

# 오디오 압축 기술의 표준 동향

김상욱, 김도형, 김미영 (삼성종합기술원)

## I. 서론

디지털 기술의 발전에 따라, 오디오 신호를 디지털 신호로 표현해 주는 일들이 진행되어 왔다. 그 결과 CD(Compact Disc)라는 새로운 제품이 탄생하였고, 디스크라는 저장 매체에 보다 많은 양을 저장하기 위하여 MPEG-1 표준<sup>[1]</sup>이, HDTV방송을 위하여 MPEG-2 표준<sup>[2]</sup>이, 인터넷과 통신서비스를 위하여 MPEG-4 표준<sup>[4]</sup>이 완성되었다.

MPEG 에서 사용되는 오디오 압축 방식은 이전의 정보이론(Information Theory)에 기반한 신호처리에 의존한 DPCM(Differential Pulse Coded Modulation)등의 압축 방식과 달리 인간의 인지특성 모델링<sup>[5][6]</sup>을 데이터 압축에 사용해 준 것이 특징이다. 인간의 소리에 대해 반응하는 특성을 주파수 성분에 대한 신호 값의 크기를 이용하여 모델링 하여, 각각의 주파수 신호 성분을 처리시 표현오차가 최소화 되도록 각 주파수 신호에 대해 다른 양자화기를 사용해 준다.

본 논문에서는 오디오 압축에 사용되는 인간의 인지특성 모델링 기술들에 대하여 2장

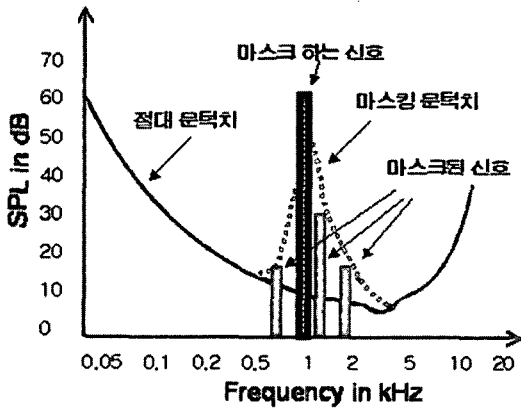
에서 다루어 준 뒤, 3장에서는 인지특성 모델링 기술들이 적용된 대표적인 압축 방식인 MPEG 오디오 방식들에 대하여 설명하고, 4장에서는 각 기술들의 성능에 대한 비교를 보인다.

## II. 오디오 압축에서 사용되는 인간의 인지특성 모델링 기술

### 1. 심리음향 모델링 (Psychoacoustic Modeling)

심리음향 모델링 기술은 인간의 소리에 대해 반응하는 특성을 주파수 성분에 대한 신호 값의 크기를 사용하여 모델링 하여, 각각의 주파수 신호 성분을 압축해 줄 수 있도록 해 주는 것을 특징으로 한다. 그림1은 마스크하는 신호에 의해 절대 문턱치가 마스킹 문턱치로 변화하고, 그 결과 마스크된 신호들이 발생하는 예를 보인다. 마스크된 신호들은 마스킹 문턱치 아래에 위치하기 때문에, 청취자에게는 들리지 않는 신호가 된다. 일

예로, 주위가 시끄럽거나, 큰 소리가 있을 때, 평소와 같은 크기로 이야기 할 경우 상대에게 들리지 않게 되는 것과 같은 현상을 말한다.



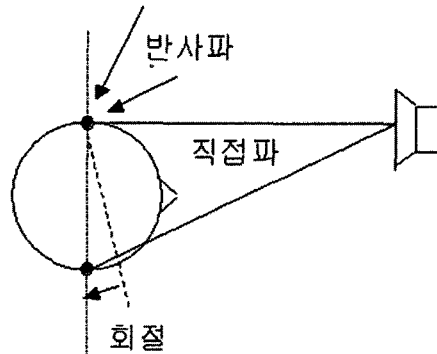
〈그림 1〉 마스크 하는 신호와 마스크된 신호

심리음향 기술을 모델링 해 적용하는 방법은 주파수 성분 별 특성을 반영하기 위해, 좌우각 필터 모델링을 해 주고, 각 주파수 성분에 대한 신호의 크기 값을 이용해, 음조 (tonality)와 마스킹 문턱치를 계산하여 사용한다.

## 2. 양이 청취 모델링 (Binaural Hearing Modeling)

귀에 도달하는 소리 신호들은 그림 2와 같이 직접파, 반사파와 회절된 신호들로 나뉜다. 어떤 음원으로 부터 소리가 양쪽 귀에 도달할 때, 음원이 존재하는 위치에 따라 소리가 왼쪽 귀와 오른쪽 귀에 도달하는 시간과 크기들이 바뀐다는 사실을 이용하여 양이 청취 모델을 적용한다. 신호 값의 차이는 ILD(Inter-aural Level Differences) 값, 신호

가 도달하는 시간 차이는 ITD(Inter-aural Time Differences) 값, 신호들간의 상관성은 IC(Inter-aural Coherence) 값을 통해 표현된다. 일반적으로, 저주파 신호일 경우에는 ITD의 영향이 ILD보다 크고, 고주파 신호일 경우에는 반대로 ILD의 영향이 ITD보다 크다고 알려져 있다.



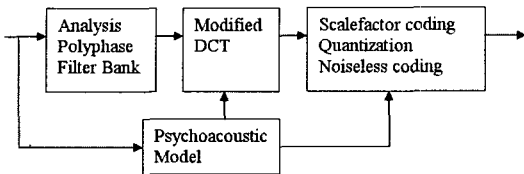
〈그림 2〉 직접파, 반사파와 회절된 신호

## III. 인지특성 모델링 기술이 사용된 오디오 압축 기술

### 1. MPEG-1 Audio

디스크라는 저장매체에 보다 많은 정보를 저장하기 위하여, MPEG-1 표준이 완성되었다. 압축을 하지 않은 경우 스테레오 오디오 신호는 초당 44100번 샘플링 되고, 16비트로 각 샘플들이 표현될 경우, 1411200 (2\*16\*44100)개의 비트가 표현에 사용되는데, MPEG-1 표준 기술에 의하여, 계층II에 의한 경우 256kbps 에서, 계층 III에 의한 경우 192kbps 에서 원음과 차이가 없는 방송용 음질을 제공할 수 있게 되었다. 여기서, 계층 II는 MUSICAM이라고도 하는 현재 유럽에서

DAB (Digital Audio Broadcasting)에 사용되는 기술이고, 계층 III은 현재 MP3라 하는 기술이다. 그림3에 MP3 부호화기의 구조를 보인다.



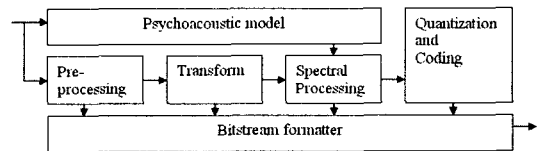
〈그림 3〉 MP3 부호화기 블록도

## 2. MPEG-2 Audio

MPEG-1에서 모노 신호와 스테레오 신호에 대하여 표준화 한 것을 MPEG-2에서는 HDTV를 대상으로 하는 5.1채널 이상의 다채널 오디오 신호를 압축하는 방식과 MPEG-1에서 지원하는 샘플링주파수 32, 44.1, 48kHz의 1/2인 16, 22.05, 24kHz까지도 지원하는 방식으로 확장하였다. 그리고, MPEG-1 방식과 호환성이 제공되는 Backward Compatible 기술<sup>2)</sup>과 호환성이 제공되지 않는 기술 Non-Backward Compatible 기술<sup>3)</sup>로 각각 표준이 완성되었다. MPEG-1과의 호환성을 조건으로 다채널 오디오 압축 방식을 개발하다 보니, 압축 방식 개발에 제약이 있어서 호환이 안되더라도 다채널 압축에 최고 성능을 제공하는 기술을 추진하였고, 그 결과 MPEG-2 AAC(Advanced Audio Coding) 표준이 개발되었다. AAC기술은 MPEG-1의 오디오 압축 기술에 TNS (Temporal Noise Shaping), prediction 기술 등을 사용하여, 압축 성능을 개선하였다.

## 3. MPEG-4 Audio AAC, BSAC

MPEG-4는 컴퓨터와 무선 그리고 통신이 결합된 환경에서 멀티미디어 응용을 위한 표준으로, 대화형 통신 기술과 Error Protection 기술이 제공되는 것이 특징이다. 대화형 통신 기술을 미디어 압축 차원에서 제공하는 것으로는 서비스 품질 제어가 가능하도록 하는 스케일러빌리티 기능이 있다. 그림 4와 같은 구조의 MPEG-4표준에 있어서 오디오 압축 방식은 MPEG-2 AAC 방식을 기본으로 사용하였고, 기능적인 틀들을 추가 하였다. AAC와 동급의 압축 품질을 제공하면서, 스케일러빌리티를 제공하는 압축 방식으로는 MPEG-4 BSAC(Bit Sliced Arithmetic Coding)가 있다.



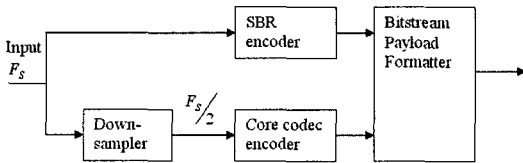
〈그림 4〉 AAC, BSAC 구성도

스펙트럴 프로세싱 부에 있어서는 MPEG-2 AAC기술에 LTP(Long Term Prediction), PNS(Perceptual Noise Substitution) 기술들이 추가로 적용되었다.

## 4. MPEG-4 HE-AAC

인간이 고주파수 음악 신호들에 대하여 상대적으로 낮은 해상도를 갖는 것과, 많은 음악 신호의 경우 반사파와 잔향 신호들을 포함한다는 것을 고려하여, 인간의 인지특성에

덜 민감한 부분을 합성된 신호로 대신 표현해 주는 주파수 대역 복제(SBR, Spectral Band Replication) 기술이 개발되었다. SBR 기술에 의해 고주파수 신호 성분들을 합성해 사용해 줌으로써 음질의 명료도 저하는 있지만, 기존의 약 50%에 해당하는 비트만 사용해서도 큰 성능 저하 없이 처리해 줄 수 있게 되었다. 그림5에 주파수 대역 복제를 사용하는 오디오 압축 방식의 예를 보인다.

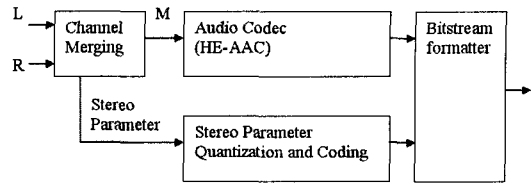


〈그림 5〉 SBR기술을 이용한 부호화기 블록도

코어 오디오 부호화기로 AAC가 사용 되는 경우 HE-AAC(High Efficiency-AAC)라 불리우고, 최근 코어 오디오 부호화기로 BSAC가 사용되어 스케일러블하게 대역 확장하는 방식이 표준화 제정 진행 중이다.<sup>[7]</sup>

### 5. MPEG-4 HE-AAC v2

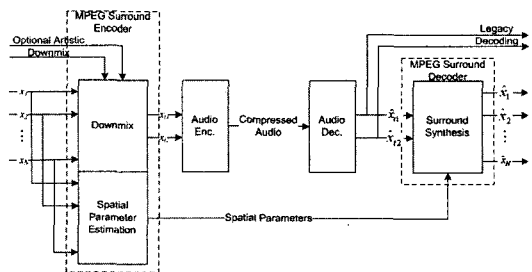
HE-AAC v2는 HE-AAC에 PS(Parametric Stereo) 기술을 사용해 준 오디오 압축 방식이다. PS 기술은 양쪽 귀를 대상으로 하는 심리음향 모델링 기술을 이용하여, 모노 신호를 가지고서 가상의 스테레오 신호를 만들어 내는 기술이다. 이러한 파라메트릭 부호화를 모노 신호에 적용해, 모노 신호로부터 스테레오 신호를 만들어 내는 방식은 그림6과 같다.



〈그림 6〉 HE-AAC v2 부호화 알고리즘

### 6. MPEG Surround

모노 신호로부터 스테레오 신호를 만들어 내는 PS기술을 다채널 오디오에 확장해, 다채널 오디오를 효과적으로 재생하고자 하는 기술이 개발 중이다. 처음에는 Spatial Audio Coding이라는 이름으로 표준 착수 되었으나, 73차 MPEG 회의에서 MPEG Surround라는 이름으로 바뀌었다. 소리가 양쪽 귀에 도달할 때, 음원이 존재하는 위치에 따라 소리가 왼쪽 귀와 오른쪽 귀에 도달하는 시간과 크기들이 바뀐다라는 사실을 이용하여 ILD, ITD, IC값들을 사용한다. 그림 7에 다채널 오디오를 위한 MPEG 서라운드 표준 부호화, 복호화 블록도를 보인다.



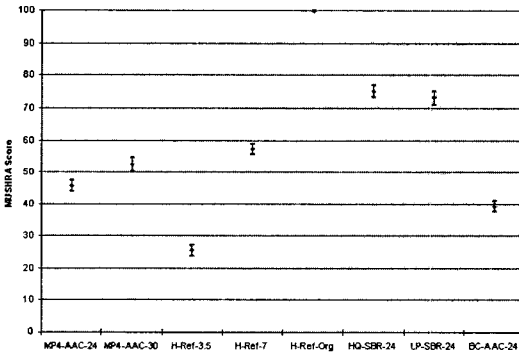
〈그림 7〉 MPEG 서라운드 부호화(Surround Coding)의 원칙

MPEG Surround의 결과로 이전에 320kbps에서만 재생되던 5.1 채널 신호를 스

테레오에 추가로 12kbps에서 32kbps의 비트율을 사용하여 유사한 효과를 제공할 수 있게 된다.

#### IV. 성능 비교

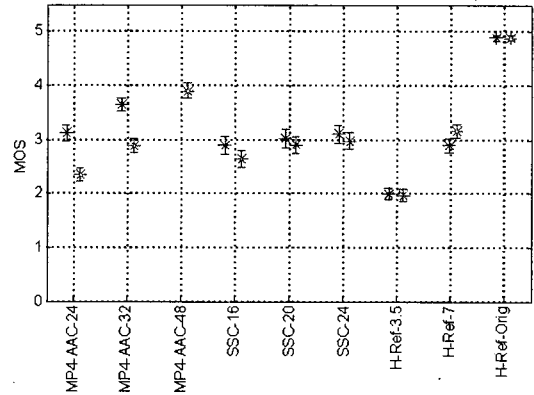
아래 그림 8의 결과는 SBR 기술에 대한 테스트 결과이다. SBR을 채용한 스테레오 24kbps의 경우(HQ-SBR-24, LP-SBR-24)가 SBR을 사용하지 않은 AAC 30kbps의 경우(MP4-AAC-30)보다 더 좋은 점수를 받았음을 볼 수 있다. H-Ref-7은 7.0kHz로 대역 제한된 원 신호(H-Ref-Org)를 말한다.



〈그림 8〉 HE-AAC 청취 실험 결과<sup>[8]</sup>

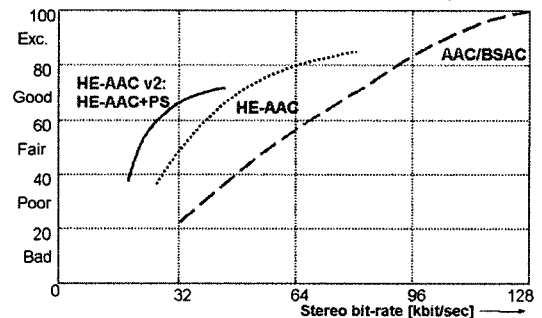
MPEG-4 AAC의 24kbps, 32kbps, 48kbps의 경우들과, PS기술이 사용된 SSC (SinuSoidal Coding)의 16kbps, 20kbps, 24kbps인 경우들을 비교한 실험 결과는 다음 그림 9와 같다. 여기서 H-Ref는 3.5kHz, 7kHz로 대역 제한된 신호들과, 원 신호를 말한다. 왼쪽의 신호는 모노신호에 대한 실험 결과이고, 오른쪽은 스테레오에 대한 실험 결과이다. 실험 결과를 보면, PS기술을 사용한 SSC방식들은 비트율의 증가에 따른 성

능 개선 정도는 크지 않지만, 20kbps에서 MP4-AAC-32kbps급의 음질을 제공하는 것을 볼 수 있다.



〈그림 9〉 PS가 사용된 SSC와 AAC간의 성능 비교 실험 결과<sup>[9]</sup>

결과들에 비추어서, AAC, BSAC, HE-AAC, HE-AAC v2의 성능을 스테레오를 표현해 주는데 사용되는 비트율과 음질의 변화 측면에서 비교해 보면 다음 그림 10과 같다. 제공될 수 있는 성능과 비트율간에 차이가 있음을 볼 수 있다.

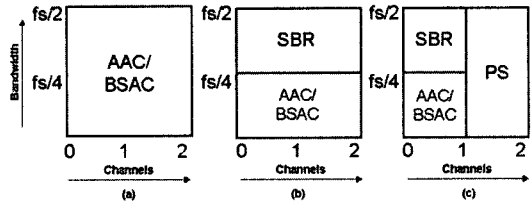


〈그림 10〉 스테레오 비트율과 음질간의 관계

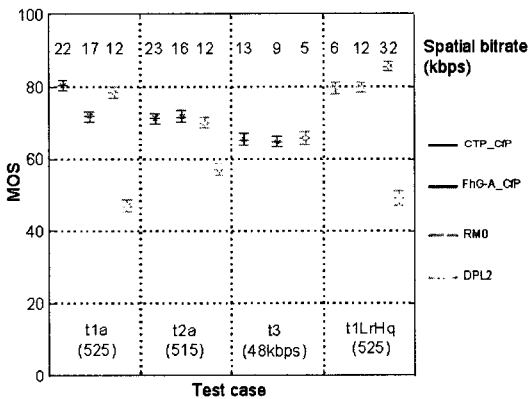
MPEG Surround의 경우, 스테레오 신호를 기반으로 하여 가상 멀티채널 신호를 만드는

돌비 프로로직과 비교를 하였다. 그림 11에서, CFP는 처음 기술 제안되었던 CTP(Philips와 Coding technologies) 기술, FhG-A(Fraunhofer와 Agere System) 기술에 대한 것을 나타내며, DLP2는 돌비 프로로직 II를 말한다. 제일 오른쪽이 DLP2이고, 그 바로 옆이 RM0에 대한 경우인데 5kbps에서 32kbps 정도의 공간정보 표현을 위한 데이터를 스테레오 압축된 결과 또는 모노 압축된 결과에 추가로 사용할 경우, 다채널 오디오 재생 효과를 제공함을 보였다.

상도가 낮은 부분을 찾아 합성 신호로 표현해, 압축 성능을 높이는 일들이 진행되고 있다. 그 결과 다음 그림12와 같이, 각 기술에 의해 처리 되는 부분이 나누어 지게 되었다.



〈그림 12〉 오디오 코덱 채널 수와 샘플링 주파수에 따른 적용 기술



〈그림 11〉 각 방식에 공간정보 표현용 비트율을 다르게 적용한 실험 결과<sup>10)</sup>

### V. 결론

디지털 오디오 압축은 저장 매체에 더 많은 곡을 저장하거나, 전송 채널의 효과적인 사용을 위하여 연구되고 있다. 특히, 대역폭이 제한된 무선 환경에 있어서 압축의 중요성이 강조되고 있다. 오디오 압축은 정보이론에 의한 신호처리적 압축에서, 심리음향 모델링 기술을 이용한 인식 기반 코덱으로 기술들이 개발 되었고, 최근에는 인간의 해

앞으로는, 정보처리 이론에 따른 신호처리 한계를 풀기 위한, 인간의 인지 특성 모델링 측면과 파라메트릭에 의한 오디오 신호의 합성 기술 측면에서 현재의 성능을 개선하고자 하는 노력이 예상된다.

### 참고 문헌

- [1] International Standard ISO/IEC 11172-3:1993
- [2] International Standard ISO/IEC 13818-3:1998
- [3] International Standard ISO/IEC 13818-7:2004
- [4] International Standard ISO/IEC 14496-3:2005
- [5] E. Zwicker and H. Fastl, Psychoacoustics, Springer-Verlag, 1990.
- [6] J. Blauert, Spatial Hearing, The MIT Press, 1983.
- [7] ISO/IEC SC29/WG11 N7373, "WD on Integration of BSAC and SBR", Poznan, Poland, July 2005.
- [8] ISO/IEC SC29/WG11 N6009, "Report on the Verification Tests of MPEG-4 High Efficiency

AAC”, Brisbane, Australia, October 2003.

[9] ISO/IEC SC29/WG11 N6675, “Report on the Verification Tests of MPEG-4 Parametric Coding for High Quality Audio”, Redmond, USA, July 2004.

[10] ISO/IEC SC29/WG11 N7138, “Report on MPEG Spatial Audio Coding RM0 Listening Tests”, Busan, Korea, April 2005.

## 저자소개



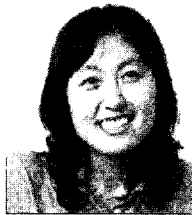
김 상 욱

1989년 2월, 연세대학교 전기공학과 학사.  
1991년 2월, 한국과학기술원 전기 및 전자공학과 석사.  
1991년 - 현재 삼성종합기술원 전문연구원.  
주관심 분야 오디오 압축 및 처리 기술



김 도 형

1992년 2월, 경북대학교 전자공학과 학사  
1994년 2월, 포항공과대학교 전자전기공학과 석사  
1994년 - 현재 삼성종합기술원 전문연구원  
주관심 분야 Digital Audio Compression, MPEG Audio, Codec Optimization, 3D Sound Processing



김 미 영

경북대학교 컴퓨터과학과 학사  
경북대학교 컴퓨터과학과 석사  
2002년 - 현재 삼성종합기술원 연구원  
주관심 분야 오디오 압축, 멀티미디어 시스템 기술