

3차원 음상정위 시스템의 구현과 주관 평가

이동우*, 김영오*, 고대식*, 강성훈**, 김현빈***

*목원대학교, **대전보건대학, ***ETRI

Implementation of Spatial Sound Localization System and Subjective Test

Dongwoo Lee*, Youngoh Kim*, Deasik Ko*, Seonghoon Kang**, Hyunbin Kim***

*Mokwon Univ., **Taejon Medical Science College, ***ETRI

dwlee@ee.mokwon.ac.kr

요약

본 논문에서는 헤드폰과 스테레오 스피커를 통하여 가상의 음상을 임의의 위치에 정위시키는 음상정위 시스템을 구현하고, 주관 평가를 통하여 음상정위 성능을 고찰하였다. 음상정위 시스템은 크게 방향감을 제어하는 컨벌루션 처리부와 공간감과 거리감을 처리하는 잔향 처리부, 그리고 스테레오 스피커를 통해 소리를 재생할 때 발생하는 크로스 토크(crosstalk)를 제거하기 위한 트랜스오럴(transaural) 필터부로 나누어진다. 구현된 시스템의 음상정위 성능은 리스링 룸에서 녹음된 유성과 메트로놈 소리를 이용하여 수평각/고도각, 정지음/이동음, 거리감등을 헤드폰과 스피커를 통하여 각각 실험한 결과 수평각 지각은 스피커 재생보다 헤드폰 재생이 우수했으며, 정지음보다 이동음의 지각 결과가, 고도각 지각은 전·후(0°~360°) 방향보다 좌·우(90°~270°) 방향의 결과가 우수하게 나왔다.

1. 서론

인간은 시각의 도움 없이 소리를 듣는 것만으로도 어느 방향에서 소리가 들리며, 음원까지의 거리가 어느 정도가 되는지, 그리고 청취공간의 크기는 얼마나 되는지를 지각할 수 있다. 이와 같이 소리를 듣는 것만으로 음원의 위치 등과 같은 3차원 정보를 지각할 수 있는 이유는 음원으로부터 발생한 소리가 두 귀에 도달하는데 걸리는 시간의 차이(Interaural Time Difference : ITD)와 두 귀에 도달한 소리의 레벨차이

(Interaural Level Difference : ILD), 그리고 소리가 공기 중을 진행할 때의 전파 특성과 물체에 부딪쳐서 반사, 굴절하는 성질 등에 의한 것이다. 이러한 조건을 인위적으로 조작할 수 있다면 실제의 청취환경에 제한을 받지 않고 원하는 청취공간을 만들어 그 안에서 원하는 위치에서 소리가 들리게 할 수 있을 것이다[1].

본 논문에서는 가상의 청취공간을 만들고 음원을 임의의 위치에 정위시키는데 필요한 3차원 음상정위 제어 알고리즘에 대하여 기술하고, 이 알고리즘들이 적용된 입체 음향 제어 시스템의 구현과 주관 평가를 통해 분석된 성능에 대하여 기술하였다.

2. 입체 음상 제어 알고리즘

2.1 방향감 제어 알고리즘

인간이 소리를 듣는 것만으로도 음원의 방향을 지각할 수 있는 이유는 두 귀에 도달한 소리의 시간차와 레벨차에 의한 것이다. 이와 같은 차이는 소리가 귀에 도달하기 전에 공기에 의해 감쇠가 일어나고 몸통, 머리, 외이 등에 의해 간섭이 생기기 때문이다. 이와 같이 방향에 따라 인간의 청각 시스템이 어떻게 반응하는지를 나타낸 것이 머리 전달 함수(Head-Related Transfer function : HRTF)이다.

방향감 제어는 음원으로부터 청취자까지의 머리 전달 함수를 이용하면 가능하다. 즉, 그림 1과 같이 음원으로부터 두 귀까지의 머리 전달 함수를 음원과 컨벌루션하면 음원은 방향 특성이 있는 바이노럴(binaural) 신호가 되어 헤드폰으로 청취하면 청취자는 소리가 머리 전달 함수가 나타내는 방향에서 들리는 것과 같이

있을 수 있다[2].

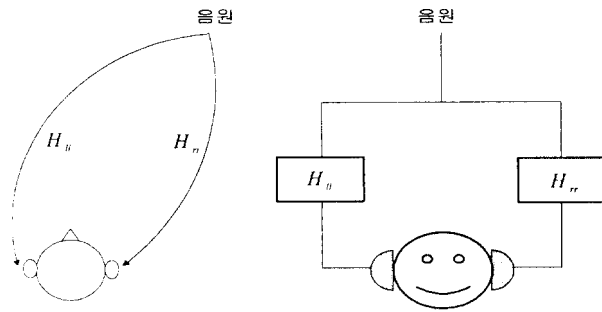


그림 1 방향감 합성원리

2.2. 거리감 제어 알고리즘

거리감은 음압 레벨, 스펙트럼 성분 그리고 잔향 성분 등의 여러 가지 요인의 종합적인 결과이다. 그 중 음압 레벨은 청취자가 거리를 지각하는데 있어 가장 기본적인 단서이다.

그림 2와 그림 3은 콘서트 홀의 각 좌석에서 측정한 압력 레벨 레스폰스를 분석한 것이다. 그림 2는 각각의 좌석에서 잔향 시간의 변화와 D-value와의 관계를 나타내고 있다. D-value는 직접음 도달 후 50ms까지의 직접음 에너지와 잔향 에너지와의 비이다. 잔향 시간은 좌석에 관계없이 거의 일정한 것을 알 수 있다. 그러나 D-value는 음원으로부터 멀어짐에 따라서 D-value가 직선적으로 감소되고 있는 것을 알 수 있다.

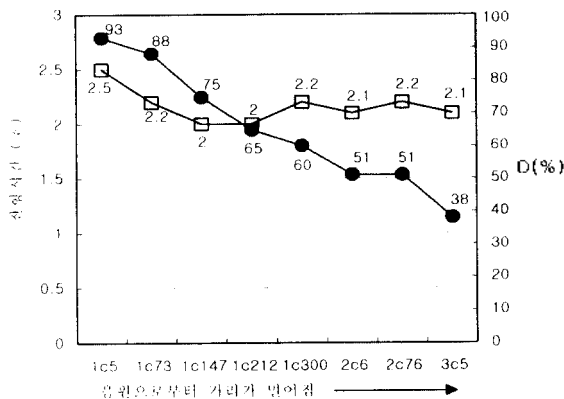


그림 2 음원으로부터의 거리변화에 따른 잔향 시간과 D-value의 변화

그림 3은 각 좌석에서의 잔향 시간의 변화와 R-value(잔향량)와의 관계를 나타내고 있다. 여기에서

도 잔향 시간은 좌석에 관계없이 거의 일정하지만, R-value는 음원으로부터 멀어짐에 따라서 직선적으로 감소되고 있는 것을 알 수 있다. 그러므로 잔향음과 직접음의 비율을 조정함으로써 거리감은 제어 가능하다[3].

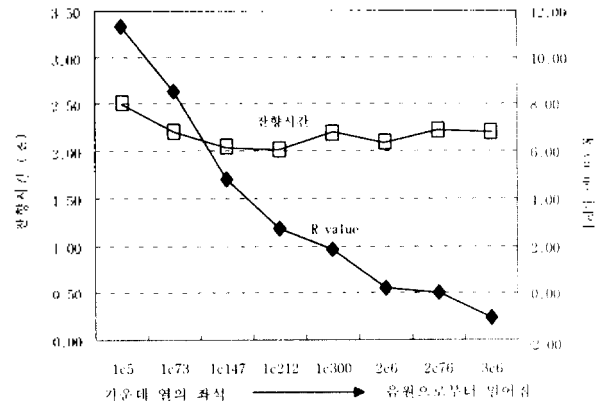


그림 3. 음원으로부터의 거리 변화에 따른 잔향 시간 및 R-value의 변화

2.3. 트랜스오럴 필터 알고리즘

트랜스오럴 필터는 방향 특성을 갖는 바이노럴 신호를 입력받아 스테레오 스피커로 재생하기 위한 것으로 원리를 살펴보면 다음과 같다. 그림 4는 청취자가 스테레오 스피커를 통해 소리를 청취할 때의 상황이다.

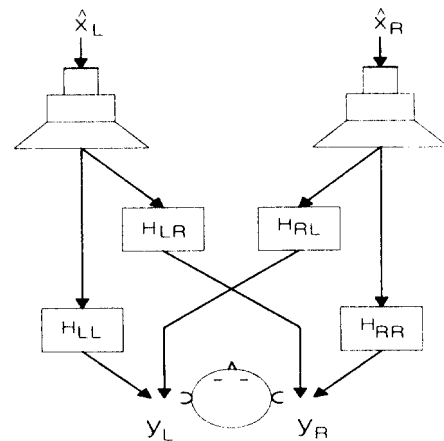


그림 4. 스테레오 스피커로부터 귀까지의 전달함수

X_R 과 X_L 은 스피커로 보내지는 바이노럴 신호이고, Y_L 과 Y_R 은 청취자의 두 귀에 도달하는 신호이다. HLL,

H_{LR} , H_{LR} , H_{RL} 은 각각 왼쪽 스피커에서 왼쪽 귀까지의 공간 전달함수, 오른쪽 스피커에서 오른쪽 귀까지의 공간 전달함수, 왼쪽 스피커에서 오른쪽 귀까지의 공간 전달함수, 오른쪽 스피커에서 왼쪽 귀까지의 공간 전달함수이다. 이러한 시스템은 벡터 식으로 표현하면 식 1과 같다.

$$y = \begin{bmatrix} y_L \\ y_R \end{bmatrix}, H = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix}, \hat{x} = \begin{bmatrix} \hat{x}_L \\ \hat{x}_R \end{bmatrix}, \quad (1)$$

$$y = H\hat{x}$$

스피커 X에서 양쪽 귀 Y로의 전달 함수는 H_{XY} 이다. 그러므로, 스피커 X에서 바이노럴 신호 x가 방출되면 공간 전달함수 H를 통과해 양쪽 귀 Y에 신호 y가 도달하는데, 귀에 도달하는 신호 y는 스피커 X에서 방출된 신호 x와 같지 않다. 즉, 오른쪽 귀에는 오른쪽 스피커에서 방출된 신호와 왼쪽 스피커에서 방출된 신호가 함께 도달하며, 왼쪽 귀에 도달하는 신호 역시 마찬가지이다. 그리고 양쪽 귀 Y에 도달한 신호 y는 전달 함수 H의 특성을 가지고 있다. 이와 같은 현상을 크로스 토크라고 하며, 크로스 토크를 제거하기 위한 일반적인 트랜스오럴 필터는 그림 5와 같다.

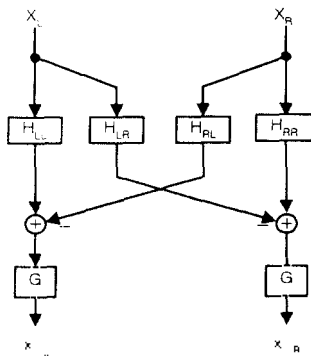


그림 5. 일반적인 트랜스오럴 필터

$$(G = 1/(H_{LL}H_{RR} - H_{LR}H_{RL}))$$

트랜스오럴 필터의 원리는 다음과 같다. x가 청취자의 귀에 도달하기 원하는 바이노럴 신호이면, 즉 스피커 X에서 방출된 신호 x를 원래의 신호 그대로 듣기 위해서는 전달함수 행렬 H는 $\hat{x} = H^{-1}x$ 와 같이 역 함수를 취해야 한다. 즉 스피커 X로부터 출력되는 신호 x와 귀 Y에 도착하는 신호 y를 같게 만들려면, 전달함수 H의 역함수를 구해야 한다. 전달함수 H의 역행렬은 식 2와 같다[4].

$$H^{-1} = \frac{1}{(H_{LL}H_{RR} - H_{LR}H_{RL})} \begin{bmatrix} H_{RR} & -H_{RL} \\ -H_{LR} & H_{LL} \end{bmatrix} \quad (2)$$

3. 실시간 음성 제어 시스템 구현

실시간 음성 제어 시스템은 제어 수단을 기준으로 음원으로부터 두 귀에 도달한 소리의 차이 즉 머리전달함수를 이용한 방향감 제어부와 인공 잔향을 이용한 거리감, 공간감 제어부, 그리고 스테레오 스피커를 통해 바이노럴 신호를 재생할 때 발생하는 크로스 토크를 제어하는 트랜스오럴 필터부로 구성된다. 실시간 처리를 위해서 4개의 범용 DSP(Digital Signal Processor)로 구성된 디지털 신호 처리기를 이용하였다.

그림 6은 실시간 음성 제어 시스템의 기본 구성도이다. 이 시스템은 크게 컨벌버(convolver), 잔향기(reverberator), 믹서(mixer), 그리고 트랜스오럴 필터로 구성된다. 각각의 알고리즘은 각 DSP에서 독립적으로 동작하며, 컨벌버와 잔향기는 병렬로 연결되고, 믹서와 트랜스오럴 필터는 차례대로 컨벌버와 잔향기의 출력에 직렬로 연결된다[5].

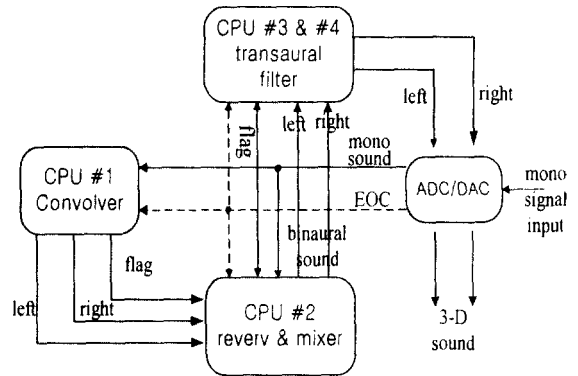


그림 6. 실시간 음성 제어 시스템의 구성도

4. 실험 및 고찰

실시간 음성 제어 시스템의 성능 실험은 주관 평가를 통하여 이루어졌다. 주관 평가를 위한 음원은 리스링 뮐에서 녹음된 남성의 음성(약 3초)과 메트로놈 소리를 이용하였다. 음성 제어 시스템은 실시간 처리가 가능하지만 실험의 용이함을 위하여 녹음된 음원을 실시간 처리하여 실험에 필요한 순서대로 MD(mini disk)에 녹음하였다. 녹음은 음성-메트로놈을 각각 다음과 같은 순서로 이루어 졌다.

헤드폰 - 정지음 - 수평각

- 헤드폰 - 정지음 - 수직각(전후방향)
- 헤드폰 - 정지음 - 수직각(좌우방향)
- 헤드폰 - 정지음 - 거리감
- 헤드폰 - 이동음 - 수평각
- 헤드폰 - 이동음 - 수직각(전후방향)
- 헤드폰 - 이동음 - 수직각(좌우방향)
- 스피커 - 정지음 - 수평각
- 스피커 - 정지음 - 수직각(전후방향)
- 스피커 - 정지음 - 수직각(좌우방향)
- 스피커 - 이동음 - 수평각
- 스피커 - 이동음 - 수직각(전후방향)
- 스피커 - 이동음 - 수직각(좌우방향)

주관 평가는 방음실(암소음 25dBA)에서 실시하였으며, 음의 재생은 헤드폰과 스피커를 이용하였다. 각각의 방향 지각 실험은 하나의 HRTF를 이용한 정지음 지각 실험과 이웃한 12개의 HRTF를 0.25초마다 변형하여 방향을 변화시킨 이동음 지각 실험으로 나누어 이루어졌다. 수직각은 전후방향(0°~180°)과 좌우방향(90°~270°)으로 나누어 실험하였다. 거리감은 산형음과 직접음의 비율을 조정하여 3개의 상대적인 거리를 표현하여 실험하였다. 실험에는 음상 정위에 대한 사전 지식이 없는 20대 초반의 남·녀 9명이 참가하였으며, 실험 전에 설문지 작성 요령만을 설명하고 실험에 임하였다.

그림 7~8은 음상 정위 평가에 대한 결과이다. 각각의 음상 정위 평가 설문지를 모아 통계를 낸 결과 수평각 정위는 헤드폰이 우수하게 나타났지만 거리감을 동시에 제어하지 않으면 음상이 머리 안에 정위되는 현상이 일어났으며 고도각 정위는 전·후 방향보다는 좌·우 방향이 우수하게 나타났다. 그리고 수평각과 고도각 모두 정지음 보다 이동음이 우수하게 나타났다.

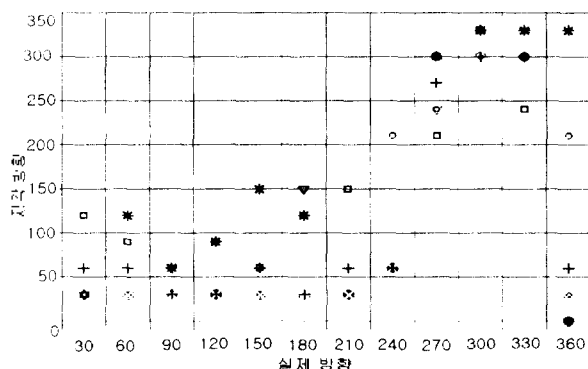


그림 7. 주관 평가 응답
(조건 : 헤드폰-정지음-수평각-스피커)

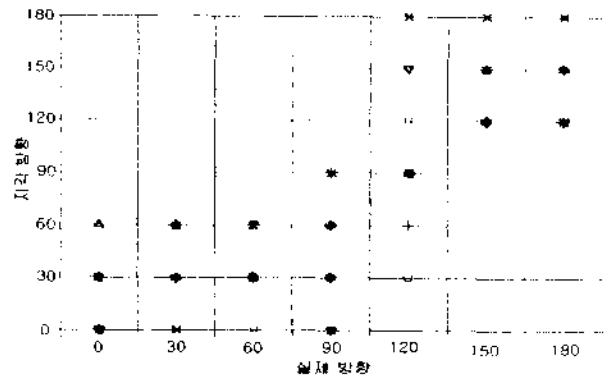


그림 8. 주관 평가 응답
(조건 : 헤드폰-정지음-수직각(좌·우)-헤드폰)

5. 결론

본 논문에서는 범용 DSP를 이용하여 실시간 음상 정위 시스템을 구현하고 주관 평가를 통하여 음상 정위 성능을 고찰하였다. 실시간 음상 정위 시스템은 음원의 방향감, 거리감을 제어할 수 있다.

헤드폰 재생에서 방향감 지각은 우수하지만 전·후면(0°·180°)의 음상 정위는 다른 방향에 비하여 혼동이 많이 일어났으며 거리감을 제어하면 히 음상이 머리내에 정위되지 않는다. 트랜스오럴 핀터를 이용한 스피커 재생은 청취 지점이 한 점에 한정되는 단점은 있지만 공간에 대한 표현이나 거리에 대한 표현 등이 헤드폰에 비해 우수했으며, 방향을 변화시킨 이동음 재생은 정지음에 비해 방향 지각이 우수하게 나타났다.

참고문헌

1. 강성훈, 강경옥, 입체음향, 기전연구소, 1997
2. D. Begault, 3-D Sound for Virtual Reality and Multimedia, Academic Press, Boston, MA, 1994
3. William Grant Gardner, The Virtual Acoustic Room, MIT Press, Cambridge, MA, 1982
4. William Grant Gardner, "Transaural 3-D audio", MIT Media Laboratory Perceptual Computing Section Technical Report No. 342, July 20, 1995
5. 이동우 외, "3차원 입체 음향 재생을 위한 실시간 디지털 신호 처리기의 구현", 한국통신학회 학술발표대회 논문집 제17권, pp. 1133~1136