

## I. 서론

오디오 압축 기술들은 크게 국제 표준 기술과 업계에서의 표준 기술로 구분 될 수 있다. 현재, 국제 표준 기술로는 모노/스테레오 처리 기술인 ISO/IEC 11172-3(MPEG-1), HDTV등의 다채널 오디오 처리용 ISO/IEC 13818-3(MPEG-2), ISO/IEC 13818-7(MPEG-2 AAC), 1999년에 표준 제정된 ISO/IEC 14496-3(MPEG-4) 등이 있다. 업계에서의 표준 기술로는 모노 또는 스테레오 처리 기술로는 마이크로소프트사의 WMA, 소니의 ATRAC 기술, 다채널 오디오 처리 기술로는 미국 Dolby Lab.의 AC-3 기술 등이 있다.

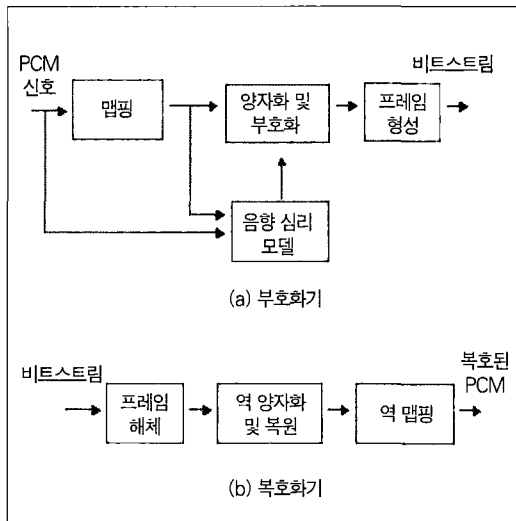
본 논문에서는 한국 지상파 DMB에서 제공하는 동영상 서비스에 있어서, 오디오 신호의 압축에 사용되는 국제 표준 기술인 ISO/IEC 14496-3

BSAC(Bit Sliced Arithmetic Coding) 방식에 대해 간략한 소개를 하고자 한다.

## II. 음향 심리 모델을 이용한 오디오 데이터 압축 방식(1)-(4)

인간의 음향 심리 모델(5)-(7)을 이용한 오디오 데이터 압축 기술들은 인간의 소리에 대한 인식 성질들 가운데 공간적인(spectral), 시간적인(temporal) 마스킹 효과에 대한 모델링을 데이터 압축에 고려해 주는데 다음(그림 1)과 같은 구조를 갖는다.

맵핑 단계에서는 입력신호를 주파수 대역 별로 분리하는 처리를 한다. 주파수 대역별로 분리를 하는 이유는 인간의 음향 심리 특성 중 신호 인식에 사용되는 단위인 바크대역(bark band)으로 신호를



〈그림 1〉 심리음향 모델을 이용한 오디오 부호화기, 복호화기의 예

분리해 주기 위함이다. 처리에 있어서는 Pseudo-QMF (Quadrature Mirror Filters)에 의한 시간영역에서의 신호 분리 또는 MDCT(Modified DCT)에 의한 주파수 영역에서의 신호분리를 많이 사용한다.

음향 심리 모델 단계에서는 대역별로 분리된 입력 신호에 대한 상대적인 중요도를 계산하는 처리를 한다. 중요도 계산에 있어서는 음향 심리 모델링 기법 가운데 마스킹(masking) 현상과 해당 주파수 신호에 대한 사람의 민감도 특성을 주로 고려한다. 마스킹 현상은 어떤 주파수의 신호가 인접한 다른 주파수의 신호를 간섭하는 현상이다. 마스킹 현상을 고려해, 존재하는 신호들 가운데, 사람이 느끼지 못하게 되는 신호를 골라 낼 수 있다. 대역별 신호에 대한 상대적인 사람의 민감도 특성을 고려해, 각 대역별 신호 표현에 사용될 비트 수를 결정하는 처리를 한다.

양자화 및 부호화 단계에서는 주파수 대역별로

분리된 신호를 할당된 비트 수를 이용해 선형 또는 비선형 양자화 처리 후, 엔트로피 방식 등을 적용해 표현해 준다.

프레임 형성부에서는 입력신호를 부호화 처리한 결과를 비트스트림으로 만들어 준다.

복호화시에는 부호화시의 역순에 의해 신호를 처리해 준다.

### Ⅲ. DMB 동영상 오디오 표준 BSAC

ISO(국제표준위원회)와 IEC(국제전기학회)가 설립한 공동기술협의회 JTC1( Joint Technical Committee 1) 산하의 SC(Sub-Committee) 가운데 SC29에는 영상 압축에 관련된 여러 WG(Working Group)들이 있다. 이 가운데, WG 11은 동화상 압축에 관련된 그룹으로 일명 MPEG이라 한다. MPEG은 1988년 5월 캐나다 오타와에서 제1회 회합을 시작한 뒤, 1992년 MPEG-1 오디오(ISO/IEC 11172-3), 1994년에는 MPEG-1과 호환성이 제공되는 MPEG-2 BC 오디오 부분(ISO/IEC 13818-3), 1997년에는 MPEG-1과 호환성이 제공되지 않는 MPEG-2 AAC 오디오(ISO/IEC 13818-7)에 대한 국제 표준을 제정하였다(8)~(11).

지상파 DMB에서 지원하는 오디오 규격인 BSAC(Bit Sliced Arithmetic Coding)은 1999년도에 제정된 MPEG-4 오디오 표준 기술(ISO/IEC 14496-3)로, ISO/IEC 14496-3 규격 중 ObjectID 22인 ER-BSAC 규격에 다음과 같은 제한 사항을 준 기술이다.

**제한 사항:**

AudioSpecificConfig()에서  
 EpConfig:0  
 FrameLengthFlag:0  
 DependsOnCoreCoder: 0  
  
 Bsc\_header()에서  
 SBA\_mode:0  
  
 General\_header()에서  
 Ltp\_data\_present:0

MPEG-4 오디오의 general audio 부호화틀은 음향 심리 모델에 기반한 오디오 부호화 방식으로 비트율이 16kbps/ch에서부터 64kbps/ch까지 지원된다. 비트율 범위내에서 "AM보다 좋은 음질"에서부터 "CD와 동등한 음질"까지 성능을 낼 수 있다.

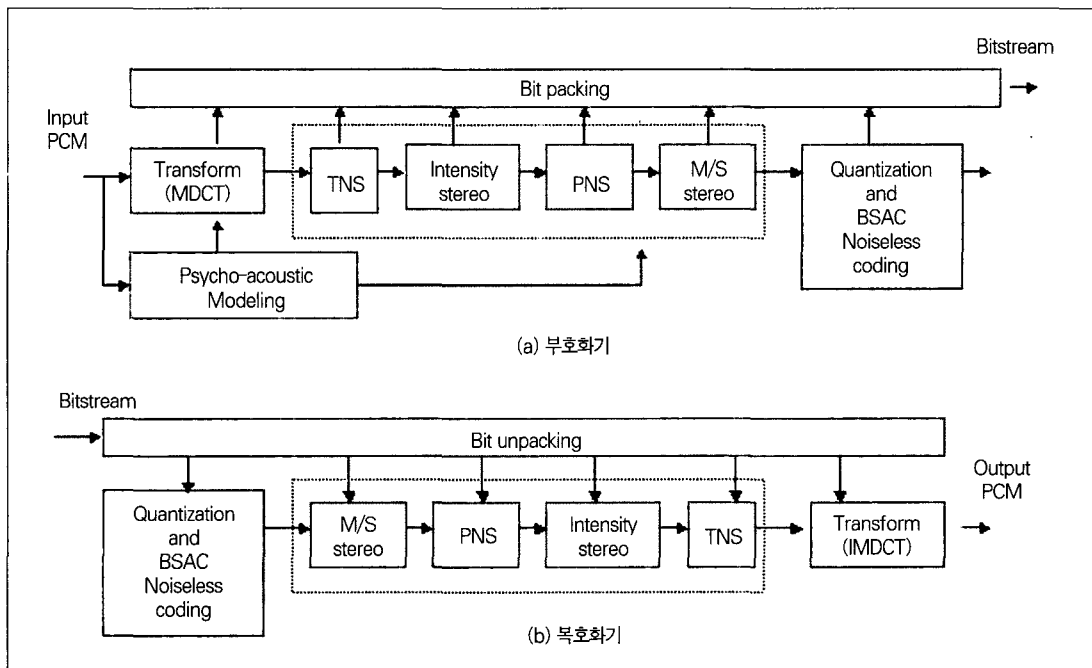
BSAC 기반 압축틀은 AAC 구조와 대부분 같으면서 미세한 계층의 스케일러블한 구조 (FGS,

Fine Grain Scalability)를 제공한다. 기본계층에는 일반적인 프레임 헤더정보와 부가정보가 포함이 되고, 강화계층에는 각 강화계층별 부가정보와 실제 부호화된 데이터가 포함이 된다. 강화계층은 1kbps/ch 또는 더 작은 단위로 계층을 만들 수 있다. 기본 아이디어는 MPEG-4 AAC에서 사용되는 scalefactor와 spectral data의 허프만 부호화 대신 산술 부호화 (BSAC, Bit Sliced Arithmetic Coding)를 사용하는 것이다.

〈그림 2〉에서 보이는, BSAC의 각 구성 요소 기술들은 다음과 같다.

**Transform**

입력 신호의 특성에 따라 long block와 short block 원도를 썬워준 뒤 MDCT(modified Discrete Cosine Transform)처리를 한다. 단위 프레임의 크기는 1024 샘플



〈그림 2〉 BSAC의 블록도

플을 사용해 주는데, 48kHz 샘플링된 신호의 경우 21ms에 해당하는 신호이다. Short 윈도우 처리의 경우 128샘플씩을 하나의 블록으로 해서, 8개 블록으로 나눠 처리를 하고, long의 경우 1024 샘플을 하나의 블록으로 한다. Short윈도우 처리는 프레임 단위 내에서 신호가 급작스럽게 변화하는 경우, 유사한 신호들을 보다 효과적으로 처리하기 위해 사용한다. 즉, pre-echo 잡음을 최소화하기 위해 사용한다.

#### TNS

TNS(Temporal Noise Shaping)은 1.5kHz 이상의 신호에 적용이 된다. 양자화된 잡음의 temporal shape를 조절해 주는 처리를 한다.

#### Intensity

스테레오 신호 처리에 적용된다. 고주파수 신호의 경우, 사람의 스테레오 음상에 대한 인지 특성은 주파수 변환된 신호의 포락선에 의존한다는 사실을 데이터 압축에 고려해 주는 부분이다. 왼쪽 채널에는 주파수 성분, 오른쪽 채널에는 스케일팩터에 대한 에너지 크기만 전달해 오디오 압축 효과를 높여준다.

#### PNS

PNS(Perceptual Noise Substitution)는 디더링 기술을 오디오 신호 처리에 적용해 준 것이다. Scalefactor 대역에 대한 에너지 크기를 전달해 주면, 난수 발생기를 이용해 전달된 크기를 고려한 신호를 발생해 사용한다.

#### M/S stereo

Left/Right 스테레오 신호를 Middle/Side로 변환해 주는 처리를 한다. MS로 변환하는 식은 다음과 같다

$$M=(L+R)/2$$

$$S=(L-R)/2$$

#### Quantization

비선형 양자화를 사용한다.

#### BSAC Noiseless Coding

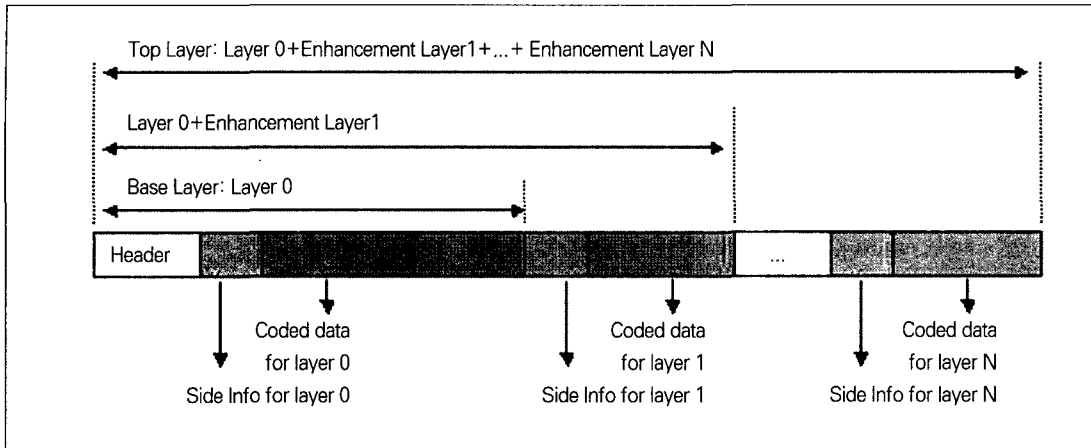
BSAC(Bit Sliced Arithmetic Coding)는 FGS(fine

grain scalability)를 제공하는 무손실 부호화 툴이다. BSAC에서는 스케일러빌리티를 제공하기 위해, bit packing시 상대적으로 중요한 데이터를 비트스트림의 앞부분에 저장하고, 이진수로 변환해 표현해 준 샘플에서 상대적으로 중요한 데이터부터 처리를 하는 비트플레인 연산을 한다. 즉, MSB(Most Significant Bit)부터 처리를 한다.

## IV. BSAC 비트스트림의 계층 구조와 비트플레인 코딩(12)

BSAC는 스케일러빌리티를 제공하기 위해 비트스트림이 계층 구조를 갖는다. 스케일러빌리티는 64kbps 서비스를 목적으로 준비한 비트스트림을 48kbps에서도 서비스가 가능하게 한다. 비트스트림의 계층 X는 계층 X-1을 포함하는 구조이다. 서비스 비트율이 결정 되면, 비트율에 제일 가까운 비트 수를 갖는 계층 Y까지가 서비스에 사용되는 비트스트림이 된다. 양자화 정보 등 신호를 복호화 하는데 필요한 정보들을 계층별로 나누어 포함 시켜 줌으로서, 일부 하위 계층의 데이터만 가지고도 음악 복원이 가능하게 되어있다.

〈그림 3〉은 BSAC의 계층구조의 예를 나타낸 것이다. 위의 계층구조는 프레임 헤더의 정보에 의해 기술 된다. 계층구조의 부가정보는 양자 정보를 위한 스케일팩터밴드 대역 단위의 스케일팩터 정보와 계층구조를 위한 코딩밴드 대역 단위의 산술 부호화 모델 정보 등이 있다. 하나의 코딩밴드는 하나의 계층에 해당하며 이는 32개의 MDCT 주파수성분에 해당한다. 무손실 부호화를 수행하기 위해 양자화된 샘플들은 비트플레인 상에 맵핑 된다. MSB는 LSB보다 상대적으로 중요



〈그림 3〉 BSAC의 계층구조

한 데이터이므로 MSB에서 LSB의 순서로 부호화 된다. 비트플레인 상에 맵핑된 이진수들은 문맥 기반 산술 부호화를 통하여 엔트로피 부호화가 수행 된다.

계층에 해당되는 모든 샘플을 부호화 후 가용 비트가 남으면 이전 계층에서 아직 부호화 안된 샘플들을 남은 가용비트만큼 할당하여 부호화 해 준다.

비트플레인에서의 데이터 처리는 다음과 같다.

〈그림 4〉와 같이 양자화 된 샘플 ( $X(i, i+3) = [2, -4, 3, 1]$ )을 부호화한다고 하면 다음 순서에 의해 수행된다.

Sign	+	-	+	+
MSB	① 0	② 1	③ 0	④ 0
MSB-1	⑤ 1	⑥ 0	⑦ 1	⑧ 0
LSB	⑨ 0	⑩ 0	⑪ 1	⑫ 1
Sample	$X(i)$	$X(i+1)$	$X(i+2)$	$X(i+3)$

〈그림 4〉 비트플레인 상에서 양자화 된 샘플의 예.



1. 코딩밴드 부가정보로부터, 0이 아닌 값이 있는 최상위 비트플레인 MSB(=3)를 구한다.
2. 0이 아닌 값이 있는 최상위 비트플레인에서, 최하위 비트플레인까지 ①②③...⑫의 순으로 이진 값들을 다음과 같은 방식으로 비트플레인 벡터 단위로 부호화 한다.
  - A. MSB에 해당하는 이진수들은 상위 비트 플레인의 이진 값들이 0이므로 ①을 부호화하기 위해서는 참조할 상위 문맥이 없다. 참조할 상위 문맥이 없는 경우, 미리 정해놓은 모델에 따라 부호화 한다.
  - B. ⑦을 부호화하기 위해서 사용되는 문맥은 상위 이진 값들, ①②③④와 자신과 같은 비트플레인에 있는 하위 이진값 ⑤와 ⑥이며 이 값들을 참고하여 산술부호화 모델을 선택해 부호화한다.
3. 계층에 할당된 사용 가능 비트 수만큼 부호화하며, 비트가 모자랄 경우 더 이상의 부호화는 하지 않는다.

그 결과, 어떤 한 샘플 값을 표현하는 데 있어서, 제일 중요한 신호인 MSB부터 처리가 가능하고, 중요도에 따라 서비스 품질을 조절해 주는 처리가 가능하게 한다.

## V. 결론

컴퓨터 기술 및 디지털 통신 기술, 디지털 신호 처리기들의 발달에 힘입은 디지털 오디오 데이터

압축/복원 방식들은 같은 용량의 기록 저장 매체에 보다 많은 정보들을 저장할 수 있고, 데이터의 전송에 드는 시간을 줄여주는 기능을 갖는다. 본 논문에서는 이러한 디지털 오디오 데이터 압축/복원 방식의 일반적인 구조와 사용 기술들을 다루어 주었고, 한국 지상파 DMB에서 제공하는 동영상 서비스에 있어서, 오디오 신호의 압축에 사용되는 국제 표준 기술인 ISO/IEC 14496-3 BSAC(Bit Sliced Arithmetic Coding) 방식에 대한 간략한 소개를 하였다.

### 참고 문헌

- [1] A.Gersho, "Advances in Speech and Audio Compression," Proc. of IEEE, vol.82, no.6, pp.900-918, June 1994.
- [2] P.Noll, "Wideband Speech and Audio Coding," IEEE Commun. Mag., vol.31, no.11, pp.34-44, Nov. 1993.
- [3] N.S.Jayant, J.Johnston, and R.Sofranek, "Signal Compression based on models of human perception," Proc. IEEE, vol.81, no.10, pp.1385-1422, Oct. 1993.
- [4] K.Brandenburg, and G.Stoll, "The ISO/MPEG-Audio Codec: A Generic Standard for Coding of High Quality Digital Audio," Presented at 92nd Audio Eng. Soc. Conv., Preprint 3336, Mar. 1992.
- [5] T.D.Rossing, The Science of Sound (2nd ed.), Addison-Wesley, Reading: Massachusetts, 1990.
- [6] J.Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization, The MIT Press: Cambridge, Massachusetts, 1983.
- [7] J.Tobias (ed.), Foundations of Modern Auditory Theory, New York: Academic, 1970.
- [8] ISO/IEC 13818-7, "MPEG-2 AAC: Advanced Audio Coding".
- [9] ISO/IEC 13818-3, "Information Technology-Generic Coding of Moving Pictures and Associated Audio: Audio".
- [10] ISO/IEC 11172-3, "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 MBit/s: Audio".
- [11] ISO/IEC 14496-3, "Information Technology-Very Low Bitrate Audio-Visual Coding: Audio".
- [12] <http://www.mpeg4bsac.com>

필자소개



김 상 욱

- 1989년 2월 : 연세대학교 전기공학과 학사
- 1991년 2월 : 한국과학기술원 전기및전자공학과 석사
- 1991년~현재 : 삼성종합기술원 멀티미디어Lab, 전문연구원
- 주관심분야 : 오디오 신호 처리, 삼차원 오디오 신호 처리, 비선형 필터