

오디오 신호처리 기술 동향

윤대회(연세대학교 전기전자공학과)

I. 서론

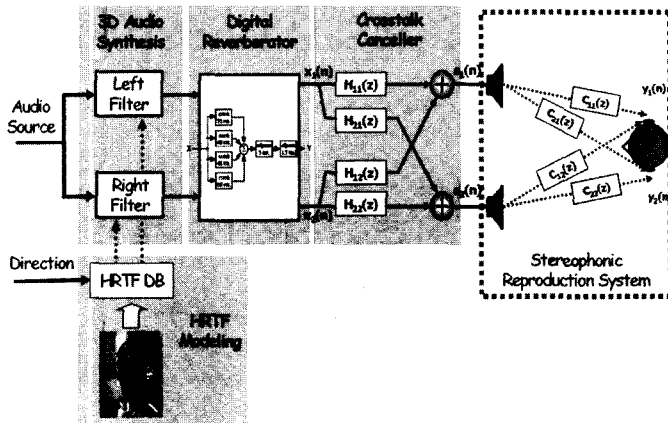
기존의 아날로그 오디오에 비해 향상된 음질과 넓은 대역폭을 갖는 디지털 오디오는 CD나 DVD 등을 비롯하여, 고선명 TV, 디지털 오디오 방송(DAB) 등 가정용 엔터테인먼트 시스템에서부터, 포터블 오디오 재생기, 이동통신 단말기 등 개인 휴대 장치에 이르기 까지 그 응용분야가 다양하다. 또한 디지털 오디오는 영화나 컴퓨터와 같은 다른 매체와의 호환성이 뛰어나기 때문에 그 응용분야는 계속 확장되고 있다[1][2][3].

최근 가장 많은 관심을 받고 있는 MP3의 경우를 살펴보더라도 휴대용 MP3 오디오 플레이어 뿐만 아니라 PDA, 핸드폰, 카 스테레오 등 멀티미디어 장비의 기본 요소가 됨으로써 1999년부터 세계시장에 휴대용 플레이어가 보급된 이후에 지금까지 약 1천만대 정도가 생산될 만큼 지속적으로 시장이 확대되고 있다.

최근까지의 디지털 오디오 기술에 대한 연구는 고품질 오디오 신호를 효율적으로 압축하기 위한 부호화 기술에 주로 집중되어 왔다. 이는 한정된 채널의 대역폭, 제한된 저장용량, 최소의 비용 등과 같은 멀티미디어 장치가 당면한 제한

조건을 만족하는 것이 시급하였기 때문이다. 심리음향 모델을 사용하여 음원을 효과적으로 부호화하기 위한 디지털 오디오 부호화 기술은 MPEG(Moving Pictures Experts Group)을 중심으로 발전을 거듭해 왔으며, 최근에는 보다 다양한 형태의 오디오 부호화 기술이 표준화되었고 상용화 시스템에까지 적용되고 있다.

한편 멀티미디어 장치의 성능과 효율이 향상되고, 전송 채널의 대역폭이 광대역화하면서 멀티미디어 과거의 오디오 품질을 담보로 한 요구 조건보다는 콘텐츠의 고급화와 실감성 향상이 좀 더 중요한 요소로 인식되고 있다. 사실 다수의 음원을 3차원 환경에서 적절한 규격의 재생 시스템을 사용하여 임의의 위치에 자유롭게 배치하는 기술은 오디오 콘텐츠를 녹음하고 재생하는 분야에서 아주 오래된 관심사이다. 그러나 이는 기본적으로 다수(4개 혹은 그 이상의) 스피커를 사용하는 재생 시스템을 필요로 하며, 또한 막대한 자원과 계산능력을 가진 하드웨어를 필요로 하기 때문에 상용 시스템을 사용하여 구현하는데 어려움이 있었다. 그러나 컴퓨터 하드웨어 및 소프트웨어 기술의 발달과 함께 시각, 청각적인 정보를 3차원의 형태로 가상 연출할 수



〈그림 1〉 3차원 오디오 모델링 시스템의 전체적인 블록도

있는 가상현실(virtual reality)의 구성이 용이해지면서 3차원적인 시청각 정보를 획득하고 재생할 수 있는 보다 효과적인 기술 개발에 대한 필요성이 더욱 높아졌다. 이러한 요구는 좀더 나가서 2개의 재생 채널을 가진 오디오 시스템을 사용하여 3차원적인 효과를 얻을 수 있는 기술의 개발에 대한 필요성을 더욱 높하게 되었다. 특히 상용화 규모가 큰 고품질의 오디오를 제공하는 소규모의 홈엔터테인먼트 시스템이나 개인휴대형 단말기에서 3차원 오디오 재생 기술에 대한 수요가 매우 높다.

3차원 오디오 표현 기술은 크게 세 가지의 모델링 과정을 거친다. 첫 번째는 음원 자체에 대한 모델링으로 자연음과 합성음을 이용하기도 하고 음원의 형태에 따른 방사형태에 대한 모델링을 하는 과정이다. 음원의 모델링 후 자연음을 부호화하기 위해 심리음향 모델을 적용한 오디오 부호화 기술이 사용될 수 있다[4][5]. 두 번째는 음을 청취하는 공간에 대한 모델링이다. 이 과정은 청취 공간의 잔향(reverberation) 특성을 반영하게 되는데, 사용되는 기술은 지각적으로 동일한 잔향을 간단한 디지털 필터로 근사화

는 인공 잔향기(artificial reverberator) 기술이다 [6][7]. 마지막 과정이 청취자에 대한 모델링으로, 청취자가 두 귀로 음의 형태와 방향 및 거리를 인식하는 과정을 반영하는 것이다[8]. <그림 1>에 3차원 오디오 구현 시스템의 구성 요소들을 도식화하였다.

두 개의 재생 채널을 사용하는 오디오 시스템에서 필요로 하는 3차원 오디오 신호 처리 기술들은 주로 청취공간을 모델링하기 위한 인공 잔향기 기술과 청취자의 3차원 음향 인식 특성을 모델링하는 기술 등이다. 청취자의 특성을 모델링한 과정에서는 특히, 머리전달함수(HRTF: head-related transfer function)를 주로 사용하기 때문에 어떻게 머리 전달함수를 잘 모델링할 것인가에 대한 연구가 주요한 과제이며[9], 이러한 모델링 과정을 거쳐 3차원 오디오를 구현할 때 발생하는 문제 채널 간의 crosstalk을 제거하는 기술[10][11]도 중요한 연구 대상이다.

본 논문에서는 디지털 오디오 장치를 위해 사용되는 최신의 디지털 오디오 신호 처리 기술에 대해 살펴본다. 특히, 최근 음원 부호화 기술의 동향과 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. 오디오 부호화 기술

1. 오디오 부호화 기술의 발전

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 MPEG-1을 1991년 제정하고, 저장 매체를 위한 2채널의 오디오 신호에 대한 압축 방식을 정의하였다[4].

MPEG-1 오디오 표준안은 구현의 복잡성과 비트율에 따라 계층-I, 계층-II, 계층-III으로 나누어진다. 이미 유럽에서는 기존의 FM 방송을 대체할 디지털 오디오 방송 계획(EUREKA-147)의 표준 오디오 압축 알고리즘으로 Philips, CCETT, IRT 등에서 제안한 MUSICAM (Masking-pattern adapted Universal Subband Integrated Coding And Multiplexing) 방식을 채택했는데 이 방식이 MPEG-1의 기본 알고리즘으로 선정되었다[12]. 한편 Dolby 연구소도 1991년 현재 북미 지역의 HDTV 오디오 압축 기술의 표준으로 채택된 AC-3 부호화 기술을 제안하였다[13]. AC-3는 SMPTE 권고안인 5.1채널을 수용하고, 32~640 kbps의 비트율을 갖는다. 압축비를 높이기 위해서 채널간 혹은 채널 내의 마스킹 특성을 이용하며, 고주파 대역의 채널 커플링을 이용하여 보다 낮은 비트율을 제공한다. 기본적인 부호화 방법은 시간축 에일리어싱 제거 방법에 기초한 MDCT 변환 부호화 방

식을 사용하였다.

MPEG-1 부호화 기술의 문제점은 HDTV와 같은 방송 매체에 적용할 경우 화질이 떨어질 뿐 아니라 오디오에 있어서도 2채널이상을 지원하지 않아 멀티채널, 음성 다중 등의 많은 부가 서비스에 미흡하다는 것이었다. 이를 해결하기 위해 ISO/IEC는 방송 등의 응용 목적에 적합한 MPEG-2 표준안을 확정하였다[5]. MPEG-2의 각 채널별 부호화 방식이 MPEG-1과 동일하고 채널간 매트릭싱을 사용하여 멀티채널 처리를 수행하며, 또한 비트열 포맷은 MPEG-1의 형식을 사용하여 확장된 멀티채널 데이터는 MPEG-1 비트열 포맷의 부가 정보 부분에 들어가도록 설계하였기 때문에 역방향 호환(Backwards Compatible)이 가능하다. 즉, MPEG-1 복호화기로 MPEG-2 비트열을 복호화할 수 있다. 또한 MPEG-2는 낮은 표본화 주파수, 음성 다중이 지원된다.

MPEG에서는 MPEG-1 포맷과의 호환이 필요없는 응용 분야에의 활용을 위한 권고안으로 1997년 새로운 멀티채널 오디오 부호화 방식의 국제 표준안으로 MPEG-2 AAC(Advanced Audio Coding)를 발표하였다[14]. AAC 표준은 전 세계의 진보된 오디오 부호화 기술을 가진 기업들이 협력한 결과이다. 음질, 메모리, 전력 요구량의 손익(tradeoff)을 고려해서 AAC 시스템은 세 가지 profile (Main profile, Low Complexity profile, Scalable Sampling Rate profile)을 지원한다. AAC는 매우 낮은 비트율에서 방송 음질 수준의 오디오를 제공하기 위해서, 고해상도 필터뱅크(high-resolution filter bank), 예측기법(prediction), 허프만 부호화(Huffman coding) 등을 결합하여 사용한다.

MPEG-4 오디오는 T/F 변환에 기반한

MPEG-2 AAC를 근간으로 오디오 압축 성능을 향상시키기 위한 확장 기능들 (Perceptual Noise Substitution, Long-Term Prediction)과 매우 낮은 비트율에서의 동작을 위한 (TwinVQ), 부호화 지연을 감소시킨 (Low Dealy AAC), 에러가 비교적 큰 채널을 위한 (error-resilience tools) 기능을 지원한다. 또한 비트율 가변성을 향상시킨 (Bit-Sliced Arithmetic Coding) 기법을 지원하여, 채널 대역폭에 따라 가변적인 비트율로 전송이 가능하게 되어 있다.

mp3PRO와 aacPlus는 오디오 신호의 고주파수 성분과 저주파수 성분간의 상관관계에 착안하여 고안된 SBR(Spectral Band Replication) 기술을 기존의 MP3와 AAC 표준과 결합하여, 부호화 효율을 높이고, 제한된 대역의 전송에서 오디오 신호의 부자연스러움을 감소시킨 최신 오디오 신호 압축 기술이다. SBR 부호화기는 저주파수 영역의 오디오 신호와 고주파수 영역의 오디오 신호간의 상관관계를 부가정보로 만들어 전송한다.

MPEG-4 BSAC은 채널당 1kbps 간격으로 fine grain scalability를 제공할 수 있다. 하나의 Base Layer 비트스트림과 여러 개의 작은 enhancement 비트스트림으로 구성되어 있다. 또한 이동통신망이나 DAB 등 error가 있는 환경에서 압축 효율에만 중점을 두어 개발된 codec에 비해서 오디오의 음질을 향상시킬 수 있다. MPEG-4 BSAC은 한국의 지상파 DMB의 오디오 신호를 압축하는 표준으로 선정되었다.

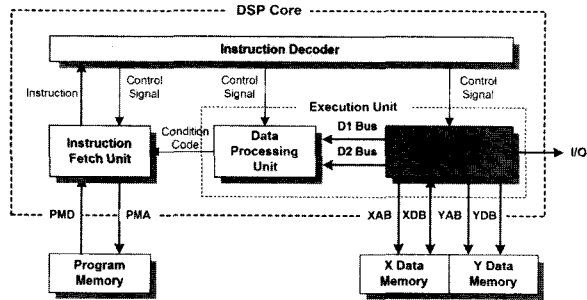
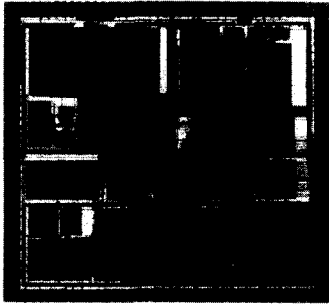
2. 오디오 부호화 기술의 구현 기술

디지털 오디오 부호화/복호화기는 응용분야에 따라 다양한 하드웨어 플랫폼이 고려되고 있

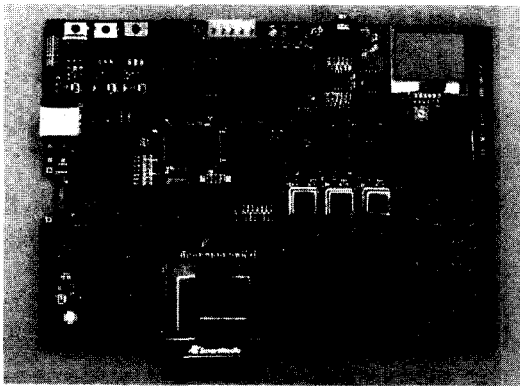
다. 즉, 범용 DSP[15][16], 오디오 전용 DSP 코어[17], 단일 RISC 코어 [18][19][20][21], DSP Core with embedded MCU[22], RISC Core with minimized DSP accelerator 등이다.

각 플랫폼을 구성하는 프로세서는 범용 DSP 프로세서, RISC 프로세서(또는 MCU), 오디오 전용 DSP 프로세서이다. 범용 DSP 프로세서를 사용하는 구현 방법은 비용면이나 효율면에서 불리한 점이 많다. 범용 DSP의 경우 범용성을 위해서 특정 응용분야에는 필요하지 않은 많은 모듈들을 포함하고 있기 때문에, 휴대용 기기와 같은 저전력 시스템을 구성하기에는 비효율적이다. 이러한 비효율성은 ASIC을 이용한 오디오 전용 DSP 프로세서를 설계하고 구현함으로써 해결할 수 있다. 특정 알고리즘에 적합한 구조를 선택하고 꼭 필요한 하드웨어 모듈로만 구성함으로써 보다 효율적이며 저전력의 프로세서를 설계할 수 있다. 반면 단일칩 RISC 프로세서를 이용한 구현의 경우, 충분한 캐쉬 메모리가 확보하는 것이 관건이다. 블록 단위의 신호처리를 요구하는 디지털 오디오의 경우, 캐쉬 메모리가 부족한 경우 잦은 캐쉬 미스로 인하여 수행 속도가 현저히 저하되는 단점이 있다.

최근에는 RISC 프로세서와 최소의 DSP 기능 블록을 하나의 칩으로 제작하여 디지털 오디오 부호화/복호화기를 구현하는 예가 증가하고 있다. 이것은 MAC 연산을 빈번하게 요구하는 FFT, MDCT, filtering 등의 연산중심 과정은 DSP를 이용하여 고속으로 처리하고, 제어중심 과정은 RISC 프로세서에 할당하여 오디오 신호 처리의 효율성을 높이고 있다. 특히 SoC 기술의 발전으로 RISC 프로세서와 DSP 기능블록이 결합한 형태의 플랫폼은 휴대용 기기를 위한 멀티미디어 프로세서로 활용되고 있다.



〈그림 2〉 오디오 전용 DSP 프로세서의 구조



〈그림 3〉 오디오 전용 DSP 프로세서로 구현된 MP3/AAC 복합 복호화기

1. 머리 전달 함수를 이용한 3차원 오디오 합성

소스로부터 3차원 효과의 사운드를 만들어내기 위해 사용되는 것이 방향성을 주는 머리전달 함수(HRTF: Head-Related Transfer Function)이다. 방향성이 없는 소스를 원하는 방향의 머리 전달함수로 필터링 시켜주고 나면 사람이 들었을 때 마치 그 방향에서 소리가 들리는 것처럼 느끼게 되는 것이다. 이러한 머리전달함수의 데이터는 무향실에서 더미헤드(dummy head)와 스피커를 이용하여 측정하며[25], 더미 헤드로부터 일정 간격 떨어져있고 특정한 방위각 θ 와 고도각 ϕ 에 위치한 소스로부터 양쪽 귀 고막까지의 임펄스 응답을 구하는 과정으로 이루어진다. 측정된 머리전달함수에는 머리주위의 반사와 산란도 고려되어져 있으며, 이러한 두 채널의 머리전달 함수가 아래식과 같이 방향성 없는 소스(S)와 컨볼루션되어 특정방향 (θ, ϕ) 의 3차원 사운드 (Z_L, Z_R) 를 만들어내게 된다.

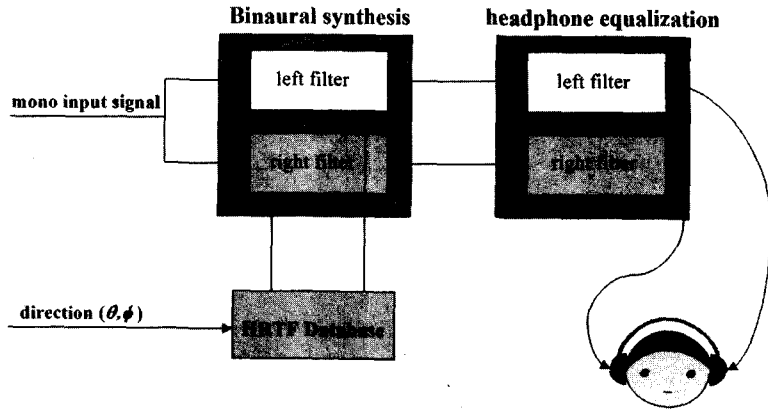
$$Z_L = S * HRTF_L(\theta, \phi) \quad (1)$$

$$Z_R = S * HRTF_R(\theta, \phi) \quad (2)$$

그런데 이러한 필터링 과정에서 고막속의 마

III. HRTF 모델링 기술

3차원의 사운드를 만들어 내는 것은 오디오 신호처리 분야에서 관심을 받고 있는 주제중의 하나이다[8]. 이러한 3차원 오디오 사운드를 만들어 내는데 있어 주요 관심사는 음색의 왜곡과 음질저하 없이 스테레오 헤드폰이나 스피커를 사용하여 자연스러운 방향성을 주는 것이다. 이렇게 만들어진 3차원의 가상 사운드는 스튜디오 녹음이나 컴퓨터 음악 등 여러 분야에서 사용되어 질 수 있다[23][24].



〈그림 4〉 헤드폰 환경에서 3차원 사운드 합성 과정

이크 또는 헤드폰, 스피커와 같은 특정 장비의 주파수 특성에 의한 왜곡이 생기게 되므로 <그림 4>와 같이 이를 보상해 주는 과정이 필요하게 된다[26]. 보상을 위한 필터는 주파수 영역과 시간 영역에서 설계가 가능한데, 주파수 영역에서의 설계 방법은 장비의 전달함수로부터 minimum phase 성분만을 따로 구해 역함수를 취해주거나[27] 최소 자승(LS: least square) 방법을 이용하여 시간 영역의 역필터를 구현하는 방법이 사용된다[28]. 한편, 머리전달함수를 설계하기 위해서는 함수가 방향을 구별해 내는 중요한 요소들, 즉, 양쪽 귀 임펄스 응답 사이의 시간 차이(IID: Interaural Time Difference), 레벨 차이(ILD: Interaural Level Difference), 컷바퀴 응답 등이 잘 보존될 수 있도록 하여야 한다[8].

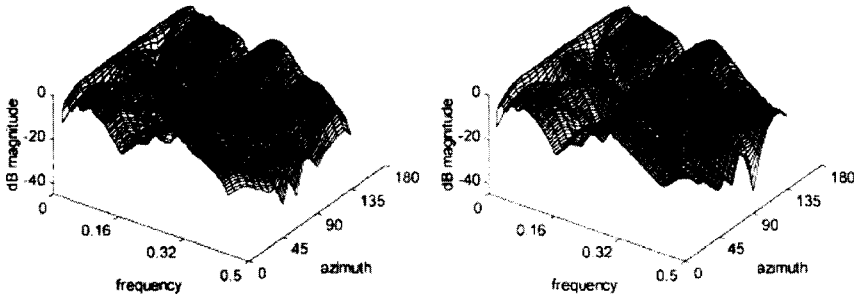
2. 머리 전달함수 필터의 설계

머리전달함수를 낮은 차수의 디지털 필터로 모델링하는 과정에서는 기본적으로 오차의 제곱을 최소화 하는 자승 평균(LS) 방식과[29][30] 오차의 최대값을 최소화 시키는 Chebyshev 방

식, 그리고 헨켈 행렬을 이용하여 시스템에 기여하는 중요도가 낮은 부분들을 잘라내는 BMT(Balanced Model Truncation)[31]등의 방식을 사용한다.

머리전달함수를 모델링하는 필터의 구조로는 FIR과 IIR 필터를 모두 고려할 수 있다. FIR 필터 모델은 측정된 머리전달함수의 임펄스 응답으로부터 특정 윈도우를 씌움으로써 간단하게 구할 수 있다[9]. 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. 대표적인 방법으로는 Prony와 Yule-walker 모델링 방법이 있는데, 시간축에서의 위상정보가 중요한 경우에는 Prony의 방법[32]이, 그리고 주파수 축에서의 스펙트럼 크기 차이의 오차를 줄이는 것이 목적인 경우에는 Yule-walker의 방법이 각각 좋은 특성을 나타내는 것으로 알려져 있다[9]. 이렇게 구현된 IIR 필터는 FIR필터에 비해 훨씬 적은 차수로 거의 비



〈그림 5〉 좌 : 128차 FIR 필터 모델 , 우 : 10차 IIR 필터 모델

스한 성능을 낼 수 있음이 여러 연구에 의해 밝혀져 있다. 이는 <그림5>의 왼쪽은 128차의 FIR 필터로 머리전달함수(HRTF)를 모델링한 결과와 오른쪽은 10차의 IIR 필터로 모델링한 결과를 비교해보면 쉽게 알 수 있다[31].

한편, 머리전달함수가 minimum phase 특성을 갖도록 필터를 추정할 수도 있다. 먼저, 머리전달함수로부터 원래와 동일한 크기 스펙트럼을 가지는 minimum phase부분을 추정해 낸 뒤, 양쪽 귀의 임펄스 응답사이 존재하는 시간차이만큼을 보상해줌으로써 minimum phase 필터를 추정할 수 있다. 이 때, minimum phase부분을 추정해 내기 위해서는 z-도메인에서 maximum phase부분을 보상해 주거나 cepstrum(cepstrum)을 이용하는 방법등이 사용된다[33][34].

이렇게 필터를 디자인 하는 과정에서 심리음향 모델에 의하여 고주파쪽보다 저주파쪽의 피크(peak)와 나치(notch)부분이 더 중요하다고 가정한다. 따라서 가장 적합한 필터를 설계하기 이전에 주파수 축을 바크(Bark) 단위로 변환하거나 주파수 왜곡(warping) 방법으로 저주파쪽 성분에 초점을 맞추는 방법들을 사용하기도 한다[35].

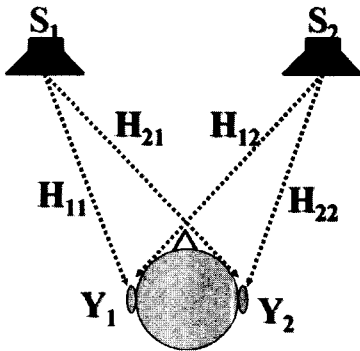
한편, 머리전달함수를 이용한 3차원 음향 시스

템의 구현성을 높이기 위한 방법들도 다양하게 제시되고 있다. 예를들면, 한쪽 귀까지의 머리 전달 함수로부터 다른쪽 귀의 머리 전달 함수를 추정해내는 방법[36]이나, 소수의 기본함수(Basis Function)의 선형 조합으로 모든 방향의 머리 전달함수를 만들어 내는 PCA (Principal Component Analysis) 방법[37]등이다.

그러나 이렇게 머리전달함수만을 적용하여 3차원 사운드를 만들어내게 되면 헤드폰 환경에서는 문제가 되지 않지만, 스피커 환경과 같이 소스로부터 왼쪽과 오른쪽 귀까지의 음향 경로(acoustic path)가 서로 간섭을 주는 경우에는 급격한 성능 저하를 가져오게 된다. 따라서 이를 해결하기 위해 crosstalk 제거 네트워크를 사용하여야 한다.

IV. Transaural 3차원 오디오 재생 시스템

특정 공간에서 스테레오로 레코딩된 음장을 재생하기 위해 헤드폰을 사용하거나(binaural system) 2 개의 스피커를 사용할 수 있다(transaural system)[25]. 헤드폰을 사용하는 경우(그림 4) 재생공간의 음향 특성을 배제할 수 있



(그림 6) 스테레오 재생 시스템

고, 각 채널의 독립적인 제어가 용이하기 때문에 비교적 우수한 효과를 연출할 수 있으나, 헤드폰을 반드시 착용해야 한다는 점과 음상의 위치가 청취자의 머리가 움직임에 따라 이동한다는 문제가 발생한다[25]. 머리 움직임과 무관하게 음상을 고정시키기 위해서는 머리의 움직임을 추적할 수 있는 장치가 필요한데 이를 위해 센서나 카메라를 이용한다[38].

헤드폰을 사용하는 경우와 달리 <그림 6>과 같이 2개의 스피커를 사용하여 3차원 음상을 재생하는 경우에는 몇 가지 추가적인 문제가 발생하는데 본 장에서는 두개의 스피커를 사용하는 transaural 시스템에서 발생하는 문제와 이를 해결하기 위한 신호 처리 기술에 대해 살펴본다.

1. 2개의 스피커를 이용한 3차원 오디오 재생 시스템

2개의 스피커를 사용하는 3차원 음상 재생 시스템에서 발생하는 문제는 크게 세 가지를 들 수 있다. 첫 번째 문제는 재생 채널 간의 음향적인 간섭(crosstalk)이 생겨 음상을 복잡하게 변형시키는 것이며, 두 번째 문제는 재생 시스템이 실

내에서 사용되는 경우 청취자 귀 주변에 형성된 음상은 재생 공간의 잔향특성이 복합될 수밖에 없다는 것이다. 그리고 세 번째 문제는 재생 시스템, 특히 스피커의 주파수 특성이 완벽하지 않기 때문에 발생한다[10].

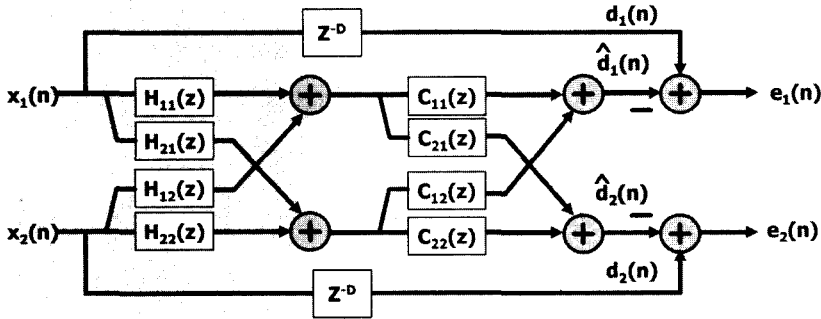
스테레오 채널 재생 시스템은 청취자의 귀 주변에서 레코딩된 원 음상을 완벽하게 재생해 내는 것을 목적으로 한다. 이러한 목적을 달성하기 위해 스피커 입력 단에 역 필터(inverse filter)를 연결하여 스피커에 의한 왜곡과 청취 공간의 음향적인 특성에 의한 변형을 보상한다. 이를 역 필터링(inverse filtering)[39]과정이라고 한다.

역 필터링을 위한 최적 필터 설계 방법은 크게 두 가지로 구분된다. 첫 번째 방법은 스피커 시스템을 통해 오디오를 재생하기 전에 미리 녹음된 데이터를 사용하여 최적의 디지털 필터를 설계하고 이 필터를 스피커 입력 단에 적용하는 방법이다. 이 방법은 멀티 채널 환경에 적용되는 필연적으로 역행렬 연산 과정을 거치게 되며, 역행렬을 연산하는 과정에서 발생할 수 있는 수치적인 불안정성을 야기할 수 있기 때문에 이에 대한 대책이 필요하다. 필터의 설계가 오프라인으로 수행되기 때문에 재생 시스템에 부가되는 계산량을 최소화할 수 있다는 장점이 있지만 주변 환경이 변화하는 경우에 대처하지 못한다는 단점이 있다.

스테레오 채널 역 필터링 문제를 도식화하면 <그림 7>과 같이 나타낼 수 있다. 그림에서 C 는 스피커 입력 신호 x 와 오차 마이크 위치에서의 전달함수, H 는 역 필터의 전달 함수, d 는 재생하고자 하는 기대 신호를 나타낸다. 마이크 출력 신호 e 은 다음과 같이 나타낼 수 있다.

$$e = d - \hat{d} \quad (3)$$

최적 필터 H 의 계수는 다음과 같은 마이크 출



(그림 7) 스테레오 재생 시스템에서의 역 필터링 문제

력 파워를 최소화하도록 결정된다.

$$\begin{aligned}
 J &= E\{e^T e\} \\
 &= E\{d^T d\} - 2E\{d^T X C h\} \\
 &\quad + h^T C^T E\{X^T X\} C h
 \end{aligned} \tag{4}$$

위 식에서 C 는 전달함수 $C(z)$ 에 해당하는 디지털 필터의 계수로 구성된 행렬이고, h 는 역 필터 $H(z)$ 의 계수로 구성된 벡터이다. 만약 입력 신호가 분산인 1인 백색잡음이고, 마이크 위치에서 재생하고자 하는 기대신호는 입력 신호가 단순히 시간 지연된 신호라고 가정한다면, 식(4)로부터 얻어지는 최적 계수벡터는 다음과 같다.

$$h_o = [C^T C]^{-1} C^T \delta \tag{5}$$

여기서, δ 는 L 만큼 지연 후에 '1' 값이 존재하고 나머지는 '0'인 벡터, 즉 $\delta = [0 \dots 0 \dots 1 \dots 0]$ 이다. 만약 C 가 정방행렬이고, 역행렬이 존재한다면 최적 계수벡터는 다음과 같이 간단하게 정리된다.

$$h_o = C^{-1} \delta \tag{6}$$

최적 계수벡터를 계산함에 있어 행렬 C 의 구조가 수치적인 안정성을 결정한다.

두 번째 방법은 적응 필터를 사용하여 주변 환경에 적응적으로 역 필터의 계수를 최적화하는 방법이다. 이런 목적을 위해 사용할 수 있는 대표적인 방법이 Multiple-Error LMS 알고리즘[10]

이다. 이 방법은 마이크 출력을 오차 신호로 되받아 출력의 파워를 최소화하도록 하는 계수를 적응적으로 찾아내는 방법으로 궁극적으로는 수식(5)의 계수벡터를 추정하도록 하는 방법이다. 최적화 방법과 다른 점은 역행렬 연산 과정이 아닌 순환적인 방법을 사용하기 때문에 수치적으로 안정한 방법이며, 환경의 변화를 고려한 시스템이라는 점이다. 그러나 순환적인 적응 과정을 통해 최적값에 수렴하기 위해서는 충분한 길이의 데이터가 필요하다는 단점과, 경우에 따라서는 추정 오차가 클 수 있다는 단점이 있다.

2. 역 필터링의 안정성

앞서 언급했듯이, 위에서 설명한 기존 알고리즘은 스피커의 위치와 주파수에 따라 crosstalk 제거 네트워크의 불안정하게 동작하는 문제점을 가지고 있다. 일반적으로 스피커의 위치에 따라 crosstalk 제거기의 안정성이 결정된다[40]. n 번째 스피커와 왼쪽 귀의 전달 함수를 다음과 같이 정의하자.

$$c_n^L = e^{j2\pi\lambda^{-1}d_n^L} \quad n = 1, 2 \tag{7}$$

여기서 s 는 신호의 파장, d_n 은 n 번째 스피커에서 왼쪽 귀(L)까지의 거리를 나타낸다. 스피커가 대칭적으로 구성되어 있는 환경을 가정하면,

$$c_2^L = c_1^R \text{ and } c_2^R = c_1^L \quad (8)$$

이 된다. a_n^R 은 a_n^L 과 유사하게 정의되는 n 번째 스피커와 오른쪽 귀 사이의 전달함수이다. 만약 n 번째 스피커와 오른쪽 사이의 거리를 $d_n^R = d_n^L + \Delta$ 로 표현하면 다음과 같다.

$$C = c_1^L \begin{bmatrix} 1 & e^{j2\pi\lambda^{-1}\Delta} \\ e^{j2\pi\lambda^{-1}\Delta} & 1 \end{bmatrix} \quad (9)$$

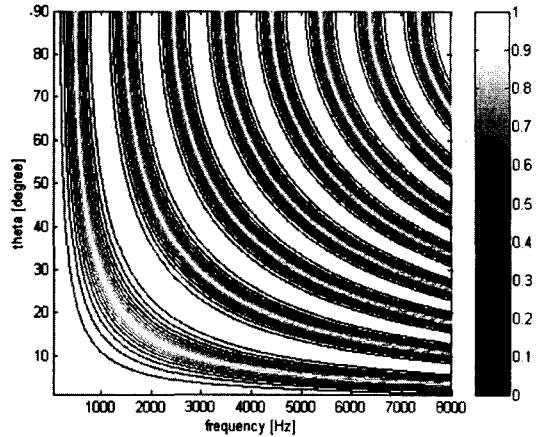
Crosstalk 제거기는 위에 정의된 행렬 C 의 역함수로 주어지기 때문에, 행렬 C 의 최소 고유치와 최대 고유치의 비인 condition number[41]를 체크함으로써 crosstalk 제거기의 안정성을 살펴볼 수 있다. 따라서

$$\Delta = \frac{\lambda}{4} + p\frac{\lambda}{2} \quad p : \text{integer} \quad (10)$$

일 때는 condition number가 가장 작은 값을 가지게 되고 crosstalk 제거기가 안정적인 특성을 가지는 반면,

$$\Delta = q\frac{\lambda}{2} \quad q : \text{integer} \quad (11)$$

인 경우에는 condition number가 가장 큰 값을 가지게 되고 crosstalk 제거기는 불안정한 특성을 가지게 된다. 다시 말하면, n 번째 스피커와 왼쪽과 오른쪽귀간의 경로차가 거의 없거나 ($\Delta \approx 0$) 입력 신호 반파장의 정수배인 경우에 crosstalk 제거 행렬(H)에 의해 만들어진 스피커 입력신호의 크기가 과도해져서 전체적인 시스템의 성능이 크게 떨어질 수 있다. 특히, 낮은 주파수 대역에서는 스피커와 양 귀간의 경로차와 상관없이 수식 (10)과 수식 (11)에 의해 행렬 C 가 singular해지는 문제가 발생한다.



<그림 8> 스피커의 위치에 따른 condition number의 역수

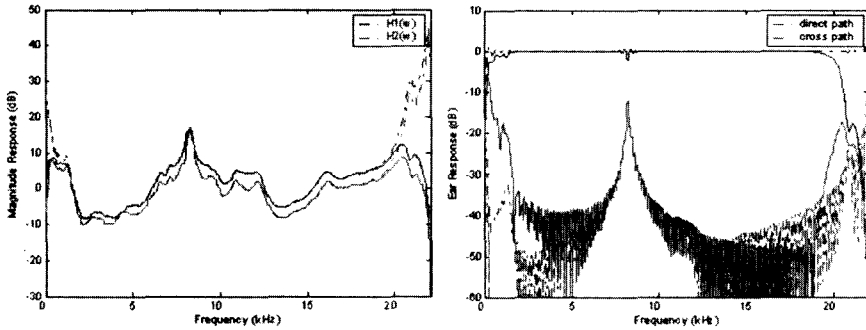
두 스피커와 귀간의 경로차 Δ 를 실제 스피커의 위치와 연관시키기 위해서 두 개의 스피커가 머리의 중앙위치에서, $(x, +\theta)$, $(x, -\theta)$ 의 위치에 있을 때를 가정하여 수치적으로 가장 안정적인 스피커의 위치와 가장 불안정한 스피커의 위치에 대한 수식을 얻을 수 있다.

$$\sin\theta = \frac{\lambda}{8r_H} = \frac{\lambda}{0.7} \text{ (best loudspeaker positions)} \quad (12)$$

$$\sin\theta = \frac{\lambda}{4r_H} = \frac{\lambda}{0.35} \text{ (worst loudspeaker positions)} \quad (13)$$

다음 그림은 스피커의 위치에 따른 condition number의 역수를 그린 것이다.

<그림 8>의 결과에서 보듯이, 스피커를 $\pm 5^\circ$ 에 배치한 경우(stereo-dipole) 비교적 넓은 주파수 대역에서 crosstalk 제거기의 특성이 안정적인 것을 알 수 있다. 이는 stereo-dipole[42]이 3kHz 이상의 대역에서 매우 이상적인 특성을 보인다는 것을 입증하는 것이다. 그러나 그림에서 볼 수 있듯이 스피커 간격이 넓을수록 더 낮은 주파수에서 안정적인 특성을 보이는 것을 확인할 수 있다. 따라서 stereo-dipole의 경우 높은 주파수 대



〈그림 9〉 Crosstalk 제거기의 성능비교 : (a) 스피커 입력 신호의 주파수 분포, (b) 왼쪽 귀와 두 입력 신호 간의 전달함수. 두 그림에서 점선과 실선은 각각 기존의 방법(식 5)과 regularization(식 17)에 의한 결과이다.

역에서는 매우 안정적이거나 저주파 대역에서의 두 채널 신호를 분리하는데 어려움이 있다.

3. 역 필터링의 안정성 개선 방법

역 필터링에 있어서 행렬의 ill-conditioned 문제는 특정 주파수 대역에서 최적 필터 계수를 구할 수 없게 만든다[43]. 특히 오디오 신호처리 분야에서는 이러한 이유로 특정 주파수 대역에서 매우 큰 이득이 발생하게 되어 신호처리의 장치들이 포화상태가 되거나 스피커나 앰프 등에 손상을 주게 된다. crosstalk 제거의 문제에서도 이 같은 문제가 발생하는데, 음향 경로의 역 필터를 구현하게 되면 특정 주파수 대역, 주로 저주파(<500Hz)와 특정 주파수 대역에서 매우 큰 이득이 발생하게 된다. 이런 역행렬 연산의 안정성 문제를 해결하기 위한 방법이 regularization이다. 이 방법은 최적 필터 계수를 구할 때 기존에 오차만을 비용 함수(cost function)에 포함시켜 자승 평균을 최소화하는 것과 달리 오차와 스피커의 입력을 함께 최소화함으로써 전달함수 행렬이 singular해지는 것을 방지하는 방법이다 [44]. 이 때 비용함수는 다음과 같이 정의된다.

$$J = E + \beta V \tag{14}$$

여기서 E는 오차의 에너지 총합이고, V는 스피커 입력 에너지의 총합이며, β 는 E와 V에 적절한 가중치를 주는 regularization 파라미터이다. 각 항을 자세히 표현하면 다음과 같다.

$$E = e^T e, \quad V = v_b^T v_b, \quad v_b = b * v \tag{15}$$

여기서 b는 regularization을 하고자 하는 주파수 대역을 선택하는 필터이며 이를 행렬식으로 표현하면 다음과 같다.

$$B = \begin{bmatrix} b(0) & 0 & 0 \\ \vdots & \ddots & 0 \\ b(N_b - 1) & \vdots & b(0) \\ 0 & \vdots & \vdots \\ 0 & 0 & b(N_b - 1) \end{bmatrix} \tag{16}$$

일반적으로 β 를 “gain factor”라 하고, b를 “shape factor”라고 하는데, 이 값들을 어떻게 선택하느냐에 따라 regularization하고자 하는 주파수 대역과 그 정도가 결정된다[45]. 이렇게 결정된 비용 함수를 최소 자승 방법을 이용하여 최적 필터계수를 찾으면 다음과 같다[44].

$$h = [C^T C + \beta B^T B]^{-1} C^T \delta \tag{17}$$

이 방법은 특정 주파수 대역에서 원치 않는 큰 이득이 나타나는 현상을 막아줌으로써 소리의 왜곡현상이나 스피커의 손상을 억제할 수 있는

장점이 있다. 그러나 regularization factor가 결국 역행렬의 정확도를 떨어뜨리기 때문에 가중치가 큰 주파수 대역에서 crosstalk 제거기의 성능이 떨어지는 단점이 있다.

<그림 9>는 기존의 crosstalk 제거기의 성능과 주파수 영역에서 regularization을 한 crosstalk 제거기의 성능을 비교하여 보여주고 있다. 그림은 웹에서 다운로드할 수 있는 MIT HRTF 데이터 베이스[46]를 사용하였으며, 스피커는 머리의 정면으로부터 $\pm 15^\circ$ 의 각도에 위치에 있으며, 스피커와 청취자간의 거리는 1.4m인 경우이다. 그림의 결과는 주파수 영역의 regularization에 의해 특정 주파수에서의 과도한 스피커 입력이 적절하게 조절될 수 있음을 보여준다. 한편 regularization에 의해 채널간의 분리 성능이 저하됨도 알 수 있다.

4. Crosstalk 제거기의 성능 향상 기법

Crosstalk 제거기를 위한 기존 알고리즘의 또 다른 문제점은 청취자의 귀의 위치가 최적화를 위해 사용된 마이크의 위치에서 조금만 벗어나도 성능이 급격히 감소되는 것이다[39]. 이러한 문제를 해결하기 위한 한 방법으로 머리가 움직이는 일정 공간에 대해 강인한 crosstalk의 제거 능력을 제공하는 기술이 필요하게 된다. 이러한 영역을 “equalization zone”은 또는 “sweet spot”이라고 하는데, 한 개의 위치만을 고려한 필터 설계가 아닌 여러 개의 머리의 위치에 대한 최적 필터를 설계하여 이 영역을 넓힐 수 있다[47].

이 방법은 여러 개의 머리 위치에 대하여 비용함수를 정의하여 최소화 하는 필터 계수를 찾는 방법이다. 이 때 다음과 같은 비용함수가 사용될 수 있다.

$$J_x(\mathbf{h}) = \frac{1}{X} \int_x |\delta - \mathbf{C}_x \mathbf{h}|^2 dx \quad (18)$$

여기서 C_x 는 머리의 중심이 x 의 위치에 있을 때의 음향 경로의 전달함수이며, X 는 여러 개의 머리 위치로 이루어진 영역이다. 이러한 비용 함수를 최소화하는 최적 필터 계수는 다음과 같이 구할 수 있다.

$$J_x(\mathbf{h}) = \frac{1}{X} \int_x |\delta - \mathbf{C}_x \mathbf{h}|^2 dx \quad (19)$$

이 때,

$$\mathbf{R} = \frac{1}{X} \int_x \mathbf{C}_x^T \mathbf{C}_x dx \quad (20)$$

$$\mathbf{Q} = \frac{1}{X} \int_x \mathbf{C}_x^T dx \quad (21)$$

이다. 이 방법은 한 개의 마이크 위치에서 필터를 최적화하는 기존의 방법에 비해 보다 넓은 영역에서 안정된 crosstalk 제거 능력을 갖고, 수식 (19)의 행렬 \mathbf{R} 의 역행렬 연산이 안정적이어서 스피커에 과도한 입력이 가해지지 않는다는 장점이 있으나, 다수의 마이크로폰을 이용해야 하고, 고정된 머리 위치에서의 두 채널간 분리도가 낮아진다는 단점을 갖고 있다.

Crosstalk 제거기 구현에 있어서 주요 문제 중 하나로 저주파 대역에서의 구현의 어려움을 들 수 있다. 저주파에서는 직접 경로 머리전달 함수와 crosstalk 경로 머리전달 함수가 매우 유사하여 역필터 연산이 불안정해지기 때문이다. 이러한 문제를 해결하기 위해 Warped FIR필터를 고려할 수 있다. 이 방법은 all-pass 필터를 딜레이 라인 대신에 사용하여 주파수 축에서 저주파 대역의 해상도를 높여 주는 방법이다. 이는 선형 FIR 필터의 딜레이 z^{-1} 를 $D(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}$ 로 대신함으로써 이뤄진다. 이 때 사용되는 λ 를 warping 파라미터라 하며 $\lambda=0.756$ ($f_s=44100\text{Hz}$)에서 바크 스케일과 유사해지게 된다[48].

최적 필터의 계수를 구하기 위해서는 스피커

와 귀간의 전달함수를 dewarping 혹은 pre-warping을 하고, 이를 적절히 잘라서 사용하게 된다[49]. 이러한 warped FIR필터를 사용하게 되면 저주파에서 주파수해상도를 높일 수 있기 때문에 저주파 대역에서 강인한 crosstalk 제거 능력을 제공할 수 있으며, 적은 수의 필터 계수를 사용하여 설계하여도 저주파에서 기존의 방법과 비슷한 성능을 얻을 수 있기 때문에 시스템의 계산량을 낮출 수가 있다[49].

V. 인공 잔향기 기술

잔향(reverberation)은 실내에서 음원에 의해 발생한 음이 어떻게 청취자에 전달되는가를 설명해 주는 음향 현상이다. 실내에서의 음은 직접 전달 경로에 의해 가장 먼저 청취자에게 도달하고, 이후 벽면과 천장, 그 밖의 물체에 의해 반사되어 청취자에게 도달한다. 이후 음파들은 반사를 계속하면서 무수히 많은 반사음으로 만들어 내고, 반사와 전파 과정에서 흡수되는 에너지만큼 크기가 연속적으로 줄게 된다. 따라서 잔향이란 청취자의 위치에 도달하는 연속적인 반사음들의 집합이라고 할 수 있다.

직접음이 전달된 이후 짧은 기간동안에 발생하는 반사음들은 방의 구조에 따라 방향성을 갖고 발생하게 되는데 이를 초기 반사음(early reflections)이라고 한다. 초기 반사음 이후에는 초기 반사음들보다 훨씬 수가 많은 방향성을 갖지 않는 반사음들이 발생하는데 이를 후기 반사(late reflections) 혹은 확산 잔향(diffuse reverberation)이라고 한다. 초기 반사음의 패턴은 방의 크기와 모양에 따라 구별되기 때문에 청취자가 위치 공간의 기하학적인 특성을 결정하는 중요한 요소이다[50].

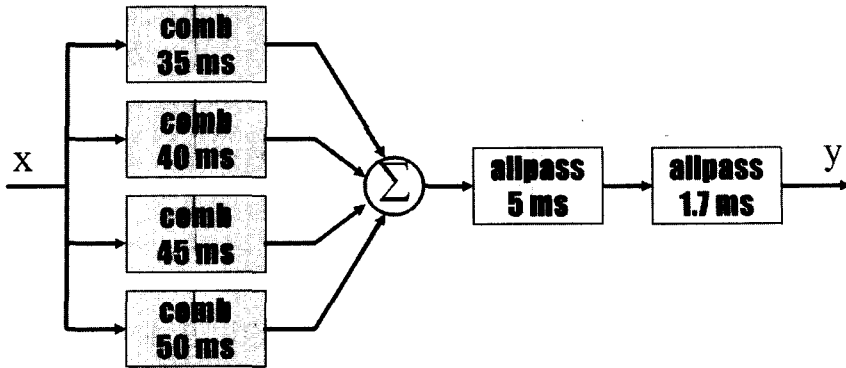
잔향에서 청취공간의 공간감(spaciousness)을 느끼는 중요한 파라미터가 두 귀에 전달되는 소리간의 통계적인 상관성(correlation)이다. 대체로 상관성이 낮을수록 공간을 크게 느끼게 된다. 또한 청취공간의 잔향 특성을 표현하는 파라미터가 잔향시간(RT: reverberation time)이다. 잔향시간은 음압이 초기 레벨의 -60dB 수준으로 떨어지는데 걸리는 시간이다. 잔향시간은 또한 주파수별로 다른 값을 가지기 때문에 주파수에 따른 잔향시간의 분포에 공간에 대한 지각적인 특성은 확연히 구별된다.

인공 잔향기(artificial reverberator)란 이렇게 실내에서 발생하는 음사 반사과정을 간단한 구조의 디지털 필터들을 사용하여 모델링하는 기술을 말한다. 인공 잔향기 기술에서는 잔향의 시간 구간에 따라 지각적인 역할이 다르다는 점 때문에 초기반사 구간과 후기 확산 잔향 구간을 나누어 다른 모델링 방법을 사용한다.

초기반사 구간을 모델링하는 방법은 주로 하나의 딜레이 채널을 이용하여 FIR 구조로 구현된다[51]. 이를 위해서는 초기 반사의 위치와 크기에 대한 정보가 필요한데 이는 실제 공간의 임펄스 응답으로부터 추출할 수도 있고 ray-tracing 방법[52]이나 image-source 방법[53]을 사용할 수도 있다. 초기 반사와 달리 후기 확산 잔향은 통계적인 접근법을 사용한다. 다음절에서는 후기 확산 잔향의 모델링 방법에 대해 자세히 살펴본다.

1. Comb 필터를 이용한 후기 반사음 모델링 기술

실내에서 확산된 반사를 효율적으로 구현하기 위해서 사용하는 기본적인 유닛은 comb 필터와 all-pass 필터이다. 자연스러운 확산 반사를 위



〈그림 10〉 Schroeder의 잔향기

해서는 잔향이 시간영역에서 충분한 밀도를 가져야 하고, 주파수 영역에서 공진주파수 사이의 간격이 좁아야 한다. 하지만, comb 필터와 all-pass 필터는 반향의 밀도가 매우 낮아 위와 같은 조건을 충족시키지 못한다. 따라서, 반향의 밀도를 증가시키기 위한 방법으로 comb 필터를 병렬로 연결하는 방법과 all-pass 필터를 직렬로 연결하는 방법을 사용한다. 1960년대 Schroeder [54][55] 이후로 많은 학자들이 comb, allpass 두 가지의 기본유닛으로 여러가지 조합을 만들어서 잔향기를 구현하여왔다.

Schroeder는 충분한 반향 밀도를 위해서 몇 개의 all-pass 필터를 직렬로 연결한 구조와, comb 필터를 병렬로 연결한 구조를 이용하여 그림 1-1와 같은 잔향기[54][55]를 제안하였다. 그림에서 딜레이가 인 경우, 초당 반향의 숫자는 25개이다. 4개의 병렬 comb 필터를 사용하는 경우에는 초당 100개의 반향이 만들어진다. 보통 8kHz의 샘플링 주파수를 사용할 경우 초당 1000개의 반향이 필요하므로, comb 필터 뒤에 all-pass 필터 2개를 이용하면 좀더 밀도가 높은 잔향기를 설계할 수 있게 된다. Schroeder의 잔향기에서는 원하는 잔향시간을 T 라고할 때, 병렬 comb

필터에서 딜레이(τ_n)를 에서 사이로 정하고, 이득 g_1, g_2, g_3, g_4 을 다음과 같이 구한다.

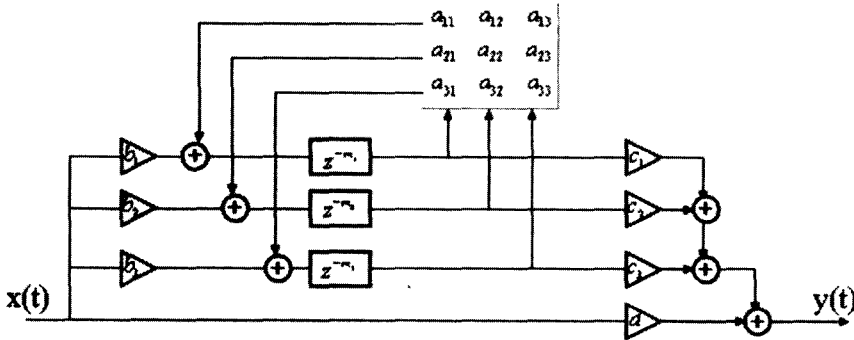
$$T = 3\tau_n / (-\log |g_n|) \quad (22)$$

한편 all-pass 필터의 딜레이는 5ms와 1.7ms로 정하고, g_5 와 g_6 는 0.7로 두었다.

Schroeder의 잔향기는 시간 영역과 주파수 영역에서 반향의 밀도가 대체로 높으나, 주파수 성분별로 감쇄하는 특징이 전부 동일하다. 하지만 실제 상황에서는 음파가 전달되면서 공기중에서 흡수되는 정도가 고주파에서 더 심하고, 음파가 벽에 부딪히면서 감쇄되는 정도도 고주파에서 더 크다[56][57][58]. 따라서 실제적인 상황과 일치하는 잔향기를 만들기 위해서는 이와 같은 음향학적인 특성을 고려해야만 한다. 즉, 주파수에 따라 다른 감쇄 특성을 줄 수 있는 방법이 필요하다. 이를 해결하기 위하여 Moorer는 기본 comb 필터안에 저역 통과 필터를 삽입하여 주파수별로 다른 감쇄를 고려한 잔향기를 제안하였다 [51]. 주파수 w 에서 잔향시간 $T(w)$ 을 갖기 위해서 삽입되는 저역 통과 필터는 다음과 같은 조건을 만족하는 주파수 응답 $h_p(e^{jw})$ 을 가져야 한다.

$$\log_{10} |h_p(e^{jw})| = -3\tau_p / T_r(w) \quad (23)$$

Moorer의 잔향기 구조에서는 실제 홀에서 측



<그림 11> 피드백 딜레이 네트워크의 블록 다이어그램

정한 임펄스 응답으로부터 초기(0~80msec) 반사패턴을 추출하여 FIR구조에 적용함으로써 초기 반사를 모델링하였고, 초기 반사들을 후기잔향 부분의 입력으로 들어가도록 하여 더욱 자연스러운 잔향을 모델링할 수 있도록 하였다.

Schroeder나 Moorer의 구조에서의 근본적인 문제인 불충분한 시간영역 반향 밀도를 해결하기 위해 J. Stautner와 M. Puckette이 FDN (Feedback Delay Network)을 제안하였다 [59]. 이것은 comb 필터를 일반화 시킨 것으로써, 상대적으로 작은 딜레이를 가지고 높은 시간영역 반향 밀도를 가지게 하는 방법이다. <그림 11>은 8개의 딜레이 라인을 갖는 FDN의 블록도이다.

FDN에서 입력과 출력 간의 전달함수를 구하면 다음과 같다.

$$H(z) = c^T [D(z^{-1}) - A]^{-1} (b) + d \quad (24)$$

위 식에서 $D(z)$ 는 z^{-m} 를 대각성분으로 갖는 대각행렬이며, A 는 $\{a_{ij}; 1 \leq i, j \leq 3\}$ 을 성분으로 갖는 피드백 이득 행렬, c 는 c_1, c_2, c_3 로 구성된 벡터이다.

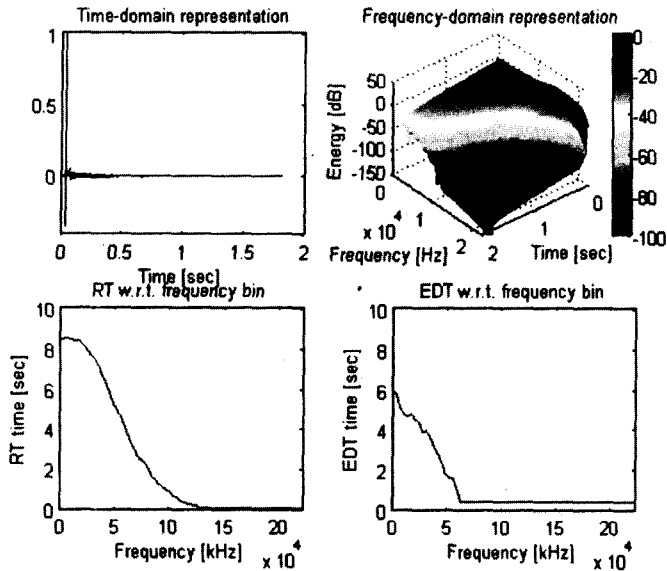
피드백 이득 행렬이 대각행렬이면 FDN은 Schroeder의 병렬 comb 필터와 같은 역할을 하게 되며, 피드백 이득 행렬이 삼각행렬이면 직렬로 연결된 all-pass 필터와 같은 동작을 한다. 일반적으로 위와 같이 특정한 행렬의 형태가 아닌

경우에 피드백 이득 행렬을 unitary 행렬을 사용한다. 피드백 이득 행렬이 unitary 행렬일 경우 FDN은 무손실 특성을 가진다[60].

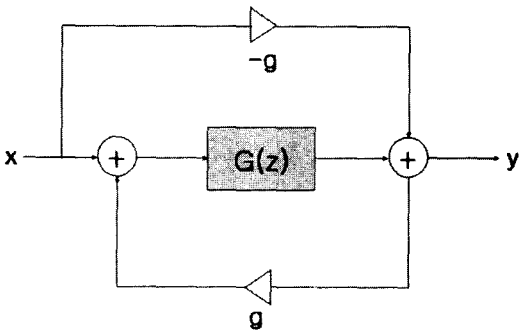
FDN에 감쇄특성을 주기 위해서는 피드백 경로에 감쇄상수를 삽입해야 한다. 그러나 특정 comb 필터가 다른 comb 필터보다 느린 감쇄 특성을 가지게 되면 FDN의 음색이 해당 comb 필터에 의해 결정되기 때문에, 이를 방지하기 위해 삽입되는 감쇄 상수 γ 는 다음과 같이 결정된다 [56][57][58][61].

$$\gamma = g_p^{1/m_p} \quad (25)$$

위 식에서 γ 는 감쇄상수이며, m_i 는 딜레이 라인의 길이이고, g 는 comb 필터의 이득값이다. 주파수별로 감쇄 특성을 조절하기 위해서는 감쇄 상수 대신에 특정한 주파수 함수를 가진 필터를 피드백 경로에 삽입한다. 실제 공간에서의 잔향은 저주파에서 빠른 감쇄특성을 보이기 때문에 주로 저역통과필터가 사용된다. <그림 12>는 FDN을 사용하는 인공 잔향기를 사용하여 만들어진 잔향의 한 예를 보여준다. 그림은 성당(cathedral)에서의 잔향을 시뮬레이션한 것으로서, 그림에서의 FDN은 20msec~35msec 사이의 길이를 갖는 8개의 딜레이 채널을 사용하였으며, 15msec 길이의 딜레이 채널에 8개의 탭을 갖는



〈그림 12〉 FDN을 사용한 인공 잔향기를 사용하여 만들어진 잔향 예



〈그림 13〉 Nested all-pass 필터의 구조

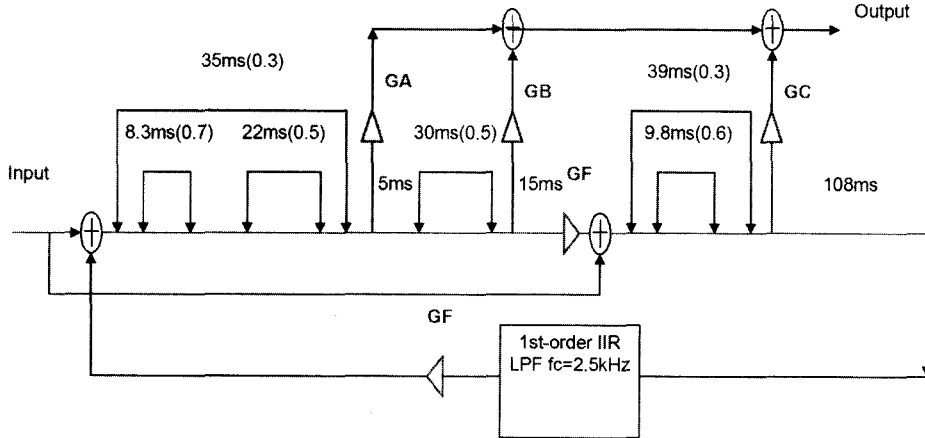
FIR구조를 적용하여 초기반사를 모델링하였다. 그림에서는 만들어진 시간영역에서의 임펄스 응답과 잔향의 주파수별 잔향시간을 동시에 보여주고 있다.

앞서 언급한 것처럼 자연스러운 후기 확산 반사를 위해서는 시간영역의 반사 밀도(time density)가 충분해야하고 주파수 영역에서 공진 주파수 사이의 밀도(modal density) 또한 충분히 높아야 한다. FDN 구조에서 이러한 조건을 만

족시키기 위해서는 딜레이 채널의 길이와 수가 충분해야하기 때문에 메모리 크기가 문제가 되고 연산량 부담 또한 상당하다. 이러한 문제를 해결하기 위해 여러 가지 형태의 저메모리, 저연산 구조의 후기 확산 잔향 방법이 제안되었다. 이는 Schroeder 잔향기 구조와 유사하게 시간영역 반사 밀도를 높이기 위해 all-pass 필터를 피드백 채널에 삽입하는 방법을 사용하고 있다[52][62].

2. All-pass 필터를 사용한 후기 반사음 모델링

앞절에서는 comb 필터를 이용한 후기 확산 잔향 모델링 방법에 대해 살펴보았다. 기본적으로 후기 잔향은 지수함수적으로 감소하는 랜덤 노이즈 시퀀스와 유사한 형태를 가지고 있다. 이러한 특성은 주파수 영역에서 크기는 일정하지만 위상은 랜덤하다는 것이다. 이러한 후기 잔향 특성을 모델링 하기위해 all-pass 필터를 사용할



<그림 14> Nested all-pass 필터를 이용한 잔향기 구조[Gardner]

수 있다.

시간영역 반사 밀도를 높이기 위해 nested all-pass 필터를 사용할 수 있다. <그림 13>은 nested all-pass 필터의 구조를 보여준다. 이는 all-pass 필터 $G(z)$ 안에 all-pass 필터가 존재하는 형태이다. Nested all-pass 필터는 다음과 같은 전달함수와 크기응답을 가진다.

$$H(z) = \frac{Y(z)}{X(z)} = \frac{G(z) - g}{1 - gG(z)} \quad (26)$$

$$|H(z)| = \sqrt{\frac{|G(z)|^2 - g(G(z) + G^*(z)) + g^2}{1 - g(G(z) + G^*(z)) + g^2|G(z)|^2}} \quad (27)$$

만일 $G(z)$ 의 크기가 1이면 $H(z)$ 의 크기는 1이기 때문에 주파수 응답은 all-pass 특성을 유지하게 된다. Nested all-pass filter의 장점은 시간에 따라 반사 밀도가 점차 증가한다는 점과 주파수 영역의 크기 응답이 주파수에 상관없이 일정하기 때문에 comb 필터와 같은 음색의 변화가 없다는 점이다.

주파수에 따른 감쇄특성을 조절하기 위해서는 comb 필터의 경우와 마찬가지로 루프 안에 감쇄 저역 통과 필터를 삽입할 수 있다[63].

Gardner는 nested all-pass 필터를 사용하여 <그림 14>와 같은 잔향기 구조를 제안하였다[7]. 이 구조는 비교적 자연스러운 고품질의 잔향을 얻을 수 있다는 장점은 있으나 잔향특성을 변화시키기 위해서는 구조를 바꾸어야하기 때문에 유연성이 떨어지는 단점이 있다.

V. 결론

본 논문에서 디지털 오디오 장치를 위해 사용되는 최신의 디지털 오디오 신호 처리 기술들의 동향에 대해 살펴보았다. 음원 부호화 기술의 동향과 이를 구현하기 위해 가능한 시스템들의 구조를 간략히 살펴보았다. 그리고 임의의 공간에 음원을 배치하기 위한 3차원 오디오 기술에 핵심적인 몇 가지 신호 처리 알고리즘을 소개하였다. 특히 최근 관심 대상이 되고 있는 두 채널 재생시스템에서 3차원 오디오를 구현하기 위해 필요한 HRTF 모델링 기술, crosstalk 제거 기술, 그리고 임의공간의 잔향을 모델링하기 위한 인공 잔향 기술 등에 관련한 신호 처리 방법들의

특징들을 관찰한다.

관련된 신호 처리 알고리즘들은 재생 시스템의 구성, 즉, 스피커의 수와 배치, 동작 환경, 등등에 따라 복잡도와 성능이 크게 변할 뿐만 아니라 신호 처리 기술을 적용할 때 고려하여야 하는 사항이 달라질 수 있다. 그러나 한 가지 확실한 사실은 본 논문에서 소개한 신호 처리 기술들은 대부분 현재 시스템에 상용화가 가능한 수준의 복잡도와 성능을 갖고 있다는 것이다. 따라서 현재의 추세대로 하드웨어와 소프트웨어 기술이 발전한다면, 보다 고품질의 3차원 오디오 기술들을 적용한 상용 제품들을 쉽게 시장에서 만날 수 있을 것이다.

참고문헌

- [1] Painter T, Spanias A, "Perceptual Coding of Digital Audio," Proc. of the IEEE, pp. 451-515, Apr. 2000.
- [2] Noll P, "MPEG Digital Audio Coding," Signal Processing Magazine, Vol. 14 No. 5, pp. 59-81, Sep. 1997.
- [3] Hans M, Schafer R.W., "Lossless Compression of Digital Audio," Signal Processing Magazine, Vol. 18, No. 4, pp. 21-32, Jul. 2001.
- [4] 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)".
- [5] ISO/IEC JTC1/SC29/WG11 No.703 "Generic Coding of Moving Pictures and Associated Audio - CD 13818-3 (Part 3. MPEG-Audio)", Mar. 1994.
- [6] W. G. Gardner, The Virtual Acoustic Room, M. S. Thesis, MIT, 1992.
- [7] 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.
- [8] Ming Zhang, Kah-Chye Tan, and M.H.Er, "Three-Dimensional Sound Synthesis Based on Head-Related Transfer Functions", J. Audio Eng. Soc., Vol. 46, No. 10, 1998.
- [9] Jesper Sandvad, Dorte Hammershoi "Binaural Auralization. Comparison of FIR and IIR Filter Representation of HIRs", Proc. 96th A.E.S. Convention(Amsterdam), 1994.
- [10] P. A. Nelson and H. Hamada, and S. J. Elliott, "Adaptive inverse filters for stereophonic sound reproduction," IEEE Trans. on Signal Processing, Vol. 40, pp. 1621-1632, Jul. 1992.
- [11] S. M. Kuo, and G. H. Canfield, "Dual-channel Audio Equalization and Cross-talk Cancellation for 3-D Sound Representation," IEEE Trans. On Consumer Electronics, Vol. 43, No. 4, pp. 1189-1196, Nov. 1997.
- [12] Y. F. Dehery, et al., "A MUSICAM source codec for digital audio broadcasting and storage," Proc. of ICASSP, pp. 3605-3608, 1991.
- [13] ISO/IEC JTC1/SC29/WG11 No.972 "Dolby AC-3 Multi-channel Digital Audio Compression System Algorithm Description," Nov. 1993.
- [14] ISO/IEC JTC1/SC29/WG11 No.1650 "IS13818-7 (MPEG-2 Advanced Audio Coding, AAC)," Apr. 1997.
- [15] <http://www.ti.com>
- [16] Mesarovic, V, et al., "MPEG-4 AAC audio decoding on a 24-bit fixed-point dual-DSP architecture," Proc. of ISCAS, pp. 706-709, May 2000.
- [17] Kyoung Ho Bang, et al. "Design optimization of a dual MP3/AAC decoder," Proc. of ICASSP, pp. 3776-3779, May 2002.
- [18] <http://www.arm.com>
- [19] Takamizawa, Y, et al., "MPEG-2 5.1-channel decoder software for a low-power embedded RISC microprocessor," Proc. of SIPS, pp. 351-360, Oct.

- 1999.
- [20] Wonchul Lee, et al. "Software optimization of MPEG audio layer-III for a 32 bit RISC processor," Proc. of APCCAS, pp. 435-438, Oct. 2002.
- [21] Keun-Sup Lee, et al. "Software optimization of the MPEG-audio decoder using a 32-bit MCU RISC processor," Proc. of CE, pp. 330-331, Jun. 2002.
- [22] Keun-Sup Lee, et al. "High Quality MPEG-audio layer III algorithm for a 16-bit DSP," Proc. of ISCAS, pp. 205-208, May, 2001.
- [23] D. Griesinger, "Binaural techniques for music reproduction", Proc. 8th A.E.S. International Conference, pp. 225-229, 1990.
- [24] S.T. Pope, L.E. Fahlen, "The use of 3-D audio in a synthetic environment : an aural renderer for a distributed virtual reality system", Proc. Int.. Computer Music Conference, pp. 146-149, 1993.
- [25] Jean-Marc Jot, Veronique Larcher and Olivier Warusfel "Digital Signal Processing Issues In the Context of Binaural and Transaural Stereophony", Proc. 98th A.E.S. convention(Paris), 1995.
- [26] Veronique Larcher, "Equalization methods in binaural technology", Proc. 105th A.E.S. Convention(San Francisco), 1998.
- [27] J. Mourjopoulos, "Digital Equalization Methods for Audio Systems", Proc. 84th A.E.S Convention (Paris), 1988.
- [28] J. Mourjopoulos, P.M. Clarkson and J.K. Hammond "A Comparative Study of Least-Squares and Homomorphic Techniques for the Inversion of Mixed Phase Signals" Proc. of ICASSP., pp. 1858-186, 1982.
- [29] J. O. Smith, "Techniques for Digital Filter Design and System Identification with Application to the Violin", Ph.D. thesis, CCRMA Tech. Rep. STAN-M-14, Stanford University(CCRMA), Stanford, CA, Jun. 1983.
- [30] T. Parks and C. Burrus, Digital Filter Design, Wiley, New York, 1987.
- [31] J. Mackenzie, J. Huopaniemi, V. Valimaki and I. Kale, "Low-Order Modeling of Head-Related Transfer Functions Using Balanced Model Truncation", IEEE Signal Process. Let., Vol. 4, No. 2, pp. 39-41, 1997.
- [32] A. Kulkarni and H. S. Colburn, "Infinite-impulse-response filter models of the head-related transfer function", J. Acoust. Soc. Amer., Vol. 97, No. 5, pt. 2, pp. 3278.
- [33] Paul D. Bauman, Stanley P. Lipshitz, Tony C. Scott, and John Vanderkooy, "Cepstral Techniques for Transducer Measurement" Proc. 76th A.E.S. Convention (New York), 1984.
- [34] Alan V. Oppenheim, Gary E. Kopec, Jose M. Tribolet, "Signal Analysis by Homomorphic Prediction" Proc. of ICASSP, Vol. 24, No. 4, pp. 327-332, 1976.
- [35] Jyri Huopaniemi and Nick Zacharov and Matti Karjalainen, "Objective and Subjective Evaluation of Head-Related Transfer Function Filter Design", J. Audio Eng. Soc., Vol. 47, No. 4, 1999.
- [36] Carlos Avendano, Richard O. Duda and V. Ralph Algazi, "Modeling the Contralateral HRTF", Proc. 16th A.E.S. International Conference, pp. 313-318, 1999.
- [37] Gregory J. Sandell, "Perceptual Evaluation of Principal- Component-Based Synthesis of Musical Timbres", J. Audio Eng., Vol. 43, No. 12, 1995.
- [38] Kyriakis, K. and Holman, T., "Video-based head tracking for improvements in multichannel loudspeaker audio," Proc. 105th A.E.S. convention (San Francisco), Preprint 4845, 1997.
- [39] P. A. Nelson, F. O. Bustamante, and H. Hamada, "Inverse Filter Design and Equalization Zones in Multichannel Sound Reproduction," IEEE Trans. On Speech and Audio Proc., Vol. 3, No. 3, pp. 185-192, May 1995.

- [40] Darren B. Ward and Gary W. Elko, "Effect of Loudspeaker Position on the Robustness of Acoustic Crosstalk Cancellation," *IEEE Signal Process. Let.*, Vol. 6, No. 5, May 1999.
- [41] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
- [42] Kirkeby, O., Nelson, P.A. and Hamada, H., "The "stereo dipole" - A virtual source imaging system using two closely spaced loudspeakers," *J. Audio Eng. Soc.*, 46(5), pp. 387-395, 1998.
- [43] Roger A. Horn and Charles R. Johnson, "Matrix Analysis," Cambridge University Press, 1985.
- [44] William H. Press, Saul A. Teukolsky, William T. Vetterling, and brian P. Flannery, "Numerical Recipes in C," Second edition, Cambridge University Press, 1992.
- [45] Ole Kirkeby, P.A. Nelson, and H. Hamada, "Digital Filter Design for Virtual Source Imaging Systems," *J. Audio Eng. Soc.*, 16-19, May 1998.
- [46] <http://sound.media.mit.edu/KEMAR.html>
- [47] Darren B. Ward, "Joint Least Squares Optimization for Robust Acoustics Crosstalk Cancellation," *IEEE Trans. On Speech and Audio Proc.*, Vol. 8, No. 2, Feb. 2000.
- [48] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, UK Laine, and J. Huopaniemi, "Frequency-Warped Signal Processing for Audio Application," *J. Audio Eng. Soc.*, Vol. 48, No. 11, Nov. 2000.
- [49] Ole Kirkeby, Per Rubak, Lars G. Johansen, and Philip A. Nelson, "Implementation of Crosstalk Cancellation Networks using Warped FIR Filters," *Proc. 16th A.E.S. International Conference: Spatial Sound Reproduction*; Apr. 1999.
- [50] M.A. Gerzon, "The design of distance panpots," *Proc. 92nd A.E.S. Convention (Vienna)*, Preprint 3308, Mar. 1992.
- [51] James A. Moorer, "About this reverberation business," *Computer Music Journal*, Vol. 3, No. 2, pp. 13-28, 1979.
- [52] Lauri Savioja, "Modeling techniques for virtual acoustics," Helsinki University of technology department, Dec. 1999.
- [53] Jont B. Allen and David A. Berkley, "Image method for efficiently simulating small room acoustics," *J. Acoust. Soc. Amer.*, Vol. 65, No. 4, Apr. 1979.
- [54] M. R. Schroeder, "Natural sounding artificial reverberation," *J. Audio Eng. Soc.*, Vol. 10, No. 3, pp. 219-223, Jul. 1962.
- [55] M. R. Schroeder and B.F. Logan, "Colorless artificial reverberation," *J. Audio Eng. Soc.*, Vol. 9, No. 3, pp. 219-223, Jul. 1961.
- [56] U. Zolzer, *Digital Audio Signal Processing*, John and Wiley & Sons, Ltd., 1997.
- [57] J. Blauert, *Spatial Hearing*, MIT Press, 1982.
- [58] Beranek, "Concert and Opera halls How to they sound," *Acoustical Society of America*, 1996.
- [59] John Stautner and Miller Puckette, "Designing multi-channel reverberators," *Computer Music Journal*, Vol. 6, No. 1, pp. 52-65, Spring 1982.
- [60] Davide Rocchesso and Jullius Smith, "Circulant and Elliptic feedback delay networks for artificial reverberators," *IEEE Trans. On Speech and Audio Proc.*, Vol. 5, No. 1, Jan. 1997.
- [61] J. M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *Proc. 90th Conv. Audio Eng. Soc.*, February, 1991, preprint 3030.
- [62] Lauri Savioja Jyri Huopaniemi Tapio Lokki and Ritta Vaananen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc.*, Vol. 47, No. 9, Sep. 1999.
- [63] Luke Dahl, Jot, Jean-Marc, "A Reverberator based on absorbent all-pass filters," *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, Verona, Italy, Dec. 7-9, 2000.

저자소개



윤 대 회

1979-1982 미국 캔자스주립대 전기공학 박사
1982-1985 미국 아이오와 대학 조교수
1985-1990 연세대학교 조교수
1990-1995 연세대학교 부교수
1995-현재 연세대학교 전자공학과 교수
2001-현재 연세대학교 공과대학 학장
주관심분야 오디오 신호처리, 음성 신호처리, 음성 및
오디오 코딩, 적응 신호처리