주 제

오디오 신호처리 기술 동향

연세대학교 **박영철**

차 례

I. 서 론

2000년대에 들어 디지털 이동통신 단말기와 포터블 멀티미디어 기기를 중심으로 온라인 디지털 컨텐츠를 제공하는 새로운 형태의 서비스가 속속 등장하고 있다. DMB(Digital Multimedia Broadcasting)폰, PMP(Portable Multimedia Player) 등등이 이런 서비스를 가능하게 하는 주역들이다. 방송 분야에서는 이미 90년대에 기존의 아날로그 방송을 대체할 DAB(Digital Audio Broadcasting), HDTV(High-Definition TV) 등의 새로운 디지털 방송 방식이 등장한 바 있으며, 오디오 기기에 있어서는 CD(Compact Disk)나 DAT(Digital Audio Tape)와 같은 디지털 오디오가 사용된 지 오래이다.

디지털 오디오 기기와 달리 방송, 통신 등과 같이 대역이 한정되어 있는 응용 분야에서 많은 데이터를 전송하는 것은 커다란 문제였다. 전송과 저장에서의 문제점을 줄이기 위해서는 신호의 압축이 불가피하였다. 부호화 과정을 거쳐 압축된 오디오 신호는 복호화한 후의 주관적 음질이 기존 음질과 거의 동일하도록 유지되는 효과적인 압축 기법이 사용돼야만 한다. 이에 80년대 후반부터 세계 각국의 여러 연구소에서는 CD 수준의 디지털 오디오 신호를 지각적인 음질을 떨어뜨리지 않고 압축하는 기술, 즉 지금까지의 아날로그 오디오 방송을 대체할 새로운 디지털 오디오 처리 기술을 개발하였다[1][2]. 이러한 기술적인 배경으로부터 현재의 고품질 오디오 부호화 기술이 등장하게 되었다.

최근 가정용 오디오 시스템과 휴대형 멀티미디어 장치의 성능이 향상되면서 보다 현실감이 높은 멀티미디어를 서비스에 대한 요구가 높아지고 있다. 이를 위해 디지털 오디오 시스템에 적용할 수 있는 3차원 오디오 환경 합성 기술에 대한 활발한 연구가 이뤄지고 있다[3][4]. 이러한 기술은 소수의 멀티미디어 장치뿐만 아니라 궁극적으로는 방송, 통신, 원격 제어 등 다양한 분야에서 활용될 수 있다[8]. 특히 실감방

송을 위한 3차원 오디오 기술은 상당히 구체적인 형태까지 논의되어 있으며, 현재는 휴대형 단말기를 중심으로 기술이 본격적으로 응용되고 있는 상태이다.

최근의 3차원 오디오 신호 처리 기술들은 주로 청취공간을 모델링하기 위한 인공잔향기 기술과 청취자의 3차원 음향 인식 특성을 모델링하는 기술에 대한 것들이다. 청취자의 특성을 모델링 하는 과정에서는 머리전달함수를 사용하여 음의 방향성을 보존하려는 시도를 주로하기 때문에 어떻게 머리 전달함수를 잘 모델링하고 적용할 것인가에 대한 연구가 주요한 과제이다. 그러나 3차원 오디오 합성 기술을 구현하기 위해서는 그 밖에도 다양한 부수적인 신호 처리 과정들을 필요로 하며, 이러한 기술들을 본 글에서 다루고자 한다.

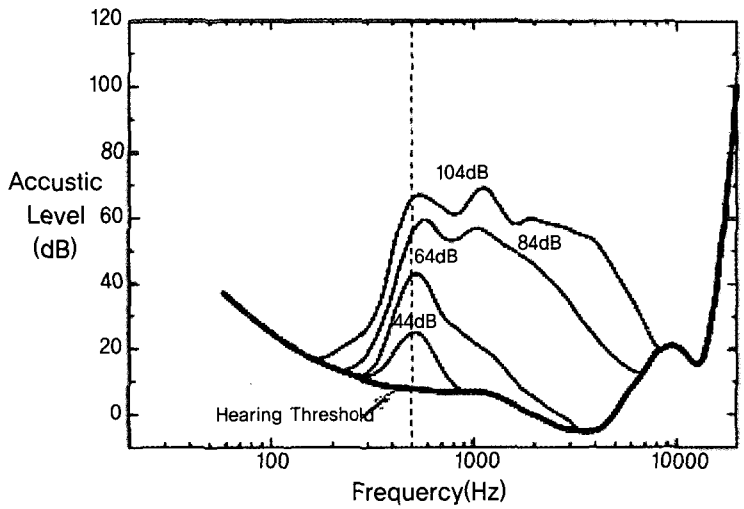
본 글에서는 디지털 오디오 장치를 위해 사용되는 최신의 디지털 오디오 신호 처리 기술에 대해 살펴본다. 특히, 고품질 오디오 부호화 기술, 3차원 오디오 구현을 위한 머리전달함수 모델링 기술, crosstalk 제거 기술, 그리고 인공 잔향 기술 등을 구성하는 신호 처리 방법들의 내용들을 살펴보고 기술적인 특징들을 관찰한다.

II. 고품질 오디오 부호화 기술

2.1 주파수 영역 마스킹의 원리

고품질 오디오 부호화 기술은 공통적으로 기존의 데이터 압축 기법에 인간의 청각 특성을 결합한 형태를 갖고 있다. 오디오 신호는 음원의 형태가 광범위할 뿐만 아니라 고음질을 필요로 하므로 음성 부호화와 같은 음원 발생 모델을 적용할 수 없다. 따라서 공통적인 수신원인 귀의 청각 특성을 이용하여 중복성을 제거하여야 하는데 여기에 주로 적용되는 특성이 마스킹(masking) 현상이다[5].

(그림 1)은 500Hz에 순음이 존재할 때 순음의 크기에 따른 마스킹 곡선을 보여준다. 그림에서 해당되는 크기의 500Hz 순음이 존재하면 곡선 이하의 신호



(그림 1) 500Hz 순음에 의한 주파수 영역 마스킹의 예

성분들은 사람의 귀에 들리지 않게 됨을 의미한다. 그림에서 굵은 선으로 나타난 곡선은 절대 가청 한계(hearing threshold)를 나타내는 것으로 그 이하의 음압을 내는 신호는 조용한 환경에서도 사람의 귀에 들리지 않음을 의미한다.

마스킹 현상은 주파수 영역에서의 동시 마스킹이 보다 효과적으로 적용될 수 있기 때문에 고음질 오디오 부호화에는 필터뱅크를 사용한 서브밴드 부호화나 DFT(Discrete Fourier Transform), MDCT (Modified Discrete Cosine Transform) 등을 사용하는 변환 부호화 방식이 사용된다[1]. 주파수 영역으로 변환된 오디오 신호는 심리 음향 모델(Psychoacoustic Model)로부터 얻어낸 마스킹 임계치에 근거하여 양자화된다[2].

2.2 MPEG 오디오 부호화 기술 표준화 역사 및 동향

국제 표준화 기구 ISO/IEC는 통신 및 저장 매체를 위한 디지털 오디오의 국제 표준을 제정하기 위해, 동화상 전문가 그룹인 MPEG(Moving Pictures Experts Group)을 구성하였다. 첫 번째 표준안으로 1992년 MPEG-1을 제정하고, 저장 매체를 위한 2채널의 오디오 신호에 대한 압축방식을 정의하였다[6].

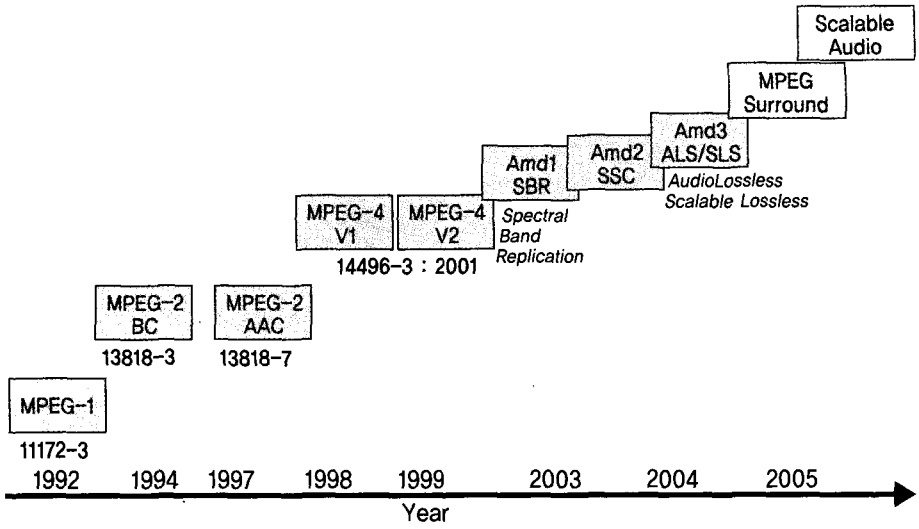
MPEG-1 오디오 표준안은 변화 부호화 방식을 사용하며, 구현의 복잡성과 음질에 따라 3개의 계층으로 나뉜다. 계층-1이 가장 간단하며, 계층-3이 가장 복잡한 반면 부호화 효율이 높다. MPEG-1은 모노와 스테레오 음원을 처리할 수 있으며, 32, 44.1, 48kHz의 샘플링 주파수를 지원한다.

MPEG-1은 1994년 MPEG-2 표준안으로 확장되었는데, MPEG-2는 5.1채널까지를 수용하며 96kHz까지의 샘플링 주파수를 지원하고, 약간의 확장성(scalability)을 지원하도록 하였다[7]. MPEG-2의 각 채널별 부호화 방식은 MPEG-1과 동일하고 채널간 매트릭싱을 사용하여 멀티채널 처리를 수행한다. 또한 비트열 포맷은 MPEG-1의 형식을 사용하여 확장된 멀티채널 데이터는 MPEG-1 비트열 포맷의 부가정보 부분에 들어가도록 설계하여 역방향 호환(Backwards Compatible)이 가능한 MPEG-2 BC 버전이 먼저 등장하였으며, 이 버전은 MPEG-1에 비해 낮은 표본화 주파수(16, 22.05, 24kHz)와 음성 다중을 지원한다.

MPEG에서는 MPEG-1 포맷과의 호환이 필요 없는 응용 분야에의 활용을 위한 권고안으로 1997년 새로운 멀티채널 오디오 부호화 방식의 국제 표준안으로 MPEG-2 NBC(Non-Backward Compatible)(이후 Advanced Audio Coding(AAC)로 명명되었다)를 발표하였다[8]. MPEG-2 BC 버전과 마찬 가지로 이 새로운 버전에서도 계층화된 구조를 사용하였는데, 음질, 메모리, 전력 요구량의 손익을 고려해서 세 가지 프로필(Main, Low Complexity: LC, Scalable Sampling Rate: SSR)을 지원한다. AAC는 매우 낮은 비트율에서 방송 음질 수준의 오디오를 제공하기 위해서, 고해상도 필터뱅크, 예측기법(prediction), 허프만 부호화 등을 결합하여 사용한다.

MPEG-4 표준화 기술은 MPEG-2 기술을 근간으로 부호화 효율을 혁신적으로 향상시킨 것으로, 다양한 부가기능을 추가하였다[9]. MPEG-4 표준안은 오디오 신호를 2kbps에서 64kbps까지의 비트율에서 최상의 품질을 얻을 수 있도록 확장성(scalability)을 일반화하였고, 동시에 여러 가지 부가기능을 지원하기 위해 파라메트릭 부호화, 음성 부호화 기술, 그리고 범용 오디오 부호화 기술을 하나의 구조안에 통합하였다.

MPEG-4에서는 MPEG-2 AAC를 근간으로 오디오 압축 성능을 향상시키기 위하여 PNS(Perceptual Noise Substitution), LTP(Long-Term Prediction)



(그림 2) MPEG 오디오 표준화 역사

와 같은 부가적인 툴을 지원하며, 매우 낮은 비트율에서 성능을 최적화한 TwinVQ, 부호화 지연을 감소시킨 LD(Low Dealy) AAC, 에러가 비교적 큰 채널을 위한 에러 복원(error-resilience) 툴을 지원함으로서 성능을 최적화하였다[10].

MPEG-4에서는 두 가지 방법으로 확장성을 제공한다. 첫 번째는 AAC Scalable로 불리우는 기술로 8kbps 단위의 비교적 큰 단위로 확장이 가능하며, 두 번째는 AAC의 무손실 부호화 기능을 채널 대역폭에 따라 1kbps 단위로 확장할 수 있는 기법으로 대신한 BSAC(Bit-Sliced Arithmetic Coding)[11] 부호화 방법을 통해서이다.

MPEG-4에서는 또한 비트율이 낮아지면 신호의 품질을 유지하기 위해서 신호의 대역폭을 제한하던 기존 부호화 방법의 문제점을 해결하기 위해 고주파 대역의 신호 성분을 낮은 주파수 대역의 스펙트럼과 고주파 대역을 표현할 수 있는 부가적인 파라메타들을 사용하여 추정하는 SBR(Spectral Band Replication) 툴을 정의하고 있다[12]. mp3PRO와 aacPlus는 SBR 기술을 기존의 MP3 및 AAC 기술과 각각 결합하여 부호화 효율을 높이고 제한된 대역의 전송에서 오디오 신호의 부자연스러움을 감소시킨 기술을 일컫는다. 일반적으로 SBR 기법은 고주파 신호 성분을 채널당 1~3kbps 정도의 정보량으로 묘사하기 때문에 기존의 T-F 변화 기반의 부호화 기법에 비해 효율이 훨씬 높다.

최근 전통적인 조인트 스테레오 기술을 확장하여 만들어진 BCC(Binaural Cue Coding)[13]는 SAC(Spatial Audio Coding)의 근간이 되는 기술로서, 관련 기술이 최근 'MPEG 서라운드' 라는 이름으로 MPEG에서 표준화가 진행 중이다. 이 기술은 다중 채널 오디오 신호를 하나 또는 두 개의 채널 신호로 결합된 신호에 인간의 지각특성을 의미하는 공간 단서(spatial cue)를 부가정보로 전송함으로써 현저히 낮은 전송률에서 다중 채널 신호를 전송할 수 있는 기술이다. 또한 다중 채널 신호로부터 역방향 호환이 되도록 기본 구조를 구성하였기 때문에 기존의 모노 채널이나 스테레오 채널을 수신할 수 있는 장치

와 완벽한 호환성을 제공한다. BCC에서 공간 단서로 제공하는 파라메타는 ICTD(Inter-Channel Time Difference)와 ICLD(Inter-Channel Level Difference), ICC(Inter- Channel Coherence) 정보이다[13].

음성-데이타 네트워크가 통합되기 시작한 이후, 오디오와 음성 대역에 걸쳐 확장성(scalability)을 가진 새로운 형태의 오디오 부호화 기술에 대한 요구는 끊임없이 제기되었다. 이런 요구를 만족시키기 위해 최근 MPEG에서 관심을 갖고 표준화를 시도하고 있는 분야가 Scalable Speech & Audio 부호화기이다[14]. Scalable 부호화 기술은 대략 최저 10kbps 정도 까지의 낮은 비트율 환경을 고려한 확장성을 요구하고 있으며, 현재 scalable 부호화 기술에 대한 여러 방향의 연구가 진행 중이며, MPEG 회의를 통해 새로운 형태의 기술이 제안되고 있는 상태이다.

III. 3차원 오디오 신호 처리 기술

3.1 3차원 오디오 신호 처리 기술의 원리

인간의 청각기관은 소리에 포함되어 있는 여러 가지 방향 정보를 이용하여 소리의 방향을 인지한다. 소리의 인지과정에 대해 아직 알려지지 않은 사실들이 많이 있으나 사람이 방향을 인지하는데 사용하는 중요한 단서(cue)들은 이미 오래전부터 알려져 있는 상태이다. 사람은 소리의 방향과 거리를 인지하기 위해 서로 다른 단서를 사용한다. 약 100년전 Rayleigh에 의해 개발된 Duplex 이론에 의하면 사람은 방향을 인지하기 위해 두 가지의 단서를 이용한다. 첫 번째는 ITD(Interaural Time Difference)이고 두 번째는 ILD(Interaural Level Difference)이다[15]. 이는 사람의 머리모양과 소리의 방향간의 기하학적인 관계에 의해 양 귀에서 지각되는 음의 도달 시간 차이와 음의 세기 차이에 의한 것이다. 양이간의 시간차이는 대략 1.2kHz정도 이하의 낮은 주파수 대역에서 인지되며, 주파수가 높아지면 시간 차이의 인지 능력이 떨어진다. 반면 사람의 머리 크기에 비해 파장이 상대적으로 큰 낮은 주파수에서는 양이간의 세기 차이가 나타나지 않으며, 1.2kHz 이상의 높은 주파수 대역에서 head-shadow 효과에 의해 세기 차이가 관찰된다. 이는 사람이 방향을 인지하는 과정에서 ITD와 ILD를 주파수 대역에 따라 상호 보완적으로 사용한다는 사실을 의미한다.

사람의 소리의 높이를 인식할 때 사용하는 단서는 귓바퀴의 모양에 의해 생기는 놋치(notch) 주파수에 의한 것으로 알려져 있으며, 대체로 사람을 음원의 높이를 인식할 때는 양이간의 정보차이를 이용하지 않는다[16]. 놋치 주파수는 사람의 귀모양에 따라 6~12kHz 대역에 걸쳐 광범위하게 분포하기 때문에, 높이 단서를 재생하는 것이 방향 단서를 재생하는 것보다 일반적으로 더 어렵다.

소리의 거리를 인식하는데 사용되는 단서는 라우드니스, motion parallex, 직접음과 반향의 비 등이다[17]. 라우드니스는 가까울수록 커지나, 절대적인 라우드니스에 따라 상대적인 거리를 인지하는 데는 한계가 있다. Motion parallex는 음원의 거리에 따라 머리의 방향을 움직였을 때 달라지는 방향감의 변화를 인지함으로써 거리를 판별한다는 이론이다. 한편 실내 공간에서 음원의 거리를 판별할 때는 주로 직접음과 반향의 비를 이용하는데, 가까운 음원의 경우 이 비가 상대적으로 크며, 멀리 떨어진 음원의 경우에는 이 비가 작아지게 된다.

3.2 3차원 오디오 시스템

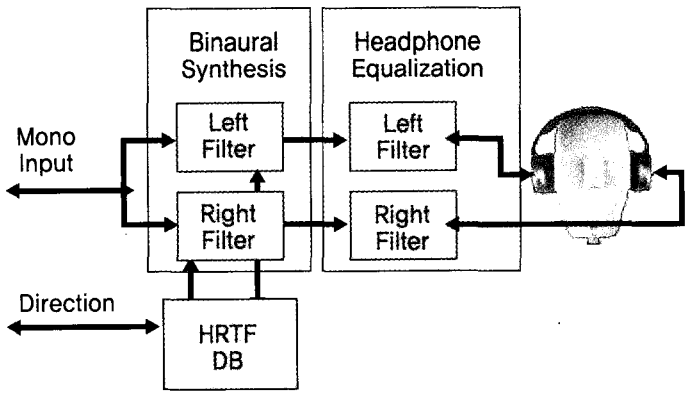
3차원 음장을 재생하기 위해 사용될 수 있는 가장

간단한 시스템은 두 개의 스피커 혹은 한 쌍의 헤드폰으로 구현되는 스테레오 시스템이다. 헤드폰을 사용하는 시스템은 마네킹의 귀에 마이크를 설치하여 무향실 환경에서 녹음한 스테레오 신호 (binaural recording)나, 또는 모노 음원에 특정 방향에 해당하는 단서를 첨가하여 만든 스테레오 신호(spatial rendering)를 헤드폰을 통해 직접 들려줌으로써, 고막에 전달되는 음압을 녹음환경에서와 동일하게 만드는 것을 목적으로 한다. 헤드폰을 사용하는 시스템을 흔히 binaural 시스템이라고 부른다[18].

헤드폰은 장시간 착용에 따른 사용자의 불편함과 헤드폰 자체의 음향 특성에 따른 왜곡이 발생할 가능성이 있으며, 소리가 귀에 너무 가까운 위치에서 전달되기 때문에 음원의 위치가 실제보다 더 가까이 들릴 가능성이 있다. 이런 문제를 해결할 수 있는 방법이 두 개의 스피커를 사용하는 transaural 시스템이다[18]. 두 개의 스피커로 구성된 transaural 시스템에서 3차원 오디오를 재생하기 위해 해결해야하는 문제는 스피커에 의해 재생되는 소리간의 crosstalk이다. 이는 두 개의 스피커에 의해 재생된 소리가 왼쪽 귀와 오른쪽 귀 모두에 전달됨으로써 청취자 고막에 전달되는 음압을 독립적으로 제어하지 못하게 하는 문제를 일으킨다. 이 문제를 해결하기 위해 transaural 시스템에서는 crosstalk 제거기를 스피커 입력단에 달아서 양이에 전달되는 소리를 분리하여 제어한다. Crosstalk 제거 기술에 대해서는 다음 절에서 설명하기로 한다.

3차원 오디오를 다중 채널 시스템에서 재생하는 경우 재생된 음장의 정확도를 너욱 높일 수 있다. 다중 채널 재생기로는 Ambisonics[19]와 WFS(Wave Field Synthesis)[20]를 예로 들 수 있다.

Ambisonics의 목표는 원래 음원이 위치한 음장의 spherical 하모닉 성분으로 청취자의 머리가 위치한 곳의 음장을 재생해 내는 것이다. 원리는 임의의 위치에서의 음장은 spherical harmonic expansion으로 나타낼 수 있으며, 하모닉 계수를 알 수 있다면 이를 이용하여 임의 공간에서 같은 음장을 재생할 수 있도록 재생 시스템의 출력을 계산할 수 있다는 것이다. 그러나 이 방법을 적용하기 위해서는 spherical harmonic을 측정할 수 있는 음장 마이크(sound field microphone)가 필요하다. 이를 구현하기 위해 1차의 harmonic을 이용하는 방법과 고차의 하모닉



(그림 3) Binaural 3차원 오디오 시스템

을 이용하는 방법이[19] 제시된 바 있다.

호이겐스의 원리(Heugens' s principle)에 의하면 음원에 의해 생성된 특정 점에서의 음압은 음원과 청취점 사이에 위치하는 평면에서의 수직 입자 속도(normal particle velocity)와 음압의 분포를 이용하여 재구성할 수 있다. 즉, 음원과 청취점 사이에 특정 평면에서 수직 입자 속도와 음압을 측정하고, 이를 여러 개의 이차 음원을 통해 다시 발생시킴으로써 원래 음원에 의해 발생한 음장을 재구성할 수 있다. WFS는 이런 이론적인 근거를 바탕으로 하고 있다.

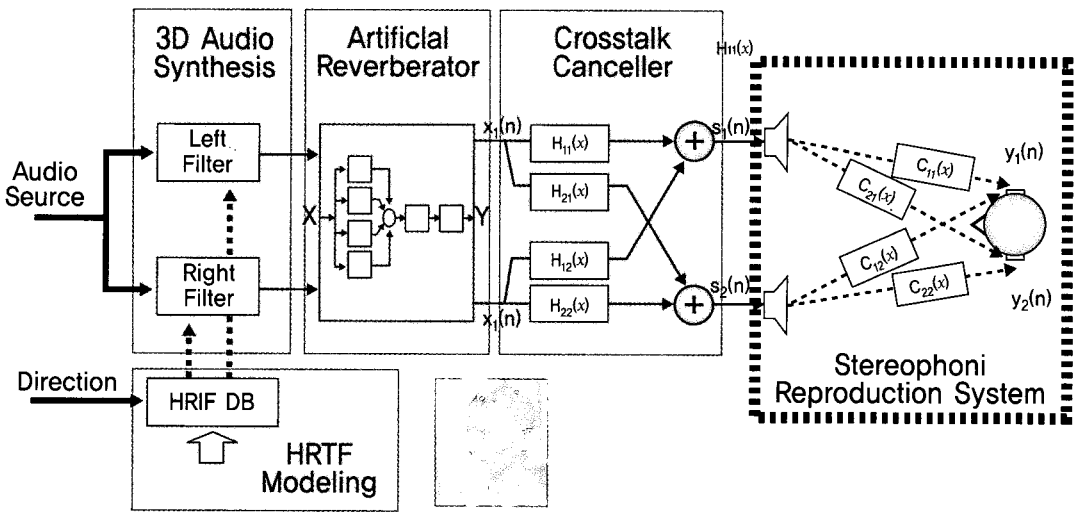
WFS에서는 여러 개의 마이크를 이용하여 음장을 녹음하고 이를 extrpolation 과정을 거쳐 스피커 어레이의 입력으로 사용함으로써 원음장을 재생한다. 이 기술의 장점은 특정 지점에서의 음장을 복원하는 것이 아니라 일정 공간에서의 음장을 재생하기 때문에 넓은 영역에서 3차원 효과를 볼 수 있다는 것이다. 그러나 이 기술은 많은 수의 스피커 어레이를 사용해야만 하며[20], 이런 점이 시스템을 구성하는데 큰 걸림돌이 되며 결과적으로 구현성을 떨어뜨리는 약점이 되기도 한다.

3.3 3차원 오디오 구현을 위한 기술 요소

(그림 4)는 transaural 시스템에서 사용될 수 있는 3차원 오디오 신호 처리 기술들 간의 연결 관계를 보여주고 있다. 모노 형태의 입력 오디오 신호를 주어진 방향 정보에 해당하는 왼쪽 귀와 오른쪽 귀의 머리 전달함수를 이용하여 왼쪽 채널 신호와 오른쪽 채널 신호를 만든 후, 청취공간의 음향 특성을 가상으로 구현할 수 있는 인공 잔향기를 거친다. 이렇게 렌더링된 신호는 transaural 시스템에서 발생하는 스피커 출력간의 크로스톡을 제거하기 위한 크로스톡 제거기를 거쳐 두 개의 스피커 채널을 통해 최종적으로 재생된다.

3.3.1 머리 전달함수(HRTF) 필터의 설계

머리전달함수를 낮은 차수의 디지털 필터로 모델링하는 과정에서는 기본적으로 오차의 제곱을 최소



(그림 4) Transautal 3차원 오디오 재생 시스템의 신호 처리 블럭들

화 하는 자승 평균(LS) 방식과 오차의 최대값을 최소화 시키는 Chebyshev 방식, 그리고 헨켈 행렬을 이용하여 시스템에 기여하는 중요도가 낮은 부분들을 잘라내는 BMT(Balanced Model Truncation)[21] 등의 방식을 사용한다.

머리전달함수를 모델링하는 필터의 구조로 FIR과 IIR 필터를 모두 고려할 수 있다. FIR 필터 모델은 측정된 머리전달함수의 임펄스 응답으로부터 특정 윈도우를 씌움으로써 간단하게 구할 수 있다[22]. FIR 필터로 머리전달 함수를 구현하게 되면 프로세스가 간단한 장점이 있지만 필터의 차수가 높은 단점이 있다(보통 44.1kHz 샘플링 율에서 100~200차 정도).

머리 전달함수의 IIR필터 구현은 ARMA(Auto regressive moving average)모델링 과정을 통하여 구현된다[22]. 대표적인 방법으로는 Prony 와 Yule-walker 모델링 방법이 있는데, 시간축에서의 위상정보가 중요한 경우에는 Prony의 방법이, 그리고 주파수 축에서의 스펙트럼 크기 차이의 오차를 줄이는 것이 목적인 경우에는 Yule-walker의 방법이 각각 좋은 특성을 나타내는 것으로 알려져 있다. 또한, 머리전달함수가 minimum phase 특성을 갖도록 필터를 추정할 수도 있다.

먼저, 머리전달 함수로부터 원래와 동일한 크기 스펙트럼을 가지는 minimum phase부분을 추정해 낸 뒤, 양쪽 귀의 임펄스 응답사이에 존재하는 시간차이만큼을 보상해줌으로써 minimum phase 필터를 추정할 수 있다. 이 때, minimum phase부분을 추정해 내기 위해서는 z-도메인에서 maximum phase부분을 보상해 주거나 켑스트럼(cepstrum)을 이용하는 방법등이 사용된다.

3.3.2 인공 잔향기 기술

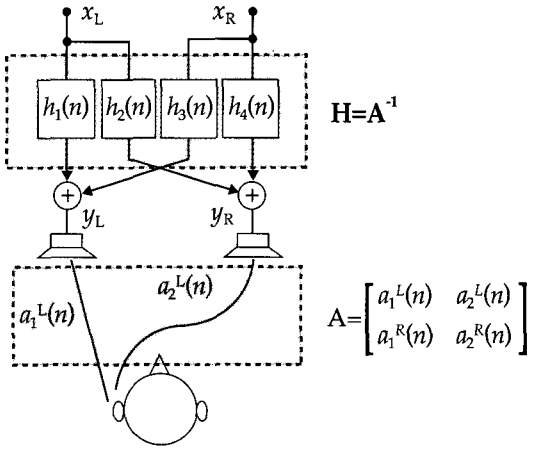
인공 잔향기(artificial reverberator)란 실내에서 발생하는 음의 반사과정을 간단한 구조의 디지털 필터들을 사용하여 모델링하는 기술을 말한다. 인공잔향기의 구조는 크게 초기반사음 모델링 부와 후기 반사음 모델링 부로 나뉜다. 초기 반사음은 흔히 FIR 필터 구조를 이용하여 모델링하는데, 긴 지연 메모리를 사용하여 초기 반사음이 발생하는 위치에 탭을 달고 적당한 계수를 곱해 줌으로써 초기 반사음을 만들어 낸다[23]. 초기 반사음 패턴은 특정 공간에서 발생한 실제 초기반사음 패턴을 이용하기도 하고, ray tracing 방법이나 image source method 방법을 통해 얻어진 가상 공간의 초기 반사음 패턴을 사용하기도 한다.

실내 공간에서 확산된 후기 반사음을 효율적으로 모델링하기 위해 사용하는 기본적인 유닛은 콤(comb) 필터와 전역통과(all-pass)필터이다. 자연스러운 확산 반사를 위해서는 잔향이 시간영역에서 충분한 밀도를 가져야 하고, 주파수 영역에서 공진주파수 사이의 간격이 좁아야 한다. 하나의 콤 필터와 전역통과 필터로는 반향의 밀도가 매우 낮아 위와 같은 조건을 충족시키지 못한다. 반향의 밀도를 증가시키기 위한 방법으로 콤 필터를 병렬로 연결하는 방법과 전역 통과 필터를 직렬로 연결하는 방법을 사용한다. 1960년대 Schroeder[24] 이후로 여러 학자들이 콤, 전역통과 필터 두 가지의 기본유닛으로 여러 가지 조합을 만들어서 잔향기를 구현하여왔다.

그러나 콤 필터를 병렬로 연결하는 구조는 근본적으로 충분한 시간영역 반향 밀도를 제공할 수 없다. 이는 자연스러운 후기반사음을 얻기 위해 반드시 해결해야 하는 문제이다. 이 문제를 해결할 수 있는 구조가 FDN(Feedback Delay Network) 구조이다[25]. 이 구조는 병렬로 배치된 콤 필터를 일반화 시킨 것으로써, 상대적으로 작은 메모리를 가지고도 높은 시간영역 반향 밀도를 갖는 후기반사음을 만들어 낼 수 있다. (그림 5)는 FDN을 이용한 인공잔향기의 구조의 예를 보여준다.

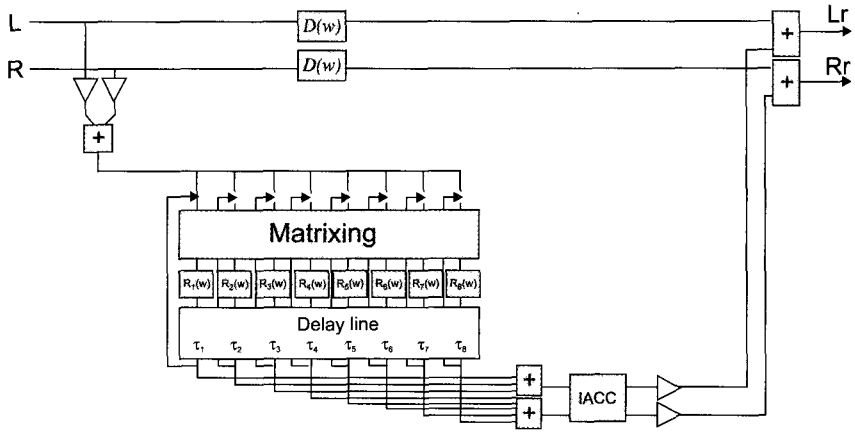
3.3.3 Crosstalk 제거 및 역 필터링 문제

(그림 6)은 Atal과 Schroeder에 의해 처음 제시된 crosstalk 제거 네트워크의 구조이다. 그림에서와 같이 만약 왼쪽 스피커(y_L)와 왼쪽 귀 오른쪽 귀 간의 경로를 각각 $a_1^L(n)$과 $a_2^L(n)$을 계수로 갖는 FIR 필터라 가정한다면, crosstalk 제거 문제는 어떻게 청취자의 왼쪽 귀, 오른쪽 귀 위치에서 원 신호 x_L과 x_R을 재생할 수 있느냐로 요약된다[26]. 이를 수식으로 정리하면 $\mathbf{A}^{-1}$를 구하는 간단한 역 필터링 문제로 귀결된다.



(그림 6) Atal-Schroeder의 crosstalk 제거 네트워크

Crosstalk 제거기는 위에 정의된 행렬 역함수로 주어지기 때문에, 행렬 A의 최소 고유치와 최대 고유치의 비인 condition number[26]에 의해 crosstalk 제거기의 안정성이 결정된다. 낮은 주파수 대역에서나 스피커와 양 귀간의 경로차가 작은 경우, 행렬 A가 singular해지는 문제가 발생한다. 스피커를 ±5°에 배치한 경우의 condition number를 구해보면 비교적 넓은 주파수 대역에서 crosstalk 제거기의 특성이 안정적인 것을 알 수 있다. 이는 stereo-dipole[27]이 3kHz 이상의 대역에서 매우 이상적인 특성을 보인다는 것을 입증하는 것이다. 그러나 낮은 주파수에서의 성능은 스피커 간격이 넓을수록 안정적인 특성을 보이는 것을 확인할 수 있다. 따라서 stereo-dipole의 경우 높은 주파수 대역에서는 매우 안정적이나 저주파 대역에서의 성능은 상당한 한계를 가지게 된다. Crosstalk 제거기에서 강인한 crosstalk의 제거 능력을 제공하는 일정 공간을 "equalization zone"은 또



(그림 5) FDN을 이용한 인공 잔향기의 구조 예

는 "sweet spot"이라고 하는데, 최소의 시스템으로 이 공간을 늘리는 것이 가장 어렵고 힘든 문제이다. 이를 위해 쓸 수 있는 방법 중 하나는 여러 위치에서의 동시에 최적화한 필터를 설계하여 사용하는 것이다[28].

녹음실에서 레코딩된 음장을 멀티 채널 재생 시스템을 사용하여 임의의 청취 공간에서 재생한다면, 특별한 신호처리 기술을 적용하지 않는 한 완벽한 녹음 환경의 원 음장을 청취자의 귀주변에 재생할 수 없다. 첫번째 이유는 위에서 언급한 바와 같이 재생 채널 간의 음향적인 간섭(crosstalk)이 생겨 음장을 복잡하게 변형시키기 때문이며, 두 번째는 청취자 귀 주변에 형성된 음장은 레코딩 공간의 잔향 특성에 재생 공간의 잔향특성이 복합된 음장일 수밖에 없기 때문이다. 그리고 세 번째는 재생 시스템, 특히 스피커의 배치 및 주파수 특성이 완벽하지 않기 때문이다. 역 필터링(inverse filtering)[29]과정이 이러한 문제를 해결하기 위해 사용되는 기술이다.

3.3.4 그 밖의 신호 처리 알고리즘

일반적으로 스피커 환경에서는 음상이 머리 밖에서 위치하게 되는 반면, 헤드폰을 사용하는 경우에는 음상이 머리 안에 맺히는 현상이 발생한다. 이러한 현상은 헤드폰 환경에서 3차원 오디오 신호처리 기술을 적용하여 머리 밖 특정 위치에 음상을 맺히게 하는 경우 중요한 문제가 된다. 이 문제를 정확하게 해결하기 위해서는 스피커로부터 청취자의 귀까지의 개인화된 전달함수(HRTF)가 반영되어야 하고, 헤드폰부터 청취자 귀까지의 전달함수가 올바르게 제거되어야한다. 그러나 이 방법은 청취자마다 일일이 전달함수들을 측정해야 하는 번거로움이 있으며, 실제 응용 시스템에도 적합하지 않기 때문에 이를 대체할 수 있는 다른 방법들이 연구 되어왔다. 그 중에 가장 효과적으로 알려진 것은 초기 반사음이나 잔향을 이용하여 상을 머리 밖에 맺히게 하는 기법이다[30].

PC나 이동통신 단말기 등에서 사용하는 보급형 스피커는 저주파 재생 능력이 현저히 떨어진다. 이런 스피커에서 저주파 음을 효과적으로 재생하기 위해 단순히 저주파의 신호를 증폭시키면, 소리의 심각한 왜곡이 발생할 뿐만 아니라 스피커나 앰프 등의 기기에 손상을 가져올 수 있다. 이러한 제약조건 하에서 저주파음을 재생할 수 있는 방법이 psychoacoustic 베이스라 불리우는 기술이다[31]. Philips의 UltraBass, SRS Lab.의 TruBass, Virtual Bass 등은 약간의 차이가 있긴 하지만 이 기술에 바탕을 두고 있다. 사람은 낮은 피치의 톤을 그것이 고조파 만으로도 지각할 수 있다는 "missing fundamental" 현상을 이용한 것이다[5]. 우리의 뇌는 기본 주파수 성분 없이 그것의 하모닉 성분만으로도 같은 음으로 인식할 수 있다. 이를 이용하면 낮은 피치에 대응하는 낮은 주파수 신호의 고조파 성분들을 만들어 줌으로써 저주파 재생 능력이 떨어지는 스피커로도 풍부한 베이스 음을 지각하게 만들 수 있다.

그 밖에도 스테레오 컨텐츠를 5.1 채널 시스템에서 3차원 음감을 최대화하여 재생하기 위해서는 가상의 서라운드 채널 신호를 적절하게 생성하기 위한 up-mix 기술에 대한 연구도 활발히 진행되고 있다[32].

IV. 결 론

고품질 오디오 압축 기술은 지난 30년간 발전을 거듭해 오고 있다. 압축 기술의 발전은 멀티미디어 컨텐츠 서비스 패러다임을 바꾸는 근간이 되었으며, MP3, AC3, AAC 등등은 이제 특정 기술을 일컫는 단어라기보다는 어디서나 흔히 접하는 일반 명사가 되어 버렸다. 그러나 공간적인 정보를 보존할 수 있는 기술이나, 음성과 오디오에 상관없이 고품질의 성

능을 제공하는 scalable 부호화 기술 등에 대한 산업적인 요구는 아직도 새로운 기술 개발에 대한 동기로 작용하고 있다. 특히 3차원 오디오 기술은 멀티미디어 컨텐츠의 가치를 현저히 높일 수 있는 수단이 될 뿐만 아니라, 사용자의 삶의 질을 높일 수 있는 중요한 사회적인 도구로 사용될 수 있을 것이다.

[참 고 문 헌]

[1] P. Noll, "MPEG Digital Audio Coding," IEEE Signal Processing Magazine, vol. 14, no. 5, pp. 59-81, Sept. 1997.

[2] M. Bosi and R. E. Goldberg, Introduction to Digital Audio Coding and Standards, Kluwer Academic Publishers, 2003.

[3] U. Zolzer, DAFX - Digital Audio Effects, John Wiley & Sons, Ltd., 2002.

[4] W. G. Gardner, 3-D Audio Using Loudspeakers, Kluwer Academic Publishers, 1998.

[5] E. Zwicker and H. Fastl, Psychoacoustics, Springer-Verlag, 1990.

[6] ISO/IEC JTC1/SC29/WG11 No.71 "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbit/s - CD 11172-3 (Part 3. MPEG-Audio)"

[7] ISO/IEC 13818-3:1998, Information technology -- Generic coding of moving pictures and associated audio information -- Part 3: Audio.

[8] ISO/IEC 13818-7:1997, Information technology -- Generic coding of moving pictures and associated audio information -- Part 7: Advanced Audio Coding (AAC).

[9] ISO/IEC 14496-3:2001. Information technology -- Coding of audio-visual objects -- Part 3: Audio.

[10] B. Grill, "The MPEG-4 General Audio Coder," *Proc. AES 17th Conf. on High Quality Audio Coding*, Florence 1999.

[11] S. H. Park, Y. B. Kim, S. W. Kim, et al., "Multi Layer Bit Sliced Bit Rate Scalable Audio Coding," *103rd AES Convention*, New York 1997, Preprint 4520.

[12] M. Dietz, L. Lijeryd, K. Kjorling, O. Kunz, "Spectral Band Replication, a Novel Approach in Audio Coding," *112nd AES Convention*, Munich 2002, Preprint 5553.

[13] J. Herre, et al., "Spatial Audio Coding: Next-generation efficient and compatible coding of multi-channel audio," *117th AES Convention*, San Francisco 2004, Preprint 6186.

[14] ISO/IEC JTC1/SC29/WG11 w7040, "Call for Information on Scalable Speech and Audio Coding," Jan. 2005, Hong Kong CN.

[15] J. Blauert, Spatial Hearing, MIT Press, 1982.

[16] M. B. Gardner, "Some Monaural and Binaural Facets of Median Plane Localization," *J. Acoust. Soc. Am.*, vol. 54, no. 6, pp. 1489-1495, 1973.

[17] M.A. Gerzon, "The design of distance panpots," *92nd AES Convention*, Vienna March 1992, Preprint 3308.

[18] Jean-Marc Jot, Veronique Larcher and Olivier Warusfel "Digital Signal Processing Issues In the Context of Binaural and Transaural Stereophony," *98th AES Convention*, Paris

1995, Preprint 3980.

[19] J. S. Bamford, "An analysis of ambisonic sound systems of first and second order," M.S. thesis, Univ. Waterloo, Waterloo, ON, Canada, 1995.

[20] U. Horbach and M. M. Boone, "Future transmission and rendering formats for multichannel sound," *AES 16th Int. Conf. on Spatial Sound Reproduction*, Helsinki 1999, Preprint s59711.

[21] J. Mackenzie, J. Huopaniemi, V. Valimaki and I. Kale, "Low-Order Modeling of Head-Related Transfer Functions Using Balanced Model Truncation," *IEEE Signal Processing Letters*, vol. 4, no. 2, pp. 39-41, 1997.

[22] Jyri Huopaniemi and Nick Zacharov and Matti Karjalainen, "Objective and Subjective Evaluation of Head-Related Transfer Function Filter Design," *105th AES Convention*, Aug. 1998. Preprint 4805.

[23] W. G. Gardner, Reverberation algorithms in Application of Digital Signal Processing to Audio and Acoustics (eds. M. Kahrs, and K. Brandenburg), Kuwer Academic Publishers, Boston/Dordrecht/London, 1998.

[24] M. R. Schroeder, "Natural sounding artificial reverberation," J. Audio Eng. Soc., vol. 10, no. 3, pp. 219-223, July 1962.

[25] John Stautner and Miller Puckette, "Designing multi-channel reverberators," *Computer Music Journal*, vol. 6, no. 1, pp. 52-65, Spring 1982.

[26] Darren B. Ward and Gary W. Elko, "Effect of Loudspeaker Position on the Robustness of Acoustic Crosstalk Cancellation," *IEEE Signal Processing Letters*, vol. 6, no. 5, May 1999

[27] O. Kirkeby, P. A. Nelson, and H. Hamada, "The :" Stereo Dipole" - A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers," *J. Audio Eng. Soc.*, vol. 46, no. 5, pp. 387-395, May 1988.

[28] Darren B. Ward, "Joint Least Squares Optimization for Robust Acoustics Crosstalk Cancellation," *IEEE Trans. on Speech & Audio Proc.*, vol. 8, no. 2, pp. 211-215, February 2000

[29] P. A. Nelson, F. O. Bustamante, and H. Hamada, " Inverse Filter Design and Equalization Zones in Multichannel Sound Reproduction," *IEEE Trans. Speech Audio Processing*, vol. 3, no. 3, pp. 185-192, May 1995.

[30] Per Rubak, "Headphone Signal Processing System for Out-of-Head Localization", *90th AES Convention*, Jan. 1991, Preprint 3063.

[31] Eric Larsen and Ronald M. Aarts "Perceiving low pitch through small loudspeakers," *108th AES Convention*, Jan. 2000, Preprint 5151.

[32] C. Avendano, and J.-M. Jot, "Ambient Extraction and Synthesis From Stereo Signal for Multi-channel Audio Up-mix," *IEEE Int. Cont. Acoustics, Speech Signal Proc, (ICASSP)*, vol. 2, pp. 1957-1960, Orlando, May 2002.

박영철

1986년 연세대학교 전자공학과 졸업
1988년 연세대학교 대학원 석사
1993년 연세대학교 대학원 공학박사
1993년 ~ 1995년 Graduate Program in Acoustics, Pennsylvania State Univ. PostDoc.
1996년 ~ 1998년 삼성전자 반도체사업부, 의공학 연구소 선임연구원
1998년 ~ 2001년 (주)인타임 시스템 LSI 연구소 연구소장
2002년 ~ 현재 연세대학교 컴퓨터정보통신공학부 교수
관심분야 : 3D 오디오 신호 처리, 적응 필터, 오디오/음성 부호화, 디지털 보청기