

# 다채널 음악파일에의 입체음향 적용

김용진\*, 송장호\*, 이동재\*, 이원돈\*  
\*충남대학교 컴퓨터학과

e-mail:guliver@brain.cnu.ac.kr

## 3D Sound Application to N channel Sound File

Yong-Jin Kim\*, Jang-Ho Song\*, Dong-Jae Lee\*, Won-Don Lee\*  
\*Dept of Computer Science, Chung-Nam University

### 요 약

본 논문에서는 다양한 채널을 가진 음악 파일에 대하여 입체 음향 효과를 줄 수 있는 시스템을 개발 하였다. 그러기 위하여 3D 사운드 기술 중에 가장 대표적으로 알려진 HRTF(머리전달 함수)를 원음에 컨볼루션(Convolution)하는 방식으로 음상정위 모형을 구현하였으며 음장감을 부여하기 위해 잔향 효과(Reverberation)효과를 추가하고 크로스토크 현상 제거를 위해 트랜스오럴(Transaural) 필터를 추가하였다. 이런 입체음향 기술을 가지고 여러 채널을 가진 음악 파일에 적용시켜서 다채널 입체음향 효과를 낼 수 있는 시뮬레이터를 구현해 보았다. 시스템 구현에는 한정된 채널이 아닌 다양한 채널에 대한 효과를 낼 수 있도록 하였으며 기본적인 실험으로는 미디를 바탕으로한 5개의 채널에 대하여 실험하여 이를 증명해 보았다.

### 1. 서 론

1)컴퓨터의 눈부신 발전 속에서 최초 과학적 탐구를 위한 목적에서 뿐만이 아닌 인간 문화 발전을 위한 부분에서도 많은 노력을 기울이고 있다. 최근 불고있는 멀티미디어의 분야는 이런 노력을 증명해주고 있다. 화려한 3D 만화 영화나 현실감 있는 음향 효과가 바로 이런 것들일 것이다. 이러한 추세로 최근 멀티미디어 기술에 대한 관심이 증대되고 있는데, 그 중 음향 처리 기술중의 하나인 입체음향에서 다 채널 사운드 트랙에 적용하기 위한 연구를 하였고 이를 위한 시뮬레이터 구현과 그 내부 알고리즘을 개발하였다.

### 2. 관련 연구 및 실험

#### 2.1 입체음향 효과

사람은 스피커에서 나오는 음원들에 대하여 위치적 감각을 느낄 수 없다. 입체 음향 효과는 이 같은 현상을 개선하기 위해 음원을 3차원 공간상에 정위시켜 마치 그 위

치에서 소리가 나게 해주는 것이다. 이런 효과를 주기 위해 본 논문에서는 HRTF(머리전달 함수)를 사용하였으며 보다 빠른 연산을 위해 원음을 주파수 영역에서 음상정위 기사는 방법을 사용하였다.[1]

#### 2.2 플레이어의 구현

이를 시험하기 위해 자체적으로 플레이어를 구현하였다. 이 플레이어는 MP3 파일 형태의 음악을 연주할 수 있게 만들었고, 실시간으로 음상정위가 가능하게 만들어졌다.

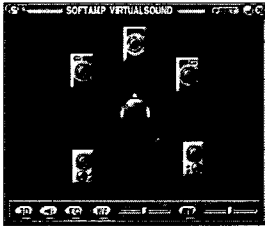
본 논문에서는 입체음향 기술을 MP3 에도 적용할 수 있게 플레이어를 구현하였으며, MP3 뿐만 아니라 여러 가지 다양한 사운드 포맷도 지원할 수 있게 구현하였다. 특히 돌비 디지털 포맷을 위한 5.1 채널 형태의 사운드에 중점을 두어 구현하였다. 구현된 모듈은 우선 음원을 재생 가능한 PCM 파일로 디코딩 해주는 입력 모듈과 디코딩된 음을 재생하는 출력 모듈, 그리고 음악에 입체 음향 효과를 넣어주는 입체음향 모듈로 구현하였다. 사용자의 편의성을 위한 인터페이스 설계를 지향하였고 프로그램은 하나의 실행 파일이 아닌 여러 개의 모듈로 구현하여, 실행파일에서 각각의 모듈을 호출하는 형식으로 이루어졌다. 또한 본 플레이어는 개방 구조로 구현하였다. 이는 객

1) "이 연구는 BK21충남대학교 정보통신인력양성사업단의 지원을 받았음."

채 지향 기법으로 인풋과 아웃풋을 명시함으로써 다른 개발자들이 본 프로그램에 어떠한 수정 없이도 원하는 기능 추가를 가능하도록 구현하였다.

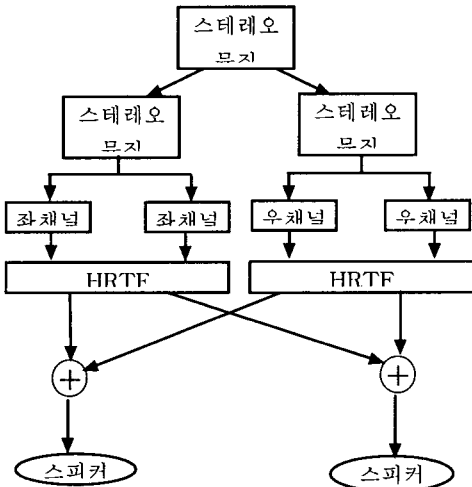
2.3 입체 음향 모듈에 대한 구조와 원리

논문을 위해 5.1채널을 기준으로 하여 개발된 플레이어의 기본 인터페이스는 다음과 같다.



<2.1> 입체음향 모듈의 인터페이스

플레이어에서 사용자는 마우스를 이용하여 쉽게 각각의 스피커를 원하는 위치에 갖다 놓을 수 있다. 그렇게 되면 스피커의 위치에 따라 음상정위가 되고 사용자는 음악을 보다 입체적으로 들을 수 있게 된다. 또한 우측 하단에 있는 잔향 여부에 따라 Room 효과와 Hall 효과를 줄 수 있다. 일반적인 음악 파일이 스테레오(2채널)로 저장되어 있는 점을 감안할 때, 이러한 스테레오 음악에 입체 음향감을 효과적으로 주기 위해서는 좌, 우 채널에 각각의 입체음향 프로세스를 가하는 것이 필요하다. 따라서 [그림 2.2]와 같이 좌, 우의 각 채널을 복사하여 각각 2채널을 만든 후 청자가 원하는 좌, 우 가상 스피커의 위치에 해당하는 HRTF를 원음에 복사된 각각의 2채널과 컨볼루션하여 2개의 가상 스피커를 구현하였다. 이는 단지 두 개의 가상 스피커 효과뿐만 아니라 n개의 가상 스피커 역시 같은 채널을 복사하여 서로 다른 HRTF를 적용하므로 구현이 가능하다. 앞의 모든 모듈들을 통과하고 나온 변환된 음악데이터는 최종적으로 적당한 데이터 값으로 스케일(scale)되어 스피커로 출력되게 된다.

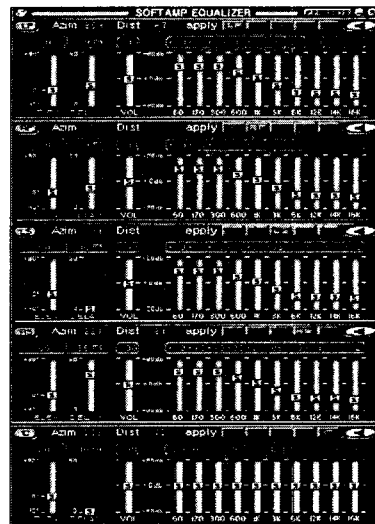


<그림 2.2> 2채널 virtual speaker 구조도

그러나 이를 스피커를 통해 듣기 위해서는 크로스토크(crosstalk)현상 때문에 그 느낌이 현저히 감소하게 된다. 이는 스피커를 통하여 듣게 될 때 왼쪽 귀에만 들려야 할 음이 오른쪽 귀에도 들리게 되고 오른쪽 귀에서도 마찬가지로 현상이 일어나기 때문이다. 이런 현상 제거를 위해 트랜스 오럴 필터(transaural filter)를 추가했다[2][3]

2.4 5개의 가상 스피커를 위한 이퀄라이저 인터페이스

[그림 2.3]은 5개의 가상 스피커를 위한 것이다. 각각의 이퀄라이저 인터페이스에는 Azim 과 Dist가 있다. 이는 0도부터 시계방향으로 360도까지의 값을 나타내며 Dist는 스피커가 듣는 사람으로부터 떨어져 있는 거리를 나타낸다. 그리고 Elev(Elevation), Delay, Volume을 보여주고 있다. Elev는 가상스피커의 높이를 나타낸다. 이는 그림 2.1의 인터페이스에서는 사용자가 2차원으로만 조절이 가능하므로 높이를 조작하기는 불가능하다. 그러나 여기서는 Elev를 수동조작으로 조절이 가능하다. Delay는 해당 스피커에서 나오는 음원의 지연을 말한다. 이는 소리의 음향감 정도를 입체적으로 변화시킬 수 있는 장점을 가지고 있다. 그 옆에 Volume 스위치는 해당 스피커의 볼륨을 조절하는 스위치다. 이는 해당 스피커에서 나오는 소리를 컨트롤 할 수가 있다.



<그림 2.3 이퀄라이저가 실행된 모습>

2.5 이퀄라이저 구현

[그림 2.3]은 시뮬레이션을 위한 이퀄라이저 모습을 나타내고 있다. 10개의 밴드로 이루어져 있고, 각각의 스위치

바를 조절함으로써 원하는 음색을 조율 할 수 있다. 이퀄라이저 효과는 시간영역과 주파수 영역에서 처리 해줄 수 있는데, 본 논문에서는 주파수 영역에서 처리하였다. 이를 위해 음원을 먼저 FFT 연산을 수행하였고, 이 데이터에 HRTF 값을 콘볼루션해서 음상정위 시킨 후 이퀄라이저 효과를 주었다. 그 다음에 다시 역 FFT 연산을 통해 시간영역으로 바꾼 다음 플레이를 하는 방식으로 구현하였다. Sampling Rate 값이 최고 44100Hz 이므로 실제 나이키스트 표본화 범칙(손실 없이 디지털화 하려면 표본화율은 적어도 최대 주파수 응답의 두 배는 되어야 한다)에 의해서 22050Hz까지 표현이 된다. 따라서 위의 이퀄라이저는 0부터 최고 22050Hz까지의 주파수 대역을 나타내고 있다. 10개의 각각의 주파수 대역을 조절하면 그 값을 가지고 나머지 주파수 영역들의 값들을 interpolation방법을 통해 적절한 값으로 정해 준다. 그러면 전체적인 값들이 곡선형태를 그리게 되고, 이 각각의 값들을 원음에 곱하여 이퀄라이저 효과를 낼 수 있게 된다.[4]

### 3. 다채널 사운드에 대한 입체음향 실험

#### 3.1 미디를 기반으로한 음상정위

다 채널로 되어있는 미디 음원을 각 음원별로 데이터를 추출한 후 추출한 웨이브 폼으로 저장한다. 이렇게 미디 음원을 웨이브로 바꿀 때 사용되는 사운드 폰트에 따라 음악의 질적 차이가 나는데 본 실험에서는 웨이브 메이커 프로그램을 사용하였다.

각각의 채널을 웨이브 파일로 만든 다음 3D 프로세싱을 한다. 프로세싱이 끝난 파일은 3D 프로세싱 하는 과정에서 왼쪽과 음의 크기 차이가 날 수 있다 따라서 이 차이를 없애주는 작업이 필요한데 이를 위해 모든 샘플의 음의 크기의 절대값을 구해서 이를 모두 더해준다.

$$\begin{aligned} \text{ABS\_NUM1} &= \text{ABS}(\text{Original Data}) \\ \text{ABS\_NUM2} &= \text{ABS}(\text{Data After 3D filter}) \\ \text{SUM\_A} &= \text{SUM\_A} + \text{ABS\_NUM1} \\ \text{SUM\_B} &= \text{SUM\_B} + \text{ABS\_NUM2} \end{aligned}$$

이 때 절대값 함수의 사용은 data의 값이 음수를 가질 수 있기 때문이다. 샘플 레이트 비율이 초당 44100이므로 스테레오 채널에서는 초당 88200개의 샘플이 된다. 따라서 위와 같은 과정을 88200번 수행하게 된다. 즉 위의 과정은 1초간의 샘플 개수만을 가지고 계산한 것이다. 이렇게 얻은 SUM\_A 와 SUM\_B의 비율을 구하기 위해 다음과 같은 연산을 수행한다.

$$\text{Alpha} = \text{SUM\_A} / \text{SUM\_B}$$

그러나 3D 프로세싱 과정에서 데이터의 특정 구간이 모두 0이거나, 원음에서 데이터 구간이 모두 0이 될 수도 있다. 그렇게 되면 Alpha값이 제대로 구해지지 않을 수 있기 때문에 특정 구간에 유효한 샘플이 몇 개 이상 존재할 때에만 위와 같은 과정으로 Alpha 값을 구해 주어야 한다. 본 실험에서는 데이터의 절대값이 30 이하인 값이 전체

88200개중에 6000개 이하로 나오는 처음 구간을 선택하였다. 즉 20 이상인 값이 28200인 경우 Alpha를 구하였다. 이는 다양한 실험에서 얻은 수치 값이다. 이렇게 구해진 Alpha를 3D 필터를 거친 데이터에 곱해주면 왼쪽과 음의 크기가 평균적으로 같은 결과를 얻을 수 있다.

다음 이렇게 구해진 데이터를 저장할 때 노멀라이즈 작업을 해준다. 이는 샘플하나의 크기가 가질 수 있는 최대 값이 있는데(2byte=32765) 오버플로우를 막기 위하여 다음과 같은 연산을 수행한다.

$$\begin{aligned} \text{MAX\_NUM} < 32765 &\Rightarrow \text{Beta} = 1 \\ \text{MAX\_NUM} > 32765 &\Rightarrow \text{Beta} = 32765/\text{MAX\_NUM} \end{aligned}$$

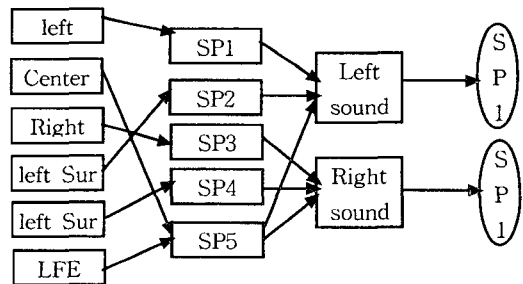
이렇게 구해진 Alpha 와 Beta를 가지고 결과 데이터를 구한다.

$$\text{result}[i] = \text{DATA2}[i] * \text{Alpha} * \text{Beta}$$

각 채널별로 위와 같은 과정을 거치면 채널수와 같은 웨이브 파일이 만들어지는데, 이 데이터를 모두 합치는 과정에서 스케일 작업과 노멀라이즈 작업을 해줬으며 특정 채널의 크기 조절을 위해 각각에 factor 값을 지정해 주었다.

#### 3.2 5.1 사운드 트랙에 대한 2채널 입체음향 시스템

5.1채널의 돌비 디지털 사운드에 대한 입체 음향을 위해 다음과 같이 각각의 가상스피커에 할당해 주었다.



<3.1>돌비 5.1 채널에 대한 입체 음향 매핑

### 4. 결론

오디오 분야에서 멀티 채널은 대세로 다가오고 있다. 특히 돌비 5.1 채널은 그 시스템 가격이 낮춰짐에 따라 더욱 많이 사용될 조짐이다. 그러나 아직도 컴퓨터에서나 기존의 시스템을 사용하는 사람이 많고 또한 MIDI음악은 멀티 채널임에도 불구하고 제대로 그 멀티 채널의 장점을 살리지 못하고 있는 것이 사실이다. 따라서 본 실험에서는 멀티 채널을 2개의 스피커로 들을 수 있도록 입체음향의 효과를 사용하여 구현하였으며 이것이 기존의 단순한 스테레오보다 보다 다양한 옵션으로 제공될 수 있도록 하였다. MIDI의 각 악기가 다른 위치에서 소리가 들릴 수 있게 함으로써 마치 청취자가 오케스트라의 지휘자가 되

어 한 가운데서 음악을 듣는 듯이 만들 수 있다. 앞으로 더욱 세밀하게 각 악기의 위치를 정할 수 있도록 하고 또한 그 위치에서 음원이 있는 것처럼 더욱 들릴 수 있게 하는 것이 과제이다. 그렇게 된다면 실제의 라이브 음악과 더욱 가까워진 양질의 음악을 감상할 시대가 오게 될 것이다.

#### 참고문헌

- [1] HRTF Measurements of a KEMAR Dummy-Head Microphone MIT Media Lab Perceptual Computin - Technical Report #280, Bill Gardner and Keith Martin, 1994
- [2] Durand R. Begault, "3-D SOUND for virtual reality and multimedia" AP Professional, 1994.
- [3] 강성훈, 강경옥, "입체음향 Spatial Audio", 첨단 과학 기술도서 출판, 1997.
- [4] William H.Press, Saul A. Teukolsky, William T.Vetterling, Brian P.Flannery,"NUMERICAL RECIPES in C", CAMBRIDGE 1992.