

# 채널 간 중복성을 이용한 다채널 오디오 부호화

\*신종민\*, 김기수\*, 한영태\*\*, 유대의\*, 차임관\*  
\*연세대학교, \*\*한국통신

## Multichannel Audio Coding using Inter-channel Redundancies

\*Jong-In Shin\*, Ki Soo Kim\*, Young Tae Han\*\*, Dae Hee Youn\*, Il Whan Cha\*  
\*Yonsei University, \*\*Korea Telecom

### 요 약

기존의 2-채널 오디오 복호화기와 호환성을 유지하기 위해서는 다채널 오디오 부호화기에 할당되는 비트율이 제한되어야 한다. 늘어난 채널 신호에 적은 양의 비트를 할당해줌으로 인해서 복원음의 음질이 저하되는 결과를 초래한다. 따라서 기존의 방법에 신호의 채널 간 중복성(inter-channel redundancies)을 이용한 다채널 부호화법을 적용시켜 낮은 비트율에서 다채널 오디오 신호의 음질을 개선시킬 필요가 있다.

본 논문에서는 채널 간 중복성을 이용하기 위한 방법으로 적응 채널 간 예측법(adaptive inter-channel prediction)을 제안하였다. 적응 채널 간 예측법은 주로 저주파 대역에서 스테레오 신호 간에 존재하는 중복된 정보를 제거하고자 하는 방법이다. 저주파 대역에서는 큰 예측 이득을 얻을 수 있어서 스테레오 음상에 불필요한 정보를 없애는데 유리하다.

기존의 방법과 제안된 예측 부호화법을 비교해 보면 두 신호의 상관 관계가 클수록 개선폭이 증가함을 알 수 있으며 낮은 비트율에서 기존의 부호화기가 처리하기 어려운 광대역 신호에 대해서 음질이 개선되는 결과를 얻었다.

### 1 장. 서 론

지금까지 오디오 신호를 압축하기 위한 방법들이 다방면으로 연구되어져 왔다. 최근에 가장 주목할 만한 방

법으로는 청각 특성을 이용한 부호화 방법[1][2]이다. 사람의 귀는 시간 영역의 오디오 신호를 주파수 영역으로 변환하여 인식하며, 음의 인식은 주파수 대역에 따라 가청 한계(hearing threshold)와 선택성(selectivity)이 다르게 나타난다[3]. 또한, 특정 주파수 대역에 큰 에너지를 가지는 신호가 존재할 때 주변 대역에 존재하는 작은 신호들은 가려져 들리지 않는 매스킹 현상(masking effect)이 존재한다[4][5]. 이런 청각 특성을 이용하여 양자화 잡음이 가청 한계 또는 원음에 의해 가려지게하면 비트율을 낮추더라도 지각적으로 원음과 동일한 복원음을 얻을 수 있다.

1991년에 제정된 국제 표준화 기구(International Standard Organization, ISO) 산하의 동영상 전문가 그룹(Moving Picture Experts Group, MPEG)에서는 서브밴드 부호화 방식 중 하나인 MUSICAM(Masking-pattern adapted Universal Subband Integrated Coding And Multiplexing) 방식[1][6]을 기본으로 하여 동영상과 CD 수준의 디지털 오디오를 15 Mbit/s로 압축할 수 있는 MPEG-1 표준안을 확정하였다. 현재 MPEG-1 표준안은 Video-CD, VOD(Video On Demand), DAB, 멀티미디어 등에 응용되고 있다.

그러나 MPEG-1 표준안은 디지털 방식의 HDTV와 같은 방송 매체에 적용될 경우 영상의 질이 떨어지며 오디오에 있어서도 다채널, 음성 다중 등의 부가 서비스를 제공하기에 부적합하다. 그래서 새로운 표준안인 MPEG-2에 대한 연구가 진행되고 있는데 방송 매체를 위한 MPEG-2 오디오 부호화 방식은 MPEG-1의 계층-2,-3 압축 방법과 다채널 부호화 방법이 결합된 오디오 압축 방식으로 약 384~1066 kbit/s의 전송율에서 뛰어난 음질을 제공한다[7].

본 논문에서는 MPEG-2에서 권고한 다채널 부호화 방법 중 채널 간 중복성을 제거하기 위한 적응 채널 간

본 연구는 한국 통신(Korea Telecom)의 연구비 지원으로 이루어졌습니다.

예측법을 연구하였다. 적응 채널 간 예측법은 저주파 영역의 중복된 정보를 제거하기 위한 것으로서 여기에 할당된 비트를 양자화 오차가 큰 다른 부분으로 옮겨 줌으로써 동일한 비트율에서 향상된 음질을 얻고자 하는 방법이다. 예측기를 사용한 부호화기와 기존의 부호화기를 성능 비교해 보았다. 그리고 예측 결정의 기준이 되는 크기 인자(scalefactor)를 여러 가지 값으로 바꾸어가면서 최적의 예측 결정 기준을 찾아보았다.

## 2 장. MPEG 오디오 압축 알고리즘

MPEG의 오디오 압축 알고리즘은 오디오 신호의 두 가지 중복성을 제거하는 방법으로 설명될 수 있다. 먼저, 통계적인 중복성을 제거하기 위해 신호를 32개의 필터뱅크를 통과시켜 각 대역의 크기에 따라 비트 할당을 달리 하는 서브밴드 부호화 방법을 사용한다. 이 신호를 더욱 압축하기 위해 마스킹 현상과 임계 대역, 절대 가청 한계 등의 청각 특성을 이용하여 지각적인 중복성을 제거한다. 위와 같은 두가지 방법을 사용하여 입력된 오디오 신호는 지각적인 음질을 떨어뜨리지 않고 큰 압축율을 가져온다. 그림 1은 MPEG 오디오 부호화기의 블록 다이어그램을 나타낸다.

### 2-1. 서브밴드 부호화

서브밴드 부호화는 신호를 대역 통과 필터를 통과시켜 여러 주파수 대역으로 나누어 부호화하는 방식으로 각 대역에 할당되는 비트 수를 신호의 크기에 따라 달리 하여 신호대 잡음비를 향상시킬 수 있다. 입력단에서 N 개의 대역 통과 필터로 구성된 분석 필터뱅크(analysis filterbank)에 채널 신호를 통과시킨다. 이 때 부호화하는 샘플의 수가 입력된 샘플의 수와 동일하도록 간축(decimation)하여 표본화율을 낮추는 과정을 갖는다. 각 대역 통과 신호는 Adaptive PCM(APCM)을 사용하여 부호화하고 부가 정보와 결합시켜 전송한다[8][9].

서브밴드 부호화에서 가장 중요한 필터 뱅크는 QMF 필터 뱅크를 비롯하여 여러 방식이 개발되었는데 최근에는 임계 대역과 유사한 모양을 갖도록 비선형적 크기를 갖는 필터뱅크를 설계하기도 한다. MPEG에서는 32개의 동일 크기를 갖는 폴리페이즈 구조(polyphase structure)의 Single Side Band(SSB) 필터 뱅크를 사용하였다[6]. 서브밴드 분석에 사용되는 필터는 512-tap 저역 통과 필터가 기본이 되며 행렬에 의해서 주파수 천이되

어 32개의 동일 크기 서브밴드가 된다.

### 2-2. 심리음향 모델

심리음향 모델은 원음이 존재할 때 양자화 잡음이 마스킹되어 들리지 않게 되는 최대값을 각 주파수 대역에서 얻는 방법이다. 따라서 서브밴드 부호화나 변환 부호화에서 심리음향 모델을 사용하면 각 밴드에서 원음에 의해 마스킹되어 들을 수 없는 최대의 잡음 레벨을 결정할 수 있다. 이 잡음 레벨, 즉 마스킹 임계값을 사용해서 각 밴드의 양자화기에 결정하는 비트 할당을 할 수 있다. 양자화 잡음의 크기는 각 밴드에 할당되는 비트수에 의해 결정되므로 비트 할당을 적절하게 하면 주관적인 왜곡을 최소화할 수 있다.

부호화기는 큰 압축율을 필요로 하고 시스템이 복잡해도 무관한 경우에는 정확한 심리음향 모델을 사용하고 낮은 압축율에서는 마스킹 곡선을 개략적으로 모델링하는 방법을 사용한다.

MPEG에서는 두 가지의 심리음향 모델을 제공하는데 두 모델을 응용 분야에 따라 적절하게 선택해서 사용할 수 있다. 첫번째 방법은 FFT 스펙트럼을 순음과 잡음 성분으로 나누어 각 성분에 의한 마스킹 임계값을 구한 후, 절대 가청 한계를 고려하여 마스킹 임계값을 구하는 비교적 간단한 방식으로 계층 1, 2에 사용한다. 두번째 방법은 FFT 스펙트럼을 청신경의 여기 모델인 스트레딩 함수와 콘볼루션하여 마스킹 임계값을 구하는 방식으로 계산량이 많지만 더 정교한 결과를 얻을 수 있으며 계층 3에 사용된다[6].

## 3 장. 직음 채널 간 예측 부호화

ISO MPEG-오디오 그룹에서 마련한 고품질 오디오 부호화 알고리즘은 128~384 kbit/s에서 스테레오 신호를 부호화할 수 있다[6] 채널 당 128~192 kbit/s 정도의 높은 비트율로 스테레오 신호를 부호화하였을 때는 주관적으로 CD 수준과 동일한 음질을 복원해 낼 수 있다. 이렇게 높은 비트율에서는 가청 잡음을 없애는데 충분한 비트를 이용할 수 있기 때문에 두 채널을 각기 따로 부호화하더라도 음질에 영향을 주지 않는다.

그러나 더 낮은 비트율에서는 두 채널을 개별적으로 부호화하는데 한계가 있기 때문에 스테레오 신호 간에 존재하는 중복성이나 스테레오와 무관한 성분을 이용한 결합 부호화(combined coding) 방법을 사용해야 한다.

3-1. 적응 채널 간 예측

스테레오 기법을 이용해서 생성된 오디오 데이터의 경우 개별적으로 생성된 모노 오디오의 경우보다 더 많은 상관 관계를 가진다. 채널 간 예측법은 두 채널 혹은 다채널 스테레오 오디오 신호에서 채널 간에 존재하는 중복성을 제거하기 위해서 사용된다. 본 논문에서 사용한 채널 간 예측은 선형 예측과 같이 과거 신호를 이용해서 현재 신호를 예측한다는 점에서는 비슷하지만 다른 채널 신호로부터 예측한다는 점과 예측에 현재값을 이용한다는 점에서 차이가 있다.

채널 간 예측은 다음과 같이 이루어진다.

$$\hat{y}(n) = \sum_{k=1}^K a_k x(n-k) \quad (1)$$

즉, 어떤 채널 신호  $x(n)$ 의 과거값과 현재값으로부터 다른 채널 신호  $y(n)$ 의 근사치를 찾는 과정이다. 여기에서  $K$ 는 예측기의 차수를 나타내고,  $a_k$ 는 예측 계수를 표현한 것이다. 예측 오차  $e(n)$ 은 다음과 같다.

$$e(n) = y(n) - \hat{y}(n) \\ = y(n) - \sum_{k=1}^K a_k x(n-k) \quad (2)$$

선형 예측에서와 같이 예측 계수  $\{a_k\}$ 를 찾는 기본적인 접근법은 유한 구간  $0 \leq n \leq N-1$ 에서 평균 예측 오차를 최소화시키는 방법이다. 단구간 평균 예측 오차는 아래 식과 같이 정의된다.

$$E = \sum_{n=0}^{N-1} e^2(n) \\ = \sum_{n=0}^{N-1} \{y(n) - \hat{y}(n)\}^2 \\ = \sum_{n=0}^{N-1} \left\{ y(n) - \sum_{k=1}^K a_k x(n-k) \right\}^2 \quad (3)$$

$E$ 를 최소화하는  $a_k$ 를 찾기 위해서

$$\frac{\partial E}{\partial a_k} = 0 \quad k=0, 1, 2, \dots, K \quad (4)$$

라고 놓고 식을 전개하면 아래와 같은 선형 방정식을 이끌어낼 수 있다.

$$\sum_{n=0}^{N-1} a_k R_{xx}(k, i) = R_{xy}(k, i) \quad i=0, 1, \dots, K \quad (5)$$

여기에서,

$$R_{xy}(k, i) = \sum_{n=0}^{N-1} p(n-k) q(n-i)$$

이다. 식 (5)는 아래와 같이 행렬식으로 나타낼 수 있다.

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \dots & R_{xx}(K-1) \\ R_{xx}(1) & R_{xx}(0) & \dots & R_{xx}(K-2) \\ R_{xx}(2) & R_{xx}(1) & \dots & R_{xx}(K-3) \\ \dots & \dots & \dots & \dots \\ R_{xx}(K-1) & R_{xx}(K-2) & \dots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_K \end{bmatrix} = \begin{bmatrix} R_{xy}(1) \\ R_{xy}(2) \\ R_{xy}(3) \\ \dots \\ R_{xy}(K) \end{bmatrix} \quad (6)$$

자기 상관값으로 이루어진 행렬은  $K \times K$  정방 대칭 (square symmetric)인 Toeplitz 행렬을 이룬다. Cholesky 분해법을 이용하여 행렬식을 풀면 예측 계수  $a_k$ 를 구할 수 있다.

3-2. 채널 간 예측을 이용한 스테레오 부호화

본 논문에서는 앞에서 제시한 채널 간 예측법을 이용하여 스테레오 신호를 부호화하고 MPEG 부호화기와 그 결과를 비교해보았다. 부호화 알고리즘들은 MPEG-1에서 제시한 부호화 방법이 주를 이루고 있으며 여기에 적응 채널 간 예측법을 도입하였다.

MPEG-2에서는 MPEG-1 복호화기와와의 호환성을 유지하기 위해서 5채널 오디오 신호를 384 kbit/s 이내로 부호화하기를 권장한다. 이 경우 각 채널당 할당되는 비트는 비트 할당 방법에 따라서 여러 가지 형태를 지닐 수 있지만 일반적으로 MPEG-1 호환 채널이 센터 채널이나 서라운드 채널보다 중요하기 때문에 이 채널에 더 많은 비트가 할당된다. 본 논문에서는 384 kbit/s 중 MPEG-1 호환 채널에 각각 96 kbit/s, 다른 부가 채널에는 각각 64 kbit/s가 할당된다고 가정하였다. 스테레오 신호 중 왼쪽 신호(L)를 호환 채널 중 하나로, 오른쪽 신호(R)를 확장 채널 중 하나로 가정하여 실험하였다.

예측기를 부호화기에 첨가시킬 때는 양자화에 의해서 생기는 효과를 고려하여야 한다. 그림 2는 양자화를 포함한 부호화/복호화의 블럭도이다. 그림에서 나타난 것처럼 신호  $x(n)$ 은 예측기에 입력되기 전에 반드시 양자화 과정을 거쳐야 한다. 이것은 복원 신호  $y_q(n)$ 의 양자화 오차가 예측 오차 신호  $e(n)$ 의 양자화 오차와 동일하도록 하기 위해서이다. 이렇게 하지 않으면 복원 신호  $y_q(n)$ 은  $x(n)$ 의 양자화 오차까지 포함하게 된다.

본 논문에서 행해지는 예측은 모두 서브밴드 신호에 대해서 행해진다. 각 서브밴드에서의 예측 이득이 크면 예측 오차 신호를 부호화하고 그렇지 않으면 원신호(R)를 부호화하였다. 원신호와 오차 신호 중 어느 쪽을 부호화할 것인가는 다음과 같은 기준에 따라서 행해졌다.

$$scf_{ch,i} \geq 2^n scf_{ch,i} \quad \begin{matrix} i=0, 1, 2 \\ n=1, 2, 3, \dots \end{matrix} \quad (7)$$

여기에서  $scf_{ch,i}$ 는 sb 서브밴드,  $i$  번째 프레임의 크기 인자이다.  $scf$ 는 원신호의 크기 인자이고,  $scf'$ 는 오차 신호의 크기 인자를 가리킨다.

$scf$ 의 크기는 양자화 오차에 직접적인 영향을 미친다.  $scf$ 의 크기가  $1/2^n$ 로 줄어들었다면 양자화 분해능(resolution)을  $1/2^n$ 로 줄여도 동일한 크기의 잡음을 얻을 수 있다. 여기에서 비트수를  $n$ 만큼 줄일 수 있으며, 이

비트들을 모아 MNR이 가장 작은 서브밴드에 재할당해 줌으로써 음질을 개선시킬 수 있다.

다음은 오차 신호의 전송 여부를 결정하는 부가적인 기준을 설정하는 기준이다. scf가 scf'보다  $2^n$  이상 클 때는 n비트를 절약할 수 있다. 하지만, 아래와 같이 scf<sub>ch, i</sub>가 scf<sub>cb, i</sub>'의 2배 이상이 되지 못할 때는

$$scf_{cb, i}' < scf_{cb, i} < 2scf_{cb, i}' \quad (8)$$

비트수를 줄일 수 없어도 원신호 대신 오차 신호를 양자화함으로써 분해능을 높일 수 있다. 동일한 비트가 할당될 지라도 scf가 줄어들었기 때문에 상대적으로 분해능이 증가한 효과를 나타내기 때문이다. 따라서, 원신호를 부호화할 때보다 양자화 오차가 줄어들어 음질 개선에 영향을 준다.

아래 식에서

$$scf < a \cdot scf_{cb, i} \quad (9)$$

a를 1.1~1.9까지 변화시키며 부호화한 신호를 복원하여 음질을 비교하였다. 그림 3은 본 논문에서 사용한 부호화 알고리즘이다.

표 1은 예측기를 사용한 부호화기와 기존의 부호화기를 성능 비교한 결과이다. 실험에 의하면 예측 차수가 2차일 때와 a가 1.1일 때 성능이 가장 좋아서 이것을 예측기의 환경으로 설정하였다. 데이터는 스테레오 신호의 상관 관계가 큰 것부터 배열하였으며 성능의 개선도 대체로 이 순서를 따른다는 것을 알 수 있다.

#### 4 상. 결과

본 논문에서는 고음질 오디오 부호화에 바탕을 둔 다채널 오디오 부호화 방법에 적용 채널 간 예측법을 도입하였다. 그리고, 채널 간 예측 부호화에서 오차 신호의 전송 여부를 결정하는 기준을 새로이 마련하였다.

적용 채널 간 예측은 채널 간 중복성을 이용한 부호화 방식으로서 결합 스테레오 부호화 방식을 개선시킨 것이다. 본 논문에서는 MPEG 계층-2 부호화기에 적용 채널 간 예측법을 적용함으로써 낮은 비트율(64 kbit/s)에서의 부호화기 성능을 향상시켰다. 예측기를 사용한 부호화기와 계층-2 부호화기를 비교해 보았는데, 모든 실험 데이터에 대해서 기존의 부호화기보다 향상된 결과를 보여주었다. 부가적인 예측 기준으로서 오차 신호의 scf가 원신호의 scf보다 작을 때는 오차 신호를 부호화하는 것이 음질 향상에 도움이 된다는 것도 살펴보았다.

앞으로 고음질 오디오 부호화에 대한 연구 방향으로서는 더 낮은 비트율에서도 현장감, 사실감 등을 재생시

킬 수 있는 부호화 방법에 대한 것이다. 이를 위해서는 더 효율적인 복합 부호화 방식이 제안되어야 하며, 복합 부호화 방식을 결합시키는 방법에 대해서도 연구가 진행되어야 한다. 이와 더불어 심리 음향 모델의 개선이나 향상된 필터뱅크에 대한 연구도 병행되어야 한다.

#### 참 고 문 헌

- [1] Y. F. Dehery, et al. "A MUSICAM source codec for digital audio broadcasting and storage," Proc. ICASSP, pp. 3605-3608, 1991
- [2] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," IEEE J. Selected Areas Comm., pp. 314-323, 1988
- [3] E. Zwicker, Psychoacoustics, Springer-Verlag, New York, 1982
- [4] J. D. Johnston, "Estimation of perceptual entropy using noise masking criteria," Proc. ICASSP, pp. 2524-2527, 1988
- [5] R. N. Veldhuis, "Bit rates in audio source coding," IEEE J. Selected Areas Comm., pp. 86-96, 1992
- [6] ISO/IEC JTC1/SC29/WG11 MPEG-Audio, 1991
- [7] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard 13818-3 "Generic Coding of Moving Pictures and Associated Audio"
- [8] R. E. Crochiere, S. A. Webber, J. L. Flanagan, "Digital Coding of Speech in Subbands," Bell Syst. Tech. J., vol. 55, pp.1069-1085, Oct. 1976
- [9] N. S. Jayant and P. Noll, Digital Coding of Waveforms : Principles and Applications to Speech and Video. Englewood Cliffs, NJ:Prentice Hall, 1984

채널 간 중복성을 이용한 다채널 오디오 부호화

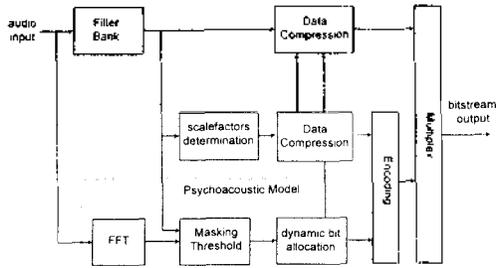


그림 1. MUSICAM 시스템의 구조

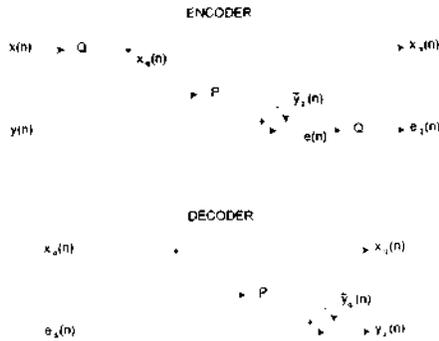


그림 2. 양자화를 포함한 예측 부호화/복호화기의 블럭도

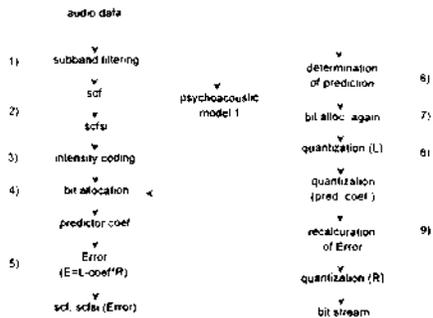


그림 3. 채널 간 예측 부호화 흐름도

표 1. 부호화가 성능 비교(NMR)

Method	계송-2(dB)	계 앞(dB)
여자 독창	9.53	8.10
남자 독창	8.50	6.22
트럼펫	8.60	8.03
바이올린	12.48	10.89
피아노	3.25	2.60