

청취자 위치 적응 실시간 사운드 재생 시스템의 개발

Development of a Listener Position Adaptive Real-Time Sound Reproduction System

이 기 승*, 이 석 필**
(Ki-Seung Lee*, Seok-Pil Lee**)

*건국대학교 정보통신대학 전자공학부, **전자부품연구원 방송통신융합 연구센터
(접수일자: 2010년 8월 3일; 채택일자: 2010년 10월 6일)

본 논문에서는 두 개의 스피커를 이용한 청취 환경에서 좌, 우 채널의 간섭 신호를 제거하기 위한 새로운 오디오 시스템을 개발하였다. 간섭 제거는 청취자의 위치에 따라 적응적으로 이루어져야 하기 때문에, 청취 위치를 추적하기 위한 기법이 적용되었다. 청취자 위치 추적은 2개의 마이크를 통하여 이루어지며 채널 간 시간 지연을 이용하여 청취자의 방향을 추정하도록 하였다. 또한 잔향 환경에서의 사용을 고려하여 선형 예측 기법을 이용한 잔향 제거 기법이 적용되었다. 좌, 우 채널의 간섭제거를 위한 음원-귀 간의 경로는 KEMAR 머리전달함수를 이용하여 나타내었다. 사용된 청취자 방향 측정 시스템의 유용성을 평가하기 위해 추정된 위치에서 채널 간섭의 성능을 평가하였다. 평가 척도로 채널 분리 비를 사용하였으며, 실험적인 결과, 사용자의 실제 위치와 추정된 위치 간에 다소 차이가 있더라도 -10 dB의 채널 분리비가 얻어짐을 확인할 수 있었다. 제안된 알고리즘은 부동소수점 디지털 신호처리 프로세서를 이용하여 실시간 구현되었으며 청취자 평균 방향 오차는 5도, 주관적 간섭 제거율은 평균적으로 80 % 얻어짐을 알 수 있었다.

핵심용어: 채널 간섭 제거, 도달각 추정, 실시간 구현

투고분야: 음향 신호처리 분야 (1.2), (1.5)

In this paper, a new audio reproduction system was developed in which the cross-talk signals would be reasonably cancelled at an arbitrary listener position. To adaptively remove the cross-talk signals according to the listener's position, a method of tracking the listener position was employed. This was achieved using the two microphones, where the listener direction was estimated using the time-delay between the two signals from the two microphones, respectively. Moreover, room reverberation effects were taken into consideration where linear prediction analysis was involved. To remove the cross-talk signals at the left-and right-ears, the paths between the sources and the ears were represented using the KEMAR head-related transfer functions (HRTFs) which were measured from the artificial dummy head. To evaluate the usefulness of the proposed listener tracking system, the performance of cross-talk cancellation was evaluated at the estimated listener positions. The performance was evaluated in terms of the channel separation ration (CSR), a -10 dB of CSR was experimentally achieved although the listener positions were more or less deviated. A real-time system was implemented using a floating-point digital signal processor (DSP). It was confirmed that the average errors of the listener direction was 5 degree and the subjects indicated that 80 % of the stimuli was perceived as the correct directions.

Keywords: Cross-talk cancellation, direction of arrival (DOA) estimation, real-time implementation

ASK subject classification: Acoustic Signal Processing (1.2), (1.5)

I. 서론

좌, 우 2개의 라우드 스피커를 이용한 스테레오 재생

환경에서는 헤드폰을 이용한 양이 재생 환경 (binaural reproduction environment) 과 달리 좌, 우채널의 소리가 각기 다른 채널의 귀에 전달된다. 이때 우측 귀에 전달된 좌 채널의 소리, 또는 좌측 귀에 전달된 우 채널의 소리를 간섭 신호 (cross-talk signal) 라 부르며, 이러한 간섭 신호는 음원의 정위감과 방향감에 영향을 끼치는 것으로

알려져 있다 [1]. 만약 좌, 우 스피커와 양쪽 귀간의 음향 경로를 알 수 있다면, 간섭 신호를 효과적으로 제거할 수 있는데, 간섭 제거 필터 (cross-talk cancellation filter) [1-3]는 이러한 목적으로 설계된 필터를 뜻한다. 간섭 제거 필터는 귀와 음원간의 경로에 의존적으로 결정되며, 이 경로는 청취자의 위치와 스피커의 위치에 따라 변화하므로 간섭 제거는 매우 제한된 영역에서만 이루어진다 [2, 3]. 보다 넓은 영역에서 간섭 제거가 이루어지기 위해서는 청취자의 위치를 실시간으로 추적하는 장치가 오디오 시스템에 포함되어야 하는데, 이를 구현하기 위한 방법으로 비디오 카메라 [4], 레이저 위치 추적기 [5], 리모트 콘트롤 [6] 을 이용한 방법들이 소개되었다. 비디오 카메라를 이용한 방법은 영상 내에 피부색에 해당하는 부분을 분할하여 청취자의 위치를 추정하는 방법으로, 실시간 추적이 가능하나 실시간 영상처리에 따르는 하드웨어 부담이 큰 단점이 있으며, 레이저 위치 추적기를 이용한 방법은 위치 추적으로 정밀도를 매우 높일 수 있으나, 고비용의 특수한 하드웨어 장치가 필요하다는 단점이 있다. 리모트 콘트롤을 이용한 방법은 적외선 센서 및 2개의 초음파 센서를 이용하여 각 신호의 도달 시간차로부터 청취자 위치 추적이 가능한, 저렴한 비용의 구현이 가능하나 청취자가 항상 리모트 콘트롤을 쥐고 있어야 하며, 실제 추정되는 위치는 리모트 콘트롤을 쥐고 있는 손의 위치로서, 귀의 위치와는 거리 차이가 있을 수 있다는 단점이 있다.

본 논문에서는 별도의 장치를 이용하지 않고 사용자가 발생시킬 수 있는 소리 (박수 소리 등) 를 이용하여 청취자의 위치를 찾을 수 있는 기법을 적용하였다. 이러한 기법은 두 개 마이크론의 도달 시간 차를 이용한 도달각 (direction of arrival, DOA) 추정을 통해 구현되며, 비교적 간단한 계산량으로 오디오 시스템에 쉽게 탑재될 수 있다는 장점이 있다.

간섭 신호를 제거하기 위해서는 음원-귀간의 경로 모델링이 필요한데, 자유 공간 모델링 기법 [7, 8] 및 머리 전달 함수를 이용한 모델링 기법 [2, 3] 등이 사용되고 있다. 자유 공간 모델링 기법은 채널 경로를 단순히 시간 지연과 음압의 감쇄만으로 표현하는 방법으로, 주파수별 상세한 표현, 즉 귀로 인한 감쇄 효과 (shading effects) 및 회절 효과 등을 표현할 수 없다는 한계점을 안고 있다. 반면 머리 전달 함수를 이용한 방법은 주파수별 상세한 경로 모델이 가능하나, 개인 별로 각기 다른 특성을 지니고 있기 때문에 개인화 (customization) 와 같은 별도의 전처리가 필요하다는 단점이 있다. 본 연구에서는 더미

헤드에서 측정된 머리 전달 함수를 이용하여 간섭 제거 필터를 설계하였으며, 측정된 청취자의 위치에 따라 적응적으로 필터 계수를 구하였다.

제안된 위치 추적 기법 및 간섭 제거 기법은 부동 소수점 연산이 가능한 범용 디지털 신호처리 프로세서를 이용하여 실시간 오디오 시스템 형태로 구현하였으며, 성능 평가를 위해 추정된 위치와 실제 위치 간의 차이, 그리고 청취자에게 음원의 방향을 얼마나 잘 인식할 수 있는가를 실험을 통해 평가하였다.

II. 청취자 위치 기반 간섭 제거 시스템

2.1. 스테레오 재생 환경에서 간섭 제거

좌, 우 두 개의 라우드 스피커를 사용하는 스테레오 재생 환경은 그림 1과 같이 총 4개의 음향 전파 경로 $H_{LL}(f)$, $H_{LR}(f)$, $H_{RL}(f)$, $H_{RR}(f)$ 를 갖는다. 여기서 $H_{ij}(f)$ 는 i -채널 스피커와 j -채널 귀에 대한 주파수 응답을 나타낸다. 이와 같은 경로는 자유 공간 상에서 전파되는 음파 모델을 이용한 자유 공간 모델 (free field model) 또는 음파와 귀, 머리, 몸통 간의 회절, 투과를 고려한 머리 전달 함수 모델 (head-related transfer function model) 을 이용하여 나타낼 수 있는데, 본 논문에서는 머리 전달 함수 모델을 이용하였다. 스피커에서 발생하는 좌, 우 채널의 소리를 각각 $S_L(f)$ 와 $S_R(f)$ 라 하고, 좌, 우 귀에 전달되는 소리를 각각 $X_L(f)$ 와 $X_R(f)$ 라 한다면 다음과 같이 나타낼 수 있다.

$$\begin{bmatrix} X_L(f) \\ X_R(f) \end{bmatrix} = \begin{bmatrix} H_{LL}(f) & H_{LR}(f) \\ H_{RL}(f) & H_{RR}(f) \end{bmatrix} \begin{bmatrix} S_L(f) \\ S_R(f) \end{bmatrix} \quad (1)$$

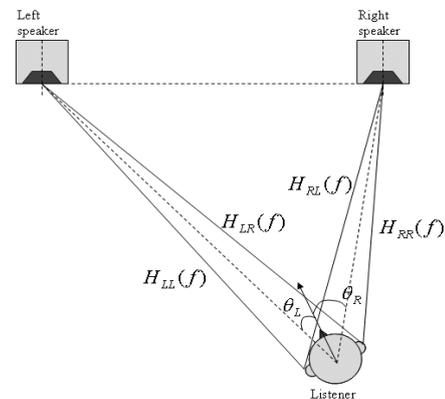


그림 1. 두 개의 라우드 스피커를 이용한 스테레오 재생환경에서 청취자와 스피커간의 전파 경로
Fig. 1. Acoustic propagation paths between the loud speakers and listener's ears under the stereo phonic environment.

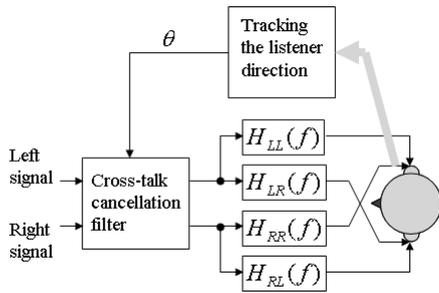


그림 2. 청취자 위치 기반 간섭 제거 기법의 블록도
 Fig. 2. The blockdiagram of the listener position-based cross-talk cancellation system.

식 (1)을 행렬식으로 표현하면 다음과 같다.

$$X = HS \tag{2}$$

여기서 $X = [X_L(f) \ X_R(f)]^T$, $S = [S_L(f) \ S_R(f)]^T$ 를 나타내며 H 는 HRTF를 포함하는 행렬을 나타낸다. 식 (1)에서 보듯이 좌측 귀에 전달되는 신호 $X_L(f)$ 에는 오른쪽 스피커에서 발생하는 신호인 $S_R(f)$ 를 포함하고 있으며, 우측 귀에도 마찬가지로 좌측 스피커에서 발생하는 신호를 포함한다. 좌, 우측 귀에 본래의 채널 신호 $S_R(f)$, $S_L(f)$ 만이 들리도록 하기 위해서는 스피커로 신호를 전달하기에 앞서서 간섭 신호를 제거하기 위한 행렬을 S 에 곱해주어야 한다. 즉,

$$Y = HCS \tag{3}$$

여기서 C 는 간섭 제거를 위한 행렬이다. 만약 귀와 스피커간의 경로를 알고 있으면 행렬 H 를 구성할 수 있으며, $C = DH^{-1}$ 로 설정함으로써 완전하게 간섭 신호를 제거할 수 있다. 여기서 D 는 시간 지연 성분을 갖는 대각 행렬 $D = \text{diag}[e^{-j2\pi ft_d} \ e^{-j2\pi ft_d}]$ 를 나타낸다.

그림 1에 제시된 바와 같이 행렬 H 의 각 성분인 머리전달함수는 청취자와 각 스피커의 상대적인 방향 θ_L, θ_R 에 따라 결정된다. 이는 간섭 제거 필터가 청취자의 위치에 의존적으로 결정됨을 의미하며, 간섭 제거는 매우 제한된 영역 내에서만 이루어짐을 의미한다 [2, 3]. 보다 넓은 영역에서 간섭제거가 이루어지기 위해서는 청취자의 현재 위치를 추적하여 적응적으로 간섭제거 필터를 설계하는 방법이 필요한데, 이와 같은 시스템의 블록도를 그림 2에 나타내었다. 이러한 시스템이 구성되기 위해서는 청취자의 위치를 실시간으로 추적하는 기술이 필요한데, 본 논문에서 적용된 위치 추적 기술을 다음 장에서 제시하였다.

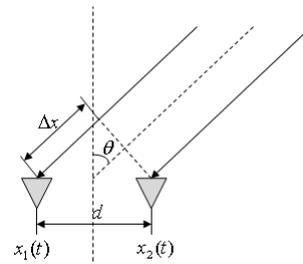


그림 3. 두 개의 마이크로폰에 도달하는 각 신호의 지연과 입사각과의 관계
 Fig. 3. A relationship between the time delay of the two microphone signals and the incident angle.

2.2. 청취자 위치 추적 기법

청취자 위치 추적 기법은 오디오 시스템에 부가적으로 탑재되므로, 오디오 본래의 기능에 큰 부담을 주지 않도록 간단하게 구현되어야 한다. 이러한 관점에서 카메라 [3]나 레이저 [4]와 같은 부가 장치를 이용한 기존의 기법은 부적당함을 알 수 있는데, 본 연구에서는 청취자가 발생시킬 수 있는 소리를 이용하여 위치를 찾는 기법을 적용하였다. 청취자의 소리를 이용하여 청취자의 방향을 찾기 위해, 본 논문에서의 두 개의 마이크로폰을 이용하는 시간 지연 추정 (time delay estimation) 방법을 사용하였다. 두 개의 마이크로폰에 도달한 각 신호의 시간 지연을 τ_d 라 놓을 때, τ_d 와 청취자 방향 θ 간의 관계를 그림 3에 나타내었으며, 이를 식으로 나타내면 다음과 같다.

$$\tau_d = \frac{\Delta x}{c} = \frac{d \sin \theta}{c} \tag{4}$$

여기서 Δx 는 두 신호의 경로 차이를, d 와 c 는 각각 두 마이크로폰간의 거리 및 상온에서 음속 ($\approx 343\text{m/sec}$)을 나타낸다. 식 (4)로부터 시간 지연 τ_d 가 주어졌을 때 청취자의 방향은 다음과 같이 구할 수 있다.

$$\theta = \sin^{-1} \left(\frac{\tau_d c}{d} \right) \tag{5}$$

두 신호의 시간차를 구하는 방법으로, 적은 계산량으로 비교적 높은 정밀도를 갖는 일반화된 상호 상관 (generalized cross correlation, GCC) 기법 [11]이 사용되었다. GCC 기법에서는 아래 식과 같이 필터링된 두 신호간의 상호 상관 함수가 최대가 되는 지점을 두 신호의 시간 지연 τ_d 로 간주한다.

$$\tau_d = \arg \max_{\Delta t} E[x_1'(t) x_2'(t + \Delta t)] \tag{6}$$

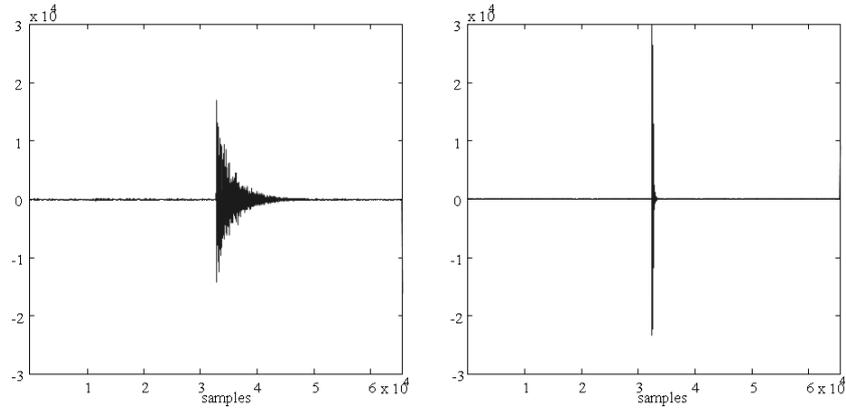


그림 4. 잔향 환경에서 녹음된 박수소리(좌) 와 무향실에서 녹음된 박수 소리(우)
Fig. 4. An example of the hand-clapping signals recorded in a room (left) and an anechoic room (right).

여기서 $x'_1(t)$ 와 $x'_2(t)$ 는 각각 필터링된 두 마이크로폰 신호를 나타낸다.

$$x'_1(t) = h_1(t) * x_1(t), \quad x'_2(t) = h_2(t) * x_2(t) \quad (7)$$

$h_1(t)$ 와 $h_2(t)$ 는 각각 두 개 마이크로폰 신호와 음원간의 경로차를 반영하여 결정된다. 오디오 기기가 사용되는 거실과 같은 일반적인 환경에서는 공간 충격파 응답(room impulse response)이 존재하는데, 이에 대한 예를 그림 4에 제시하였다. 그림 4의 오른쪽은 무향실에서 녹음된 박수소리의 파형이고, 왼쪽은 4 x 5 x 3 M (WDH) 크기를 갖는 잔향 공간에서 녹음된 박수소리의 파형이다. 무향실에서의 박수소리는 그림에서 보듯이 충격파 응답과 상당히 유사한 모양을 갖으며, 따라서 잔향이 있는 공간에서 녹음된 박수 소리는 녹음이 이루어지는 공간의 충격파 응답과 유사하다고 가정할 수 있다. GCC 기법에서 상호 상관값이 계산되는 신호는 충격파 응답의 모양에 가까울수록 시간 지연 측정의 정확도가 높아지는 것으로 알려져 있으며 [11], 따라서 $h_1(t)$ 와 $h_2(t)$ 는 공간 충격파 응답의 역필터(inverse filter)로 주어져야 한다. 이러한 역필터를 구현하기 위해 본 논문에서는 공간 충격파 응답을 자기 회귀(auto regressive, AR) 모델로 근사화하여 표현하였다.

AR 모델에서는 주어진 충격파 응답을 몇 개의 극점(pole)을 이용하여 표현하는데, 이는 주어진 충격파 응답을 무한 길이 충격파 응답(infinite impulse response, IIR) 필터로 근사화함을 의미한다. 그림 4에서 보듯이 공간 충격파 응답은 상당히 긴 구간(44.1 kHz 샘플링 주파수에서 약 10,000 샘플에 걸쳐 서서히 감쇄하며, 이와 같은 긴 길이의 충격파 응답은 AR 모델을 이용하여 표현

하는 것이 적합하다고 볼 수 있다. AR 모델에서의 역필터링은 아래와 같이 구현된다.

$$e_i(t) = x_i(t) - \sum_{p=1}^P a_i(p)x_i(t-p), \quad i = 1, 2 \quad (8)$$

여기서 $e_i(t)$ 는 i -번째 마이크로폰 신호에 대한 역필터 신호를, $a_i(p)$ 와 P 는 각각 선형 예측 계수(linear prediction coefficient) 및 예측 차수를 나타낸다. 예측 차수는 본 연구에서 50으로 설정하였다. 식 (8)에서 역필터 신호는 실제적으로 선형 예측 잔차 신호(linear prediction error signal)로 주어짐을 알 수 있으며, 이와 같은 선형 예측 잔차 신호를 이용한 시간 지연 추정 방법은 잔향 환경에서 우수한 성능을 나타내는 것으로 알려져 있다 [10, 12]. 따라서 본 논문에서는 선형 예측 잔차 신호를 이용하여 시간 지연을 추정하였다. 이 경우, 시간 지연의 추정 식은 아래와 같이 주어진다.

$$\tau_d = \arg \max_{\Delta t} E[e_1(t)e_2(t + \Delta t)] \quad (9)$$

2.3. 청취자의 방향을 이용한 간섭 제거 필터의 설계

스테레오포닉 재생 환경에서 청취자의 방향, 각 스피커와 청취자와의 방향을 그림 5에 나타내었다. 그림에서 d_s 는 두 스피커 간의 거리, θ_u 는 청취자의 방향, y_u 는 두 스피커가 놓인 직선과 청취자간의 수직 거리, 그리고 θ_L 과 θ_R 은 청취자에 대한 좌, 우 스피커의 방향을 나타낸다. 2.1절에 제시한 바와 같이 이러한 조건에서의 간섭 제거 필터는 각도 θ_L 과 θ_R 에 대응하는 좌, 우 채널의 머리 전달함수를 이용하여 구현할 수 있다. 청취자의 방향 θ_u 와 거리 y_u 가 주어져 있을 때, θ_L 는 θ_R 다음과 같이 구할 수 있다.

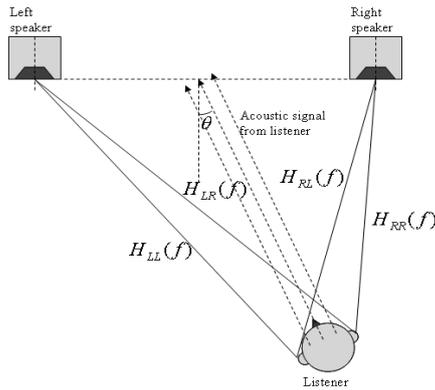


그림 5. 스테레오 재생환경에서 청취자의 방향과 각 스피커의 방향
 Fig. 5. Direction of the listener and locations of each speaker under the stereo phonic environment.

$$\theta_L = \tan^{-1}\left(\frac{d_s}{2y_u} - \tan\theta_u\right) \quad \theta_R = \tan^{-1}\left(\frac{d_s}{2y_u} + \tan\theta_u\right) \quad (10)$$

(10)에서 θ_u 는 2.2절에 제시한 청취자 방향 추정 기법을 통해 얻어지며, 스피커 간 거리 d_s 는 일반적으로 스피커를 설치하고 난 뒤 이동하지 않는다고 가정하여 고정된 값으로 셋팅이 가능하나 y_u 는 청취자의 위치이기 때문에 고정된 값으로 셋팅이 불가능하다. 본 논문에서는 청취자의 y -축 변위 Δy_u 에 따라 간섭 제거 성능이 어떻게 변화하는지 살펴봄, 이로부터 청취자의 y -축 위치를 고정시키는 경우, 허용 가능한 청취자의 y -축 변위를 조사하였다. 만약 청취자의 방향이 θ_u 로 주어지고 y -축의 실제 청취자 위치가 $y_u + \Delta y_u$ 인데 셋팅된 값이 y_u 이라면, 간섭 제거 필터 행렬식은 아래와 같이 주어진다.

$$Y = HCS = H(y_u + \Delta y_u)H^{-1}(y_u)S \quad (11)$$

식 (11)에서 음원-귀 간의 경로를 나타내는 행렬 H 와 간섭 제거 필터 C 에 공통으로 포함되는 변수 d_s , θ_u 를 생략하여 나타내었다. 식 (11)에서 두 행렬의 H 와 C 는 서로 역행렬 관계를 갖지 못하므로 두 행렬의 곱 HC 는 대각 행렬 형태를 갖지 못하고, 비 대각 성분을 갖게 된다. 이 때 행렬 곱 $G = HC$ 의 비 대각 성분(간섭 신호)과 대각 성분(주신호) 간의 비를 채널 분리 비(channel separation ratio, CSR) [2, 3]라 하며, 다음과 같이 나타낼 수 있다.

$$CSR_L = 10 \log_{10} \frac{\int_{f_{\min}}^{f_{\max}} |G_{11}(f)|^2 df}{\int_{f_{\min}}^{f_{\max}} |G_{12}(f)|^2 df},$$

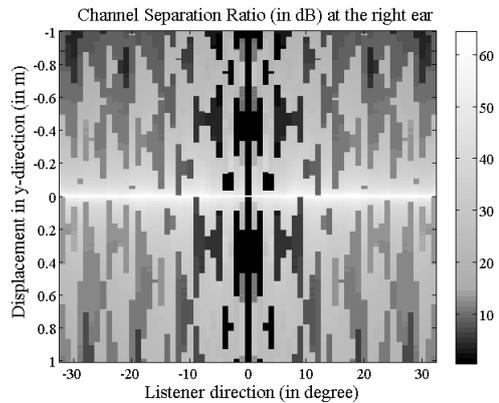
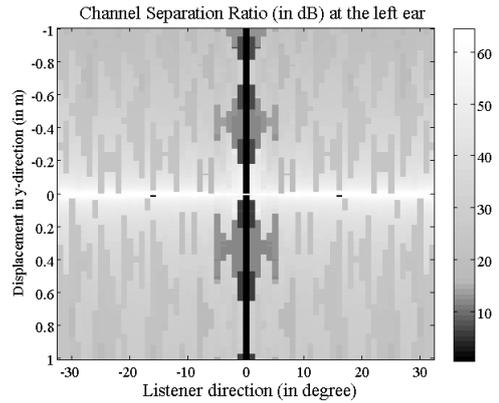


그림 6. 스피커 간 거리 5 M, 청취자 위치 4 M에서 청취자 y -축 변위에 따른 채널 분리도
 Fig. 6. The channel separation ratios for displacement in y -direction and for each listener angles, in case when distance between the two loudspeakers is 5 M and y -location of the listener is 4 M.

$$CSR_L = 10 \log_{10} \frac{\int_{f_{\min}}^{f_{\max}} |G_{11}(f)|^2 df}{\int_{f_{\min}}^{f_{\max}} |G_{12}(f)|^2 df} \quad (12)$$

여기서

$$G = \begin{bmatrix} G_{11}(f) & G_{12}(f) \\ G_{21}(f) & G_{22}(f) \end{bmatrix} = HC$$

(12)에서 CSR이 구해지는 주파수 범위 $[f_{\min}, f_{\max}]$ 는 head shading effect 영향을 덜 받는 20 Hz ~ 3 kHz 로 설정하였다. 청취자의 y -축 변위 Δy_u 에 따른 CSR_L 과 CSR_R 을 그림 6에 나타내었다. 그림 6에 제시된 결과는 MIT의 KEMAR 머리전달함수 [9]를 이용하였으며, 스피커 간의 거리 $d_s = 5M$, 청취자의 y -축 값을 4 M로 셋팅한 경우 얻어진 것이다. KEMAR 머리전달 함수를 측정할 환경과 실험 환경이 다르기 때문에, 본 논문에서는 청취자의 중심 거리와 스피커 간의 거리를 이용하여 KEMAR

머리전달함수의 주파수 특성을 보정하였다. 그림에서 세로축은 청취자의 셋팅 위치 4 M를 기준으로 +/-1 M 이동한 변위값을, 가로축은 청취자의 방향 θ_u 을 나타내며, 이들 두 변수에 따른 CSR_L 과 CSR_R 값을 그레이 스케일로 나타내었다. CSR이 10 dB를 넘는 경우 만족할 만한 간섭 제거의 효과를 얻을 수 있다고 가정하는 경우 [3] 왼쪽 채널은 청취자가 0도에 놓인 경우를 제외하고 대부분 위치에서 만족할 만한 간섭 제거가 이루어짐을 알 수 있다.

반면 오른쪽 채널에 대해서는 왼쪽 채널에 비해 CSR이 다소 낮게 분포되어 나타나지만, 왼쪽 채널과 마찬가지로 청취자의 y-축 변위가 +/-1 M 이내로 국한될 경우, 10 dB 이상의 CSR을 얻을 수 있음을 나타내었다. 실제 청취 환경에서는 청취자의 위치 변동은 주로 x-축 방향에서 이루어지며, y-축 변위는 상대적으로 크지 않기 때문에 y-축 변위를 +/-1 M로 제한하는 것은 실용상 큰 문제가 없을 것으로 판단된다. 따라서 본 논문에서는 청취자의 y-축 방향을 스피커의 거리와 함께 프리-셋팅하는 기법을 적용하였다.

2.3. 하드웨어 시스템

제작된 하드웨어 시스템의 블록도를 그림 7에 제시하였다. 간섭 제거 및 청취자 방향 추정을 수행하기 위해 32비트 부동소수점 디지털 신호처리 프로세서인 DSP6713 (225 Mhz) [13]가 사용되었으며 사용자 인터페이스를 위해 ARM9 마이크로프로세서가 사용되었다. 사용자 위치 추적을 위한 센서로 2개의 마이크로폰이 사용되었으며 저잡음 앰프 및 2채널 시그마-델타형 A/D변환기를 통해 디지털화된 데이터를 DSP에 전송하도록 하였다. 마이크로프로세서 보드에는 운영시스템과 응용 프로그램 탑재를

위한 64 MBytes NOR-Flash 메모리 및 32 MBytes D-RAM이 탑재되어 있으며 외부 사용자와의 데이터 전송을 위한 Wireless LAN이 연결되어 있다. 디스플레이는 진공냉음극관 (Vacuum Fluorescent Display, VFD)이 사용되어 현재 오디오 기기의 상태가 표시되도록 하였다. 사용자 명령은 적외선 리모트 콘트롤에 의해 전달되며 이에 따라 적외선 수신기 및 리모트 콘트롤 인터페이스가 마이크로프로세서 보드와 연결되어 있다. 마이크로프로세서는 사용자 인터페이스 기능 외에 MP3 플레이어 기능을 포함하고 있으며, MP3 파일은 프로세서 보드와 연결된 1G SD메모리 카드에 저장되며, 무선 LAN을 외부로부터 다운로드가 가능하도록 설계하였다.

마이크로프로세서와 DSP간의 통신 IIC 버스를 통해 이루어지며, 음원은 내부에 장착된 MP-3 재생기 외에 2개의 외부 입력을 선택할 수 있도록 하였다.

그림 8은 제작된 시스템의 기능적 블록도를 나타낸 것으로, 마이크로프로세서 보드에서는 사용자 인터페이스 프로그램 및 MP-3 복호화 프로그램이 실행되며 DSP 보드에서는 시간지연 추정을 바탕으로 하는 청취자 방향 추적 프로그램 및 간섭 제거 프로그램이 실행된다. 마이크로프로세서에서 실행되는 프로그램은 WinCE 운영체제에서 Embedded Visual C++를 이용하여 작성되었으며, DSP용 프로그램은 Code Composer Studio v3.3 환경 [14]에서 C-언어로 작성되었다. 간섭제거 필터는 고속컨볼루션 기법을 이용하여 채널당 128탭 유한길이필터 (finite impulse response filter)로 근사화하여 설계하였으며 샘플링 주파수 44.1 kHz로 설정하였다. 간섭 제거 필터링은 DSP의 eDMA 기능을 이용하여 오디오 데이터가 버퍼가 다 채워진 이후 수행되도록 하였으며, 버퍼가

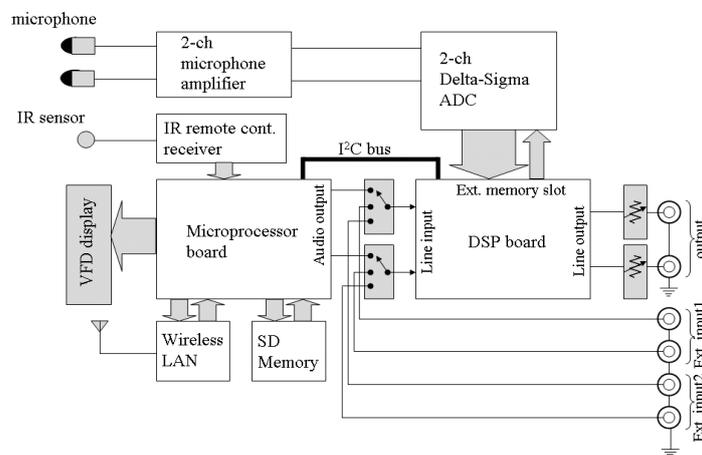


그림 7. 하드웨어 시스템의 블록도
Fig. 7. The blockdiagram of the hardware system.

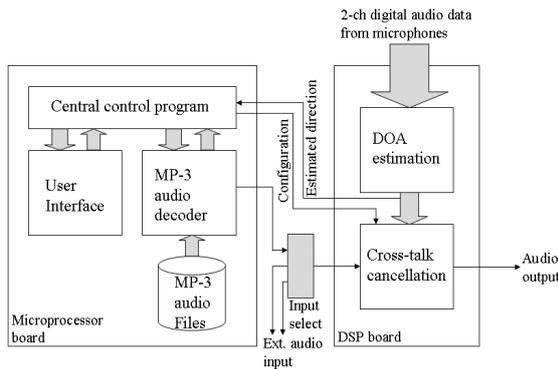


그림 8. 제작된 시스템의 기능적 블록도
Fig. 8. The functional blockdiagram of the developed system.

채워지고 있는 동안은 사용자 위치 추적을 위한 연산이 수행되도록 하였다. 사용자 위치 추적은 약 2초 간격으로 반복적으로 이루어지며, 위치 추적 전 에너지 비교를 수행하여 박수소리와 같은 큰 에너지의 신호가 입력되었을 때만 위치 추적을 수행하도록 하였다. DSP 프로그램의 전체 길이는 약 180 KBytes로서, DSP의 내부 메모리 (192 KBytes)에 탑재될 수 있었다.

그림 9는 제작된 시스템의 내부 및 전면 패널 사진으로, 새시는 두께 1 mm의 방자형 철판으로 구성되고 전면 패널은 헤어라인 가공된 두께 10 mm의 알루미늄으로 제작되었다. 전면은 VFD와 입력 선택, 볼륨 조절 및 위치 추적용 2개 마이크로폰, 리모콘 IR 센서가 장착되어 있다.

III. 결과

3.1. 청취자 방향 추적의 성능 평가

청취자 방향 추적은 지정된 위치에서 청취자가 박수를 치게 하고, 청취자의 방향을 추정하는 실험을 통해 그 성능을 평가하였다. 방향 추정 실험은 160 Hz의 최저 차단 주파수를 갖는 크기 4 x 3 x 1.5 M (WDH)의 무향실과, 크기 8 x 6 x 2.5 M (WDH)의 일반 거실에서 이루어졌다. 청취자의 방향은 -70도~70도 사이를 10도 간격으로 나눈 총 15종류의 방위각으로 설정하였으며, 각 방위각에 대해 5번 반복하여 박수 소리를 취득하였다. 따라서 실험에 사용된 총 샘플 수는 75개였다. 청취자의 손바닥과 두 마이크로폰의 중심 거리는 4 M로 일정하게 유지하였다. 식 (8)로 주어지는 잔향 제거는 거실에서 측정된 데이터에 대해서만 적용하였으며, 적용된 잔향 기법의 유용성을 평가하기 위해 박수 소리를 직접적으로 방향 추적에 사용하는 경우와 선형 예측 후 잔차 신호를 방향 추적에 사용하는 경우, 2가지에 대해 실험을 수행하였다.

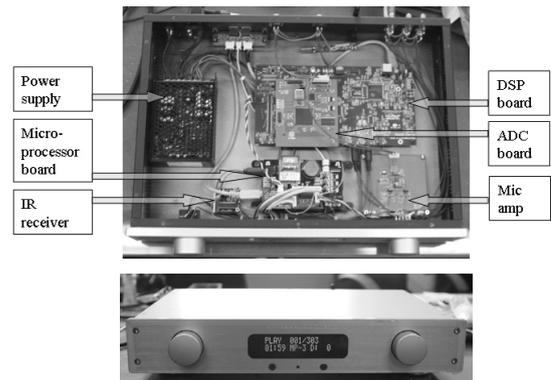


그림 9. 제작된 시스템의 내부 구조(상) 및 전면 사진(하)
Fig. 9. The internal photograph of the developed system (top) and front view (bottom).

방향 예측의 성능은 실제 방향과 추정된 방향간의 상관 계수 및 제공된 평균 제곱 오차 (root mean squared error, RMSE)를 통해 평가하였다.

그림 10에 각 환경에서 취득된 데이터에 대한 실제 방향과 추정된 방향간의 상관도를 나타내었다. 각 상관도의 상단에 상관값과 RMSE값을 함께 나타내었다. 무향실에서 추정된 방향이 실제 방향과 가장 높은 상관값 (0.9936) 및 가장 낮은 RMSE (4.95도)를 나타내었다. 잔향 환경인 거실에서는 무향실과 비교하여 다소 낮은 0.9829의 상관값과 8.10의 RMSE값을 나타내었는데, 그림 10에 제시한 잔향실의 결과는 배경 잡음이 거의 없는 조용한 환경에서 얻은 것으로서, 배경 잡음의 영향이 거의 없다고 가정하면 성능 저하는 주로 잔향의 영향에 의한 것이라고 판단된다. 이는 그림 10의 오른쪽에 제시한 잔향 기법이 적용된 결과를 통해서도 확인할 수 있는데, 선형 예측에 바탕을 둔 잔향 제거 기법이 적용된 경우 상관값은 0.9933, RSME=5.0771로서 무향실에서 측정된 결과와 거의 유사하게 나타났다. 무향실의 결과와 차이점은, 0도 근방의 방위각에 대해서 잔향 제거 기법이 적용된 방법은 매우 우수한 성능을 보인 반면, 방위각이 커짐에 따라 그 편차가 다소 크게 나타남을 그림 10을 통해 확인할 수 있다. 결론적으로 선형 예측 후 잔차 신호를 이용한 청취자 방향 추정 방법은 전체적으로 무향실에서 측정된 결과와 거의 근접한 결과를 나타내어, 잔향 환경에서 효과적인 청취자 방향 추적 방법임을 알 수 있었다.

3.2. 간섭 제거 기법의 주관적 성능 평가

개발된 오디오 시스템의 주관적인 성능 평가를 위해 14개의 방위각에 대해 주관적 청취 테스트를 수행하였다. 테스트에는 총 20명의 정상 청력을 가진 실험자가 참여하

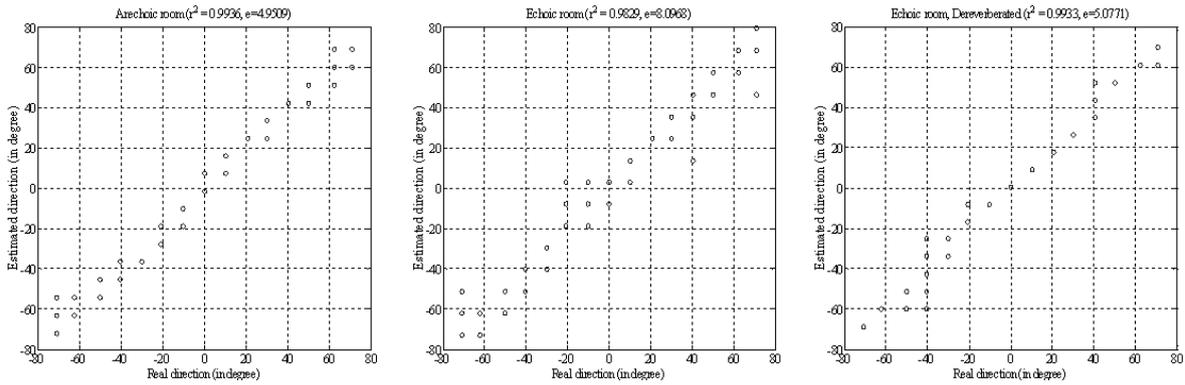


그림 10. 각 방법에 대한 청취자 방향 추정 상관도 (좌: 무향실, 중: 잔향제거가 사용되지 않은 거실, 우: 잔향제거가 사용된 거실)
 Fig. 10. The correlation plots for each listener direction (Left: Anechoic room, Mid: Echoic room without de-reverberation, Right: Echoic room with de-reverberation.

였으며, 10초의 길이를 갖는 5종류의 샘플 (클래식, 팝송, 피아노솔로, 가요, 남자음성)이 사용되었다. 따라서 각 방위각에 대한 테스트 샘플 수는 총 100개였다. 주관적 청취 테스트는 먼저 특정 방위각에 피실험자를 위치시키고, 간섭 제거 전의 소리와 간섭 제거 후의 소리를 들려 준 뒤 어느 소리가 간섭이 없는 기준 소리에 가까운 지를 비교하도록 하였다. 헤드폰으로 청취되는 소리를 간섭이 없는 기준 소리로 간주하였으며, 피실험자는 판단을 내리기 전까지 횟수에 제한없이 헤드폰 소리와 비교할 수 있도록 하였다. 피실험자가 간섭이 제거된 소리를 기준소리와 가깝다고 판단한 경우, 올바르게 간섭 제거가 이루어졌다고 판정하였으며, 이로부터 14개 각 방향에 대한 간섭 제거율을 구하였다. 테스트에 사용된 라우드 스피커는 스튜디오 모니터 스피커인 8020A (Genelec, Finland)가 사용되었으며, 기준 소리의 청취에는 DT990PRO (Beyerdynamic, German)이 사용되었다. 간섭 제거를 위한 청취자 위치 정보는 잔향 제거가 적용된 청취자 방향 예측 기법을 통해 얻었으며, 청취자의 y-축 위치는 고정되어 있다고 가정하였다. 오디오 시스템에 장착된 두 마이크로폰의 중심과 청취자간의 거리는 4M로 고정하였으며, 두 스피커간의 거리는 5M였다. 간섭 제거 필터는 역행렬 형태로 존재하며, 이에 따른 역문제 (inverse problem)가 발생되는데 [15], 본 논문에서는 이와 같은 문제를 해결하기 위해 행렬의 대각 성분에 shape factor를 더하여 역행렬을 구하였다. shape factor는 재생음의 음색이 크게 변화하지 않도록 주파수 대역별로 적절히 조정된 값을 사용하였다 [15].

그림 11은 제안된 오디오 시스템의 각 방향에 대한 간섭 제거율을 원의 크기로 표현하여 나타낸 것으로, 큰 원은 높은 인식율을 나타낸다. 그림에서 보면 방위각이 0도에 가까울수록 간섭 제거율이 비교적 낮게 나타나며, 좌,

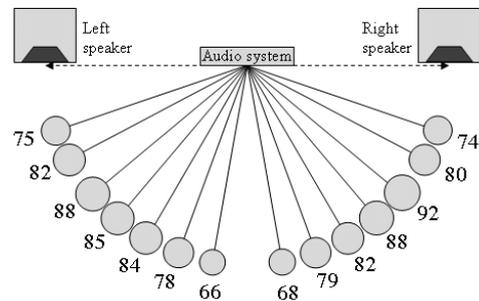


그림 11. 제안된 오디오 시스템의 각 방향에 대한 간섭 보정율
 Fig. 11. Correct adjustment ratio of the stimuli direction for the proposed audio system.

우 스피커에 가까울수록 높게 나타나는데, 이는 0도 근방에서 간섭 제거의 성능이 다소 저하됨을 의미한다. 이러한 결과는 0도 근방에서는 간섭 신호와 주 신호간의 구분이 명확하게 이루어지기 어렵다는 점으로 설명될 수 있는데, 실제로 실험 참여자들은 소리의 방향이 0도 방위각에 가까울수록 간섭이 제거된 신호와 제거되지 않은 신호간의 구분이 어려움을 표시하였다. 반대로 양쪽 스피커에 가까운 방향에 대해서는 간섭 제거의 여부에 따라 소리 구분이 비교적 뚜렷하였으며, 간섭이 제거된 소리가 헤드폰으로 청취한 신호와 비교적 가까움을 나타내었다.

피실험자가 제기한 문제점으로, 청취자가 특정 채널의 스피커에 가깝게 위치할수록 음원이 해당 스피커의 방향으로 이동하는 이른바 “precedence effect” [7] 로 인하여 간섭 제거의 효과가 명확하게 나타나지 않는 경우가 종종 발생하였다. 이는 본 연구에서 HRTF만으로 음원과 귀간의 경로를 모델링하며, 음원-귀 간의 거리에 따른 음압 레벨의 감소가 고려되지 않는 것에 그 원인이 있는 것으로 풀이된다. 또한 피실험자에 따른 간섭 제거율의 편차가 다소 높게 나타났는데, 각 피실험자별로 개인화된 HRTF

를 사용하지 않는 것에 원인이 있는 것으로 판단된다. 따라서 스피커-청취자 간의 거리와 함께 HRTF의 개인화가 고려되면 주관적 청취 관점에서 더 높은 간섭 제거 성능을 얻을 것으로 기대된다.

IV. 결론

본 논문에서는 좌, 우 2개의 라우드 스피커를 사용하는 스테레오 재생 환경에서 좌, 우 채널에 대한 간섭 신호를 제거하는 기법을 살펴보고, 이를 실시간 하드웨어로 구현하는 과정을 제시하였다. 제안된 시스템은 별도의 복잡한 하드웨어 없이 청취자의 위치를 추적하기 위한 방법으로 청취자가 발생시킨 박수소리를 이용하는 방법을 제안하였다. 수평 방향으로 일정 거리만큼 떨어진 두 개의 마이크로폰에 가해진 두 소리의 시간 지연을 이용하여 청취자의 방향을 추적하도록 하였으며, 잔향 공간에서의 사용을 고려한 지연 시간 추정 기법이 사용되었다.

범용 마이크로프로세서 및 범용 DSP를 이용하여 실시간 시스템을 구현하였으며, 구현된 시스템에 대한 성능 평가를 수행하였다. 성능 평가 결과, 제안된 시스템은 추정된 방향과 실제 방향간의 상관 계수가 0.99 이상, RMSE가 5도로서, 매우 우수한 결과를 나타내었으며, 특히 선형 예측 기법에 근거한 잔향 제거 기법을 통해 무향 환경의 근접한 결과를 얻을 수 있음이 실험 결과를 통해 확인되었다. 주관적인 청취 테스트 결과에서도 제안된 시스템이 간섭 신호를 효과적으로 제거함을 확인할 수 있었다.

본 연구에서는 하드웨어 제작의 편의성 등을 고려하여 청취자 위치를 추적하기 위한 마이크로폰을 전면에 위치시켰는데, 보다 높은 정밀도로 청취자의 위치를 추적하고 특히 2차원 평면상의 청취자 위치를 추적하기 위해서는 마이크로폰의 개수 및 위치에 대한 연구가 더욱 진행되어야 할 것으로 판단된다.

감사의 글

본 논문은 2009학년도 건국대학교의 지원에 의하여 연구되었음.

참고 문헌

1. J. J. Lopez, F. Orduna and A. Gonzalez, "Modeling and measurement of cross-talk cancellation zones for small displacements of the listener in transaural sound reproduction with different loudspeaker arrangements", *AES 109th Convention*, 2000.
2. M. R. Bai and C.-C. Lee, "Comprehensive analysis of loudspeaker span effects on crosstalk cancellation in spatial sound reproduction", *AES 120th Convention*, 2006.
3. J. Rose, P. Nelson, B. Rafaely and T. Takeuchi, "Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations," *J. Acoust. Soc. Am.*, vol. 112, no. 5, pp. 1992-2002, 2002.
4. C. Kyriakakis, T. Holman, J.-S. Lim, H. Hong and H. Neven, "Signal processing, acoustics, and psychoacoustics for high quality desktop audio," *J. Visual Comm. and Image Rep.*, vol. 8, no. 1, pp. 51-61, 1998.
5. P. G. Georgiou, A. Mouchtaris, S. I. Roumeliotis and C. Kyriakakis, "Immersive sound rendering using laser-based tracking", *AES 109th Convention*, 2000.
6. S. Kim, S. Jang, D. Kong and S. Bang, "Adaptive virtual surround sound rendering method for an arbitrary listening position", *AES 30th Int. Conference*, 2007.
7. S. Merchel and S. Groth, "Analysis and implementation of a stereophonic play back system for adjusting the sweet spot to the listener's position", *AES 126th Convention*, 2009.
8. N. Sakamoto, T. Gotoh, T. Kogure and M. Shimbo, "Controlling soundimage localization in stereophonic reproduction", *J. Audio Eng. Soc.*, vol. 29, no. 11, pp. 794-799, 1981.
9. B. Gardner and K. Martin, *KEMAR HRTF data*, <ftp://sound.media.mit.edu/pub/Data/KEMAR> (last viewed 8/21/08), 1994.
10. M. Jian, A. C. Kot and M. H. Er, "Performance study of time delay estimation in a room environment", *Proc. IEEE ICASSP*, pp. V554-V557, 1998.
11. C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay", *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320-327, 1976.
12. J. R. Hopgood and P. J. W. Rayner, "A probabilistic framework for subband autoregressive model applied to room acoustics", *Proc. IEEE Signal Processing Workshop on Statistical Signal Processing*, pp. 492-495, 2001.
13. TMS320C6713 *Floating-point digital signal processor*, Texas Instruments, Nov., 2002.
14. TMS320C6000 *Programmer's Guide*, Texas Instruments, Aug., 2002.
15. O. Kirkeby, P. A. Nelson and H. Hamada, "Fast deconvolution of multichannel systems using regularization", *IEEE Trans. Speech and Audio Process.*, vol. 6, no. 2, pp. 189-194, 1998.

저자 약력

•이 기 승 (Ki-Seung Lee)



1991년 2월: 연세대학교 전자공학과 (공학사)
 1993년 2월: 연세대학교 대학원 전자공학과 (공학석사)
 1997년 2월: 연세대학교 대학원 전자공학과 (공학박사)
 1997년 3월 ~ 1997년 9월: 연세대학교 신호처리 연구 센터 선임 연구원
 1997년 10월 ~ 2000년 9월: AT&T Shannon Lab 연구원
 2000년 11월 ~ 2001년 8월: 삼성종합기술원 HCI Lab 전문연구원
 2001년 9월 ~ 2007년 8월: 건국대학교 정보통신대학 전자공학부 조교수
 2007년 9월 ~ 2010년 8월: 건국대학교 정보통신대학 전자공학부 부교수
 2010년 9월 ~ : 건국대학교 정보통신대학 전자공학부 교수
 ※ 관심분야: 생체 신호 처리, 오디오 신호 처리 등

•이 석 필 (SeokPhil Lee)



1990년 2월: 연세대학교 전기공학과 (공학사)
 1992년 8월: 연세대학교 대학원 전기공학과 (공학석사)
 1997년 8월: 연세대학교 대학원 전기공학과 (공학박사)
 1997년 7월 ~ 2002년 2월 : 대우전자 (주) 선임연구원
 2002년 3월 ~ 현재 : 전자부품연구원 방송통신융합 연구센터장
 ※ 관심분야: 비디오 신호 처리, 오디오 신호 처리