

실감방송을 위한 3차원 오디오 처리 기술

이태진, 강경옥

한국전자통신연구원 방송미디어연구그룹

tjlee@etri.re.kr, kokang@etri.re.kr

3D Audio Processing Technology for Realistic Broadcasting

Taejin Lee, Kyeongok Kang

Electronics and Telecommunications Research Institute

요약

디지털 방송기술이 진보함에 따라 사용자들의 실감방송에 대한 욕구도 증가하고 있다. 실감방송의 목적은 3차원 AV 기술을 통해 사용자가 마치 녹음/녹화한 현장에 있는 것과 같은 느낌을 주는 것이다. 청취자가 녹음한 환경에 있는 것과 같은 느낌을 주는 3차원 오디오 신호의 획득을 위해 더미헤드를 많이 이용한다. 인간의 머리형태를 띤 더미헤드의 특성 때문에 더미헤드를 이용하여 획득한 신호를 헤드폰을 통해 재생하는 경우 입체감을 느낄 수 있다. 하지만 더미헤드의 형태 및 크기 때문에 공공장소에서 사용하기가 어렵고, 획득한 신호를 멀티채널 신호로 변환하기가 힘들며, 재생 시 전/후방 혼동현상이 많이 발생한다. 본 논문에서는 이러한 더미헤드의 단점을 극복하기 위해 머리형태를 구체로 간략화 하고, 구체 위에 다수개의 마이크를 배치함으로써 3차원 오디오를 획득하고, 후처리 과정을 통해 다양한 재생환경에 적절한 재생신호를 생성할 수 있는 3차원 오디오 획득 및 재생 시스템을 제안한다. 제안한 시스템의 성능평가를 위해 무향실에서 주관적 방향성 평가 실험을 수행한 결과, 더미헤드 기술의 단점인 전/후방 혼동현상을 현저하게 줄일 수 있었다. 본 논문에서 제안한 3차원 오디오 시스템은 3D-TV나 실감방송 등에서 입체음향 콘텐츠 획득에 이용 가능하다.

I. 서론

실감방송은 사용자가 마치 녹음/녹화한 현장에 있는 것과 같은 느낌을 주기 위한 것으로, 이를 위해 3차원 AV기술이 필요하다. 본 논문은 이러한 실감방송 서비스를 위한 3차원 오디오 기술을 제안한다. 일반적으로 3차원 오디오 신호의 획득을 위해 더미헤드를 사용한다. 인간의 머리형태를 한 더미헤드의 특징 때문에 획득한 신호를 헤드폰을 통해 재생하는 경우 입체감을 느낄 수 있다 [1]. 더미헤드를 통해 획득한 바이노럴 신호를 스피커를 통해 재생하기 위해서는 역필터를 이용하여 크로스토크를 제거해 주어야 하는데 이러한 시스템을 트랜스오럴(Transaural) 시스템이라 한다 [2]. 더미헤드를 이용하면 풍부한 입체감을 얻을 수 있지만, 후방 음상의 상승, 음색의 변화 및 전/후방 혼동현상 등의 단점이 있다. 이러한 단점 중 가장 보편적인 단점은 스피커 재생인 경우 후방의 음상을 전방의 음상으로 착각하고 헤드폰 재생인 경우 전방의 음상을 후방의 음상으로 착각하는 전/후방 혼동현상이다. 이러한 전/후방 혼동현상의 원인은 다양한 인간의 머리형태를 더미헤드가 정확하게 묘사하지 못하고, 인간의 머리움직임을 처리할 수 없기 때문이다. 인간이 음원의 방향을 판단하기 위해서는 두 귀에 도달하는 음원의 시간차이(ITD Inter-aural Time Difference)와 세기 차이(ILD Inter-aural Level Difference)를 이용한다. 하지만 양 귀를 중심으로 전/

후 같은 각도/거리를 가지는 위치의 음원은 ITD와 ILD가 같기 때문에 전/후 방향을 판단하기가 어렵다. 이 경우 머리를 좌/우로 움직임으로써, ITD와 ILD를 변경해 가면서 전/후 방향을 판단한다 [3]. 하지만 더미헤드는 고정된 형태로 되어 있기 때문에 이러한 인간의 머리움직임 정보를 이용할 수 없어서, 전/후방 혼동현상이 발생한다. 이러한 더미헤드의 단점을 해결하기 위해 본 논문에서는 구체 위 수평면에 마이크를 배치하여 입체음을 획득하고 후처리를 통해 헤드폰, 스테레오, 스테레오 다이폴, 4채널, 5채널 재생환경 등 다양한 재생환경에서 재생 가능한 신호를 생성할 수 있는 3차원 오디오 시스템을 제안한다. 구체위의 마이크는 전면의 음상을 강조하기 위한 중앙 마이크와 인간의 머리움직임을 이용하기 위한 측면 마이크로 이루어진다 [4]. 후처리는 멀티채널 3차원 마이크 출력신호와 다양한 역필터의 콘볼루션 연산을 실시간으로 처리하기 위해 H/W로 구현하였다. 시스템의 성능평가를 위해 주관적 방향성 평가실험을 수행하였다. 실험결과 5채널과 4채널 재생환경에서 전/후방 혼동현상을 현저하게 줄일 수 있었다.

본문의 구성은 다음과 같다. 먼저, II장에서는 본 논문에서 제안하는 3차원 오디오 시스템의 전체 구성 및 구현방법에 관하여 기술한다. 다음 III장에서는 3차원 오디오 시스템을 이용한 주관적 방향성 평가 실험 방법 및 결과를 기술하고, 마지막으로 IV장에서 결론을 맺는다.

표. 3차원 오디오 획득 및 재생 시스템

본 논문에서 제안하는 3차원 오디오 시스템의 구성은 그림 1과 같다. 멀티채널 3차원 마이크는 입체음원의 획득을 위해 구체 위 수평면에 다섯 개의 마이크를 배치하여 제작하였고, 후처리 H/W는 멀티채널 3차원 마이크를 통해 획득한 신호를 다양한 재생환경에서 재생하기 위한 역필터링 연산을 실시간 처리하기 위해 DSP H/W를 이용하여 제작하였다.

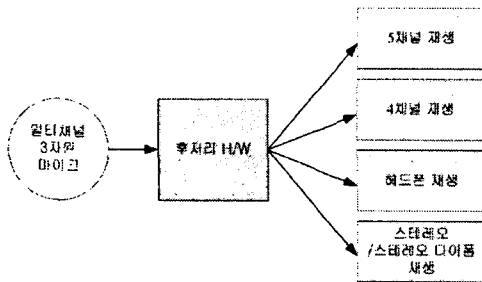


그림 1. 전체 시스템 구성도

1. 멀티채널 3차원 마이크 제작

그림 2는 구체를 이용하여 입체음향을 획득하기 위한 멀티채널 3차원 마이크의 형상이다. 정면의 음상을 증가시키기 위한 중앙 마이크는 구체의 수평면 상의 정중앙에 위치하고, 인간의 머리움직임 특성을 이용하기 위한 측면 마이크는 좌우측에 30도의 간격을 두고 배치하였다.

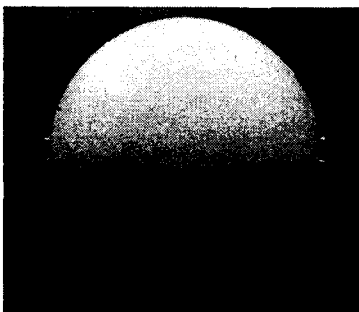


그림 2. 구체 마이크로폰 형상

2. 다양한 재생신호 생성을 위한 후처리

후처리는 멀티채널 3차원 마이크의 출력신호와 역필터 사이의 콘볼루션을 통해 다양한 재생환경에 적절한 신호를 생성하기 위한 것이다. 다양한 재생환경에 따른 크로스토크 제거를 위한 역필터의 생성을 위해 구체의 임펄스 응답을 측정하였다. 구체의 임펄스 응답은 무향실에서 구체를 수평 방향으로 5도씩 회전해 가면서 총 72개를 측정하였다. 임펄스 응답의 측정에는 임펄스 응답측정을 위한 S/W가 사용되었고, 스피커와 마이크의 특성을 보상하기 위해 스피커와 마이크 사이의 임펄스 응답을 측정하여 구체의 임펄스 응답 데이터에서 제거하였다. 그림 3은 측정된 구체의 임펄스 응답의 예이다. 그림에서 알 수 있듯이 전/후 동일한 각도에서 측정된 구체의 임펄스 응답은 동일하므로, 구체의 좌/우에 배치한 마

이크를 이용하여 입체음원을 획득하는 시스템을 한 단계 업그레이드하여 전/후방 혼동현상이 더 심하게 발생한다.

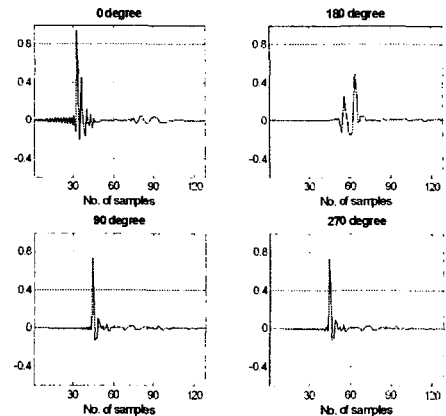


그림 3. 구체의 임펄스 응답

3차원 오디오 재생을 위한 역필터링은 크로스토크를 제거하기 위한 것이다. 바이노럴 신호를 스피커를 통해 재생하는 경우, 좌측 스피커 신호가 우측 귀에 들리고, 우측 스피커 신호가 좌측 귀에 들리는 크로스토크 현상이 발생한다. 이러한 크로스토크 현상을 제거하기 위해 다양한 방법을 사용하는데, 본 논문에서는 Fast Deconvolution 방법을 사용한다 [5]. 멀티채널 디콘볼루션의 목적은 주어진 S 개의 스피커를 이용하여 공간상의 R 위치에서 음장을 가능한 정확하게 재생하기 위한 것이다. 크로스토크 제거를 위한 최적의 역필터 $H(z)$ 를 구하기 위해 Fast Deconvolution 방법은 수식 (1)을 이용한다. 수식에서 $C(k)$ 는 음장의 특성을 의미하고 β 는 역필터의 특성을 조절하기 위한 파라미터이며, H 는 Hermitian matrix를 의미한다 [6].

$$H_f(k) = [C^H(k)C(k) + \beta I]^{-1}C^H(k) \quad (1)$$

5채널 재생환경을 위한 역필터는 멀티채널 3차원 마이크의 출력신호를 그림 4의 좌측과 같이 재생환경의 중앙에 위치한 구체의 5개의 위치에서 가능한 정확하게 재생하기 위한 것이다. 4채널 재생환경을 위한 역필터는 5채널 재생환경 중 중앙채널을 제외한 4개의 스피커를 이용하여 구체 위의 4개의 위치에 원하는 신호를 재생하기 위한 것이고, 2채널 재생환경을 위한 역필터는 그림 4의 우측과 같이 좌/우 스피커 출력신호를 재생환경의 중앙에 위치한 구체의 좌/우 90도 위치에 가능한 정확하게 재생하기 위한 것이다.

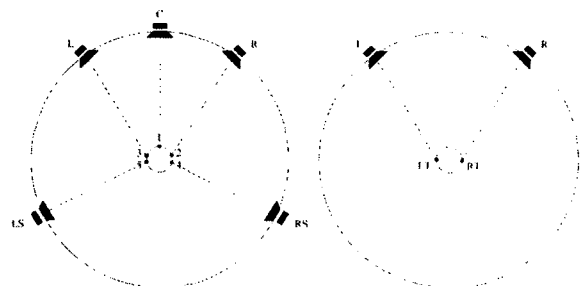


그림 4. 역필터링을 위한 스피커 배치 및 Target

후처리 과정은 5채널, 4채널, 스테레오/스테레오 다이폴 재생환경에서의 크로스트크 제거를 위한 역필터와 다양한 입력 신호의 콘볼루션 연산을 통해 각각의 재생환경에 적합한 재생신호를 생성한다. 5채널 재생은 ITU의 5.1채널 재생환경에서 LFE를 위한 subwoofer를 제외한 5개의 스피커를 통해 입체음향을 재생하기 위한 것으로 멀티채널 3차원 마이크의 출력신호와 5×5 역필터의 콘볼루션을 통해 5채널 재생신호를 생성한다.

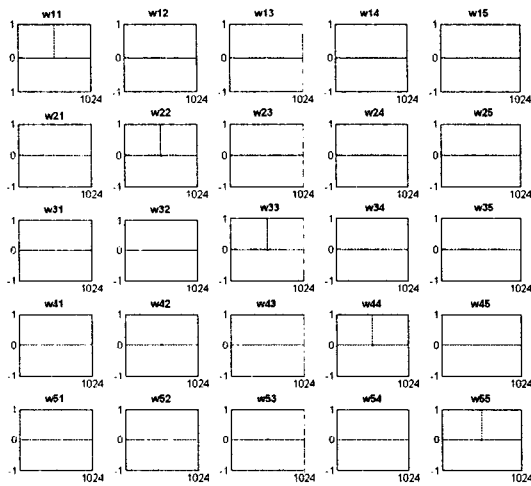


그림 5. 5채널 재생을 위한 역필터의 시간영역 특성

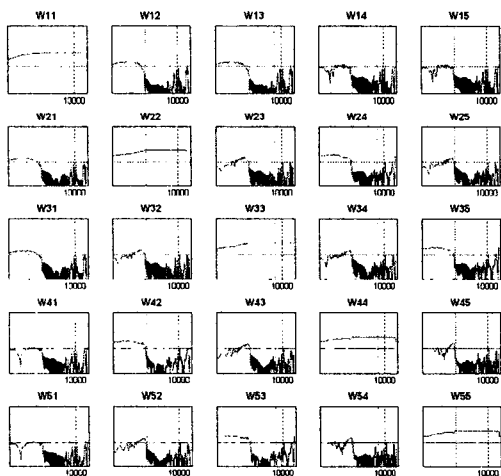


그림 6. 5채널 재생을 위한 역필터의 주파수영역 특성

그림 5, 6은 5채널 재생을 위한 역필터의 특성을 보여준다. 그림 5는 5채널 재생을 위한 역필터와 5채널 재생환경의 음장특성, 즉 $C(k)$ 를 콘볼루션한 결과이다. 그림에서 알 수 있듯이 대각선 성분, 즉 각각의 스피커와 구체 위의 특정 위치에서만 신호가 관찰되고 나머지 성분, 즉 크로스트크는 없음을 알 수 있다. 그림 6은 그림 5의 결과를 주파수 영역에서 분석한 것이다. 그림에서 대각선 성분, 즉 재생할 신호는 주파수 특성이 평탄하고, 나머지 크로스트크 성분은 신호가 많이 감쇠함을 알 수 있다. 따라서 본 시스템에 사용하는 역필터가 효율적으로 크로스트크를 제거하고 원하는 위치에서 원하는 스피커의 신호만을 재생할 수 있음을 알 수 있다.

4채널 재생환경과 스테레오/스테레오 다이폴 재생환경을 위한 재생신호 역시 각각 4×4 , 2×2 역필터를 이용하여 생성하였고, 이들 역필터 역시 5채널 재생을 위한 역필터와 같이 크로스트크 성분을 효율적으로 제거하도록 설계하였다.

헤드폰 재생신호는 멀티채널 3차원 마이크의 출력신호와 5×5 역필터의 콘볼루션을 이용하여 생성한 5채널 재생환경을 위한 5채널 재생신호와 재생환경의 중앙에 위치한 청취자의 두 귀와 각각의 스피커 사이의 임펄스 응답을 이용하여 생성한다. 그림 7은 헤드폰 재생신호의 생성을 위해 필요한 스피커와 청취자의 두 귀 사이의 임펄스 응답의 종류를 보여준다. 예를 들어 중앙의 채널에서 재생하기 위한 신호와 중앙 스피커와 좌/우측 귀까지의 임펄스 응답을 이용하여 좌/우측 스피커에서 재생하기 위한 신호를 생성한다.

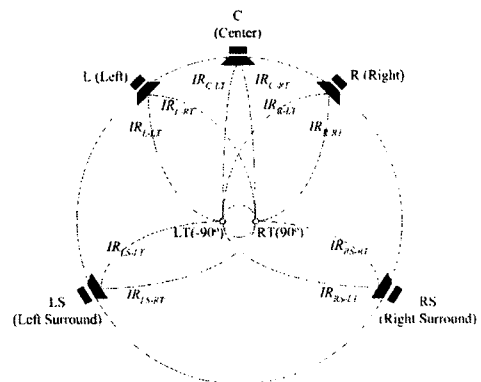


그림 7. 헤드폰 재생신호의 생성

3. 후처리 H/W 구현

다양한 재생환경에 따른 적절한 재생신호의 생성을 위한 후처리는 멀티채널 3차원 마이크의 출력신호와 다양한 역필터 사이의 콘볼루션 연산으로 이루어진다. 이러한 콘볼루션 연산은 많은 계산량을 요구한다. 특히 5채널 재생신호의 생성을 위해 멀티채널 3차원 마이크의 출력인 5채널 신호와 5×5 역필터 사이에 콘볼루션을 수행해야 하기 때문에 실시간 처리를 위해 DSP를 이용하여 후처리 H/W를 제작하였다. 후처리 H/W는 멀티채널 3차원 마이크의 5채널 출력신호를 입력받아 두 개의 DSP에서 각각의 채널을 사용자가 전면의 기능선택 버튼으로 선택한 동작에 따른 출력신호의 생성을 위한 입력신호와 역필터 사이의 콘볼루션 연산을 수행한 후, 출력신호를 전송한다. 후처리 H/W 전면부의 기능선택 버튼은 5채널, 4채널, 스테레오, 스테레오 다이폴 및 헤드폰 재생을 선택할 수 있다.

III. 주관적 방향성 평가 실험

본 논문에서 제안한 3차원 오디오 획득 및 재생시스템의 성능을 평가하기 위해 무향실에서 주관적 방향성 평가실험을 수행하였다. 방향성 평가 실험은 일본 농경전기 대학(Tokyo Denki University)의 무향실에서 청력이 정상인 10명의 학생을 대상으로 수행하였다. 실험에 사용한 콘텐츠는 모노 음

원과 각각의 방향에 따른 구체의 임펄스 응답의 콘볼루션 연산을 이용하여 제작하였다. 콘텐츠는 남성의 목소리와 클래식 음악으로 구성하였다. 실험에 사용한 콘텐츠의 길이는 30초이고, 방향은 전면 0도에서 후면 180도까지 15도의 간격을 가지고 있다. 멀티채널 방향성 실험에서 피험자는 그림 8과 같이 무향실에 배치한 멀티채널 스피커 재생환경의 중앙에 위치하고 임의의 순서로 재생한 각각의 콘텐츠에 대해 피험자가 느끼는 음원의 방향을 지적하는 방법으로 수행하였다. 주관적 방향성 평가 실험 시 실험자는 자신의 머리를 자유로이 움직여 음원의 방향을 판단하였다.

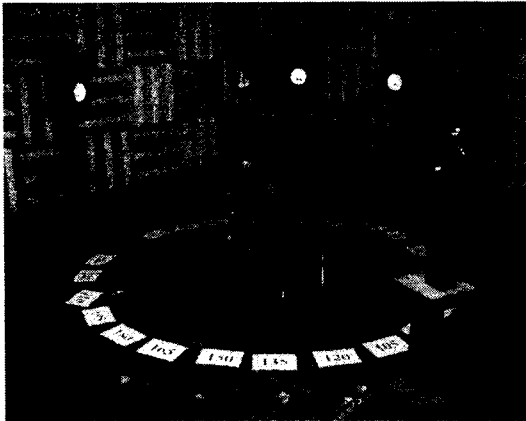


그림 8. 방향성 평가 실험을 위한 스피커 배치

5채널 재생실험을 위해 ITU 5.1 스피커 배치 중 LFE를 위한 subwoofer를 제외한 5개의 스피커를 무향실에 그림 8과 같이 설치하고 실험을 수행하였다. 4채널 재생실험에서는 5채널 재생환경 중 중앙의 스피커를 제외한 4개의 스피커를 이용하였다. 그림 9는 5채널과 4채널 주관적 방향성 평가 실험의 결과를 보여준다. 그림 9에서 수평축은 실제 콘텐츠의 방향을 의미하고 수직축은 피험자가 느낀 방향을 의미한다. 5채널과 4채널에 대한 방향성 실험결과 전/후방 혼동현상이 현저하게 줄어들었음을 알 수 있다.

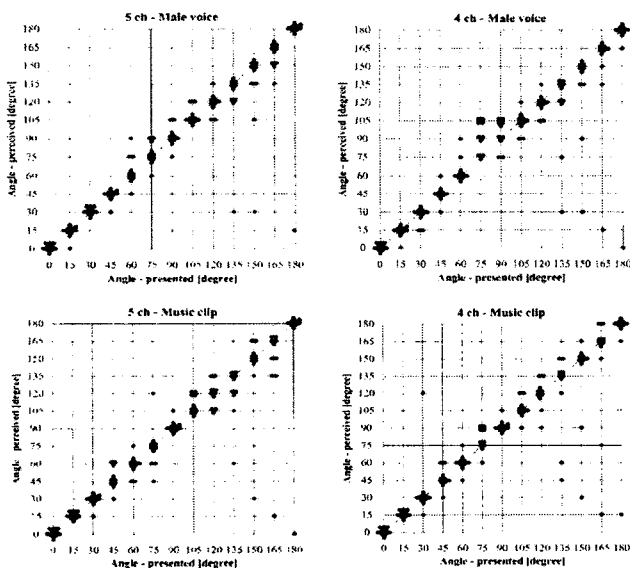


그림 9. 멀티채널 방향성 실험 결과

본 논문에서는 구체 위에 배치한 다수 개의 마이크를 이용하여 제작한 멀티채널 3차원 마이크를 이용하여 3차원 오디오 신호를 획득하고 후처리 H/W를 통해 다양한 재생환경에서 재생 가능한 3차원 오디오 시스템에 대해 기술하였다. 멀티채널 3차원 마이크는 중앙의 음원을 강조하고 인간의 머리움직임을 이용하기 위한 위치에 마이크를 배치하여 제작하였다. 후처리 과정은 다양한 재생환경에서의 스피커 배치에 따른 크로스토르크를 제거하는 과정으로써, 재생환경에 적합한 역필터를 이용하여 각각의 크로스토르크를 제거한다. 멀티채널 3차원 마이크의 성능을 검증하기 위한 주관적 방향성 평가 실험 결과 5채널과 4채널 등 멀티채널 재생환경에서 기존 더미헤드 기술의 문제점인 전/후방 혼동현상을 현저하게 줄일 수 있었다. 본 논문에서 제안한 3차원 오디오 시스템은 향후 실감방송이나 3DTV등에서 사용할 입체음향의 획득에 이용 가능할 것이다.

감사의 글

본 논문은 정보통신부의 지능형통합정보방송(SmarTV) 기술개발 과제를 통해 수행한 한국전자통신연구원과 일본동경전기대학(Tokyo Denki University)의 공동연구 결과물입니다.

참고 문헌

1. H. Moller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, 1992.
2. D. H. Cooper and J. Bauck, "Prospects for transaural recording," *J. Audio Eng. Soc.* 37 (1/2), 3-19, 1989.
3. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT, Cambridge, MA, 1983.
4. Yuvi Kahana, Philip A. Nelson, Ole Kirkeby, and Hareo Hamada, "A multiple microphone recording technique for the generation of virtual acoustic images," *J. Acoustic. Soc. Am.*, 105 (3), 1503-1516, 1999.
5. Ole Kirkeby, Philip A. Nelson, Hareo Hamada, and Felipe Orduna-Bustamante, "Fast Deconvolution of Multichannel Systems Using Regularization," *IEEE Trans. On Speech and Audio Processing*, 6 (2), 189-194, 1998.
6. Hareo Hamada, Hironori Tokuno, Yuko Watanabe and P. A. Nelson, "3D Sound Generation using Two Loudspeakers Stereo Dipole System and Its Applications," *AES 15th International Conference*, 1998.