한국형 HRTF를 이용한 입체음향 구현

(*)김재현, 정상배, 양희식, 한민수
한국정보통신대학원대학교 공학부
대전광역시 유성구 화암동 58-4

Implementation of 3-D Audio using Korean-Type HRTF

(*)Jaehyun Kim, Sangbae Jeong, Heesik Yang, Minsoo Hahn
Information and Communications University(ICU)
E-mail: (*)hijhkim@icu.ac.kr

요약

입체음향의 구현은 21 세기 멀티미디어 콘텐트 관련 산업의 핵심기술 중 하나로 인식되고 있으며, 그 응용분야가 매우 넓기 때문에 이에 대한 투자가 점차 늘어가고 있는 실정이다. 본 논문은 한국인의 표준형 두상에 맞는 HRTF(Head-Related Transfer Function)를 이용한 입체음향의 구현 및 현장효과의 인공적 재현 방법에 대한 연구 결과이다.

1. 서론

다양한 멀티미디어 콘텐트 중에서 가장 핵심이 되는 요소는 영상, 음성 및 음향 정보라고 할 수 있으며 인터넷과 같이 다수의 사용자가 공유하는 가상 공간에서 중요한 매개체로 자리매김하고 있다. 멀티미디어 콘텐트 중에서 음성 및 음향정보를 담고 있는 경우의 예를 들자면, 가상현실, 게임 소프트웨어, 일반적인 음반 CD, 디지털 TV 방송에서의 음성음향 정보 등을 말할 수 있다. 이미 미국, 일본 같은 선진국에서는 멀티미디어 콘텐트 관련 산업을 21 세기의 정보산업의 핵심으로 인식하고 그 투자를 늘려가고 있는 실정이다.

입체음향은 「원래의 음장을 확실히 재현하고 음의 고저, 음색 뿐만 아니라 방향이나 거리감까지도 재생하여 임장감을 가지게 하는 음향」으로 정의가 되어 있다. 이러한 음향을 구현하기 위해서는 인간의 두상에 맞는 HRTF 를 이용하여, 마치 특정위치에서 소리가 들리는 것처럼 해 주는 3 차원 음향의 구현이 필수적이다. 현재, 국내외에서 구현된 입체음향 생성기는 대부분 MIT-Media Lab. 에서 측정된 HRTF 를 사용하고 있다. 그러나, 외국인의 표준형 두상에 맞는 HRTF 를 사용한 입체음향의 구현은 한국인에게는 잘 맞지 않을 가능성이 크다. 따라서 본 연구에서는 한국인의 표준형 두상에 맞게 측정된 HRTF 를 이용한 입체음향의 구현을 주 목표로 한다.

2. 한국형 HRTF

2-1. HRTF 의 정의 및 측정법

HRTF 는 음원에서 고막까지 소리가 전파될 때의 음의 전달경로에 의한 단위 충격 응답(impulse response)으로 정의된다. 수식적인 정의는 식 (1)과 같다. φ, θ 는 각각 방위각과 고도를 나타낸다.

$$\frac{P_2}{P_1}(\varphi,\theta) = \frac{\text{막혀있는 외이도 입구에서의 음압}}{\text{머리 중심에서의 음압}} \quad (1)$$

HRTF 의 값으로서, 주파수 영역에서의 응답을 사용하지 않고 시간영역에서의 값을 사용하는 이유는 실내의 공간적 음향특성을 헤드폰이나 오디오 시스템에 쉽게 인가할 수 있기 때문이다. 이

러한 HRTF 는 음원에서 고막까지의 경로에 존재하는 머리, 몸통, 귀 등에서의 반사와 회절에 대한 효과를 포함하고 있다. 사람이 음원의 위치를 파악하는 메커니즘이 물리적으로 정확히 밝혀진 바는 없으나, 앞서 언급한 반사와 회절에 밀접한 관계가 있음이 실험적으로 확인되고 있다. 이와 같은 특성은 소리가 귀에 입사하는 방향에 따라서 틀려지게 된다. 따라서 HRTF 의 측정은 구좌표계에서 단위 원 내에 미리 정해진 각 점에서 측정이 이루어져야 한다. MIT-Media Lab. 에서는 1994 년에 미국인의 표준형 두상을 모델링한 KEMAR Dummy Head 를 이용해서 측 HRTF 를 측정한 후에 전세계에 인터넷으로 그 데이터를 공유하고 있다[1]. 국내에서는 1996 년에 한국전자통신연구원(ETRI)에서 한국인의 표준형 두상에 맞는 HRTF 를 최초로 측정하였다[2]. MIT HRTF 와 ETRI HRTF 는 다음의 측정사양에서 동일하였다. 그것을 Table 1 에 정리하였다.

Table 1 : HRTF 의 측정사양

	측정사양
Sampling Rate	44.1 kHz
Resolution	16 bit
Number of measuring Positions	710 positions
Length	512 points

2-2. MIT 및 ETRI HRTF 의 비교

MIT 및 ETRI 에서 측정된 HRTF 의 비교를 위해서 그림 1-(a),(b)에서 시간영역 및 주파수영역에서의 응답특성을 나타내었다.

그림 1 에서 알 수 있듯이 MIT HRTF 는 2~4 kHz 및 10~15kHz 대역에서 ETRI HRTF 보다 큰 이득을 보이고 있다. 원신호를 HRTF 에 통과시켰을 때에는 주파수 영역에서의 왜곡을 피할 수 없겠지만, MIT HRTF 를 이용하였을 때, 특정 주파수 대역에서의 왜곡의 양이 더 커짐을 예측할 수 있다. 실

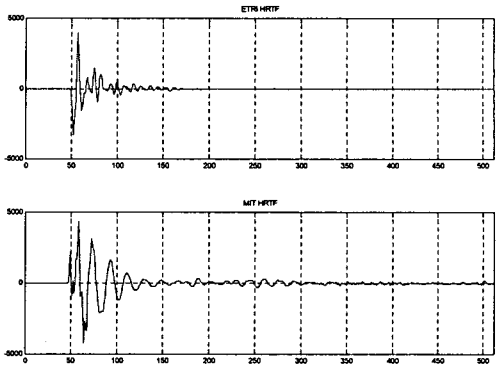


그림 1-(a) : 고도 0, 방위각 45 에서의 HRTF

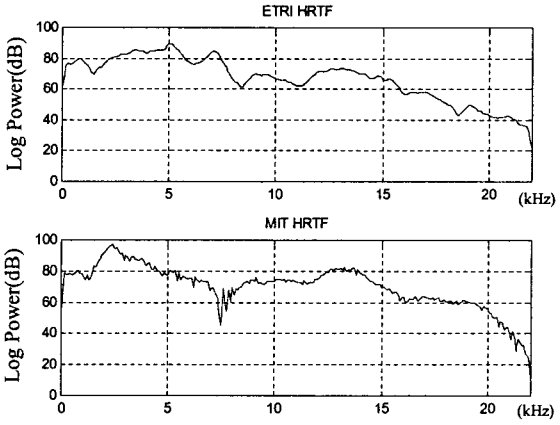


그림 1-(b) : 그림 1-(a)의 주파수 응답

제로 MIT HRTF 를 이용하여 입체음향을 생성하였을 경우에 ETRI HRTF 를 이용하였을 때보다 고주파 성분이 강조됨을 느낄 수가 있다. 이는 앞서 언급한 특정 주파수 대역에서의 이득이 상대적으로 높아졌기 때문이다. 따라서 본 연구에서는 주파수 왜곡이 상대적으로 적고 한국인의 두상에 맞게 측정된 ETRI HRTF 를 이용한 입체음향의 구현을 목표로 한다.

3. HRTF 를 이용한 입체음향의 구현

3-1. HRTF 데이터의 축소

입체음향의 구현은 주로 선형 복적분(linear convolution) 연산을 통해서 이루어 지며, 선형 복적분은 곱셈 연산과 덧셈 연산의 조합이다. 예를

들어, 입력되는 오디오 신호의 표본화율이 44.1 kHz 이고, 각 위치에서 주어지는 512 개의 HRTF 계수를 모두 사용한다고 가정한다면, 실시간 구현을 위해서 100 Mips 급의 고성능 DSP(Digital Signal Processor)를 필요로 하게 된다. 그렇지만, 그림 1-(a)에서 확인할 수 있듯이, HRTF 데이터는 시간지연의 양이 적은 부분에서는 큰 값이 분포하고 시간지연의 양이 큰 부분에서는 미소한 값이 분포한다. 따라서, 본 연구에서는 512 개 중에서 앞부분의 20 개를 버린 후에 64 개만을 입체음향 구현에 이용하기로 한다. 데이터 축소에 따른 주파수 왜곡은 저주파에서 클 것으로 예상할 수 있으나, 그림 2 에서 확인할 수 있듯이 그 양은 아주 적었다. 그림 2 는 그림 1-(a)에서 데이터 축소를 시행하였을 때의 주파수 특성이다.

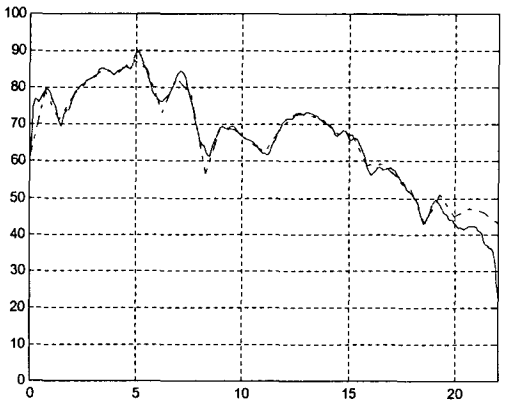


그림 2: 데이터 축소에 따른 주파수 왜곡
(실선: 512 points, 점선: 64 points)

3-2. HRTF 의 선형 보간

HRTF 의 측정은 3 차원 공간 상에서 미리 정해진 710 곳에서 측정되었기 때문에 임의의 위치에 대한 HRTF 의 응답을 이용하고자 할 경우에는 그 주위의 이미 알고 있는 HRTF 정보를 이용한 추정이 필요하다. 이러한 과정을 HRTF 의 보간이라 한다.

HRTF 의 보간은 추정하고자 하는 위치에서 미리 측정된 HRTF 4 곳까지의 거리를 측정하고, 최종적으로 4 개의 HRTF 에 거리의 역수에 해당하는 가중치를 곱해서 더해주는 방법으로 구할 수 있다. 그림 3 에서, Q 에서의 HRTF 를 추정한다고 가정하자. A, B, C, D 는 Q 와 가장 가까운 기지의 HRTF 측정 위치이다.

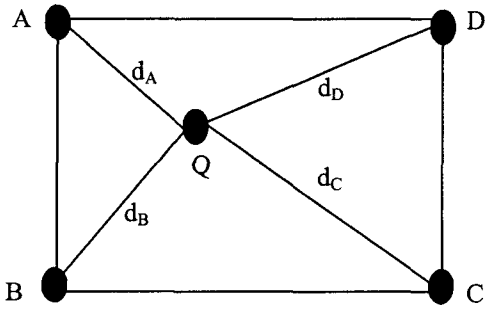


그림 3: HRTF 의 보간 예

각 위치, A, B, C, D 에서의 HRTF 값을 $HRTF_A$, $HRTF_B$, $HRTF_C$, $HRTF_D$ 라고 하면, 위치 Q 에서의 HRTF 는 식 (2)와 같이 구할 수 있다.

$$HRTF_Q = \frac{\frac{HRTF_A}{d_A} + \frac{HRTF_B}{d_B} + \frac{HRTF_C}{d_C} + \frac{HRTF_D}{d_D}}{\frac{1}{d_A} + \frac{1}{d_B} + \frac{1}{d_C} + \frac{1}{d_D}} \quad (2)$$

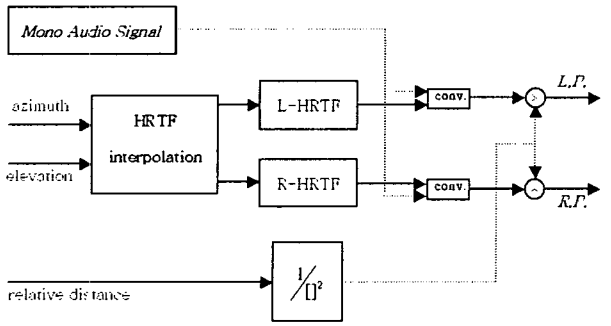
이러한 선형 보간의 단점은 여현 이득 왜곡(cosinusoidal magnitude distortion)이 발생한다는 것이다[3][6]. 여현 이득 왜곡은 시간 지연이 다른 두 신호를 합하여 새로운 신호를 추정할 때 생기는 현상인데, 추정된 신호의 주파수 응답이 원래의 주파수 응답에 여현 신호가 곱해진 형태가 되는 것을 말한다. 그렇지만, 본 연구에서 시행한 HRTF 의 보간은 시간지연의 양이 대부분 1 샘플 미만이므로 여현 이득 왜곡의 양은 무시할 정도로 적으리라고 추측할 수 있다.

3-3.거리 증감 효과의 구현

거리 증감 효과의 구현은 단순히 전력제어의

개념을 도입하여 사용하고 있다. 입체음향의 전력제어는 음원과 청취자 사이의 상대적인 거리의 증감을 계산하여 구현한다. 음향의 강도(intensity)는 무향실(anechoic chamber)에서 측정하였을 때, 거리가 2 배 증가하면, 약 6 dB 정도 감소함이 알려져 있다. 이것을 역 제곱 법칙(inverse square law)라고 한다.

그림 4 에 전체적인 입체음향의 구현법을 도시하였다.



conv. : convolution operation

R.P. : Right Program Signal , *L.P.* : Left Program Signal

그림 4 : 전체적인 입체음향의 생성과정

4. 현장감있는 입체음향의 구현

현장감 있는 입체음향의 구현을 위해서는, 재현하고자 하는 음장 환경의 단위 충격 응답을 이용해야 한다. 일반적으로 특정 음장에서의 단위 충격 응답은 초기 반사(Early Reflection)와 추후 반사(Late Reverberation)로 나누어서 분석을 한다[4][7]. 초기 반사는 전체 단위 충격 응답과 비교하여 상대적으로 큰 값을 가지게 되며, 음장효과 구현에서 중요한 역할을 하게 된다. 추후 반사는 음원에서 발생한 신호가 여러 경로를 통해 반사와 반사를 거듭하여 단위 충격 응답을 잡음으로 모델링 가능한 영역을 말한다. 특정 음장에서의 단위 충격 응답은, 초기 반사의 경우에 특정 위치에서 상대적으로 큰 값을 가지는 충격 함수 열로 모델링을 하고, 추후 반사의 경우에는 지수적으로 감소

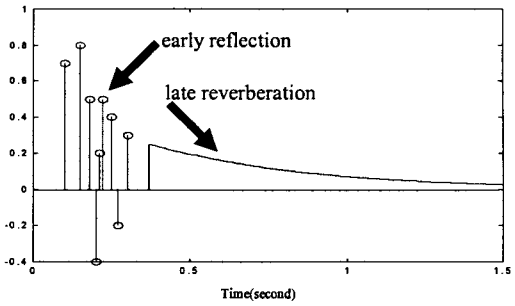


그림 5 : 단위 충격 응답의 모델링 예

하는 잡음 신호를 이용하여 근사화 시킨다. 그림 5 는 단위 충격 응답의 모델링을 나타낸 것이다.

5. GUI 환경에서의 입체음향 구현

입체음향의 구현을 위해서 음향제어의 용이성 및 현장감을 살릴 수 있는 10 ~ 20 초 분량의 모노 오디오 신호를 입력으로 하는 시스템을 구현하였다. 입력 오디오 신호의 표본화율 및 양자화율은 44.1 kHz – 16 bit 였다. 기본적으로 입체음향 생성엔진은 그림 4 를 바탕으로 Visual C++ 를 이용하여 제작되었다.

현장감있는 입체음향의 생성을 위해서 성당, 오페라홀, 터널 등에서의 단위 충격 응답을 분석하였다. 초기 반사는 50 개의 충격 함수로 모델링하였으며, 추후 반사는 복적분시의 약간의 성능저하를 감수하고 계산량 감축을 위해서 200 개의 난수(Uniform Random Variable)를 발생시킨 후에 원하는 음장의 단위 충격 함수의 길이에 맞도록 등간격으로 맞추어 근사화 시켰다.

입체음향 생성엔진의 음원제어를 확인하기 위해서 여러 가지 시나리오에 따른 음원이동 모드 역시 구현되었다.

일반적으로 스피커를 통한 입체음향의 구현을 위해서는 크로스토크 제거기의 설계가 필수적이다. 크로스토크 제거기가 제대로 동작하기 위해서는 스피커가 놓여 있는 환경에서의 단위 충격응답의 실시간 등화 및 청취자가 이동하였을 때의

상황을 고려하여 적응성을 띈 알고리즘의 사용이 필수적이다. 이러한 알고리즘의 구현은 기술적으로 큰 문제점이 없지만, 본 연구에서는 연구 환경의 제약상 무향실 및 청취자가 고정되어 있다는 가정하에서의 크로스토크 제거기를 구현하였다[5]. 구현된 시스템을 그림 6 에 나타내었다.

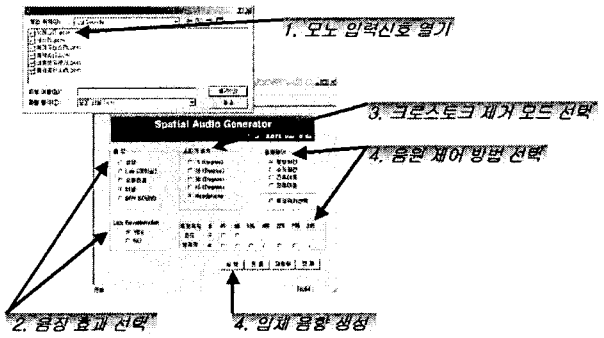


그림 6 : GUI 환경에서 구현된 시스템

6. 실험결과 및 결론

구현된 시스템을 이용하여 입체음향을 생성하였을 때, 미리 정해 놓은 음원 제어 시나리오에 합당한 2-channel 오디오 신호가 생성되고 있음을 확인할 수 있었으며, 음장효과를 첨가하였을 경우에는 원래의 해당하는 음장의 단위 충격 응답을 모두 사용하였을 때보다 성능 저하가 약간 느껴지지만 만족할 만한 효과를 나타내었다.

본 시스템의 성능을 높이기 위해서는 복적분 전용 DSP 를 이용한 음장감 개선이 필요하고, 청취자의 위치를 적외선 센서를 사용하여 실시간으로 추적하고 청취자가 움직였을 때에 해당 청취 환경에서의 단위 충격 응답의 실시간 측정과 이를 고려한 크로스토크 제거기의 구현이 필요하다고 하겠다.

참고문헌

[1] ftp://sound.media.mit.edu/pub/Data/KEMAR

[2] KRISS-96-124-IR, *한국인의 표준 HATS 제작과 머리전달함수 측정 연구*, 한국전자통신연구원

[3] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*, 1994, ISBN 0-12-084735-3

[4] John Garas, *Adaptive 3D Sound System*, Kluwer academic Publishers, 2000.

[5] William G. Gardner, 3-D Audio using Loudspeakers, Kluwer academic Publishers, 1998

[6] E. M. Wenzel, S. H. Foster, "Perceptual Consequences of Interpolating Head-Related Transfer Functions During Spatial Synthesis," *IEEE Workshop '93*

[7] D. R. Begault, "Perceptual Effects of Synthetic Reverberation on Three dimensional Audio Systems,"*J. Audio Eng. Soc.*, Vol. 40, No. 11, pp. 895-904, Nov. 1992.