

멀티채널 3차원 마이크를 이용한 입체음향 처리 기술

Spatial Audio Signal Processing Technology Using Multi-Channel 3D Microphone

이 태 진*, 강 경 옥*
(Taejin Lee*, Kyeongok Kang*)

*한국전자통신연구원 방송미디어연구그룹
(접수일자: 2004년 11월 1일; 채택일자: 2005년 2월 4일)

입체음향 시스템의 목적은 청취자에게 음원을 획득한 장소에 있는 것과 같은 느낌을 주는 것이다. 이를 위해 일반적으로 더미헤드가 많이 사용되고 있다. 인간의 머리형태를 한 더미헤드의 특성 때문에 더미헤드를 통해 획득한 음원을 헤드폰을 통해 청취하는 경우 현장감을 느낄 수 있다. 하지만 더미헤드의 형태 및 크기는 공공장소에서 사용하기에는 제약이 있고 더미헤드를 통해 획득한 신호는 멀티채널로 확장하기가 어렵기 때문에 본 논문에서는 이러한 더미헤드를 대체로 간략화 한 후 구체 위에 다수 개의 마이크를 배치하여 입체음원을 획득하기 위한 멀티채널 3차원 마이크 기술에 대해 제안한다. 본 논문에서 제안하는 멀티채널 3차원 마이크는 구체 위의 수평면 상에 5개의 마이크를 배치하여 입체음원을 획득한 다음 후처리 과정을 통해 헤드폰, 스테레오, 스테레오 다이폴, 4채널 및 5채널 재생환경 등에서 재생이 가능하다. 다양한 재생신호의 생성을 위한 후처리 과정은 많은 연산량을 필요로 하기 때문에, H/W로 제작하였다. 멀티채널 3차원 마이크의 성능을 검증하기 위해 방향성 실험을 수행한 결과, 멀티채널 재생환경에서 더미헤드 기술의 단점인 전/후방 혼동현상을 현저하게 줄일 수 있었다.

핵심용어: 3차원 오디오, 더미헤드, 멀티채널 3차원 마이크, 크로스토크 제거

투고분야: 전기음향 분야 (3.5)

The purpose of a spatial audio system is to give a listener an impression as if he were present in a recorded environment when its sound is reproduced. For this purpose a dummy head microphone is generally used. Because of its human-like shape, dummy head microphone can reproduce spatial images through headphone reproduction. However, its shape and size are restriction to public use and it is difficult to convert the output signal of dummy head microphone into a multi-channel signal for multi-channel environment. So, in this paper, we propose a multi-channel 3D microphone technology. The multi-channel 3D microphone acquire a spatial audio using five microphones around a horizontal plane of a rigid sphere and through post processing, it can reproduce various reproduction signals for headphone, stereo, stereo dipole, 4ch and 5ch reproduction environments. Because of complex computation, we implemented H/W based post processing system. To verify the performance of the multi-channel 3D microphone, localization experiments were performed. The result shows that a front/back confusion, which is the one of common limitations of conventional dummy head technology, can be reduced dramatically.

Keywords: 3D Audio, Dummy-head, Multi-channel 3D Microphone, Crosstalk Cancellation

ASK subject classification: Electro-Acoustics (3.5)

I. 서론

이상적인 입체음향 시스템은 청취자가 녹음한 환경에 있는 것과 같은 느낌을 주는 것이다. 실제 인간의 귀에

probe 마이크를 장착하여 바이노럴 녹음이 가능하지만, 고음질의 마이크를 인간의 귀 속에 장착하는 것이 쉽지 않고 녹음하는 중간에 인간에 의해 발생하는 잡음이나 머리 움직임 등 때문에 자주 이용되지 않고 있다. 이러한 단점 때문에 인간의 머리형태를 한 더미헤드의 귀에 소형 마이크를 장착하여 입체음원을 획득하는 바이노럴 녹음기술은 음장의 평가나 가상현실 시스템 등에서 많이

이용되고 있다[1-3]. 더미헤드를 통해 획득한 신호는 일반적으로 헤드폰을 통해 재생하는 경우 입체감을 느낄 수 있고, 스피커를 통해 재생하기 위해서는 역필터를 이용하여 크로스토크를 제거해 주어야 하는데 이러한 시스템을 트랜스오럴 (Transaural) 시스템이라 한다[4]. 더미헤드를 이용하면 풍부한 입체감을 얻을 수 있지만, 후방 음상의 상승, 음색의 변화 및 전/후방 혼동 등의 단점이 있다. 이러한 단점 중 가장 보편적인 단점은 헤드폰 재생 시 전방의 음원을 후방의 음원으로 착각하고, 스피커 재생 시 후방의 음원을 전방의 음원으로 착각하는 전/후방 혼동현상이다. 이러한 전/후방 혼동현상의 원인은 다양한 인간의 머리형태를 더미헤드가 정확하게 묘사하지 못하고, 인간의 머리움직임을 고려할 수 없기 때문이다. 인간은 약간의 머리 움직임을 통해 신호를 변경하여 전/후 방향을 판단하는데, 더미헤드는 고정된 형태로 되어 있기 때문에 전/후방 혼동현상이 발생한다[5].

인간의 머리형태를 구체로 간략화 하여 입체음을 획득하는 기술은 머리의 shadow 효과 등을 묘사할 수 있고, 더미헤드를 통해 녹음한 신호를 스피커를 통해 재생할 때 발생하는 pinna filtering 효과를 제거할 수 있다. 일반적으로 더미헤드를 통해 녹음한 바이노럴 신호는 녹음할 때와 스피커를 통해 재생할 때 두 번의 pinna filtering 효과가 발생하기 때문에 음색의 왜곡현상이 발생한다[1].

한편, 구체 위의 양 끝에 두 개의 마이크를 장착하여 스테레오 재생 시 입체감을 향상시키는 제품이 시판되고 있다[6, 7]. 하지만 구체의 대칭적인 특징 때문에 이 경우 전/후방 혼동현상이 더 발생한다. 이를 극복하기 위해 구체 위에 4개의 마이크를 배치하여 전/후방 혼동현상을 줄이기 위한 연구가 발표되었다[5]. 이 시스템은 전/후방 혼동현상을 줄일 수 있었지만 단지 4채널 재생환경만을 고려하였다.

본 논문은 전/후방 혼동현상을 극복할 수 있고, 다양한 재생환경에서 적용할 수 있는 입체음향 획득 및 재생 시스템을 제안한다. 본 시스템은 구체 위 수평면 상에 다섯 개의 마이크를 배치하여 입체음을 획득하고 이 신호를 후처리 과정을 통해 헤드폰, 스테레오, 스테레오 다이폴, 4채널, 5채널 재생환경 등 다양한 재생환경에서 재생 가능한 신호를 생성한다. 후처리 과정에서는 다양한 재생환경에서 크로스토크 제거를 위해 Fast Deconvolution 방법을 이용하여 역필터를 생성한다[8]. 이러한 역필터들은 재생환경의 특정 위치에서 원하는 신

호가 재생될 수 있도록 설계하였다. 시스템의 성능평가를 위해 주관적 방향성 평가실험을 수행하였다. 실험결과 5채널과 4채널 재생환경에서 전/후방 혼동현상을 현저하게 줄일 수 있었고, 헤드폰 재생인 경우 더미헤드와 비슷한 결과를 보였다.

본문의 구성은 다음과 같다. 먼저, II장에서는 구체 위에 다수개의 마이크를 배치하여 입체음을 획득하는 방법을 기술하고, III장에서는 후처리 H/W에 관하여 기술한다. IV에서는 주관적 방향성 실험과정과 결과를 기술하고 마지막으로 V장에서 결론을 맺는다.

II. 멀티채널 3차원 마이크로폰

2.1. 구체의 임펄스 응답 측정

후처리 과정에서 재생환경에 따른 다양한 역필터를 구하기 위해 구체의 임펄스 응답을 측정하였다. 임펄스 응답은 무향실에서 (크기: 6.2m×5.4m×4.95m (W×D×H), 압소음: 19dB(A) 이상, 역제품 특성: JIS Z 8732 및 ISO 3745 준거) 그림 1과 같이 1.2m 높이에 구체를 설치한 후, 스피커와 1.4m 거리를 두고 구체를 수평면에서 5도씩 회전하면서 총 72개의 임펄스 응답을 측정하였다.

구체 위에 설치한 마이크와 스피커의 특성을 제거하기 위해 자유음장 임펄스 응답을 측정한 후 역필터를 구해서 측정된 구체의 임펄스 응답에서 제거하였다. 그림 2는 측정된 임펄스 응답의 예이다. 구체의 특성상 전/후방 같은 방향에서 입사한 임펄스 응답은 동일하므로 구체의 좌/우 90도 위치에 마이크를 설치하여 획득한 음원은 전/후방 혼동현상이 더미헤드보다 더 발생한다. 따라

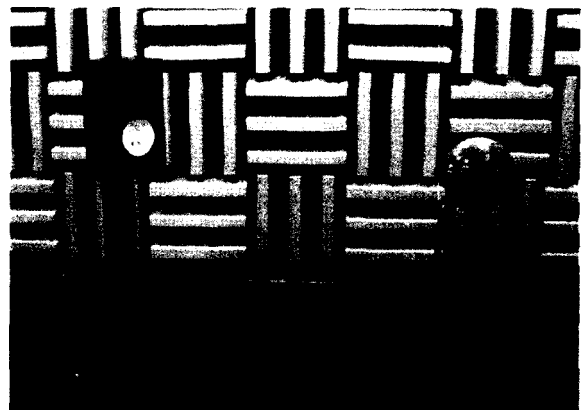


그림 1. 무향실에서 구체의 특성 측정

Fig. 1. The feature measurement of a rigid sphere.

서 본 논문에서 제안하는 시스템은 구체의 좌/우 90도에 마이크를 위치시키지 않고, 특정한 위치에 다수개의 마이크를 위치시킴으로써, 전/후방 혼동현상을 감소시킨다.

2.2. 구체 위의 마이크 배치

그림 3은 구체 위에 5개의 마이크를 배치하여 제작한 멀티채널 3차원 마이크로폰이다. 구체 위의 5개의 마이크 중, 중앙 마이크는 구체의 수평면 상의 정 중앙에 위치하고, 나머지 마이크는 좌우에 30도의 간격을 두고 배치하였다. 중앙 마이크는 정면 음상을 향상시키기 위한 것이고 나머지 마이크는 머리 움직임을 고려하기 위한 것이다. 인간이 음원의 방향을 지각하기 위해서는 머리를 좌/우로 회전시켜서 음원의 ITD (Interaural Time Difference)와 ILD (Interaural Level Difference)를 변화시킨다[9,10]. 머리의 정 중앙면 (median plane)에서 음이 입사하는 경우 머리를 움직이지 않으면 ITD와 ILD가 동일하기 때문에 방향을 판단하기가 어려워진다. 일반적으로 인간이 방향을 판단하기 위해 회전하는 머리의 각도가 20~40도 이기 때문에 본 논문에서 제안하는 멀티채널 3차원 마이크 시스템은 구체 위에 30도의 각도를 두고 좌/우에 두 개의 마이크를 위치시켜서 입체음원을 획득한다.

2.3. 다양한 재생환경을 위한 후처리

2.3.1. Fast Deconvolution

더미헤드를 통해 녹음한 바이노럴 신호를 스피커를 통해 재생하는 경우 좌측 스피커 신호가 우측 귀에 들리고, 우측 스피커 신호가 좌측 귀에 들리는 크로스토크 현상

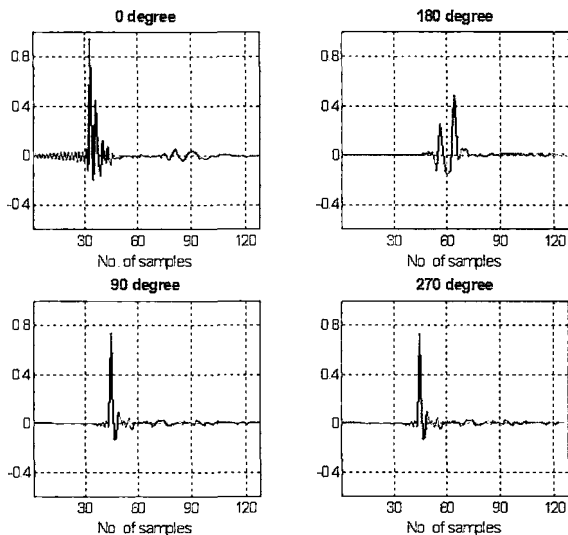


그림 2. 구체 임펄스 응답의 예
Fig. 2. Example of a rigid sphere's impulse responses.

이 발생한다. 이러한 크로스토크 현상을 제거하기 위해 다양한 방법을 사용하는데 [8,11-13]. 본 논문에서는 Fast Deconvolution 방법을 사용한다[8]. 멀티채널 디콘볼루션의 목적은 주어진 S 개의 스피커를 이용하여 공간상의 R 위치에서 음장을 가능한 정확하게 재생하는 것이다. 본 논문에서는 5×5 , 4×4 와 2×2 역필터를 이용하여 5채널, 4채널 및 스테레오/스테레오 다이플 [14,15] 재생환경에서 크로스토크를 제거한다.

그림 4는 멀티채널 재생환경을 위한 Fast Deconvolution 방법을 설명한다. Fast Deconvolution 방법의 목적은 주어진 $C(z)$ 음장환경의 R 위치에서 S 개의 입력신호 $u(z)$ 를 가능한 정확하게 재생하기 위하여 $C(z)$ 의 영향을 보상하는 역필터 $H(z)$ 를 구하는 것이다. 그림 4에서 $v(z)$ 는 입력신호 $u(z)$ 와 역필터 $H(z)$ 를 이용하여 생성한 스피커 출력신호를 의미하고, $w(z)$ 는 R 위치에서 실제 재생되는 T 개의 신호를 의미한다. $d(z)$ 는 R 위치에서의 이상적인 재생신호이고 $e(z)$ 는 $d(z)$ 와 $w(z)$ 의 차이를 의미한다.

최적의 역필터 $H(z)$ 를 구하기 위해 Fast Deconvolu-

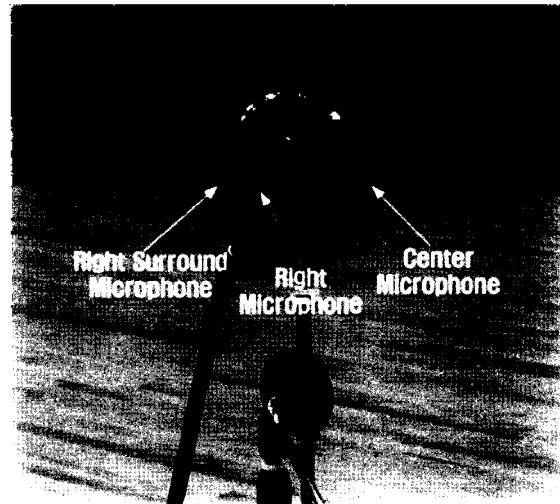


그림 3. 멀티채널 3차원 마이크로폰
Fig. 3. The multi-channel 3D microphone.

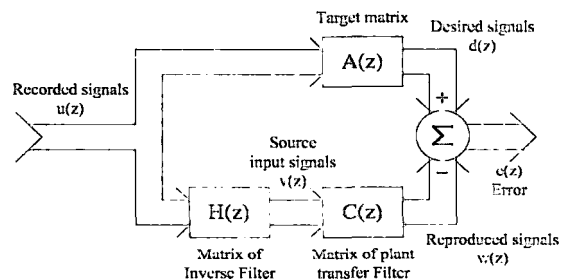


그림 4. 크로스토크 제거를 위한 Fast Deconvolution 방법
Fig. 4. Fast Deconvolution Algorithm for crosstalk cancellation.

tion 방법은 그림 4의 각각의 벡터를 이용하여 유도한 수식 (1)을 이용한다[8].

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

수식 (1)에서 $C(k)$ 는 음장의 임펄스 응답신호 $c_{RS}(n)$ 을 $R \times S$ FFT를 취한 값이다. $C(k)$ 와 $C(k)$ 의 Hermition 매트릭스 $C^H(k)$ 값을 이용하여 $R \times S$ 매트릭스 형태의 $H_I(k)$ 를 구한다. 본 논문에서는 $S = R = T = 5, 4, 2$ 인 경우의 역필터를 이용하여 5채널, 4채널, 스테레오/스테레오 다이폴 재생을 위한 신호를 생성한다. Fast Deconvolution 방법의 특징은 역필터를 구할 때 수식 (1)의 β 를 이용하여 역필터의 특성을 조절할 수 있다는 점이다.

5채널 재생환경을 위한 역필터는 멀티채널 3차원 마이크의 출력신호 u_{1-5} 를 재생환경의 중앙에 위치한 구체 위의 5개의 위치에서 가능한 정확하게 재생하기 위한 것이다.

4채널 재생환경을 위한 역필터는 멀티채널 3차원 마이크의 출력신호 중 u_1 을 제외한 u_{2-5} 신호를 재생환경의 중앙에 위치한 구체 위의 4개의 위치에서 가능한 정확하게 재생하기 위한 것이다. 스테레오/스테레오 다이폴 재생을 위한 역필터는 멀티채널 3차원 마이크의 출력신호를 이용하여 생성한 헤드폰 재생신호를 재생환경의 중앙에 위치한 구체 수평면상의 좌/우 90도 위치에 가능한 정확하게 재생하기 위한 것이다.

멀티채널 3차원 마이크의 후처리 과정에서 사용한 다양한 재생환경을 위한 역필터의 특성은 다음절에서 기술한다.

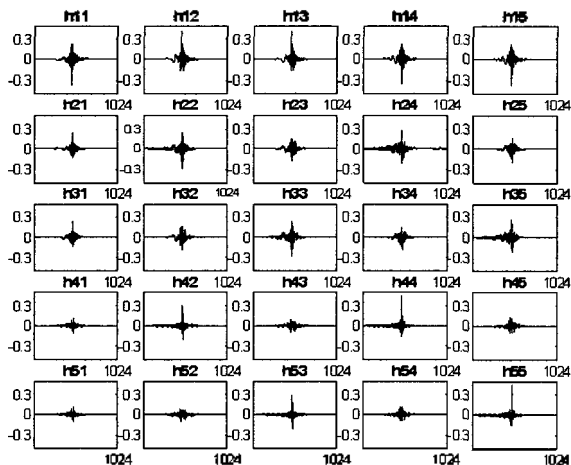


그림 5. 5채널 재생을 위한 역필터
Fig. 5. Inverse filters for 5ch reproduction environments.

2.3.2. 재생환경에 따른 역필터의 구성

후처리 과정에서는 5채널, 4채널, 스테레오/스테레오 다이폴 재생환경에서의 크로스토크 제거를 위한 역필터와 다양한 입력신호의 콘볼루션 연산을 통해 각각의 재생환경에 적합한 재생신호를 생성한다. 또한 멀티채널 3차원 마이크 출력신호와 구체 임펄스 응답의 변환필터를 이용하여 헤드폰 재생을 위한 신호를 생성한다.

가. 5채널 재생을 위한 역필터

5채널 재생은 ITU의 5.1채널 재생환경에서[16] LFE를 위한 subwoofer를 제외한 5개의 스피커를 통해 입체음향을 재생하기 위한 것으로 멀티채널 3차원 마이크의 5채널 출력신호 u_{1-5} 와 5×5 역필터의 콘볼루션을 통해 5채널 재생신호 $v_C^{5ch}, v_L^{5ch}, v_R^{5ch}, v_{LS}^{5ch}, v_{RS}^{5ch}$ 를 생성한다. 그림 5는 5채널 재생신호의 생성을 위한 5×5 역필터를 나타내고, 그림 6은 5×5 역필터의 주파수 특성을 보여준다.

나. 4채널 재생을 위한 역필터

4채널 재생은 ITU의 5.1채널 재생환경에서 중앙 스피커와 LFE를 위한 subwoofer를 제외한 4개의 스피커를 통해 입체음향을 재생하기 위한 것이다. 4채널 재생을 위해 멀티채널 3차원 마이크의 출력신호 중 중앙마이크의 출력신호 u_1 을 제외한, u_{2-5} 와 4×4 역필터의 콘볼루션을 통해 4채널 재생신호 $v_L^{4ch}, v_R^{4ch}, v_{LS}^{4ch}, v_{RS}^{4ch}$ 를 생성한다. 그림 7은 4채널 재생신호 생성을 위한 4×4 역필터를 나타내고 그림 8은 4×4 역필터의 주파수 특성을 보여준다.

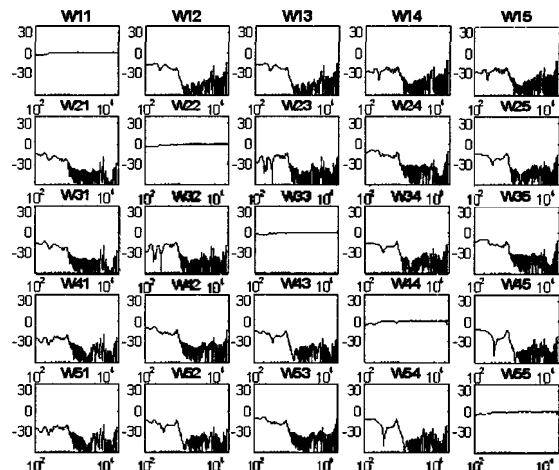


그림 6. 5채널 재생을 위한 역필터의 주파수 특성
Fig. 6. Frequency characteristics of 5ch inverse filters.

다. 스테레오/스테레오 다이폴 재생을 위한 역필터
 스테레오/스테레오 다이폴 재생을 위해서는 먼저 멀티 채널 3차원 마이크의 출력신호를 변환필터를 통해 헤드폰 재생을 위한 2채널 신호로 변환한 후, 이 신호와 2×2역필터의 콘볼루션을 통해 2채널 재생신호를 생성한다. 멀티채널 3차원 마이크의 출력신호와 변환필터를 이용하여 헤드폰 재생을 위한 2채널 신호를 생성하는 방법은 다음 절에서 기술한다. 그림 9는 스테레오 재생을 위한 2×2역필터와 2×2역필터의 주파수 특성을 보여주었고 그림 10은 스테레오 다이폴 재생을 위한 2×2역필터와 2×2역필터의 주파수 특성을 보여준다.

위에서 설명한 5×5, 4×4, 2×2역필터는 모두 재생 환경에 따른 각각의 스피커와 구체 위에 설치한 마이크 사이의 임펄스응답을 이용하여 구하였다. 역필터의 주파수 응답을 보면 대각선 부분 즉 스피커와 마이크의 직접 신호는 평탄한 특성을 보이고 나머지 크로스토크 신호는

효율적으로 제거됨을 알 수 있다. 각각의 그림에서 h_{RS} 는 S 스피커에서 구체 위의 R 위치까지의 역 필터를 의미한다. w_{RS} 는 S 스피커에서 재생한 신호가 구체 위의 R 위치에 도달한 경우의 신호를 의미한다.

2.3.3. 헤드폰 재생신호의 생성

헤드폰 재생을 위한 2채널 신호의 생성을 위해 멀티채널 3차원 마이크의 5채널 출력 신호와 구체 임펄스 응답의 변환필터를 이용한다. 수식 (2)는 구체의 임펄스 응답을 이용하여 헤드폰 신호의 생성을 위한 변환필터를 생성하기 위한 것이다. 수식 (2)에서 SIR_n (Sphere Impulse Response for n degree)은 스피커와 구체 위의 마이크 사이의 각도가 n도인 경우의 구체 임펄스 응답을 나타내고, SIR_0' 는 구체 위의 마이크와 스피커 사이의 각도가 0인 구체 임펄스 응답의 역필터를 의미한

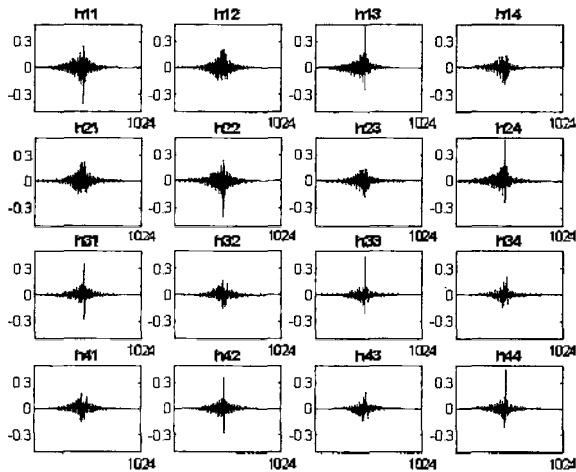


그림 7. 4채널 재생을 위한 역필터
 Fig. 7. Inverse filters for 4ch reproduction environments.

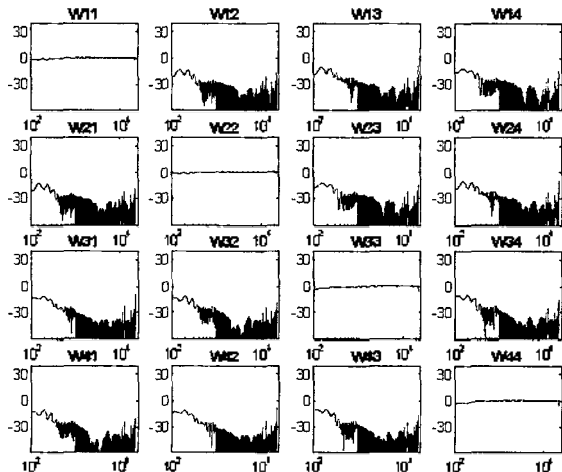


그림 8. 4채널 재생을 위한 역필터의 주파수 특성
 Fig. 8. Frequency characteristics of 4ch inverse filters.

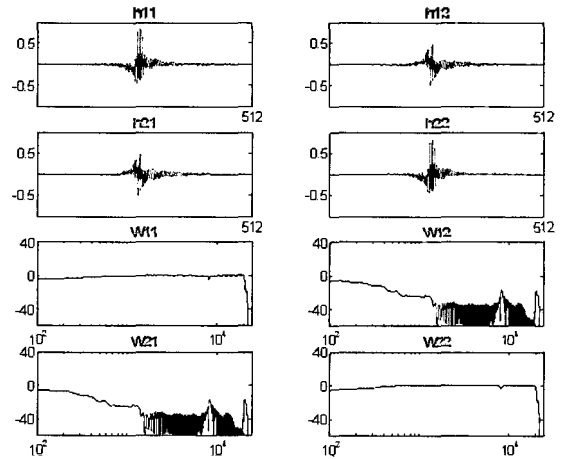


그림 9. 스테레오 재생을 위한 역필터와 역필터의 주파수 특성
 Fig. 9. Inverse filters and its frequency characteristics for stereo reproduction environments.

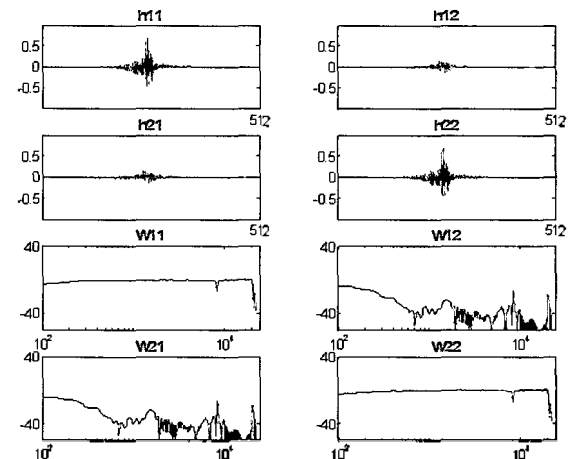


그림 10. 스테레오 다이폴 재생을 위한 역필터와 역필터의 주파수 특성
 Fig. 10. Inverse filters and its frequency characteristics for stereo dipole reproduction environments.

다. 수식 (2)는 SIR_0^{-1} 와 각각의 구체 임펄스 응답과의 콘볼루션 연산을 통해 구체 임펄스 응답의 변환 필터 SCF_{0-355} (Sphere Conversion Filter)를 생성하는 것을 의미한다.

$$SCF_{0-355} = conv(SIR_{0-355}, SIR_0^{-1}) \quad (2)$$

수식 (2)를 이용하여 생성한 구체 임펄스 응답의 변환 필터와 멀티채널 3차원 마이크의 5채널 출력신호 사이의 콘볼루션 연산을 통해 5채널 출력신호를 구체 위의 LT (좌측 90도)와 RT (우측 90도)로 변환한다.

수식 (3)은 멀티채널 3차원 마이크의 출력신호 u_{1-5} 와 구체 임펄스 변환 필터 SCF_{0-355} 를 이용하여 2채널 헤드폰 재생신호를 생성하기 위한 것이다.

$$\begin{aligned} v_L^{HP} &= conv(u_1, SCF_{1-LT}) + conv(u_2, SCF_{2-LT}) + conv(u_4, SCF_{4-LT}) \\ v_R^{HP} &= conv(u_1, SCF_{1-RT}) + conv(u_3, SCF_{3-RT}) + conv(u_5, SCF_{5-RT}) \end{aligned} \quad (3)$$

헤드폰 재생을 위한 좌측 신호는 u_1, u_2, u_4 와 $SCF_{1-LT}, SCF_{2-LT}, SCF_{4-LT}$ 사이의 콘볼루션 연산을 통해 구해지고, 우측 신호는 u_1, u_3, u_5 와 $SCF_{1-RT}, SCF_{3-RT}, SCF_{5-RT}$ 사이의 콘볼루션 연산을 통해 구해진다.

구체 임펄스 응답의 변환 필터 SCF를 이용하지 않고 단순히 좌측 마이크 출력신호(u_2, u_4) 두 개와 우측 마이크 출력신호(u_3, u_5) 두 개를 더해서 헤드폰 재생을 위한 2채널 신호를 생성하는 경우, 좌측과 우측에서 입사하는 음원에 대해 음상이 머리 안에서 위치하는 현상이 발생한다. 하지만 수식 (2)를 통해 구한 구체 임펄스 응답의 변환필터를 이용하면 이 현상을 줄일 수 있다.

III. 후처리 H/W

멀티채널 및 스테레오/스테레오 다이폴 재생환경 등 다양한 재생환경에 따른 재생신호의 생성을 위해 멀티채널 3차원 마이크로폰을 통해 획득한 신호와 다양한 역필터 사이의 콘볼루션 연산이 필요하다. 특히 멀티채널 재생신호의 생성을 위해서는 입력신호와 역필터 사이의 5×5 콘볼루션 연산이 필요하기 때문에, 실시간 처리를 위해 후처리 H/W를 제작하였다. 후처리 H/W는 두 개의

TMS320C6416 DSP를 이용하여 구현하였다. TMS320C6416은 Integer 연산을 위한 DSP이기 때문에 다양한 역필터를 Integer 형태로 변환해 주어야 한다. Float 형태의 필터계수 및 연산을 Integer로 변경하는 경우, Integer변환에 따른 오차가 발생하는데, 이를 최소화하기 위해 필터계수의 스케일을 조절하였다. 그림 11은 Float 필터와 Integer 필터를 이용하여 콘볼루션 연산을 수행한 결과 및 오차를 나타낸다. 그림에서 알 수 있듯이 Integer 변환의 오차가 매우 작기 때문에 변환에 의

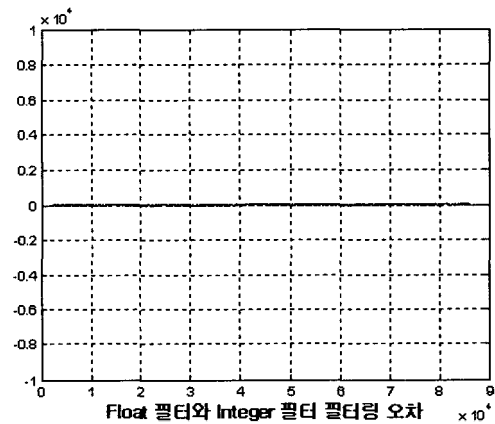
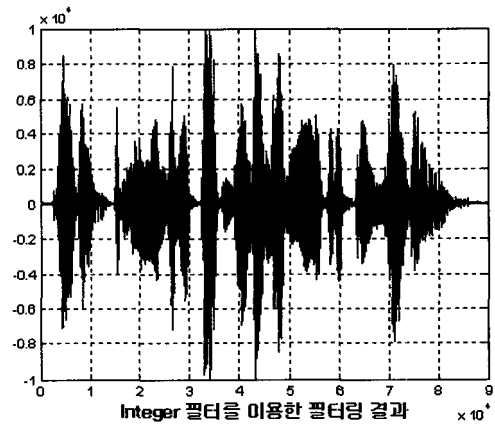
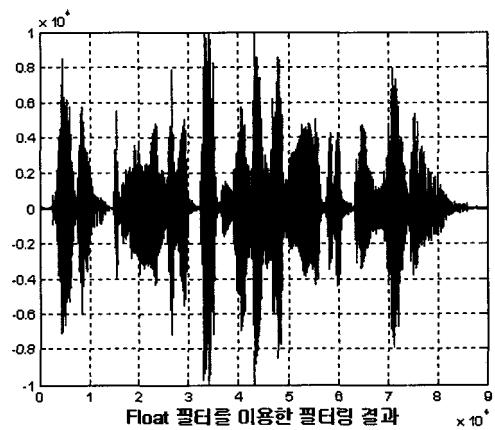


그림 11. Float 필터와 Integer 필터 계수의 비교
Fig. 11. The comparison of float filter and integer filter coefficients.

한 오차는 무시할 수 있다.

멀티채널 3차원 마이크와 후처리 H/W 사이에는 아날로그 처리를 위한 모듈을 두었고 아날로그 처리 모듈은 멀티채널 3차원 마이크 입력신호를 후처리 H/W에 ADAT 인터페이스를 통해 전달한다. 후처리 H/W는 입력신호에 대해 사용자가 선택한 재생환경에 따른 재생신호의 생성을 위한 연산을 수행한 후 다시 ADAT 인터페이스를 통해 아날로그 모듈에 연산결과를 전달하고, 아날로그 모듈은 이 신호를 스피커나 저장장치에 전송한다. 즉 아날로그 처리모듈의 기능은 멀티채널 3차원 마이크의 출력신호를 후처리 H/W에 전달하고, 후처리 H/W의 출력신호를 스피커나 외부저장장치로 전달하는 것이다. 그림 12는 멀티채널 3차원 마이크로폰 및 후처리 H/W이다. 후처리 H/W의 전면부에는 다양한 재생환경을 선택하기 위한 스위치와 스위치 선택에 따른 동작상태를 보여주기 위한 디스플레이가 있다. 후면부에는 멀티채널 3차원 마이크로폰 및 아날로그 모듈과의 연결을 위한 ADAT 인터페이스가 있다.

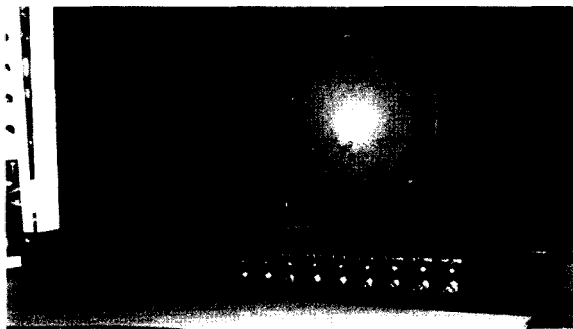


그림 12. 멀티채널 3차원 마이크 및 후처리 H/W
Fig. 12. The multi-channel 3D microphone and post processing H/W.

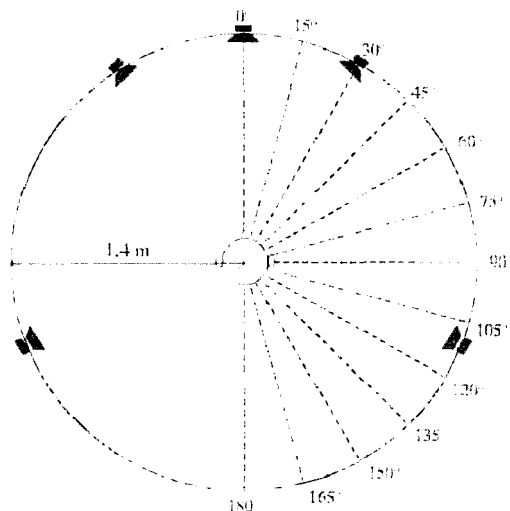


그림 13. 방향성 평가 실험을 위한 스피커 배치 및 실험 콘텐츠의 방향
Fig. 13. Layout of loudspeakers for localization experiments.

IV. 주관적 방향성 평가 실험

본 논문에서 제안한 멀티채널 3차원 마이크의 성능을 평가하기 위해 헤드폰과 멀티채널 재생환경에서 주관적 방향성 평가실험을 수행하였다. 실험은 일본 동경전기 대학 (Tokyo Denki University)의 무향실에서 청력이 정상인 10명의 학생을 대상으로 수행하였다. 실험에 사용한 콘텐츠는 모노 음원을 구체의 임펄스 응답을 통해 변환하여 제작하였다. 3개의 콘텐츠는 남성과 여성의 목소리와 클래식 음악으로 구성하였다. 실험에 사용한 콘텐츠의 길이는 30초이고, 방향은 그림 13과 같이 전면 0도에서 후면 180도까지 15도의 간격을 가지고 있다. 멀티채널 방향성 평가 실험은 피험자가 무향실에 배치한 멀티채널 스피커 재생환경의 중앙에 위치하고 임의의 순서로 재생한 각각의 콘텐츠에 대해 피험자가 느끼는 음원의 방향을 지적하는 방법으로 수행하였다.

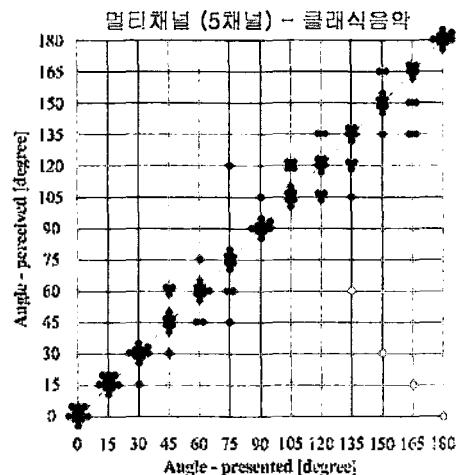
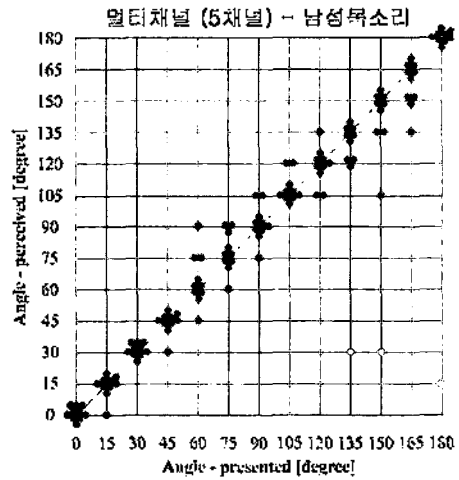


그림 14. 멀티채널 방향성 평가 실험 결과: 5채널
Fig. 14. The result of localization experiments for 5ch reproduction environment.

4.1. 멀티채널 환경에서 방향성 평가 실험 결과

5채널 재생실험을 위해 ITU 5.1 스피커 배치 중 LFE를 위한 subwoofer를 제외한 5개의 스피커를 무향실에 그림 13와 같이 설치하고 실험을 수행하였다. 4채널 재생실험에서는 5채널 재생환경 중 중앙의 스피커를 제외한 4개의 스피커를 이용하였다. 그림 14는 5채널 방향성 실험의 결과를 보여준다. 그림 14에서 수평축은 실제 콘텐츠의 방향을 의미하고 수직축은 피험자가 느낀 방향을 의미한다. 그림 14에서 5채널 방향성 실험결과 전/후방 혼동현상이 현저하게 줄어들었음을 알 수 있다.

그림 15는 4채널 재생환경에서의 방향성 실험결과를 보여준다. 그림 15에서 알 수 있듯이 4채널 재생환경에서도 전/후방 혼동현상이 현저하게 줄어들었다. 하지만 5채널 재생환경에서 보다 전방의 음상이 많이 약하다는 피험자들의 의견이 있었다. 5채널 재생인 경우, 전방 0도에서 입사하는 음원의 생성을 위해 중앙 스피커에서

많은 에너지를 방출하지만, 4채널 재생인 경우, 좌측/우측 스피커에서 동일한 크기의 에너지를 방출하기 때문에 5채널 재생인 경우보다 전방의 음상이 약해지고 피험자가 느끼는 음원의 거리감도 가까워진다.

멀티채널 재생환경에서의 방향성 실험 후, 몇몇 피험자는 음원의 전/후 방향에 따라 음색이 달라지는 현상이 발생한다는 의견이 있었는데, 이는 역필터의 주파수 특성 때문이다. 그림 6과 그림 8의 역필터의 주파수 특성에서 전체 주파수 영역에 대해 크로스토크를 제거하는 비율이 약간씩 차이가 있기 때문에 음원의 방향에 따른 약간의 음색변화가 발생하였다.

4.2. 헤드폰을 이용한 방향성 평가 실험 결과

헤드폰을 이용한 방향성 실험을 위해, 수식 (3)을 이용하여 생성한 헤드폰 재생신호와 KEMAR 더미헤드의 HRTF를 이용하여 생성한 헤드폰 재생신호를 이용하여

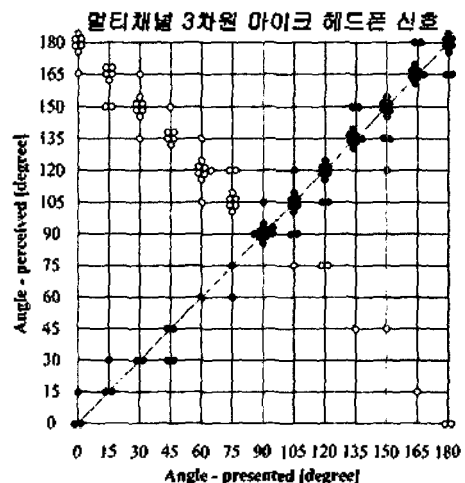
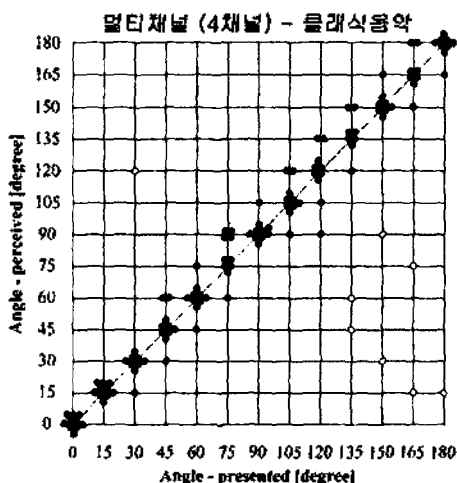
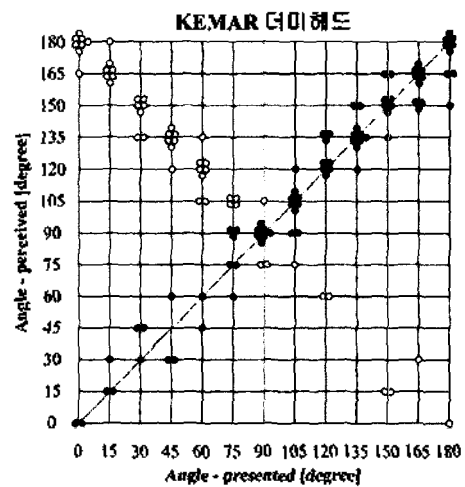
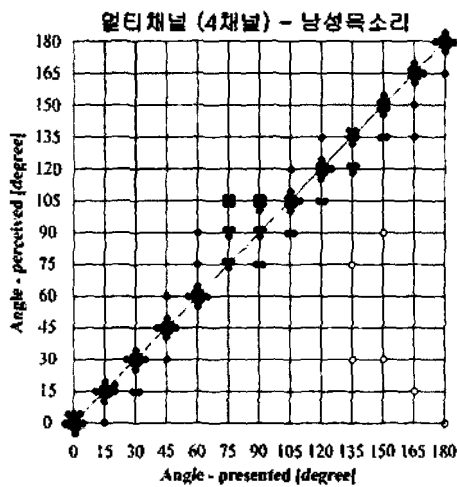


그림 15. 멀티채널 방향성 평가 실험 결과: 4채널
Fig. 15. The result of localization experiments for 4ch reproduction environment.

그림 16. 헤드폰을 이용한 방향성 평가 실험 결과
Fig. 16. The result of localization experiments for headphone reproduction environment.

실험을 수행하였다. 실험은 멀티채널 재생환경에서의 방향성 실험과 동일한 방법을 사용하였다. 그림 16은 헤드폰을 이용한 방향성 평가 실험 결과를 보여주는데 두 경우 모두 멀티채널 재생실험에 비해 전/후방 혼동현상이 많이 발생하였다. 헤드폰 재생 시 전/후방 혼동현상이 발생한 주된 원인은 멀티채널 환경과는 달리 실험자의 머리움직임을 반영할 수 없기 때문이다. 멀티채널 재생 환경인 경우, 피험자가 머리를 움직임에 따라 좌/우측 귀에 입사하는 신호가 변화하면서 전/후 방향에 대한 정보를 주지만, 헤드폰을 통해 재생하는 경우 이러한 정보를 줄 수 없기 때문에 전/후방 혼동현상이 많이 발생한다.

4.3. 실험결과와 분석

표 1은 멀티채널 및 헤드폰 재생실험결과를 분석한 것이다. 표의 값은 실험에 사용한 각각의 콘텐츠의 실제 각도와 피험자가 지각한 각도의 평균오차 값을 의미한다. 멀티채널 재생인 경우, 5채널 환경에서는 평균 7.4도의 각도오차가 있었고, 4채널 재생인 경우 평균 9.6도의 각도오차가 있었다. 헤드폰 재생인 경우 평균 42.7도의 각도오차가 있었고, KEMAR 더미헤드인 경우 평균 41도의 각도오차가 있었다. 따라서 멀티채널 재생인 경우 (5채널, 4채널) 기존 더미헤드에 비해 전/후방 혼동현상이 현저하게 줄어들었고, 헤드폰 재생 시는 비슷한 결과를 보임을 알 수 있다.

표1. 주관적 방향성 평가실험 결과
Table 1. The result of subjective localization experiments.

각도 \ 환경	5채널	4채널	헤드폰	KEMAR
0	0	0	126	142.5
15	5.5	6	103.5	108
30	0.5	5	84	44
45	3.8	4	58.5	66
60	6	5.5	48	48
75	8.3	7.5	28.5	18
90	3	5.5	1.5	4.5
105	4.5	3.5	7.5	7.5
120	6	8	12	16.5
135	15	15.5	12	4.5
150	14.3	20	16.5	31.5
165	15.8	25.5	18	18
180	17.3	19	39	24
평균	7.4	9.6	42.7	41

기 위한 방향성 평가 실험 결과 멀티채널 재생환경에서 기존 더미헤드 기술의 단점인 전/후방 혼동현상을 현저하게 줄일 수 있었다. 본 논문에서 제안한 입체음향의 획득을 위한 멀티채널 3차원 마이크관련 기술은 향후 실감방송이나 3DTV등에서 사용할 입체음향의 획득에 활용 가능하다.

본 시스템은 재생환경의 중앙에 위치한 청취자만을 위한 시스템으로, 향후 청취영역의 확대 및 재생 시 전/후방 음원의 음색 변화에 대한 연구가 필요하다.

V. 결론

본 논문에서는 구체 위에 배치한 다수 개의 마이크를 이용하여 제작한 멀티채널 3차원 마이크를 이용하여 3차원 오디오 신호를 획득하고 후처리 과정을 통해 다양한 재생환경에서 재생 가능한 입체음향 획득 시스템에 대해 기술하였다. 멀티채널 3차원 마이크는 중앙의 음원을 강조하고 인간의 머리움직임을 고려하기 위한 위치에 마이크를 배치하여 제작하였다. 후처리 과정은 다양한 재생환경에서의 스피커 배치에 따른 크로스토크를 제거하는 과정으로써, 재생환경에 적합한 역필터를 이용하여 각각의 크로스토크를 제거한다. 후처리 과정은 다양한 콘볼루션 연산을 위해 많은 계산량이 필요하기 때문에 H/W로 제작하였다. 멀티채널 3차원 마이크의 성능을 검증하

감사의 글

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

참고 문헌

1. Francis Rumsey and Tim McCormick, *Sound and Recording: An Introduction*, Focal Press, 2002.
2. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT, Cambridge, MA, 1983.
3. 김경옥, 장대영, 서정일, 정대권, "객체기반 3차원 오디오 방송 시스템 설계", *한국음향학회지*, 22 (7), 592-602, 2003.
4. D. H. Cooper and J. Bauck, "Prospects for transaural

recording," J. Audio Eng. Soc. 37 (1/2), 3-19, 1989.

5. 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.
6. <http://www.schoeps.de/E/kfm6.html>
7. <http://www.theaudio.com/bs-3d.html>
8. 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.
9. Corey I. Cheng and Gregory H. Wakefield, "Introduction to HRTFs: Representations of HRTFs in Time, Frequency, and Space," J. AES, 49 (4), 231-249, 2001.
10. 구자현, "Rayleigh의 소리 방향 지각 연구에 대한 과학서적 고찰", 한국음향학회지, 21 (8), 695-702, 2002.
11. P. A. Nelson and S. J. Elliott, *Active Control of Sound*, Academic Press, 1992.
12. 최갑근, 방승범, 정완섭, 김순협, "5.1채널 스피커 시스템에서 트랜스오울 필터 구현에 관한 연구", 한국음향학회지, 21 (3), 245-255, 2002.
13. 김래훈, 임준석, 성광모, "3차원 오디오 시스템에서 합과 차 여파가를 이용한 새로운 간섭신호 제거 방법", 한국음향학회지, 20 (4), 17-21, 2001.
14. 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.
15. Ole Kirkeby, Philip A. Nelson and Hareo Hamada, "The Stereo Dipole - A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers," J. AES, 46 (5), 387-395, 1998.
16. ITU-R Recommendation BS. 755: "Multi-channel stereophonic sound system with or without accompanying picture" 1993.

저자 약력

• 이 태 진 (Taejin Lee)



1973년 9월 21일생
 1996년 2월: 전북대학교 전자공학과 공학사
 1998년 2월: 전북대학교 대학원 전자공학과 공학석사
 1998년 2월~2000년 5월: (주)모멘스 영상음성
 기술연구원 연구원
 2002년 10월~2003년 9월: 일본 Tokyo Denki Univ.
 객원연구원
 2000년 5월~현재: 한국전자통신연구원 방송미디어연구
 그룹 연구원

• 강 경 옥 (Kyeongok Kang)



1962년 11월 17일생
 1985년 2월: 부산대학교 물리학과 이학사
 1988년 2월: 부산대학교 대학원 물리학과 이학석사
 2004년 2월: 한국항공대학교 대학원 항공전자공학과 공학
 박사
 1991년 2월~현재: 한국전자통신연구원 방송미디어연구
 그룹 3D미디어연구팀장, 책임연구원