

## 다채널 오디오 시스템을 위한 음향 신호처리

김래훈, 전재진, 이신렬, 김세웅(서울대학교 전기공학부), 임준석(세종대학교 교수), 성평모(서울대학교 교수)

### 1. 서론

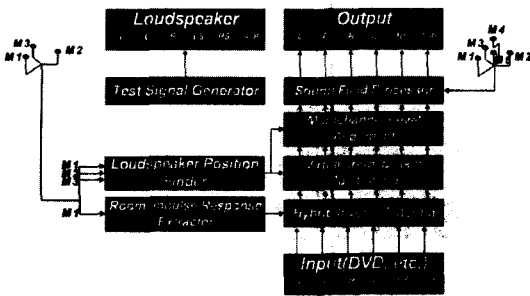
근래 들어 홈시어터 시스템과 같은 다채널 오디오 시스템의 보급이 활발해지면서 이러한 시스템에서의 음장 특성의 최적화가 중요한 이슈로 부각되기 시작하였다. 음장 특성의 최적화를 위해서는 여러 가지 신호처리기법의 사용이 필수불가결한데, 본 논문에서는 그림 1.1에서 보이는 바와 같은 자동 음장 튜닝 및 가상 음장 시스템의 신호처리 기법에 대해서 자세히 다루면서, 몇몇 기법을 새롭게 제안하겠다.

첫째로 각 스피커에서 청자의 위치까지의 공간 응답의 영향을 제거할 수 있는 역 필터링에 대하여 다룬다. 여기서는 현재까지 제안되어 있는 방식들의 문제점을 지적하고 이를 극복하기 위해 실제적인 측면을 고려하는 역 필터링 방식을 제안한다. 공간 응답을 역 필터링하기 위해서 본 논문에서는 저주파 부분과 고주파 부분을 나누어 다른 방식을 적용하는 새로운 방식을 제안한다 [1]. 저주파 부분은 청취자의 머리이동에 의한 측정 점의 이동에 비해 파장이 길기 때문에 상대적으로 공간 응답의 변화가 적음에 주목하여 최소 사승법을 이용하여 역 필터링을 수행하

였다. 고주파 부분에서는 우리 귀가 인지하는 방식과 가장 가까운 1/3 옥타브 밴드 간략화 (smoothing)를 수행한 후 이 때 측정 점의 위치 변화에 따른 주파수 크기 응답의 변화가 적음을 주목하여 이 주파수 크기 응답의 역을 크기로 갖고 이에 따른 선형 위상을 부가하는 방식으로 역 필터링을 수행하였다. 저주파 부분과 고주파 부분의 경계 주파수는 식 1.1과 같은 슈뢰더의 큰 방 한계라고 불리는 슈뢰더 공식 (Schroeder formula of Schroeder's large room limit)에 따라 결정되는 것이 일반적이다 [2].

$$f_s = c \sqrt{\frac{6}{A}} \approx 2000 \sqrt{\frac{T}{V}} \quad (1.1)$$

여기서 사빈의 잔향시간 (Sabine's reverberation time) 공식인  $T=0.163V/A$ (초),  $V$  ( $m^3$ ) 는 공간의 부피  $A(m^2)$ 은 넓이를 의미하며  $c$  는 음파 전달 속도를 말한다. 이 때 일반적인 거실의 부피를  $60 m^3$  잔향시간을 0.4초라고 하면 경계주파수는 약 163Hz 정도 된다. 이 주파수를 경계주파수로 하여 본 논문에서 제안하는 역 필터링을 수행할 경우 우퍼 (woofer)이외의 스피커에서는 그다지 이득을 얻지 못한다. 이는 우퍼이외의 다른



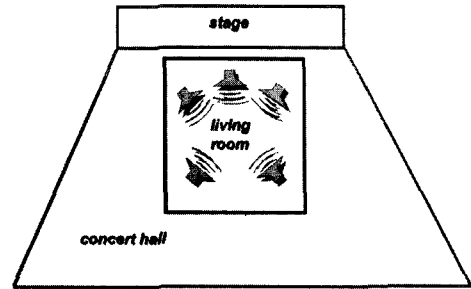
〈그림 1.1〉 자동 음장 튜닝 및 가상 음장 시스템의 개념도

스피커가 일반적으로 약 100Hz 이상 대역만을 사용하기 때문이다 [3]. 따라서 본 논문에서는 이 경계 주파수를 1kHz 근방 대역까지 올리기 위해서 측정 점을 청취 점을 중심으로 여러 곳으로 선정하고 후처리를 통해 이 여러 측정 응답의 대표 값을 구해내어 이를 역 필터링의 대상으로 선정하는 방법을 도입한다.

둘째로 다채널 스피커의 위치를 파악하여 최적 위치와의 차이를 자동적으로 보상할 수 있는 방법에 대하여 다룬다. 이 방법은 청취자의 위치에서 현재 스피커의 배치 위치를 3채널 마이크로폰 시스템을 이용하여 상대적으로 구해내어 최적 위치와의 차이만큼을 수정하는 방식이다.

현재까지의 다채널 재생 시스템의 경우 한 평면상에 스피커를 위치시키는 것이 일반적이기 때문에 이렇게 평면상에 배치되어 있는 세 개의 마이크로폰을 이용하여 그 위치를 파악할 수 있지만, 만약 다른 고도를 갖는 스피커의 위치를 파악해야 한다면 6장에서 소개되는 5 채널 마이크로폰 시스템을 이용하면 가능할 것이다.

최적 위치와의 차이 보상은 쉽게는 파악된 차이 만큼에 대해서 스피커를 실제적으로 위치 이동시켜 만족시킬 수도 있으나, 이것이 불가능할



〈그림 1.2〉 다채널 재생 시스템에서의 음장 재현 방식

경우가 많을 것이므로 본 논문에서는 그림 1.1에서 보이는 가상 음원 위치 방식(virtual loudspeaker positioning)에 의한 자동 보정 방법을 제안한다. 가상 음원 위치 방식은 새롭게 제안되는 개선된 인접 스피커 사이의 신호 크기 값을 조정하여 음상을 정위하는 방식(ECPP: enhanced constant power panning law)을 적용시킨다 [4].

셋째로 그림 1.1에서 보이는 인간의 인지적인 측면을 고려하는 다채널 스피커로부터의 에너지 레벨 정렬(Multi-channel level alignment) 방식에 대하여 다룬다. 기존에는 인간의 방향에 따른 에너지 레벨 인식의 차이를 무시하고 물리적으로 같은 에너지 레벨로 정렬하는 방식이 많이 사용되거나, 방향에 따라 다른 방식을 적용하더라도 권장되는 기준 상황에 대해서 정해진 보상 값을 적용하는 방식을 선택하여 왔다. 그러나 실제적인 배치 상황은 그와 많이 다를 수 있으므로 여기서는 각 실제적인 배치위치에 해당되는 적합한 보상 값을 적용시키기 위해 먼저 전 방향에 따른 양귀 응답 레벨을 구하고 상황에 따라 적용하는 방식을 제안한다.

넷째로 특정 공간의 반사음의 분포 패턴을 구

하여 이를 일반적인 청취공간에서 재현해 내는 음장 재현 방식(sound field reproduction)에 대하여 다룬다. 우리가 일반적으로 5.1채널과 같은 다채널 재생 시스템을 이용하는 곳은 공간이 협소한 거실과 같은 공간이다. 그러나 실황 연주 등의 DVD 타이틀을 이를 이용하여 감상할 때는 마치 콘서트홀과 같은 큰 공간에 있는 듯한 느낌을 갖게 되길 원하기 마련이다. 이러한 공간의 크기와 같은 입장감(spatial impression)은 음장 공간의 초기반사음(early response) 특성에 많은 영향을 받는다는 것이 널리 알려져 있다 [6]. 여기서는 그림 1.2에서 보는 것과 같이 다채널 시스템을 마치 콘서트홀과 같은 원하는 음장 공간에 놓고 있는 듯한 느낌을 갖도록 하는 것을 목표로 한다.

마지막으로 요약을 하면서 본 논문을 마치겠다.

## II. 에너지 가변적 사인 스위프(Sine Sweep)를 이용한 음향 공간의 충격응답 측정 기법

음향 공간에서 공간의 전달함수와 그에 따른 충격응답을 제대로 구하는 것은 매우 중요하다. 충격응답을 구하는 과정에서 중요한 점 중 하나는 충분한 신호대잡음비(Signal to Noise Ratio)를 확보하는 것이다. 기존에는 총이나 풍선, 박수 소리 등의 충격음을 사용하여 공간의 충격응답을 구하였는데, 이것은 위에서 말한 충분한 신호대잡음비를 확보하지 못하는 단점이 있다. 그리고 근래에도 많이 사용되는 MLS(Maximum Length Sequence) 역시 이 점에서는 한계를 가지고 있다 [9].

충분한 신호대잡음비를 확보하기 위해서는 저주파 대역의 에너지가 강조된 형태의 음원이 바

람직하다. 위의 충격음이나 MLS의 대안으로 저주파가 강조된 형태의 음원으로는 로그함수 형태를 가지는 사인 스위프(logarithmic sine sweep)이 있다. 본 장 시간 영역에서 일정한 엔벨로프(envelop)를 가지면서, 공간의 주파수에 따른 주변 소음 크기(ambient noise level)를 고려한 사인 스위프를 만들어, 각 주파수 대역에서, 가능한 한 비슷한 신호대잡음비를 확보할 수 있는 방법을 제시하고, 옥타브 밴드(octave band)별 신호대잡음비 확보 정도의 비교를 통하여 이 방법의 장점을 보이겠다.

### 1. 기존의 측정방법

#### 가) 충격음(impulse) 기법

충격(impulse) 음원을 입력 신호로 사용하여 그 응답, 즉 충격 응답을 바로 구하는 방법으로. 예전에는 아날로그적인 방법으로 충격 음원을 만들어 사용하였으나 요즘에는 임펄스 형태의(Dirac style) 펄스를 디지털로 만들어 이를 D/A(Digital to Analog) 변환기를 통해 증폭하여 내보내고 이를 다시 A/D(Analog to Digital) 변환기를 통해 받아들여 충격 응답을 구하는 방식이 선호되고 있다. SNR을 높이기 위한 방법으로 펄스를 주기적으로 인가하여 각 주기의 응답을 더하는 방법이 있다. 이 방법을 사용할 경우, 평균하는 주기의 수가 두 배가 될 때마다 약 3dB의 상관관계가 없는 잡음의 감소 효과가 있는 것으로 알려져 있다. 또한 스피커와 마이크론 사이의 전달 시간만큼의 시간 지연을 제거하고 윈도우링(windowing) 하는 것으로도 SNR을 높일 수 있다.

전달 함수는 충격 응답을 FFT(Fast Fourier Transform)함으로써 구할 수 있는데, 측정의 정

확도를 높이려면 그 결과를 측정 시스템의 입력과 출력을 연결하여 얻은 전달 함수의 역과 곱해주면 되는데 이는 입력 신호를 배경 잡음에 대해 전처리 강조(preemphasis) 한 효과를 갖게 한다.

이러한 충격음 기법은 측정과정 전체가 완전히 전기적일 때에나 가능하고 또한 빨리 결과를 얻어야 하는 경우에 가장 간단하고 알맞은 방법이다.

나) Maximum Length Sequence(MSL)

MLS는 N 개의 쉬프트 레지스터(shift resistor)와 하나의 XOR(Exclusive OR) 게이트를 통해 생성되는 간단한 이진 행렬로 2N 개의 상태(state)를 갖는다. 이중 "all-0"의 경우를 제외한 2N-1의 길이를 갖는데 이들의 자기상관도는 백색 스펙트럼을 가리키는 '임펄스(impulse)'에 가깝다. 같은 진폭을 갖는 단일 펄스와 비교할 때, 여진 신호가 측정 주기 전체에 걸쳐 있기 때문에 훨씬 많은 에너지가 측정하고자 하는 시스템에 인가되어 결과적으로 높은 SNR을 얻을 수 있다.

MLS의 단점은 비선형 왜곡과 시간에 따른 시스템의 변화에 (time-variance)에 매우 취약하다는 것이다. 또한 대부분의 음향학적 측정에 비백색 잡음이 적합한 것으로 알려진 것에 비해 MLS가 백색 스펙트럼 갖는다는 것 또한 단점으로 지적될 수 있다.

다) 사인 스위프(Sine Sweep)

사인 스위프는 시간에 따라 사인파의 주파수가 증가되는 음원이다. 사인파를 증가시키는 방법에 따라 선형 스위프(linear sweep)과 로그형태의 스위프(logarithmic sweep) 두 가지로 나눌 수 있다. 사인 스위프를 시간 축에서 합성하는 식은 다음과 같다.

(선형 스위프)

$$x(t) = A \sin(\Phi) \tag{2.1}$$

$$\Phi = \Phi + \Delta\Phi \tag{2.2}$$

$$\Delta\Phi = \Delta\Phi + inc\Phi \tag{2.3}$$

식(2.1), (2.2), (2.3)에서 보는 것과 같이 시간 축에서 선형 스위프를 합성하기 위해서는 스위프 신호의 위상  $\Phi$ 를 시간에 따라 고정된  $inc\Phi$  만큼 증가시키면 가능하다. A는 합성되는 스위프신호의 엔벨로프 크기를 결정한다.

(로그 형태의 스위프)

$$\Delta\Phi = \Delta\Phi \cdot Mul\Phi \tag{2.4}$$

$$\Delta\Phi_{start} = 2\pi \cdot f_{start} / f_s \tag{2.5}$$

$$Inc\Phi = 2\pi \frac{f_{stop} - f_{start}}{f_s \cdot N} \tag{2.6}$$

$$Mul\Phi = 2^{\log_2(f_{stop} - f_{start}) / N} \tag{2.7}$$

여기서 합성식과 위상 식은 식 (2.1), (2.2)과 동일하고,  $f_{start}$  는 시작 주파수를  $f_{stop}$  마지막 주파수를 의미한다. 그리고 N은 합성될 스위프 신호의 샘플 수를 의미한다. 시간 축에서 합성된 스위프의 경우 시간 축에서는 보면 완벽한 엔벨로프(envelope)를 갖게 되지만, 주파수 영역에서는 갑작스런 스위치 on/off 때문에 스펙트럼 모양에 원하지 않는 물결 모양의(ripple)문제가 생긴다.

FFT 스펙트럼의 크기(magnitude)와 군 지연 시간(group delay)을 이용하여 주파수 도메인에서 스위프 신호를 합성할 수도 있다. 여기서 군 지연 시간(group delay)  $\tau_g(f)$ 는 스위프 신호에서, 개별적인 특정 주파수가 나타나게 되는 시간을 의미하게 된다. 주파수 축에서의 합성 하는 식은 다음과 같다.

(선형 스위프)

$$\tau_G(f) = \tau_G(0) + f \cdot k \quad (2.8)$$

$$k = \frac{\tau_G(f/2) - \tau_G(0)}{N} \quad (2.9)$$

(로그 형태의 스위프)

$$\tau_G(f) = A + B \cdot \log_2(f) \quad (2.10)$$

$$B = \frac{\tau_G(f_{end}) - \tau_G(f_{start})}{\log_2(f_{end} - f_{start})} \quad (2.11)$$

$$A = \tau_G(f_{start}) - B \cdot \log_2(f_{start}) \quad (2.12)$$

주파수 축에서 합성하는 경우에는 시간 축에서 합성하는 경우와는 반대로 합성되는 스위프 신호가 갑자기 시작되고 끝나게 되는 이유로 시작점과 끝점이  $\tau_G(f_{start})$ 와  $\tau_G(f_{end})$ 로 한정되지 않는다.

선형 스위프 신호는 MLS와 같은 백색 잡음이지만, 같은 에너지로 발생시 MLS와는 달리 비선형 왜곡에 의한 영향 없이 발생시킬 수 있고, 로그 형태의 스위프 신호를 이용하여 저주파에 보다 더 많은 에너지를 부가할 수도 있다. 또한 MLS에 비해 시간에 따른 측정 하고자하는 시스템의 변화에 영향을 덜 받는다고 알려져 있다. 그러나 로그 형태의 스위프를 사용한다 하더라도 개개의 공간에서 소음의 주파수특성을 개별적으로 반영할 수 없고, 모든 공간에서 똑같은 형태로 사용할 수밖에 없다는 단점을 갖는다. 따라서 배경 소음의 주파수별 에너지 특성을 반영하여 저주파가 강조된 음원을 사용하는 것이 충분한 신호대잡음비를 구하는 데 있어서 최적의 방법이다.

## 2. 배경 소음의 주파수별 에너지 특성을 반영할 수 있는 에너지 가변적 사인 스위프

그림 2.1의 에너지 가변적 사인 스위프를 만드는

흐름도(Flow Chart)이다. 흐름도에 따라 우선, 공간의 배경 소음을 측정하고, 이것을 FFT한 다음 중심 주파수가 125Hz부터 시작하는 선형 위상을 갖는 유한 충격 응답 필터(FIR: Finite Impulse Response)로 옥타브 밴드 평균을 구하면, 그림 2.2가 된다. 이 때 그림 2.2의 주파수별로 정규화된 크기값 (Normalized frequency magnitude response)을  $H(f)$ 의 절대 값으로 하고 다음과 같이 균 지연시간을 구하는데 쓰인다.

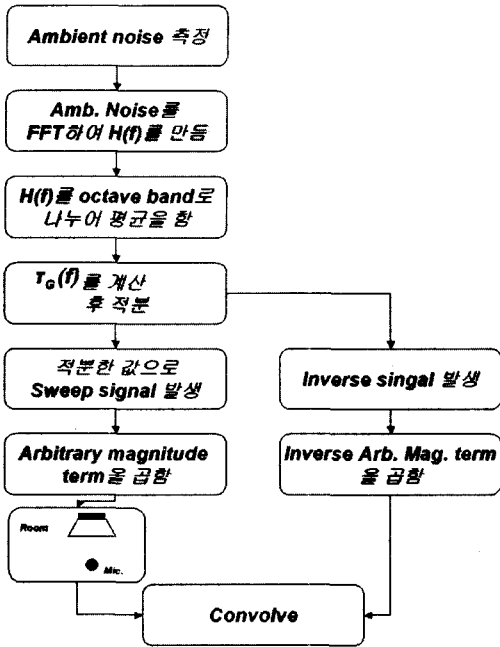
$$\tau_G(f) = \tau_G(f - df) + C \cdot |H(f)|^2 \quad (2.13)$$

$$C = \frac{\tau_G(f_{end}) - \tau_G(f_{start})}{\sum_{f=0}^{f_{end}} |H(f)|^2} \quad (2.14)$$

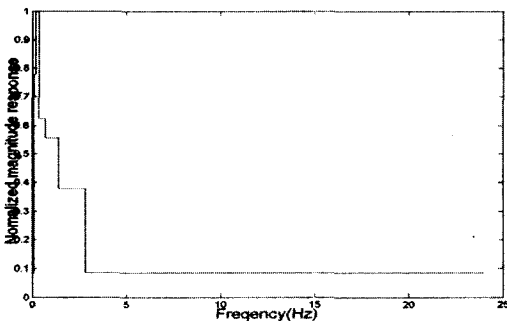
여기서 식 (2.13)은 반복문을 이용하여 각 주파수축 상의 모든 샘플 값에서 행하게 된다. 'df'는 따라서 각 샘플간의 간격을 의미한다. 위의 식 (2.13)에서 구한 균 지연시간을 이용하여 주파수축 상에서 스위프 신호를 합성하고, 여기에  $H(f)$ 의 절대 값을 곱한 후, 역 푸리에 변환하면 원하는 스위프신호를 얻을 수 있다. 이때 측정된 신호에 콘볼루션하여 충격 응답을 얻을 때 사용하게 될 역 신호는 그림 2.3과 같다. 이렇게 발생된 에너지 가변적 사인 스위프를 음원으로 사용하여, 공간의 주파수별 신호대잡음비를 구하고, 선형 스위프와 로그형태의 스위프를 이용하여 측정한 결과와 비교하면, 표 2.1과 같다.

에너지 가변적 sine sweep을 음원으로 사용하여 공간의 각 주파수별 신호대잡음비를 구해보면, 표 2.1의 'Arbitrary' 항목에서 보는 것처럼 선형 신호(linear)나 로그형태의 신호(log sweep)보다 전 주파수 대역에서 더 나은 신호대잡음비를 확보할 수 있는 것을 볼 수 있다.

균 지연시간을 조절하여 만드는 에너지 가변

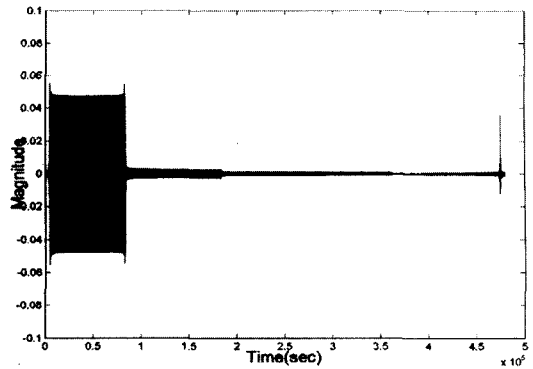


〈그림 2.1〉 배경 소음의 주파수별 에너지 특성을 반영할 수 있는 에너지 가변적 사인 스위프 신호 합성의 흐름도



〈그림 2.2〉 옥타브 밴드 필터 통과 후 각 밴드별 평균을 구해 얻은 H(f)의 정규화된 크기 응답

적 사인 스위프는 공간의 잡음 주파수특성을 반영할 수 있어서, 저주파수에 에너지밀도를 높여 줄 수 있다. 동시에 시간 축상의 엔벨로프는 거의 유지할 수 있는 장점을 가지고 있다. 이러한 일정한 엔벨로프의 유지는 기존의 선형 스위프 또는 로



〈그림 2.3〉 합성된 역 사인 스위프 신호

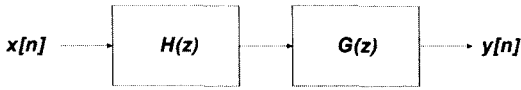
〈표 2.1〉 주파수 별 신호대 잡음비 비교 결과

$f_c$ (Hz) 주파수	125	250	500	1000	2000	4000
Linear (dB)	32	34	65	43	40	40
Log	39	34	70	43	42	45
Arbitrary	39	41	72	51	50	45

그형태의 스위프가 일정한 엔벨로프를 가짐으로써 얻을 수 이득을 그대로 갖게 되는 이유가 된다.

### III. 인간의 인지적 특성을 고려한 혼합적 공간 역 필터링 기법

인간의 인지적 측면을 고려한 혼합적 방식의 공간 역 필터 설계는 청취자가 측정 점으로부터 벗어나는 경우에도 등화(Equalization) 효과를 극대화 하는 것에 목표를 두었다. 일반적으로 청취중의 약간의 머리 이동은 필수적인 것이므로



〈그림 3.1〉 종래의 최소 자승 역 필터의 다이어그램

어느 정도 이상의 고주파 대역까지 LS 기법을 이용하여 위상(phase)까지 등화하는 것은 오히려 음질의 손질을 유발시킴을 확인 하였다. 여기서는 그 기준 주파수를 1.3781kHz로 산정하였다. 이 주파수를 기준으로 아래 대역에서는 LS 기법을 이용한 완전(perfect) 역 필터링을 위 대역에서는 1/3 옥타브 간략화(1/3 octave smoothing) 결과의 역을 주파수 응답으로 갖는 선형 위상 필터링을 수행하였다. 따라서 이 제안하는 방법을 이용하여 전 주파수 영역에서 평탄화된 응답을 얻을 수 있었고, 1.3781kHz 이하의 영역에서는 위상 응답의 보정까지 가능하여 공간에 의한 영향을 받지 않는 결과를 얻을 수 있었다.

### 1. 기존의 방법

#### 가) Least square 기법 기술 구성

그림 3.1은 최소 자승 기법(LS 기법)을 이용한 공간 균등화(room equalization) 시스템을 그린 것이다. 만약 역 필터를 쓰지 않는다면, 입력 신호  $x[n]$ 은 공간 전달 함수  $G(z)$ 으로 인해 왜곡된다. 이를 보상하기 위해서는 공간 역필터(room inverse filter)를 거쳐야 한다. LS방법은 가장 많이 쓰이는 방법 중 하나이다. 공간 전달 함수  $G(z)$ 가 역필터  $H(z)$ 를 거치면 이상적으로는 다음과 같이 표현되는 단위 충격 응답(unit impulse function)이 되어야 한다.

$$G(z)H(z) = 1 \quad (3.1)$$

그러나 실제로는 공간 전달 함수가 여러 위상 특성을 포함하고 있기 때문에, 역 필터 설계 시

에는 두 가지를 고려해야 한다. 신호의 비 최소 위상(non-minimum phase)을 보상하기 위해 모델링 딜레이(Modeling delay)의 도입이 필요하다. 그리고 너무 낮거나 높은 주파수영역은 스피커 특성상 재생될 수 없으므로, 실제 레퍼런스(reference) 신호로는 식 (3.1)과 같은 임펄스가 아닌 밴드 패스된 신호를 사용해야 한다. 이 역 필터는 다음과 같이 표현된다.

$$\sum_{i=0}^{M-1} g[n-i]h[i] = p[n] \quad (3.2)$$

역 필터  $h[n]$ 의 길이는  $M$ ,  $p[n]$ 은 시간지연을 거치고 주파수 응답이 공간 충격 응답과 스피커의 특성을 따르는 레퍼런스 신호이다. 이것은 행렬형태로는 다음과 같이 표현될 수 있다.

$$Gh = p \quad (3.3)$$

$$G = \begin{bmatrix} g[0] & 0 & \dots & 0 & 0 \\ g[1] & g[0] & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ g[N-1] & g[N-2] & \ddots & g[0] & 0 \\ 0 & g[N-1] & \ddots & \vdots & g[0] \\ \vdots & \vdots & \ddots & g[N-1] & \vdots \\ 0 & 0 & 0 & 0 & g[N-1] \end{bmatrix} \quad (3.4)$$

$$h = [h[0], \dots, h[M-1]]^T \quad (3.5)$$

$$p = [p[0], \dots, p[N+M-2]]^T \quad (3.6)$$

LS방법은 다음 비용 함수(cost function)를 최소화하는  $h$  필터를 찾을 수 있다.

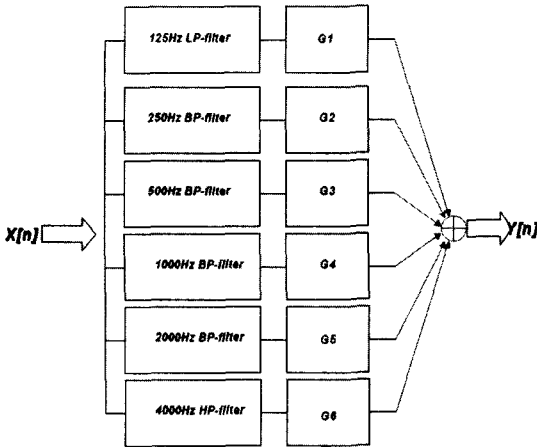
$$J = (Gh - p)^T (Gh - p) \quad (3.7)$$

역필터는 다음과 같다.

$$h = (G^T G)^{-1} G^T p \quad (3.8)$$

#### 나) 옥타브 이득 제어 기법 (octave gain control) 기술 구성

그림 3.2는 6밴드 옥타브 필터 뱅크를 통과하여 각 밴드별로 이득(gain)을 제어하는 방식의 간단한 등화기(equalizer) 구성도 이다. 이 경우

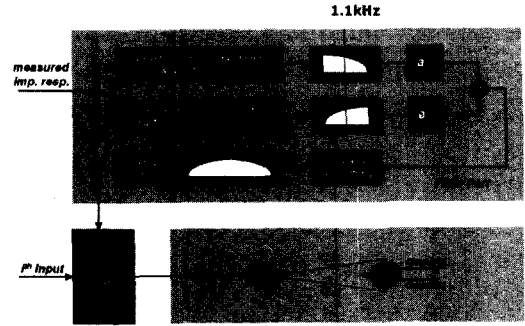


〈그림 3.2〉 종래의 6밴드 옥타브 이득 제어 기법  
(octave gain control) 구성도

목표 주파수 크기 응답 곡선을 설정한 후 이와 공간 전달 함수의 주파수 크기 응답 곡선의 옥타브 밴드 별 평균에너지의 차를 보상하는 방식으로 역 필터링이 행해진다. 보상은 G1부터 G8까지의 이득 값을 조정하는 방법으로 행해진다.

#### 다) 기존 방법의 문제점

현재 최소 자승법과 옥타브 이득 제어 기법 (octave gain control)은 등화기법으로서 많이 사용되고 있다. LS 방법은 음향 소스와 청취자의 위치나 사물의 배치 등 환경적 변화가 전혀 없다는 가정 아래 음향 소스가 역 필터를 통과하였을 때 목표 신호에 대한 에러의 자승이 최소화 되게끔 만들어 주는 기법이다. 이 기법은 혼합 위상 (mixed-phase) 특성을 지닌 임펄스 응답의 위상도 보상이 가능하다는 이점이 있다. 하지만 역 필터를 계산하는 과정에 많은 계산량을 요구한다는 것과 소스나 청취자의 위치가 바뀌는 경우 오히려 신호가 더 많이 왜곡될 수 있다는 것이 단점이다. 그 외에 이득 제어 방법은 신호를 각 밴드로 구별한 후 각 밴드의 파워에 따라 각기



〈그림 3.3〉 제안 알고리즘의 개략도

다른 이득을 설정하여 전체적인 주파수 응답을 평탄하게 만드는 기법이다. 이것은 필터의 길이가 짧고 계산량을 많이 요구하지 않으며 원음을 최소한으로 처리하여 음질 손상을 방지할 수 있다는 장점이 있다. 하지만 주파수 해상도가 매우 낮기 때문에 음향 청취공간의 특성을 제대로 보정할 수가 없고 위상 특성에 대한 고려가 전혀 불가능하다는 단점이 있다.

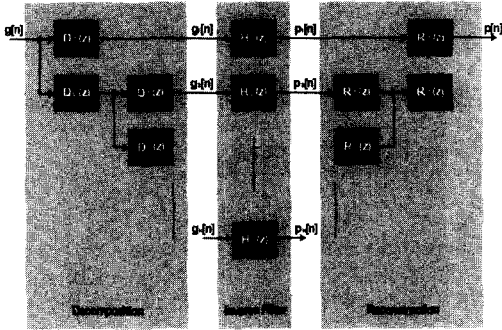
## 2. 제안하는 알고리즘

일반적으로 청취중의 약간의 머리 이동은 필수적인 것이므로 제안하는 알고리즘은 청취자가 측정 점으로부터 벗어나는 경우에도 등화 (Equalization) 효과를 극대화 하는 것에 목표를 둔다. 전반적인 제안 알고리즘의 전체적인 구성은 그림 3.3과 같다. 저주파 대역과 고주파 대역에 다른 방식을 적용시키는 것을 확인할 수 있다. 각 부분에 대한 자세한 설명은 아래와 같다.

#### 가) Wavelet 기반 LS 알고리즘

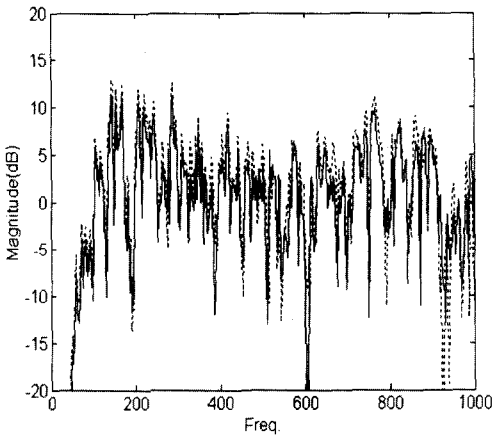
제안 알고리즘은 신호를 wavelet 필터를 사용하여 각 밴드별로 나눈 후 각 밴드의 특성에 맞는 신호처리를 적용한 후 다시 합성하여 역 필터





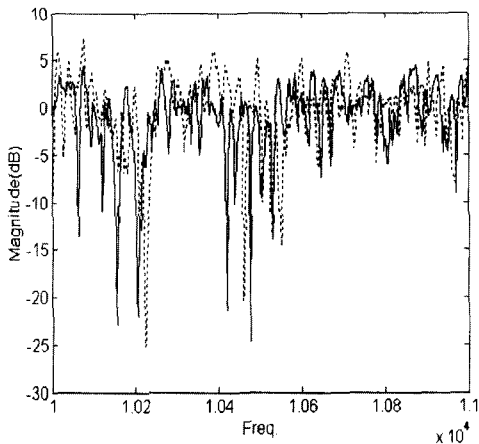
〈그림 3.4〉 Wavelet 기반 LS 알고리즘의 개략도

를 구하는 방법이다. 제안한 알고리즘의 개략도는 그림 3.4와 같다. 여기서  $g[n]$ 은 공간 임펄스 응답,  $g_k[n]$ 는 임펄스 응답이 wavelet 필터를 거쳐 얻어진  $k$  번째 밴드의 신호,  $p[n]$ 은 임펄스 응답이 역 필터를 거친 후 최종적으로 얻을 수 있는 목표 신호, 그리고  $p_k[n]$ 는 각 밴드별 신호가 해당 필터를 통과한 후 얻어지는 밴드별 목표 신호이다. 이제 역 필터를 구하는 문제는 각 밴드에 해당하는 임펄스 응답에 대한 역 필터를 구하는 과정으로 요약된다.

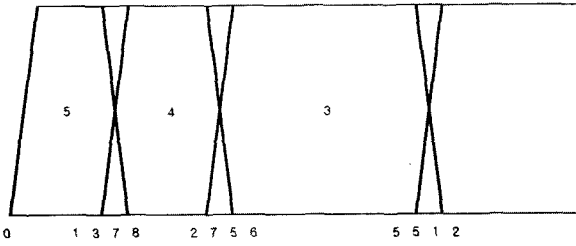


〈그림 3.5〉 6cm 떨어진 두 점에서 측정된 0-1kHz 구간의 측정된 임펄스 응답 신호 비교

Wavelet 변환을 한번 통과 할 때마다 각 서브밴드는 이전 밴드 폭의 1/2에 해당하는 신호를 얻을 수 있다. 샘플링 주파수가 44100인 경우 wavelet 변환을 4번하여 얻어지는 저역 통과 신호는 전체 밴드의 1/16, 즉 1.3781kHz 까지의 성분을 포함한다. 청취자 머리 가운데서 공간 전달 함수를 측정할 경우 머리가 움직이지 않는다는 가정 하에서도 귀와 측정점 간의 거리는 6센티미터 이상이다. 임펄스 응답이 거의 변화가 없는 영역은 주파수에 따라 다르다. 위 경우 파장의 1/10 정도 움직임이 허용된다면 560Hz 이하의 성분만이 변화가 없다. 그림 3.5와 그림 3.6은 실제 측정지점을 6센티미터 이동하여 임펄스 응답을 측정한 후 주파수 분석을 한 것이다. 여기서 확인 할 수 있듯이 1kHz 까지의 임펄스 응답은 거의 변화가 없는 반면에 10kHz 이상의 신호에서는 영점과 극점의 위치가 반대로 나타나는 경우도 확인 할 수 있다. 따라서 1.3781kHz 이상에서는 크기뿐 아니라 위상까지 정확히 역 필터링을 해주는 것은 오히려 신호를 더 왜곡시킬 가능성이 있다. 따라서 측정점을 주변의 여러 곳으로 하여 측정하고 이 여러 충격응답의 대표 응답(representative impulse response)을 구하는 것이 보다 머리이동에 강인한 역 필터링을 하는데 필요하다. 위에서 제안한



〈그림 3.6〉 6cm 떨어진 두 점에서 측정된 10-11kHz 구간의 측정된 임펄스 응답 신호 비교



〈그림 3.7〉 Wavelet 분석 4번 실행한 후의 대역 분포

알고리즘은 신호를 대역별로 구별하여 각각 다른 필터링 기법을 적용한다.

Wavelet 분석을 4번 실행하고 난 후 대역 분포는 그림 3.7과 같다. 여기서 5번 대역은 기존의 LS 기반의 역 필터를 설계하여 적용한다. 이 필터는 주파수 축에서 크기뿐만 아니라 위상까지도 정확히 보정해 주는 필터로서 다음과 같은 수식으로 표현될 수 있다.

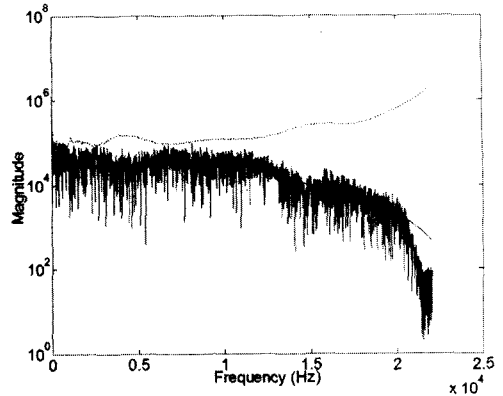
$$h_k = (G_k^T G_k)^{-1} G_k^T p_k \quad (3.9)$$

여기서  $k$ 는  $k$ 번째 subband를 의미하고  $h_k$ ,  $G_k$ ,  $p_k$ 는 각각  $k$ 번째 대역에 해당하는 역 필터, 공간 전달 함수, 목표 신호가 된다. 2,3,4번 대역은 기존의 옥타브 간략화(octave smoothing) 기법을 이용하여 간략화(smoothing) 된 주파수 크기를 보상해 주는 선형 위상 필터를 설계한다. 이렇게 구해진 각 필터를 더하면 각 밴드마다 우리가 원하는 특성을 가지는 필터를 구성할 수 있다.

나) 1/3 octave smoothing 기반 선형위상 보정 알고리즘

간략화에 대한 기본적인 수식은 아래와 같다. (이후  $\otimes$ 는  $\oplus$  즉 circular convolution을 의미함.)

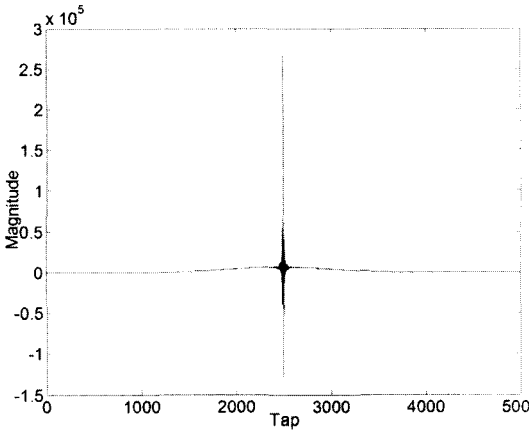
$$H_{sm}^{-2}(k) = |H^2(k)| \otimes W_{sm}(k) = \sum_{i=0}^{N-1} |H[(k-i) \bmod N]|^2 \cdot W_{sm}(i) \quad (3.10)$$



〈그림 3.8〉 1/3 octave frequency magnitude smoothing 결과 (실선)와 그 역(점선)

간략화의 대략적인 순서는 음향신호를 FFT한 후에 이것의 파워  $|H^2(k)|$ 를 구하고 로패스필터  $W_{sm}(k)$ 와 순환 컨볼루션하여, 즉 식 (3.10)에서 보이는 것과 같이 이동평균(moving averaging)하는 것이다. 그러나 FFT는 선형적인 스케일이므로 인간의 청각을 고려한다면, 로패스필터를 주파수의존적으로 하거나 파워자체를 로그스케일로 변형시켜야 한다. 이 중에 간단한 방법은 로패스필터를 상수 대역폭으로 설계하고, 파워를 로그스케일로 인터폴레이션(interpolation)하여 필터링하는 방법이다. 이 후에 다시 원래의 선형 스케일(linear scale)로 환원하게 되면 원하는 정도의 주파수 크기 응답 간략화 효과를 얻을 수 있다 [10].

그림 3.8은 1/3 옥타브 주파수 크기 간략화(1/3 octave frequency magnitude smoothing) 결과 (실선)와 그 역 (점선)을 보였다. 원래의 주파수 응답이 고주파로 갈수록 더욱 간략화 되어가는 결과를 확인 할 수 있다. 기존의 그래픽 등화기(graphic equalizer)와 같은 옥타브 이득 제어 기법이 대역간의 파워 차이를 각 대역에 해당하는 몇 개의 게인 조절을 통해 보상하는 방법이라면 이 방법은 측정된 신호를 바탕으로 주파수 응답

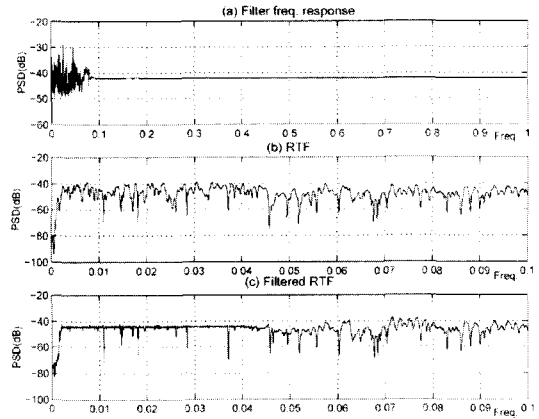


〈그림 3.9〉 선형위상inverse filter

상의 연속적인 평균값을 이용하는 보상 방법이므로 보다 평탄화된 양상의 주파수 응답 보상이 가능하다. 이 경우에는 단순히 레퍼런스 즉 목표 응답(target response)을 평탄화된 응답으로 상정하고 그대로 역을 취하였지만, 만약 원하는 크기 응답 패턴이 정해져 있다면 이를 분모로 하여 역을 취하면 된다. 이렇게 역을 취한 후 이런 크기 응답 패턴을 갖는 선형위상 필터를 설계하여 그림 3.9에 보였다. 이 방법은 주파수 응답 패턴의 간략화된 형태가 측정 위치와 상관없이 거의 일정함을 고려할 때 넓은 청취 지역에서 그 보정 효과를 들을 수 있는 방법이다. 이 경우 0.08 이하의 대역은 고려 대상 영역이 아니므로 그 대역은 역이 아닌 평탄화된 응답을 그대로 두고 구한 필터이다.

다) Wavelet 기반 LS 필터

그림 3.10은 wavelet 기반 LS 필터의 성능을 보였다. 그림 3.7에서 5번 대역에 LS 필터를 적용하고 나머지 대역은 아무런 필터도 적용하지 않았다. 그림 3.10(a)에서 이 필터의 주파수 응답을 살펴보았다. 그 결과 이 필터는 0.08 이하의 신호



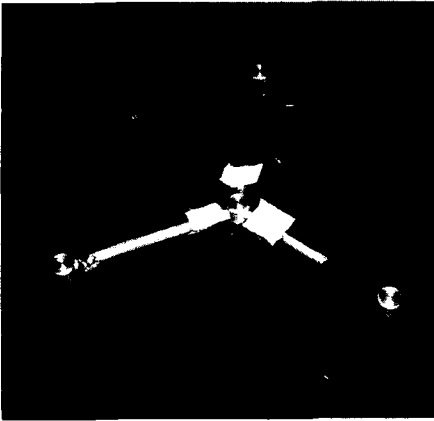
〈그림 3.10〉 wavelet 기반 LS 필터의 성능

에만 영향력이 있다는 것을 알 수 있다. 그림 3.10(b)와 (c)는 공간 전달 함수와 보정된 공간 전달 함수를 0에서부터 0.1까지의 대역만 자세히 살펴본 것이다. (b)에서처럼 많은 극점과 영점을 보이던 공간전달 함수가 필터를 통과한 후에는 (c)에서처럼 평탄한 주파수 응답 특성을 보인다.

IV. 스피커 위치 추적을 이용한 가상스피커 배치 기술

1. 3채널 마이크로폰 시스템 이용한 스피커 위치 추정 알고리즘 개발

홈시어터 시스템과 같이 여러 개의 스피커를 사용하는 시스템의 경우 각 스피커의 최적 위치가 결정되어 있는 것이 일반적이다. 그러나 일반적으로 사용자가 자신의 청취실에 이러한 시스템을 설치할 경우 그 최적위치에 적합하게 설치하기는 쉽지 않다. 이를 해결하기 위해 여기서는 자신의 청취 룸에서 청취 위치 대비 각 스피커의 위치를 파악해 주는 알고리즘을 제안한다. 이를 통해 사용자로 하여금 각 스피커의 최적 위치 선



(그림 4.1) 3-채널 마이크로폰 시스템

정을 용이하게 해주고, 이것이 불가능 할 경우에는 마치 최적위치에 위치해 있는 것과 같은 효과를 낼 수 있도록 튜닝해 주기 위한 기본 데이터로 활용할 수 있을 것이다.

### 가) 3 채널 마이크로폰 시스템을 이용한 측정

그림 4.1에서 보는 것처럼 특성이 거의 유사한 3개의 마이크로폰을 통해 스피커로부터 발생되는 사인 스윙 음원을 녹음한다. 일반적인 공간의 충격 응답 측정을 각기 다른 세 개의 채널에서 동시에 행하기 때문에 각 채널 별 충격 응답들의 직접 음 시간 지연 차이를 구해낼 수 있고, 이러한 시간 지연 정보 즉 스피커로부터 각 마이크로폰까지의 거리 차(이는 시간 지연정보와 소리의 공기 중 전파속도  $340\text{m/s}$ 를 곱해서 얻을 수 있다.)와 3-채널 시스템의 기하학적 구조를 이용하여 서론 부에 언급한 것과 같이 2차원 평면상에서의 스피커의 위치를 구해 낼 수 있다. 이러한 방법은 음원의 평면 파 전달을 전제로 한다. 스윙 사인 음원의 녹음은 멀티트랙 녹음(multi track recording)이 가능한 하드웨어와 소프트웨어를 사용하여 수행 할 수 있다.

### 나) 스피커 위치 추정 알고리즘

#### · 극점 찾기 알고리즘

(Peak detecting algorithm)

이 알고리즘은 기준채널에서의 직접 음 수음 시점을 구해내는 알고리즘이다. 가장 편리하기로는 처음으로 들어오는 가장 큰 극점(peak)의 시점을 구해내기만 하면 되지만 이 방법을 이용하면 각 반사 음 중에서의 두드러지는 극점들까지도 함께 구하는 것이 가능하다. 이러한 정보는 6장에서 다루는 음장 데이터 확보를 위해 이용될 수 있을 것이다. 보다 정확한 수음 시점을 구해내기 위해 스플라인 보간(spline interpolation)을 10배로 수행하였다.

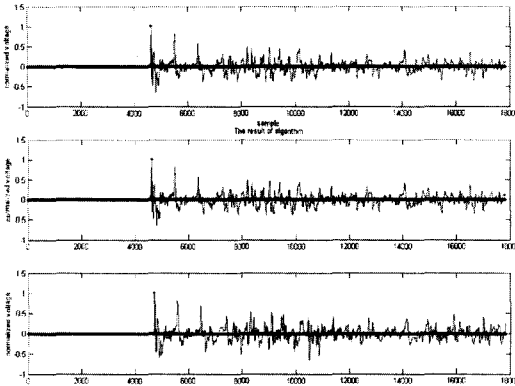
직접 음 및 두드러진 반사음들의 극점들을 찾기 위해서는 각 신호들의 극대점들을 찾아가는 방식을 사용하였다. 이러한 극대점들 찾기를 반복적으로 수행할수록 더욱더 두드러지는 극점들만 남게 된다.

#### · 쌍 찾기 알고리즘(Pair matching algorithm)

우선 수행된 극점 찾기 알고리즘으로부터 구해진 기준 채널의 수음 시점으로부터 다른 채널의 극점 도달 시점을 구해주는 알고리즘이다. 이것은 각 채널간의 상호상관(cross-correlation)을 수행하고 가장 그 정도가 높은 시점을 시간 지연차로 추정하는 방식을 이용한다. 스피커 응답 길이를 고려한 블록(block)을 기준 채널의 수음 시점을 중심으로 설정되고 같은 크기의 다른 채널의 블록(block)을 시점을 이동해 가면서 상호상관을 수행한다. 이 알고리즘의 수행 결과는 그림 4.2에서 확인 할 수 있다.

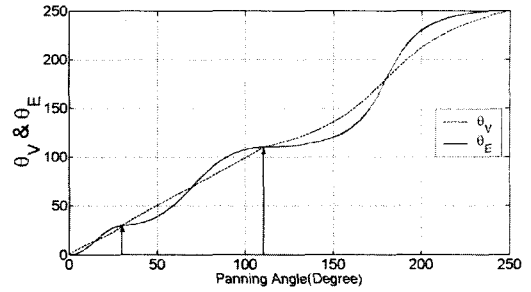
## 2. 가상 스피커 배치 알고리즘 개발

가상 음원을 스피커 입력 신호의 이득(gain)



〈그림 4.4〉 pair matching algorithm 실행 예

을 이용하여 정위시키기 위한 다양한 패닝 방법(panning law)이 존재한다. 정위시키고자 하는 방향의 좌우 스피커만을 이용하는 pair-wise 정위 기법과, 3개의 스피커를 이용하는 triple-wise 기법 그리고 모든 스피커를 이용하는 앰비소닉(ambisonic) 기법 등이 사용되고 있다. 이러한 기법상의 다양성에도 불구하고 대부분의 시스템의 경우 정확한 음상 정위에 대한 필요성의 부족으로 구현이 용이한 pair-wise 기법만을 사용하고 있다. 하지만 오늘날 대용량 디지털 미디어 기술의 발전으로 서라운드 음향이 보편화 되면서 서라운드 음상 정위에 대한 많은 연구가 이루어지고 있으며 기존 정위 기법의 문제점을 극복하기 위한 많은 연구가 진행되고 있다. 정위 기술은 크게 두 가지로 구분할 수 있는데 각 스피커에 입력 게인 만을 조절하는 게인 패닝(gain panning)과 인간의 양 귀에 주파수 특성을 고려한 HRTF 패닝 등으로 나눌 수 있다. HRTF 패닝 방법은 음상 정위에 있어서는 효과를 나타내지만 음질의 저하가 심해 아직 널리 쓰이지 않고 있다. 또한 게인 패닝(gain panning)의 경우도 많은 방식들이 있지만 대부분의 경우 일정 파워 정위 방식(constant power



〈그림 4.5〉 CPP law에서의 분석

panning)을 사용한다.

#### 가) Constant power panning

Pair-wise 정위의 경우 정위시키고자 하는 모든 위치에서 일정한 파워(power)를 갖는 것이 무엇보다 중요하기 때문에 서라운드 시스템에서 일정 파워 정위 기법에 대한 연구가 활발히 진행되고 있다. 지금까지 널리 사용되고 있는 일정 파워 정위 기법은 아래와 같이 나타낼 수 있다.

$$\sin^2 \theta + \cos^2 \theta = 1 \quad (4.1)$$

한 채널의 게인은 에 따라 변하고 다른 채널의 게인은 에 따라 변하게 되어 전체 power는 항상 일정하게 된다.

$$\theta_m = \frac{\theta_{pan} - \theta_1}{\theta_2 - \theta_1} \times 90 \quad \text{where } \theta_2 \geq \theta_{pan} \geq \theta_1 \quad (4.2)$$

$$g_1 = \cos \theta_m, \quad g_2 = \sin \theta_m \quad (4.4)$$

이 방법은 사용자가 정위하고자 하는 각도  $\theta_{pan}$ 과 실제 정위되는 각도  $\theta$ 와는 차이를 나타내는 문제점을 가지고 있다. 이러한 현상은 “detant effect”라고 명시되어 있을 정도로 널리 알려져 있는 문제점이다. 이러한 문제점은 스테레오 시스템에서는 크게 문제가 되지 않아 지금까지 무리 없이 사용되어 왔지만 서라운드 시스

템에서는 문제점이 대두되었다.

나) 방향심리인자(Directional Psychoacoustic parameter)

방향심리인자에 대한 연구는 영국의 음향학자 Gerzon에 의해 체계화 되었으며 서라운드 녹음, 재생 그리고 시스템 설계에 이르기까지 다양한 분야에 적용되고 있다 [11]. Gerzon은 패닝 방법을 최적화하기 위해 속도 벡터(velocity vector), 에너지 벡터(energy vector)의 개념을 도입하였다. 이 중에서 에너지 벡터는 음상 정위에 있어 가장 중요한 부분을 차지하고 있다.  $i$ -번째 스피커의 gain  $g_i$ 에 대한 전체 스피커로부터의 크기의 합

$$\text{전체 에너지} \quad \sum_{i=1}^n g_i^2 \quad (4.5)$$

$$\text{벡터 } x\text{축 성분} \quad \sum_{i=1}^n g_i^2 \cos \theta_i \quad (4.6)$$

$$\text{벡터 } y\text{축 성분} \quad \sum_{i=1}^n g_i^2 \sin \theta_i \quad (4.7)$$

이 벡터를 전체 크기의 합으로 나누어줌으로써 단위 벡터로 정규화 시킬 수 있다. 이때의 단위 벡터의  $x$ 축 성분은

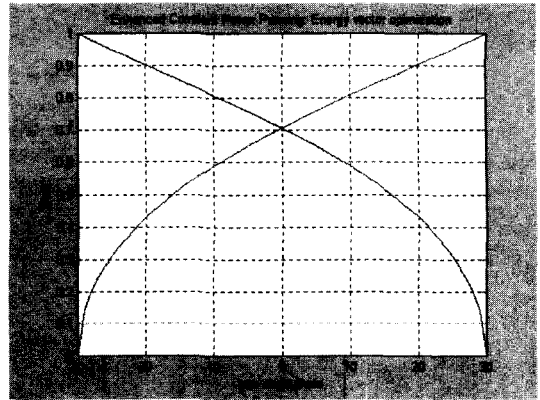
$$r \cos \theta = \frac{\sum_{i=1}^n g_i^2 \cos \theta_i}{\sum_{i=1}^n g_i^2} \quad (4.8)$$

$y$ 축 성분은

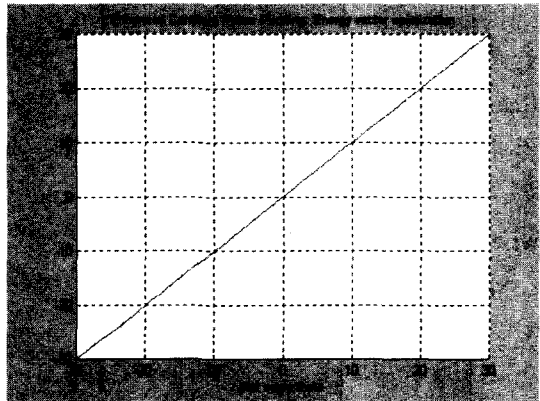
$$r \sin \theta = \frac{\sum_{i=1}^n g_i^2 \sin \theta_i}{\sum_{i=1}^n g_i^2} \quad (4.9)$$

이때  $\sum_{i=1}^n g_i^2$ 을 전체 에너지 게인(total energy gain)이라 하고  $r$ 을 에너지 벡터 크기(energy vector magnitude)라고 하며  $\theta$ 를 에너지 벡터 방향(energy vector direction)이라고 한다.

$$\theta_E = \tan^{-1} \left( \frac{\sum_{i=1}^n g_i^2 \sin \theta_i}{\sum_{i=1}^n g_i^2 \cos \theta_i} \right) \quad (4.10)$$



(a)

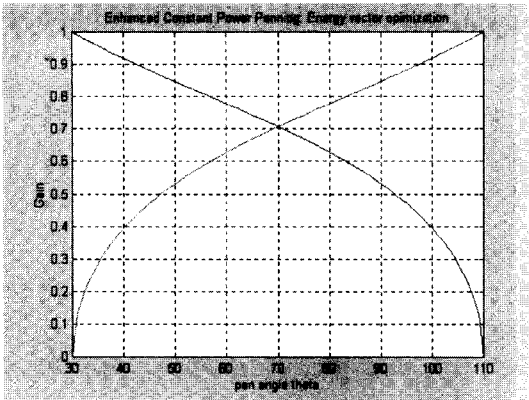


(b)

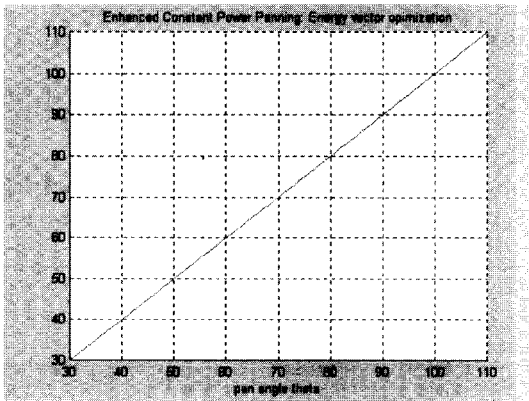
〈그림 4.6〉 전면 스피커 배치에서의 제안하는 패닝 알고리즘 (a) 양 스피커의 각도에 따른 게인 분포도 (b) 패닝 각도에 따른 에너지 벡터 방향

식 (4.10)에 보이는 에너지 벡터 방향은 700Hz 이상 대역에서의 음상 정위를 설명하고 있으며 청취자 이동시의 모든 주파수 대역을 포함하고 있어 방향음향심리인자의 최하위개념에서 가장 중요한 부분을 차지하고 있다.

다) 개선된 일정한 파워를 유지하는 패닝 알고리즘(Enhanced Constant Power Panning Algorithm)



(a)



(b)

〈그림 4.7〉 측면 스피커 배치에서의 제안하는 패닝 알고리즘 (a) 양 스피커의 각도에 따른 개인 분포도 (b) 패닝 각도에 따른 에너지 벡터 방향

에너지 벡터 방향을 최적화시키기 위한 알고리즘은 다음과 같다.

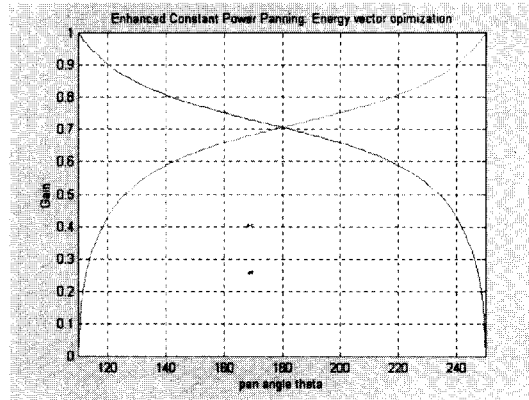
$$\theta_E = T(\theta) \quad (4.11)$$

이때  $\theta_E = \theta_{pan}$  을 만족하는  $\theta_{target}$  을 찾는다.

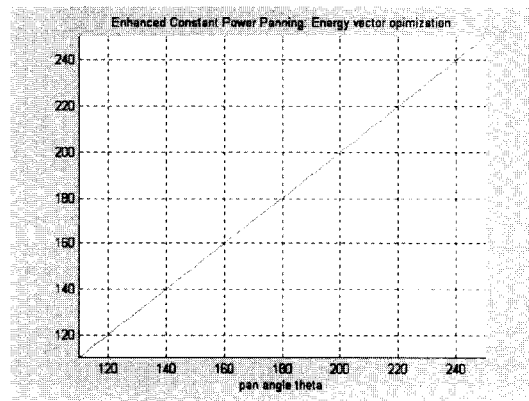
$$T(\theta_{target}) = \theta_{pan} \quad (4.12)$$

따라서  $\theta_{target}$  은 아래의 식으로 구할 수 있다.

$$\theta_{target} = T^{-1}(\theta_{pan}) \quad (4.13)$$



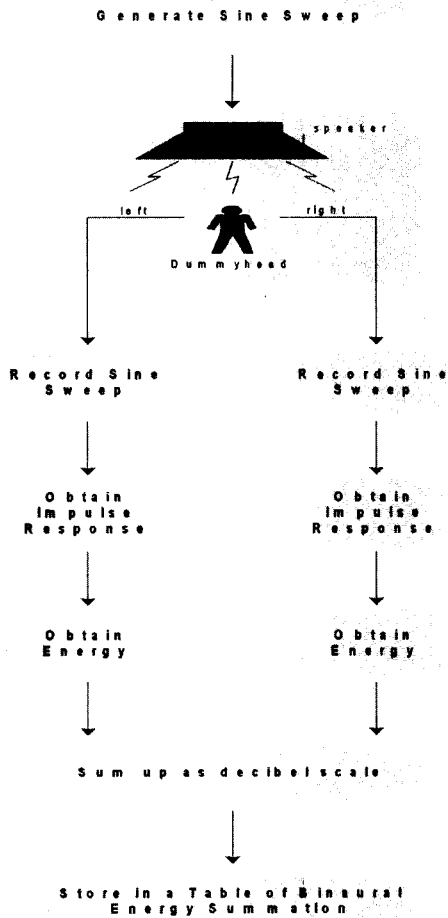
(a)



(b)

〈그림 4.8〉 후면 스피커 배치에서의 제안하는 패닝 알고리즘 (a) 양 스피커의 각도에 따른 개인 분포도 (b) 패닝 각도에 따른 에너지 벡터 방향

위의 과정을 통해서  $\theta_{pan} = \theta_E$  를 만족하도록 하여 방향심리인자를 최적화 시키는 패닝 알고리즘을 찾을 수 있다. 이 알고리즘은 또한 일정한 파워를 유지하기 때문에 기존의 모든 패닝 법칙을 대체할 수 있으며 음상 정위가 스피커 방향으로 치우치는 문제점 또한 제거할 수 있다. 그림 4.6, 4.7, 4.8을 통해 문제가 해결되고 있음을 확인할 수 있다.



〈그림 5.1〉 각 방향에서의 바이노럴 에너지 합 측정 흐름도

## V. 다채널 레벨 보정 알고리즘 개발

기존의 레벨 보정은 각 채널의 절대적인 크기를 같게 보정하였다. 하지만, 인간은 소리를 들을 때 같은 크기의 소리라도 방향에 따라 다른 크기로 듣는다. 따라서 기존의 경우는 거리에 대한 레벨 보정만 할 뿐, 방향에 대한 보정은 하지 않는다. 다 채널의 경우는 각 스피커가 서로 다른 방향에 위치하고 있으므로 방향에 대한 레벨 차이를 레벨 보정에 적용하여 기존의 것보다 보

다 현실적인 보정을 할 수 있다. 스피커의 위치는 4장에서 살펴본 스피커 위치 추적 알고리즘을 통해 파악할 수 있다. 따라서 방향에 대한 레벨 차이를 보정하기 위해 미리 배치 가능한 위치에서의 바이노럴 에너지 합(total binaural energy)을 구하여 표를 만들고, 하나의 마이크론으로 측정된 에너지와 비교하여 레벨 보정을 수행한다.

### 1. 실험 방법

실험은 그림 5.1의 흐름도를 따른다.

· 측정조건 (무향실, 잔향실)

거리 : 0.5m에서 2.5m까지 0.25m간격으로  
(0.5m : 일반적인 거실에서 서라운드 스피커가 가장 가까이 배치 될 수 있는 거리)

각도 : 0도에서 180도까지 10도 간격으로

측정신호 : 사인 스윕

측정 마이크론 : 더미헤드

· 다음 식처럼 모든 위치에서 더미헤드의 양쪽 귀 IR을 구하여 에너지의 합을 구한다.

$$E_i = (\sum_1^n d_l^2 + \sum_1^n d_r^2) / n \quad (5.1)$$

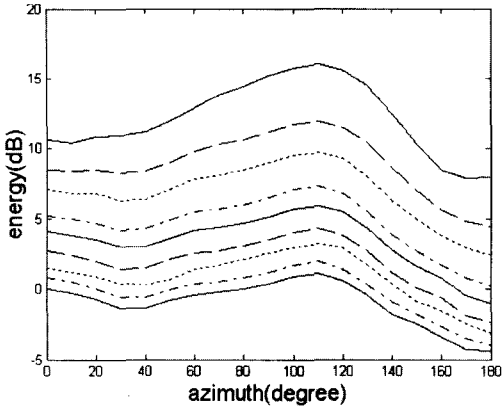
· 모든 데이터를 구하여 테이블을 만든다.

· 스피커 위치를 알면 이 테이블에서 각각의 레벨 차이를 구하여 각 스피커의 계인에 적용한다.

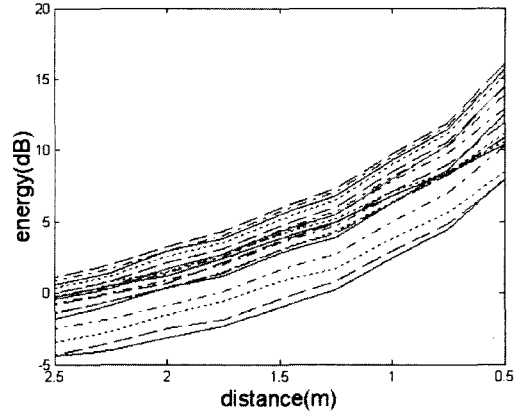
### 2. 측정 결과 및 분석

그림 5.2, 5.3에서 보듯이 110도에서 가장 큰 레벨을 가진다. 이는 뒤에 자리하는 서라운드 채널에 대해 가장 민감하여 서라운드 채널의 레벨

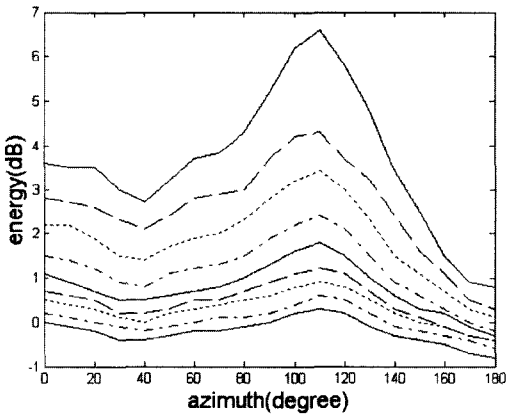




〈그림 5.2〉 무향실에서 각도에 따른 바이노럴 에너지 합



〈그림 5.4〉 무향실에서 거리에 따른 바이노럴 에너지 합

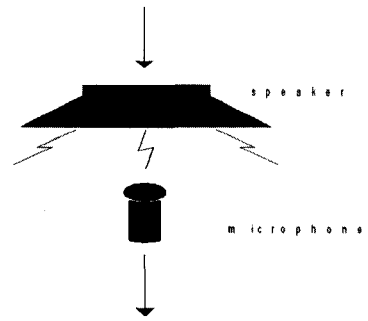


〈그림 5.3〉 잔향실에서 각도에 따른 바이노럴 에너지 합

이 메인 채널보다 작아야 한다는 것을 보여준다. 그리고 잔향실에서 측정한 그림 5.3을 보면, 40도에서도 작아지는 경향을 볼 수 있는데, 이는 메인 L, R 채널이 센터 채널보다 약간 커야 함을 보여준다. 하지만 이는 상대적으로 작은 값이기 때문에 서라운드 채널보다는 중요하지 않다.

그림 5.3, 5.4는 거리에 따른 그래프인데 무향실에서는 레벨이 정확히 감소하는 경향을 보이지만, 잔향실에서는 잔향의 영향으로 민감하지 않은 방향에서의 레벨은 그렇게 큰 차이를 보이지 않고, 민감한 방향에서는 무향실과 같은 경향

Generate Sine Sweep



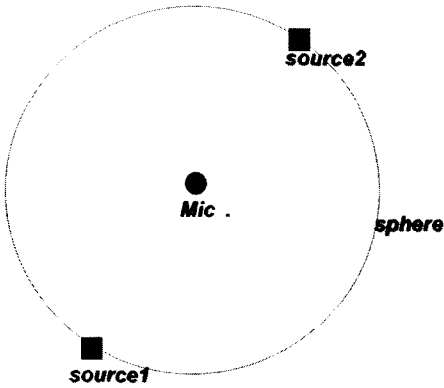
Record Sine Sweep

Obtain Impulse Response

Obtain Energy

Compensate the Difference according to the Table

〈그림 5.5〉 레벨 보정 과정



〈그림 6.1〉 4개의 수음점만 가지는 경우에 발생할 수 있는 문제점

을 보인다.

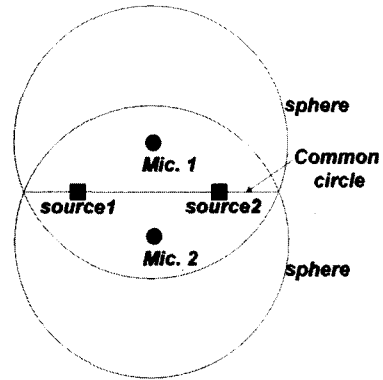
### 3. 레벨 보정 과정

스피커의 위치를 알고 있으므로 그림 5.5와 같은 방법으로 하나의 마이크로폰으로 각 채널의 레벨을 측정하여 미리 측정해 놓은 테이블에서 각각의 위치에 맞는 값을 찾아 그 만큼씩 보정해 주면 된다.

## VI. 가상 음장 시스템의 구현

### 1. 5채널 마이크로폰 시스템을 이용한 공간에서의 직접음 및 반사음의 분포 패턴 파악

실재감 있는 입장감 제어를 위해서는 모사하고자 하는 송신 음장 공간에서 직접음과 반사음의 위치를 찾아내어 이를 재생환경에서 비슷하게 구현해 낼 수 있어야 한다. 직접음 및 반사음의 위치를 찾기 위해 기존에는 4개의 마이크로폰을 이용하는 방법이 제안되었고 널리 이용되

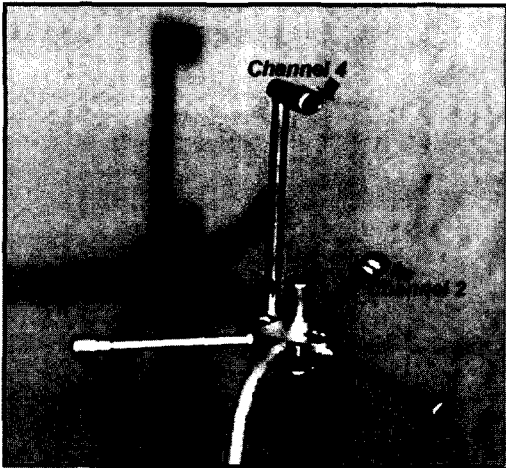


〈그림 6.2〉 한 개의 부가 마이크가 존재하는 경우에 발생할 수 있는 문제점

어 왔다 [12], [13]. 이 방법은 3개의 독립적인 기저를 이용하여 3차원상의 임의의 점을 표현할 수 있다는 이론을 그 기반으로 제안되었다. 그러나 이렇게 4개만의 마이크로폰을 사용할 때는 그림 6.1에서 보이는 상황과 같이 같은 거리에 존재하는 두 개 이상의 음원의 위치를 따로 분별해 내지 못하는 단점을 갖는다.

따라서 이론적으로는 4개의 마이크로폰으로 이루어진 3개의 독립적인 기저만 있으면 되지만, 현실적으로는 그 이상의 마이크로폰을 필요로 하는 것이다. 만약 한 개의 여분 마이크로폰을 허용한다면 이러한 현상이 두 개 이상의 마이크로폰에서 동시에 발생하는 경우만 문제가 되고 그림 6.2에서와 같이 이때는 두 구의 교선인 공통 원 상에 다수의 음원이 존재할 경우에만 발생한다. 더 나아가서 두 개의 여분의 마이크로폰을 허용한다면 두 원의 교점인 두 개의 점에 음원이 존재하는 경우에만 문제가 발생하고, 세 개의 여분의 마이크로폰을 허용한다면 이러한 문제점을 완벽하게 해결할 수 있다.

따라서 7개 이상의 마이크로폰을 가진 구조를 가져야 복합 음원이 겹쳐져 보이는 현상을 해결



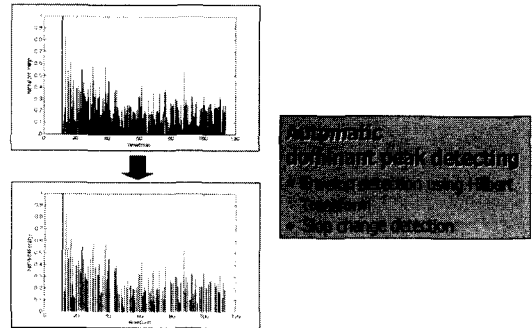
〈그림 6.3〉 5채널 마이크로폰 시스템

할 수 있다. 물론 이러한 여유분의 채널을 가진 구조의 어느 4점을 선택하여도 세 개의 독립된 기저를 가져야 한다. 그러나 그림 6.3에서 보이는 한 개의 여분 마이크가 부가된 5 채널 마이크로폰 시스템을 이용하여 충분히 좋은 효과를 얻을 수 있음이 연구되었다 [7], [13].

이러한 시스템을 이용하여 직접음 및 반사음원의 위치를 파악하기 위해서는 각 마이크로폰 채널로 들어오는 음원의 시간 위치를 각각 파악하여야 한다. 각 채널별 음원의 시간 위치를 자동으로 파악할 수 있게 만들어주는 알고리즘은 4장에서 살펴보았다[13]. 이의 수행 결과를 그림 6.4에서 확인할 수 있다. 또한 이러한 기준 채널에서 파악된 음원 위치 근처에서 다른 채널과의 유사도 분석을 통해 그 음원의 각 채널 도달 시간위치를 추정해 낸다. 그림 6.5는 이 알고리즘 수행후의 결과를 나타낸 것으로 검은 점으로 표시된 부분이 파악된 각 채널별 음원들의 시간 위치이다.

파악된 시간 위치  $T_i(i=1,2,3,4,5)$ 와 음파 전달 속도  $C$ 를 이용하여 아래의 수식을 이용하여  $X-$

Post processing of 5-channel microphone system (Peak detection)



〈그림 6.4〉 극점 찾기 알고리즘의 개략도

$Y-Z$  좌표상의 음원의 위치를 구해낼 수 있다. 여러 가지 다른 양상의 수

식이 존재할 수 있다. 채널 1,2,3,4의 정보를 이용할 경우는

$$\begin{aligned} X &= (T_2^2 - T_3^2) \times C^2 / 2L \\ Y &= (-2T_1^2 + T_2^2 + T_3^2) \times C^2 / 2\sqrt{3}L \\ Z &= (-3T_4^2 + T_1^2 + T_2^2 + T_3^2) \times \sqrt{3}C^2L/9 - \sqrt{3}X/36 \end{aligned} \quad (6.1)$$

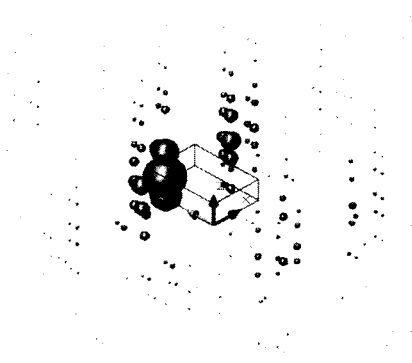
1,2,3,5의 정보를 이용할 경우는

$$\begin{aligned} X &= (T_2^2 - T_3^2) \times C^2 / 2L \\ Y &= (-2T_1^2 + T_2^2 + T_3^2) \times C^2 / 2\sqrt{3}L \\ Z &= \sqrt{T_5^2 \times C^2 - X^2 - Y^2} \end{aligned} \quad (6.2)$$

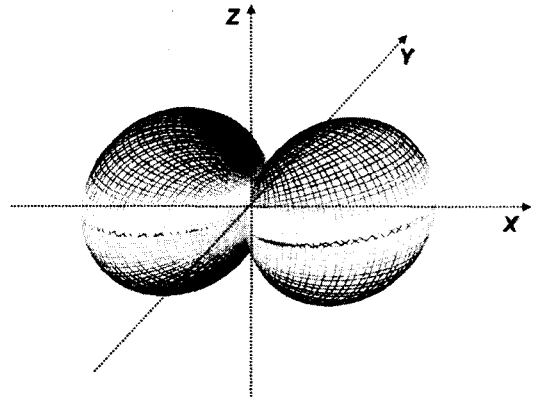
1,2,4,5의 정보를 이용할 경우는

$$\begin{aligned} X &= (T_2^2 - T_1^2) \times C^2 / L - \sqrt{3}Y \\ Y &= ((T_3^2 - T_1^2) \times C^2 + 7/24 \times L^2 + L/\sqrt{6} \times Z) / L/2 \times \sqrt{3} \\ Z &= (T_5^2 - T_4^2) \times C^2 / (\sqrt{6} \times L/2) + \sqrt{6} \times L/8 \end{aligned} \quad (6.3)$$

1,3,4,5의 정보를 이용할 경우는



〈그림 6.6〉 직접음 및 반사음의 위치와 에너지 정보



〈그림 6.7〉 8차형 가중 함수

$$\begin{aligned}
 X &= (T_1^2 - T_3^2) \times C^2 / L + \sqrt{3} \times Y \\
 Y &= ((T_3^2 - T_1^2) \times C^2 + 7/24 \times L^2 + L \sqrt{6} \times Z) / L 2 \times \sqrt{3} \\
 Z &= (T_3^2 - T_4^2) \times \sqrt{3} / 2L + L / 2\sqrt{3}
 \end{aligned}
 \tag{6.4}$$

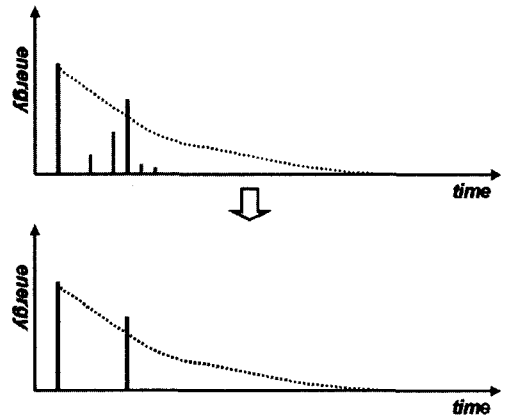
2,3,4,5의 정보를 이용할 경우는

$$\begin{aligned}
 X &= (T_1^2 - T_3^2) \times C^2 / 2L \\
 Y &= \sqrt{T_3^2 \times C^2 - X^2 - Z^2} \\
 Z &= (T_3^2 - T_4^2) \times \sqrt{3} / 2L + L / 2\sqrt{3}
 \end{aligned}
 \tag{6.5}$$

이렇게 얻어진 직접음 및 반사음의 위치와 에너지 정보를 그림 6.6과 같이 표현할 수 있다.

## 2. 가상 음장의 입장감 구현

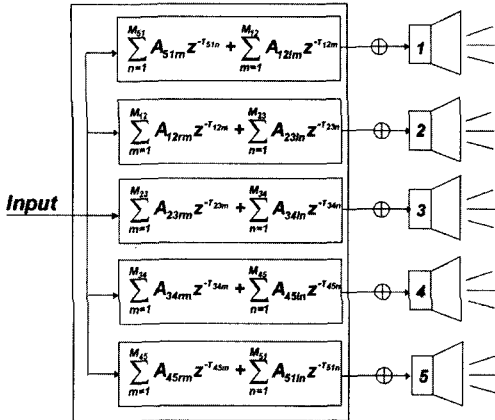
앞장에서 얻어진 직접음 및 반사음의  $(r, \phi, \theta)$  와 에너지 정보를 이용하여 입장감을 구현하는 방식은 전 절에서 살펴본 음상 정위 방식을 그대로 이용할 수가 있다. 그러나 전 방향 음원에 대해서 그대로 에너지를 적용하지 않고 공간감을 느끼는데 있어서 중요한 음원의 방향 성분에 보다 더 가중을 두어 적용하게 된다. 이는 인간의 인지적 측면을 이용하여 간략화(simplification) 작업을 하는 것으로 이러한 간략화 작업은 특히나



〈그림 6.8〉 사후 마스크를 고려한 의미 있는 도달 음원의 추출

다채널 환경의 경우 그 필요성이 높아지는데, 이는 모든 채널에 동시에 그 간략화 정도가 적용되기 때문이다.

이를 위해 첫 번째로 반사 음원의 방향에 따라서 다른 가중 방식을 적용한다. 일반적으로 공간감 인지를 위해서는 측면에서 오는 반사음이 가장 중요함이 알려져 있으므로 [6], 그림 6.7과 같은 가중 함수를 방향에 따라 우선적으로 적용시킨다. 즉 건축음향 분야에서 널리 쓰이는 방법처럼 측면 방향에 보다 더 큰 가중을 주기 위해 8차형 가중 함수를 적용시키는 것이다. 이 경우 천



〈그림 6.9〉 다채널 시스템에서 가상 음장 구현을 위한 개선된 CPP 법칙을 이용하는 경우에 대한 개념도

장이나 바닥 또는 정면이나 후방에서 오는 반사음들의 효과는 상대적으로 감소되고 측면 방향의 효과를 극대화하는 효과를 얻을 수 있다.

두 번째로는 인간 귀의 시간적 마스킹(temporal masking) 효과를 이용한다. 시간적 마스킹 효과는 그림 6.8에서 보이는 것처럼 마스킹 신호가 생기기 전이나 없어진 후에 발생하는데, 일반적으로 생긴 후에 발생하는 약 50~200ms의 사후 마스킹(post-masking)을 이용한다. [14]. 이러한 시간적 마스킹을 이용하여 한계치(threshold)를 넘지 않는 경우의 신호는 제거하는 방식으로 의미 있는 반사음 위치를 선택하는 과정을 거치게 된다. 그림 6.8은 이를 도식적으로 보여주고 있다. 만약 웨이브렛(wavelet) 등의 필터 뱅크(filter bank)를 이용하는 분석(analysis)과 합성(synthesis)이 가능하다면 이러한 압축 정도를 주파수 의존적으로 수행하여 보다 높은 효과를 얻을 수 있을 것이다. 물론 음상 정위 정보를 그대로 유지하기 위해서는 각 채널에서 이러한 효과를 동시에 추구해야 한다. 따라서 인접

두 채널 간에 하나라도 마스킹 한계를 넘는 경우가 발생하면, 간략화를 수행하지 않는다.

이후에 X-Y 평면상에만 설치되어 있는 재생 시스템을 이용하여 개선된 CPP 법칙은 적용하게 되는 경우 고도 정의가 불가능하므로 이를 X-Y 평면상으로 사영(射影, projection) 한다. 아니면 머리전달함수를 이용하여 각각의 직접음 및 반사음원을 시간 지연을 이용하여 정위시킬 수 있을 것이다. 그림 6.9는 다채널 시스템에서 개선된 CPP 법칙을 이용하는 경우에 대한 개념도이다. 여기서  $M_{ij}$ 는 인접한  $ij$  스피커 사이에 정위할 수 있는 반사음의 개수를 의미하며  $A_{ijlm}$ 에서  $ij$ 는 인접한  $ij$  스피커임을  $l$ 은 이 두 스피커의 왼쪽(left)에 해당됨을  $m$ 은 음원의 개수에 대한 인덱스에 해당됨을 의미한다. 또한  $ijm$ 에서  $ij$ 는 인접한  $ij$  스피커임을  $m$ 은 음원의 개수에 대한 인덱스에 해당됨을 의미한다. 이것은 각 직접음 및 반사음원의 위치에 따라 이를 재생할 인접 스피커  $ij$ 를 선택하고 이 두 스피커의 신호 입력 크기  $A$ 를 개선된 CPP 법칙을 적용하여 각도에 따라 정해주고 음원의 상대적 위치를 지연시간을 통해 구현하는 방식이다.

### Ⅶ. 요약

본 논문에서는 다채널 오디오 시스템에서 사용되어지는 음향신호처리에 대하여 저자들이 그간 제안하였던 방법들을 중심으로 다뤘다. 다룬 내용은 첫째로 각 스피커에서 청자의 위치까지의 공간 응답의 영향을 배재할 수 있는 역 필터링이고, 둘째로 다채널 스피커의 위치를 파악하여 최적 위치와의 차이를 자동적으로 보상할 수 있는 방법이다. 셋째로 인간의 인지적인 측면을 고려하는 다채널 스피커로부터의 에너지 레

벨 정렬 방식에 대하여 다뤘고, 마지막으로 특정 공간의 반사음의 분포 패턴을 구하여 이를 일반적인 청취공간에서 재현해 내는 음장 재현 방식에 대하여 요약 하였다.

### 참고문헌

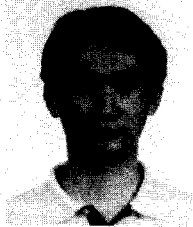
- [1] L. -H. Kim, J. -J. Jeon, S. -I. Lee, and K. -M. Sung, "Hybrid Equalization of Room for Home Theatre System" Audio Engineering Society 115th convention, October 2003.
- [2] L. -H. Kim, J. -S. Lim, C. Choi and K. -M. Sung, "Equalization of Low Frequency Response in Automobile", IEEE Trans. on Consumer Electronics, Vol. 49, no. 1, pp. 243-252, February 2003.
- [3] M. Nakahara, "Multichannel Monitoring Tutorial Booklet", rev 2.3.0 2002 Yamaha Co. & 2002 Sona Co., June 2002.
- [4] S. -I. Lee, K. -Y. Han, S. -R. Lee, and K. -M. Sung, "Enhanced Constant Power Panning Law for Surround Sound Systems", submitted to IEEE Transaction on Consumer Electronics
- [5] M. A. Gerzon, "General metatheory of auditory localization", 92nd Convention of the Audio Engineering Society, J. Audio Eng. Soc., preprint 3306, May 1992.
- [6] H. Kuttruff, ROOM ACOUSTICS Third Edition (Elsevier applied science, London and New York, 1991.
- [7] C. Choi, L. -H. Kim, S. Doo, Y. Oh, and K. -M. Sung, "Measurement of Early Reflections in a Room with 5 Microphones System", IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, Vol. e 86-A, no. 12, pp. 3283-3287, December 2003.
- [8] D. Lee, and K. -M. Sung, "Sound field processor for creating virtual concert hall impression in home", IEEE Trans. on Consumer Electronics, Vol. 41, no.2, May 1995.
- [9] I. S. Muller and P. Massarani, "Transfer-Function Measurement with Sweeps", J. Audio Eng. Soc., Vol. 49, no. 6, June 2001.
- [10] 심 환, 김래훈, 성평모, "Constant Bandwidth Filtering을 이용한 Complex Fractional-Octave Smoothing", 한국음향학회 하계 학술대회 논문집, pp. 627-630, July 2003.
- [11] M. A. Gerzon, "General metatheory of auditory localization", 92nd Convention of the Audio Engineering Society, J. Audio Eng. Soc., preprint 3306, May. 1992.
- [12] Y. Yamazaki, and T. Ito, "Measurement of spatial information in sound field by closely located 4-point microphone method", J. Acoustics. Soc. Jpn. (E), 10(2), 101-110, 1989.
- [13] 김래훈, "실내의 초기반사음 측정을 위한 향상된 알고리즘의 개발", 서울대학교 전기컴퓨터 공학부, 석사학위 논문, 2001.
- [14] J. G. Beerends, and J. A. Stemerdink, "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation", J. Audio Eng. Soc., Vol. 40, No. 12, 1992 December.

저자소개



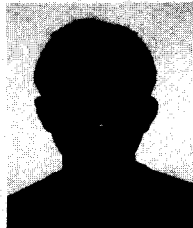
김래훈

1999년 서울대학교 전기공학부 학사  
 2001년 서울대학교 전기공학부 석사  
 2002년 - 현재 서울대학교 전기공학부 박사과정



전재진

1998년 서울대학교 전기공학부 학사  
 2001년 서울대학교 전기공학부 석사  
 2001년 - 현재 서울대학교 전기공학부 박사과정



이신렬

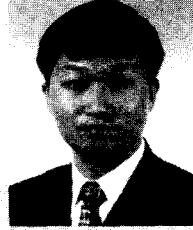
1999년 광운대학교 컴퓨터공학과 학사  
 2002년 서울대학교 전기공학부 석사  
 2002년 - 현재 서울대학교 전기공학부 박사과정



김세웅

2002년 서울대학교 기계공학부 학사  
 2004년 - 현재 서울대학교 전기공학부 석·박사 통합과정

저자소개



임준석

1986년 서울대학교 전자공학과 학사  
 1988년 서울대학교 전자공학과 석사  
 1996년 서울대학교 전자공학과 박사  
 1988년 - 1993년 국방과학연구소  
 1996년 - 1998년 LG종합기술원  
 1998년 - 현재 세종대학교 전자공학과 부교수



성평모

1965년 - 1971년 서울대학교 전자공학과  
 1977년 독일 RWTH Aachen  
 Dipl.-Ing. in communication  
 engineering  
 1982년 독일 RWTH Aachen Dr.-Ing.  
 degree in acoustic  
 1983년 - 현재 서울대학교 전기공학부 교수