

일반논문 (Regular Paper)

방송공학회논문지 제21권 제3호, 2016년 5월 (JBE Vol. 21, No. 3, May 2016)

<http://dx.doi.org/10.5909/JBE.2016.21.3.369>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

채널 간 위상차 파라미터 생략 기법을 이용한 오디오 부호화

김 현 휘^{a)}, 김 인 철^{a)†}

An Audio Coding Technique Employing the Inter-channel Phase Difference Skip

Hyun-Hwi Kim^{a)} and Rin-Chul Kim^{a)†}

요 약

본 논문에서는 음성/음악 통합 압축 기술(USAC)에 적용된 공간 오디오 부호화 기법인 MPEG 서라운드에서 채널 간 위상차(IPD) 파라미터를 효과적으로 생략하는 기법에 대해 다룬다. 기존의 방법에서는 파라미터 밴드의 IPD 파라미터가 모두 작은 경우에도 전체를 처리하여 전송한다. 이러한 경우 생략을 통해 비트 효율을 향상시킬 수 있다. IPD 파라미터의 심리음향적 민감도를 고려하여 청취 환경에서 인지하지 못하는 문턱 값을 추정하고, 문턱 값 아래의 위상차를 발생시키는 IPD 파라미터는 0으로 설정하고 전송하지 않는다. 제안하는 채널 간 위상차 파라미터 생략 기법을 적용하면 기존에 비해 38% 정도의 위상 파라미터 절감 효과를 얻을 수 있고, MUSHRA 청취 실험을 통해 복호화된 오디오의 음질 하락이 없음을 보인다.

Abstract

This paper deals with an efficient method for skipping inter-channel phase differences (IPD) in the MPEG surround of the unified speech and audio coding (USAC). Based on the psycho-acoustic sensitivity on the IPD, we estimate a threshold on IPD, below which we can not notice degradation in spatial cue. We propose an IPD skip method, in which any IPDs within the threshold are set to zero and are not transmitted. The proposed IPD skip method gives about 38% savings in terms of bit amount for IPD. Nevertheless, in the MUSHRA test, the proposed method does not show any noticeable degradation in the decoded audio quality.

Keyword : MPEG Surround(MPS), Unified speech and audio coding(USAC), Inter-channel phase difference(IPD), Overall phase difference(OPD)

a) 서울시립대학교 전자전기컴퓨터공학부(Dept. of ECE, Univ. of Seoul)

† Corresponding Author : 김인철(Rin-Chul Kim)

E-mail: rin@uos.ac.kr

Tel: +82-2-6490-2336

ORCID: <http://orcid.org/0000-0002-7643-8717>

※ 이 논문은 2014년도 서울시립대학교 교내학술연구비에 의하여 지원되었음.

※ 이 논문의 연구 결과 중 일부는 “2015년 한국방송공학회 하계학술대회”에서 발표한 바 있음.

· Manuscript received September 4, 2015; Revised May 4, 2016; Accepted May 23, 2016.

1. 서 론

MPEG 서라운드(MPS : MPEG Surround)^[1]는 공간 오디오 부호화(SAC : spatial audio coding) 기법의 한 종류로 가장 최근 표준화된 코덱인 음성/음악 통합 압축 기술(USAC : unified speech and audio coding)^[2]에도 적용되어 있다.

MPS에서는 스테레오 채널을 모노 채널로 다운믹스하는

과정에서 공간 파라미터를 추출하는데 여기는 채널 간 레벨차(CLD : inter-channel level difference), 채널 간 상관도(ICC : inter-channel correlation), 채널 간 위상차(IPD : inter-channel phase difference), 전체 위상차(OPD : overall phase difference)가 있다. IPD 파라미터는 채널 간 시간차(ITD : inter-channel time difference)로 나타낼 수 있는데 ITD 파라미터는 SAC의 한 종류인 양이 단서 부호화(BCC : binaural cue coding)^[3]에서 적용되던 파라미터로 이와 관련해 시간차로 인한 방향 인지의 청취 민감도를 추정하는 연구가 활발히 진행되어 왔다.

본 논문에서는 MPS에서의 스테레오 복원과 복원 시 실제 발생하는 시간차에 대해 알아보고 이를 통하여 청취 민감도를 고려한 허용 가능한 위상 변화를 결정한다. 허용 가능한 위상 변화를 제안하는 생략 기법에 적용하고 비트 효율과 주관적 음질평가를 통해서 그 성능을 평가한다.

본 논문의 구성은 다음과 같다. 2장에서는 스테레오 복원 과정과 채널 간 시간차와 그에 따른 청취 민감도를 설명하였다. 3장에서는 제안하는 허용 가능한 위상 변화와 생략 기법을 설명하였으며 4장에서는 제안한 기법의 성능을 실험을 통해서 확인한다. 마지막으로 5장에서는 본 논문에 대한 결론을 맺는다.

II. 스테레오 복원

1. 스테레오 파라미터 추출

MPS의 부호화기는 스테레오 정보를 추출하고 모노로 다운믹스하기 위해 여러 주파수 대역을 통합해 하나의 파라미터를 가지게 만든 파라미터 밴드 별로 앞서 설명한 공간 파라미터를 계산한다^[4].

CLD 파라미터는 채널 간의 레벨 차이를 의미한다. 하이브리드 부 대역의 왼쪽 신호를 $x_{1,m}[k]$, 오른쪽 신호를 $x_{2,m}[k]$ 라 하고 m 은 시간, b 는 파라미터 밴드의 인덱스, k_b 는 파라미터 밴드의 시작 인덱스라 하면 b 번째 파라미터 밴드의 CLD 파라미터는 식 (1)과 같이 정의된다.

$$CLD[b] = 10 \log_{10} \left(\frac{\sum_m \sum_{k=k_b}^{k_{b+1}-1} x_{1,m}[k] x_{1,m}^*[k]}{\sum_m \sum_{k=k_b}^{k_{b+1}-1} x_{2,m}[k] x_{2,m}^*[k]} \right) \quad (1)$$

계산된 CLD 파라미터 값은 -15부터 15까지 해당하는 31개의 인덱스로 로그 스케일 비선형 양자화되어 전송된다.

ICC 파라미터는 채널 간의 상관도를 의미한다. ICC 파라미터를 추정하는 방법은 두 가지인데 하나는 `bsPhaseCoding = 0`으로 IPD 파라미터를 사용하지 않을 때, 다른 하나는 `bsPhaseCoding = 1`로 IPD 파라미터를 사용할 때이다. 먼저 IPD 파라미터를 사용하지 않을 때의 ICC 파라미터는 식 (2)와 같이 정의된다.

ICC 파라미터는 -1부터 1까지의 값을 가지게 되며 위상 정렬되지 않은 형태로 사용된다. -1은 π 의 위상차를 나타내며, 0은 무상관, 1은 동일한 신호임을 나타낸다. 복호화기에서 ICC 파라미터를 통해 다운믹스 신호와 디코릴레이션 신호를 조합하여 채널 간의 위상을 포함한 상관도를 복원한다. IPD와 그에 따른 OPD 파라미터를 사용하지 않으므로 다운믹스 신호를 기준으로 상대적인 복원만이 가능하며, 0 이하의 값을 가질 경우 인위적으로 생성된 디코릴레이션 신호가 다운믹스 신호보다 많이 포함되어 음질 하락을 야기한다. 다음으로 IPD 파라미터를 사용할 때의 ICC 파라미터는 식 (3)과 같이 정의된다.

$$ICC[b] = Re \left\{ \frac{\sum_m \sum_{k=k_b}^{k_{b+1}-1} x_{1,m}[k] x_{2,m}^*[k]}{\sqrt{\sum_m \sum_{k=k_b}^{k_{b+1}-1} x_{1,m}[k] x_{1,m}^*[k]} \sqrt{\sum_m \sum_{k=k_b}^{k_{b+1}-1} x_{2,m}[k] x_{2,m}^*[k]}} \right\} \quad (2)$$

$$ICC[b] = \frac{\left| \sum_m \sum_{k=k_b}^{k_{b+1}-1} x_{1,m}[k] x_{2,m}^*[k] \right|}{\sqrt{\sum_m \sum_{k=k_b}^{k_{b+1}-1} x_{1,m}[k] x_{1,m}^*[k]} \sqrt{\sum_m \sum_{k=k_b}^{k_{b+1}-1} x_{2,m}[k] x_{2,m}^*[k]}} \quad (3)$$

ICC 파라미터는 0부터 1까지의 값을 가지게 되고 이는 위상 정렬 이후의 채널 간 상관도 값이다. IPD 파라미터로 위상 정보를 처리하므로 상관도를 다르게 추정하는 방식이다. 0은 무상관, 1은 동일한 신호임을 나타낸다. 복호화기에서 ICC 파라미터를 통해 채널 간 상관도를 복원하고 IPD와 OPD 파라미터를 통해 채널 간 상대적인 위상 차이를 복원할 수 있다. 식 (2)와 (3)의 결과는 공통적으로 8개의 인덱스로 비선형 스케일 양자화되어 전송된다.

IPD 파라미터는 채널 간 위상차를 의미한다. IPD 파라미터는 식 (4)와 같이 정의된다.

$$IPD[b] = \angle \left(\sum_m \sum_{k=k_b}^{k_{b+1}-1} x_{1,m}[k] x_{2,m}^*[k] \right) \quad (4)$$

양자화는 Coarse와 Fine, 두 가지 방식 중 하나를 선택하게 되며 각각 8개, 16개의 인덱스로 선형 스케일 양자화되어 전송된다. IPD 파라미터는 채널 간의 상대적인 위상차를 의미하기 때문에 기준이 어느 점이 되는지는 알 수 없다. 이 기준을 명시하기 위해 OPD 파라미터가 사용된다.

OPD 파라미터는 두 개의 신호를 하나의 모노 신호로 다운믹스했을 때, 다운믹스 신호와 왼쪽 채널 신호 간의 위상차를 의미한다. $s_{1,m}$ 을 다운믹스된 신호라 하면 OPD 파라미터는 식 (5)와 같이 정의된다.

$$OPD[b] = \angle \left(\sum_m \sum_{k=k_b}^{k_{b+1}-1} x_{1,m}[k] s_{1,m}^*[k] \right) \quad (5)$$

OPD 파라미터는 다른 공간 파라미터와 다르게 부호화

단계에서 계산되지 않고 복호화 단계에서 전송된 CLD와 IPD 파라미터에 의해 예측된다^[5]. 즉, 직접 전송되지 않는다.

2. 스테레오 복원

계산된 공간 파라미터는 모노로 다운믹스된 신호와 함께 복호화기에서 스테레오로 복원된다. 이 과정에서 복호화기는 업믹스 행렬을 구성하여 사용하게 된다. c_1 은 왼쪽 채널의 크기, c_2 는 오른쪽 채널의 크기, α 는 상관도에 대한 회전 파라미터, β 는 채널 에너지에 대한 회전 파라미터이고 c_1 , c_2 , α , β 는 각각 식 (6), (7), (8), (9)와 같이 정의된다.

$$c_1 = \sqrt{\frac{\frac{CLD[b]}{10}}{1 + \frac{CLD[b]}{10}}} \quad (6)$$

$$c_2 = \sqrt{\frac{1}{1 + \frac{CLD[b]}{10}}} \quad (7)$$

$$\alpha = \frac{1}{2} \arccos(ICC[b]) \quad (8)$$

$$\beta = \arctan\left(\frac{c_2 - c_1}{c_2 + c_1} \tan(\alpha)\right) \quad (9)$$

$s_{1,m}$ 은 다운믹스된 신호, $D(s_{1,m})$ 은 디코릴레이션된 다운믹스 신호, $x_{1,m}^r$ 은 복원된 왼쪽 신호, $x_{2,m}^r$ 은 복원된 오른쪽 신호이고 업믹스 행렬을 이용한 스테레오 복원은 식 (10)과 같이 정의된다.

우선, 출력 신호는 서로 직교하는 다운믹스된 신호와 디코릴레이션된 다운믹스 신호, 그리고 CLD와 ICC 파라미터에 의해 본래의 크기와 상관도를 가지는 스테레오 신호로 복원된다. 이는 그림 1과 같이 벡터 형태의 표현이 가능하다.

$$\begin{bmatrix} x_{1,m}^r \\ x_{2,m}^r \end{bmatrix} = \begin{bmatrix} e^{jOPD} & 0 \\ 0 & e^{j(IPD-OPD)} \end{bmatrix} \begin{bmatrix} c_1 \cos(\beta+\alpha) & c_1 \sin(\beta+\alpha) \\ c_2 \cos(\beta-\alpha) & c_2 \sin(\beta-\alpha) \end{bmatrix} \begin{bmatrix} s_{1,m} \\ D(s_{1,m}) \end{bmatrix} \quad (10)$$

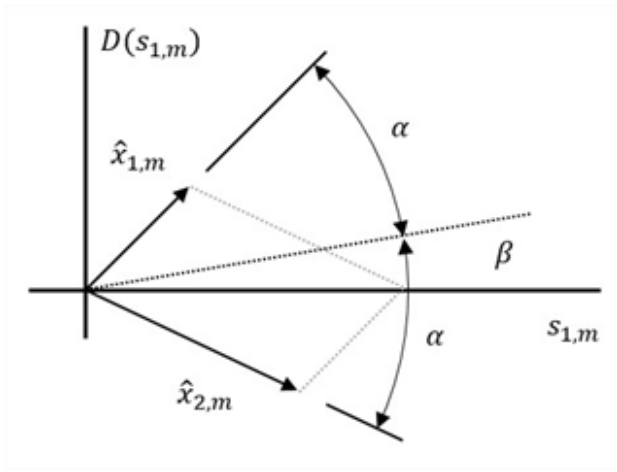


그림 1. 위상 파라미터를 제외한 업믹싱 표현
Fig. 1. Up-mixing representation without phase

여기에 마지막으로 IPD 파라미터와 계산된 OPD 파라미터에 의해 위상 정정되는 과정을 거친다. OPD 파라미터로 왼쪽 신호의 본래 위상을 복원하고, 왼쪽 신호의 본래 위상인 OPD 파라미터에 상대적인 위상차인 IPD 파라미터의 차를 이용하여 오른쪽 신호의 본래 위상을 복원한다.

3. 복원 시 발생하는 시간차

앞서 언급한대로 채널 간 위상차인 IPD 파라미터는 ITD 파라미터로 표현이 가능한데, 이는 파라미터 밴드 인덱스의 중심 주파수와 밀접한 관련이 있다. ITD는 IPD가 사용되기 이전에 사용되던 파라미터이나 2π 이상의 위상 관계를 표현함에 있어 문제를 내포하여 IPD 파라미터로 대체되었다⁶⁾. 예를 들면 ITD 파라미터가 0.6ms이고 중심 주파수가 1.66kHz일 때 2π 인근의 위상 관계를 표현하는데 이는 근사적으로 정위상의 관계이다. ITD 파라미터는 시간 딜레이로 양 신호의 위상을 표현하기 때문에 주파수에 따른 폴이 관계를 명확하게 하지 않으면 ITD 파라미터에 의해 본래와 다른 위상으로 복원이 이루어질 수 있다. 스테레오 신호의 방향 인지에 관한 연구는 주로 채널 간 시간차를 통해 이루어진다. 이를 참조하기 위해 여기서는 IPD 파라미터에 의해 정상적인 스테레오 복원을 마친 후의 위상차를 ITD 파라미터로 변환하여 사용할 것이다. b 는 파라미터 밴드의

인덱스, $f_c[b]$ 는 해당 파라미터 밴드의 중심 주파수라고 할 때, IPD와 ITD 파라미터의 관계는 식 (11)과 같이 정의된다.

$$IPD[b] = 2\pi f_c[b]D[b] \tag{11}$$

스테레오 복원을 통해 실제 청취 환경에서 인지하게 되는 위상차를 ITD 파라미터로 변환하고 ITD 문턱 값과 비교함으로써 허용 가능한 위상 변화를 결정할 것이다.

IPD 파라미터가 0 인근의 값일 경우 상관도 정정 이후 위상차가 없는 0을 기점으로 연속된 파라미터 밴드에서 \pm IPD가 혼재된 모습을 보인다. 이는 IPD 파라미터의 지속성을 분석하여 파악이 가능하다. IPD 파라미터의 지속성 분석과 4장에서 비트 효율 분석, 청취 실험에 사용될 샘플들은 표 1과 같다. 표 1의 샘플들은 제 84차 MPEG 회의에서 USAC 코덱들의 성능 평가를 위해 선정된 샘플들 중 일부이다.

표 1. 실험 샘플
Table 1. Test samples

Sample	Note
es01	Speech
louis_raquin_15	
Arirang_speech	
twinkle_ff51	Mixed
SpeechOverMusic_1	
SpeechOverMusic_4	
HarryPotter	Music
te15	
Music_1	
Music_3	

IPD 파라미터의 지속성이란, 연속된 파라미터 밴드에서 단일 채널이 일괄적으로 선행하는 경우와 그렇지 않은 경우를 파악한 것이다. 연속된 파라미터 밴드의 수는 단일 파라미터 밴드를 기준으로 양 옆의 파라미터 밴드를 고려하기 위해 3개로 설정하였다. 20개의 파라미터 밴드를 사용하

면 위상 파라미터의 전송은 위상이 청취자의 방향 인지에 지대한 영향을 미치는 약 1.5kHz 미만의 10개 파라미터 밴드만 전송하게 되는데 지속성 분석을 위해 설정된 3개의 연속된 파라미터 밴드 수를 고려하면 한 프레임 당 8번의 카운트를 하게 된다. IPD 파라미터의 지속성을 분석한 결과는 표 2와 같다.

연속된 3개의 파라미터 밴드가 왼쪽 혹은 오른쪽 채널만을 일괄적으로 선행시킨다면 지속된 파라미터 밴드로, 일괄적이지 않다면 지속되지 않은 파라미터 밴드로, 위상차가 없는 경우 위상차 없음으로 포함시켰다. 또한 지속되거나 지속되지 않은 파라미터 밴드의 경우 위상차 값의 평균을 구해 어느 위상차를 기준으로 분포하였는지를 파악하였다.

표 2를 보면 대부분의 샘플이 인접한 파라미터 밴드 간의 IPD 파라미터가 특정 채널을 지속적으로 선행시키지 않는다는 것을 알 수 있다. 하지만 Music_3 등 일부 샘플에 대해서는 지속된 파라미터 밴드가 그와 비견될 정도로 많은 특성을 보인다.

지속되지 않은 파라미터 밴드의 위상 평균을 보면 근사적으로 0에 가까운 값이다. 이는 0을 기점으로 \pm IPD가 혼재됨을 뜻한다. 연속된 파라미터 밴드는 청취자가 완전히 분리하여 듣지 못하므로 위상차가 단일 채널을 일괄적으로

선행시키지 않을 경우 단일 파라미터 밴드에서 가지는 IPD 파라미터로 변환된 ITD 파라미터에 의한 시간차보다 더 작은 시간차를 인지하게 된다.

채널 간 레벨차가 없다고 가정하면 다운믹스 신호에 대한 왼쪽 신호와 오른쪽 신호의 위상 정정이 같아지게 된다. $PB(i)_{Left,Right}$ 를 i 번째 파라미터 밴드의 왼쪽과 오른쪽 위상, $PB(i+1)_{Left,Right}$ 를 그 다음 파라미터 밴드의 왼쪽과 오른쪽 위상이라고 할 때 단일 파라미터 밴드가 가지는 시간차를 그림으로 나타내면 그림 2(a)와 같다. 지속되지 않은 파라미터 밴드를 가정하여 $PB(i)$ 는 왼쪽 신호를 선행시키고 $PB(i+1)$ 은 오른쪽 신호를 선행시키면 실제 청취 환경에서 인지하게 되는 시간차를 그림 2(b)와 같이 나타낼 수 있다. 지속된 파라미터 밴드의 경우 $PB(i+1)_{Left,Right}$ 를 $PB(i)_{Left,Right}$ 보다 더 높은 주파수 대역의 파라미터 밴드라고 가정하였을 때 단일 파라미터 밴드의 경우보다는 시간차가 감소하는 것을 볼 수 있는데 그림으로 나타내면 그림 2(c)와 같다. 연속하는 파라미터 밴드를 대상으로 하므로 더 낮은 주파수 대역의 파라미터 밴드 또한 고려해야 하는데 $PB(i-1)_{Left,Right}$ 를 더 낮은 주파수, 이전 파라미터 밴드의 왼쪽과 오른쪽 위상이라고 할 때, $PB(i-1)_{Left,Right}$ 와 $PB(i+1)_{Left,Right}$ 의 IPD 파라미터에 따라서 단일 파라미터 밴드의 경우보다 시간차가 증가할 수도, 감소할 수도

표 2. IPD 파라미터의 지속성
 Table 2. Sustainability of IPD Parameters

Sample	Sustained PB	Non-sustained PB	Zero phase	Sustained PB phase avg.	Non-sustained PB phase avg.
es01	221	457	1218	1.634765	-0.376736
louis_raquin_15	0	0	1896	-	-
Arirang_speech	327	924	645	-0.179337	0.027954
twinkle_ff51	44	327	1525	0.916298	0.074800
SpeechOverMusic_1	158	804	934	0.270084	0.070590
SpeechOverMusic_4	229	174	1493	-0.401273	0.042951
HarryPotter	0	35	1861	-	-0.067707
te15	403	1141	352	0.639232	0.075727
Music_1	303	920	673	0.415596	0.097358
Music_3	794	887	215	-0.231795	0.204515

있다. 증가하는 경우를 그림으로 나타내면 그림 2(d)와 같다.

따라서 단일 파라미터 밴드에서 ITD 파라미터를 얻는다면 IPD 파라미터에 의한 양쪽 채널의 상대적인 위상차만 필요로 하지만 인접한 파라미터 밴드를 고려한다면 위상차가 지속되지 않은 경우 매우 작은 값을 가지며 위상차가 지속되는 경우는 기존 파라미터 밴드의 양 옆의 파라미터 밴드의 IPD 파라미터에 따라 달라지게 된다.

일반적인 청취 환경에서 위상적으로 선행하는 귀의 소리

가 먼저 인지되고 크기 역시 더 큰 값을 가지기 때문에 IPD, OPD 파라미터에 의한 위상 정정시 정위상의 다운믹스 신호에 가까운 값을 가지게 되어 레벨차가 없다고 가정한 위 경우보다 시간차는 더욱 감소하게 된다.

4. 시간차의 민감도

시간차의 민감도는 양 귀에 들리는 신호의 ICC 파라미터가 작을수록 둔감해지게 된다. 또한 잡음에 의해서 역시 둔

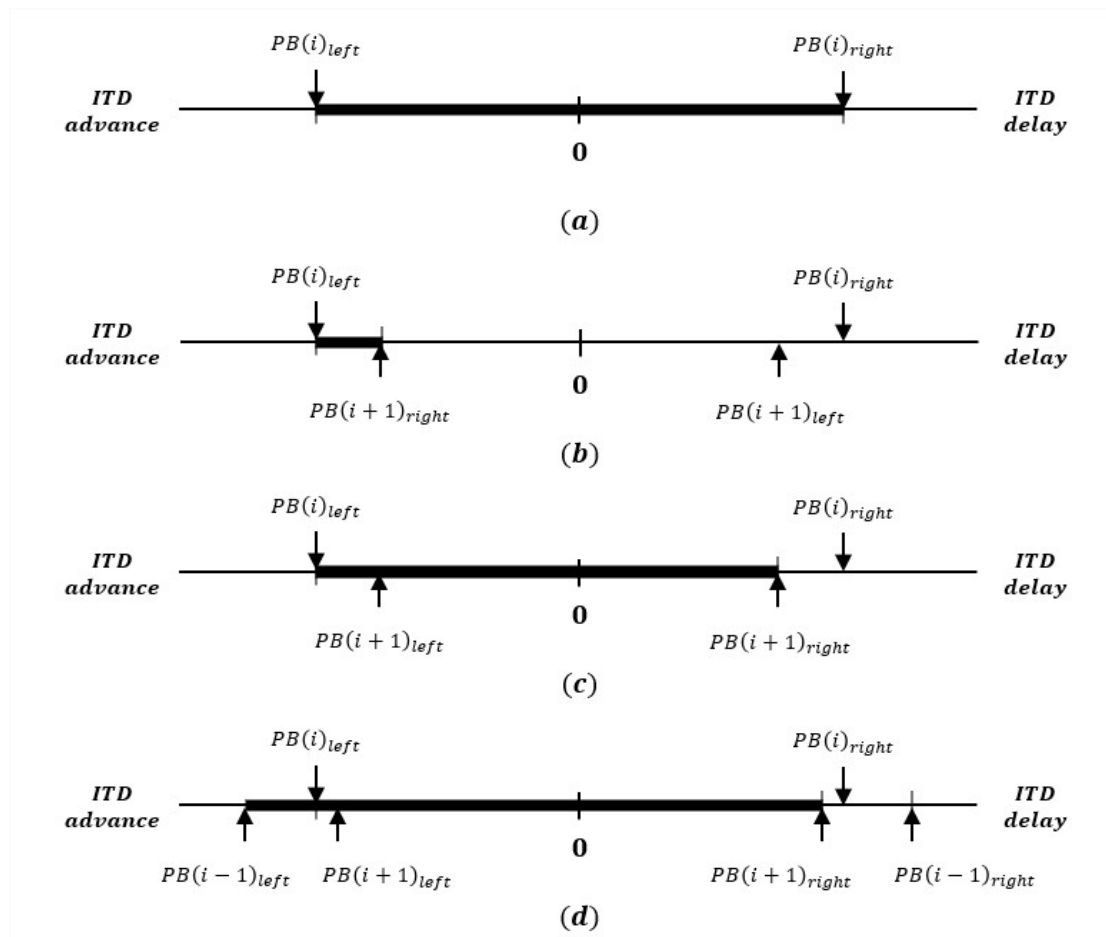


그림 2. 실제 청취 환경에서 인지하게 되는 시간차. (a) 단일 파라미터 밴드의 시간차. (b) 위상차가 지속되지 않는 연속된 파라미터 밴드의 시간차. (c) 위상차가 지속되는 연속된 파라미터 밴드의 시간차. (d) 양 옆 파라미터 밴드의 위상차를 고려한 시간차가 증가하는 경우.

Fig. 2. Perception of time difference in actual listening environment. (a) Time difference of single parameter band. (b) Time difference of series parameter bands with non-sustained IPD parameters. (c) Time difference of series parameter bands with sustained IPD parameters. (d) Case of increasing time difference with series parameter bands

감해지게 되는데, 이러한 환경에서 중심 주파수가 225Hz일 땐 $125\mu s$, 715Hz일 땐 $50\mu s$ 이상의 ITD 문턱 값을 가진다^{[7][8]}. 표 1의 샘플들의 경우 ICC 파라미터 값이 전체적으로 작고 다양한 주파수 성분이 포함되어 있으므로 ITD 문턱 값은 이보다 높다.

위상차 파라미터를 전송하는 파라미터 밴드 중 상대적인 저주파 대역인 파라미터 밴드 2~4를 그룹 1, 상대적인 고주파 대역인 파라미터 밴드 5~7을 그룹 2라고 할 때 20개의 파라미터 밴드를 사용하는 경우에 채널 간 레벨차가 없고 동일한 위상차로 \pm 가 교차하면 기존의 IPD Fine 모드로 $-\pi/4$ 이상, $\pi/4$ 이하에서 발생할 수 있는 최대 시간차는 그룹 1에서 $130\mu s$, 그룹 2에서 $52\mu s$ 이다. 이와 같은 시간차는 ITD 문턱 값 이하의 값일 것으로 예상되며 생략할 수 있다.

IPD 파라미터가 단일 채널이 일괄적으로 선행하도록 지속되는 경우는 단일 파라미터 밴드보다 시간차가 증가할 수도, 감소할 수도 있으나 증감폭이 크지 않다. 따라서 항상 ITD 문턱 값 이상의 값을 발생시키며 같은 ITD 문턱 값

이상의 값이라고 해도 그 크기에 따라서 방향 인지에 영향을 미치지 때문에 임의로 수정될 경우 공간감의 손상이 발생할 수 있다.

따라서 단일 채널이 일괄적으로 선행하지 않은 경우에 한해서 몇 가지 조건에 의해 IPD 파라미터를 생략하여 0으로 처리가 가능하다.

III. 제안된 생략 기법

기존의 방법에서는 한 프레임 전체의 IPD 파라미터가 0인 경우에도 이를 모두 처리하는데 제안하는 생략 기법을 사용하면 기존에 고려되지 않던 인지하지 못하는 정도의 시간차를 발생시키는 경우의 IPD 파라미터 조합을 USAC의 IPD 코딩 플래그인 bsPhaseMode를 0으로 만들어서 복호화기에서 한 프레임 전체의 IPD 파라미터가 0인 경우와 동일한 효과를 낼 수 있다. IPD 코딩을 하지 않은 경우를 뜻하기 때문에 위상차 파라미터를 생략하게 되어 전체적인

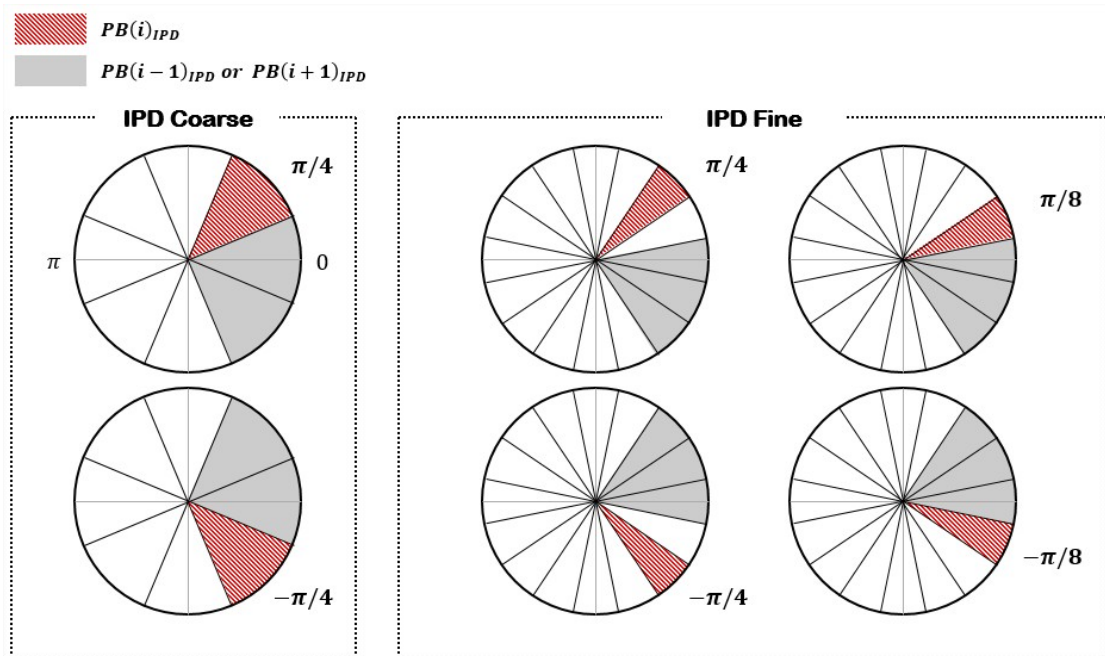


그림 3. 위상 변화 조건
 Fig. 3. Phase alteration terms

비트 효율이 향상된다.

인지하지 못하는 정도의 시간차 조합은 생략하고 외의 경우 보존해야하기 때문에 IPD 파라미터의 변화 조건이 필요하다. 조건을 만족하지 못하면 인지할 수 있는 정도의 시간차 조합을 발생시키는 것으로서 2장에서 설명한 지속되는 경우의 파라미터 밴드이다. 첫 번째로 연속되는 모든 파라미터 밴드는 $-\pi/4$ 이상, $\pi/4$ 이하의 IPD 파라미터 값을 가져야 한다. 이 외의 값은 상대적으로 큰 시간차를 가지기 때문에 IPD 파라미터의 지속성 분석에서 큰 부분을 차지하는 인접한 파라미터 밴드의 IPD 파라미터 값이 0인 경우에 ITD 문턱 값 이상의 시간차를 발생시킨다. 두 번째로 인접한 파라미터 밴드끼리는 동일한 IPD 파라미터 값이나 동일한 채널을 선행시키는 IPD 파라미터 값을 가질 수 없다. 단일 채널을 선행시키는 지속되는 경우의 파라미터 밴드 조합이므로 역시 ITD 문턱 값 이상의 시간차를 발생시킨다. 이 조건을 그림으로 나타내면 그림 3과 같다.

IPD 파라미터 지속성 분석에서 단일 채널이 일괄적으로 지속되지 않은 경우와 IPD 파라미터가 0인 경우가 많은 샘플이 조건을 만족시키기 쉬워진다. 이런 샘플은 표 1에서 대체로 음성 신호 혹은 혼합 신호인 경우가 많으나 녹음 환경이나 주파수 성분 분포, 위상차 크기 등의 변수로 인해 음성/혼합 신호가 반드시 이 조건을 만족하는 것은 아니다.

생략 기법의 적용을 위해 비트 스트림 구조를 살펴보면 그림 4와 같다. 그림 4(a)는 성공적으로 생략되었을 때의

위상 파라미터 비트 스트림 구조이며 그림 4(b)는 생략하지 못했을 경우이다. 기존의 비트 스트림 구조는 그림 4(c)와 같다.

생략 기법을 사용하기 위해 그림 3의 조건을 한 프레임 전체가 만족하는 경우 bsPhaseMode를 0으로 설정한다. 복호화기에서는 bsPhaseMode가 0이기 때문에 IPD 코딩을 하지 않는 것으로 판단하여 위상 정정을 하지 않는다. 이 경우 부호화기에서 처음 판단된 한 프레임의 IPD 파라미터들이 모두 0이 아니었다고 해도 그림 3의 조건을 만족하여 인지할 수 없는 정도의 시간차를 발생시켰기 때문에 실제 청취 환경에서 음질의 손상이 없다. 20개의 파라미터 밴드를 사용하면서 생략 기법을 사용하지 않을 경우 한 프레임에서 IPD 코딩이 Coarse일 때 38비트, Fine일 때 48비트가 필요한 처리를 생략 기법을 사용하면 1비트로 줄일 수 있다. 조건을 만족하지 못해 생략하지 못했을 경우에는 비트 소모, 음질 상의 어떠한 차이도 보이지 않는다.

IV. 모의 실험

채널 간 위상차 파라미터 생략 기법의 비트 효율을 표 1의 실험 샘플을 통해 확인한 결과는 표 3과 같다. IPD 코딩이 Fine일 때는 Coarse일 때와 흡사한 동향을 보이기 때문에 생략하였다. 괄호 안의 수는 기존 USAC에 IPD 파라

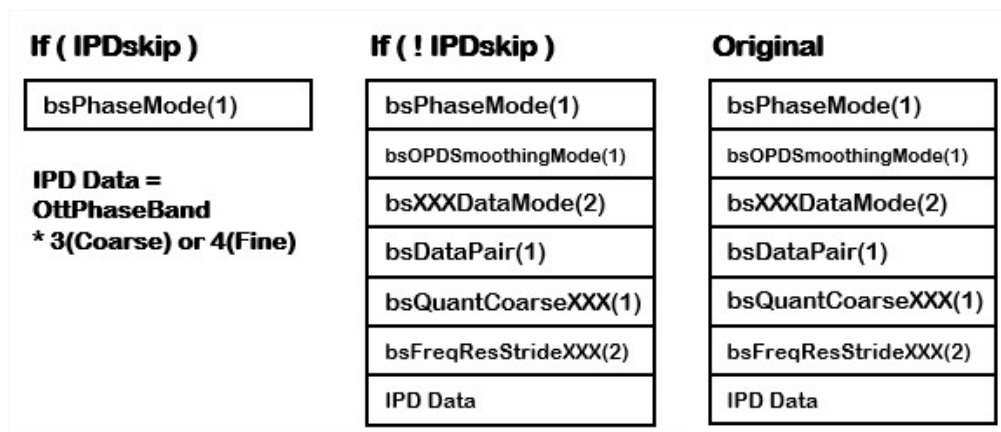


그림 4. 제안된 비트 스트림 구조
 Fig. 4. Proposed bit stream structure

표 3. 비트 효율
 Table 3. Bit efficiency

	Output type	Current coarse IPD	Proposed coarse IPD		Output type	Current coarse IPD	Proposed coarse IPD
es01	Skip frame	0 (54)	159	Harry Potter	Skip frame	0 (207)	226
	Bit cost	9044	3002		Bit cost	9044	456
	Percentage	100%	33.20%		Percentage	100%	5.04%
louis_raquin_15	Skip frame	0 (238)	238	te15	Skip frame	0 (0)	1
	Bit cost	9044	0		Bit cost	9044	9006
	Percentage	100%	0%		Percentage	100%	99.58%
Arirang_speech	Skip frame	0 (0)	0	Music_1	Skip frame	0 (0)	4
	Bit cost	7562	7562		Bit cost	9044	8892
	Percentage	100%	100%		Percentage	100%	98.32%
twinkle_ff51	Skip frame	0 (0)	127	Music_3	Skip frame	0 (0)	1
	Bit cost	9044	4218		Bit cost	9044	9006
	Percentage	100%	46.64%		Percentage	100%	99.58%
Speech Over Music_1	Skip frame	0 (5)	54	Average	skip frame percentage	0 (21)%	37.51%
	Bit cost	9044	6992		bit cost	8896	5559
	Percentage	100%	77.31%		bitcost percentage	100%	62.49%
Speech Over Music_4	Skip frame	0 (0)	68				
	Bit cost	9044	6460				
	Percentage	100%	71.43%				

미터의 변화 조건을 적용하지 않고 한 프레임의 전체 파라미터 밴드의 IPD 파라미터가 0인 경우에 생략 가능한 프레임 수이다. louis_raquin_15, HarryPotter와 같이 채널 간 위상차가 대부분 0으로 판단되는 샘플의 경우 위상 변화 조건 없이 생략만 가능하게 하여도 비트 효율을 증가시킬 수 있으나 위상 변화 조건을 적용하면 HarryPotter에서 거의 모든 프레임을 생략할 수 있음을 보인다. es01, twinkle_ff51, SpeechOverMusic_1, 4의 경우 위상 변화 조건 없이는 생략 가능한 프레임이 적으나 적용하면 많은 프레임을 생략할 수 있는 샘플들이다. 특히 twinkle_ff51은 괄호 안의 수 0에서 127 프레임으로 절반 가까이 비트 소모가 감소한다. 이는 0은 아니지만 작은 위상차가 대부분이고 \pm IPD가 혼재된 샘플임을 뜻한다. Arirang_speech, te15, Music_1, 3의 경우 한 프레임에 다양한 위상 성분을 포함하고 있어 생략이 거의 불가능하다. 결과를 통해 약 38% 정도의 위상 파라미터를 감소시킬 수 있음을 확인하였고 전체적으로 볼 때

24kbps 전송률에서 총 비트의 1.39%를 감소시킬 수 있음을 확인하였다.

청취 평가에는 MUSHRA 실험 방법^[9]을 이용하였으며, 13명의 피실험자가 참여하였다. 표 1의 실험 샘플 전체를 이용하지 않고 특징적인 몇 가지 샘플을 대상으로 하였다. IPD 파라미터를 사용하지 않은 경우까지 총 3개의 경우를 비교하였으며, 6개의 실험 샘플에 대해 Sennheiser HD640 헤드폰으로 실험한 결과는 그림 5와 같다.

청취 실험에서 제외된 샘플들 중 louis_raquin_15, Harry-Potter는 위상차가 작은 성향이 강한 아이템으로 기존 USAC에서도 대부분의 IPD 파라미터를 0으로 판단하기 때문에 전체 위상을 0으로 처리하여도 이론적으로 구분이 불가능하다. SpeechOverMusic_1, Music_1은 각각 SpeechOverMusic_4, Music_3와 흡사하기 때문에 제외하였다. 청취 실험에 사용된 샘플들 중 제안된 생략 기법의 장점을 잘 설명해주는 샘플들로는 SpeechOverMusic_4, twinkle_ff51이 있다. 이

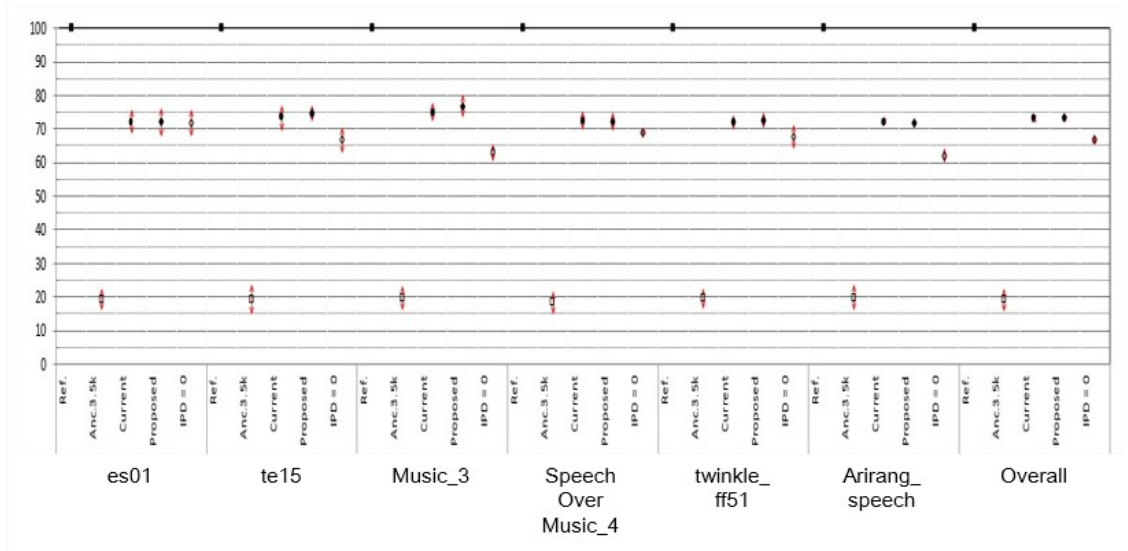


그림 5. MUSHRA 청취 실험 결과
Fig. 5. MUSHRA test result

샘플들은 기존 USAC에서 IPD 파라미터를 대부분 0으로 판단하지 않고, 전체 위상을 0으로 처리하는 경우 음질 하락을 발생시키며 제안된 생략 기법을 적용하였을 때 음질 하락 없이 비트 효율만 증가하는 샘플들이다. es01은 많은 프레임이 생략되나 전체 위상을 0으로 처리하여도 주관적 음질 평가에서 눈에 띄는 차이를 보이지 않았다. Arirang_speech, te15, Music_3는 생략이 거의 불가능하여 비트 효율 역시 기존과 흡사하였다. 참고할 점은 Arirang_speech, Music_3는 전체 위상을 0으로 처리하면 주목할 만한 음질 손상을 발생시킨다는 점이다.

V. 결론

본 논문에서는 심리음향적 특성을 이용해 채널 간 위상차 파라미터를 생략하기 위한 기법에 대해 다루었다. 이를 위해 스테레오 복원에 대해 분석하고, 스테레오 복원 시 실제 발생하는 위상차에 따른 시간차를 청취 민감도와 비교하였다. 이를 통해 위상 변화 조건을 결정하고 생략 기법을 제안해 비트 효율과 MUSHRA 청취 실험을 통해 결과를 확인하였다. 제안된 생략 기법을 적용하면 음질 하락 없이 비트 소모 효율이 증가할 수 있음을 확인하였다.

참고 문헌 (References)

- [1] ISO/IEC 23003-1 : Information technology - MPEG audio technologies - Part 1 : MPEG Surround, 2006.
- [2] ISO/IEC 23003-3 : Information technology - MPEG audio technologies - Part 3 : Unified speech and audio coding, 2011.
- [3] C. Faller and F. Baumgarte, "Binaural cue coding - part I : psychoacoustic fundamentals and design principles," IEEE Trans. On Speech and Audio Proc., vol. 11, no. 6, pp. 509-519, 2003.
- [4] J. Breebaart, G. Hotho, J. Koppens, E. Schuijers, W. Oomen, and S. van de Par, "Background, concept, and architecture for recent MPEG surround standard on multichannel audio compression," J. Audio Eng. Soc. Vol. 55, no. 5, pp. 331-351, 2007.
- [5] J. Kim, E. Oh, and J. Robilliard, "Enhanced stereo coding with phase parameters for MPEG unified speech and audio coding," AES 127th Convention, paper no. 7875, 2009.
- [6] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "Parametric coding of stereo audio," EURASIP J. Appl. Signal Process., vol. 9, pp. 1305-1322, 2004.
- [7] B. Rakerd and W. Hartmann, "Localization of sound in room. V. Binaural coherence and human sensitivity to interaural time differences in noise," J. Acoust. Soc. Am. Vol. 128, no. 5, pp. 3052-3063, 2010.
- [8] McFadden D., Jeffress L. A., and Lakey J. R, "Differences in interaural phase and level in detection and lateralization : 1000 and 2000 Hz," J. Acoust. Soc. Am. 52, 1197-1206, 1972.
- [9] ITU-R BS.1534-1, Method for the subjective assessment of intermediate sound quality(MUSHRA), 2011.

저 자 소 개

김 현 휘



- 2014년 : 한성대학교 정보통신공학과(공학사)
- 2016년 : 서울시립대학교 전자전기컴퓨터공학부(공학석사)
- 2016년 2월 - 현재 : KT 미디어사업본부 미디어기술담당
- ORCID : <http://orcid.org/0000-0002-1629-9831>
- 주관심분야 : 오디오 신호처리, 멀티미디어 통신, 음성 인식, IPTV 등

김 인 철



- 1985년 : 서울대학교 제어계측공학과(공학사)
- 1987년 : 서울대학교 제어계측공학과(공학석사)
- 1992년 : 서울대학교 제어계측공학과(공학박사)
- 1992년 8월 - 1994년 3월 : 대우전자 영상연구소
- 1994년 3월 - 1999년 8월 : 한성대학교 정보전산학부 조교수
- 1999년 9월 - 현재 : 서울시립대학교 전자전기컴퓨터공학부 교수
- ORCID : <http://orcid.org/0000-0002-7643-8717>
- 주관심분야 : 오디오 및 비디오 신호처리, 멀티미디어 통신, 실시간 신호처리 등