

## 다시점 방송 서비스를 위한 사용자 위치추적 기반 다채널 오디오 패닝 시스템 구현

### Implementation of a Person Tracking Based Multi-channel Audio Panning System for Multi-view Broadcasting Services

김용국, Yong Guk Kim\*, 양종열, Jong Yeol Yang\*\*, 이영한, Young Han Lee\*\*, 김홍국, Hong Kook Kim\*\*\*

**요약** 본 논문에서는 다시점 방송 서비스(multi-view broadcasting service)를 위한 사용자 위치추적 기반 다채널 오디오 패닝 시스템을 제안한다. 다시점 방송이란 각기 다른 시점에 해당하는 카메라들을 이용하여 촬영한 영상을 렌더링 해주는 방송 서비스이며 다채널 오디오 패닝 기법은 이러한 다시점 방송의 실감효과를 증대시키기 위한 오디오 렌더링에 필요한 기술이다. 이런 실감 오디오 기술을 다시점 방송 서비스에 적용하기 위해서 사용자의 시점정보를 얻을 수 있는 사용자 추적 기술 역시 필요하다. 이를 위해서 본 논문에서 제안한 기술은 크게 두 가지로 나눌 수 있다. 첫째로, 초음파 위성 및 수신기를 이용하는 사용자 추적 기술을 제안한다. 제안한 사용자 추적 기술을 이용해 약 10 mm 근방의 높은 해상도와 약 0.15초의 빠른 간격으로 사용자의 위치정보를 얻을 수 있었다. 둘째로, 공간 파라미터 기반 다채널 오디오 패닝 기술을 제안한다. 제안한 오디오 패닝 기법은 MPEG Surround로 압축된 오디오에 대해 MPEG Surround의 공간 파라미터를 이용하여 복호화 시 패닝된 오디오를 얻는 방법이다. 제안된 방법의 성능을 평가하기 위해 음질 측면과 음상 정위 측면에서 성능을 평가하였다. 음질 측정은 MUSHRA 방법으로 이루어졌으며, 제안된 방식은 약 90점에 해당하는 MUSHRA 점수로 평가되었다. 또한, 음상 정위 성능은 디미헤드를 이용하여 각 패닝 오디오에 대한 각도별 주파수응답을 측정하여 방사패턴을 분석하였다. 그 결과 본 논문에서 제안된 방식은 기존 파라미터 기반 오디오 패닝 방식에 비해 향상된 음질을 보임과 동시에 음상이 올바른 위치에 정위됨을 보일 수 있었다. 마지막으로 제안된 사용자 추적 기법 및 오디오 패닝 기법을 다시점 영상 재생 시스템과 결합시킴으로써, 사용자 추적 기반 다시점 방송 시스템의 프로토타입을 구현하였다.

**Abstract** In this paper, we propose a person tracking based multi-channel audio panning system for multi-view broadcasting services. Multi-view broadcasting is to render the video sequences that are captured from a set of cameras based on different viewpoints, and multi-channel audio panning techniques are necessary for audio rendering in these services. In order to apply such a realistic audio technique to this multi-view broadcasting service, person tracking techniques which are to estimate the position of users are also necessary. For these reasons, proposed methods are composed of two parts. The first part is a person tracking method by using ultrasonic satellites and receiver. We could obtain user's coordinates of high resolution and short duration about 10 mm and 150 ms. The second part is MPEG Surround parameter-based multi-channel audio panning method. It is a method to obtain panned multi-channel audio by controlling the MPEG Surround spatial parameters. A MUSHRA test is conducted to objectively evaluate the perceptual quality and measure localization performance using a dummy head. From the experiments, it is shown that the proposed method provides better perceptual quality and localization performance than the conventional parameter-based audio panning method. In addition, we implement the prototype of person tracking based multi-view broadcasting system by integrating proposed methods with multi-view display system.

**핵심어:** Audio Panning, Multi-view Broadcasting Service, 3DTV, Person Tracking

본 연구는 지식경제부 및 정보통신연구진흥원의 대학 IT연구센터 육성·지원사업의 연구결과로 수행되었음. (IITA-2008-C1090-0801-0017)

\*주저자 : 광주과학기술원 정보통신공학과 e-mail: bestkyg@gist.ac.kr

\*\*공동저자 : 광주과학기술원 정보통신공학과 e-mail: {jyyang, cpumaker}@gist.ac.kr

\*\*\*교신저자 : 광주과학기술원 정보통신공학과 교수; e-mail: hongkook@gist.ac.kr

## 1. 서론

최근 들어 멀티미디어 기술은 DMB와 같은 통신망 응용, MP3나 디지털 카메라와 같은 가전기기 등의 분야에서 급속히 발전해 왔다. 특히 영상분야에 있어서 고화질의 영상 재생을 위한 high definition television (HDTV)의 보급과 함께 초고화질 영상 재생을 위한 ultra-high definition television (UHDTV)에 대한 관심 역시 증대되고 있다 [1]. 이러한 고품질 기반의 영상재생에 관한 연구와 함께 보다 실감나고 사실적인 영상 콘텐츠를 시청자에게 서비스하기 위한 three-dimensional television (3DTV)에 대한 연구와 표준화 작업 역시 활발히 진행되고 있다 [2]. 3DTV 기술의 입체영상 재생을 위한 대표적인 기술로써 다시점 기반 방송 시스템이 있다. 다시점 기반 방송 시스템은 다시점 카메라 (multi-view camera)를 이용해 획득한 다시점 영상을 다시점 모니터 (multi-view monitor) 혹은 다시점 TV를 통해 재생함으로써 시청자가 다양한 시점 위치에 해당하는 입체 영상 서비스를 제공받을 수 있도록 하는 기술이다 [3].

다시점 방송에 있어서, 시청자의 위치에 따라 달라지는 입체영상 효과와 함께 재생되는 오디오 역시 시청자의 위치에 해당하는 지점의 오디오 효과를 시청자에게 다르게 제공하여 줌으로써 시청자가 느끼는 몰입도를 극대화할 수 있다. 이와 같은 오디오 서비스를 위하여 크게 두 가지 기술이 필요하다. 먼저 사용자의 시점 정보, 즉, 사용자가 현재 어느 곳에 위치해 있는지 위치를 추적하여 위치정보를 얻을 수 있는 기술이 필요하고, 다음으로 이 위치정보에 해당하는 오디오 효과를 위치 정보에 따라 렌더링 시켜줄 수 있는 기술이 필요하다.

다시점 방송 시스템에서 적용될 수 있는 위치 추적 기술로써, 특정 시청 공간상에서 사용자의 현재 위치를 예측해 낼 수 있는 다양한 추적 기술이 사용될 수 있다. 또한 방송 시스템에 적용될 수 있는 다채널 오디오의 효과적인 렌더링 방법 중의 하나로는 다채널 오디오 패닝 기법이 있다. 다채널 오디오 패닝 기법은 같은 신호가 스피커에서 다른 크기나 지연시간을 갖고 출력됨으로써 음원이 스피커 사이의 특정 지점에 음상이 정위되게 하는 기술이다 [4]. 이를 통해 마치 오디오의 전체채널이 회전된 듯한 효과를 느낄 수 있도록 해 준다.

본 논문에서는 위와 같은 다시점 방송 서비스를 위한 사용자 위치추적 기반 다채널 오디오 패닝 시스템을 제안하고 이를 구현한다. 사용자 위치추적을 위해 초음파기술을 이용한 위치추적 기술을 제안하였으며, 오디오 패닝을 위해 신호 기반 오디오 패닝 기법 및 파라미터 오디오 패닝 기법을 제안하였다. 본 논문에서 제안된 시스템은 개인용 다시점 방송 서비스에 효과적으로 사용될 수 있다.

본 논문의 구성은 다음과 같다. 2장에서 본 논문에서 다

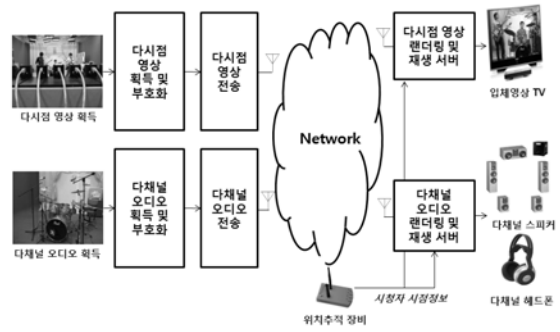


그림 1. 다시점 영상 및 다채널 오디오 방송 시스템

루고자 하는 다시점 방송시스템의 개요를 소개한다. 그리고 3장에서는 다시점 방송에서의 개인화된 오디오 서비스를 위해 사용자 추적을 통한 위치정보 획득 방법을 기술한다. 4장에서는 다채널 오디오를 패닝 하는 방법에 대해 자세히 살펴본 뒤, 5장에서는 이에 대한 성능평가 방법 및 그 결과에 대해 설명한다. 6장에서는 제안한 기법을 이용하여 구현한 시스템 프로토타입에 대해 논의하고, 마지막으로 7장에서 결론을 맺는다.

## 2. 다시점 영상 및 다채널 오디오 방송 시스템

다시점 영상 및 다채널 오디오를 이용한 방송 시스템의 구성은 그림 1과 같다. 그림 1에서 볼 수 있듯이 방송 시스템은 크게 다시점 영상 및 다채널 오디오 콘텐츠를 획득하는 부분과 이를 부호화하여 전송하는 부분, 전송받은 영상 및 오디오를 렌더링 및 재생하는 부분으로 나눌 수 있다.

다시점 영상 콘텐츠는 시청자의 하나의 시점에 해당하는 카메라 쌍의 array를 이용해 촬영되고, 이를 H.264와 같은 영상 부호화기를 이용해 부호화된다. 다채널 오디오 콘텐츠는 다채널 마이크 또는 마이크 array를 이용해 획득된 뒤, AAC 등 오디오 부호화기를 통해 부호화된다.

각각의 영상 및 오디오 콘텐츠는 network을 통해 각 시청자의 가정으로 전송이 되며, 전송된 다시점 영상 콘텐츠는 시청자 시점기반 렌더링 처리 과정을 거쳐 입체영상 TV를 통해 재생된다. 이와 마찬가지로 다채널 오디오 콘텐츠 역시 시청자의 시청 시점에 따라 렌더링 처리를 통해 5.1채널 및 7.1채널 등과 같은 다채널 스피커 또는 다채널 헤드폰을 통해 재생되게 된다.

이와 같은 다시점 기반의 방송 시스템에서는 1장에서 언급한 바와 같이 시청자가 느끼는 실감효과를 증대시키기 위하여 시점에 따른 영상 및 오디오의 렌더링 기술이 필요하다. 다시 말해, 다시점 TV를 이용하는 경우 시청자와 영상 콘텐츠간의 상대적인 각도에 따라 달라지는 입체영상 효과를 얻을 수 있다. 하지만 입체영상효과와 함께 시청자가 느

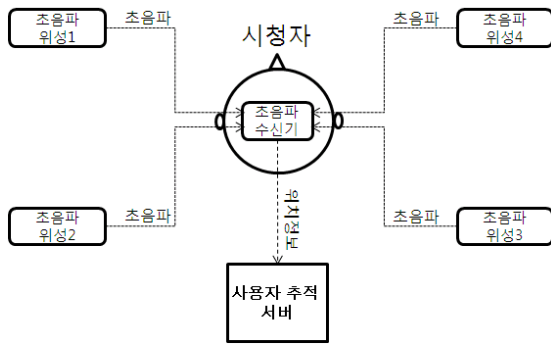


그림 2. 시점 정보 획득을 위한 사용자 위치 추적 과정

끼는 실감성 및 몰입도를 높이기 위하여 시청자의 시점에 따라 오디오 효과를 다르게 해주는 오디오 렌더링 기술이 필요하다.

오디오 렌더링 기술을 위해서 시청자의 시점에 따라 재생되는 오디오의 전체채널을 상대적으로 회전시켜주는 오디오 패닝 기술을 적용할 수 있다. 시청자가 다시점 영상 콘텐츠를 시청하면서 오른쪽 또는 왼쪽으로 이동하는 경우 재생되는 오디오 역시 전체 채널을 상대적으로 오른쪽 또는 왼쪽으로 이동한 위치에 맞게 패닝시켜 줌으로써 재생되는 오디오 콘텐츠로부터 느끼게 되는 실감효과가 증대된다. 위와 같은 오디오 패닝 기술을 다시점 방송 서비스에 적용하기 위하여 시청자의 현재 시점정보를 얻을 수 있는 위치추적 기술 역시 필요하다. 각 시청자의 배치된 위치추적 장비를 통해 현재 시청자가 위치해 있는 위치정보를 얻고, 얻은 위치정보를 영상 렌더링 장치 또는 오디오 렌더링 장치로 전달해 줌으로써 보다 사실적인 사용자 시점기반 영상 및 오디오 렌더링 처리가 가능해진다.

### 3. 시점 정보 획득을 위한 사용자 위치추적 방법

본 장에서는 본 논문에서 구현한 시스템에서 사용자 시점 정보 획득 방법에 대해 자세히 기술하도록 한다. 1장과 2장에서 언급한 바와 같이, 시청자에게 각기 다른 오디오 효과를 제공하기 위해서는, 각 시청자가 현재 다시점 영상이 재생되는 다시점 TV로부터 어디에, 얼마만큼 떨어져 있는지를 알아야 한다. 이를 위해서 시청자의 위치를 추적하는 기술이 필요하다

사용자의 위치를 추적하는 기술으로써는 다양한 방법이 사용될 수 있다. 사용자 위치추적 기술은 크게 사용자의 눈이나 얼굴 등을 추적하는 비전(vision) 기반 위치추적 기술[5]과 사용자가 active badge나 RFID, 초음파 센서 등을 직접 소지하거나 혹은 이들을 탑재된 장비를 장착함으로써 사용자의 현재위치를 추적하는 센서 기반 위치추적 기술로 크게 나눌 수 있다[6,7]. 비전 기반 위치추적 기술의 경우 영상

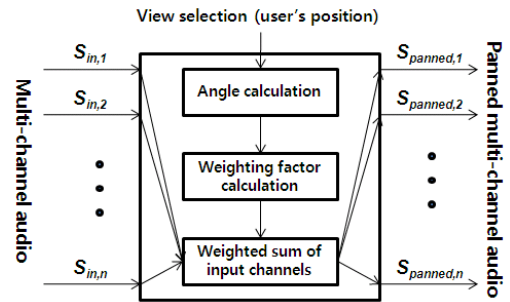


그림 3. 신호 기반 오디오 패닝

처리 방법에 주로 의존하여 사용자의 시선 위치 등을 추적하기 때문에 처리 속도도 늦고 많은 사용 제약을 가지는 문제점이 있는 반면, 센서 기반 위치 추적 방법은 이러한 처리 속도 면에서 장점이 있다. 특히 초음파를 이용하는 방법이 RFID 등과 같은 tag 기반 장비에 비해서는 비교적 높은 해상도를, 그리고 active badge의 방법에 비해서는 저렴하고 다양한 환경에 설치가 용이하다는 장점이 있다. 따라서 본 논문에서는 초음파 위성과 수신기를 이용하여 사용자의 현재 위치를 추적하는 방법을 사용하였다.

본 논문에서 사용된 위치추적 장비는 한국 LPS 사의 USAT[7]를 사용하였으며 위치추적 과정은 그림 2와 같다. 그림 2에서 볼 수 있듯이, 먼저 특정 공간의 천정에 위치한 4개의 초음파 위성에서 발생된 초음파 신호를 시청자의 헤드폰이나 옷 등에 부착된 초음파 수신기에서 수신한다. 수신된 각 위성까지의 거리 데이터는 블루투스 통신을 통해 사용자 추적 서버로 전달되고, 사용자 추적 서버에서 삼각측량을 이용하여 현재 시청자의 위치를 예측한다.

### 4. 다채널 오디오 패닝 방법

다채널 오디오 패닝 기법은 다채널 오디오 신호를 직접 이용하는 신호기반 오디오 패닝과 MPEG Surround[8]와 같은 부호화기를 이용하여 음원에 대한 공간 파라미터 영역에서 오디오를 패닝 하는 방법으로 크게 나눌 수 있다. 각각의 방법에 대해서 아래의 4.1절과 4.2절에서 자세히 기술하도록 한다.

#### 4.1 신호기반 오디오 패닝

신호기반 오디오 패닝 기법은 각 다채널 입력샘플에 대하여 오디오 패닝 이득의 적용을 통해 오디오 패닝 효과를 적용하는 방법으로써 기본 구조는 그림 3과 같다. 그림 3에서 보는 바와 같이 다채널 오디오 신호와 함께 3장에서 기술한 방법으로부터 얻어진 사용자의 시점정보에 해당하는 사용자의 현재 좌표로부터 오디오 패닝 각도를 계산한다. 그리고 나서, 오디오 패닝 각도에 따라 각 스피커의 가중치

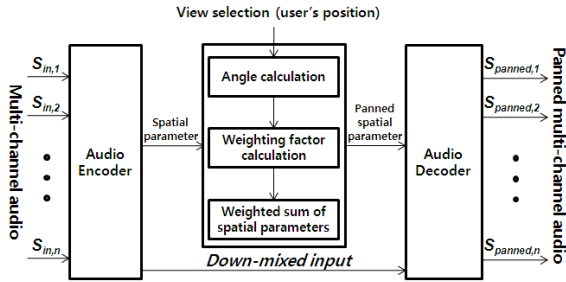


그림 4. 파라미터 기반 오디오 패닝

(weighting factor)를 계산, 이를 각 채널 입력 샘플에 적용하여 패닝된 오디오를 얻게 된다.

본 논문에서 제안하는 신호기반 오디오 패닝을 적용하는 세부 과정은 다음과 같다 [9]. 먼저 사용자의 위치좌표 ( $x, y$ )로부터 사용자와 각 스피커 위치에 대한 상대 각도  $\theta_s$  및 다시점 TV 중심으로부터의 사용자까지의 거리  $d$ 를 다음 식과 같이 계산할 수 있다.

$$\theta_s = \tan^{-1} \left( \frac{x_{spk_s} - x}{y_{spk_s} - y} \right) \text{ for } s = 1, \dots, 5 \quad (1)$$

$$d = \sqrt{(x - x_{screen})^2 + (y - y_{screen})^2}$$

여기서 ( $x_{spk}, y_{spk}$ ) 및 ( $x_{screen}, y_{screen}$ )은 스피커의 평면상의 좌표 및 TV의 평면상 좌표를 각각 의미한다.

다음으로 사용자와 각 스피커 위치에 대한 상대 각도  $\theta_s$ 를 통해 오디오 재생 스피커를 선택하게 되는데, 패닝 전 음원을 재생하던  $Speaker_s$ 가 패닝 후 재생되는  $Speaker_m$ 과  $Speaker_{m+1}$  사이에 위치할 때, 각 스피커의 가중치  $\alpha_{m,s}$  및  $\alpha_{m+1,s}$ 을 아래 식을 통해 계산하게 된다.

$$\alpha_{m,s} = 1 - \left| \frac{\theta_m - \theta_s}{\theta_{m+1} - \theta_m} \right| \text{ for } s = 1, \dots, 5 \quad (2)$$

$$\alpha_{m+1,s} = 1 - \left| \frac{\theta_{m+1} - \theta_s}{\theta_{m+1} - \theta_m} \right| \text{ for } s = 1, \dots, 5$$

여기서  $m$  및  $m+1$ 은 패닝 후 입력신호의 재생을 담당하는 두 스피커의 인덱스를 의미한다.

식 (2)에서 구한 각 스피커의 가중치를 각 채널 입력에 곱함으로써 시점 변화에 따른 패닝된 오디오 신호를 다음 식과 같이 구할 수 있다.

$$S_{panned,m} = \sum_{s=1}^5 \alpha_{m,s} S_{in,s} \text{ for } m = 1, \dots, 5 \quad (3)$$

#### 4.2 파라미터 기반 오디오 패닝

파라미터 기반 오디오 패닝 기법은 그림 4와 같이 다채널 오디오를 다운믹스 신호와 공간 파라미터로 부호화하고, 공

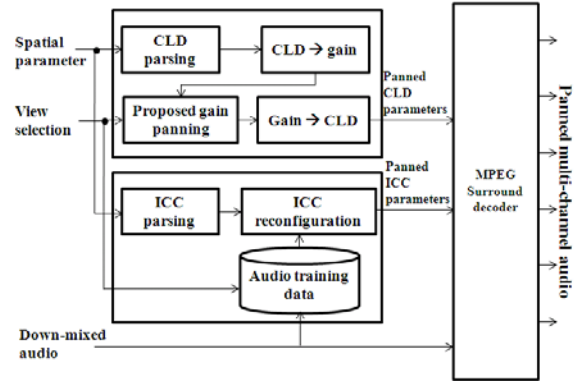


그림 5. 제안한 MPEG Surround 파라미터 기반 오디오 패닝

간 파라미터에 패닝 각도에 따른 가중치를 적용하는 구조로써 결과적으로 복호화기에서 복호화 시에 패닝된 오디오를 얻을 수 있게 된다. 이러한 공간 파라미터를 이용한 오디오 패닝 방법으로써 MPEG Surround의 공간 파라미터를 이용하는 오디오 패닝 기법 등이 제안되어 왔다 [10-12].

본 논문에서는 MPEG Surround의 부호화에 사용되는 대표적인 공간 파라미터인 채널 간 크기차이 파라미터(channel level difference)와 채널 간 상관도 파라미터(inter-channel correlation)를 이용한 패닝 기법을 제안한다. 제안 기법의 전체적인 구조는 그림 5와 같다. 그림 5에서 볼 수 있듯이 제안된 패닝 방법은 MPEG Surround로 부호화된 신호의 채널 간 크기차이 파라미터에 대하여 패닝각도에 따라 계산된 weighting을 적용하는 이득 패닝을 적용하는 방법과, 동시에 인공 신경망(artificial neural network)을 적용하여 채널 간 상관도 파라미터를 패닝된 오디오에 맞도록 변환하는 방법의 두 가지로 구성된다.

채널 간 크기차이(CLD) 파라미터를 이용한 오디오 패닝을 위해, MPEG Surround 복호화기에서 우선 CLD 파라미터를 분리해낸다. 그 다음, 분리된 CLD 파라미터를 아래의 식을 통해 선형적인 채널 이득으로 변환한다.

$$G_n = \frac{1}{\sqrt{1+c^2}}, G_{n+1} = \frac{c}{\sqrt{1+c^2}} \quad (4)$$

여기서  $n$ 은 채널 인덱스,  $G_n$ 과  $G_{n+1}$ 은 MPEG Surround 복호화기에서 one-to-two (OTT) 박스에서의 left 출력과 right 출력을 각각 의미한다.

채널 이득을 계산한 후 채널 이득에 오디오 패닝을 적용하는 세부 과정은 그림 6과 같으며 이는 4.1절에서 설명한 신호기반 오디오 패닝 과정과 유사하다. 먼저, 5.1 채널 스피커 상에 존재하는 5개의 스피커 쌍이 이루는 사이각과 입력된 패닝 각도를 비교하게 된다. 패닝 각도가 각 스피커 쌍의 사이각보다 큰 경우 먼저 출력되는 스피커 배열을 재배치를 하게 된다. 다음으로는 설정된 스피커 configuration과 비교하여 각 스피커 쌍이 이루는 사이각 중에 가장 작은 값

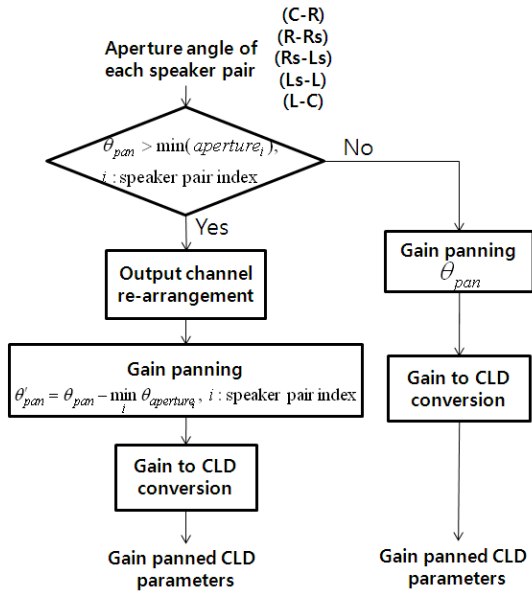


그림 6. 제한한 채널간 크기차이 파라미터 기반 오디오 패닝

에 해당하는 위치로 출력 채널을 배정한다. 그리고 입력된 패닝 각도  $\theta_{pan}$  으로부터 최소 스피커 쌍이 이루는 사이각을 뺀 잔여각도에 해당하는 만큼을 새로운 패닝 각도,  $\theta'_{pan}$ , 를 다음과 같은 식으로 얻을 수 있다.

$$\theta'_{pan} = \theta_{pan} - \min \theta_{aperture_i}; i: speaker pair index \quad (5)$$

오디오 패닝 각도가 최소 인접 스피커 사이각보다 작은 경우에는 기존 패닝 각도에 대하여 이득 패닝을 적용하게 된다. 이는 최소 인접 스피커가 이루는 사이각보다 오디오 패닝 각도가 크게 되는 경우 잘못된 위치에 음상이 정위되는 문제점을 해결하기 위함이다 [12].

위와 같이 구한  $\theta'_{pan}$  을 이용해 constant power panning을 적용하는 과정은 다음과 같다.

패닝 이득을 계산하기 위해 패닝 각도와 스피커간의 각도  $\theta_m$  을 각 스피커 쌍에 대하여 아래와 같이 구한다.

$$\theta_m = \frac{\theta'_{pan} - \theta_b}{\theta_{aperture} - \theta_b} \cdot \frac{\pi}{2} \quad (5)$$

여기서  $\theta_b$  는 스피커 쌍에서의 왼쪽 스피커의 각도를,  $\theta_{aperture}$  는 스피커 쌍의 사이각을 의미한다. 식 (5)와 같이 구한  $\theta_m$  으로부터 새로운 신호의 이득을 아래 식과 같이 구할 수 있다.

$$\begin{aligned} G_{n(pair_i)} &= G_n \cos(\theta_{m(pair_i)}) \\ G_{n+1(pair_i)} &= G_{n+1} \sin(\theta_{m(pair_i)}) \end{aligned} \quad (6)$$

여기서  $G_n$  과  $G_{n+1}$  은 패닝 전 오디오 채널의 이득과 패닝이 적용된 후의 오디오 채널 이득을 의미한다. 그리고  $i$  는 스피

표 1. 인공 신경망을 위한 configuration 상세

No. of input node	20
No. of hidden layer	1
No. of hidden node	40
No. of output node	20
Threshold for stopping the training	0.001
Maximum number of iteration for the training	1000
Learning rate	0.7

커 쌍의 인덱스를 의미한다.

이득 패닝이 적용된 후 각 채널 이득은 식 (7)을 통하여 다시 채널 간 크기 파라미터로 재변경을 하게 된다.

$$\begin{aligned} CLD_0^{panned} &= 10 \log_{10} \left( \sum \frac{pG_{ch}^2}{pG_C^2 + pG_{lfc}^2} \right) \\ CLD_1^{panned} &= 10 \log_{10} \left( \frac{pG_L^2 + pG_{Ls}^2}{pG_R^2 + pG_{Rs}^2} \right) \\ CLD_2^{panned} &= 20 \log_{10} \left( \frac{pG_C}{pG_{lfc}} \right) \\ CLD_3^{panned} &= 20 \log_{10} \left( \frac{pG_L}{pG_{Ls}} \right) \\ CLD_4^{panned} &= 20 \log_{10} \left( \frac{pG_R}{pG_{Rs}} \right) \end{aligned} \quad (7)$$

여기서  $i$  는 one-to-two (OTT) 박스 인덱스를  $CLD_i^{panned}$  ( $i=1, \dots, 4$ )는 패닝된 채널 간 크기차이 파라미터를 각각 의미한다. 그리고  $pG_X$  ( $X$ 는 R, L, C, Rs, Ls 중 하나)는 계산된 각 채널의 패닝 이득을 의미한다. 이와 같이 얻은 패닝된 CLD 파라미터를 기존의 CLD 파라미터와 대체하여 MPEG Surround 복호화를 수행하고, 그 결과 패닝된 다채널 오디오 신호를 얻게 된다.

공간 파라미터를 이용하여 다채널 오디오 패닝을 적용하는 데에 있어서 채널 간 크기 차이만을 고려하는 경우 패닝된 오디오의 품질 및 음상 정위 성능이 저하되게 되는 문제점이 있다 [10,11]. 이러한 문제점을 해결하기 위하여 채널 간 상관도(ICC) 파라미터 역시 패닝된 각도에 따라서 적절한 변환을 해주어야 한다.

본 논문에서는 패닝된 오디오에 대해 ICC 파라미터 변환을 위하여 인공 신경망을 이용한 기법을 제안한다.

이를 적용하기 위한 인공 신경망의 훈련을 위하여 먼저 패닝이 적용되지 않은 오디오 샘플로부터 채널 간 상관도 파라미터를 추출하여 input 데이터로 사용하고, 신호 기반 오디오 패닝을 적용한 후 패닝된 오디오에 대하여 역시 채널 간 상관도 파라미터를 추출하여 output 데이터로 사용하였다. 이를 위하여 약 10 분 길이에 해당하는 5.1 채널 오디오로부터 추출된 12,920 개 프레임에 대하여 훈련되었다. 표 1은 본 논문에서 사용된 인공 신경망의 configuration의

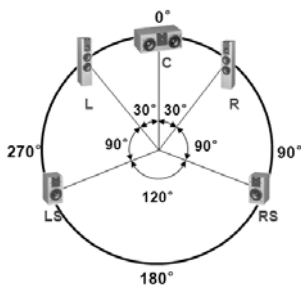


그림 7. 청취 실험을 위한 스피커 배치도

요약을 나타낸다.

본 장에서는 신호 및 파라미터 기반에서의 다채널 오디오를 패닝 하는 방법을 살펴보았다.

일반적으로 신호기반 오디오 패닝 기법은 패닝 효과의 적용에 의한 원음의 음질 및 음상정위 성능저하가 최소가 되는 장점이 있다. 하지만 오디오 패닝 적용을 위해 압축된 오디오의 복호화 과정이 선행되어야 한다는 점과, 전체 채널 샘플에 대해 오디오 패닝 이득 적용을 해야 하는 단점이 있다.

그에 비해 MPEG Surround를 이용하여 공간 파라미터 영역에서 오디오를 패닝하는 방법은 낮은 비트율로 표현할 수 있는 공간 파라미터에 패닝을 적용함으로써 적은 계산량을 사용하여 패닝된 다채널 오디오를 얻을 수 있다는 장점이 있다. 하지만, 공간 파라미터를 이용한 부호화 및 복호화 성능에 따라 복원 및 패닝된 오디오 품질에 영향을 끼칠 수 있다는 단점이 있다 [13].

## 5. 성능평가

본 논문에서 제안된 오디오 패닝 기법의 성능을 검증하기 위하여 오디오 품질 및 음상정위 성능을 각각 비교하였다. 성능을 검증하기 위하여 본 논문에서 제안된 다채널 오디오 패닝 기법을 5.1 채널 음원에 적용하여 원하는 위치에 음상을 정위 시켰다. 성능 측정은 특정 잡음이 존재하지 않은 밀폐된 녹음실에서 진행되었으며, 실험을 위한 스피커 배치환경은 그림 7과 같다.

### 5.1 오디오 품질평가

본 논문에서 제안한 기법의 품질성능을 검증하기 위하여 multiple stimuli with hidden reference and anchor (MUSHRA) 청취 실험[14]을 실시하였다. 본 실험을 위해 rock 음악, classic 음악 및 metal 음악 등 총 3가지 장르에 해당하는 음악을 채택하였으며, 청각적 장애가 없는 10명의 피실험자가 청취실험에 참여하였다.

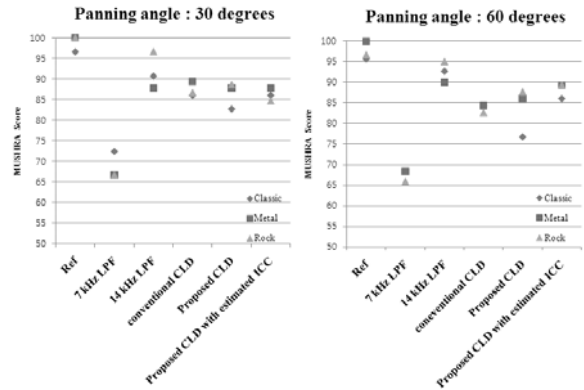


그림 8. 패닝각도 30° 및 60° 에 대한 MUSHRA 청취 실험 결과

MUSHRA 청취실험을 위하여 신호기반 오디오 패닝 기법을 적용한 오디오 음원을 reference로 하였으며 비교 음원은 아래와 같다.

- (1) Hidden reference
- (2) 7 kHz 저역통과 필터링된 앵커
- (3) 14 kHz 저역통과 필터링된 앵커
- (4) 기존 채널 간 크기차이 파라미터 기반 오디오 패닝 처리된 음원[10,11]
- (5) 제안한 채널 간 크기차이 파라미터 기반 오디오 패닝 처리된 음원
- (6) 제안한 채널 간 크기차이 파라미터 및 예측된 채널 간 상호상관도 파라미터 기반 오디오 패닝 처리된 음원

그림 8은 패닝 각도, 30°, 60° 에 대한 음질측정 성능결과를 각각 보여 준다. 그림 8에서 볼 수 있듯이 패닝 각도가 30° 인 경우, 즉 스피커 쌍이 이루는 사이각보다 작은 경우, 제안한 기법은 기존의 기법과 유사한 성능을 보임을 확인할 수 있었다. 하지만 스피커 쌍이 이루는 사이각보다 패닝 각도가 큰 경우, 그림 8과 같이 제안한 기법이 기존 기법보다 높은 오디오 품질 성능을 나타냄을 확인할 수 있었다. 이는 기존기법의 잘못된 음상정위로 인해 부자연스러운 음원의 재생으로 인한 결과로 분석될 수 있다.

### 5.2 음상정위 성능평가

본 논문에서 구현된 시스템의 성능 측정을 위해 center 채널에만 신호가 존재하는 5.1 채널 신호 오디오 신호를 실험용 콘텐츠로 정하여 방향성 패턴을 분석하는 방법[15]을 사용하여 음상정위 성능을 분석하였다. 패닝 각도는 패닝을 적용하지 않은 음원(0°)과 30°, 60° 만큼 패닝된 음원에 대하여 방향성 패턴을 분석하였다. 이는, 현실적으로 다시점 영상 콘텐츠를 시청하는 데 있어서 최적 시청 영역이 60° 도를 벗어나지 않으므로 실험을 위한 패닝 각도를 이와 같이 설정하였다. 그리고 방향성 패턴 측정을 위하여 각 패닝된

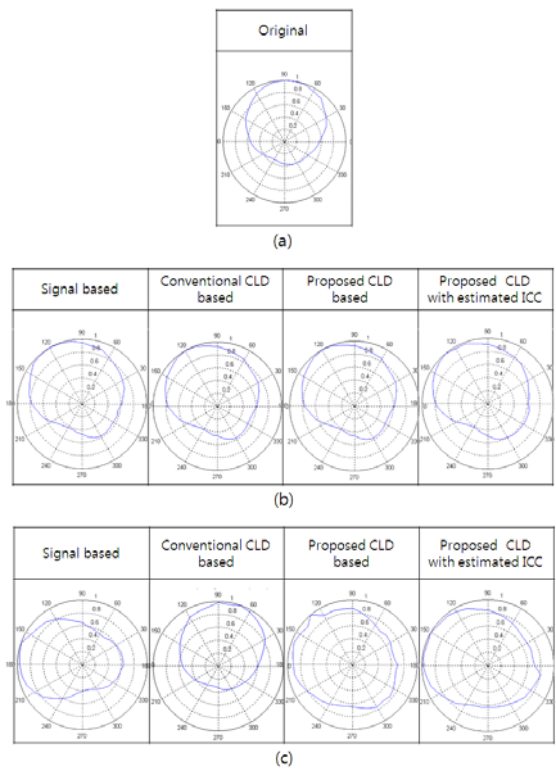


그림 9. 패닝각도 (a) 0°, (b) 30° 및 (c) 60° 에 대한 방향성 패턴

음원 재생 시 더미헤드를 10° 간격으로 회전시키면서 500 Hz 대역의 주파수 응답 크기를 측정하였다. 본 실험에 사용된 더미헤드는 Georg Neumann GMBH 사의 KU100[16] 모델을 사용하였다.

그림 9는 각 패닝 각도에 대하여 측정된 더미헤드의 왼쪽 귀에 해당하는 방향성 패턴을 나타낸다. 방향성 패턴으로부터 음상 정위 위치를 예측하기 위하여 전 방위각에 대해서 가장 큰 power가 측정되는 방위각에 음상이 정위된다고 가정하였다. 그림 9에서 볼 수 있듯이 오디오 패닝을 적용하지 않은 경우 방향성 패턴이 정면 방향(0°)에 해당하는, 즉 더미헤드가 시계방향으로 90° 만큼 회전한 경우에 power가 가장 크게 측정됨을 확인할 수 있다. 이와 마찬가지로 30° 만큼 패닝된 오디오의 경우 모든 방식이 패닝을 적용하지 않은 음원에 대해 약 30° 만큼 증가된 위치, 120° 에서 power가 가장 크게 측정됨을 확인할 수 있었다.

하지만 최소 스피커 사이각보다 큰 60° 만큼 패닝된 오디오의 경우에는 기존 패닝 기법으로 처리된 음원의 방향성 패턴이 잘못된 위치에서 power가 가장 크게 측정됨을 확인할 수 있었고, 그에 비해 제안된 기법으로 처리된 음원의 방향성 패턴의 경우에는 어느 정도 오차는 있지만 약 60° 만큼 증가된 위치, 120° 근방에서 power가 가장 크게 측정됨을 확인할 수 있었다. 이와 같은 방향성 패턴 분석실험 결과를 통해 각 채널 출력이 해당 시점에 따라 올바른 위치에

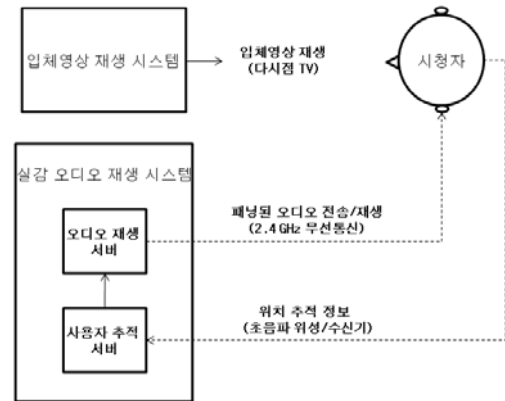


그림 10. 사용자 추적 기반 다시점 방송 시스템

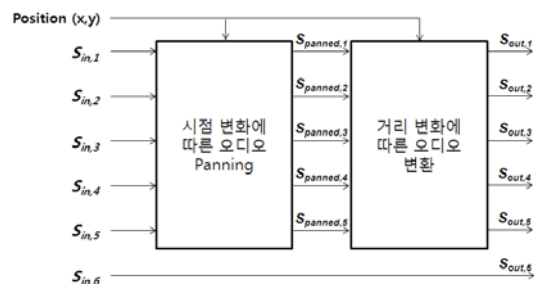


그림 11. 다채널 오디오 패닝

정위된 음원이 사용자의 헤드폰으로 전달됨을 확인할 수 있었다.

## 6. 다채널 오디오 패닝 시스템 구현

본 논문에서 구현한 사용자 추적 기반 다시점 방송 시스템의 전체구조는 그림 10과 같다. 그림에서 볼 수 있듯이 전체 시스템은 디코딩된 다시점 영상을 다시점 모니터를 통해 재생하는 입체영상 재생 시스템과 사용자 추적 및 패닝된 오디오 신호를 재생하는 실감 오디오 재생 시스템으로 나뉜다.

실감 오디오 재생 시스템은 3장에서 기술한 위치추적 장비를 이용하여 사용자 위치 좌표 정보를 얻게 되고, 이를 이용해 사용자 시점 기반 오디오 패닝을 적용한 후에 패닝된 오디오 신호를 무선 헤드폰을 이용하여 재생하게 된다. 프로토타입의 구현에 사용된 오디오 패닝 기법은 실시간 재생 및 패닝으로 인한 음질왜곡을 최소화시키기 위하여 신호기반 오디오 패닝 기법을 사용하였다.

프로토타입의 구현에 적용된 신호 기반 오디오 패닝은 그림 11과 같이 크게 사용자의 위치 좌표 및 다채널 오디오 신호를 입력 받아 시점 변화에 따라 오디오를 패닝하는 과정과 거리변화에 따른 오디오를 변환하는 과정으로 나눌 수 있다. 시점 변화에 따른 오디오 패닝은 4.1절에서 제안한 패닝 기법과 동일한 패닝 기법을 적용 하였으며, 그 결과로 얻

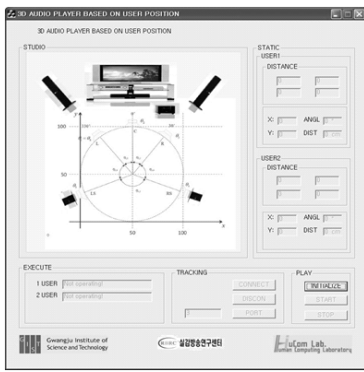


그림 12. 오디오 패닝 시뮬레이터 snapshot

어진 각각의 패닝된 오디오 채널 신호  $S_{panned, n}$ 에 대하여 거리 변화에 따른 오디오 변환 함수  $f(d)$ 를 식 (8)과 같이 적용하게 된다.

$$S_{out, n} = f(d) \cdot S_{panned, n} \text{ for } n = 1, \dots, 5 \quad (8)$$

여기서  $f(d)$ 는 inverse square law에 따라서 거리  $d$ 에 대한 재생되는 오디오의 출력 크기를 변화시켜주는 부분으로써 아래 식과 같이 구할 수 있다.

$$f(d) = \frac{1}{d} \quad (9)$$

식 (8) 및 식 (9)에서 볼 수 있듯이 입체영상 재생 디스플레이로 부터의 떨어진 거리,  $d$ 가 작을수록 오디오 출력 크기는 반비례 관계에 의해 커지게 된다.

본 논문에서 제안된 오디오 패닝 시스템의 실제 구현된 오디오 패닝 시뮬레이터의 snapshot은 그림 12와 같다. 그림 12에서 볼 수 있듯이 오디오 재생 시스템은 최대 두 명의 시청자까지 개인화된 오디오 서비스를 지원하도록 설정하였고, 실시간으로 각 사용자의 위치좌표를 표시함과 동시에 각 위치에 해당하는 오디오 패닝을 적용하여 무선 헤드폰으로 패닝된 오디오를 전송하도록 구현하였다.

특히 추적된 사용자의 현재 위치정보는 보다 자연스러운 패닝 효과 적용을 위하여 0.15초 주기로 위치추적 서버로 사용자의 현재좌표를 송신토록 하였으며, 위치추적 서버는 수신 받은 사용자의 현재좌표를 이용해 사용자와 다시점 모니터간의 상대적인 각도  $\theta$ 와 거리  $d$ 를 계산하여 오디오 재생 프로그램으로 전달하게 되고 오디오 재생 프로그램은 현재 재생되고 있는 5.1 채널 오디오에 그림 11과 같은 다채널 오디오 패닝을 적용한 후 2.4 GHz 무선 통신을 통해 무선 헤드폰으로 패닝된 오디오 신호를 전달하게 된다.

## 7. 결론

본 논문에서는 다시점 방송 서비스를 위한 사용자 위치추

적 기반 다채널 오디오 패닝 시스템을 구현하였다. 이를 위해 초음파 위성 및 수신기를 이용한 위치추적 기술을 제안하였고, 제안한 기술을 통해 사용자의 현재 위치 좌표를 추적해냄으로써 사용자의 시점 정보를 얻을 수 있었다. 또한 다채널 오디오 패닝을 위하여 파라미터 기반 오디오 패닝 기법을 제안하였고, MUSHRA 청취 실험 및 방향성 패턴분석을 통한 음상 정위 성능평가를 통해 제안된 기법이 기존 기법에 비해 높은 품질 성능 및 음상 정위 성능을 보임을 확인할 수 있었다.

## 참고문헌

- [1] 안충현, "UHD(Ultra High Definition) TV 기술개발 동향," *한국방송공학회지*, 제13권, 제1호, pp. 97-112, 2008년 3월.
- [2] 김규현, 안충현, 김성규, 유지상, "3D 기술 표준화 동향," *한국방송공학회지*, 제13권, 제1호, pp. 16-26, 2008년 3월.
- [3] K. -J. Oh, *et al.*, "Multi-view video and multi-channel audio broadcasting system," in *Proc. of IEEE 3DTV Conference 2007*, Kos Island, Greece, pp. 001-004, May 2007.
- [4] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 56, no. 6, pp. 456-466, June 1997.
- [5] M. L. Cascia, S. Sclaroff, and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 4, pp. 322-336, Apr. 2000.
- [6] R. Want and A. Hopper, "Active badges and personal interactive computing objects," *IEEE Trans. on Consumer Electronics*, vol. 38, no. 1, pp. 10-20, Feb. 1992.
- [7] [http://www.korealps.co.kr/bbs/usat\\_a105.html](http://www.korealps.co.kr/bbs/usat_a105.html).
- [8] ISO/IEC FDIS 23003-1:2006(E), *MPEG Audio Technologies-Part 1: MPEG Surround*, 2004.
- [9] Y. H. Lee, H. K. Kim, and M. Kim, "View-dependent multichannel audio processing for a multi-view camera system," in *Proc. of International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Busan, Korea, pp. 949-950, July 2007.
- [10] S. K. Baeck, J. I. Seo, I. S. Jang, and D. Y. Jang, "Multichannel sound scene control for MPEG Surround," in *Proc. of 29th AES International Conference*, Seoul, Korea, pp. 63-66, Sept. 2006.
- [11] S. K. Baeck, J. I. Seo, T. J. Lee, and D. Y. Jang, "Spatial Cue Based Sound Scene Control For MPEG Surround," in *Proc. of IEEE International Conference on Multimedia and Expo 2007*, pp. 1886-1889, Beijing, China, July 2007.
- [12] 김용국, 김홍국, "채널 간 크기 차이 파라미터를 이용한 MPEG Surround 다채널 오디오 패닝 기법," *2008년도 한국통신학회 추계종합학술발표회 논문집*, pp. 4A-2, 2008년 11월.
- [13] A. Brinker, E. Schuijers, and A. Oomen, "Parametric coding for high-quality audio," in *Proc. of 112th AES Convention*, Munich, Germany, preprint 5554, May 2002.
- [14] ITU-R Recommendation BS. 1534-1, *Method for the Subjective Assessment of Intermediate Quality Levels of Coding System*, Jan. 2003.
- [15] 김용국, 김홍국, 박창수, "방향성 패턴 분석을 통한 음상의 인지 방향 추정," *2008년도 한국통신학회 김춘추정이론연구회 하계워크샵 논문집*, pp. 37-40, 2008년 8월.
- [16] Georg Numann GmbH, *Product Information KU 100*, Germany, Nov. 2000.