

머리 전달 함수장 재현을 통한 광대역 입체 음향 구현

HRTF-field reproduction for robust virtual source imaging

최정우†
Joung-Woo Choi

Key Words : Virtual acoustics(입체음향), Head-related transfer function(머리전달함수), HRTF-field(머리전달함수장)

ABSTRACT

A hybrid technique that combines the advantages of binaural reproduction and sound field reproduction technique is proposed. The concept of HRTF-field, which is defined as the set of HRTFs corresponding to the various head dislocations, enables us to realize virtual source imaging over a wide area. Conventional 2x2 definition is redefined as a MIMO system composed of multiple control sources and multiple head locations, and HRTF variations corresponding to various head movement are quantified. Through the direct control of HRTF-field, reproduction error induced by head dislocation can be minimized in least-square-error sense, and consequential disturbances on the virtual source image can be reduced within a selected area. Simple lateralization examples are investigated, and the reproduction error of the proposed technique is compared to that of Higher-order Ambisonics.

1. 서론

청취자 위치에서 가상 음원의 소리를 재현하기 위한 입체 음향 연구는 크게 두 가지 관점에서 시도되어 왔다. 그 첫번째는 음장(sound field)을 직접 제어하는 방법으로서, 기본적으로 다수의 라우드스피커의 간섭(interference)현상을 이용하여 유한한 넓이의 영역에서 원하는 음장을 복제하고자 하는 것이다. 보다 구체적으로는 단순 역문제를 푸는 방법[7]-[9]과 기저 함수로 전개한 후 재현하는 Ambisonics 기반 방법[10-12] 혹은 적분 방정식에 기초를 둔 방법[13-15] 등이 이에 속한다. 다음으로, 머리전달함수(HRTF)에 기반한 바이노럴(binaural) 재현법[1-6]을 생각할 수 있는데, 이는 인체에 의해 산란된 후 최종적으로 양 귀에 입사하는 음파, 즉 머리전달함수를 재현하는 것이다.

두 방법은 상이한 장단점을 가지는데, 음장을 직접 재현하는 방식의 경우, 상대적으로 많은 수의 라우드스피커를 필요로 하며, 고주파 영역에서 유한한 스피커 개수로 인한 음상의 깨어짐 혹은 공간에일리어싱 현상이 필연적으로 발생하게 되는 단점이 있다. 반면, 머리전달함수를 이용하는 경우에는 두 귀 위치의 신호만을 제어하므로 비교적 작은 수의 스피커로도 고주파 영역까지 재현이 가능하다. 하지만, 고정된 청취자의 두 귀 위치만을 고려하는 머리전달함수 재현법은 청취자의 위치 이동에 영향 받기 쉬운 단점을 안고 있다.

본 논문에서는 보다 넓은 공간상의 영역에 걸쳐 입체 음향을 구현하기 위한 머리전달함수장(HRTF-field)의 재현 방법을 제안한다. 이는 선택한 공간 내부에서 자유로이 움직이는 청취자의 각 위치에 따라 서로 다르게 발생하는 머리전달함수 전체를 고려하여 재현하는 것으로, 두 귀 위치의 신호만을 제어하는 바이노럴 재현 방법과 공간상의 선택 영역에 걸쳐 음장을 재현하는 음장 재현법(sound field reproduction)을 복합한 것이라 할 수 있다.

본 연구의 핵심은, 머리전달함수를 사용하는 기존 바이노럴 재현 기법에 기반을 두되, 선택한 청취 공간 내에서 발생할 수 있는 청취자의 이동을 고려하는 것이다. 청취자 위치 이동에 따라 필연적으로 발생하게 되는 머리전달함수 변화를 예측하고, 이러한 변화를 다출력 시스템으로 파악하여, 다수의 라우드스피커, 즉 다수의 입력을 사용해 제어를 수행한다. 다음은 이러한 다입출력 시스템을 수학적으로 표현하는 방법에 대하여 설명한다.

2. 문제 정의

2.1 머리전달함수장(HRTF-field)

그림 1에 도시된 바와 같이, \vec{r}_s 에 위치하는 음원이 선택한 공간 A_h 내의 청취자의 양쪽 귀로 소리를 전달하는 경우를 고려하여 보자. 만일 음원이 입력 신호 s 로 가진된다면 \vec{r}_{sr} , \vec{r}_{sl} 위치의 청취자의 양쪽 귀에서 측정되는 음압 p 는 다음과 같이

† 삼성종합기술원, Computing & Intelligence Laboratory
E-mail : jw010.choi@samsung.com
Tel : (031) 280-1734, Fax : (031) 280-9257

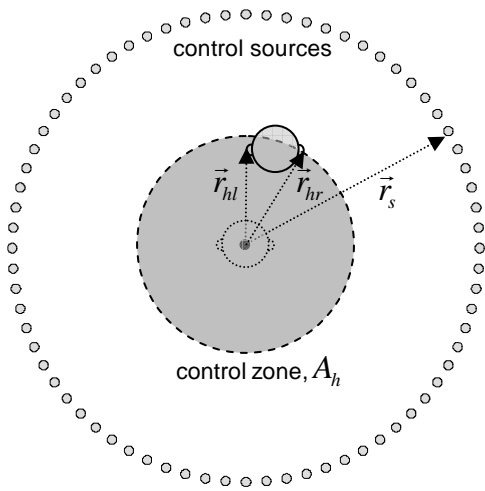


Figure 1 Overview of HRTF reproduction system

기술할 수 있다.

$$\begin{bmatrix} p(\vec{r}_{hr}, \omega) \\ p(\vec{r}_{hl}, \omega) \end{bmatrix} = \begin{bmatrix} h(\vec{r}_{hr} | \vec{r}_s, \omega) \\ h(\vec{r}_{hl} | \vec{r}_s, \omega) \end{bmatrix} s(\vec{r}_s, \omega), \quad (1)$$

여기서, h 는 음원과 청취자가 수음하는 음압 사이의 전달함수를 나타내며, 이는 머리전달함수라는 이름으로 잘 알려져 있다. 음압 p 는 항상 양쪽 귀의 위치에서 관측되므로, 이를 묶어 하나의 그룹으로 표현하자.

$$\mathbf{p}(\vec{r}_h) = \mathbf{h}(\vec{r}_h | \vec{r}_s) s(\vec{r}_s) \quad (2)$$

\vec{r}_h 는 양쪽 귀(\vec{r}_{hr} , \vec{r}_{hl})의 두 음압을 표현하는 기호이며, 볼드체로 표시된 변수는 벡터임을 나타낸다. 편의상 주파수 표기 ω 는 생략되었다.

(1) Binaural Reproduction

기존의 바이노럴(2×2) 시스템을 앞서 정의한 표기에 따라 표기하면 다음과 같다.

$$\begin{bmatrix} p(\vec{r}_{hr}) \\ p(\vec{r}_{hl}) \end{bmatrix} = \begin{bmatrix} h(\vec{r}_{hr} | \vec{r}_{sr}) & h(\vec{r}_{hr} | \vec{r}_{sl}) \\ h(\vec{r}_{hl} | \vec{r}_{sr}) & h(\vec{r}_{hl} | \vec{r}_{sl}) \end{bmatrix} \begin{bmatrix} s(\vec{r}_{sr}) \\ s(\vec{r}_{sl}) \end{bmatrix} \quad (3)$$

이 경우, 두 제어 음원이 중앙면을 중심으로 대칭으로 위치하므로, 두 귀 위치 표현과 같이 두 음원의 위치 \vec{r}_{sr} 과 \vec{r}_{sl} 을 기호 \vec{r}_s 를 사용하여 다음과 같이 그룹 형태로 표현한다.

$$\mathbf{p}(\vec{r}_h) = \mathbf{H}(\vec{r}_h | \vec{r}_s) \mathbf{s}(\vec{r}_s), \quad (4)$$

여기서, \mathbf{H} 는 두 입력과 출력 사이의 전달함수를 표현하는 2×2 행렬이다. 바이노럴 재생 시스템

의 목적은 재현하고자 하는 음압 신호 벡터를 청취자의 두 귀 위치에서 재현하는 것이며, 이는 달리 말하면 다음과 같이 재현하고자 하는 신호와 실제 발생된 신호의 오차

$$e^2 = \|\mathbf{p}(\vec{r}_h) - \mathbf{p}_d(\vec{r}_h)\|^2 \quad (5)$$

를 최소화 하는 입력 신호 $\mathbf{s}(\vec{r}_s)$ 를 찾는 것으로 정리될 수 있다. 여기서, $\|\cdot\|$ 는 벡터의 2-norm을 표현한다. 식 (5)에서 주어지는 오차는 벡터 요소 제곱의 합과 같으므로, 이를 최소화 하는 해를 통상적으로 최소 자승 오차(minimum-square-error; MSE)라 하며, 수학적으로 다음과 같이 주어진다.

$$\mathbf{s}_{MSE}(\vec{r}_s) = \mathbf{H}(\vec{r}_h | \vec{r}_s)^+ \mathbf{p}_d(\vec{r}_h) \quad (6)$$

여기서, 위 첨자 $+$ 는 의사역행렬(pseudo-inverse)을 나타낸다.

바이노럴 문제의 경우, 재현하고자 하는 신호 \mathbf{p}_d 와 동일한 개수의 제어 음원을 사용하므로 특이점을 제외하면 오차 없이 재현이 가능하다. 하지만, 전달함수행렬 \mathbf{H} 가 특이점(singularity)을 가질 때는 시스템이 불안정하여 최적해를 역추산하는 과정에서 많은 오차가 유발될 수 있으며, 청취자의 움직임에 매우 민감한 특성을 지닌다. 따라서 바이노럴 시스템을 설계하는데 안정성(stability)은 매우 중요하게 간주되며, 안정된 시스템을 구현하기 위해 많은 노력이 기울여져 왔다. 대표적인 방법이 청취자의 움직임에 의한 출력 변화를 최소화하는 음원 배치를 찾는 것으로, Takeuchi[6] 등은 condition number를 지표로 하는 6개 음원 최적배치 방법을 제안한 바 있다.

언급된 안정화 방법은 청취자의 미세한 움직임에 대해서는 유용하나, 다음과 같은 근본적인 문제를 안고 있다. 그림 2에 도시된 바와 같이, 청취자의 위치가 원래의 위치 A에서 A'으로 이동한

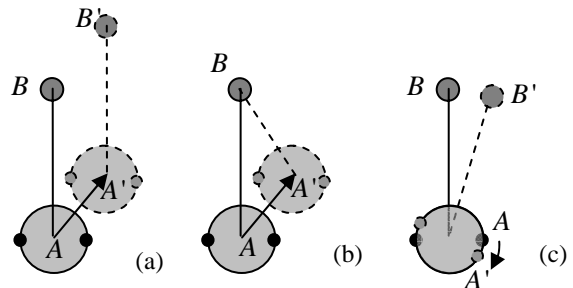


Figure 2 Dislocation of virtual source's image due to head movement

경우를 고려하자. 이 때 청취자가 이동하였으므로, 청취자가 느끼는 상대적인 음원의 위치는 기하학적으로 이동함이 옳지만, 기존의 안정성(stability)에 기반한 제어 방법은 청취자가 움직이더라도 두 귀 위치의 신호가 변하지 않으므로 결과적으로 청취자는 자신의 이동과 함께 가상 음원의 위치가 이동(B')하는 것으로 인식하게 된다. 이 현상은 청취자의 머리 회전(그림 2(c))에 대해 특히 문제가 되는데, 인간은 청취중에 동적으로 머리를 움직여 그 변화로부터 정확한 음원의 위치를 찾아내기 때문이다[16,17]. 머리 회전의 경우, 수평 이동에 비하여 작은 움직임만으로도 청취 신호의 큰 변화를 유발하므로, 기존의 안정성 제어 방법으로는 이러한 동적 인지과정에 부합할 수 없게 된다.

따라서, 청취자의 움직임에 따라 새로운 가상 음원 신호를 제공할 수 있는 재현 방법이 구현되어야 한다. 다음 절에서는 청취자의 이동을 고려한 새로운 제어 기법에 대해 논한다.

(2) 청취자 이동의 표현

청취자의 이동을 수학적으로 표현하여 보자. 먼저, 청취자가 위치할 수 있는 공간상의 영역을 A_h 로 정의한다. 이 공간 내에서 청취자는 자유롭게 움직일 수 있으며, 우리는 이 공간내의 임의의 위치에 존재하는 청취자가 가상 음원의 머리전달함수를 전달 받을 수 있도록 제어하여야 한다. 따라서, 공간 A_h 내의 임의의 위치 $\tilde{\mathbf{r}}_h$ 에서 식 (2)가 성립하여야 함은 명백하다. 이를 위해, 공간을 이산화하여 다수의 청취자 위치 $([\tilde{\mathbf{r}}_h^{(1)}, \dots, \tilde{\mathbf{r}}_h^{(M)}])$ 를 고려하자. 여기서 다수의 청취자 위치라 함은, 청취자의 머리중심위치 뿐 아니라, 동일한 중심위치에서의 제자리 회전도 포함하는 것이다. 다수의 위치에 대하여 식 (2)를 재기술하면 다음과 같다.

$$\begin{bmatrix} p(\tilde{\mathbf{r}}_{hr}^{(1)}) \\ p(\tilde{\mathbf{r}}_{hl}^{(1)}) \\ \vdots \\ p(\tilde{\mathbf{r}}_{hr}^{(M)}) \\ p(\tilde{\mathbf{r}}_{hl}^{(M)}) \end{bmatrix} = \begin{bmatrix} h(\tilde{\mathbf{r}}_{hr}^{(1)} | \tilde{\mathbf{r}}_{sr}) \\ h(\tilde{\mathbf{r}}_{hl}^{(1)} | \tilde{\mathbf{r}}_{sr}) \\ \vdots \\ h(\tilde{\mathbf{r}}_{hr}^{(M)} | \tilde{\mathbf{r}}_{sr}) \\ h(\tilde{\mathbf{r}}_{hl}^{(M)} | \tilde{\mathbf{r}}_{sr}) \end{bmatrix} \mathbf{s}(\tilde{\mathbf{r}}_s), \quad (7)$$

이는 앞서의 행렬 표현을 사용하여 다음과 같이 간략화 될 수 있다.

$$[\mathbf{p}(\tilde{\mathbf{r}}_h^{(1)}) \cdots \mathbf{p}(\tilde{\mathbf{r}}_h^{(M)})]^T =$$

$$[\mathbf{h}(\tilde{\mathbf{r}}_h^{(1)} | \tilde{\mathbf{r}}_s) \cdots \mathbf{h}(\tilde{\mathbf{r}}_h^{(M)} | \tilde{\mathbf{r}}_s)]^T \mathbf{s}(\tilde{\mathbf{r}}_s). \quad (8)$$

이를 보다 더 단순히 표기하기 위하여 다음과 같은 표현법을 정의하자.

$$\mathbf{p}(\tilde{\mathbf{r}}_h) = \mathbf{h}(\tilde{\mathbf{r}}_h | \tilde{\mathbf{r}}_s) \mathbf{s}(\tilde{\mathbf{r}}_s) \quad (9)$$

괄호안에 볼드체로 표기된 변수 $\tilde{\mathbf{r}}_h$ 는, 행렬 $\mathbf{p}(\tilde{\mathbf{r}}_h)$ 가 여러 청취자 위치 $\tilde{\mathbf{r}}_h = [\tilde{\mathbf{r}}_h^{(1)}, \dots, \tilde{\mathbf{r}}_h^{(M)}]^T$ 에서 측정되는 음압 행렬 $\mathbf{p}(\tilde{\mathbf{r}}_h^{(m)})$ 를 적층(partitioning)함으로써 정의됨을 의미한다.

결과적으로, 청취 공간 A_s 를 자유로이 움직일 수 있는 청취자가 $\tilde{\mathbf{r}}_s$ 위치의 음원을 지각하도록 하는 것은, 식 (9)의 청취 음압 $\mathbf{p}(\tilde{\mathbf{r}}_h)$ 가 $\tilde{\mathbf{r}}_s$ 위치의 음원에 의한 전달함수 $\mathbf{h}(\tilde{\mathbf{r}}_h | \tilde{\mathbf{r}}_s)$ 와 동등하게 되는 것을 의미한다. 하지만, 이 경우 매우 많은 측정점을 제어하여야 하므로 우리는 다수의 음원을 제어하여야 할 필요가 있다. 만일 다수의 음원이 식 (4)에서 정의된 바와 같이 $(\tilde{\mathbf{r}}_h^{(1)}, \dots, \tilde{\mathbf{r}}_h^{(M)})$ 위치에 위치한다면, 관계식은 다음과 같이 확장된다.

$$\begin{bmatrix} \mathbf{p}(\tilde{\mathbf{r}}_h^{(1)}) \\ \vdots \\ \mathbf{p}(\tilde{\mathbf{r}}_h^{(M)}) \end{bmatrix} = \begin{bmatrix} \mathbf{H}(\tilde{\mathbf{r}}_h^{(1)} | \tilde{\mathbf{r}}_s^{(1)}) & \cdots & \mathbf{H}(\tilde{\mathbf{r}}_h^{(1)} | \tilde{\mathbf{r}}_s^{(N)}) \\ \vdots & \ddots & \vdots \\ \mathbf{H}(\tilde{\mathbf{r}}_h^{(M)} | \tilde{\mathbf{r}}_s^{(1)}) & \cdots & \mathbf{H}(\tilde{\mathbf{r}}_h^{(M)} | \tilde{\mathbf{r}}_s^{(N)}) \end{bmatrix} \times \begin{bmatrix} \mathbf{s}(\tilde{\mathbf{r}}_s^{(1)}) \\ \vdots \\ \mathbf{s}(\tilde{\mathbf{r}}_s^{(N)}) \end{bmatrix} \quad (10)$$

주목할 것은, 이제 더 이상 관계식이 바이노럴 관계로 주어지지 않고, 다수의 입력과 출력을 갖는 MIMO(multiple-input-multiple-output; MIMO) 시스템으로 정의된다는 점이다. 식 (9)의 행렬 표기를 사용하면 다음과 같이 단순화된 표기를 얻을 수 있다.

$$\mathbf{p}(\tilde{\mathbf{r}}_h) = \mathbf{H}(\tilde{\mathbf{r}}_h | \tilde{\mathbf{r}}_s) \mathbf{s}(\tilde{\mathbf{r}}_s) \quad (11)$$

이는 단순히 바이노럴 시스템을 다수의 머리 위치와 음원에 대하여 확장한 것이지만, 다수의 음원을 최적화하여 입체 음향 구현이 가능한 공간의 크기를 확장하는 기초가 된다. 또한, 음장을 제어하기 보다는, 산란된 결과물인 머리 전달함수를 직접 제어한다는 점에서 일반적인 음장 재현법과도 차별화 된다.

3. HRTF-field 의 재현

3.1 최소 자승 오차해

앞서 정의된 확장된 표현을 사용하여 다수의 음원을 최적화 하는 방법을 제안한다. 구체적으로, \vec{r}_h 에 위치한 다수의 제어 음원을 사용하여 \vec{r}_v 위치의 가상 음원을 재현하는 문제를 생각해보자. 이 경우, 청취자 위치에서 재현할 음장은 \vec{r}_v 위치에서 \vec{r}_h 까지의 머리 전달함수가 된다. 즉,

$$\mathbf{p}_d(\vec{r}_h) = \mathbf{h}(\vec{r}_h | \vec{r}_v) \quad (12)$$

로 쓸 수 있다. 이 음장이 \vec{r}_s 위치의 제어 음원에 의하여 재현되어야 하므로, 식 (11)로 주어지는 재현 음장이 $\mathbf{p}_d(\vec{r}_h)$ 와 가능한 유사하게 되어야 한다. 이 역시 원하는 음장과 실제 재현된 음장 사이의 오차를 최소화하는 최소자승오차 문제로 생각할 수 있으며, 그 오차는 다음과 같이 기술될 수 있다.

$$e = \|\mathbf{p}_d(\vec{r}_h) - \mathbf{H}(\vec{r}_h | \vec{r}_s) \mathbf{s}(\vec{r}_s)\|. \quad (13)$$

실제로는 청취자의 위치가 제어 음원의 개수보다 많은 것으로 간주하는 것이 일반적이므로 ($M > N$), 정의된 문제는 over-determined 문제로 볼 수 있다. 이와 같은 문제에 대하여, 최소자승 오차해는 다음과 같은 관계식으로 주어진다.

$$\mathbf{s}(\vec{r}_s) = \left\{ \mathbf{H}(\vec{r}_h | \vec{r}_s)^H \mathbf{H}(\vec{r}_h | \vec{r}_s) \right\}^{-1} \mathbf{H}(\vec{r}_h | \vec{r}_s)^H \mathbf{p}_d(\vec{r}_h) \quad (14)$$

식 (12)에 따르면, 원하는 음장은 머리 전달함수의 형태로 주어지므로, 식 (14)를 다음과 같이 정리할 수 있다.

$$\mathbf{s}(\vec{r}_s) = \mathbf{H}(\vec{r}_h | \vec{r}_s)^+ \mathbf{h}(\vec{r}_h | \vec{r}_v). \quad (15)$$

마찬가지 논리로, 다수의 가상 음원에 대한 재현 오차를 표현할 수 있다. 가상 음원의 위치들을 그룹 행렬 \vec{r}_v 로 표현하면, 각각의 가상 음원 위치에 대한 최소자승오차해들을 다음과 같이 행렬 형태로 정리할 수 있다.

$$\mathbf{S}(\vec{r}_s) = \mathbf{H}(\vec{r}_h | \vec{r}_s)^+ \mathbf{H}(\vec{r}_h | \vec{r}_v) \quad (16)$$

위에서 주어진 최적해 행렬 ($\mathbf{S}(\vec{r}_s)$)의 각 열은 서로 다른 가상 음원의 위치에 대응되는 해를 나타

낸다. 위 최적해는 아래의 재현 오차를 제어함으로써 얻어지는 해임은 자명하다.

$$e = \|\mathbf{E}\| = \|\mathbf{H}(\vec{r}_h | \vec{r}_v) - \mathbf{H}(\vec{r}_h | \vec{r}_s) \mathbf{S}(\vec{r}_s)\| \quad (17)$$

오차 행렬 \mathbf{E} 의 행과 열 방향은 각각 청취자의 위치와 가상 음원의 위치에 대한 오차를 나타낸다. 위의 기술은 모두 주파수 영역에서 표현된 것이므로, 오차 행렬은 주파수의 함수이며, 최적해 역시 주파수에 따라 독립적으로 주어진다.

오차 행렬 \mathbf{E} 를 도입함으로써 공간 위치별, 가상 음원별로 재현 오차를 정량화 할 수 있으며, 또한 이 오차에 다양한 가중치를 부여하여 보다 섬세한 제어가 가능하게 된다. 오차 행렬에 부여되는 가중치는 다음 절의 예제를 통하여 보다 상세하게 다루어 질 것이다.

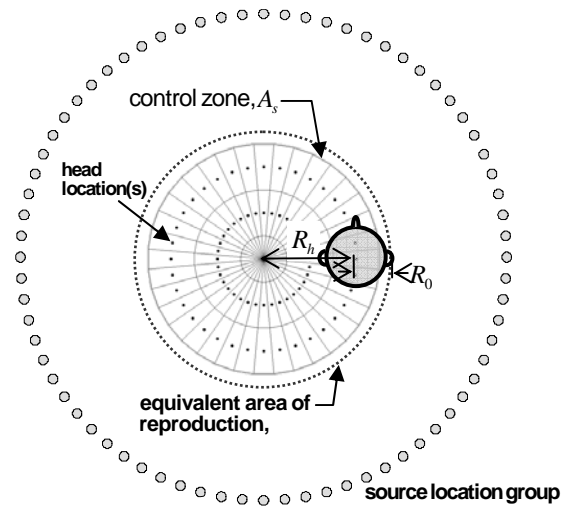


Figure 3 Experimental configuration (R_h : distance of the head from origin, R_0 : radius of a head, A_s : control zone)

4. 제어 예제

제안된 방법의 효용성을 검증하기 위하여 다음과 같은 모의 실험을 수행하였다. 실험에 사용된 머리 전달함수는 KEMAR DB 66 더미헤드를 사용하여 측정된 데이터 베이스를 활용하였으며, 외이를 장착한 상태에서 2m 거리에 위치한 음원을 사용하여 전달함수를 측정하는 것이다. 데이터베이스는 수평방향으로 0° 에서 35° 까지 총 72 각도에서 측정되었으며, 44.1kHz 로 샘플링 된 것이다.

4.1 실험 설정

간단한 예로, 그림 3 에 도시된 바와 같이 청취자가 위치하는 공간을 제어 공간(평면) A_s 로, 그리고 가상 음원과 제어 음원이 그 평면과 동일한 높이에 원주 형태로 분포하는 수평 방향 재현 문제를 설정하였다.

다양한 제어 음원의 수 (N_s)에 따른 재현 오차를 살펴보았으며, 이 때 제어 음원의 위치(\vec{r}_s)는 HRTF 가 측정된 72 개 위치 중에서 등간격으로 선택된다. 반면에, 재현하고자 하는 가상 음원의 위치(\vec{r}_v)는 가능한 많은 위치를 고려하여야 하므로 측정된 72 개 모두를 고려하였다.

청취자의 다양한 위치 변화(\vec{r}_h)에 따라 측정되지 않은 HRTF 정보도 요구되므로, 측정된 HRTF 데이터 베이스를 사용하여 시간 영역(time domain)에서 내삽(interpolation)을 수행하였다. HRTF 는 청취자와 음원의 상대적인 위치 관계에 따라 결정되므로, 먼저 각각의 청취자 위치에서 음원과의 상대 위치를 계산하고, 이를 이용하여 HRTF 예측을 수행하였다.

정밀한 HRTF 예측을 위한 여러가지 기법들 [18][19]이 존재하지만, HRTF 예측기법 자체는 본 연구의 주요 관심사가 아니므로, 본 예제에서는 음원이 단극 음원으로 주어지고, 제어 공간으로부터 원거리에 위치한 간단한 경우를 고려한다. 이와 같은 경우에 원거리 음장 가정(Far-field approximation)을 적용하여 HRTF 예측을 수행할

수 있는데, 이는 음원과의 상대거리 및 각도 (azimuth)에 의해 HRTF 가 결정된다는 가정이다.

원거리 음장의 가정에 따라, 청취자의 다양한 위치에 대하여 음원과 청취자의 상대 수평각 (azimuth)을 추정하고, 추정된 각도에 따라 HRTF 를 내삽한 후, 상대 거리 차이에 따른 시간 지연 (time delay) 및 감쇠(attenuation)치를 보정하였다.

그림 4 는 HRTF 내삽에 따른 오차의 크기를 살펴본 것으로서, 측정된 HRTF 데이터베이스 중에서 짝수번째 방위각에 해당하는 HRTF 만으로 홀수번째 방위각의 HRTF 를 예측하고 실측치와 비교한 것이다. 도시된 상대 오차값은 대부분 20dB 이하의 크기를 보이며, 이보다 큰 상대 오차를 나타내는 영역 또한 HRTF 의 절대 크기가 무시될 수 있을 정도로 작은 구간에서 발생하는 것이다.

4.2 가중치 부여: 공간 윈도우(spatial window)

이와 같이 추산된 HRTF 를 사용하여 오차행렬 (식 (17))을 계산하고, 이를 최소화하는 최적해로 제어를 수행하였다. 하지만, 여기서 유한한 음원 개수로 인해 발생하는 물리적 한계인 공간 에일리어싱(spatial aliasing) 현상으로 인해, 재현을 수행할 수 있는 공간의 크기에 한계가 있으므로, 이 한계점 내에서 최적의 제어를 수행하는 것이 중요하다. 즉, 제어를 수행할 수 없는 크기의 청취 공간에 대한 최적해를 찾는 것은 무의미하므로, 물리적인 한계 내에서 적절한 최적해를 탐색하도록 제어 공간을 제한할 필요가 있다.

이를 위해 공간 윈도우(spatial window)를 오차행렬에 적용하였다. 공간 윈도우는 HRTF 재현을 수행할 때 발생하는 오차 중에서, 공간의 외곽쪽에서 발생하는 에러가 공간의 중심에서 발생하는 에러보다 낮은 비중을 갖도록 가중치를 부여하는 역할을 수행한다. 적용된 후에는 공간의 외곽 지점을 제어에 참여시키지 않음으로써, 물리적 한계 이상의 영역은 제어에 고려되지 않는다.

본 예제에서는 다음과 같이 N_s 개의 음원으로 제어를 수행하였을 때, 공간 에일리어싱이 발생하는 지점[11]

$$R_r = (N_s - 1) / 2k \quad (18)$$

에서 -20dB 의 크기를 갖는 exponential window 를 공간 윈도우로 사용하였다. 여기서, k 는 파수, R_r 은 에일리어싱이 발생하지 않는 영역의 반지름을 나타낸다.

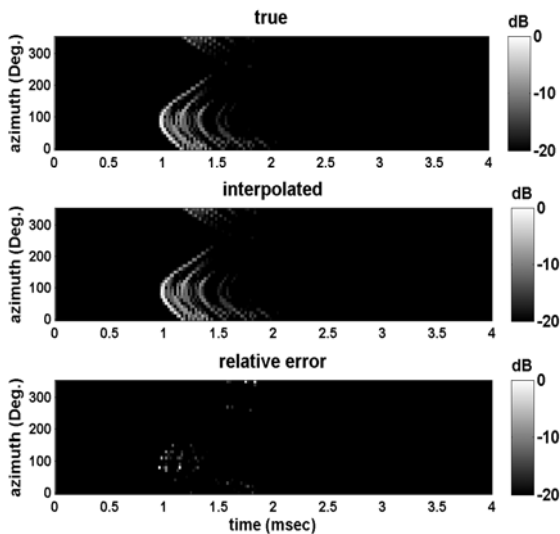


Figure 4 Recreation performance of HRTF interpolation algorithm (From the top: true(measured), interpolated (reconstructed), relative error of the HRTFs)

$$w(kR_h^{(m)}, N_s) = \exp\left(-2.3 \frac{k(R_h^{(m)} + R_0)}{(N_s - 1)/2}\right) \quad (19)$$

위 윈도우는, 중심으로부터의 청취자 머리중심 위치($R_h^{(m)} = |\vec{r}_h^{(m)}|$)에 따라 지수적으로 감소하며, 청취자의 바깥쪽 귀 위치가 반지름 R_r 과 일치할 때 ($R_r = R_h^{(m)} + R_0$, R_0 : 머리 반지름) -20dB 의 감쇠를 갖게 된다. 윈도우 함수는 다음과 같이 행렬 형태($\mathbf{W}_{N_s}(\vec{r}_h) \in \mathbb{R}^{M \times M}$)로 기술할 수 있으며,

$$\begin{aligned} [\mathbf{W}_{N_s}(\vec{r}_h)]_{(n,m)} &= w(kR_h^{(m)}, N_s) \quad \text{if } n = m, \\ &= 0 \quad \text{otherwise.} \end{aligned} \quad (20)$$

다음과 같이 윈도우 함수가 적용된 오차를 최소화 하는 최적해를 찾게 된다.

$$\begin{aligned} e_w &\equiv \|\mathbf{W}_{N_s}(\vec{r}_h) \mathbf{E}(\vec{r}_h | \vec{r}_v)\| \\ &= \|\mathbf{W}_{N_s}(\vec{r}_h) [\mathbf{H}(\vec{r}_h | \vec{r}_v) - \mathbf{H}(\vec{r}_h | \vec{r}_s) \mathbf{S}(\vec{r}_s | \vec{r}_v)]\|, \end{aligned} \quad (21)$$

그림 5 는 제안된 최적 제어 기법의 제어 성능을 상대 오차량을 통해 도시한 것이다. 청취자 머리의 이동(R_h)은 7cm 까지 관찰하였으며, 이 경우 머리 반지름을 8cm 로 가정하면 약 15cm 의 반지름을 갖는 공간을 제어하는 것으로 생각할 수 있다. 기존 방법과의 성능 비교를 위해, 음장재현 기법 중 HOA(Higher-order Ambisonics) 기법의 일종 [12]으로 제어한 경우의 상대 오차량과 함께 도시하였다. 상대 오차량은 각 주파수(x 축) 및 공간상의 청취자 위치(y 축)에 대하여 도시되었다.

그래프 상에서 어둡게 표현되는 영역이 보다 낮은 재현 오차를 갖는 영역으로, 최압부의 경우 -20dB이하의 오차로 재현이 이루어진 영역이다. 가는 점선은 이론적으로 계산되는 공간 에일리어싱에 의한 한계점을 나타내며, 좌측 그래프의 Ambisonics 결과에서 볼 수 있듯이 음장 재현 기법이 정상적으로 작동할 수 있는 한계와 일치한다. 우측의 그래프에 도시된 굵은 점선은, 적용된 공간 윈도우의 크기(magnitude)가 -20dB이하로 감소되는 영역으로, 공간 에일리어싱 커브와 일치하게 설계되었다.

상단에서부터 순서대로 2, 4, 6, 8, 12 개의 제어 음원(N_s)을 구동하는 경우의 최적화 결과를 도시하였다. 제어 음원의 개수에 관계없이 공통적으로

관측되는 현상은, 청취자의 위치가 원점 근처일 때 매우 높은 주파수 대역까지 재현이 가능하다는 점이다. 이는 HRTF 를 직접 재현함으로써 얻어지는 장점으로서, Ambisonics 와 같은 음장 재현 기법은 고주파수 대역에서 머리 보다 작은 크기의 영역에서만 음장 재현이 가능하기 때문에, 비록 청취자가 원점에 정확히 위치하더라도 재현 가능한 주파수 대역이 제한된다.

청취자의 위치가 원점에서 멀어짐에 따라서 재현 가능한 주파수 대역은 서서히 감소하게 되지만, 기존의 음장 재현 기법보다 넓은 주파수 대역에서 제어가 가능함을 볼 수 있다. 청취자의 머리 이동이 과도하게 발생함에 따라 제안된 기법은 Ambisonics 의 성능으로 수렴하게 된다.

최우측 그래프의 경우, 청취자의 머리 이동외에 회전이 발생하는 경우를 고려하여 최적화 한 결과를 나타낸 것이다. 청취자 머리의 작은 회전 ($-15^\circ \sim 15^\circ$; 5° 간격)을 함께 고려하여 \vec{r}_h 를 설정하고 최소자승 오차해를 계산하였다. 결과적으로 머리 회전이 존재하지 않는 경우에 비해 더 많은 오차가 발생하지만, 여전히 Ambisonics 기법에 비해 중심점 근처에서 보다 고주파수 대역까지 낮은 오차로 재현 가능함을 보이고 있으며, 과도한 머리 이동시에 수렴하는 동일한 양상을 나타내고 있다.

결과적으로, 제안된 방법은 청취자의 머리 이동에 따른 재현 오차를 정량화 함으로써, 같은 거리의 머리 이동에 대하여 보다 높은 주파수 대역까지 입체 음향 재현이 가능함을 보여주고 있으며, 향후 공간 윈도우의 설계 개선 및 제어 음원의 배치 방법 변화 등에 따라 보다 향상된 성능을 보여 줄 수 있을 것으로 기대된다.

5. 결론

보다 넓은 공간상의 영역에 걸쳐 입체 음향을 구현하기 위한 머리전달함수장(HRTF-field)의 재현 방법을 제안하였다. 입체 음향 문제를 다수의 스피커 입력과 다수의 청취자 위치에 따른 머리전달함수 출력으로 정의하고, 이들이 서로 상관관계를 이루는 다입력-출력(Multiple-Input-Multiple-Output; MIMO) 시스템으로 파악하여, 선택한 공간 내부의 어느 지점에 청취자가 위치하더라도 최소의 출력 오차로 청취자에서 입체 음향을 전달할 수 있는 다수의 스피커 입력을 산출하는 방법을 탐색하였다. 이를 위해 다출력계의 역문제 계산 방법(MIMO Inverse theorem; MINT)을 적용하여 최적의

스피커 어레이 제어 방법을 도출하는 방법을 개발하였다.

제안된 방법은 고정된 청취자의 위치, 혹은 매우 미소한 이동만을 허용하는 기존의 머리전달함수(HRTF) 기반 바이노럴 재현법에 비해 보다 넓은 공간에 걸쳐 입체 음향 구현이 가능하게 한다. 또한, 음장 제어 기법과의 직접적인 비교를 위해 살펴본 가상음원 수평각 제어 예제를 통하여, 제안된 방법이 음장 재현 방식의 하나인 HOA에 비해 양 귀에 입사되는 결과물인 머리전달함수를 제어하므로 보다 고주파 영역까지 제어가 가능함을 보였다

후 기

이 논문은 2006년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임을 밝힙니다. (KRF-2006-214-D00123)

참고문헌

- (1) M. R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," IEEE Conv. Rec., pt. 7, 150-155 (1963)
- (2) B. S. Atal and M. R. Schroeder, "Apparent sound source translator," U.S. Patent 3 236 949, Feb. (1966)
- (3) J. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," J. Audio Eng. Soc., 44, 683-705 (1996)
- (4) 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., 46, 387-395 (1998)
- (5) D. B. Ward and G. W. Elko, "Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation," IEEE Sig. Proc. Lett., 6(5) (1999)
- (6) T. Takeuchi, P. A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers," J. Acoust. Soc. Am., 112 (6), 2786-2797 (2002)
- (7) O. Kirkeby, P. A. Nelson, "Reproduction of plane wave sound fields," J. Acoust. Soc. Am., 94, 2992-3000 (1993)
- (8) O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante and H. Hamada, "Local sound field reproduction using digital signal processing," J. Acoust. Soc. Am., 100, 1584-1593 (1996)
- (9) O. Kirkeby, P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," J. Audio. Eng. Soc., 47, 583-595 (1999)
- (10) M. A. Poletti, "A unified theory of horizontal holographic sound systems," J. Audio. Eng. Soc., 48, 1155-1162 (2000)
- (11) D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," IEEE Transactions on speech and audio processing, 9(6), 697-707 (2001)
- (12) M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," J. Audio. Eng. Soc., 53(11), 1004-1025 (2005)
- (13) A. J. Berkout, D. D. Vries and P. Vogel, "Acoustic control by wave field synthesis," J. Acoust. Soc. Am., 93, 2764-

2778 (1993)

(14) S. Ise, "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems," Acoustica – acta acoustica, 85, 78-87 (1999)

(15) S. Takane, Y. Suzuki and T. Sone, "A new method for global sound field reproduction based on Kirchhoff's integral equation," 85, 250-257 (1999)

(16) H. Wallach, "The role of head movements and vestibular and visual cues in sound localization," J. Exp. Psychol. 27, 339-368 (1940)

(17) F. L. Wightman, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," J. Acoust. Soc. Am., 105(5), 2841-2853 (1999)

(18) M. J. Evans and J. A. S. Angus, "Analyzing head-related transfer function measurements using surface spherical harmonics," J. Acoust. Soc. Am., 104(4), 2400-2411 (1998)

(19) R. Duraiswaini, D. N. Zotkin, N. A. Gumerov, "Interpolation and range extrapolation of HRTFs," in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004 (ICASSP '04), 17-21 May 2004, 4, iv 45- iv 48 (2004) ISSN: 1520-6149

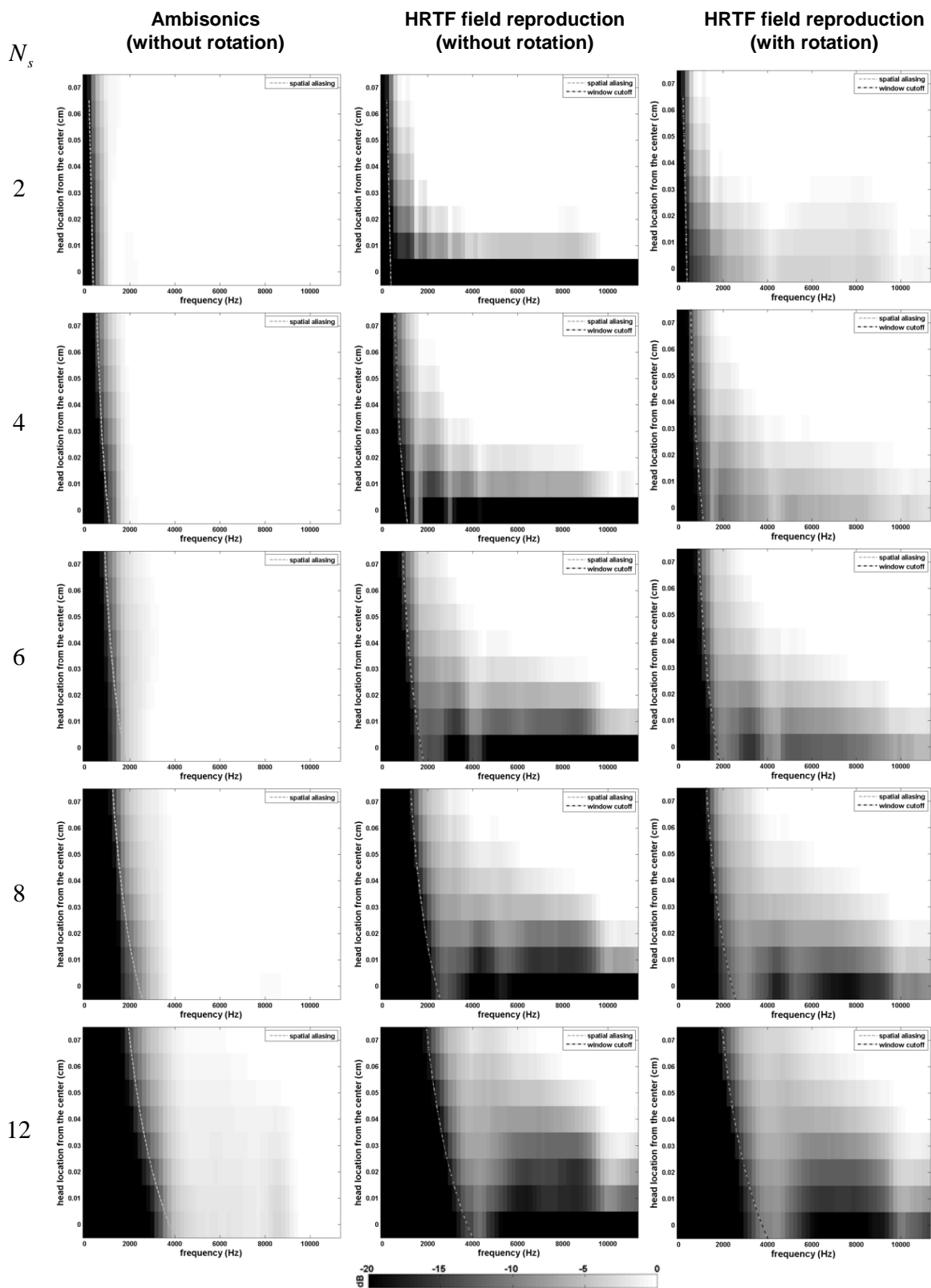


Figure 5 Reproduction error with respect to head location from the center and frequency ($R_h = 7\text{ cm}$) (from left to right: reproduced by Ambisonics mode-matching solution, HRTF-field reproduction with/without head rotation, from top to bottom: increasing number of sources [2,4,6,8,12])