트럼 밀도는 귀의 기저층의 에너지 분포를 시뮬레이션하는 스프레딩 함수와 콘볼루션된다. 그리고 마지막으로 각 주파수 계수마다 원하는 신호대 잡음비를 구하기 위해 원래의 전력 스펙트럼 밀도에서 계산된 마스크 커브를 빼서 그 차이값에 따라 비트할당을 하게 된다. 이와같은 심리음향 모델의 매개변수 조정은 인코더에서만 계산된다.

MPEG에서는 두가지의 심리음향 모델을 사용하는 데[15], 두 모델은 응용 분야에 따라 적절하게 선택되어 사용되어질 수 있다. 각 계층에 지정된 모델이 있는 것은 아니지만, 일반적으로 계층-1과 계층-2는 모델 1을 사용하고, 계층-3은 모델 2를 사용한다. 모델 1은 FFT 스펙트럼을 순음과 잡음 성분으로 나누어 각 성분에 의한 마스크 임계값을 구한 후 절대가치 한계를 고려하여 마스크 임계값을 구한다. 모델 2는 FFT 스펙트럼을 청신경의 여기 모델인 스프레딩 함수와 콘볼루션하여 마스크 임계값을 구하는 것으로 많은 계산량이 필요하지만 마스크 특성을 보다 정확히 모델링할 수 있으므로 높은 압축률을 필요로 하는 응용분야에 이용한다.

5. AC-3의 현재와 미래

오늘날 HDTV, 디지털 위성방송 등 디지털 방송의 등장이 가시화되면서 방송의 환경이 점차 아날로그에서 품질이 뛰어난 디지털 환경으로 바뀌어가고 있다. 특히 여기에서 언급한 여러 디지털 오디오 압축 알고리즘이나 디지털 영상 압축 알고리즘으로 이러한 물결을 주도하고, 또한 가속화시키고 있는 상황이다. 특히 Dolby AC-3는 ATSC에 의해 북미 HDTV 표준을 위한 오디오 부호화 시스템으로 채택되었을 뿐만 아니라, 차세대 디지털 오디오-비주얼 매체로 새롭게 등장한

DVD (Digital Video Disc)의 오디오 표준으로 사용되면서 그 전망이 더욱 밝아지고 있다.

현재 AC-3는 DVD Player가 등장하면서 이미 상용화가 되어 일반인들에게 사용되고 있다. AC-3 알고리즘이 복잡하고 많은 연산을 요구하지만, 고성능의 프로세서들이 등장하면서 실시간으로 디코딩이 가능한 시스템들이 개발되었다. DVD Player와 같은 AC-3 디코딩 시스템에는 범용 DSP 프로세서가 사용되기도 하지만, 많은 시스템이 AC-3 디코딩 전용 프로세서를 사용한다. 이러한 전용 프로세서는 현재 여러 DSP 프로세서 관련 회사에서 개발되어 있는데, 이러한 프로세서를 사용하면 AC-3 디코딩 시스템을 개발하는데 많은 시간과 비용을 절약할 수 있으므로 AC-3 시스템의 수요가 늘어날수록 이와같은 프로세서의 수요 또한 늘어날 것으로 예상된다. 따라서 구조가 간단하며 크기가 작으면서도 높은 성능을 가진 디코딩 프로세서를 설계하는 것이 AC-3 관련 연구의 중요한 주제로 떠오르고 있다.

본 연구실에서는 Dolby AC-3의 5.1 채널 비트열을 디코딩 할 수 있는 디코더 프로세서에 대한 연구가 진행되어 있으며, 또한 Dolby AC-3와 MPEG-2 오디오 모두를 디코딩 할 수 있는 공용 디코더에 대한 연구도 되어 있다.[2] Dolby AC-3 와 MPEG-2 오디오 압축 알고리즘은 기본적으로 비슷한 구조를 가지며 유사한 연산을 행하므로, 이와같은 공용 디코더의 설계가 가능하다. 따라서 디지털 오디오 방식의 저장매체나 디지털 오디오 방송 등 두 표준안이 공용되는 분야에서 단일 시스템으로 두 복호화 과정을 모두 지원할 수 있으므로 디코더를 구성하는데 필요한 비용을 줄일 수 있다.

이 글에서는 Dolby 사의 AC-3 오디오 압축 알고리즘 전반에 대해 알아보았다. Dolby AC-3가 나오기까지의 역사와 알고리즘 상의 특징에 대해 알아보았으며, 또다른 표준인 MPEG 오디오 압축 알고리즘과도 비교해 보았다. 위에서 살펴본 바와 같이 80년대 후반부터 90년대 초반에 이르기까지 오디오 압축 알고리즘의 발전은 눈부신 것이었다. 이제는 이러한 발전을 바탕으로 뛰어난 음질의 오디오를 보다 저렴한 비용으로 저장하고 전송할 수 있게 되었다.

현재 여러 디지털 매체의 등장으로 여러 오디오 압축 알고리즘의 상업적 가치가 높아가고 있다. 특히 Dolby AC-3는 최근 DVD의 오디오 표준으로 사용되면서 앞으로 많은 사용자를 확보할 것으로 예상되며, HDTV, 위성방송, 그리고 케이블 디지털 텔레비전 등의 여러 분야에서도 상업적 발전이 계속될 전망이다.

참고 문헌

- [1] Advanced Television System Committee(ATSC) Standard Doc. A/52, "Digital Audio Compression Standard(AC-3)", Nov., 1994
- [2] 고우석, "AC-3와 MPEG-2 오디오 공용 복호화기의 VLSI 설계", 연세 대학교석사 학위 논문, Dec., 1997
- [3] M. Iwadare, et al. "A 128 kbit/s hi-fi audio codec based on adaptive transform coding with adaptive block size MDCT." IEEE J. Selected Areas Comm. pp. 138-144, 1992.
- [4] N. S. Jayant and P. Noll, "Digital Coding of Waveforms : Principles and Applications to Speech and Video". Englewood Cliffs, NJ: Prentice Hall, 1984.
- [5] John P. Princen, "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation." IEEE. Tans. on Acoustics, Speech, and Signal Processing, Vol ASSP-34 No. 5, pp. 1153-1161 Oct., 1986.
- [6] J. P. Princen, A. W. Johnson and A. B. Bradley, "Subband/Transform Coding Using Filter Bank Designs Based On Time Domain Aliasing Cancellation." Proc. ICASSP pp. 2161-2164, 1987.
- [7] T. D. Rossing, The Science of Sound, Addison Wesley, 1990.
- [8] D. Meares "High-Definition Sound of High Definition Television" AES/SMPTE Joint Television Conference, Detroit, Feb. 1-2, 1991
- [9] Craig C. Todd, et al. "AC-3: Flexible Perceptual Coding for Audio Transmission and Storage," Audio Engineering Society, 1993.
- [10] Joseph H. Rothweiler, "Polyphase Quadrature Filters - A New Subband Coding Technique", Proc. ICASSP, pp. 1280-1283, 1983
- [11] R. N. J. Veldhuis, M. Breeuwer and R. G. van der Waal, "Subband coding of digital audio signals," Philips J. Res., vol. 44, no. 2-3, pp. 329-343, 1989
- [12] R. G. Waal and R. Veldhuis, "Subband Coding of Stereophonic Digital Audio Signals," IEEE ICASSP, pp. 3601-3604, 1991
- [13] 정남훈, "VLSI를 이용한 MPEG-2 오디오 부호화기 설계", 연세 대학교석사 학위 논문, Dec., 1996
- [14] Y. F. Dehery, et al. "A MUSICAM source codec for digital audio broadcasting and storage." Proc. ICASSP pp. 3605-3608, 1991
- [15] ISO/IEC JTC1/SC29/WG11 No.71 "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s - CD 11172-3 (Part3. MPEG-Audio)"
- [16] ISO/IEC JTC1/SC29/WG11 No.703 "Generic Coding of Moving Pictures and Associated Audio - CD 13818-3 (Part3. MPEG-Audio)", Mar., 1994
- [17] K. Brandenburg, M. Bosi, "Overview of MPEG-Audio: Current and Future Standards for Low Bit-Rate Audio Coding", The 99th AES Convention, Vienna, Preprint No.4130, Oct., 1995

필자 소개



윤 대 희

- 1951년 5월 25일생
- 1977년 2월 연세대학교 전자공학과 졸업 (공학사)
- 1979년 8월 Kansas State University 전기공학과 졸업 (공학석사)
- 1982년 8월 Kansas State University 전기공학과 졸업 (공학박사)
- 1982년 8월 ~ 1985년 6월 University of Iowa 전기공학과 조교수
- 1985년 9월 ~ 현재 연세대학교 전자공학과 교수



이 근 섭

- 1975년 2월 1일생
- 1997년 2월 연세대학교 전자공학과 졸업 (공학사)
- 1997년 8월 ~ 현재 연세대학교 전자공학과 석사과정