

# 청취자 위치 기반 사운드 렌더링 기술

이기승 (건국대학교)

## I. 서론

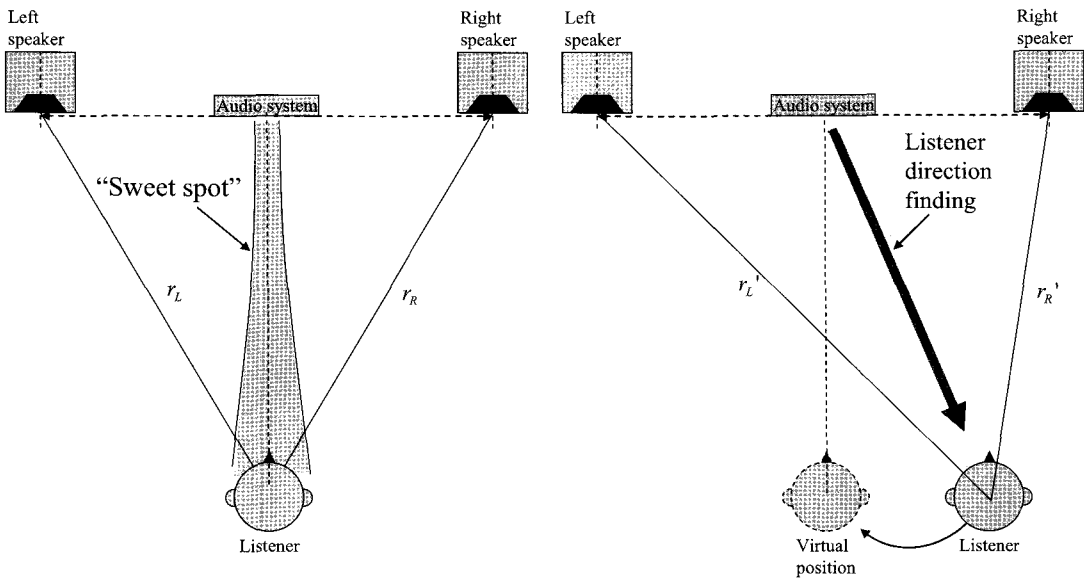
좌, 우 두 개의 스피커를 이용한 오디오 시스템에서 청취자에게 인지되는 스테레오감은 한정된 영역에 국한되어 나타난다. 이러한 영역을 “sweet spot” 이라 칭하며 좌, 우 스피커와 동일한 거리를 갖는 영역, 즉 두 스피커가 놓인 지점을 잇는 직선의 수직 이등분선 상에 존재한다<sup>[1]</sup>. 청취자가 sweet spot 으로부터 멀어지게 되면 청취자가 느끼는 음상 (sound image) 이 근접한 스피커에 가깝게 이동하는 “precedence effect” 가 발생하게 된다. 이는 청취자가 sweet spot에서 벗어나게 되면 좋은 스테레오감을 느낄 수 없음을 의미한다.

최근 디지털 신호 처리 기술의 발전에 따라 청취자가 sweet spot에서 벗어나더라도 좋은 스테레오감을 유지시킬 수 있는 오디오 재생 기법들이 제안되고 있다<sup>[2~4]</sup>. 이러한 기술은 <그림 1>과 같이 현재 청취자의 위치를 자동적으로 인식하고, 인식된 위치에서 sweet spot이 형성되도록 오디오 신호를 적응적으로 제어하는 방법을 통해 구현된다. 따라서 청취자 위치 기반 오디오

재생은 청취자의 위치를 자동으로 인식하는 기술과 위치에 따라 오디오 신호를 제어하는 기술을 필요로 한다.

청취자의 위치를 자동적으로 인식하는 기술은 영상 정보를 이용한 방법<sup>[2]</sup>, 적외선과 초음파와 같은 2차적인 신호를 이용하는 방법<sup>[3]</sup>, 그리고 청취자로부터 발생하는 신호를 이용하는 방법<sup>[4]</sup>들이 제안되고 있으며, 본 고에서는 이러한 방법들의 장, 단점을 살펴보고 오디오 제어 기술 관점에서 고려되어야 할 요인들에 대해 논의하기로 한다.

오디오 신호 제어 기술은 청취자가 인식하는 소리의 위치, 즉 음상을 제어하는 기술로 간주할 수 있으며, 이에 따라 두 개의 채널을 갖는 환경에서 음상을 제어하는 여러 기술들이 적용되고 있다. 대표적인 것으로 머리전달함수 (head-related transfer function, HRTF) 를 이용한 방법<sup>[3~4]</sup>이 있는데 이를 보다 간략화한 자유 공간 모델<sup>[5]</sup> (free field model) 등도 널리 사용되고 있다. 본 논문에서는 현재까지 발표된 여러 음상 제어 기술을 소개하며 향후 실용화되기 까지 해결되어야 할 여러 기술적인 문제에 대해 논의하고자 한다.



〈그림 1〉 스테레오 재생 환경에서 sweet spot 영역 (좌) 및 청취자 위치 추적을 통한 sweet spot 교정 (우)

## II. 사용자 위치 기반 오디오 시스템

두 개의 채널을 이용한 오디오 재생 시스템에서 청취자가 인지하는 소리는 위치에 따라 다르게 나타난다. 이는 <그림 1>에 제시된 바와 같이 청취자의 위치 (x,y) 에 따라 각 채널의 스피커와 청취자의 귀간의 거리가 달라지며 이에 따라 청취자에게 도달되는 시간 지연 및 소리의 레벨차가 달라지기 때문이다. 또한 오른쪽 채널의 소리가 왼쪽 귀에 들리고, 왼쪽 채널의 소리가 오른쪽 귀에 들리는 “crosstalk” 현상이 사용자의 위치에 따라 각기 다르게 나타난다<sup>[1]</sup>. 일반적으로 좋은 스테레오감을 느끼기 위해서는 각 채널의 소리가 양쪽 귀에 도달하는 시간과 레벨 차이가 동일하게 유지되는 sweet spot 영역에 청취자가 존재해야 하며, 따라서 기존의 오디오 시스템은 사용자가 sweet spot에서 음악을 청취할 것을 권장하고 있다. 사용자가 sweet spot에서 벗어나 음악을 감상하는 경우, 기존 오디오 시스템

에서는 밸런스 컨트롤, 즉 좌, 우 채널의 음량을 변화시킴으로써 스테레오감이 유지되도록 한다. 이와 같이 단순히 좌, 우 채널의 레벨만을 조절하는 경우 매우 제한적으로 스테레오감을 유지시킬 수 있으며, 청취자의 위치가 바뀔 때 마다 밸런스를 조절해야 한다는 번거로움이 따른다.

청취자 위치 기반 오디오 재생 시스템은 청취자의 위치를 자동적으로 인식하고, 인식된 위치에서 최적의 스테레오감을 느낄 수 있도록 오디오 신호를 처리하는 기술이 결합된 기기로 볼 수 있다. 이러한 오디오 시스템의 구현에 필요한 제반 기술에 대해 살펴보면 다음과 같다.

### 1. 청취자 위치 추적 기법

- 영상 데이터를 이용한 위치 추적<sup>[2]</sup>

THX (Tomlinson Holman eXperiment) 기술을 개발한 미국 University of Southern California 의 Tomlinson Holman 은 비디오 카

메리를 이용한 위치 추적 기법이 적용된 오디오 시스템을 개발하였다. 이 시스템은 소형의 스피커와 스피커간 거리가 비교적 짧은 (1m 이하) 탁상형 오디오 재생 환경에 적합하도록 개발되었다. 청취자의 위치 또한 오디오 시스템과 근접한 거리로서, 비디오 카메라를 통한 청취자의 인식이 용이한 환경이 사용되었다.

카메라를 통한 청취자 위치인식은 1초당 30프레임으로 구성된 동영상을 통해 이루어지며, 첫 단계로서 인접한 두 프레임간의 차이를 구하여 청취자의 변위를 찾는다. 다음으로 차 영상의 히스토그램을 구하여 청취자에 해당하는 부분만을 세그먼트로 분할하고, 다시 피부색에 해당하는 부분만을 추출한 후 얼굴과 귀 부분의 위치를 인식하는 일련의 과정을 통해 청취자의 위치를 인식하게 된다.

이와 같은 방법은 비디오 카메라가 탑재된 멀티미디어 재생환경에서 용이하게 구현될 수 있다는 장점이 있으나, 재생 환경의 조도와 같은 외부적 환경의 영향을 받을 수 있다는 단점이 있다. 또한 실시간 영상처리에 따른 계산량의 증가와 전용 하드웨어의 필요성 등이 문제가 될 수 있는데, 실제로 Tomlinson에 의해 구현된 시스템은 영상 처리를 위한 워크스테이션을 필요로 하였다. 그러나 최근에는 고속의 프로세서 및 제반 기술들이 발달하여 이에 대한 문제가 많이 해결되었다고 볼 수 있다.

#### • 초음파와 적외선을 이용한 위치 추적<sup>[3]</sup>

김 등은 가정용 디지털 텔레비전의 사운드 렌더링 기술의 사용자 위치 추적 방법으로 초음파와 적외선을 이용한 방법을 제안하였다. 이 방법에서는 리모트 컨트롤에 초음파 발생기와 적외선 발생기를 부착하여 사용자가 특정 버튼을 누

르면 두 신호가 동시에 발생하고, 텔레비전에서는 수신된 두 신호를 이용하여 청취자의 위치를 추정하도록 하였다.

청취자와 텔레비전간의 직선 거리는 초음파와 적외선 간의 시간 차이를 이용하였다. 즉, 적외선은 빛의 속도로 텔레비전에 전달되므로 시간 지연이 거의 없으며 초음파는 상대적으로 긴 시간 동안 전파되므로, 두 신호간의 시간 차이를 알면 청취자가 위치한 거리를 알 수 있다. 청취자와 텔레비전이 이루는 각도는 텔레비전의 좌, 우 양쪽에 설치된 초음파 센서를 이용하여, 좌, 우 센서의 신호간 시간 지연을 이용하여 측정하였다.

이러한 사용자 위치 추적 기법은 초음파와 적외선을 사용하므로 빛이나 소음과 같은 외부적인 환경에 영향을 덜 받으며, 기존의 리모트 컨트롤에 비교적 저렴한 비용으로 쉽게 탑재될 수 있다는 장점이 있다.

#### • 청취자로 부터의 신호를 이용한 위치 추적<sup>[4]</sup>

리모트 컨트롤을 이용하는 방법은 위치 추적시 항상 리모트 컨트롤을 휴대해야 한다는 제한점이 있는데, 이를 보완하기 위해 박수 소리와 같이 청취자가 발생시키는 소리를 이용한 위치 추적 방법이 제안되었다. 이 방법은 충격파 응답(impulse response) 과 비슷한 파형을 갖는 박수 소리를 이용하는 방법이다. 두 개 이상의 마이크로폰으로 소리를 입력 받아 각 마이크로폰의 신호간 시간 지연을 추정하여 청취자의 (x,y) 좌표를 추정하도록 하였다. 소리를 이용한 위치 추적 방법은 청취자의 위치 (x,y) 가 각 마이크로폰의 소리 도달 시간의 함수로 주어진다라는 사실에 바탕을 둔 것이다. 각 마이크로폰의 지연 시간과 청취자 위치값은 비선형식으로 표현되며, 따라서 청취자 위치값은 Gauss-Newton 방법 등

을 적용하여 구하게 된다.

이와 같은 방법은 위성항법장치 (global positioning system) 에 사용된 측위 기술을 음향 신호에 적용시킨 예로서 영상, 초음파, 적외선과 같은 별도의 신호를 사용하지 않고 청취자가 자체적으로 발생시킬 수 있는 신호를 사용하여 위치를 추적하는 기술로 볼 수 있다. 음향 신호를 사용하기 때문에 배경 잡음, 현재 재생되고 있는 소리, 재생 환경의 잔향 등에 의해 영향을 받을 수 있다. 이러한 영향은 현재까지 발표된 많은 잡음 제거 기술 및 잔향 제거 기술 등을 적용하여 부분적으로 극복하고 있다.

## 2. 사운드 랜더링 기법

### • 머리전달함수를 이용한 렌더링 기법<sup>[3-4]</sup>

두 개의 스피커를 이용한 재생 환경에서 sweet spot 조정 문제는 청취자가 인지하는 음상이 이동시키는 기법을 통해 해결할 수 있다. 청취자가 인지하는 음상의 위치 제어는 방향상을 부여하는 주파수전달함수를 통해 구현될 수 있으며, 이에 대한 대표적인 함수가 머리전달 함수(Head Related Transfer Function, HRTF) 이다. 머리 전달 함수는 자유 공간 상에서 청취자의 귀와 음원의 위치간 전달 특성을 나타내는 함수로서, 음원과 청취자간 고도각 (elevation angle) 및 방위각 (azimuth angle) 에 따라 각기 다른 값을 갖는다.

HRTF를 이용한 sweet spot 보정은 좌, 우 채널의 소리가 청취자로 하여금 서로 대칭인 방향에서 들리도록 하고 좌, 우 채널의 각소리가 좌, 우의 귀에만 전달되도록 처리하는 기술 (crosstalk cancellation) 을 통해 구현된다. 즉, <그림 2>에서 주파수 도메인으로 나타낸 좌, 우

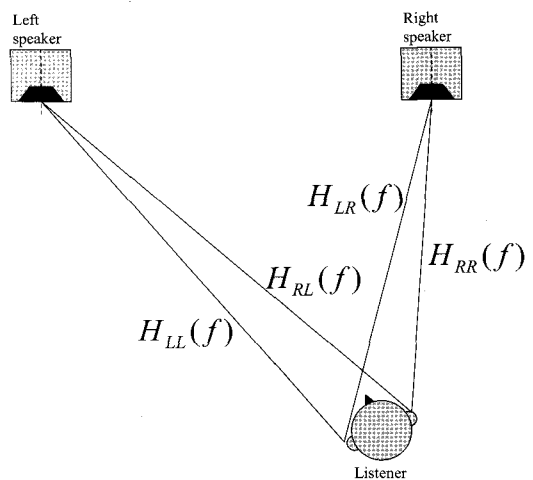
채널의 각 소리를  $S_L(f)$ ,  $S_R(f)$ , 청취자의 귀에 도달한 소리를  $X_L(f)$ ,  $X_R(f)$  라 하면, 다음과 같이 나타낼 수 있다.

$$\begin{bmatrix} X_L(f) \\ X_R(f) \end{bmatrix} = \begin{bmatrix} H_{LL}(f) & H_{LR}(f) \\ H_{RL}(f) & H_{RR}(f) \end{bmatrix} \begin{bmatrix} S_L(f) \\ S_R(f) \end{bmatrix} \quad (1)$$

여기서  $H_{LL}$  과  $H_{LR}$  은 각각 왼쪽 귀로부터 왼쪽 스피커, 오른쪽 스피커간의 HRTF를 나타내며 마찬가지로  $H_{RL}$ ,  $H_{RR}$  은 오른쪽 귀에서 측정된 왼쪽 스피커와 오른쪽 스피커 간의 HRTF를 나타낸다. 식(1)을 행렬식으로 표현하면 다음과 같다.

$$X = HS \quad (2)$$

여기서  $X = [X_L(f) \ X_R(f)]^T$ ,  $S = [S_L(f) \ S_R(f)]^T$  를 나타내며  $H$ 는 HRTF를 포함하는 행렬을 나타낸다. Crosstalk는  $X$ 로부터  $S$ 를 얻기 위한 행렬  $C$ 를 구하는 것으로,  $C$ 는 다음과 같이 나타낼 수 있다.



<그림 2> HRTF를 사용한 청취자의 귀-스피커간 경로 모델

$$S = CX = CHX = H^{-1}HX \quad (3)$$

즉, 행렬  $H$ 의 역행렬을 구함으로써 crosstalk가 구현될 수 있다.

이와 같은 HRTF를 이용한 sweet spot 보정 기술은 청취자의 음원과 귀 간의 경로를 반영하여 정확하게 음상을 제어할 수 있다. 그러나 HRTF가 청취자의 귀, 머리 모양에 따라 각기 다르게 나타나기 때문에 청취자별로 측정된 HRTF를 사용해야 하며, 사용자 자신의 HRTF를 사용하지 못한 경우 올바르게 음상을 제어할 수 없다. 일반적으로 HRTF를 측정하기 위해서는 무향실과 양이마이크로폰과 같은 전문화된 설비와 장비가 필요하며, 개인이 음악을 청취하는 환경에서는 측정하기 어렵다. 따라서 HRTF를 이용한 sweet spot 보정 기술은 실제 구현이 매우 어렵다는 단점이 있다.

최근에는 이러한 문제를 부분적으로 해결하기 위해 사용자별로 측정된 HRTF를 사용하지 않고, 타인으로부터 측정된 HRTF를 개인화(customization)<sup>[6~7]</sup> 하여 사용하는 방법이 제안되고 있다. 개인화의 방법으로서 다양한 귀와 머리 모양을 갖는 복수의 사람들로부터 여러 종류의 HRTF를 측정하고, 사용하고자 하는 사람의 HRTF와 가장 가까운 특성을 갖는 HRTF를 선택하는 방법이 있다. 대표적인 예로서 Zotkin<sup>[6]</sup>은 해부학적인 파라미터와 이에 대응하는 HRTF가 함께 제공되는 HRTF 데이터 베이스를 이용하여 해부학적 파라미터의 유클리디언 거리가 가장 적게 나타나는 HRTF를 선택하는 방법을 제안하였다.

HRTF를 개인화 시키는 또 다른 방법은 비개인화된 HRTF를 변형시켜 사용하는 방법이다. 예로서 Middlebrooks<sup>[7]</sup>은 주파수 영역에서

HRTF를 청취자의 귀 모양에 따라 압축/신장 기법을 제안하였다.

이러한 방법을 통해 개인화된 HRTF를 사용하는 경우, 타인의 HRTF를 사용하는 경우와 비교하여 음상 오차 면에서 우수한 성능을 보인다고 보고되어 있으나, 개인별로 측정된 HRTF를 사용하는 경우와 비교하여 아직까지 많은 오차를 나타내고 있다. 이는 sweet spot의 제어 시 청취자 자신의 HRTF를 사용하지 않는 한 어느 정도의 오차가 항상 발생할 수 있음을 의미한다.

- 자유 공간 전파 모델을 이용한 음상 제어 기법

비개인화된 HRTF의 사용에 따른 여러 문제점에 따라 보다 단순화된 경로 모델을 사용하는 기법들이 제안되고 있다. 이 중 자유 공간 모델을 사용한 기법에서는 스피커-청취자간 경로를 주파수 영역의 전달함수로 세분하여 표현하는 대신, 단순화된 근사식을 사용하고 있다. 자유 공간에서 음원과의 거리가  $r$ 만큼 떨어진 위치에서 음압 레벨  $p(r)$ 은 다음과 같이 나타낼 수 있다<sup>[5]</sup>.

$$p(r) = \frac{j\omega\rho_0 q e^{-jkr}}{4\pi r} \quad (4)$$

여기서  $k$ ,  $\rho_0$ ,  $q$ 는 각각 파동수, 매질의 밀도, 음원의 유효 크기를 나타낸다. 이러한 자유 공간 모델을 이용하여 <그림 2>의 각 경로를 나타내면 다음과 같다.

$$H_{free\ field} = \frac{\rho_0}{4\pi} \begin{bmatrix} \frac{e^{-jkr_{LL}}}{r_{LL}} & \frac{e^{-jkr_{LR}}}{r_{LR}} \\ \frac{e^{-jkr_{RL}}}{r_{RL}} & \frac{e^{-jkr_{RR}}}{r_{RR}} \end{bmatrix} \quad (5)$$

여기서  $r_{LL}$  과  $r_{LR}$  은 각각 왼쪽 귀로부터 왼쪽 스피커, 오른쪽 스피커간의 거리를 나타내며 마찬가지로  $r_{RL}$ ,  $r_{RR}$  은 오른쪽 귀로부터 각 스피커간의 거리를 나타낸다.

자유 공간 모델은 식 (4)와 (5)에 나타난 것과 같이, 귀와 스피커간의 거리만으로 공간 특성을 나타낸 것이며, HRTF와는 달리 개인별 특성이 반영되어 있지 않다. 따라서 개인별로 특성화된 sweet spot 제어가 필요하지 않다는 장점이 있다. 그러나 음원을 단순히 시간 지연과 레벨 차이를 변경하는 방법을 통해 구현하므로 주파수별로 세분화된 음상 제어가 불가능하며, 음원과 머리간의 전달 특성을 반영하지 못한다는 단점이 있다.

그러나 Rose 등의 연구에서는 HRTF를 이용하여 sweet spot을 제어하는 방법과 자유 공간 전파 모델을 이용하는 방법간에 제어 가능한 sweet spot의 크기면에서 큰 차이를 나타내지 않았으며 (반경 25cm) 최근의 Ward<sup>[8]</sup>, Merchel<sup>[9]</sup> 등의 연구에서도 단순화된 음상 제어 방법이 청각적으로 큰 차이를 보이지 않음을 보고하였다.

### 3. 청취자 위치 기반 사운드 렌더링 시스템의 예

- Desktop audio system with head tracking<sup>[2]</sup>

미국 University of southern california 의 Integrated Media system center에서 개발된 오디오 시스템으로서, 비디오 카메라를 이용한 청취자 위치 추적 장치와 sweet spot을 보정하는 부분으로 구성되어 있다.

청취자 위치 추적은 비디오 카메라가 연결된

SGI Indy 워크스테이션에서 이루어지며 sweet spot 보정을 위한 렌더링은 아날로그 디바이스사의 32비트 디지털 신호처리 프로세서인 ADSP-2106x 을 통해 구현되었다. 렌더링에는 HRTF를 사용하지 않고 단순히 양쪽귀의 도달 시간을 제어하는 방법을 사용하였는데, cross-talk 영향 역시 고려되지 않았다. Sweet spot은 양쪽 귀의 각 채널에 해당하는 신호가 동일한 지연 시간을 갖고 도달하도록 제어되었으며 채널 간 최대 시간 지연은 340  $\mu$ sec 로 설정하였다.

본 시스템은 초당 30장 발생하는 동영상을 이용하여 청취자 위치를 인식하므로 청취자 위치 변동을 매우 빠르게 감지할 수 있다는 장점이 있는 반면, 워크스테이션을 사용하므로 실용상으로 어려움이 있다.

- Immersive audio rendering system using laser-based tracking<sup>[10]</sup>

미국 USC의 Integrated Media System Center 내 Immersive Audio 연구소에서 제안한 시스템으로서 레이저를 이용한 청취자 위치 추적과 HRTF를 이용한 사운드 렌더링 기법이 사용되었다.

2차원 레이저 레이더 시스템을 사용하여 원거리 (최대 8m) 에 있는 청취자를 비교적 높은 정밀도 (방위각 0.5도 간격) 로 측정이 가능하였음을 보고하였다.

렌더링에는 HRTF를 이용한 역필터링 기법이 적용되었는데, HRTF가 최소 위상 시스템 (minimum-phase system) 을 보장하지 못하므로 역필터의 안정성을 보장하기 위한 몇 가지 방법이 제안되었다.

또한 청취자의 급격한 위치 변동이 있는 경우, HRTF의 급격한 변동으로 인한 클릭 잡음을 제

거하기 위해 공간 상태를 모델링하기 위한 기법이 적용되었다.

- Adaptive virtual surround sound rendering system<sup>[3]</sup>

국내 삼성전자에서 제안한 방법으로 디지털 텔레비전에 탑재를 목적으로 개발되었다. 사용자 위치 추적은 청취자 위치에서 발생하는 적외선과 초음파를 이용하였으며, 별도의 리모트 컨트롤을 사용하였다.

렌더링은 5.1채널 서라운드 신호를 2채널 스테레오 신호로 변환시키고, 변환된 신호를 방향성 필터를 통과시켜는 방법으로 구현하였다. HRTF를 기반으로하는 crosstalk 제거 방법이 사용되었으며, 동일한 방향에 대한 양쪽 귀의 HRTF를 1개의 HRTF와 시간 지연, 크기 보정 계수로 표현하는 방법을 제안하였다.

적외선과 초음파를 이용한 위치 추적 기법의 정밀도는 수평 방향 평균 오차 2.9cm, 방향 오차 2.0도로 보고되었으며, 비교적 저렴한 가격의 마이크 보드를 텔레비전내에 별도로 추가하여 시스템을 구현할 수 있다고 보고하였다.

#### 4. 청취자 위치 기반 사운드 렌더링 시스템의 향 후 전망

- 청취자 위치 추적

본 고에서 살펴본 오디오 시스템의 청취자 위치 추적은 일반적인 위치 추적 시스템과 비교하여 다음과 같은 차별성을 갖고 있어야 한다.

- 1) 오디오 시스템에 서브 시스템 형태로 탑재되어야 하므로 하드웨어 구조가 간단해야 하며 되도록이면 기존 시스템의 하드웨어

를 활용하여 구현할 수 있어야 한다.

- 2) 위치 추적의 정밀도는 추정 오차가 청취상으로 차이를 나타내지 않는 범위내에 있어야 한다. 즉, 절대적인 오차보다는 청취상의 오차가 더 중요한 요인이다.
- 3) 사용되는 환경의 변동 요인에 강인한 위치 추적 기술 필요.
- 4) 위치 추적에 따른 사용자의 불편함을 야기시키지 않는 기술 필요.

1)번과 2)번을 고려할 때, 위치 추적을 기반으로 하는 오디오 렌더링 시스템은 매우 높은 정밀도로 청취자의 위치를 추정할 수 있는 고성능 알고리즘 보다 어느 정도의 오차를 유지하면서 적은 계산량과 메모리로 구현될 수 있는 알고리즘이 더 선호됨을 알 수 있다. Rose 등의 연구<sup>[5]</sup>에 따르면 청취자가 비대칭 영역에 존재하는 경우, 제어 가능한 최대 sweet spot의 반경은 25cm로 보고하였는데 이는 청취자 위치 추적의 최대 허용 오차로 간주할 수 있다.

청취자 위치 추적은 어떤 신호를 사용하여 위치를 추정하는가에 따라 그 정확도가 변동될 수 있다. 예로서 카메라로부터 얻어지는 영상 신호를 사용하는 경우, 주변 조도에 따라 영상의 콘트라스트가 변동하며 이는 청취자 위치 추적에 영향을 끼칠 수 있다. 또한 음향 신호를 기반으로 하는 위치 추적 기법은 사용된 환경의 잔향, 배경 잡음에 따라 다른 결과를 나타낼 수 있다. 김 등이 제안한 초음파와 적외선 이용 방법 및 Georgiou<sup>[10]</sup> 등이 제안한 레이저 추적 시스템은 환경 변화에 영향을 받지 않는 신호를 별도로 사용함으로써 위치 추정의 정밀도를 높이는 방법으로 간주할 수 있다. 그러나 이와 같은 방법은 별도의 신호를 발생하거나 수신하기 위한 부가

적인 하드웨어가 필요하다.

청취자가 발생하는 소리를 이용한 최근의 청취자 위치 추적 방법은 환경 변화 요인에 다소 영향을 받을 수 있으나 보다 향상된 잔향 제거 기법 및 배경 잡음 감쇄 기법을 통해 그 정밀도가 증가될 것으로 기대되고 있다.

또한 최근의 오디오 시스템이 사용상의 편리성이 강조되고 있기 때문에, 위치 추적을 위해 청취자가 별도의 행위를 취하지 않아도 되는 방법이 선호되고 있다.

#### • 위치 기반 사운드 렌더링 기법

HRTF를 이용하는 렌더링 기법은 자유 공간 전파 모델 기법과 비교하여 객관적, 주관적으로 높은 성능이 기대됨에도 불구하고 여러 기술상의 난제로 인하여 실제 시스템에의 적용이 어려운 형편이다.

이러한 난제 중 하나는 전술한 바와 같이 사용자의 특성에 적합한 HRTF를 현실적으로 사용하기 힘들다는 점이다. 이와 같은 난제를 부분적으로 극복하기 위해 현재까지 제안된 HRTF 개인화 방법은 사용자의 귀와 머리 모양을 필요로 하며, 이에 따라 별도의 영상 정보를 요구한다. Kyriakakis<sup>[2]</sup> 등이 제안한 위치 기반 렌더링 시스템에서는 장착된 카메라를 통해 청취자의 위치를 인식하는 한편, 청취자의 귀를 촬영함으로써 적합한 HRTF를 얻도록 하였다. 귀의 모양과 머리 모양을 이용한 HRTF 개인화 방법은 인간의 청각과 관련된 신체의 해부학적 구조가 HRTF와 깊은 연관성을 갖고 있다는 가정을 전제로 한다. 이는 HRTF의 특성이 주로 외이(outer ear)의 형태에 의존적임을 의미하는 것이다. 그러나 인간의 청각 시스템은 외이 뿐이 아니고 심리 음향 모델 (psychoacoustic model)

을 포함하는 내이(inner ear)의 특성이 포함된다. 따라서 외이 특성만을 이용하여 개인화된 HRTF는 실제 HRTF와 비교하여 차이를 나타낼 수 있다. 따라서 높은 신뢰성과 적은 계산량으로 비개인화된 HRTF를 개인화된 HRTF로 변환시키는 기법이 필요하다.

HRTF를 사용하는 렌더링 기법의 또다른 문제는 sweet spot의 보정이 가능한 영역 크기가 제한적이라는 점이다. 청취자 위치 기반 오디오 렌더링 시스템은 추정된 위치에 적합한 HRTF를 사용하여 sweet spot를 보정한다. 따라서 sweet spot의 보정 위치는 저장 가능한 HRTF의 방위각, 고도각 수로 제한된다. 예로서 MIT에서 제공하는 KEMAR HRTF<sup>[11]</sup>를 사용하는 경우, 수평면에서의 HRTF 개수는 72가지로서 방위각 5도 간격으로 sweet spot 보정이 가능하다. 1개의 HRTF를 저장하기 위한 메모리 크기는 1KB이며, 양쪽 귀에 대한 72종류의 HRTF를 저장하기 위해서는 총 144KB의 메모리 공간이 필요하다. 이와 같은 메모리 크기는 최근의 메모리 기술을 고려할 때 큰 사이즈라고는 볼 수 없으나 hand-held device 및 DSP의 내부 메모리만을 사용하는 어플리케이션에서는 적지 않은 부담으로 작용할 수 있다. 모든 지점의 HRTF를 저장하지 않고 몇 개의 대표 HRTF만을 저장하고 나머지 지점은 보간(interpolation)을 통해 HRTF를 얻는 방법<sup>[12]</sup>은 보다 세분화된 sweet spot 보정이 가능한 향후 오디오 시스템에 적용될 것으로 판단된다.

## 5. 결론

결론적으로, 청취자 위치 기반 사운드 렌더링 기술은 측위 기술과 오디오 신호 처리 기술이 결



합되어 새로운 개념의 오디오 시스템을 구현하는 기술로 간주할 수 있으며, 과거 최적의 스테레오감을 얻기 위해 사용자가 일일이 시스템을 조정해야 하는 불편함을 해소할 수 있는 기술로 볼 수 있다. 최근 발표되고 있는 디지털 신호처리 분야의 관련 기술들이 오디오 시스템에 적용됨으로써 보다 나은 성능의 청취자 기반 오디오 시스템이 개발될 것으로 기대된다.

#### ===== 참고문헌 =====

- [1] N. Sakamoto, T. Gotoh, T. Kogure and M. Shimbo, "Controlling sound-image localization in stereophonic reproduction," *J. Audio Eng. Soc.*, Vol.29, No.11, pp.794-799, 1981.
- [2] C. Kyriakakis, T. Holman, J.-S. Lim, H. Hong and H. Neven, "Signal processing, acoustics, and psychoacoustics for high quality desktop audio," *J. Visual Comm. and Image Rep.*, Vol.8, No.1, pp.51-61, 1998.
- [3] S. Kim, S. Jang, D. Kong and S. Bang, "Adaptive virtual surround sound rendering method for an arbitrary listening position," *Proc. 30th AES Int. Conference*, 2007.
- [4] 유승수, 김영문, 김선용, 이기승, 이석필 "위치정보와 머리전달함수를 사용한 적응형 사운드 시스템 설계", 제어로봇 시스템 학회 한국자동제어학술회의 논문집, pp. 364-365, 2009.
- [5] J. Rose, P. Nelson, B. B. Rafaely and T. Takeuchi, "Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations," *J. Acoust. Soc. Am.*, Vol.112, No.5, pp.1992-2002, 2002.
- [6] D. N. Zotkin, R. Duraiswami and L. S. Davis, "Rendering localized spatial audio in a virtual auditory space," *IEEE Trans. Multimedia*, Vol.6, No.4, pp.553-564, 2004.
- [7] J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Am.*, Vol.106, No.3, pp.1480-1492, 1999.
- [8] D. B. Ward and G. W. Elko, "A new robust system for 3D audio using loudspeakers," *Proc. IEEE ICASSP*, pp.781-784, 2000.
- [9] S. Merchel and S. Groth, "Analysis and implementation of astereophonic play back system for adjusting the sweet spot to the listener's position," *Proc. 126th AES Convention*, 2009.
- [10] P. G. Georgiou, A. Mouchtaris, S. I. Roumeliotis and C. Kyriakakis, "Immersive sound rendering using laser-based tracking," *Proc. 109th AES Convention*, 2000.
- [11] MIT media lab, KEMAR HRTF data, <ftp://sound.media.mit.edu/pub/Data/KEMAR>
- [12] 이기승, 이석필, "원형 머리 모델을 이용한 머리 전달 함수의 보간" 한국음향학회지, pp.333-341, 제27권, 제7호, 2008년 10월.

## 저자소개



이 기 승

1991년 2월 연세대학교 전자공학과(공학사)  
1993년 2월 연세대학교 대학원 전자공학과(공학석사)  
1997년 2월 연세대학교 대학원 전자공학과(공학박사)  
1997년 10월~2000년 9월 미국 AT&T Labs-Research  
2000년 11월~2001년 8월 삼성전자(주) 종합기술원  
2001년 9월~현재 건국대학교 정보통신대학 전자공학부

주관심 분야 : 오디오 신호처리, 실시간 구현, 생체 신호  
처리