

가상현실 환경에서의 3차원 사운드 생성을 위한 거리 변화에 따른 구조적 머리전달함수 모델*

이영한(GIST), 김홍국(GIST)

<차 례>

- | | |
|-------------------------------------|------------------------------------|
| 1. 서론 | 3.1. 레벨 표준화 |
| 2. 거리 변화에 따른 HRTF 특징 | 3.2. 원거리 HRTF 모델 |
| 2.1. 원거리 HRTF | 3.3. 근거리 HRTF 모델 |
| 2.2. 근거리 HRTF | 4. 제안된 HRTF의 가상현실에서의 구현
및 성능 평가 |
| 3. 제안된 거리 변화에 따른 구조적 머리
전달 함수 모델 | 5. 결론 |

<Abstract>

A Range Dependent Structural HRTF Model for 3-D Sound Generation in Virtual Environments

Young Han Lee, Hong Kook Kim

This paper proposes a new structural head-related transfer function(HRTF) model to produce sounds in a virtual environment. The proposed HRTF model generates 3-D sounds by using a head model, a pinna model and the proposed distance model for azimuth, elevation, and distance that are three aspects for 3-D sounds, respectively. In particular, the proposed distance model consists of level normalization block, distal region model, and proximal region model. To evaluate the performance of the proposed model, we setup an experimental procedure that each listener identifies a distance of 3-D sound sources that are generated by the proposed method with a predefined distance. It is shown from the tests that the proposed model provides an average distance error of 0.13 ~ 0.31 meter when the sound source is generated as if it is 0.5 meter ~ 2 meters apart from the listeners. This result is comparable to the average distance error of the human listening for the actual sound source.

* Keywords: 3D sound, HRTF, Range dependence, Structural model, Virtual reality

* 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업(IITA-2006-C1090-0603-0017)과 광주과학기술원 실감콘텐츠 연구센터(ICRC)를 통한 과학기술부 특정연구개발사업의 지원으로 수행되었음.

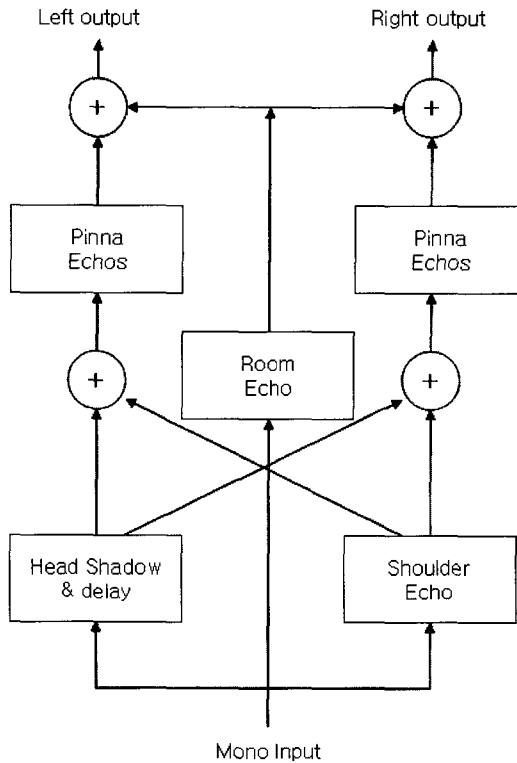
1. 서 론

가상현실 연구는 지난 수십 년 동안 많은 발전을 이루어 왔다. 가상현실 실현을 위한 중요한 구성요소로서의 3차원 사운드는 영상과 결합할 경우 가상현실의 몰입도를 높여주는 역할을 한다[1]. 다시 말해서 가상현실 구현의 완성도를 높일 수 있다는 측면에서 가상현실에서의 3차원 사운드는 매우 중요한 요소로 자리매김을 하고 있다.

그 동안 3차원 사운드에 관한 연구는 주로 음원(sound source)의 위치와 청취자의 귀 사이의 관계를 찾는 데 집중되어 왔다[2]. 일반적으로 이러한 음원과 청취자 사이의 관계를 머리전달함수(head-related transfer function, HRTF)라 부르고 이것을 통해 3차원 사운드를 생성한다.

인간은 양 귀의 고막에서 들리는 음원크기의 차이인 interaural level difference (ILD)와 양 귀의 고막에 도달하는 시간 차이인 interaural time difference(ITD)로 음원의 위치를 인지하는 것으로 알려져 있다[2]. 음원의 방위각(azimuth)과 고도(elevation), 거리(distance)에 따라 ITD와 ILD는 변하게 되고, 이 변화를 이용하여 인간은 후천적 학습으로 음원의 위치를 인지하게 된다. 하지만 ITD나 ILD는 음원의 위치뿐만 아니라 음원의 주파수 성분에 따라 달라지며, 청취자의 머리의 크기, 형태, 귀의 위치, 귓바퀴의 모양 등 개인의 신체조건에 민감하게 변화한다. 신체조건에 따라 HRTF가 달라지기 때문에 많은 연구자들은 통계적인 방법을 통해 인간의 표준 머리모델을 만들고 이것을 이용하여 방위각, 고도, 거리에 따른 HRTF를 측정해 왔다. 하지만 이와 같이 측정된 HRTF를 이용한 3차원 사운드의 구현은 모든 각도 및 모든 거리에 대한 HRTF가 필요하다는 단점을 갖는다. 또한 위에서 언급한 것처럼 개인의 신체조건에 민감하기 때문에 표준 머리모델을 대상으로 측정한 HRTF를 이용하여 3차원 사운드를 생성할 경우, 그 효과가 감소될 수밖에 없다. 따라서 효과를 높이기 위해서는 개인에 적합한 HRTF를 각각 측정해야 하지만 측정비용 및 안전 등의 어려움이 있다. 이러한 단점을 보완하기 위해 제안되어온 방법이 HRTF 모델링이다. 많은 연구자들은 측정된 HRTF를 근거로 이를 수학적으로 표현하는 HRTF 모델링 방법을 연구해왔다[3]-[5]. 이러한 모델링 방법은 방위각, 고도, 거리를 변수로 가지기 때문에 모든 경우에 대하여 HRTF를 측정하지 않아도 될 뿐만 아니라 3차원 사운드를 생성하는데 있어서 메모리를 효율적으로 사용할 수 있는 장점을 갖는다. 뿐만 아니라, 인간의 신체조건을 변수로 표현하여, 청취자에 따라 적합한 HRTF를 만들 수 있다.

HRTF를 모델링하는 방법 중에서 본 논문에서는 구조적 머리전달 함수 모델(structural HRTF model)을 기반으로 하여 거리 변화에 따른 모델을 제안한다. 구조적 HRTF 모델은 <그림 1>과 같이 머리에 의한 효과, 귀에 의한 효과 등으로 여러



<그림 1> 기존의 구조적 머리전달 함수 모델

개의 블록으로 조합으로 표현되며 이를 통해 3차원 사운드를 생성한다[3]. 먼저 방향성이 없는 입력 신호는 ‘Head Shadow & Delay’ 과정을 거치면서 방위각에 대한 HRTF가 적용된다. 그리고 나서, 고도를 느끼게 해주는 귓바퀴 효과(pinna echo) 및 청취환경(room echo), 흉부 효과(shoulder echo)를 거쳐 측정된 HRTF를 이용한 3차원 사운드와 유사한 효과를 얻게 된다. 이를 이용하면 방위각, 고도에 대한 HRTF를 구현할 수 있지만, 거리 변화에 대한 HRTF 변화는 고려하지 않았다.

가상현실 실현을 위한 3차원 사운드 생성에 있어서 방위각, 고도에 따른 음의 변화가 필요하지만, 거리에 따른 변화의 구현 없이는 가상현실에 적용하였을 때 사용자의 몰입도가 떨어지게 된다. 따라서 본 논문에서는 가상현실 사용자가 가상현실에 몰입하는 것을 돕기 위한 거리 변화에 대한 HRTF 모델을 제안한다.

본 논문의 구성은 다음과 같다. 제 2 절에서는 거리에 따른 HRTF의 특징을 살펴보고, 제 3 절에서는 HRTF의 특징을 이용하여 거리 변화에 따른 구조적 머리전달함수를 제안한다. 그리고 제 4 절은 제안한 구조적 머리전달함수를 적용한 3차원 사운드 생성기에 대해 기술하고 청취 실험을 통해 제안한 모델의 성능을 측정한다. 마지막으로 제 5 절에서는 결론을 맺도록 한다.

2. 거리 변화에 따른 HRTF의 특징

인간은 음원의 거리를 인지하기 위해 음의 크기와 잔향(reverberation)을 이용한다. 하지만 이중에서 잔향은 청취자의 가상현실 환경, 예를 들면 운동장, 산속, 콘서트 홀 등에 따라 크게 바뀌기 때문에 본 논문에서는 거리 변화에 있어서 음의 크기만을 고려하기로 한다.

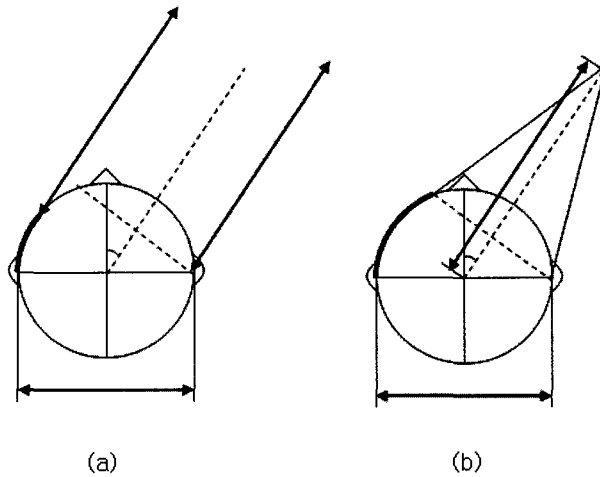
거리 변화를 고려한 HRTF 모델을 만들기 위해 음원과 청취자 사이의 거리에 따라 원거리와 근거리로 구분할 필요가 있다. 그 이유는 거리에 따라 모델링하는 방법을 다르게 해야만 하기 때문이다[6]. 일반적으로 머리에서 1 m 내의 거리를 근거리 지역(proximal region)으로, 그리고 1 m 밖의 거리를 원거리 지역(distal region)으로 나눈다. 원거리에 대한 모델링은 <그림 2> (a)와 같이 ITD의 변화 없이 ILD의 변화만을 모델링함으로써 구현할 수 있다. 이를 위해 inverse square law를 적용하여 거리 변화에 따른 ILD를 모델링한다. 하지만 음원이 근거리에 있다면 모델링하는 방법은 달라진다. <그림 2> (b)에서 보는 것과 같이 근거리일 경우는 가까워질수록 양귀에 도달하는 음파가 평행하지 않기 때문에 ITD가 증가하고, 머리에 의해 발생하는 음영 효과로 인해 ILD도 증가한다. 따라서 본 논문에서는 거리 변화를 고려한 HRTF 모델을 거리에 따라 근거리와 원거리로 나누어 제안한다.

2.1 원거리 HRTF

인간은 음원의 거리를 인지하기 위해 앞에서 언급한 것처럼 ILD와 ITD를 이용한다. 음원의 위치가 머리에서 충분히 멀다고 가정할 경우, 음파가 양귀에 동일한 각도로 평행하게 전파(propagation)되기 때문에 음원과의 거리가 다르더라도 동일한 ITD를 가지게 된다고 가정할 수 있다. 구조적 HRTF 모델의 경우 ‘Head Shadow & Delay’를 모델링을 하기 위해 충분히 멀리 떨어져 있는 음원에서 발생하는 신호를 모델링한다고 가정한다. 따라서 ITD는 변화를 주지 않고 전체적인 음의 크기 조절을 통해 만족할 만한 3차원 사운드를 얻을 수 있다. 본 논문에서는 음의 크기 조절을 위해 inverse square law를 사용한다.

2.2 근거리 HRTF

근거리에서의 HRTF 특징을 설명하기 위해 두 귀중에서 음원에 가까이 있는 귀를 near ear로, 멀리 있는 귀를 far ear로 부르기로 한다. 근거리의 경우 <그림 2>에서 보는 것처럼 음파의 진행경로가 다르기 때문에 원거리의 경우와 비교하여 ITD의 변화가 생긴다.



<그림 2> 거리에 따른 ITD 변화; (a) 원거리에서 ITD 차이 (b) 근거리에서 ITD 차이

하지만 동일 각도에서 원거리와 근거리의 ITD의 변화가 인간이 인지하지 못 할만큼 미미하기 때문에 원거리의 ITD 계산을 이용한다[6]. ILD은 기본적으로 inverse square law를 따른다. 하지만 근거리의 경우 원거리와는 달리 머리에 의해 far ear에 음영 효과로 인해 음의 크기가 작아지게 되고, 고주파 성분의 감쇄가 일어난다[7]. 또한 near ear는 방위각이 커질수록 음의 크기가 증가하는 현상이 일어난다.

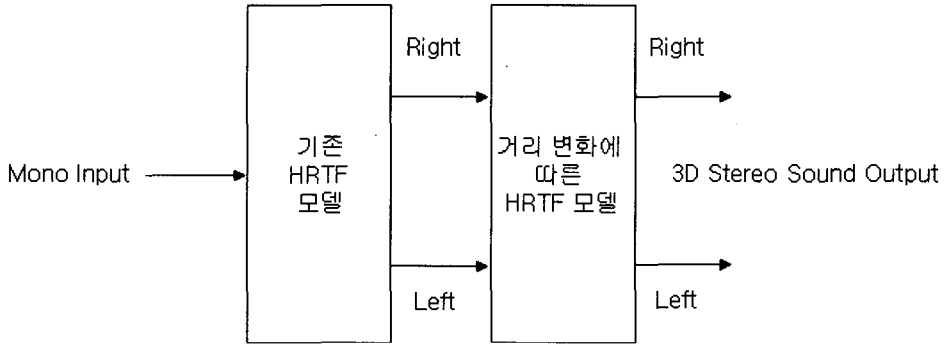
3. 제안된 거리 변화에 따른 구조적 머리 전달 함수 모델

본 논문에서 제안된 거리 변화에 따른 구조적 머리전달함수 모델은 <그림 3>과 같이 기존 구조적 머리전달함수의 후처리 단으로 연결된다. 제안된 모델은 세부적으로 크게 3개의 블록으로 나누어진다. <그림 4>에서 보는 것과 같이 제안한 모델은 음의 레벨을 표준화하는 블록(Level Normalization Block)과 제 2 절에서 설명한 바와 같이 원거리 HRTF 모델 블록과 근거리 HRTF 모델 블록으로 나뉜다. 그리고 WAVE 파일형식의 표현범위의 한계를 고려하여 생성 가능한 거리의 범위는 0.25 m에서 5 m로 제한한다.

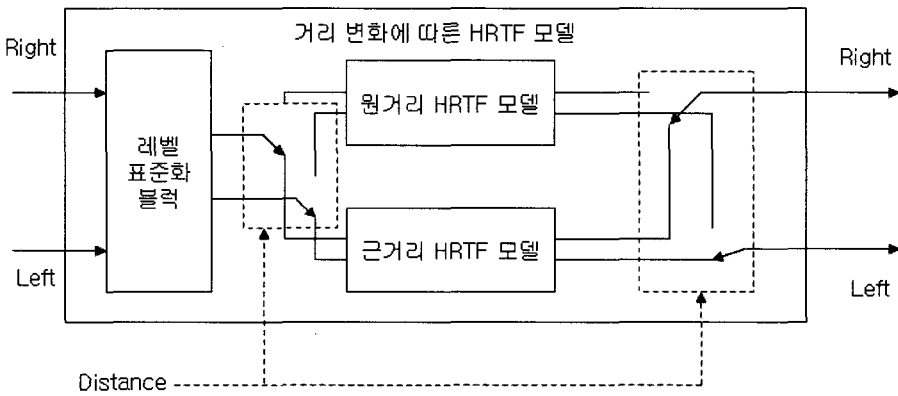
3.1 레벨 표준화

거리 변화에 따른 구조적 HRTF 모델의 입력은 16 비트 WAVE 파일이다. WAVE 파일은 -32768에서 32767까지 음원 신호의 표현이 가능하기 때문에 제안하는 HRTF 모델에서 음의 크기를 조절할 경우 표현 가능한 범위를 넘는 경우가 발생하게 된다. 이러한 overflow 문제를 해결하기 위해 WAVE 파일 샘플을 최대값이

4096이 되도록 전체적으로 음원의 크기를 조절하여 준다.



<그림 3> 제안된 구조적 머리전달함수 모델



<그림 4> 제안된 모델의 세부 블록

3.2 원거리 HRTF 모델

거리 변화에 따른 원거리 HRTF 모델은 기존 구조적 HRTF 모델로 생성된 stereo 사운드에 inverse square law를 적용한다. 이를 위한 사운드 크기를 다음 식을 통해 얻는다.

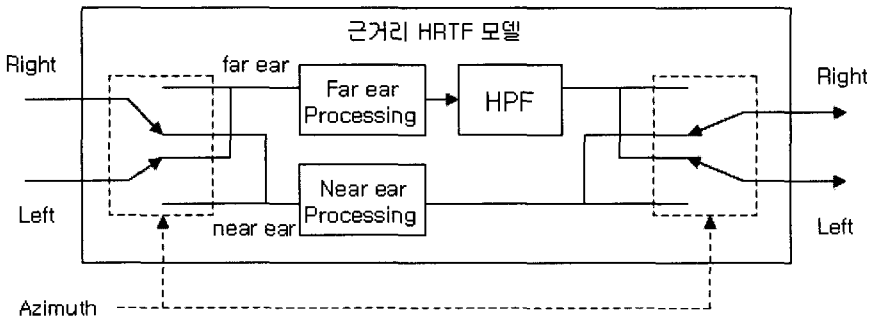
$$\frac{I_m}{I_{ref}} = \left(\frac{d_{ref}}{d_m} \right)^2 \tag{1}$$

여기서 I_m 는 거리에 따른 음의 크기이고, d_m 는 음원과의 거리, d_{ref} 는 기준 거리

인 1 m, I_{ref} 는 제 3-1 절의 레벨 표준화 블록에 의해 표준화된 음의 크기이다.

3.3 근거리 HRTF 모델

근거리 HRTF 모델의 경우 머리에 의해 shadowing 효과를 나타내는 블록과 음원의 level을 조정하기 위한 블록으로 나눈다. 다음 <그림 5>는 근거리 HRTF 모델을 나타낸 것이다.



<그림 5> 근거리 HRTF 모델

근거리에서는 far ear의 음의 크기가 거리 비율에 비례하고 고주파가 감쇄하는 특성이 있다[7]. 근거리에서 far ear에 들리는 음의 크기를 효율적으로 구하기 위하여 다음과 같은 수식을 이용하여 모델링 한다.

$$\frac{I_{m, far\ ear}}{I_{ref}} = \left(\frac{d_{ref}}{d_m} \right) \quad (2)$$

여기서, $I_{m, far\ ear}$ 는 far ear 음의 크기이다. 측정된 HRTF로 근거리에 대한 음의 크기를 조절해준 후 far ear에 대해 고주파감쇄를 위해 고역통과필터(HPF)를 far ear에 적용한다. 이를 위해 차단 주파수 4 kHz를 가지는 2차 IIR Butterworth 필터를 사용하였다.

아래의 식 (3)은 near ear를 위한 사운드 크기 조절 방식을 나타낸다. Near ear의 경우 측정된 HRTF 음의 크기가 거리 비율에 비례하고 각도에 의해 영향을 받음을 알 수 있다.

$$\frac{I_{m, near\ ear}}{I_{ref}} = \left(\frac{d_{ref}}{d_m} \right) \times 2^{\frac{|azimuth|}{90}} \quad (3)$$

여기서, $I_{m, \text{near ear}}$ 는 near ear 음의 크기이고 *azimuth*는 방위각을 나타내며 정면을 0 도로 하여 near ear로 향하는 각도로써 범위는 -90 도에서 +90 도로 한다.

4. 제안된 HRTF의 가상현실에의 구현 및 성능 평가

본 논문에서 제안한 방식으로 구현된 HRTF 모델을 실감모델 중 하나인 MP3 플레이어 조작에 관한 가상현실에 다음과 같이 적용하였다. 먼저, MP3로 압축을 위한 음원 데이터를 수집한 후, 16 비트 44.1 kHz, mono 형식의 WAVE 파일로 저장한다. 그리고 가상현실에서 사용자가 MP3 플레이어를 조작하다가 그로 인한 이벤트가 발생한다면, 가상현실 시스템으로부터 구현한 HRTF 모델은 MP3 플레이어의 현재의 위치, 즉 방위각, 고도, 거리를 입력 받는다. 이 때 위치 정보는 구좌표계(spherical coordinate)로 표현되며, 좌표계의 각각의 축은 방위각, 고도, 거리에 해당한다. 이렇게 입력된 위치 정보로부터 3차원 사운드를 생성한다. 이러한 HRTF 모델은 앞에서 언급한 것처럼 모든 위치에 대한 HRTF를 필요로 하지 않기 때문에, 저장공간 측면에서 효율적이다.

가상현실에서 MP3 플레이어 출력음을 생성하는 거리변화에 따른 구조적 HRTF 모델의 성능을 테스트하기 위해 본 논문에서는 제안한 HRTF 모델을 이용하여, 일반인을 대상으로 거리에 대한 청취 실험을 하였다. 실험에서 고도는 고정된 채, 방위각을 0 도, 30 도, 60 도, 90 도로 변화시키고, 거리는 0.25 m, 0.50 m, 1 m, 1.5 m, 2 m로 변화시켰을 때 생성된 3차원 사운드를 헤드폰을 이용하여 청취하였다. 실험 방법은 제안된 HRTF 모델로 생성한 3차원 사운드를 1 m 기준음과 비교 청취 후 생성된 3차원 사운드의 거리를 예측하도록 하였다. 실험에 사용한 음원은 16 비트, 44.1 kHz, mono WAVE 파일 형식으로 모든 주파수 성분을 고루 가지고 있는 총소리를 이용하였다. 청각적 장애가 없는 20세 이상의 남녀 10명을 대상으로 실험을 진행하였으며 실험 결과는 <표 1>과 같다.

<표 1>에서 보는 바와 같이, 0.25 m, 0.5 m, 1.5 m, 2.0 m의 거리에 대해 생성한 3차원 사운드는 평균 오차 0.13 m, 0.16 m, 0.30 m, 0.31 m를 각각 가지는 것을 알 수 있다. 인간에게 실제로 위와 동일한 거리만큼 떨어진 곳에서 음원을 들려줬을 경우, 평균 인지 오차가 0.15 m, 0.15 m, 0.2 m, 0.4 m인 것을 감안한다면 제안한 모델은 만족할 만한 성능을 가진다고 할 수 있다 [2]. 특히 오른쪽으로 60 도일 경우 가장 근사한 결과를 얻었고, 방위각에 따른 큰 편차를 보이지 않음을 알 수 있다.

<표 1> 3 차원 사운드를 이용한 거리 인지 실험

(단위 : m)

생성방위각	생성 거리	인지 평균	인지오차 평균
0 도	0.25	0.32	0.11
	0.5	0.70	0.20
	1.5	1.81	0.33
	2.0	2.22	0.28
우현 30 도	0.25	0.35	0.14
	0.5	0.65	0.15
	1.5	1.55	0.33
	2.0	2.02	0.44
우현 60 도	0.25	0.34	0.13
	0.5	0.67	0.17
	1.5	1.32	0.28
	2.0	2.08	0.26
우현 90 도	0.25	0.34	0.13
	0.5	0.65	0.15
	1.5	1.41	0.31
	2.0	1.87	0.25

제안된 근거리 모델의 성능을 분석하기 위해 기존의 방식인 원거리 HRTF 모델과 비교하였다. 실험을 위하여 기존의 방식과 제안된 방식에 대해 생성 거리 0.25 m, 0.5 m에서 3차원 사운드를 각각 생성하였고, 실험 결과는 <표 2>와 같다.

<표 2>에서 보는 바와 같이, 제안된 방식의 인지 오차 평균값이 기존의 방식에 비해 작은 것으로 나타났다. 특히 생성 거리 0.5m에서 인지 오차 평균값이 20% 감소된 것을 확인 할 수 있었다. 이는 기존 방식에 비해 제안된 방식이 거리 인지에 대한 성능을 향상시키는 것을 의미한다.

<표 2> 기존 모델과 제안된 모델의 거리 인지 성능 비교

(단위 : m)

생성방위각	생성 거리	기존(원거리) 방식		제안된 방식	
		인지 평균	인지오차 평균	인지 평균	인지오차 평균
0 도	0.25	0.12	0.13	0.32	0.11
	0.5	0.30	0.20	0.70	0.20
우현 30 도	0.25	0.11	0.14	0.35	0.14
	0.5	0.30	0.20	0.65	0.15
우현 60 도	0.25	0.11	0.14	0.34	0.13
	0.5	0.30	0.20	0.67	0.17
우현 90 도	0.25	0.10	0.15	0.34	0.13
	0.5	0.32	0.18	0.65	0.15

5. 결 론

본 논문에서는 실감 모델의 위치 변화에 따른 효과적인 3차원 사운드를 생성하기 위해 거리 변화에 따른 구조적 머리전달함수를 제안하였다. 그리고 제안한 머리전달함수로 생성한 3차원 사운드를 이용한 청취 실험을 통하여 그 성능을 측정하고 오차를 분석하였다. 제안한 알고리즘의 오차는 인간이 실제 거리를 인지할 때 발생하는 오차 범위 내에 있음을 확인하였다. 제안된 HRTF 방식을 실감 모델의 하나인 MP3 플레이어에 적용한 결과, 기존 구조적 HRTF 모델에 비해 거리 변화에 따른 사운드를 제공함으로써 가상현실 사용자에게 몰입감을 높이는데 효과적인 것으로 평가되었다.

참 고 문 헌

- [1] 강성훈, “오디오와 심리음향”, *국제음향공학위크샵*, pp. 1-11, 1996년 7월.
- [2] J. Blauert, *Spatial Hearing*, Revised Edition, MIT Press, Cambridge, MA, 1997.
- [3] C. P. Brown and R. O. Duda, “An efficient HRTF model for 3-D sound”, *Proc. ASPAA*, pp. 298-301, Oct. 1997.
- [4] D. J. Kistler and F. L. Wightman, “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction”, *Journal of the Acoustical Society of America*, Vol 91, No 3, pp. 1637-1647, Mar. 1992.
- [5] N. M. Cheung, S. Trautmann, and A. Horner, “Head-related transfer function modeling in 3-D sound systems with genetic algorithms”, *Journal of the Audio Engineering Society*, Vol 46, No 6, pp. 531-539, Jun. 1998.
- [6] R. O. Duda and W. L. Martens, “Range dependence of the response of a spherical head model”, *Journal of the Acoustical Society of America*, Vol. 104, No. 5, pp. 3048-3058, Nov. 1998.
- [7] D. S. Brungart and W. M. Rabinowitz, “Auditory localization of nearby sources. Head-related transfer functions”, *Journal of the Acoustical Society of America*, Vol 106, No 3, pp. 1465-1479. May 1999.
- [8] D. R. Begault, *3D Sound for Virtual Reality and Multimedia*, Academic Press, Cambridge, MA, 1994.
- [9] W. G. Gardner and K. D. Martin, “HRTF measurements of a KEMAR”, *Journal of the Acoustical Society of America*, Vol. 97, No 6, pp. 3907-3908, Jun. 1995.

접수일자: 2006년 8월 15일

게재결정: 2006년 9월 25일

▶ 이영한(Young Han Lee)

주소: 50-712 광주광역시 북구 오룡동 1번지 광주과학기술원
소속: 광주과학기술원(GIST) 정보통신공학과 휴먼컴퓨팅 연구실
전화: 062) 970-3121
E-mail: cpumaker@gist.ac.kr

▶ 김홍국(Hong Kook Kim) : 교신저자

주소: 500-712 광주광역시 북구 오룡동 1번지 광주과학기술원
소속: 광주과학기술원(GIST) 정보통신공학과 휴먼컴퓨팅 연구실
전화: 062) 970-2228
E-mail: hongkook@gist.ac.kr