

확산계수를 이용한 파라메트릭 멀티채널 코딩 방식에 관한 연구

이철우, 문한길, 이남숙
삼성전자 디지털미디어 연구소

Study on parametric multi-channel coding method using diffuseness coefficient

Chul-woo Lee, Han-gil Moon, Nam-suk Lee
Digital Media R&D Center, Samsung Electronics

Abstract - 본 논문은 멀티채널 오디오 신호를 두 채널의 다운믹스 신호와 서브밴드별로 추출된 부가 정보로 표현하는 파라메트릭 멀티채널 압축 방식에 관한 연구이다. 기존의 파라메트릭 멀티채널 방식에서 사용하는 부가 정보로는 ICLD, ICC, OPD, IPD 등이 있다. 본 논문에서는 부가 정보로서 phase관련 파라미터를 사용하지 않고, 채널간의 level difference에 상응하는 파라미터와 새로 제안한 diffuseness 계수를 사용하여 우수한 압축 성능을 보이는 파라메트릭 멀티채널 코딩 방식을 제안하고, 그 성능을 기존 방법과 비교하고자 한다.

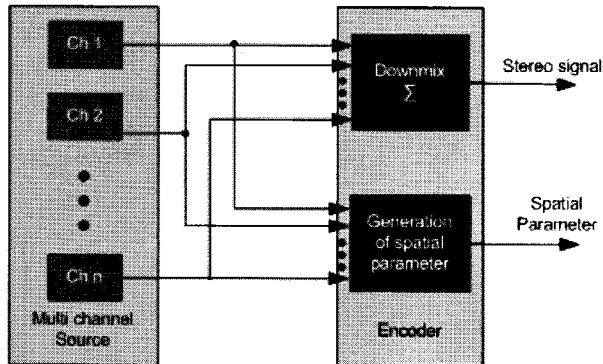
1. 서 론

본 논문에서 소개하는 파라메트릭 멀티채널 코딩 방식은 n채널(n>2)의 멀티채널 오디오 신호를 보다 효율적으로 코딩하기 위해 두 채널의 오디오 신호와, n채널로 복원하는데 필요한 공간 파라미터로 표현하는 방식이다. 파라메트릭 멀티채널 코딩 방식에서 흔히 쓰이는 공간 파라미터로는 ICLD(Inter Channel Level Difference), ICC(Inter channel correlation), OPD(Overall Phase Difference), IPD(Interchannel Phase Difference) 등이 있다. 본 논문에서는 ICLD와 같은 채널간의 level difference를 표현하는 angle파라미터와 diffuseness계수를 사용하여 우수한 압축 성능을 보이는 새로운 방식을 제안하고, 기존의 방법과 성능을 비교하고자 한다. 성능 평가를 위해서는 현재 MPEG 오디오 그룹에서 멀티채널 오디오 코덱의 테스트에 사용되고 있는 11개의 테스트 음원을 사용하였다.

2. 본 론

2.1 파라메트릭 멀티채널 오디오 코딩

파라메트릭 멀티채널 오디오 코딩 방식은 아래 <그림 1>에서와 같이 n개(n>2)의 채널에서 두 채널(또는 한 채널)로 다운믹싱을 수행하고, 이 두 채널을 이용하여 원래의 n채널로 복원하는데 필요한 공간 파라미터를 추출한다. 일반적으로 사용되는 공간파라미터로는 심리 음향적 근거에 의해 ICLD(Inter Channel Level Difference), ICC(Inter channel correlation), OPD(Overall Phase Difference), IPD(Interchannel Phase Difference)등을 사용한다. 이 가운데 ICLD는 채널간에 존재하는 오디오 신호의 크기 차이에 대한 정보를 의미하고, OPD는 두 개의 스테레오 신호를 하나의 모노 신호로 다운믹싱 했을 때, 다운믹스된 신호와 왼쪽 채널 신호 사이의 위상 차이를 의미한다. IPD는 두 채널 신호 사이의 위상 차이를 의미하고, ICC는 두 채널 사이의 유사도를 의미하는 파라미터이다.



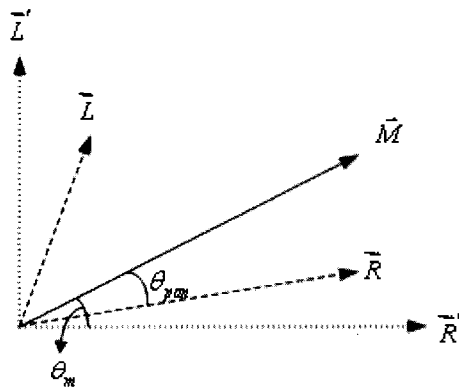
<그림 1> general parametric multi-channel audio encoder

2.2 VSLI

2.1에서 설명한 공간정보중 ICLD는 채널간의 크기 또는 파워의 차이를 비율로 나타낸 파라미터로서 그 정의는 아래 식 (1)과 같다.

$$\Delta \bar{L} = \lim_{l \rightarrow \infty} 10 \log_{10} \left(\frac{\sum_{k=-l}^l x_2^2(k)}{\sum_{k=-l}^l x_1^2(k)} \right) \quad (1)$$

여기서 k는 시간 index이고 $X_1(k)$ 와 $X_2(k)$ 는 각 채널의 특정 서브밴드 신호이다. 그러나 ICLD를 사용할 경우 dynamic range가 넓어서 양자화에 사용되는 비트가 많이 소요된다. In-phasing이 이루어진 두 신호는 청각 사건 축 상에 수없이 많은 가상의 음상을 만들어낸다. 이 많은 음상들은 서브밴드 단위로 분리가 가능하고, 특정 서브밴드에 속한 음상은 청각 사건 축 상의 특정 위치에 분포하게 된다. 특정 음상의 정확한 위치는 CPP(Constant power panning) 방식 혹은 CGP(Constant gain panning) 방식을 활용하여 벡터 형식으로 표현할 수 있다.[2][3] 즉, 특정한 음상의 위치를 극좌표 형태로 표현이 가능하다. 이러한 사실을 이용하여 특정 음상의 가상 위치정보(VSLI:Virtual Source Location Information)를 angle값으로 표현한다[4]. 이 파라미터는 ICLD를 대체할 뿐만 아니라, angle의 값은 dynamic range가 한정되어 있으므로 양자화를 수행하는데 있어서도 용이하다. angle값 추출 과정은 아래 <그림 2>와 같다.



<그림 2> angle 정보를 구하는 방법

$$\sin^2 \theta + \cos^2 \theta = 1$$

$$\theta_m = \frac{\theta_{pan} - \theta_{\bar{R}}}{\theta_{\bar{L}} - \theta_{\bar{R}}} \times 90 \quad \theta_{\bar{L}} \geq \theta_{pan} \geq \theta_{\bar{R}}$$

$$g_{\bar{R}} = \cos \theta_m \quad g_{\bar{L}} = \sin \theta_m$$

2.3 Diffuseness Coefficient

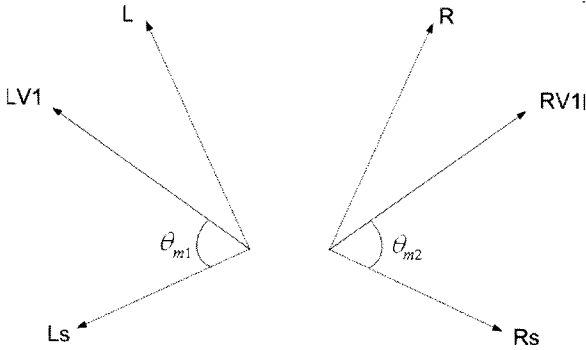
<그림 2>에서와 같은 방법으로 다운믹스 채널 \vec{M} 을 구하고, 공간정보로서 θ_m 을 전송하면, 디코더에서는 이 정보를 이용하여 다운믹스 이전의 채널 \vec{L} 과 \vec{R} 을 복원할 수 있다. 그러나 다운믹스 과정에서 inphasing으로 인해 채널간의 liveness가 소실되고, 코딩 효율을 높이기 위해 심리 음향 이론을 바탕으로 서브밴드를 구분하고 각각의 서브밴드마다 하나의 θ_m 만을 사용하기 때문에 원래의 채널에서 느끼는 음상의 확장감이 떨어지게 된다.

만일 phase관련 공간정보를 사용하면 이러한 현상이 감소하기는하나 코딩 효율이 떨어지므로, 본 논문에서는 liveness를 보강하기 위하여 diffuseness 계수를 사용하였다.

아래 <그림 3>에서와 같은 멀티채널 구조에서 L채널을 구하기 위한 방법으로, 기존의 angle만 사용할 경우에는 다음 식과 같은 방법으로 구한다.

$$L = \overrightarrow{LV1} \sin \theta_{m1}$$

그러나, 채널간의 유사도가 떨어지는 음원일 경우 이렇게 구한 L채널은 위에서 설명했듯이 신호의 intensity 정보만이 반영되었으므로 채널간의 liveness가 사라져서 채널간의 분리도가 떨어지고 음상이 좁아지는 현상이 발생한다.



<그림 3> 멀티 채널 구성

이러한 단점을 극복하기 위하여 diffuseness 계수를 적용하여 아래 식과 같은 방법으로 L채널을 구할 수 있다.

$$L_s = \overrightarrow{LV1} \cos \theta_{m1}$$

$$L_a = \overrightarrow{LV1} \sin \theta_{m1}$$

$$L_b = LV1 - L_s$$

$$L = D_L \times L_a + (1 - D_L) \times L_b$$

여기서 L_a 는 다운믹스 채널과 gain정보만으로 구한 pseudo L채널 신호이고, L_b 는 다운믹스 채널에서 L_s 를 빼는 방법으로 구한 pseudo L채널 신호이다. 경우에 따라 L_b 가 L_a 에 비해 신호의 명료도가 떨어지는 경우는 있으나, 다운믹스 채널에서 L_s 신호를 잘 분리해 내는 경우도 있어서, diffuseness 계수(D_L)를 사용하여 두 신호를 적절히 혼합할 경우, 디코더에 de-correlation 블록을 사용하지 않고서도 음질 향상 효과를 얻을 수 있다. L_s 채널 신호도 이와 같은 방법으로 구할 수 있으며, R채널 신호를 구하는 방법도 이와 상응하는 방법으로 아래 식을 이용하여 구할 수 있다.

$$R_s = \overrightarrow{RV1} \cos \theta_{m2}$$

$$R_a = \overrightarrow{RV1} \sin \theta_{m2}$$

$$R_b = RV1 - R_s$$

$$R = D_R \times R_a + (1 - D_R) \times R_b$$

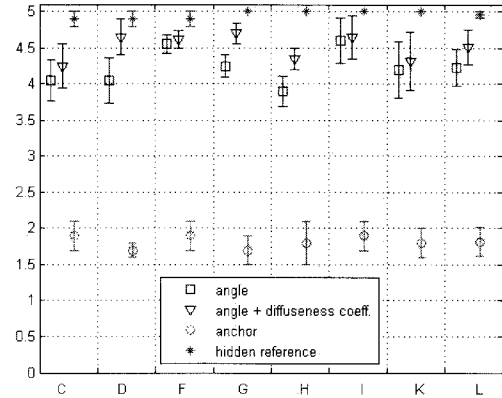
2.4 실험 결과

새로 제안하는 diffuseness 파라미터와 angle정보를 사용한 방법을 활용하여 기존의 angle정보만을 사용한 방법과 음질 비교를 수행하였다. 평가에 사용된 테스트 음원은 MPEG 오디오 그룹에서 멀티채널 오디오 코덱의 평가에 사용되는 11개의 음원을 사용하였다.([표 1] 참조)

<표 1> Test Materials

Index	Material Name	Category
A	BBC Applause	Pathological & Ambience
B	ARL Applause	Pathological & Ambience
C	Chostakovitch	Music(back:direct)
D	Fountain music	Pathological & Ambience
E	Glock	Pathological & Ambience
F	Indie2	Movie sound
G	Jackson1	Music(back:ambience)
H	Pops	Music(back:direct)
I	Poulenc	Music(back:direct)
J	Rock concert	Music(back:ambience)
K	Stomp	Movie sound
L	Total average	N/A

실험 방법은, 2.2에서 설명한 angle정보만을 사용하여 encoding/decoding을 수행한 그룹과, 새로 제안한 diffuseness 파라미터를 front Left, front Right 채널에 적용하여 encoding/decoding을 수행한 그룹으로 구분하여 두 그룹의 음질을 MUSHRA 테스트[5]로 평가하였다. 여기서 diffuseness 파라미터(D)는 0 ~ 1까지를 0.2 단위로 나누어 가장 성능이 좋은 D값을 heuristic한 방법으로 찾아서 적용하였다. 또한 D값이 front Left, front Right 두 채널 모두 0인 4개(A,B,E,J)의 음원은 제외하고, 나머지 7개의 음원에 대해서만 평가를 수행하였다. 평가 결과는 아래 <그림 4>과 같이 3개의 음원(D,G,H)은 음질이 향상되었고, 4개의 음원에서는 비슷한 수준이었으며 평균적으로 MUSHRA test score 0.3의 향상이 있었다. 음질이 향상되는 3개의 음원은 다른 음원에 비해 채널간의 유사도가 떨어지거나 음상의 이동이 많은 특징이 있다.



<그림 4> result of quality test

3. 결론 및 향후 연구 방향

본 논문에서는 파라메트릭 멀티채널 오디오 코딩 방식에 있어서, 다운믹스 과정에서 사라지는 채널간의 liveness를 보강해주는 diffuseness 파라미터를 사용하는 방법을 제안하였다. 이 방법은 추가적인 비트 소요가 거의 없으면서 음질이 향상되거나 거의 비슷한 수준으로 유지됨을 확인할 수 있었다. 현재는 최적의 계수값을 실험적인 방법으로 찾아서 적용을 하였으나, 앞으로의 연구 방향은 인코더에서 최적의 계수를 선정하는 정량적인 측정 방법을 찾는 것이다. 또한 추가적인 비트 소요가 허락되는 경우에는 프레임마다 최적의 계수값을 찾아 변화시켜주는 방법에 대한 연구도 진행되어야 할 것이다.

[참고 문헌]

- [1] J.Breebaart, J.Herre, C.Faller, J.Roden, F.Myburg, S.Disch, H.Purnhagen, G.Hotho, M.Neusinger, K.Kjorling, W.Oomen, "MPEG Spatial Audio Coding / MPGE Surround: Overview and Current Status", 119th Conv. Aud io Eng. Soc., Oct. 2005
- [2] V.pulkki, M.Karjalainen, "Localization of Amplitude-Panned Virtual Source I:Stereophonic Panning", J. Audio Eng. Soc., vol 49, no. 9, pp. 739-752, Sep. 2001
- [3] V.pulkki, "Localization of Amplitude-Panned Virtual Source II:three-dimensional panning", J. Audio Eng. Soc., vol 49, no. 9, pp. 753-767, Sep. 2001
- [4] Han-gil Moon; Jeong-il Seo; Seungkwon Baek; Koeng-Mo Sung; "A multi-channel audio compression method with virtual source location information for MPEG-4 SAC"Consumer Electronics, IEEE Transactions on Volume 51, Issue 4, Nov. 2005 Page(s):1253-1259
- [5] ITU-R Recommendation BS.1534-1, "Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)", ITU, Geneva, Switzerland, 2001