

# 입체음향(3D 오디오) 기술과 원리

□ 최 범 석 · 홍 진 우 / 한국전자통신연구원 음향기술연구팀

## 1. 서 론

오늘날 멀티미디어 기술의 발전은 영상과 음향에 대한 전반적인 기술의 발전을 가져왔다. 특별히 영상분야에 있어서는 화려한 색상과 넓은 화면, 고선명 화질, 여기에 3D 영상까지 눈부신 발전을 보여 왔다. 이에 따라 컴퓨터 음향분야에서 역시 좀 더 현실감 있는 사운드를 만들기 위하여 많은 노력이 있어왔다. 단순히 리듬만을 연주해 주던 컴퓨터 음악에서 사람의 말소리와 여러 가지 효과 음향의 지원, 그리고 더 나아가서 원음에 가까운 소리를 연주해 주는 기능과 좀 더 웅장하고 실제적인 분위기를 연출할 수 있는 소리를 요구하게 된 것이다. 입체음향(3D sound)는 이러한 요구를 가장 최적으로 맞추어 줄 수 있는 음향에 있어서의 최첨단 기술이라고 할 수 있다. 본고에서는 입체음향의 관련 연구 동향과 원리 그리고 문제점

등을 중심으로 간략히 살펴보고자 한다.

오늘날의 멀티미디어 컴퓨터는 CD음질 수준의 스테레오 사운드를 제공하고 있다. 일반적인 스테레오는 오른쪽 스피커와 왼쪽 스피커 사이에서 어느 곳이나 음을 위치시킬 수 있다. 그러나 입체음향은 좌, 우 두 개의 스피커로 음원을 좌, 우, 아래, 위, 가까운 곳 먼 곳, 어느 위치나 정위(음상정위:sound image location)시킬 수 있다. 이는 입체음향 기술을 이용한다면 음장의 완벽한 제어가 가능하다는 것을 뜻한다. 입체음향에서는 단순한 음상의 분리뿐만 아니라, 잘 들리지 않는 아주 작은 레벨에 대한 분해 능력이 있게 된다. 또한 소음 속에서도 음상(sound image)을 추출하기도 하고, 또는 소음 자체를 식별하여 제거하기도 하며 음의 풍부함이나 유쾌함까지 개선되게 한다. 이러한 현상은 단일 채널의 파형 전송으로는 실현할 수 없으며, 입체음향 고유의 현상이다. 입체음향의

이러한 특징은 실감음향의 구현에 있어서 없어서는 안 될 특징이라 할 수 있으며 실제로 많은 음향관련 응용분야에서 이를 적용하고 있는 것이 사실이다.

입체음향을 응용할 수 있는 분야는 사실 음향이 사용되는 모든 분야라 할 수 있다. 종래의 스테레오 음향에 비하여 그 효과가 탁월하기 때문에 입체음향을 적용한 시스템을 한번이라도 사용해 본 유저라면 종래의 시스템으로 다시 돌아가기를 꺼려할 것은 자명하다. 입체음향을 도입하기 시작한 분야중 대표적인 분야는 게임, 가상현실, 오디오, 영화, 음향통신 등의 분야라 할 수 있으며 이 중 가장 앞서서 도입하고 있는 곳은 주로 게임에서의 효과음과 실감나는 음향이 생명이라 할 수 있는 영화음향 부분이라 할 수 있다. 뛰어난 그래픽과 화려한 사운드의 완벽한 조화가 필요한 게임의 세계는 멀티미디어의 정점에서 있다고 해도 과언이 아니며, 그런 만큼 각종 최신 기술이 빠르게 도입되고 있다. 칠흠 같은 어둠 속에서 보이지 않는 괴물의 비명소리는 플레이어의 간담을 서늘하게 만들며, 괴물들이 문을 여닫는 소리는 그 위치를 판단하는 데 중요한 단서가 된다. 이와 같이 디지털 적인 처리로 3D 사운드를 합성해 내는 것은 가상현실 분야에서 큰 관심거리로 부각되고 있다. 또한 이미 다채널 음향을 사용하고 있는 극장에서만 느낄 수 있었던 실감나는 음향을 입체음향을 사용하면 일반 가정의 소규모 장소에서 단지 두개의 일반 스피커를 가지고도 느낄 수 있게 할 수 있다. 이는 일반적으로 좋은 음질을 얻으려면 비싼 오디오 하드웨어를 사야만 한다는 부담을 제거하므로 비용 절감 측면에서 그 파급효과가 크다고 할 수 있다.

## 2. 관련연구 동향

현재, 멀티채널 음향과 음향 시뮬레이션 기술은 원 음장의 고 정밀도의 재현이 가능하며, VR 기술에서 head mount display에 장착된 헤드폰 재생은 헤드 트래커라고 하는 청취자의 머리 움직임의 검출을 이용하여 머리 전달 함수를 컴퓨터로 제어하는 방식을 채용한 본격적인 바이노럴 재생 시스템이다. 그러나 음향 이론적으로는 실현 가능한 기술도, 기본적으로는 최적 위치에서의 실현이고, 기술적, 경제적 제약이 항상 존재하게 되며, 순수하게 청각만의 관점으로부터는 개인차의 문제가 남아 있다. 따라서, 오히려 VR 이라고 하는 지각의 복합 기술에서 그 응용 가치를 높이는데 의미가 있고, 또 A 와 V의 상승효과를 활용한 복합 시스템으로서의 실용성을 주목받고 있다.

원 음장의 충실한 재현을 위하여 멀티채널에 의한 수음 전송을 하기 위해서는 적어도 20채널 이상이 필요하다. 이것을 일반 가정에서 구성하는 것은 기술적, 경제적으로 볼 때 불가능하다. 1970년대에 처음으로 등장한 4 채널 스테레오는 오디오 소프트/하드와 유저 요구가 일치되지 않았고, 이로 인해 상업적으로도 실패로 끝나고 말았다. 그러나, 70년대 말에 Dolby사가 4채널 영화 서라운드 시스템을 개발하였고, 80년대 이후에는 VTR, Video Disc에 의한 본격적인 AV home theater용 돌비 서라운드, Dolby Prologic 인코더, 디코더에 의해서 앞 방향 3채널(L,R,C), 뒤 방향 1채널로서 정착되었다. 또한, Dolby Prologic을 기본으로 하고, 영화의 서라운드에 가까운 음장을 재현하기 위하여, 재생 기기의 음향 특성을 규정한 미국의 루카스 필름사의 THX 시스템이 영화관과 가정

용으로서 개발되었다.

NHK의 HiVision의 음향도 앞 방향 3 채널, 뒤 방향 1 채널로 구성된 멀티채널을 도입하고 있다. Dolby 사에서는 차세대 디지털 AV 미디어로서 기대를 한 몸에 담고 있는 DVD의 음향 방식으로, 앞 방향에 3 채널, 뒤 방향에 2 채널, 중저음용 0.1 채널, 합계 5.1 채널의 Dolby Digital을 개발하였다.

종래의 2채널 스테레오 음원, 특히 AV 프로그램의 재생시에 화면과 일치된 음장 공간 정보를 DSP 제어에 의해 추가하는 방식으로서 선구적인 Yamaha의 DSP, JVC의 DAP, Matsushita의 DSFC 등이 있다. 또한, 호주의 Lake DSP 사의 Huron(256K tap, 샘플링 주파수 48kHz)은 초고속 디지털 오디오 버스를 병렬 구성하여 지연 제로, 최대 262144포인트의 임펄스 응답과 콘볼루션 처리를 실현한 audio workstation을 이용하여, 청취자 위치, 규정한 실내의 반사, 음성의 데이터 속도 등을 실시간으로 분석하여, 청취자의 동작에 일치하여 3차원의 가상 음장 시뮬레이션과 재생이 가능하다. 또, 건축 음향 시뮬레이션 소프트웨어를 사용하면, 상정하는 실내의 형상이나 규모, 건축 소재 등의 정보를 입력하는 것만으로 잔향 효과 등의 음장 시뮬레이션을 쉽게 할 수 있다. 또 재생 스피커를 4개 이상으로 하면 청취 영역을 확대할 수 있다.

입체 영상 방송을 위한 현장감이 높은 전송 시스템으로서 NHK가 개발한 음장 능동 제어 방식에서는 1 채널 당 36개의 스피커를 부착한 판을 측벽과 후벽에 이용하여, 음성 추적법 시뮬레이션에 의하여 초기 반사 및 잔향음을 벽면으로부터의 반사음 레벨, 방향 및 지연 시간의 제어를 기본으로 추가하는 방법이 제안되고 있다.

이것에 의하여, 스피커 wall 부근에서도 스피커의 존재를 의식하지 못하고, 청취 영역에 제한이 없으며, 자연스러운 음장을 실현할 수 있다.

입체 음향 시스템의 구성은 음원을 3차원으로 배치하는 convolotron과 실제 음원을 발생시키는 CD, MEAX II(digital sound process) 및 시스템 전체의 host computer인 Machintosh로 구성되어 있다. Machintosh는 센서인 3 space tracker로부터 받은 머리 위치, 방향, 음원의 위치를 convolotron에 전달한다. 이 머리의 위치와 방향, 음원의 위치 정보는 컴퓨터 그래픽스를 rendering 하는데도 이용되어, 가상 공간을 생성한다. 즉, 3차원 CAD에 의해서 미리 그려진 3 차원 모델을 머리 위치와 방향으로부터 실시간으로 eye phone(head mount display)에 그린다. 가상 현실분야에서 주목되고 있는 청각 디스플레이는 Binaural Hearing을 이용한 음상의 3차원 정위가 가능한 시스템이다. 음원으로부터 두 귀의 고막까지의 전달 함수를 미리 몇 군데 위치에 대하여 계측하여 저장하고, 보간 계산에 의해 순시치를 구하여, 이것을 기본으로 고막 상에서의 음압을 실시간으로 계산하면, 실질적으로는 존재하지 않는 음원을 근사적으로 생성할 수 있다. 이 시스템을 이용하면 헤드폰을 사용할 때만 3차원적인 음장을 재현할 수 있다. 대표적인 시스템으로는 convolotron이 있다. 이것은 IBM PC를 host computer로 하여 동작하고, TMS/C25 을 탑재한 spectrum board와 이 보드와 parallel port로 통신하는 convolution card로 구성되어 있으며, 미국의 Crystal River Engineering사가 개발하였다. Convolotron에서는 입체음향을 생성하기 위하여 Foster에 의한 머리전달함수를 도입하였다.

시뮬레이션에 의한 음 발생의 모델을 이용하여, 예를 들면, 물체와 물체의 충돌 음이나 마찰음 등의 파형을 합성하며, 그것을 음원으로 하여, 헤드폰에 가해질 파형을 계산하는 것이 음향 분야의 랜더링 기술이다. 사운드 랜더링은 Takala에 의해 1992년에 제안된 것이다. 많은 CG 작품에서 영상은 계산기에 의해 수동적으로 생성되고 있지만, 사운드 트랙은 보통의 수작업 애니메이션이나 사진의 영상과 똑같은 수법으로 만들어진다. 이것을 영상과 같은 레벨로 계산기로 합성하는 것이 사운드 랜더링의 기본 사고방식이다.

### 3. 입체음향 기술

#### 3.1 음상 정위(Sound image location)

음상 정위 기술이라 함은 임의의 음원을 제작자 마음대로 3차원 공간상에 위치시키는 기술을 말한다. 이는 입체음향 기술만이 갖는 가장 큰 특징이자 장점이라 할 수 있다. 여기서는 음상 정위 기술의 간단한 원리와 구현방법을 언급하고자 한다.

사람은 두 귀로 3차원 공간상의 음원의 위치를 판별할 수 있다. 이러한 음원의 위치판별에 가장 큰 영향을 주는 요소가 바로 음이 두 귀에 도달하는 시간차(ITD: Interaural Time Delay)와 두 귀에서 느끼는 음의 레벨(에너지)차(IID: Interaural Intensity Difference) 그리고 스펙트럼 차(ISD: Interaural Spectrum Difference)라고 할 수 있다. 이러한 원리를 이용하여 헤드폰 또는 두 개의 스피커로 음원의 위치를 3차원 공간상에 어느 곳에나 정위(locate)시킬 수 있는 기술을 입체음향기술 이라고 한다. HRTF(이하 머리전달함수)는 이

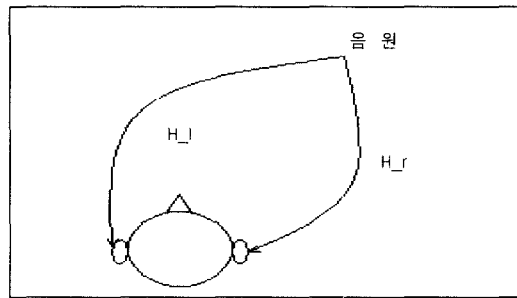


그림 1. 방향감 합성 원리

러한 음원의 위치를 판별하는데 단서가 되는 두 귀간 시간차, 음압 차, 스펙트럼 차의 정보를 모두 담고 있는 impulse response 이다. 그림 1 에서 H-L과 H-R은 각각 모노음원이 좌, 우측 귀에 입사하기까지의 전달 경로를 나타낸다.

더미헤드 양옆의 모조 귀 안에 마이크로폰을 설치하고 일정한 거리에서 impulse 파를 내보내면 이 파가 더미헤드의 머리와 귓바퀴의 특성을 타고 귓속에서 미리 설치된 마이크로폰에 녹음이 되는데, 음원의 위치를 변화시켜 측정함에 따라 왼쪽 귀와 오른쪽 귀에 녹음되는 response 파의 특성이 서로 다르다. 따라서 인간은 이 두 귀에 입사해 들어오는 시그널의 차에 의하여 음원의 위치를 판별하는 것이다.

입체음향 생성 기술 중 대표적인 방법이 바로 원음에 머리전달함수를 콘볼루션(convolution)하는 방식이다.[3] 다음은 콘볼루션의 일반식이다.

식 (3.1.1)은 시간영역에서의 콘볼루션 식이

$$y[n] = \sum_{i=0}^{i=N-1} x[n-i] * h[i] \quad (3.1.1)$$

$$y[i] = X[i] * H[i] \quad (3.1.2)$$

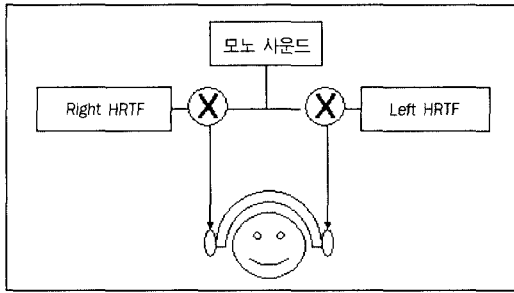


그림 2. 모노 사운드의 음상 정위

며, 식 (3.12)는 주파수 영역에서의 콘볼루션 식이다. 일반적으로 주파수 영역에서 콘볼루션을 수행할 때 연산 시간을 줄일 수 있으므로 대부분의 입체음향 시스템은 입력신호와 머리전달 함수를 주파수 영역으로 변환한 후에 콘볼루션하여 구현한다.

그림 2는 일반적인 모노음을 두 귀에 해당되는 머리전달함수와 각각 콘볼루션하므로 음상을 한곳에 정위시키는 시스템의 다이어그램이다.

이동음을 나타내기 위해서는 각 지점에 해당하는 머리전달함수를 원음과 연속적으로 콘볼루션하여 처리하게 된다. 그러나 머리전달함수 자체가 불연속적인 지점 상에서 측정된 것이므로 한 지점에서 다른 지점 사이의 경계에서 클릭 현상이 일어날 수 있다. 이를 위하여 선형 보간법 등이 쓰일 수 있다.

이와 같이 음상정위 처리를 통하여 나오는 음은 청취자가 헤드폰을 사용하여 듣게 될 때 음상이 3차원 공간상의 임의의 위치에 놓여지게 된다. 그러나 무향실에서 녹음된 원음의 경우 음상이 머리 안에 맺히게 되기 때문에 이를 방지하기 위하여 인공 잔향기를 부과한다.

### 3.2 음장효과(Sound field effect)

실내에서 음을 방사시키면 그림 3에 나타난

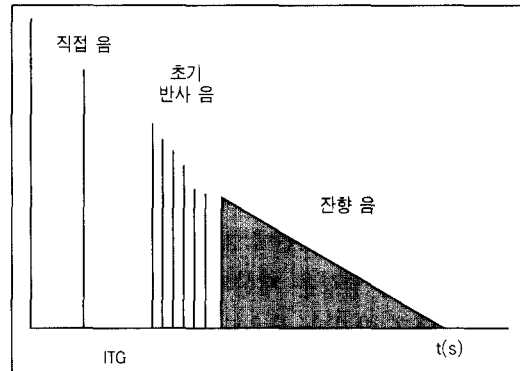


그림 3. 실내에서의 반사음 분류

바와 같이 직접음 이외에 다수의 반사음이 겹쳐진다. 이들 반사음은 반사면의 성질에 따라 진폭이나 위상이 변화되고, 동시에 직접음에 비해 시간적으로 늦게 도달하는 성질이 있다. 실내에서 박수 소리와 같이 지속 시간이 짧은 음을 내면, 그 반사음 그룹의 울림을 들을 수 있다. 이것이 잔향(reverberation)이며, 반사음이 분리되어 들릴 경우에는 에코(echo)라고 한다.

잔향은 청취자로 하여금 음의 공간감을 느끼게 하는데 중요한 역할을 한다. 이미 이 분야에 대한 연구가 많이 진행 되어있다.

잔향은 음원이 정지된 후에 60dB 감쇠할 때까지의 시간을 잔향 시간이라 하지만, 실제로는 잔향감은 반드시 잔향 시간에 반비례하지 않는다. 잔향시간은 식(3.2.1)과 같이 용적 V에 비례하고 실내 표면적 S와 평균 흡음률에 반비례한다.

$$T = K \frac{V}{aS} \quad (3.2.1)$$

여기서 V는 실내의 체적, S는 실내 표면적,  $\bar{\alpha}$ 는 평균 흡음률, K는 상수이다. 따라서 잔향시

간을 2배로 하기 위해서는  $V$ 를 두 배로 하거나  $\alpha$ 를 반으로 하면 된다. 즉, 공간을 두 배로 늘리거나 흡음 처리하여 평균 흡음률을 절반으로 한다는 의미이다. 그런데 이 양자는 완전히 다른 물리 현상이며, 같은 잔향 시간이라도 청감적인 인상은 전혀 다를 수 있다.

잔향기는 주로 초기반사음(early-reflection) 생성기, 콤필터(comb filter), 전역통과필터(allpass filter)로 이루어져 있는데 룸 모델링에 따라 각각에 영향을 주는 delay time과 gain값에 차이가 있다. 잔향기는 주로 룸 모델링을 얼마나 잘하느냐에 따라 잔향기의 질에 영향을 미친다.

먼저 초기반사음 생성기는 1차 벽면을 반사하여 청취자에게 도달하는 음을 생성한다. 이는 음원으로부터 나온 음이 1차 반사체에 의하여 반사되어 타임  $m$  만큼 지연되고 열화(attenuated)되어 타임  $m$  후에 음원으로부터 나온 음과 함께 청취자의 귀에 입사되는 음을 모델화 한 것이다. 초기반사음은 음원으로부터 직접 도달하는 음을 보강해주므로 청취자로 하여금 음의 친밀감과 명료함을 주로 느끼게 하는 역할을 한다고 알려져 있다. 그림 4는 간단한 초기반사음 생성기의 구조를 나타낸다.

다음은 콤 필터로 초기 반사음 후의 잔향을

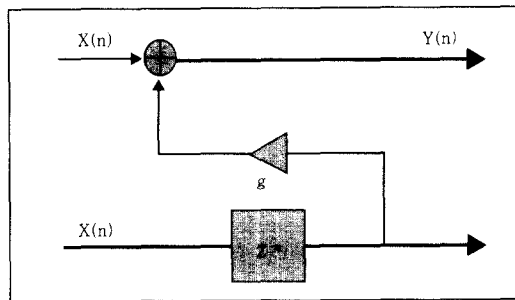


그림 4. 초기 반사음 생성기

모델화 한 것으로 음원으로부터 나온 음이 여러 가지의 장애물에 의하여 반사되어 감소되면서 피드백 되어 다시 청취자의 귀로 들어가는 음을 생성해 낸다. 음이 무한정 피드백 되는 것을 막기 위하여 중간에 이득( $g$ ) 값을 적용하여 피드백 되는 음을 감소 시킨다. 음이 지연되는 시간과 이득 값의 크기에 따라서 청취자는 음의 울림 현상을 크게 또는 작게 느낄 수 있다. 그림 5는 콤필터를 도식화한 그림이다.

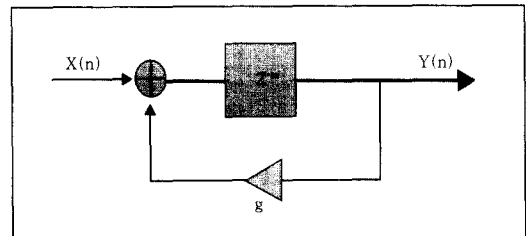


그림 5. 콤 필터

마지막으로 전역통과필터인데, 이 필터는 초기반사음 필터와 콤필터를 통해서 나온 잔향 성분이 복잡한 자연 세계에서의 음의 반향과는 달리 일률적인 성향을 갖게 되므로 이를 랜덤하게 섞어 주는 역할을 한다. 그림 6은 일반적인 전역통과필터의 도식도이다.

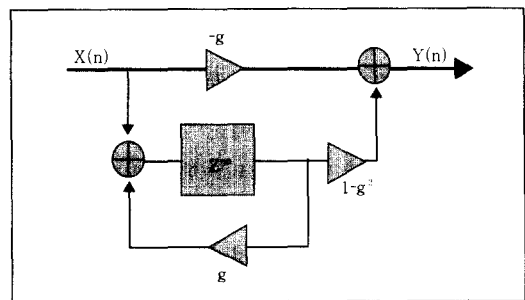


그림 6. 전역통과 필터

잔향기는 설계하는 사람마다 다르게 디자인 될 수 있다. 아래 그림은 일반적인 인공 잔향기의 모습으로 1개의 초기 반사음 생성기, 6개의 콤필터와 2개의 전역통과필터로 이루어져 있으며 그 구성은 그림 7 과 같다.

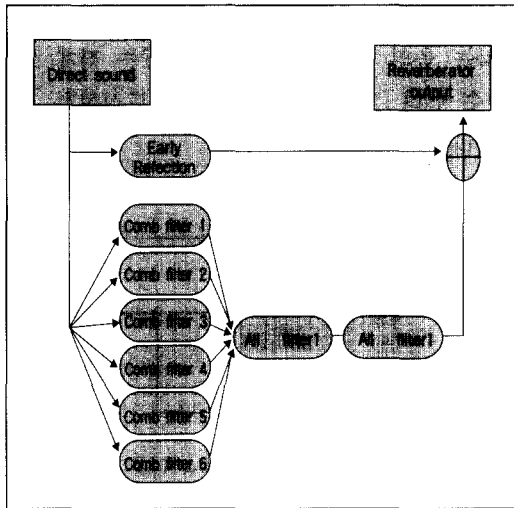


그림 7. 잔향기 구조의 예

### 3.3 크로스토크 제거(Crosstalk cancellation)

잔향기를 통과하여 출력되는 음은 헤드폰을 통하여 들을 때 그 음상이 머리 밖의 한곳에 정위 되는 것을 느낄 수 있다. 그러나 이를 스피커를 통해 듣는다면 크로스토크(crosstalk) 현상 때문에 그 느낌이 현저히 감소하게 된다. 이는 스피커를 통하여 듣게 될 때 왼쪽 귀에만 들려야 할 음이 오른쪽 귀에도 들리게 되고 마찬가지로 오른쪽 귀에만 들려야 할 음이 왼쪽 귀에도 들리게 되기 때문이다. 크로스토크 현상은 헤드폰이 아닌 스피커에서 입체음향의 완벽한 구현을 막는 최대의 걸림돌이라 할 수 있다. 이제까지의 연구 결과에서는 일반적인 자연 세계에서는 크로스토크를 완벽하게 제거

하는 것은 불가능 한 것으로 알려져 있다. 그러나 트랜스오럴 필터(Transaural filter)를 이용하면 어느 정도 크로스토크 현상을 막을 수 있다.

트랜스오럴 필터의 원리는 다음과 같다.  $x$  가 청취자의 귀에 도달하기 원하는 binaural 신호이면, 즉 스피커 X에서 방출된 신호를 원래의 신호 그대로 듣기 위해서는 전달 함수 행렬  $H$ 는  $\hat{x} = H^{-1}x$  와 같이 역 함수를 취해야 한다. 즉 스피커 X로부터 출력되는 신호  $x$ 와 귀 Y에 도착하는 신호  $y$ 와 같게 만들려면, 전달함수  $H$ 의 역함수를 구해야 한다.

트랜스오럴 필터를 구현하는 방법은 다음과 같다.

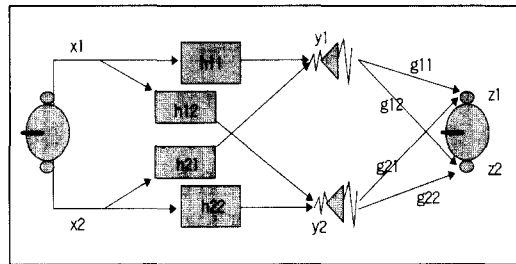


그림 8. 크로스토크 제거

그림 8 에서 스피커를 통해서 청취자의 오른쪽 및 왼쪽 귀에서 최종적으로 재생되는 신호를 각각  $z1(n)$ ,  $z2(n)$ 라 하면,

$$z1(n) = g11(n)y1(n) + g21(n)y2(n) \quad (3.3.1)$$

$$y1(n) = h11(n)x1(n) + h21(n)x2(n) \quad (3.3.2)$$

$$z2(n) = g22(n)y2(n) + g12(n)y1(n) \quad (3.3.3)$$

$$y2(n) = h22(n)x2(n) + h12(n)x1(n) \quad (3.3.4)$$

이 되며 식 (3.3.1)부터 식(3.3.4)까지를 다시 정리하면 다음과 같다.

$$z(n) = \begin{bmatrix} z1(n) \\ z2(n) \end{bmatrix} = \begin{bmatrix} g11(n) & g21(n) \\ g12(n) & g22(n) \end{bmatrix} \begin{bmatrix} y1(n) \\ y2(n) \end{bmatrix} = \begin{bmatrix} g11(n) & g21(n) \\ g12(n) & g22(n) \end{bmatrix} \begin{bmatrix} h11(n) & h21(n) \\ h12(n) & h22(n) \end{bmatrix} \begin{bmatrix} x1(n) \\ x2(n) \end{bmatrix}$$

따라서  $z1(n)$ 과  $z2(n)$ 에서의 신호가  $x1(n)$ 과  $x2(n)$ 의 신호와 같게 하기 위해서는 행렬  $H$ 가 행렬  $G$ 의 역행렬이 되어야 한다. 이를 수식으로 표현하면,

$$H(n) = G^{-1}(n) = \begin{bmatrix} g22(n)/g00(n) & -g12(n)/g00(n) \\ -g12(n)/g00(n) & g11(n)/g00(n) \end{bmatrix}$$

$$g00(n) = g11(n)g22(n) - g12(n)g21(n)$$

이렇게 트랜스오럴 필터를 사용하여 크로스토크 제거를 하는 경우, 주위 벽면에 의한 반사음이 전혀 없어야 한다는 조건이 있어야 한다. 따라서 청취자는 무향실이나 자유음장과 같은 가능한 한 반사음이 적은 곳에서 청취를 하여야 하며, 만일 청취자가 자유로이 움직이는 가운데 음상정위를 시키고자 한다면 헤드 트래커를 사용하여야 한다. 이는 실시간으로 청취자의 머리 위치를 추적하여 청취자가 머리를 움직일 때 마다 다른 트랜스오럴 필터 계수를 계산할 필요가 있으므로 이에 따른 연산량으로 인한 실시간 처리의 어려움이라는 또 하나의 문제가 있다.

### 3.4 청각 거리 조절

입체음향 처리는 음원의 거리 조작 역시 포함한다. 여기서 거리란 귀 축의 가운데 점에서

음상까지의 거리라고 정의한다. 음상의 거리가 머리의 반경보다 작은 경우, 바꿔 말하면, 음상이 머리 가운데에 생기는 경우를 머리 내 정위라 한다. 머리 내 정위에 관한 문제는 헤드폰에 의해 신호를 재생하는 전기 음향 시스템의 발전과 함께 생겨났다.

그림 9는 Bekesy가 실제의 회화 음성을 자극으로 하여 정면 방향 10m까지의 음상 거리의 대응을

구한 결과이다. 약 1.5m까지는 일치하고 있지만, 그 이상이 되면 차이가 커지고,

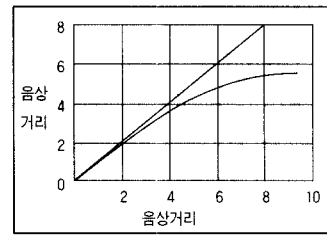


그림 9. 음원거리 대 음상거리의 비교

고, 음원 거리가 증가하여도 음상 거리는 그렇게 증가하지 않는 것을 알 수 있다. 음상은 광범위한 거리에는 생기지 않고, 청각 공간에는 한계가 있다는 것을 추론 할 수 있다. 그러나 짧은 거리에 있어서는 어느 정도 그 원근감을 소리의 크기로 표현할 수 있다. 음원의 출력이 일정하고, 그 거리를 바꾸면, 청취점의 음압 레벨이 변하여 소리의 크기가 바뀌게 된다. 소리의 크기는 거리의 제곱 단서로서 제일 먼저 들 수 있는 것으로 많은 연구가 있다. 물론 소리의 크기가 증가함과 동시에 음상 거리는 가깝게 지각된다. Gardner에 의한 실험에서 피험자의 정면 방향 3m에서 9m 사이에 등 간격으로 5개의 스피커를 설치하여, 피험자는 가장 가까운 스피커를 볼 수 있게 하였다. 녹음된 음성이 여러 종류의 음압 레벨이고, 3m와 9m의 위치에 있는 2개의 스피커에서 재생되었다. 피험자는 지각된 음상의 거리를 응답하도록 하였다. 그



결과, 음상의 거리는 음원의 거리와 관계없이 음압 레벨에만 의존하고 있는 것을 알 수 있었다. 자유 공간에서는 점 음원과 청취점 간의 물리적인 거리가 배가되면, 음압 레벨은 6dB 감소하는데 비하여, 실제적으로 사람이 느끼는 음상 거리를 배로 하기 위해서는 20dB 이상의 감쇠가 필요하다는 것을 나타내고 있다.

#### 4. 입체음향의 문제점

입체음향에 대한 연구는 이미 1930년대부터 진행되어 왔으며 그 구현방법 역시 정립된 지 오래라 할 수 있다. 그럼에도 불구하고 아직까지 실제적인 범용화가 느려지고 있는 이유는 입체음향 처리자체가 가지고 있는 몇 가지 중요한 문제점이 남아 있기 때문이다. 여기서는 이러한 문제점들 중 가장 대표적인 몇 가지만 언급하도록 하겠다.

먼저 음향 지각은 아직 완전히 구조화된 것이 아니므로 개발자는 입체음향을 디자인하기 위한 적당한 가격 대 효용비를 제대로 결정할 수 없다. 더군다나 시스템이 잘 작동하지 않을 때 개발자는 그 이유를 잘 설명할 수가 없다. 몇 가지 알려진 음향 지각 큐(양 귀간 음의 지연시간, head shadow, shoulder echoes, pinna effects 등)가 있으나, 그들 간의 상대적 중요도는 아직 알려져 있지 않다. 더욱이 아직 알려지지 않은 큐들이 얼마나 중요한 역할을 하는지도 모르고 있다. 다행히 현재 전 세계의 인지과학자들이 계속적으로 음상 지각에 대한 이해를 넓히고 있으므로 앞으로 이에 대한 연구가 진행됨에 따라 좀더 효율적인, 비용이 덜 드는 입체음향 시스템이 만들어 질 것으로 예상된다.

또 다른 문제점은 사람마다 머리전달 함수의

특성이 다르다는 것이다. 관련 논문들에 의하면 이러한 차이는 개개인의 머리 모양, 크기, 신장, 귓바퀴의 모양과 크기에서 온다고 밝혀져 있다. 때문에 다른 사람의 머리전달 함수로 처리된 입체음향을 청취 시 그 정확성이 크게 떨어지게 된다. 따라서 완벽한 입체음향을 구현하기 위해서는 사용자 고유의 머리전달 함수를 측정하여 이를 입체음향 프로세싱에 적용시켜야 한다. 그러나 이것은 사실상 불가능하다고 할 수 있다. 먼저 개인의 머리전달 함수를 측정하는 일이 매우 번거로운 일이고, 일반인이 사용하게 될 애플리케이션 프로그램으로서 사람마다 다양한 조건을 충족시키는 것은 어렵기 때문이다. 최근의 이 분야에서는 머리전달 함수의 차이에서 오는 음상정위의 오차를 최소화시키려는 연구가 진행되고 있다.

마지막으로 입체음향의 음질에 대한 문제이다. 머리전달 함수 자체가 측정 시 미디어의 특성에 심하게 좌우되며 입체음향을 재생할 때 사용하는 스피커에 따라서 음질이 크게 달라지는 것이 보통이다. 이로 인하여 머리전달 함수와 원음의 콘볼루션 시 음색의 변형을 가져올 수 있다. 따라서 이를 충분히 고려하여야만 완벽한 음질의 입체음향을 감상할 수 있다.

#### 5. 결 론

디지털 방송 시대가 눈앞에 온 현재, 영상的高화질화와 음향의 다 채널화 및 고음질화를 가장 중요한 요소로 삼고 있다. 특히 음향의 다 채널화와 고음질화는 실감음향을 그 목표로 하고 있다. 한 예로, 앞으로 디지털 방송의 음향 부분 표준으로 채택될 가능성이 유력한 음향 압축방식인 MPEG-2 AAC를 보더라도 기본적으로 5.1채널을 지원하고 있다. 종전의 실감음

향을 위해서는 각 채널당 여러 대의 스피커를 설치하여야만 입체 서라운드 효과를 느낄 수 있었다. 이는 일반가정의 경제 사정을 볼 때 공간적으로나 경제적으로 여의치 않으므로 주로 극장이나 콘서트홀에서만 가능한 시스템이었다. 그러나 입체음향 기법을 이용한다면 적은 비용과 좁은 공간에서도 극장에서나 느낄 수 있었던 웅장하고 실감나는 사운드를 즐길 수 있다. 이는 사람이 단지 2개의 귀로 입사되는 음향으로 모든 음감을 지각한다는 점을 감안하여 음원으로부터 인간의 고막까지를 전달 경로

를 모델링 하여 원음을 모델링 된 시스템에 통과 시켜 두 귀에 각각 들려주므로 마치 모델링 된 음장에서 들리는 음처럼 느끼게 해주는 실감음향기술에 있어서 최첨단 기술이라 할 수 있다. 그러나 이러한 청각 모델링이 상당히 주관적인 특색을 가지므로 모든 사람들에게 똑같은 효과를 느끼게 해 주기에는 한계가 있다. 요즘 이러한 청각 모델링의 오차를 낮추기 위한 여러 가지 연구가 진행되고 있으므로 곧, 범용 애플리케이션에 적용 가능한 시스템적인 청각 모델링 방법이 제안될 것으로 보인다.

● 참고 문헌 ●

- [1] 강성훈, 강경욱, "입체음향 Spatial Audio", 기전연구소, 1997
- [2] 이동우, 고대식, 강성훈, 김현빈, 김종민, "TMS320-C6x를 이용한 실시간 3차원 음성처리", 제 3회 한국음향학회 전기음향 학술대회 논문집, pp.188 - 191, 1999.
- [3] Durand R. Begault, "3D SOUND for virtual reality and multimedia" AP Professional, 1994.
- [4] Alan V. Oppenheim, Ronald W. Schaefer, John R. Buck, "DISCRETE-TIME SIGNAL PROCESSING -G", PRENTICE HALL, 1997.
- [5] John G. Proakis, Dimitris G. Manolakis, "Digital Signal Processing The Third Edition", Prentice Hall, 1996.
- [6] HRTF Measurements of a KEMAR Dummy-Head Microphone MIT Media Lab Perceptual Computing - Technical Report #280, Bill Gardner and Keith Martin, 1994.

필자 소개



홍진우

- 1982년 2월 광운대학교 응용전자공학과 졸업 (공학사)
- 1984년 2월 광운대학교 대학원 전자공학과 졸업 (공학석사)
- 1993년 8월 광운대학교 대학원 전자계산기공학과 졸업 (공학박사)
- 1998년 ~ 1999년 독일 프라운호퍼연구소 (교환연구원)
- 1984년 3월~현재 한국전자통신연구원 음향기술연구팀장
- 2000년 1월~현재 한국음향학회 홍보이사, 뉴미디어음향 학술분과위원장
- 주관심분야 : 오디오 신호처리 및 부호화, 디지털 콘텐츠 보호, 디지털 오디오 방송



최범석

- 1997년 2월 충남대학교 컴퓨터공학과 졸업(이학사)
- 2001년 2월 충남대학교 컴퓨터공학과 졸업 (이학석사)
- 2001년 3월~현재 한국전자통신연구원 음향기술연구팀 연구원
- 주관심분야 : 입체음향처리, 디지털 방송 콘텐츠 보호 시스템