

입체음향효과 향상을 위한 스테레오-10.2채널 블라인드 업믹스 기법

Stereo-10.2Channel Blind Upmix Technique for the Enhanced 3D Sound

최선웅 · 현동일 · 이석필** · 박영철*** · 윤대희

(Sunwoong Choi, Dong-il Hyun, Suk-Pil Lee**, Young-cheol Park***, and Dae Hee Youn)

연세대학교 전기전자공학과, **전자부품연구원(KETI) 디지털미디어 연구센터,
***연세대학교 컴퓨터정보통신공학부

(접수일자: 2012년 2월 29일; 수정일자: 2012년 5월 2일; 채택일자: 2012년 5월 16일)

초 록: 본 논문에서는 입체음향효과의 향상을 위한 스테레오-10.2채널의 블라인드 업믹스 알고리즘을 제안하였다. 최근에, 소비자들은 더 나은 입체음향효과나 3D 사운드를 즐기길 원하고 다양한 멀티채널 포맷의 등장으로 업믹스 알고리즘들이 연구 되어 왔다. 그러나 기존의 업믹스 알고리즘들은 공간정보를 왜곡하는 문제점을 가지고 있었다. 이러한 문제점을 해결하기 위해 전·후방 채널에 대한 이득 조절 및 10.2 채널의 각 채널별 믹싱 알고리즘을 제안한다. 기존의 상용화된 멀티채널 업믹스 알고리즘들과의 주관적 평가 실험결과 제안한 알고리즘은 입력신호의 공간정보를 유지하면서 입체음향효과를 향상시킨 것으로 확인되었다.

핵심용어: 스테레오, 멀티채널, 업믹스, 다운믹스, PCA, CCP, 비상관기

투고분야: 음향 신호처리 분야(1)

ABSTRACT: In this paper, we proposed the stereo-10.2channel blind upmix algorithm for the enhanced 3D sound. Recently, consumers want to enjoy better sound and the use of a various of multichannel configuration has been steadily improved. Thus, upmix algorithms have been researched. However, conventional upmix algorithms have the problem that distorts the spatial information of original source. To solve this problem and enhance the spatial sound quality, we proposed front and rear channel gain adjustment and 10.2 channel upmix algorithm for each additional channel. The listening test results show that it maintains spatial information of stereo input and enhances 3D sound effects unlike other conventional upmix algorithms.

Key words: Stereo, Multichannel, Upmix, Downmix, PCA, CCP, Decorrelator

ASK subject classification: Acoustic Signal Processing (1)

1. 서 론

오늘 날 멀티미디어 기술의 발전은 영상과 음향에 대한 전반적인 기술의 발전을 가져 왔다. 화려한 색상과 넓은 화면, 고선명 화질, 실감나는 입체 음향 등에 있어 활발한 진행과 발전을 가져왔다. 최근들어 일반 가정에서도 영화관과 같은 입체음향효과를 즐기기 위해 홈 씨어터와 같은 멀티채널 오디오 시스

템에 대한 수요가 급증하고 있다. 멀티채널 오디오 시스템은 실감나는 입체음향효과를 극대화 시키는 데 그 목적이 있다. 이러한 목적을 달성하기 위해 그동안 활발한 연구가 진행되어 왔고, 다양한 스피커 배치와 개수의 사용으로 오디오 렌더링 기술을 꾸준히 발전시켜 왔다^[1-2]. 하지만 무비 사운드 트랙이나 사운드 레코딩들과 같은 멀티채널 포맷으로 이루어진 레코딩은 여전히 제한적으로 사용되고 있고, 대부분의 오디오 콘텐츠들은 2채널의 스테레오(stereo)로 녹음 및 합성 되어있는 문제가 있다. 따라서 스테

*Corresponding author: 최선웅 (sunw224@dsp.yonsei.ac.kr)
120-749 서울특별시 서대문구 신촌동 연세대학교 B공학관 601
디지털신호처리연구실
(전화: 02-2123-4534; 팩스: 02-364-4870)

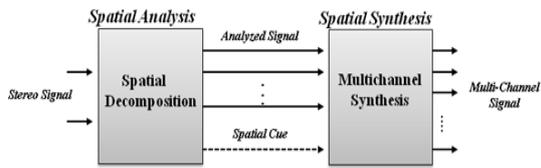


그림 1. 멀티채널 업믹스 시스템: 스테레오-멀티채널
Fig. 1. Multichannel upmix system:
Stereo-to-Multichannel.

레오로 레코딩된 신호를 멀티채널 오디오 시스템에 재생시키기 위한 업믹스(upmix)기술이 필요하다^[3-6]. 일반적인 오디오 신호처리의 과정은 음원 분석(spatial analysis)과 음원 재합성(spatial synthesis)의 두 과정으로 구분^[7] 지을 수 있는데 스테레오 신호에서 멀티채널 신호로의 경우에는 그림 1과 같이 스테레오 음원 분석과 분석된 음원의 재합성의 일련의 과정으로 정의할 수 있다. 음원 분석은 스테레오 입력 신호의 공간단서(spatial cue)를 분석하고 입력신호에 포함되어 있는 음원을 분리(source separation)하는 과정이다. 먼저 입력 2채널 신호를 주파수 도메인에서의 신호모델로 정의하면 식 (1)과 같다.

$$\begin{aligned} X_L &= a_L S + N_L \\ X_R &= a_R S + N_R \end{aligned} \quad (1)$$

스테레오 신호의 좌·우 채널은 패닝게인(panning gain) a_L, a_R 에 의해서 방향성을 가지게 되는 음원 S 와 잔향성분에 해당되는 N_L, N_R 성분이 좌·우 채널에 각각 섞여 있다. 이 신호모델을 바탕으로 음원 분리 과정에서는 음원 S 와 N_L, N_R 성분을 분리하고 공간정보에 해당되는 패닝게인 a_L, a_R 을 분리해 낸다. 이때 분리된 패닝게인은 음향공간에서 음원이 가지고 있는 방향에 해당되고 음원 재합성 과정에서도 왜곡 없이 재현되어야 한다. 또한 패닝게인은 에너지 정규화(energy normalization)를 위한 관계를 갖기 위해 식 (2)의 조건을 만족한다.

$$a_L^2 + a_R^2 = 1 \quad (2)$$

음원 재합성과정은 분석된 공간정보와 분리된 음원을 임의의 멀티채널 오디오 시스템에 적합하도록 음

원을 재합성하고, 입력신호가 가지고 있는 공간정보를 재현하는 과정이다. 스테레오에서 임의의 멀티채널 오디오 시스템으로 변환하는 업믹스 기술의 목적은 기존의 스테레오 입력이 가지고 있는 공간정보의 왜곡을 최소화 하면서 임의의 멀티채널 포맷을 통해서 실감나는 입체음향효과와 풍부한 음장감을 재생시키는 것이라고 할 수 있다. 본 논문의 채널 업믹스 알고리즘의 입력 신호로는 가장 범용적으로 사용되고 있는 2채널의 스테레오 신호를 가정하였고 출력 신호는 USC 10.2채널을 목표로 스테레오 신호를 멀티채널로 변환하는 업믹스 알고리즘을 기본 구조로 하고 있다^[8]. 식 (2)에 해당되는 패닝게인의 추정으로 입력 스테레오가 가지고 있는 평면상의 음원 방향각도(azimuth)를 추정할 수 있고, 추정된 패닝게인은 음원 재합성 과정에 해당되는 스테레오에서 임의의 배치의 멀티채널 신호로 재합성하는 과정에서 방향성 있는 음원을 만들어낼 수 있다. 패닝 알고리즘은 기본적인 레벨 패닝(amplitude panning)방법^[9]을 적용하였다.

본 논문에서는 앞선 연구의 기존 알고리즘이 가지고 있는 채널 포맷 변환 과정 중 입력신호가 가지고 있는 공간정보의 왜곡을 최소화 하는 전방채널 재합성 방안^[10]을 이용하고 이를 기반으로 효과적인 입체음향과 풍부한 음장 효과를 극대화하기 위한 업믹스 알고리즘을 제안한다. 제안된 알고리즘들은 목표로 하고 있는 USC 10.2채널 외에도 다양한 채널 포맷과 채널 추가 과정에서도 이용될 수 있는 호환성을 지니고 있다. 식 (1)과 같은 스테레오 신호모델은 기존의 많은 채널변환 연구에서 공통적으로 가정하고 있으며, 이 신호모델을 기반으로 단 구간 푸리에 변환(short time Fourier transform)에 의해 얻어진 주파수 신호를 청각특성을 기반으로 하는 서브밴드단위인 임계대역(critical band)^[11]으로 나누어서 입력 신호모델을 분석하였다. 본 논문의 구성은 2장에서 멀티채널 환경과 음원 분석에 대한 소개와 음원 분석 알고리즘을 설명하고 3장에서는 멀티채널 포맷 변환을 위한 음원 재합성 과정을 채널별로 제안하였다. 4장에서는 제안된 업믹스 알고리즘의 성능 평가를 위해 주관적 청취 평가 실험을 실시하였으며, 5장에서는 결론으로 맺음을 짓는다.

II. 멀티채널 환경과 음원 분석

멀티채널 오디오 시스템에서 각각의 스피커에 어떤 신호가 전달되어야 할 것인가에 대한 정의를 하기 위하여 먼저 멀티채널 레코딩 환경을 정의한다. 멀티채널 레코딩 시에 멀티 마이크로폰이 녹음실 안의 각각의 지정된 위치에 있다고 가정한다. ‘주요성분’을 녹음하기 위한 개개의 악기나 보컬 신호 근처에 위치한 몇 개의 마이크로폰과, ‘잔향(ambient)(잔향, 박수, 바람, 다양한 배경음) 성분’을 녹음하기 위한 몇 개의 마이크로폰이 적절히 위치한다. 녹음된 음원을 이용하여 믹싱할 때에는 주요성분은 전방채널, 잔향성분들은 후방채널로 전달된다고 가정한다^[12]. 잔향 성분은 입력 스테레오 신호가 가지고 있는 잔향, 배경음에 해당하는 음원이며, 공간적 특성을 나타내는 성분이다. 잔향성분은 써라운드 또는 후방채널에 전달되어 공간감을 극대화하고 풍부한 사운드를 만드는데 사용된다. 이러한 믹싱 방법은 청취자가 다양한 무대나, 콘서트 홀의 스테이지가 앞에 위치하는 것 같은 공간정보를 줄 수 있다. 스테레오에서 임의의 멀티채널 시스템으로의 업믹싱은 스테레오 신호로부터 잔향성분과 주요성분을 추출하고 이를 전방, 및 후방 채널로 전달하는 것을 기본으로 한다. 음원 분석 알고리즘은 방향감을 가지는 주요성분과, 잔향성분의 분리로 크게 구분할 수 있으며 음원 분리 알고리즘은 업믹싱 알고리즘의 전체 성능에 영향을 미칠 수 있는 중요한 과정이다. 다음 장에서는 스테레오 신호에서 주요성분, 잔향성분, 패닝게인의 분리, 분석 알고리즘을 설명하겠다.

2.1 패닝 게인 추정 알고리즘

식 (1)과 같이 스테레오 신호는 레벨차이를 가지는 공통된 음원과 잔향이나 기타 배경음의 합으로 이루어져 있다고 가정 하였다. 여기서 스테레오 좌·우의 신호가 레벨 차이만을 가지는 음원들로부터 구성되어 있다면 좌·우의 채널의 정규화된 상호 상관도(normalized cross correlation)는 ‘1’이 된다. 하지만 잔향 및 기타 다른 성분들이 함께 존재하면서 좌·우의 채널의 상호 상관도를 떨어뜨리게 된다. 본 연구에서는 이러한 특성을 이용하여, 주요성분을 분리하

기 위해 Principal Component Analysis(PCA)알고리즘을 이용하였다^[13]. PCA알고리즘을 이용하여 스테레오 신호의 고유 값(eigen value)과 고유 벡터(eigen vector)를 얻을 수 있고, 이를 이용해 스테레오의 패닝게인, 최종적으로 스테레오의 공간정보에 해당되는 방향 각도(θ)를 얻을 수 있다. 추정된 방향 각도(θ)는 스테레오 배치 시 좌·우 채널의 사이 공간 범위(청자를 중심으로 좌·우 30°)에 해당하는 스테레오 입력이 가지고 있는 평면 공간 정보이다.

2.2 주요성분 과 잔향성분 분리 알고리즘

스테레오 신호에서 주요성분의 분리를 위해서는 식 (3)의 신호 합성 모델을 정의한다^[13].

$$X = A \cdot S, \text{ where } A = \begin{bmatrix} a_L & 10 \\ a_R & 01 \end{bmatrix} \quad (3)$$

$X = [X_L X_R]^T$ 은 스테레오 모델이고 A 는 믹싱행렬(mixing matrix), $S = [S_L S_R]^T$ 는 스테레오 좌·우 채널의 레벨이 다른 주요성분과 잔향 및 배경음에 속하는 잔향성분을 나타낸다. A 의 성분들은 식 (1)에서 나타난 스테레오 신호모델에서 S 에 적용된 패닝게인이다. 스테레오 신호에서 주요 성분인 S 를 분리하기 위해서는 식 (3)에서 A 에 대한 역행렬(inverse matrix)이 필요하다. 하지만 수식 (3)에서 A 의 행렬은 under-determined, degenerate 경우에 해당되기 때문에 이러한 음원을 분리하기 위해서 유사역행렬(pseudo-inverse matrix)방법을 이용하였다. 이때 A 성분은 모두 양의 실수 값을 가지기 때문에 식 (4)와 같이 전치행렬(transpose matrix)로 대치될 수 있다.

$$A^+ = A^T (A^T A)^{-1}, A^+ = \frac{1}{a_L^2 + a_R^2 + 1} \begin{bmatrix} a_L & a_R \\ a_R^2 + 1 & -a_L a_R \\ -a_L a_R & a_L^2 + 1 \end{bmatrix} \quad (4)$$

그 결과 스테레오 입력신호에서 방향성을 가지는 음원 즉, 주요성분은 유사역행렬을 통해 식 (5)와 같이 분리해낼 수 있다.

$$\hat{S} = A^+ X \quad (5)$$

under-determined 경우에 해당하는 일반적인 스테레오 모델은 위의 음원 분리 알고리즘을 적용하더라도 음원을 분리할 수 없다. 그렇기 때문에 분리하고자 하는 음원 이외에 원치 않은 음원 및 주요성분들이 간섭(interface) 혹은 잡음의 형태로 분리된 음원에 존재하게 된다. 이런 문제점은 기존의 음원 알고리즘(source separation)에서도 계속 나타나고 있다. 하지만 본 연구와 같은 입체음향 신호처리에서는 음원 분리 성능의 극대화보다 방향정보의 유지가 우선이 되어야 한다. 따라서 제안된 음원 분리 알고리즘의 장점은 이전 방법과 달리, 음원의 에너지 정보에 영향 받지 않고 공간상에서의 방향정보만을 반영하여 음원을 분리할 수 있다는 점이다. 잔향 성분 분리 알고리즘에서도 주요성분 분리 알고리즘에서 나타나는 문제점이 나타나는데, 여러 음원 분리 알고리즘을 적용하더라도 완벽하게 원하는 음원만을 분리할 수 없고 원치 않은 음원들 또는 주요성분들이 포함된다. 멀티채널 오디오 포맷에서 주요성분이 상당량 포함되어 있는 잔향성분이 후방채널에서 재생된다면 전·후방에서 같은 주요성분이 재생되기 때문에 정확한 음상 정위가 일어날 수 없고 스테레오 입력의 공간정보가 왜곡될 수 있다. 특히 스테레오 입력 좌·우의 한쪽 채널로 심하게 치우쳐 패닝(hard panning)된 경우 좌·우 채널의 상관도가 매우 떨어지게 된다. 따라서 주요성분도 잔향성분으로 인지하게 되고 정확한 잔향성분 분리가 어렵게 되고 위와 같은 문제점이 발생하게 된다. 여러 음원 분리 알고리즘 중 스테레오 채널 간의 교차 채널 예측 방법(cross-channel prediction)은 상관도를 이용하여 주요잔향성분을 분리하는 것이 아니라 상호 채널간의 교차 추정방식을 이용하기 때문에 주요성분이 한쪽으로 치우쳐서 패닝된 경우에도 주요성분을 잘 분리해 낼 수 있다. 그렇기 때문에 본 논문에서는 음원 분리 성능과 채널 변환 성능을 극대화시키기 위해 상호 교차 채널 추정 방법을 잔향분리 알고리즘으로 사용하였다. 잔향성분 분리를 위해 사용되는 교차 채널 예측 방법 블록도는 그림 2에 나타나 있다. 기본적으로 스테레오 좌·우 채널에 대해서 상관도를 떨어뜨리는 잔향성분을 추출하기 위해 교차 필터의 계수 W_L , W_R 을 사용한다. 두 채널사이의 공통적인 성분을 제

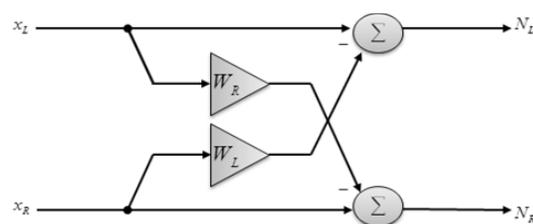


그림 2. 수교차 채널 예측 방법
Fig. 2. Cross-Channel Prediction.

거하고 상관도가 적은 좌·우 채널의 잔향성분을 남기는 것이다. W_L , W_R 은 적응형 알고리즘(adaptive algorithm)을 통해서 업데이트할 수 있고, 프레임별 자기 상관도와 상호 상관도 값을 구하여 교차 필터 계수인 W_L , W_R 을 갱신한다^[13-14]. 그림 2에서 N_L 과 N_R 는 좌·우 채널에서 분리된 잔향성분에 해당되며 이 잔향성분은 음원 재합성의 과정에서 써라운드 및 후방채널에 보내져 입체감이나 청자로 하여금 음장 감에 휩싸이는 듯한 포위감(envelopment)을 극대화시킬 수 있다.

III. 멀티채널 포맷 변환을 위한 신호 재합성

이전 장에서는 멀티채널 오디오 신호로의 변환과정 중 음원 분석과정을 통해 좌·우 두 채널이 공통적으로 가지고 있는 주요성분, 패닝계인, 잔향성분을 분리하기 위한 알고리즘들을 소개 하였다. 이 장에서는 변환하고자 하는 채널 포맷에서 스테레오 입력을 가지는 공간정보의 유지와 확장된 스피커 채널 환경에서 입체음향효과를 극대화시키기 위한 음원 재합성 과정에 대해서 제안한다. 변환하고자 하는 채널포맷은 그림 3에 나타나 있는 USC 10.2채널 써라운드 시스템으로 미국의 음향 시스템 업체인 THX와 USC대학이 공동 연구, 제안 하였다^[13]. 멀티채널 포맷으로는 가장 범용적으로 사용되는 ITU3/2 채널^[15]을 기반으로 "twice as good as 5.1"의 슬로건을 목표로 하는 USC 10.2채널 포맷은 ITU3/2 채널 포맷에 스피커를 추가하는 형태를 가지고 있다. ITU3/2 채널은 USC 10.2채널과 비교했을 때 가장 눈에 띄는 차이는 ITU3/2 채널은 모든 스피커가 동일 평면상에 배치된

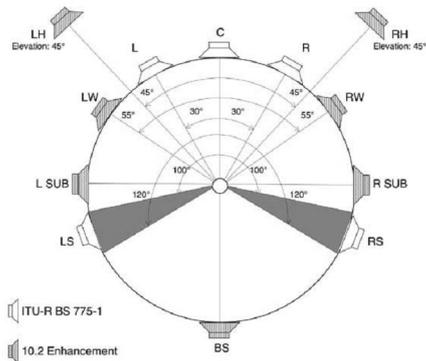


그림 3. ITU 3/2 및 USC 10.2 채널 시스템의 스피커 배치 [8]
 Fig. 3. ITU 3/2 and USC 10.2 Channel Configuration [8].

반면에, USC 10.2채널에서는 전방의 상향 45도 위치에 고도채널(LH, RH)을 배치한 것이다. 10.2 채널에서의 고도채널은 콘서트 홀이나 공연무대에서 상향 반사판과 같이 관객의 위치에서 위쪽에서 전해오는 천장 반사판과 같은 역할을 하고, 본 논문에서는 천장 초기 반사음(ceiling early reflection)의 이용을 제안한다. 천장 초기 반사음은 콘서트홀의 천장 반사판과 같은 역할을 만족시키고 풍성한 사운드를 전달한다. 이를 통해 청자의 몰입도를 높이고 청자가 음장 감에 둘러싸이는 듯한 포위감을 높일 수 있다. 또한 전방 채널과 후방 써라운드 채널 사이에 생기는 넓은 공간을 보상하고 음원의 방향성을 넓히기 위해 측면채널(WL, WR)을 추가로 배치하였다. 측면채널은 사운드의 방향성을 넓히고 풍성함의 극대화를 위해 고도채널과 마찬가지로 측면 초기 반사음(wide early reflection) 사용을 제안한다. 후방 써라운드 채널 간의 유격현상을 보상을 위해 후방 센터 채널(BS)이 추가되었고 저주파 사운드를 보강하기 위해 우퍼 채널(Lsub, Rsub)을 좌·우에 배치함으로써 보다 풍부한 저주파 사운드를 제공한다. ITU 3/2 채널은 평면상의 2차원적인 음향효과를 극대화시켰다면 USC 10.2채널에서는 추가적인 채널 배치를 통하여 3차원적인 음향효과를 만드는 것이 목적이다.

3.1 전방 채널의 음원 합성과 렌더링

스테레오에서 멀티채널 포맷으로 채널 변환을 할 때 가장 큰 변화는 전방 스피커 개수의 증가이다. 그림 3과 같이 ITU 3/2 채널은 전방에 세 개의 채널이 배

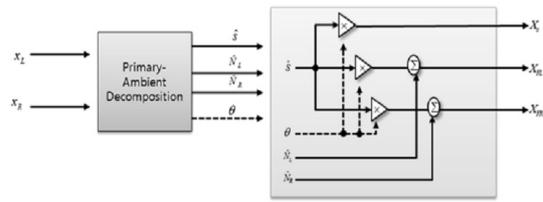


그림 4. 전방 채널(FL, FR, C)의 음원 합성과 렌더링 과정
 Fig. 4. Synthesis and Rendering of Front Channel.

치되어 있다. 일반적인 스테레오 배치(좌·우 30°)에 C채널이 정면에 추가 배치된 형태이다. 이때 C채널의 추가는 전방 음상의 안정성을 가져오고 영화나 음악에서의 가사나 영화의 대사 등의 명료도를 높이는 역할을 제공한다^[16]. 전방 3채널에 대한 패닝 알고리즘은 정면 C채널을 기본으로 좌·우로 패닝된 경우 C채널과 FL 또는 FR채널을 두 쌍의 스피커(pair-wise)를 묶어 패닝계인을 계산하는 Amplitude Panning 방법이 일반적이거나 pair-wise 패닝 방법은 전방 C채널을 기준으로 좌·우 두 채널 가운데 하나의 채널만을 이용하기 때문에 청자가 전방 음상을 지각할 때 정면 혹은 좌·우의 측면 스피커 쪽으로 음원이 몰리는 현상이나, 음상의 불연속적인 느낌을 초래하는 문제점이 있다. 따라서 전방의 세 채널에 패닝계인을 적용하여 세 채널을 모두 이용하면서 Vector Base Amplitude Panning(VBAP)과 같은 간단한 연산이 가능한 패닝 방법을 적용하였다^[9,17]. 다음의 그림 4는 전방 세 채널에 해당하는 음원 합성과 렌더링 과정을 보여주고 있다.

앞 장에서 각각의 알고리즘을 이용해 입력 스테레오 신호에서 주요성분과 잔향성분을 분리해 내고 분리된 각 성분들과 패닝계인을 전방 세 채널에 대해 적절히 합성, 렌더링 하면서 기존 스테레오 입력이 가지고 있는 공간적 정보를 왜곡하지 않고 C채널의 추가로 인해 전방 음상의 명료도와 음상에 대한 몰입감을 높일 수 있다.

다음은 고도채널과 측면채널의 음원 합성과정이 다. 그림 5와 6은 고도 채널과 측면 채널을 생성하는 블록도이다. 고도 채널과 측면 채널에 사용되는 음원은 가상의 룸 임펄스 응답(Room Impulse Response: RIR)을 Image Method^[18]를 이용해서 만들고 그림 7에 나타나 있는 것처럼 룸 임펄스 응답 중에 고도 및 측

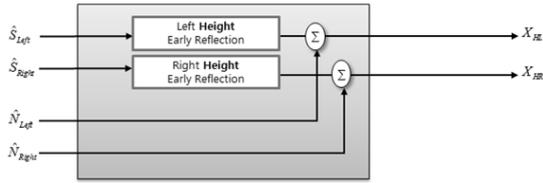


그림 5. 고도 채널(HL, HR) 생성 및 제어
Fig. 5. Height Channel Generation and Control.

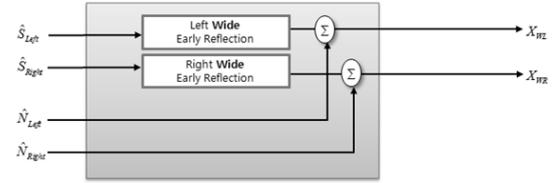
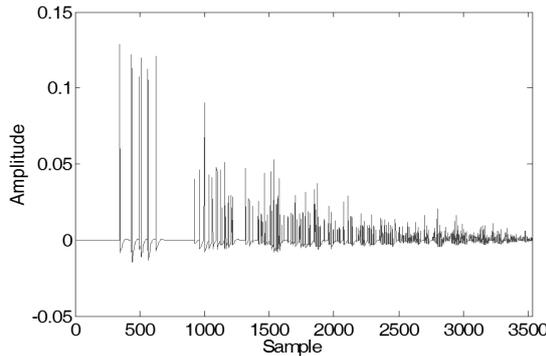
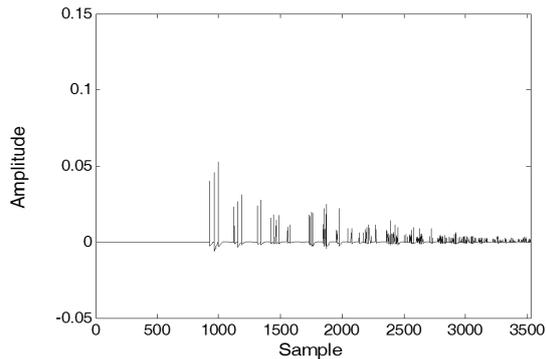


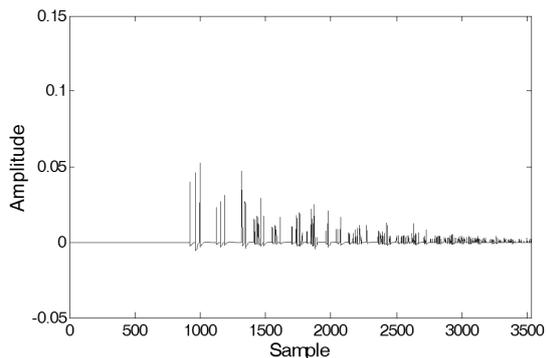
그림 6. 측면 채널(WL, WR) 생성 및 제어
Fig. 6. Wide Channel Generation and Control.



(a)



(b)



(c)

그림 7. 룸 임펄스 응답과, 측면 및 천장 임펄스 응답. (a) 전체 룸 임펄스 응답, (b) 천장 임펄스 응답, (c) 측면 임펄스 응답

Fig. 7. Room Impulse Response, Wide and Ceiling Impulse Response. (a) Total Room Impulse Response, (b) Ceiling Impulse Response, (c) Wide Impulse Response.

면에서 청자로 오는 초기 반사음(early reflection)을 분리한다. 각 채널 영역에 해당되는 초기 반사음들은 각각의 채널에 전달되어 청자의 몰입도를 높이고 포위감을 높이는 역할을 한다.

3.2 써라운드 및 후방 채널의 생성과 제어

ITU3/2채널은 본 논문에서 목표로 하고 있는 USC 10.2채널 뿐만 아니라 채널 업믹스 알고리즘에서는 써라운드 효과를 높이기 위해 후방채널 SL, SR채널을 생성한다. 앞 장에서 소개한 잔향성분 분리 알고리즘을 이용해 입력 스테레오에서 잔향 성분만을 분리한 후에 청자로부터 후방에 해당되는 채널에서 잔향 성분이 재생 되게 된다. 전방 채널이 스테레오 입력 신호가 가지고 있는 공간감을 유지, 재현 하였다면 후방 채널에서는 잔향성분을 통해 써라운드 효과와 청자가 음장감에 둘러싸이는 듯한 포위감을 극대화 시킴으로써 입체 음향 효과를 만들기 위한 채널 업믹스 알고리즘의 목적을 충족시키게 된다. 그림 8은 써라운드 채널생성에 대한 전반적인 블록도이다. 교차 채널 추정 방법 알고리즘은 분리된 잔향성분이외에 간섭신호나, 주요성분이 가장 많이 제거되는 면에서 가장 뛰어난 알고리즘으로 실험적으로 평가되었지만 주요성분이 전혀 포함되지 않고 완벽한 잔향성분만을 분리해 내는 것은 불가능한 일 이라고 할 수 있다. 그러므로 교차 채널 추정 방법 알고리즘을 이용해 분리해낸 잔향성분에서 소량의 주요성분이 새어 나오는 것은 불가피한 일이다. 후방 써라운드 채널은 그림 8처럼 입력 스테레오에서 분리된 잔향성분만을 이용해서 생성되는데, 분리된 잔향성분을 써라운드 채널 생성에 그대로 이용하면 전방채널에서 음원 합성 및 렌더링을 위해 사용한 주요성분이 전, 후방 채널에서 동시에 재생이 되게 된다. 채널 변환 알고리즘에서 가장 중요한 것은 스테레오 입력

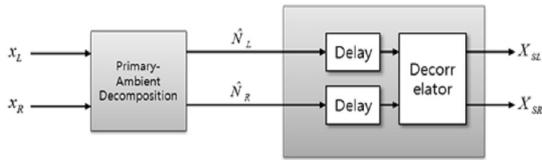


그림 8. 써라운드 채널(SL, SR)의 생성과 제어 과정
Fig. 8. Generation and Control of Surround Channel.

의 공간정보를 왜곡시키지 않는 범위에서 써라운드 효과를 극대화시키는 것인데, 위의 경우에는 입력 스테레오 기존 입력 스테레오 신호가 가지고 있는 공간정보를 왜곡시키는 경우이기 때문에 이런 현상을 없애기 위해 탭딜레이와 역상관기(decorrelator)를 통해 음상이 청자의 위치로 몰리게 되는 현상을 없앨 수 있다. 탭딜레이는 *summing localization*^[18]에 의해 전방 음원에 영향을 미치지 않는 약 1.5 ms 이상의 딜레이를 갖도록 하는데 탭딜레이는 *Haas effect*^[19-20] 현상이 일어나기 때문에 딜레이가 적용된 후방 채널보다 전방채널에 집중할 수 있게 된다. 탭딜레이 후 처리 과정으로 비상관기를 사용하였는데, 이는 전, 후방 채널의 상호 상관도를 떨어뜨려 청자는 전, 후방채널을 비교 했을 때 비상관(*uncorrelated*)된 소리로 인지하고 조금 더 넓은 음장감을 느낄 수 있다^[21]. USC 10.2채널에서는 ITU 3/2채널과 다르게 SL, SR 채널 이외에 후방 써라운드(BS)라고 하는 전방 센터 채널 반대편 후방 센터 채널이 추가되는데, 이는 상대적으로 SL, SR채널간의 유격이 크기 때문에 이를 보완하기 위해 추가 되었다. 이 후방 써라운드 채널은 그림 9에 나와 있는 것처럼 생성된 써라운드 채널에 대해서 다운믹스(downmix) 알고리즘^[4]에 의해서 생성된다. 써라운드 채널을 이용해서 후방 써라운드 채널을 생성하기 때문에 써라운드 채널과 후방 써라운드채널 또한 상호 상관도가 높을 수 있다. 그렇기 때문에 써라운드 채널을 생성 모듈처럼 다운믹스된 후방 써라운드채널에 역상관기를 적용하여 상관도를 떨어뜨리고 더 넓은 음장을 재생 한다.

3.3 공간정보 유지를 위한 전·후방 채널 이득 조절

멀티채널 오디오 시스템은 써라운드 효과와 입체감을 극대화시키기 위해서 후방 채널 SL, SR 신호를

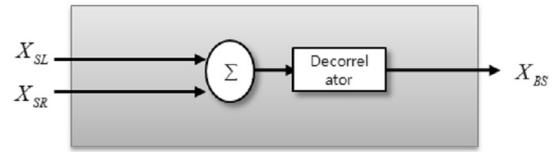


그림 9. 후방 채널(BS)의 생성과 제어 과정
Fig. 9. Generation and control of Rear Channel.

생성한다. 하지만 써라운드 채널은 스테레오 음원의 특성에 따라 써라운드의 효과음의 세기의 차별화를 두기 위해 잔향성분에 대한 레벨 조절이 필요하다^[22]. 기존에 진행되었던 연구에서는 식 (6)과 같이 후방 채널에 대한 이득을 구할 때 서브밴드 단위로 계산된 상호 상관도 ϕ_{LR} 을 이용하여 조절한다.

$$G_{SL} \approx G_{SR} = 1 - |\phi_{LR}| \quad (6)$$

하지만 좌·우의 25도 이상의 하드 패닝된 신호의 경우 패닝 신호임에도 불구하고 수식 (6)에 사용되는 상호 상관도인 ϕ_{LR} 은 '0'값에 가까워진다. 이런 경우 패닝 신호의 에너지 대부분이 후방 채널로 빠져 그대로 재생되는 문제가 발생되고 기존의 잔향성분을 분리하는 알고리즘이 가지는 한계로 인해 이 같은 하드 패닝된 신호의 경우 패닝 신호가 후방채널에 많이 남게 된다. 돌비(Dolby)에서는 이런 문제점을 지적하면서 기존의 상호 상관도를 이용한 방법 이외 별도의 편향 추정기(bias estimation)를 사용하였다^[23]. 다음의 수식 (7), (8), (9)은 전·후방 채널의 이득 조절을 위해 상호 상관도와 편향 추정기를 나타내는 식이다.

$$\rho_{LR}(m, b) = \frac{|E\{L(m, b) \cdot R(m, b)^T\}|}{\sqrt{E\{L(m, b) \cdot L(m, b)^T\} \cdot E\{R(m, b) \cdot R(m, b)^T\}}} \quad (7)$$

$$\phi_{LR}(m, b) = \frac{|E\{L(m, b) \cdot R(m, b)^T\}|}{\max\{E\{L(m, b) \cdot L(m, b)^T\}, E\{R(m, b) \cdot R(m, b)^T\}\}} \quad (8)$$

$$\beta(m, b) = \begin{cases} \rho_{LR}(m, b) & \phi_{LR} \geq \mu_0 \\ \rho_{LR}(m, b) + \frac{(\mu_0 - \phi_{LR}(m, b))}{\mu_0} & \phi_{LR} \leq \mu_0 \end{cases} \quad (9)$$

수식 (8)은 위에서 언급하였던 Dolby사에서 제안한

편향 추정기에 해당된다. 편향 추정기에 의해 신호의 하드 패닝 정도를 검출해서 수식 (10)과 같이 전방 및 후방 채널의 이득을 결정한다.

$$G_F(m,b) = \hat{\alpha}_0 + (1 - \hat{\alpha}_0) \sqrt{\beta(m,b)}, \quad G_B(m,b) = \sqrt{1 - G_F(m,b)} \quad (10)$$

수식 (10)과 같이 Dolby사의 전·후방 채널 게인 조절의 제안은 위에서 언급한 하드 패닝된 신호의 경우 후방 채널로 패닝된 신호가 빠져버리는 문제는 어느 정도 해결할 수 있었다. 하지만 여전히 문제는 하드 패닝된 정도를 파악하는 편향 추정기를 사용함으로써 추가적인 계산이 필요하고 또한 이 편향 추정기는 하드 임계점(hard threshold)를 사용하기 때문에 잔향 성분이 많이 섞여 있는 경우 편향 추정기는 정확한 역할을 하기 어렵다. 따라서 본 논문에서는 앞장에서 주요성분의 분리 분석을 위해 사용했던 PCA 알고리즘을 기반으로 고유 값의 비율인 γ 를 이용하여, 후방 채널의 이득을 조절한다.

$$0 \leq \gamma = \frac{\lambda_{\min}}{\lambda_{\max}} \leq 1 \quad (11)$$

식 (11)에 나와 있는 γ 값은 스테레오 두 채널에 잔향 성분만이 존재하게 될 경우 '1'에 가까워지고 그렇지 않을 경우 패닝 정도, 잔향성분의 에너지에 의해 0~1 사이의 값을 가질 수 있다. 이 값은 스테레오 두 채널에 잔향 성분이 존재하게 될 경우 '1'에 가까워지고 그렇지 않을 경우 패닝 정도, 잔향성분의 에너지에 의해 0~1 사이의 값을 가질 수 있고, 이 γ 값을 이용해 식 (12)와 같이 전·후방 채널의 이득을 얻을 수 있다.

$$G_B(n,k) = \sqrt{\gamma}, \quad G_F(n,k) = \alpha + (1 - \alpha) \sqrt{1 - G_B^2(n,k)} \quad (12)$$

제안된 전·후방 채널 게인은 γ 값을 이용하기 때문에 패닝 된 신호의 패닝 각도에 상관없이 일정한 이득을 얻을 수 있다. 또한 기존의 PCA 분석 시 사용하였던 파라미터를 그대로 사용하고 편향 추정기를 사용하지 않게 됨으로써 추가적인 계산을 필요로 하지 않으며 하드 패닝 된 신호에 대해 Soft한 이득을 구할

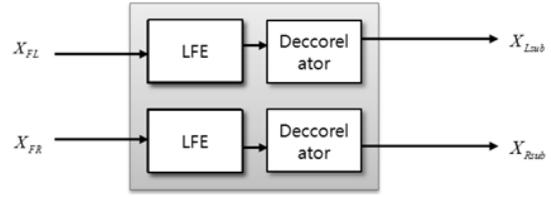


그림 10. 서브 우퍼 채널(Lsub, Rsub)의 생성과 제어
Fig. 10 Generation and Control of Subwoofer Channel.

수 있어 제안한 전·후방 채널 이득 조절은 간단한 계산 량으로 하드 패닝된 신호에 대한 여러 문제점을 개선해주는 장점이 있다.

3.4 서브우퍼채널(Lsub, Rsub)의 생성과 제어

서브우퍼채널(subwoofer)은 보통 저주파 대역을 증폭시켜주고^[24], 웅장감을 제공하는 채널이다. 서브우퍼 채널에서 사용되는 음원은 상대적으로 낮은 주파수의 음원이기 때문에 방향성을 가지지 않는다^[25]. 그렇기 때문에 5.1채널이나, 다른 임의의 멀티채널 오디오 포맷에서의 서브우퍼채널은 특정한 위치를 가지고 있지 않았다. 하지만 USC 10.2채널에서는 우퍼 채널의 좌·우에 배치하고 좌·우 스테레오 채널의 음원을 각각 처리한다. 멀티채널 오디오 시스템에서 하나의 서브우퍼채널보다 상호 상관도가 없는 두 개의 서브우퍼채널은 청자에게 더 넓은 웅장감을 전해줄 수 있다^[26]. 10.2 채널 업믹스 알고리즘 이용되는 서브우퍼 채널은 그림 10과 같이 나타낼 수 있으며 전방 좌·우 채널을 입력으로 200 Hz이하의 음원을 저역 통과 필터링(low pass filtering)한다.

그리고 두 채널간의 상호 상관도를 줄이기 위해 비상관기를 적용하는 일련의 과정을 통해서 만들어 진다.

IV. 주관적 음질 평가 실험

4.1 실험환경 및 과정

제안된 채널별 모듈을 이용하여 생성된 10.2채널을 기존의 상용화된 업믹싱 알고리즘들과 비교하는 주관적 음질 평가 실험을 실시하였다. 실험 방법으로는 10~30초의 길이를 가지는 각기 다른 특징을 가지는 5가지 샘플에 대해 스테레오 입력 음원을 기준

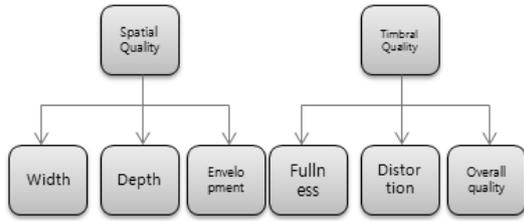


그림 11. 성능 평가에 사용된 비교 군
Fig. 11. Comparison groups used in performance evaluation.

으로 5가지의 업믹싱 방법을 그림 11에 나와 있는 것처럼 입체 공간적 성능(spatial quality)와 음색적 성능(timbral quality)으로 나누어 평가하도록 하였으며 피 실험자는 표 1에 나와 있는 기준과 같이 -3 ~ +3점의 상대적인 선호도를 평가하도록 하였다.^[27]

본 실험에서는 외부잡음이 차단된 녹음실에서 행 되었고, 청자와 스피커간의 거리는 1.5 m로 통일 하였다. 실험은 총 10명의 피 실험자가 참가하였고, 비교 업믹싱 방법으로는 Dolby PrologicII(5.1채널)^[1], Dolby PrologicIIX(7.1채널)^[1,28], 그리고 본 논문에서 제안한 채널별 모듈을 통해 생성된 5.1, 7.1, 10.2채널에 대해서 USC 10.2채널에 정의된 배치를 준수한 멀티채널 오디오 시스템들을 비교하였다. 실험에 쓰인 장비는 Yamaha의 NS-M125 5.1채널용 스피커와 rx-v 리시버 두 대를 이용했는데, 실험에 쓰인 리시버는 멀티채널 입력을 지원하기 때문에 앞에서 제안한 채널별 모듈을 통해 생성된 각 채널별 음원이 그대로 증폭되어 비교 성능 평가가 가능하다. 또한 비교 업믹싱 방법은 실험에 쓰인 리시버에서 제공되는 정보를 이용하여 업믹싱 기법간의 성능 평가가 가능하도록 하였다. 실험을 진행할 때에는 비교 대상이 되는 업믹싱 방법을 피 실험자에게 알려주지 않고 임의로 들려주는 비교 평가를 하게 하였다.

4.2 주관적 음질 평가 실험 결과

5가지 샘플을 이용한 주관적 음질 평가 실험의 결과가 그림 12에 나타나 있고 실험 결과에 색인된 업믹싱 알고리즘은 표 2에 설명되어 있다. 청취 실험 결과, 제안된 모든 업믹싱 기법들은 기준인 스테레오에 비해 모두 높은 성능의 점수를 받았다. 추가된 채널로 인해 스테레오가 가지고 있지 않은 음원들의

표 1. 주관적 평가 실험에 사용된 채점 등급
Table 1. The grading scale used in Subjective test.

Performance	Grade
Much Better	3
Slight Better	2
Better	1
About the Same	0
Worse	-1
Slight Worse	-2
Much Worse	-3

표 2. 주관적 평가 실험에 사용된 알고리즘
Table 2. Upmix algorithms used in subjective test.

Method	Description
5.1	Proposed 5.1ch
Dolby5.1	Dolby Prologic ii (5.1ch)
7.1	Proposed 7.1ch
Dolby7.1	Dolby Prologic ii x (7.1ch)
10.2	Proposed 10.2ch

입체감이 증가하고 음장감이 확대되면서 업믹싱 기법들은 전체적으로 스테레오보다 써라운드 효과 및 입체적인 음향효과가 향상되었음을 알 수 있다. 그러나 기준이 되는 스테레오와 업믹싱 기법들을 비교했을 때, 음질의 열화(degradation)를 비교한 청취 실험 결과를 보게 되면 0 ~ -3점의 성능 점수로 평가된 것을 볼 수 있는데, 이는 업믹싱 기법들은 기준이 되는 스테레오가 가지는 방향감이나, 공간감을 유지하는 것과 추가 채널생성 과정에서 음질의 열화를 평가하는 것으로 0 ~ -3점으로 평가 되었다. 음질의 열화의 성능 평가 결과를 보게 되면, Dolby5.1, Dolby7.1 보다 본 논문에서 제안된 채널 별 모듈로 생성된 업믹싱 기법들이 좋은 평가를 받은 것을 볼 수 있다. 이는 제안된 알고리즘인 음원 분석과정에서 실험적으로 가장 좋은 음원 분리 분석 알고리즘을 평가, 사용하고 스테레오가 가지는 공간감을 유지시키기 위한 음원 합성 과정으로 인해, 상대적으로 스테레오에 가까운 공간감을 유지하고 입체 음향효과를 극대화 시켰기 때문이라고 할 수 있다. 각 비교 군에 대해 5가지 샘플의 실험결과의 평균을 비교해 보았을 때 제안된 5.1채널과 Dolby5.1은 거의 거의 비슷한 성능을 보이고 있고, 제안된 7.1채널은 Dolby7.1과 비교해

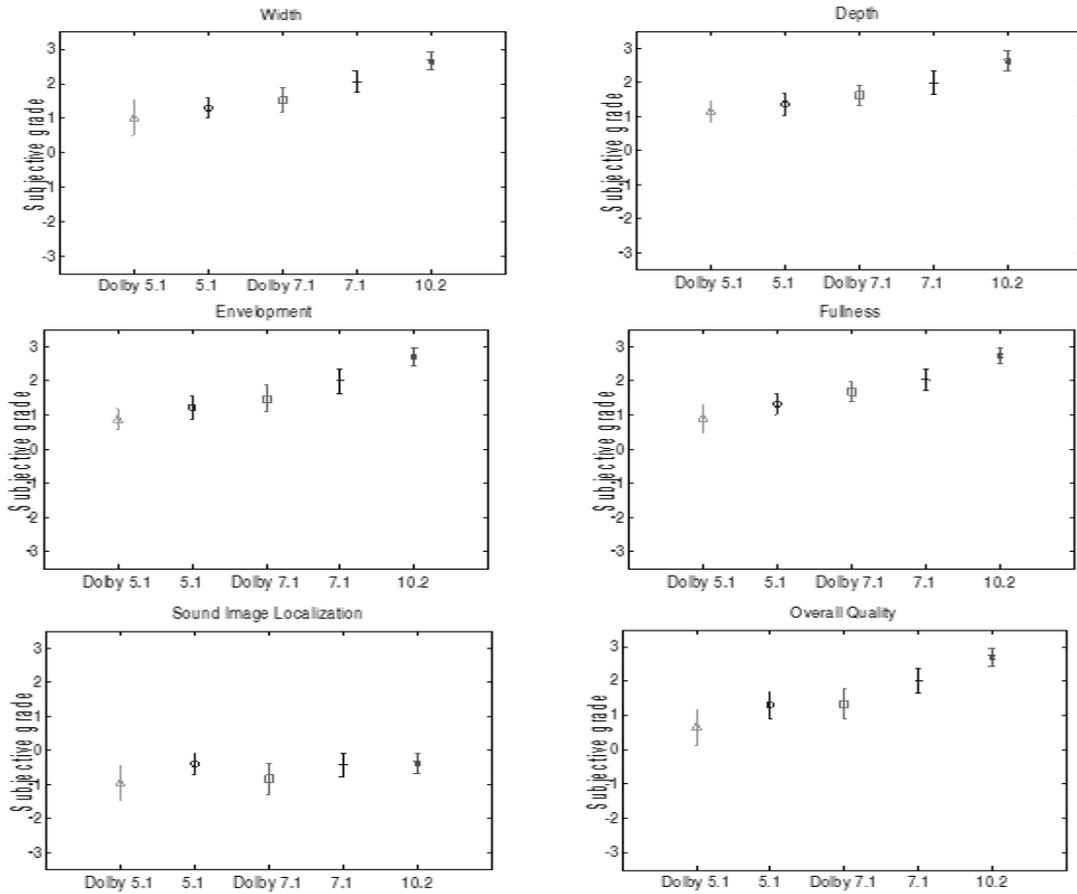


그림 12. 주관적 성능 평가
Fig. 12. Subjective test results.

상대적으로 조금 더 좋은 성능의 점수를 받았다. 특히 본 논문에서 제안한 채널 별 생성 모듈을 이용한 10.2채널 업믹싱 기법은 비교 대상이 되는 다른 업믹싱 기법들 보다 좋은 성능의 평가를 받았다. 이는 7.1채널에서 10.2채널로 채널 확장과정에서 고도채널 및 측면채널, 후방채널과 상호 상관도가 없는 서브우퍼를 추가 사용함으로써 2차원적인 음향효과와 더불어 3차원적인 음향효과까지 극대화시켰음을 알 수 있다.

V. 결 론

본 논문에서는 스테레오를 이용해 멀티채널 포맷 변환을 위한 업믹싱 기법을 제안하고 있다. 음원 재합성 과정에서는 본 논문에서 제안하고 있는 채널별 모듈을 이용해 실험적으로 스테레오와 기존의 업믹

싱 기법들을 주관적 음질 평가를 통해 비교하였다. 그 결과, 본 논문에서 제안된 채널 모듈을 이용한 업믹싱 기법은 현재 가장 많이 사용되고 있는 Dolby PrologicII, Dolby PrologicIIX에 비해 기존의 스테레오가 가지고 있는 공간정보의 유지와 전체적인 음질 면에서 좋은 평가를 받았으며, 입체 음향적 측면에서도 좋은 평가의 점수를 받았다. 특히 본 논문에서 최종적으로 제안하고 있는 10.2채널 업믹싱 기법은 5.1, 7.1채널과 비교해서 크게 향상된 성능을 보이는 것으로 보여 졌다.

감사의 글

이 논문은 2011년도 정부(지식경제부)의 재원으로 산업원천기반구축개발사업의 지원을 받아 수행된 연구임(No. 10037244).

참고문헌

1. Dolby Laboratories, *Dolby Surround Prologic II Decoder*, Principles of Operation., http://www.dolby.com/uploaded/Files/zz_Shared_Assets/English_PDFs/Professional/209_Dolby_Surround_Pro_Logic_II_Decoder_Principles_of_Operation.pdf
2. DTS Laboratories, *An Overview of DTS NEO:6 Multi-Channel*, <http://www.dts.com/media/upload/pdfs/DTS%20Neo6%20Overview.pdf>, 2010
3. R. Irwan and R. M. Aarts, "Two-to-Five Channel Sound Processing," *J.Audio Eng. Soc.*, vol.50, no.11, pp. 914-926, 2002.
4. M. R. Bai, G.-Y. Shih, and J.-R. Hong, "Upmixing and downmixing two-channel stereo audio for consumer electronics," *IEEE Trans. on Consumer Electronics*, vol. 53, no. 3, pp. 1011-1019, 2007.
5. C. Avendano and J.-M. Jot, "A frequency domain approach to multichannel upmix," *J.Audio Eng. Soc.*, vol. 52, no. 7/8, pp. 740-749, 2004.
6. C. Faller, "Multiple-loudspeaker playback of stereo signals," *J. Audio Eng. Soc.*, vol. 54, no. 11, pp. 1051-1064, 2006.
7. J. Breebaart and E. Schuijers, "Phantom Materialization: A Novel Method to Enhance Stereo Audio Reproduction on Headphones," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 8, pp. 1503-1511, 2008.
8. <http://en.wikipedia.org/wiki/10.2>
9. V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456-466, 1997.
10. S. W. Jeon, Y. C. Park, S. P. Lee, and D. H. Youn, "Robust representation of spatial sound in stereo-to-multichannel upmix," *AES 128th Conv.*, London, 2010.
11. J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization.*, Cambridge, MA : MIT Press, Cambridge, Massachusetts, USA, revised edition, 1997.
12. T. Holman. "Mixing the Sound," *Surround Magazine*, pp. 35-37, 2001.
13. S. W. Jeon, D. G. Hyun, J. G. Seo, Y. C. Park, and D. H. Youn, "Enhancement of principal to ambient energy ratio for PCA-based parametric audio coding," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 385-388, 2010.
14. S. Haykin, *Adaptive Filter Theory*, 4th. upper Saddle River, NJ: Prentice-Hall, 2002.
15. ITU-Recommendation ITU-R BS.775-1: Multichannel Stereophonic Sound System with and without Accompanying Picture, Geneva, 1992-1994.
16. Earl Vickers, "Frequency-Domain Two-to Three Channel Upmix for Center Channel Derivation and Speech Enhancement," *AES 127th Convention*, no. 7917, 2009.
17. S-W. Jeon, Y-C. Park, S-P. Lee, and D-H. Youn "Virtual Source Panning using Multiple-Wise Vector Base in the Multispeaker Stereo Format," in *Proc in 19th European Signal Processing Conference*, pp. 1337-1341, 2011.
18. J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943-950, 1979.
19. R.Y. Litovsky and H. S. Colburn, "The precedence effect," *J. Acoustic. Soc. Am.*, vol. 106, no. 4, pp. 277-303, 1999.
20. Rice, JJ. May BJ, Spirou GA, and Young, ED. "Pinna-based spectral cues for sound localization in cat," *Hearing Res*, pp. 132-152, 1992.
21. G.S. Kendall, "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery," *Computer Music Journal.*, vol. 19, no. 4, pp. 71-87, 1995.
22. 전세운, 박영철, 이석필, 윤대희, "다채널 포맷 변환과 공간적인 입체 음향 정보의 효과적인 유지에 대한 연구," *전자공학회 2010년도 하계종합학술발표회 논문집*, 2010.
23. M.S. Vonton, M. F. Davis, and C. Q. Robinson, "Signal models and Upmixing Techniques For Generating Multichannel Audio," *AES 127th International Conference*, no. 7917, 2009.
24. <http://en.wikipedia.org/wiki/Subwoofer>
25. J. Borenus, "Perceptibility of direction and time delay errors in subwoofer reproduction," presented at the *AES 79th Convention*, no. 2290, 1985.
26. W. Martens, "The Impact of Decorrelated Low Frequency Reproduction on Auditory Spatial Imagery : Are Two Subwoofers Better than One?" *AES 16th International Conference*, pp. 67-77, 1999.
27. ITU-R BS.562.3, "Subjective assessment of sound quality", *International Telecommunications Union*, Geneva, Switzerland, 1990.
28. http://en.wikipedia.org/wiki/Dolby_Pro_Logic#Dolby_Pro_Logic_IIx

저자 약력

▶ 최 선 웅(Sunwoong Choi)

2010년: 강원대학교 전기전자공학과(학사)
 2010년~현재: 연세대학교 전기전자공학과(석사과정)
 <관심분야> 오디오 신호처리, 3D 오디오, 멀티채널 사운드 시스템

▶ **현 동 일(Dong-il Hyun)**

2005년: 연세대학교 기계전자공학과(학사)
 2007년: 연세대학교 전자공학과(석사)
 2007년~현재: 연세대학교 전자공학과(박사과정)
 <관심분야> 디지털 신호처리, 오디오 신호처리, 음성 신호처리, 적응신호처리

▶ **이 석 필(Seok-Pil Lee)**

1990년: 연세대학교 전기공학과(학사)
 1992년: 연세대학교 전기공학과(석사)
 1997년: 연세대학교 전기공학과(박사)
 2009년 ~ 현재: 전자부품연구원(KETI) 디지털미디어 연구센터 센터장
 <관심분야> 디지털방송통신융합시스템, 멀티미디어 신호처리

▶ **박 영 철(Young-cheol Park)**

1986년: 연세대학교 전자공학과(학사)
 1988년: 연세대학교 전자공학과(석사)
 1993년: 연세대학교 전자공학과(박사)
 2002년 ~ 현재: 연세대학교 컴퓨터정보통신공학부 교수
 <관심분야> 디지털 신호처리, 오디오 신호처리, 음성 신호처리, 적응신호처리

▶ **윤 대 희(Dae Hee Youn)**

1977년: 연세대학교 전자공학과(학사)
 1979년: Kansas State University(석사)
 1982년: Kansas State University(박사)
 1985년 ~ 현재: 연세대학교 전기전자공학과 교수
 <관심분야> 디지털 신호처리, 적응 신호처리, 음성 신호처리, 오디오 신호처리