

주 제

MPEG-4 오디오 구현 기술 동향

연세대학교 박영철

차례

I. 서론

II. MPEG 오디오 부호화 기술 표준화 역사 및 동향

III. MPEG-4 오디오 구현 기술

IV. 결론

I. 서론

2000년대에 들어 디지털 이동통신 단말기와 포터블 멀티미디어 기기를 중심으로 고품질의 온라인 디지털 콘텐츠를 제공하는 새로운 형태의 서비스가 등장하고 있다. 디지털 오디오 기기와 달리 온라인 서비스 응용 분야에서는 제한된 채널 대역으로 보다 많은 데이터를 전송하는 것은 중요한 문제이기 때문에, 서비스되는 디지털 오디오 신호의 지각적인 음질을 떨어뜨리지 않고 최대한으로 압축하는 고품질 오디오 부호화 기술의 필요성이 절실하다.

최근 사용되고 있는 대부분의 고품질 오디오 부호화 기술은 ISO/IEC MPEG(Moving Pictures Experts Group)을 중심으로 제정된 표준 기술들이다. 1992년 제정된 MPEG-1으로부터 가장 최근 제정된 MPEG-4에 이르기까지 상당한 부호화 기술의 발전이 이루어져 왔다. MPEG-4 오디오는 뛰어난

부호화 효율성 때문에 이전의 표준 기술보다 훨씬 다양한 분야에서 응용되고 있다. 가능한 응용분야는 인터넷 스트리밍 서비스, 디지털 라디오 방송, 디지털 위성 및 케이블 방송, 휴대용 멀티미디어 플레이어, 데이터 저장 장치, 3세대 이동전화, 무선 네트워크 멀티미디어 서비스, 양방향 통신 등등 매우 다양하다.

MPEG-4 오디오 기술이 보다 많은 분야에 활용됨에 따라 복호화기/부호화기도 다양한 형태의 하드웨어에서 적응적으로 구현되어야 할 필요성이 대두되고 있다. 이동통신 단말기와 DMB와 같은 포터블 기기의 발달과 함께 최근의 경향은 주로 저전력으로 동작하는 전용 프로세서에서 동작하는 시스템 개발이다. MPEG-4 오디오 기술이 이러한 저전력 기기에서 고품질의 서비스를 구현하기 위해서는 부호화 및 복호화 알고리즘의 연산량 및 메모리 사용량의 최적화가 선행되어야만 한다. 최적화는 구현 하드웨어의 특성을 최대한 반영하는 소극적인 방법뿐만 아니라

성능대비 알고리즘을 복잡도를 최소화하는 적극적인 방법까지 동원해야 하는 경우가 대부분이다. 본 글에서는 현재 가장 활발하게 응용되고 있는 MPEG-4 오디오 부호화기에 적용되고 있는 구현 기술들의 동향 및 특성을 소개하기로 한다.

II. MPEG 오디오 부호화 기술 표준화 역사 및 동향

MPEG-1 오디오 표준안[1]은 변화 부호화 방식을 사용하며, 구현의 복잡성과 음질에 따라 3개의 계층으로 나뉜다. 계층-1이 가장 간단하며, 계층-3이 가장 복잡한 반면 부호화 효율이 높다. MPEG-1은 모노와 스테레오 음원을 처리할 수 있으며, 32, 44.1, 48kHz의 샘플링 주파수를 지원한다.

MPEG-1은 1994년 MPEG-2 표준안을 확장되었는데, MPEG-2는 5.1 채널까지를 수용하며 96kHz 까지의 샘플링 주파수를 지원하고, 약간의 확장성(scalability)을 지원하도록 하였다[2]. 이후 MPEG에서는 MPEG-1 포맷과의 호환이 필요 없는 응용 분야에의 활용을 위한 권고안으로 1997년 새로운 멀티채널 오디오 부호화 방식의 국제 표준안으로 MPEG-2 NBC(Non-Backward Compatible)(이후 Advanced Audio Coding(AAC)로 명명되었다)를 발표하였다[4]. MPEG-2 BC 버전과 마찬가지로 이 새로운 버전에서도 계층화된 구조를 사용하였는데, 음질, 메모리, 전력 요구량의 손익을 고려해서 세 가지 프로파일(Main, Low Complexity: LC, Scalable Sampling Rate: SSR)을 지원한다. AAC는 매우 낮은 비트율에서 방송 음질 수준의 오디오를 제공하기 위해서, 고해상도 필터뱅크, 예측기법(prediction), 허프만 부호화 등을 결합하여 사용한다.

MPEG-4 표준화 기술은 MPEG-2 기술을 근간

으로 부호화 효율을 혁신적으로 향상시킨 것으로, 다양한 부가기능을 추가하였다[3]. MPEG-4 표준안은 오디오 신호를 2kbps에서 64kbps까지의 비트율에서 최상의 품질을 얻을 수 있도록 확장성(scalability)을 일반화하였고, 동시에 여러 가지 부가기능을 지원하기 위해 파라메트릭 부호화, 음성 부호화 기술, 그리고 범용 오디오 부호화 기술을 하나의 구조안에 통합하였다. MPEG-4에서는 오디오 압축 성능을 향상을 위해 PNS(Perceptual Noise Substitution), LTP(Long-Term Prediction)와 같은 부가적인 툴을 지원하며, 매우 낮은 비트율에서 성능을 최적화한 TwinVQ, 부호화 지연을 감소시킨 LD(Low Delay) AAC, 에러가 비교적 큰 채널을 위한 에러 복원(error-resilience) 툴을 지원함으로써 성능을 최적화하였다[5].

MPEG-4에서는 두 가지 방법으로 확장성을 제공한다. 첫 번째는 AAC Scalable로 불리우는 기술로 8kbps 단위의 비교적 큰 단위로 확장이 가능하며 [5], 두 번째는 AAC의 무손실 부호화 기능을 채널 대역폭에 따라 1kbps 단위로 확장할 수 있는 기법으로 대신 한 BSAC(Bit-Sliced Arithmetic Coding) [6] 부호화 방법을 통해서이다. MPEG-4에서는 또한 비트율이 낮아지면 신호의 품질을 유지하기 위해서 신호의 대역폭을 제한하던 기존 부호화 방법의 문제점을 해결하기 위해 고주파 대역의 신호 성분을 낮은 주파수 대역의 스펙트럼과 고주파 대역을 표현할 수 있는 부가적인 파라메타들을 사용하여 추정하는 SBR(Spectral Band Replication) 툴을 정의하고 있다[7]. mp3PRO와 aacPlus는 SBR 기술을 기존의 MP3 및 AAC 기술과 각각 결합하여 부호화 효율을 높이고 제한된 대역의 전송에서 오디오 신호의 부자연스러움을 감소시킨 기술을 일컫는다. 일반적으로 SBR 기법은 고주파 신호 성분을 채널당 1~3kbps 정도의 정보량으로 묘사하기 때문에 기존

의 T-F 변화 기반의 부호화 기법에 비해 효율이 훨씬 높다.

최근 전통적인 조인트 스테레오 기술을 확장하여 만들어진 BCC(Binaural Cue Coding)는 SAC(Spatial Audio Coding) [8]의 근간이 되는 기술로서, 관련 기술이 최근 'MPEG 서라운드'라는 이름으로 MPEG에서 표준화가 진행 중이다. 이 기술은 다중 채널 오디오 신호를 하나 또는 두 개의 채널 신호로 결합된 신호에 인간의 지각특성을 의미하는 공간단서(spatial cue)를 부가정보로 전송함으로써 현저히 낮은 전송률에서 다중 채널 신호를 전송할 수 있는 기술이다. 또한 다중 채널 신호로부터 역방향 호환이 되도록 기본 구조를 구성하였기 때문에 기존의 모노 채널이나 스테레오 채널을 수신할 수 있는 장치와 완벽한 호환성을 제공한다. BCC에서 공간 단서로 제공하는 파라메타는 ICTD(Inter-Channel Time Difference)와 ICLD(Inter-Channel Level Difference), ICC(Inter-Channel Coherence) 정보이다[8].

III. MPEG-4 오디오 구현 기술

MPEG-4 오디오 부호화/복호화기 구현을 위해 다양한 하드웨어 플랫폼이 선택될 수 있다. 크게는 DSP 칩을 이용한 구현[9], ARM 코어와 같은 MCU를 이용한 소프트웨어 구현[10], ARM 코어에 DSP 기능을 갖는 하드웨어 가속기를 결합한 하드웨어/소프트웨어 코디자인 방법[11], 그리고 마지막으로 ASIC 기술을 이용한 전용 프로세서 개발[12]과 같은 선택을 할 수 있다. 각 방법이 갖는 장단점을 활용한다면 상당히 다양한 알고리즘들이 시도될 수 있으나 본 글에서는 모든 종류의 하드웨어에 일반적으로 적용될 수 있는 구현 기술들에 대해 살펴본다.

MPEG-4 부호화기 중 가장 활용도가 높은 부호화기는 AAC 부호화기이다. 그리고 MPEG-4 BSAC은 국내 지상파 DMB(Digital Multimedia Broadcasting)의 표준으로 채택되면서 그 활용도가 높아진 상태이다. 이 두 부호화 기술은 알고리즘의 기본 구조와 툴들을 공유하고 있기 때문에 구현 기술상 차이는 크지 않다. 다만 비트할당과정에서 서로 차이를 보이고 있으며, 연산량 차이는 대체로 3(AAC):4(BSAC) 정도의 비율이고, 메모리 사용량은 최대 3(AAC):8(BSAC) 정도이다. 따라서 AAC와 BSAC을 구별하여 구현기술을 관찰하는 것은 적절하지 않을 수 있다. 본 글에서는 MPEG-4 AAC 부호화기를 중심으로 BSAC에도 적용 가능한 구현 기술들을 살펴보기로 한다.

3.1. MPEG-4 오디오 부호화기 툴

MPEG-4 오디오 부호화기는 MPEG-2 AAC에서 사용하는 허프만 부호화, MDCT 필터뱅크, 비선형 양자화기, TNS(Temporal Noise Shaping), 조인트 스테레오 부호화 등과 함께 부호화 효율을 개선하기 위한 새로운 기능들의 툴들을 제공한다.

PNS(Perceptual Noise Substitution)

PNS 툴은 입력 오디오 신호의 파형 자체를 부호화하는 대신 매개적 표현이 가능한 잡음 신호로 나타냄으로써 AAC의 부호화 효율을 높일 수 있는 툴이다. 잡음이라고 판단되는 스케일팩터 대역은 잡음 치환 플래그와 스펙트럼 계수의 전체 파워로 표현되며, 복호화기에서는 전달된 전체 파워를 갖는 랜덤 신호를 발생시켜 스펙트럼 계수를 대체한다.

LTP(Long-Term Prediction)

AAC의 부호화 효율을 향상시키기 위해 사용되는

또 다른 틀이 LTP 틀이다. LTP는 현재와 이전 프레임 사이에 후방 예측을 수행함으로써 프레임 간 정보들 간의 시간적인 중복성을 제거하기 위한 것이다. 이전 프레임의 스펙트럼 계수들은 시간 영역 신호로 변환되고 역 TNS 필터에 의해서 필터링 된 후, 딜레이와 이득으로 나타나는 예측 매개변수의 최적값을 추정하게 된다.

예측된 신호 성분은 다시 스펙트럼 정보로 변환되고 TNS 필터링 된 후, 원래의 스펙트럼 계수와의 차이를 구함으로써 잔차성분을 얻게된다. 각각의 스케일팩터 밴드에서 원래 신호를 선택할 것인지 잔차신호를 선택할 것인지에 대한 결정은 주파수 선택 스위치(frequency selective switch)에 의해 이뤄지며, 이 과정을 통해 더 작은 비트율을 필요로 하는 신호를 선택하여 부호화하게 된다.

TwinVQ

TwinVQ는 극 저전송율에서 좋은 부호화 효율을 얻기 위해 만들어진 부호화 틀이다. 먼저 스펙트럼 계수들은 정해진 크기 범위에 들도록 정규화되고, 재배치(interleaved)된 후 서브벡터 단위로 나뉜다. 양자화 잡음을 마스킹 곡선 아래로 제어하기 위해 저주파 계수 양자화를 위한 비트 수요가 고주파 계수 양자화를 위한 수요보다 더 커지게 된다. 따라서 일정한 양자화 비트를 가진 서브벡터를 얻을 수 있도록 스펙트럼 계수들이 재배치된다. 이 과정이 끝나면 서브벡터들은 가중 왜곡 측정(weighted distortion measure) 방법을 통해 선택된 최적 토드북을 사용하여 양자화된다.

저지연(Low Delay) AAC

MPEG-4 T/F 부호화 알고리즘에 의한 시간지연은 수백밀리에 해당하기 때문에 양방향 통신에 적합하지 못하다. 이런 문제를 해결하기 위해 MPEG4-

AAC는 프레임 길이를 512/480으로 줄인 (원래길이 1024/960) 저지연 오디오 부호화 모드를 지원하고 있다. 윈도우 스위칭 정보 추출을 위해 필요한 사전 지연(look-ahead delay)을 피하기 위해 저지연 AAC는 윈도우 스위칭을 지원하지 않는다. 입력 신호가 여기 상황일 때 TNS의 성능을 향상시키기 위해 중첩율이 낮은 윈도우가 사용되며 시간 지연을 감소시키기 위해 비트저장(bit reservoir) 기법은 최소한으로 사용되거나 사용되지 않는다.

에러 강인성

에러 강인성을 향상시키기 위해 MPEG-4에서는 VCB11(Virtual codebook tool)을 사용한다. 이 방법은 큰 스펙트럼 계수를 갖는 스케일팩터 대역에서의 에러 복원력을 높임으로써 지각적으로 쉽게 인지되는 대역의 비트에러율을 낮추는 방법이다. 지나치게 큰 값을 발생시키는 에러들을 검출하기 위해 각기 다른 최대값을 갖는 virtual 코드북이 사용된다. 또한 가장 중요한 스펙트럼 계수에서의 에러 전파를 피하기 위한 기법들이 사용되고 있다.

3.2. MPEG-4 AAC 복호화기 구현 기술

저전력 시스템으로 복호화기를 구현하는 경우, 반드시 필요한 과정이 제한된 유효자리수를 갖는 프로세서에서 성능을 평가하기 위한 인티저 프로그래밍이다. 최적 유효자리수는 출력 오디오 샘플이 만족해야 하는 기준[13]을 바탕으로 결정된다. 여러 가지 실험에 의하면[14], 대체로 24비트 유효자리수를 갖는 프로세서를 사용할 경우 기준을 만족할 수 있으며, 유효자리가 20비트 정도인 경우에도 복호화기의 품질을 기준에 근사화시킬 수 있다. 그러나 범용 프로세서의 경우, 유효자리수를 임의로 변경할 수 없으므로

상당한 double-precision 연산을 동반하기도 한다.

MPEG-4 AAC 복호화기 구현을 위한 시작점은 MPEG-2 AAC 메인 및 LC(Low Complexity) 프로파일을 구현하기 위해 최적화된 복호화기 소프트웨어일 것이다. MPEG-2/4 AAC에서 메인 프로파일과 LC 프로파일의 차이는 예측 틀의 사용여부이다. 예측 틀을 제외하면 나머지는 허프만 디코딩, 역양자화, 스테레오 프로세싱, MDCT 합성으로 구성된다. <표 1>에 MPEG-2 AAC LC 복호화기를 구현하였을 때 연산량을 전체 연산량 대비 각 연산블럭이 차지하는 연산량의 비율로 정리하였다. [10]

<표 1> MPEG-2 AAC LC 복호화기 계산량

연산블럭	Bitstream Deformatting	Huffman Decoding	TNS	MDCT 합성
전체연산량 대비 비율 (%)	18	6	22	54

MPEG-2/4 AAC 복호화 단계 중 MDCT 합성 과정은 몇 가지 알고리즘 중 선택을 가능하게 한다. 이는 사용하는 구현 시스템의 특성에 따라 최상의 성능을 발휘할 수 있는 알고리즘이 달라질 수 있기 때문이다. 즉, 구현하는 프로세서의 메모리 접근 방식이나 연산기의 구조, 메모리 사용량 등을 고려하여 분석함으로써 종합적인 알고리즘의 구현 성능을 평가할 수 있다. <표 2>에 프로세서 구조에 따른 MDCT 알고리즘의 구현 결과를 비교하고 있다[15]. DCT 정의식을 직접 구현하는 경우와 고속 구현 알고리즘을 프로세서별로 각각 구현하여 그 결과를 비교한 것이다.

DSP의 경우, MAC 연산을 기반으로 하는 직접 구현 방법은 프로세서의 구조가 달라지더라도 수행 사이클 면에서 크게 차이가 나지 않았지만, 고속 알고리즘과 같이 메모리 접근이 빈번하고 불규칙적인 연산에서는 메모리 접근 방식에 따라서 많은 차이를 보인다. 한편 MCU는 캐쉬 기반의 메모리 구조로 메모리

<표 2> 프로세서 구조에 따른 MDCT 구현 결과 비교[15]

연산알고리즘	고속 DCT 알고리즘[*] 기반			직접 구현 DCT 기반		
곱셈	80			1024		
덧셈	209			1024		
Processor	A	B	C	A	B	C
수행 사이클	149	2192	3693	1323	1312	17258

A type : 평행 연산(parallel instruction), 평행 메모리 접근(parallel move) 지원, 개선된 하바드 구조를 사용
 B type : 제한된 평행 연산, 평행 메모리 접근 지원 하지 않음, 하바드 구조 사용
 C type : 캐쉬 기반으로 메모리 접근 지연, 평행 연산 지원하지 않음, 폰 노이만 구조 사용

접근하는데 지연이 발생 할 수 있고, MAC를 사용할 수 없는 구조이므로 직접 구현 방법의 알고리즘 규칙성이 장점이 되지 못하며, 오히려 많은 메모리 접근과 절대적으로 많은 연산량으로 인하여 MCU에서는 고속 DCT를 기반으로 하는 구현방법에 비해 소요되는 클럭 수가 현저히 증가함을 알 수 있다.

MPEG-2 AAC 디코더에 PNS 틀을 위한 디코딩 블럭을 추가하면 일단 MPEG-4 AAC 복호화기를 위한 소프트웨어가 구성된다. PNS 틀은 잡음 발생기와 단순한 스케일링로 구성되기 때문에 상대적으로 간단하게 구현된다. 추가적인 연산량은 대체로 잡음 발생기를 구현하는 과정에서 생기지만 전체 연산량에 비하면 미미하다. 하지만 LTP 틀을 사용하는 경우에는 MDCT 변환을 채널마다 추가적으로 수행해야 하기 때문에 계산량이 상당히 증가하게 되는데, 구현 방법에 따라 전체 계산량의 약 25~60% [16] 정도가 추가된다.

3.3. MPEG-4 AAC 부호화기 구현 기술

MPEG-4 부호화기는 임의의 오디오 비트스트림에 대해서도 완벽한 복원을 할 수 있어야 하기 때문에 표준에서 정의한 모든 프로파일, 비트율, 샘플율을 지원하도록 설계되어야만 한다. 하지만 부호화기의 경우는 이와 다르다. 부호화기는 특정 응용분야에서 요

구되는 비트율과 샘플율, 그리고 프로파일에 따라 표준에서 정의한 규격에 맞는 비트스트림을 생산해 낼 수 있으면 된다. 따라서 복호화에 비해 유연성이 높다고 할 수 있다.

하지만 부호화를 구현함에 있어 가장 큰 문제는 복잡도이다. MPEG-4 AAC 부호화기의 복잡도는 복호화기의 수배에 이르기 때문에 저전력 응용에 적합하지 않은 면이 있다. 이런 문제를 해결하기 위해 최근 많은 저전력 부호화 알고리즘이 개발되었다. <표 3>에 MPEG-4 AAC의 복잡도의 비율을 ISO reference 코드를 기준으로 각 실행 과정별로 정리하였다[17]. 비슷한 연산량 분석을 다른 문헌에서도 살펴볼 수 있다[18]. <표 3>의 분석 결과를 살펴보면 부호화기의 복잡도를 증가시키는 주요한 요인이 심리음향모델(PAM)과 비트 할당 루프임을 알 수 있다. 따라서 저전력 부호화 알고리즘들은 PAM과 비트 할당 루프의 연산량을 최소화하기 위한 방법들을 제안하고 이를 바탕으로 부호화기를 구현하고 있다.

<표 3> MPEG-4 AAC 부호화기 연산량 분포

연산블럭	PAM	MDCT 분석	비트할당	Huffman Encoding 및 기타
전체연산량 대비 비율 (%)	39	5	55	1

PAM의 간략화

MPEG-4에서 사용하는 PAM-II 모델은 크게 DFT 스펙트럼 연산, tonality 검출, 마스킹 곡선 연산 과정으로 나눌 수 있다. 각각의 과정의 복잡도를 연산량 측면에서 측정하여 정리하면 <표 4>와 같다. 상당한 연산량이 DFT 스펙트럼 연산과 tonality 검출 과정에 집중되어 있음을 알 수 있다. MPEG-4 PAM-II 모델의 DFT 스펙트럼 연산 과정을 간략화하기 위해 여러 가지 알고리즘이 제안되었다. 알고리즘은 DFT와 MDCT, MDST(Modified Discrete

Sine Transform)간의 결합 관계를 이용하는 방법 [17,19] 과 MDCT 스펙트럼만으로 입력의 크기 스펙트럼을 추정하는 방법[20]으로 나눌 수 있다. 전자의 경우는 MDST만을 추가적으로 수행함으로써 DFT와 동일한 스펙트럼 계수를 얻을 수 있다는 면에서 정확도가 높은 방법인 반면, 후자는 DFT와의 스펙트럼 차이는 있으나 분석 필터 과정에서 얻어진 MDCT 계수만을 사용하기 때문에 상당한 계산상의 이득이 있는 방법이다. MDCT 계수만을 사용하는 경우, PAM-II가 필요로 하는 연산량의 50%만으로 마스킹 곡선을 구할 수 있으며 [20], MDCT 계수와 MDST 계수를 결합하여 DFT 연산을 하는 방법의 경우, DFT 연산량을 절반으로 줄일 수 있다[17].

<표 4> MPEG-4 AAC 부호화기 연산량 분포

연산블럭	DFT	Tonality 검출	마스킹 곡선 연산	기타
전체연산량 대비 비율 (%)	52	37	10	1

Tonality 검출 과정 또한 DFT 연산에 견줄만한 연산을 필요로 한다. PAM-II의 tonality 검출은 이웃하는 프레임의 스펙트럼 bin간의 예측을 통해서 이뤄지기 때문에 계산량이 많을 뿐만 아니라 메모리 또한 상당한 양을 필요로 한다. 이 문제를 해결하기 위해서는 PAM-I에서 사용하는 스펙트럼의 크기 비교에 의한 톤 성분 검출 방법을 사용할 수 있다[21]. 크기 비교에 의한 톤 검출 결과를 tonality 인자로 변환할 수 있는 방법을 사용함으로써 연산량을 원래의 1/5로 줄이면서도 우수한 성능을 기대할 수 있다 [21].

비트할당 과정의 효율적인 구성

비트할당 과정은 양자화 잡음을 제어하는 외부 루프와 비트율을 제어하는 내부 루프로 구성되어 있다.

외부루프에서는 각 스케일팩터 대역별로 적절한 스케일팩터를 조절하는 기능을 수행하고, 내부루프에서는 글로벌 계인을 조정함으로써 전체 비트율을 조정하는 역할을 수행하며, 이 두 루프는 이중으로 구성되어 있어서 루프 실행 횟수에 따라 계산량의 증가가 급격해진다[5].

MPEG 오디오에서 사용하는 비트할당 과정은 이중 루프 구조 때문에 계산량이 절대적으로 많은 뿐만 아니라, 계산량이 변화가 프레임마다 크기 때문에 연산량의 기준선을 책정하기가 어렵다는 문제가 있다. 이런 문제를 해결하기 위해 가장 일반적으로 시도되고 있는 방법이 이중 구조의 루프는 분리하여 처리하는 기술이다. <표 5>은 루프 분리 과정을 통해서 얼마나 연산량을 줄일 수 있는지를 예시하고 있다.

루프분리를 위해서는 스케일팩터를 적절히 예측할 수 있는 모델이 필요하다. 양자화 잡음은 외부 루프의 스케일팩터에 의해 제어될 뿐만 아니라 내부 루프에서 조정되는 글로벌 계인에 의해서도 영향을 받는다. 따라서 내부 루프를 수행하지 않은 상태에서 최적의 스케일팩터를 찾을 수는 없다. 그럼에도 불구하고 적절한 통계적인 모델을 도입함으로써 최적의 스케일팩터에 비교적 근사한 값들을 추정할 수 있다 [22,23].

<표 5> 루프분리에 의한 비트할당 과정의 연산량 감소

구분 \ 연산량	내부루프 (M-MIPS, K회) / 외부루프 (N-MIPS, J회)
이중루프 방법(ISO 방법)	$J \times M \times K + J \times N$ [MIPS / frame]
분리된 루프를 사용하는 경우	$M \times K + J \times N$ [MIPS / frame]
고속 알고리즘의 경우	$\alpha \ll (J \times N) + \beta \ll (K \times M)$ [MIPS / frame]

내부루프에서 최적의 글로벌 계인을 찾는 과정 또한 반복루프에 의해 실행된다. 내부루프에서 비선형적으로 양자화된 부호들의 전체 비트율이 허프만 부호화 과정을 거치는 동안 바뀌기 때문에, 반복적인 방법으로 글로벌 계인을 조정함으로써 목표치에 근접

한 비트율로 조절한다. 이 과정의 효율성을 개선할 수 있는 방법으로 실제 사용한 비트와 목표 비트율 간의 차이를 이용하여 글로벌 계인을 적응적으로 조정하는 방법[22]을 사용하거나, 글로벌 계인과 사용된 비트간의 관계를 모형화한 선형 모델을 사용하여 최적의 글로벌 계인을 예측하는 방법[23]을 사용할 수 있다. 이러한 고속 비트 할당 알고리즘을 사용할 경우, 기존의 ISO 방법에 비해 연산량을 최대 1/12 정도까지 줄일 수 있다.

PAM을 사용하지 않는 부호화 알고리즘

MPEG-4 오디오 부호화 알고리즘의 연산량을 최소화 하는 방법은 PAM을 사용하지 않는 것이다. PAM은 특히 낮은 비트율에서 양자화 잡음을 마스킹 곡선에 맞추어 조절함으로써 오디오의 음질을 주관적인 관점에서 최선으로 유지하는 역할을 한다. 따라서 PAM을 사용하지 않는 경우, 특히 낮은 비트율에서 음질의 열화가 발생할 수 있다. 그러나 단순화된 PAM과 효율적인 비트할당 방법을 사용하는 경우에도 부호화기의 전체 연산량이 50~60MIPS 수준이기 때문에 여전히 포터블 기기에 부담이 된다.

MPEG-4 PAM에서 제공하는 정보는 양자화기의 기준이 되는 마스킹 곡선, 순간적으로 오디오의 크기가 커지는 프레임에서 발생하는 프리에코를 방지하기 위한 윈도우 스위칭 정보, 그리고 부가적인 툴들을 위한 PE(Perceptual Entropy) 정보 등이다. 만약 윈도우 스위칭을 위한 정보를 AC-3와 같이 입력 오디오 프레임에서 직접 얻어내고[17], 비트 할당을 위한 스케일팩터를 입력 신호의 에너지로부터 근사화 시킨다면 PAM을 사용하는 부호화기와 유사한 성능을 가진 고속의 알고리즘을 만들 수도 있다. faac[24]가 그 대표적인 부호화기이다. 그러나 PAM을 사용하지 않는 경우, 낮은 비트율에서의 음질 열화가 불가피하기 때문에, 이를 보완하기 위해 비

트율에 반비례하게 신호의 대역폭을 조절하는 방법을 사용하고 있다. (그림 1)은 MPEG-4 AAC 부호화기에서 대역폭을 적절히 조절함으로써 주관적인 음질이 어떻게 보상되는 보여주고 있다. 스테레오 96kbps 환경에서 평가한 결과로서 12.5kHz의 대역폭을 갖는 경우(중간)에 비해 대역폭을 10.5kHz로 줄였을 때(오른쪽) 약 0.23 정도의 MOS 향상이 있음을 확인할 수 있다. 이 때 faac는 10.5kHz의 대역폭을 유지한다(왼쪽). 스테레오한편 PAM을 사용하지 않는 부호화기의 경우 복호화기에 비해 150%이내의 계산량 범위에서 구현이 가능하다.

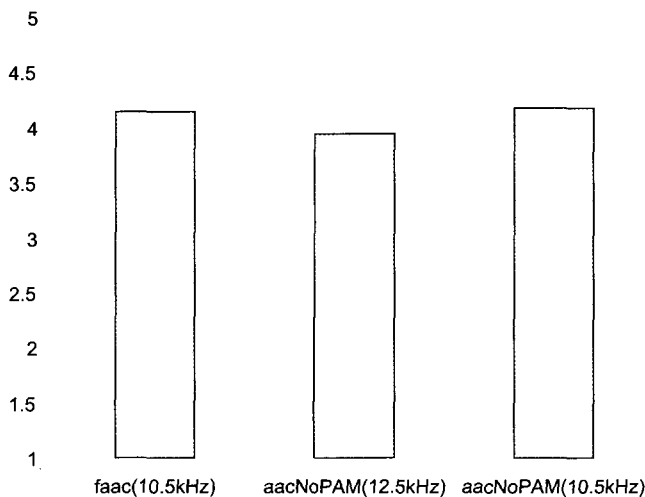
3.4 MPEG-4 HE AAC 복호화기 구현 기술

HE(High Efficiency) AAC는 AAC 기술에 parametric 스테레오 기술과 SBR(Spectral Band Replication) 기술을 결합한 고효율의 부호화 기술이다. HE AAC는 24kbps 정도에서도 고음질의 스테레오 오디오를 얻을 수 있는 부호화 기술로 알려져 있

다. HE AAC는 입력 신호를 동일한 대역폭의 상,하 위 대역으로 분할한 후, 하위 코어대역은 기존의 AAC 부호화 기술을 사용하고, 상위대역은 SBR 기술을 적용한다. SBR 기술은 계산량이 다소 많은 서브밴드 분석,합성 과정을 기반으로 하기 때문에 복호화기의 경우 계산량은 코어대역 AAC의 1.5~2 배 정도이다. 그러나 대역 분할 과정을 통해 샘플링율이 절반으로 줄어들기 때문에 전대역 신호를 부호화하는 AAC에 비해 전체적인 계산량의 증가가 크지 않다. 24비트 연산구조를 갖는 DSP로 구현하는 경우, AAC 부호화기가 스테레오 음원을 약 20MIPS의 연산량으로 복호화하는데 비해 HE AAC 복호화기는 25~30MIPS 정도로 구현된다[25].

3.5 부호화기 구현을 위한 최적 고정 소숫점 연산

저전력 부호화기를 구현하기 위해서는 고정 소숫점 연산 구조를 갖는 프로세서를 사용해야만 한다. 부호화기에서 연산의 정확도는 복호화기 보다 더 중요



(그림 1) 오디오 대역폭 변화에 따른 부호화기의 성능 비교

한 문제이다. 특히 PAM과 비트 할당 과정에서 요구되는 많은 초월함수 연산의 정확도가 충분하지 않을 때, PAM 자체의 성능에 문제가 발생할 수도 있기 때문에 부호화기의 음질 저하를 막기 위해서는 정확한 연산이 매우 중요하다. 그러나 연산의 정확도를 높이기 위해 전체 연산량을 과도하게 높이는 것 또한 바람직하지 않다. 초월함수를 급수 표현법으로 구현하는 경우, 이런 문제가 발생할 수 있다.

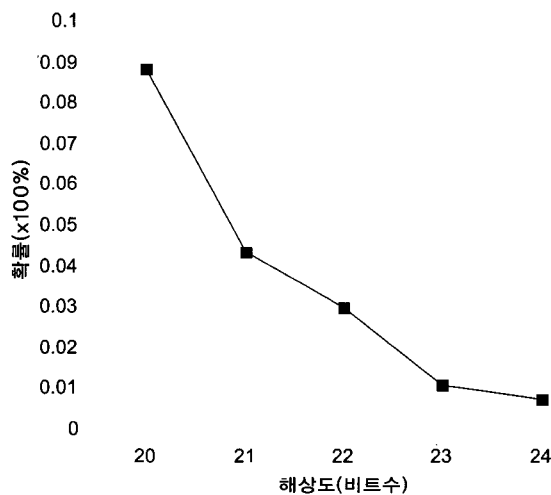
초월함수의 연산 정확도를 높이고 연산량을 최소화할 수 있는 방법이 테이블을 이용하는 방법이다 [17,26]. 그러나 이 방법은 피연산 값의 범위가 제한적일 때 유효하며, 피연산 값의 범위가 예측 불가능할 경우에는 적당한 통계적인 측정 과정을 통해 피연산 값의 범위를 제한한 다음, 이 범위를 벗어나는 값들에 대해서는 interpolation 혹은 extrapolation의 방법으로 구한 추정치를 사용할 수 있다.

고정 소숫점 연산에 의한 오차를 확률적인 모델을 사용하여 예측한 결과 [17]에 의하면, 연산기가 24비트의 해상도를 가질 때 지각적으로 큰 의미가 없는 정도의 작은 연산 오차로 PAM을 구현할 수 있음을 보

여준다. (그림 2)는 연산기의 해상도에 따른 PAM의 최대 오차와 허용 가능한 오차를 비교하여 보여주고 있다. 그림에서 세로축은 고정소수점 연산에 의해 PAM의 결과가 달라질 확률을 나타낸다. 따라서 24비트 연산기를 사용할 경우 1% 미만의 확률로 PAM 결과가 달라질 수 있음을 알 수 있다.

IV. 결 론

고품질 오디오 압축 기술은 지난 30년간 발전을 거듭해 오고 있다. 압축 기술의 발전은 멀티미디어 콘텐츠 서비스 패러다임을 바꾸는 근간이 되었으며, 끊임없이 새로운 형태의 기술들이 등장하고 있다. 전통적인 고품질 오디오 부호화 기술은 심리음향 모델을 기반으로 하는 변환 부호화 기술을 근간으로 하고 있으며, 이러한 기술의 완성도가 MPEG-4에 이르러 정점에 도달해 있다. 현재 MPEG-4의 산업적인 응용성과 서비스의 확장성을 고려하면, 이를 저전력 저비용 시스템으로 구현하는 것은 매우 중요한 문제가



(그림 2) 연산기 해상도에 따른 고정 소수점 연산 오차(점선)와 허용 가능한 최대 오차(실선)의 비교

다. 본 글에서는 MPEG-4 오디오 부호화, 복호화 기술을 저전력 저비용 시스템으로 구현할 때 사용할 수 있는 다양한 구현 기술들에 대해 살펴보았다.

[참 고 문 헌]

- [1] ISO/IEC JTC1/SC29/WG11, No.71 “Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbit/s – CD 11172-3 (Part 3. MPEG-Audio)”
- [2] ISO/IEC 13818-3:1998, Information technology -- Generic coding of moving pictures and associated audio information -- Part 3: Audio.
- [3] ISO/IEC 14496-3:2001. Information technology -- Coding of audio-visual objects -- Part 3: Audio.
- [4] ISO/IEC 13818-7:1997, Information technology -- Generic coding of moving pictures and associated audio information -- Part 7: Advanced Audio Coding (AAC).
- [5] B. Grill, “The MPEG-4 General Audio Coder,” *Proc. AES 17th Conf. on High Quality Audio Coding*, Florence 1999.
- [6] S. H. Park, Y. B. Kim, S. W. Kim, et al., “Multi Layer Bit Sliced Bit Rate Scalable Audio Coding,” *103rd AES Convention*, New York 1997, Preprint 4520.
- [7] M. Wolters, K. Kjolring, D Homm, H. Purnhagen, “A closer look into MPEG-4 High Efficiency AAC,” *115th AES Convention*, NY, USA, Oct. 2003, Preprint 5871.
- [8] J. Herre, et al., “Spatial Audio Coding: Next-generation efficient and compatible coding of multi-channel audio,” *117th AES Convention*, San Francisco, Oct. 2004, Preprint 6186.
- [9] B. Grill, S. Geyersberger, J. Hilpert, and B. Teichmann, “Implementation of MPEG-4 audio components on various platforms,” *109th AES Convention*, Los Angeles, Sept. 2000, Preprint 5270.
- [10] Keun-Sup Lee, Young-Cheol Park, Dae Hee Youn, “Software optimization of the MPEG-audio decoder using a 32-BIT MCU RISC processor,” *IEEE Transactions on Consumer Electronics*, Vol. 48, No. 3, pp. 671-676, August 2002.
- [11] 방경호, 이근섭, 박영철, 윤대회, “DSP 가속기가 내장된 RISC 프로세서 기반 MPEG/Audio 복호화기의 구현”, *한국통신학회논문지*, 제29권, 12C호, pp. 1617-1622, 2004. 12
- [12] S. Hashimoto, et al., “VLSI implementation of portable MPEG-4 audio decoder,” *Proc. 13th Annual IEEE Intl. ASIC/SOC Conf.*, pp. 80-84, Sept. 2000.
- [13] ISO/IEC 13818-4:2004. Information technology -- Generic coding of moving pictures and associated audio information -- Part 4: Conformance testing.
- [14] K. Bang, et al., “Design and VLSI implementation of a digital audio-specific DSP core for MP3/AAC,” *IEEE Trans. on Consumer Electronics*, Vol. 48, No. 3, pp.

- 790-795, August 2002.
- [15] 이재성, 박영철, 윤대희, “하드웨어 구조에 따른 고속 DCT 알고리즘의 성능 비교”, 제17회 신호처리합동학술대회, 2004. 10.
- [16] A. Servette, A. Rinotti, J. C. Martin, “Fast implementation of the MPEG-4 AAC main and low complexity decoder,” Proc. IEEE ICASSP '05, Vol. 5, pp. 249~252, March 2005.
- [17] K. S. Lee, “A Simplified MPEG-4 AAC Encoding Algorithm for Real-time Embedded Applications,” Ph.D. Thesis, Yonsei University, Feb. 2005.
- [18] Y. Lu, C. Shen and C. Chen, “A Novel Hardware Accelerator Architecture for MPEG-2/4 AAC Encoder,” Proc. IEEE ICME 2004, Taipei, Taiwan.
- [19] Y. Wang, L. Yaroslavsky, M. Vilermo, “On the relationship between MDCT, SDFT and DFT,” Proc. ICSP2000, pp. 44~47, Beijing, China 2000.
- [20] T. Tsai, Y. Wang, and S. Huang, “An MDCT-based psychoacoustic model co-processor design for MPEG-2/4 AAC audio encoder,” Proc. DAFX '04, pp. 335~338, Naples, Italy, Oct. 2004.
- [21] K. Lee, K. Yeon, Y. Park and D. H. Youn, “Effective tonality detection algorithm based on spectrum energy in perceptual audio coder,” 117th AES Convention, San Francisco, Oct. 2004, Preprint 6220,
- [22] K. H. Bang, K. S. Lee, Y. C. Park and D. H. Youn, “Fast bit allocation method for MP3/AAC encoders,” 118th AES Convention, Barcelona, May 2005, Preprint 6381.
- [23] C. Lee, et al., “A fast audio bit allocation techniques based on a linear R-D model,” IEEE Trans. Consumer Electronics, Vol. 48, No. 3, pp. 662~670, August 2002.
- [24] <http://www.audiocoding.com/>
- [25] <http://www.codingtechnologies.com/products/productsheets.htm>
- [26] S. Huang, L. Chen, and T. Tsai, “Memory and computationally efficient psychoacoustic model for MPEG AAC on 16-bit fixed-point processors,” Proc. IEEE ISCAS '05, pp. 3155~3158, 2005.



박영철

1986년 연세대학교 전자공학과 졸업

1988년 연세대학교 대학원 석사

1993년 연세대학교 대학원 공학박사

1993년 ~ 1995년 Graduate Program in Acoustics,
Pennsylvania State Univ. PostDoc.

1996년 ~ 1998년 삼성전자 반도체사업부, 의공학

연구소 선임연구원

1998년 ~ 2001년 (주)인타임 시스템 LSI 연구소 연구소장

2002년 ~ 현재 연세대학교 컴퓨터정보통신공학부 교수

관심분야 : 오디오/음성 부호화, 3D 오디오 신호처리, 적응필터, 디지털 보청기