

3차원 합성 입체음향 환경에서의 음향반향제거기

정회원 김 현 태*, 박 장 식**

An Acoustic Echo Canceled under 3-Dimensional Synthetic Stereo Environments

Hyun-Tae Kim*, Jang-Sik Park** *Regular Members*

요 약

본 논문에서는 다자간 화상회의 시스템에서 합성 입체 음향을 재현하는 방법과 음향반향제거 방법을 제안한다. 합성 입체 음향은 HRTF(head related transfer function)을 이용하여 구현하고 반향제거를 위하여 주변잡음에 대하여 강건한 적응 알고리즘을 제안한다. 제안하는 알고리즘은 SMAP(set-membership affine projection)을 변형한 것으로 적응필터의 계수를 갱신할 때 입력신호와 추정오차신호의 상호상관을 입력신호의 자기상관 행렬과 투영 차수를 곱한 추정오차신호 전력의 합으로 정규화한다. 제안하는 적응알고리즘은 SMAP 알고리즘과 비교하여 투영차수와 주변잡음에 대하여 강건한 특성을 갖는다. 컴퓨터 시뮬레이션을 통해 제안하는 합성 입체음향 반향제거기의 성능이 효과적으로 반향을 제거할 수 있음을 보인다.

Key Words : synthetic stereo environments, acoustic echo canceler, modified SMAP, HRTF.

ABSTRACT

This paper proposes a method of implementing synthetic stereo and an acoustic echo cancellation algorithm for multiple participant conference system. Synthetic stereo is generated by HRTF and two loudspeakers. A robust adaptive algorithm for synthetic stereo echo cancellation is proposed to reduce the weight misalignment due to near-end speech signals and ambient noises. The proposed adaptive algorithm is modified version of SMAP algorithm and the coefficients of adaptive filter is updated with cross correlation of input and estimation error signal normalized with sum of the autocorrelation of input signal and the power of the estimation error signal multiplied with projection order. This is more robust to projection order and ambient noise than conventional SMAP. Computer simulation show that the proposed algorithm effectively attenuates synthetic stereo acoustic echo.

I. 서 론

최근 DSP 기술이 발전함에 따라 음성통신 시스템은 다양하게 발전하고 있다. 원격화상회의에서는 입체음향 기술 등을 적용하여 편리한 음성통화 환경을 제공하기 위한 연구가 진행되고 있다. 입체음향 기술을 활용하여 참여자들의 음성을 각각 다른 위치에 정위시킴으로써 발언자를 쉽게 확인할 수 있어

효율적으로 원격화상회의의 진행할 수 있다. 원격화상회의에서는 입체음향을 다채널 방식으로 구현하고 있으며 편리하게 회의를 진행하기 위하여 핸드프리 단말기를 사용하고 있다. 핸드프리 단말기를 이용한 음성통신에서는 음향반향(acoustic echo)이 생겨 통화의 품질이 저하된다. 음향반향은 스피커를 통하여 출력된 상대방의 음성이 마이크를 통하여 입력되어 상대방에게 재전송되어 발생한다. 입체음향 기술을

* 동의대학교 멀티미디어공학과(htaekim@dongeui.ac.kr), ** 동의공업대학 영상정보과(jsipark@dit.ac.kr)

논문번호 : 030244-0526, 접수일자 : 2003년 5월 26일

*이 논문은 2002년도 한국학술진흥재단의 지원에 의하여 연구되었음(KRF-2002-003-D00243)

포함한 고품질의 음성서비스를 위해서는 음향반향은 반드시 제거되어야 하기 때문에 입체음향 환경에서의 반향제거를 위한 연구가 진행되고 있다[1-3].

Yensen 등은 원단화자의 음성신호를 4 채널 공간화합수를 이용하여 참여자의 음성을 특정 위치에 정위시키고 다채널 음향반향을 제거하기 위하여 채널간의 상호상관도에 의하여 성능이 저하되지 않는 음향반향제거기 구조를 제안하였다[2,3]. 적응알고리즘은 간단하고 안정적인 수렴을 하는 NLMS 알고리즘을 사용하였다. 그러나 NLMS 알고리즘은 음성신호와 같은 유색신호에 대하여 수렴속도가 느려진다. 최근 반향제거를 위하여 유색신호에 대하여 수렴속도가 빠른 AP(affine projection) 알고리즘[4-6] 혹은 SMAP 알고리즘이 적응알고리즘으로 제안되고 있다[7]. 그러나 수렴속도와 계산량 측면에서 AP 혹은 SMAP 알고리즘이 효과적이지만 투영 과정에서 추정오차신호 속에 포함된 주변잡음에 의한 잡음증폭 현상이 있어 반향제거 성능이 저하된다[6].

본 논문에서는 다자간 원격화상회의를 위하여 HRTF를 이용하여 2 채널만으로 입체음향을 재현하고 합성 입체음향 반향을 제거하기 위한 적응알고리즘을 제안한다. 2 개의 스피커만으로 입체음향을 구현함으로써 사용자가 편리하게 설치할 수 있을 뿐만 아니라 헤드폰 혹은 이어폰을 착용한 상태에서 가상음원을 적용할 수 있다[8-11]. 합성 입체음향의 반향은 원단화자의 음성신호가 HRTF와 반향경로를 거쳐 마이크로 입력되어 발생한다. 따라서 합성 입체음향 환경에서 요구되어지는 음향반향제거기는 HRTF와 반향경로를 동시에 추정하여야 반향을 적절히 제거할 수 있다. 본 논문에서는 유색신호에 대하여 수렴이 빠르면서 주변잡음에 강건한 특성을 갖는 적응알고리즘을 제안한다. 제안하는 적응알고리즘은 SMAP 알고리즘을 정성적으로 분석하여 계수 오조정이 투영차수와 주변오차신호의 분산에 비례함으로 확인하고 입력신호의 자기상관과 투영차수를 곱한 추정오차신호의 전력을 더하여 정규화함으로써 주변잡음에 강건하도록 한다. 시뮬레이션 결과 제안하는 알고리즘이 대표적인 주변잡음 신호인 근단화자의 음성신호에 대하여 계수오조정이 감소되어 안정적으로 반향을 제거할 수 있음을 확인하였다.

II. 합성 입체음향 반향제거기

원격화상회의에서 합성 입체음향은 단일 채널로 전송

된 신호를 공간화합수를 통과시켜 근단화자를 중심으로 일정한 위치에 가상음원을 정위시키는 것이다. 원격화상회의를 위한 합성 입체음향 시스템은 1 개의 마이크와 2 개 이상의 스피커로 구성할 수 있다. 그림 1은 4 개의 스피커를 이용하여 5 영역의 가상음원을 설정하는 예이다. 참여자들의 음성신호를 공간화합수, $G_{i,FL}$, $G_{i,FR}$, $G_{i,BL}$ 그리고 $G_{i,BR}$ 를 이용하여 근단화자를 중심으로 임의의 5 영역에 가상음원을 정위시킨다. $G_{i,FL}$, $G_{i,FR}$, $G_{i,BL}$ 그리고 $G_{i,BR}$ 는 각각 앞쪽의 왼쪽과 오른쪽 그리고 뒷쪽의 왼쪽, 오른쪽 스피커를 위한 공간화합수이며 $i=1, \dots, 5$ 는 가상음원을 나타내는 첨자이다. 근단화자는 참여자들의 음성이 각각 다른 방향에서 들리므로 화자를 쉽게 인식할 수 있어 회의의 진행이 편리하게 된다.

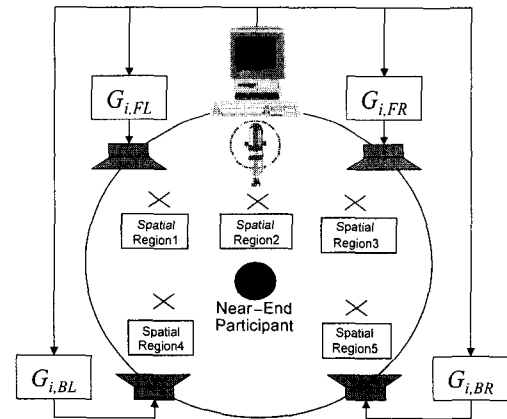


그림 1. 합성 입체음향을 이용한 원격화상회의의 개념도

그런데 그림 1과 같은 핸드프리 환경에서는 4 개의 스피커에서 출력된 소리가 근단화자의 마이크로 입력되어 반향이 생긴다. 통화의 상대방인 원단화자는 일정 시간 후 자신의 음성을 다시 듣게 됨으로 통화가 부자연스럽게 된다. 따라서 음향반향제거기는 전이중 방식(full-duplex)의 핸드프리 통신 시스템에서는 필수적이다.

그림 2는 Yensen 등이 제안한 원격화상회의를 위한 4 채널 입체음향 재생 시스템과 입체음향 반향을 제거하는 구성도이다. $s_j(n)$ 참여자의 각 음성신호이며 $v_i(n)$ 는 특정한 공간으로 사상된(mapping) 음성신호이다. i 와 j 는 각각 M 영역의 가상음원과 N 명 참여자를 나타내는 첨자이다.

G_i 는 단일 채널신호인 $v_i(n)$ 를 i 위치에 정위시키는 공간화 함수로써 $G_{i,FL}$, $G_{i,FR}$, $G_{i,BL}$ 그리고 $G_{i,BR}$ 로 구성된 행렬이다.

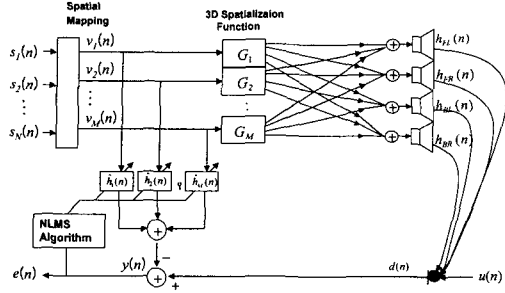


그림 2. 4 채널 입체음향 재생 시스템과 음향반향제거기

$h_{FL}(n)$, $h_{FR}(n)$, $h_{BL}(n)$ 그리고 $h_{BR}(n)$ 는 스피커에서 마이크 사이의 반향경로를 나타낸다.

$\hat{h}_i(n)$ 는 반향경로와 공간화함수를 모델링하는 적응필터이며 $d(n)$ 과 $e(n)$ 은 마이크 입력신호와 추정오차신호이다. $u(n)$ 과 $y(n)$ 은 각각 근단화자의 음성신호를 포함한 주변잡음 신호와 적응필터의 출력신호이다. Yensen 등은 적응필터 $\hat{h}_i(n)$ 를 식 (1)과 같이 NLMS 알고리즘으로 계수를 갱신하였다.

$$\hat{h}_i(n) = \hat{h}_i(n) + \mu(n)e(n)v_i(n), \quad i = 1, \dots, M \quad (1)$$

$$\mu(n) = \frac{\alpha}{L \sum_{i=1}^M \sigma_{v_i}(n)}$$

α 는 적응상수이며 L 은 적응필터의 차수이다. $\sigma_{v_i}(n)$ 은 사상된 입력신호의 $v_i(n)$ 의 분산 즉, 전력이다. $\hat{h}_i(n)$ 은 적응필터의 계수벡터이며 $v_i(n)$ 은 사상된 입력신호 벡터인 동시에 적응필터의 입력신호 벡터이다.

Yensen 등이 제안한 합성 입체음향 반향제거기는 사상된 입력신호 $v_i(n)$, 즉 상호상관이 없는 신호를 적응필터에 입력시켜 적응시킴으로써 최적해에 수렴할 수 있도록 하였다. 그러나 일반적으로 NLMS 알고리즘으로 적응필터를 적용하는 경우 수렴속도가 늦고 근단화자 등의 주변잡음신호에 의하여 적응필터의 계수가 오조정되어 반향제거 성능이

크게 저하된다. 2차 이상의 투영차수를 갖는 AP 혹은 SMAP 알고리즘으로 적응필터를 적응시킴으로써 수렴속도는 개선할 수 있지만 주변잡음에 의한 영향이 투영 차수가 증가함에 따라 증가한다.

III. 제안하는 합성 입체음향 재생과 반향 제거기

다채널 입체음향을 구현하기 위해서는 사용자가 여러 개의 스피커를 배치하여야 하기 때문에 불편한 점이 있다. HRTF를 이용하면 사람이 음원의 위치를 인지하는 것과 같은 원리로 입체음향을 재생할 수 있다. HRTF를 이용하면 두 개의 스피커를 배치하여 입체음향을 재현할 수 있기 때문에 설치하는데 있어 편리할 뿐만 아니라 헤드셋을 착용한 상태에서도 가상음원을 특정한 위치에 정위시킬 수 있다.

본 논문에서는 원격화상회의를 위한 합성 입체음향을 HRTF를 이용하여 구현하고 입체 음향반향제거를 위한 적응알고리즘을 제안한다. 그림 3은 제안하는 원격화상회의를 위한 합성 입체음향 시스템과 음향반향제거기의 구조이다. $s_j(n)(j=1, \dots, N)$ 는 참여자의 각 음성신호이며 $v_i(n)(i=1, \dots, M)$ 는 특정한 공간으로 사상된 음성신호이다. $G_i(i=1, \dots, M)$ 는 가상음원을 정위시키기 위한 HRTF이며 왼쪽 스피커와 오른쪽 스피커를 위한 임펄스응답 $g_{iL}(n)$, $g_{iR}(n)$ 로 구성된다. 스피커 출력신호 $x_L(n)$, $x_R(n)$ 은 아래 식과 같이 사상된 음성신호 $v_i(n)$ 과 공간화함수 HRTF로부터 계산된다.

$$x_L(n) = \sum_{i=1}^M v_i^T(n) g_{iL}(n) \quad (2)$$

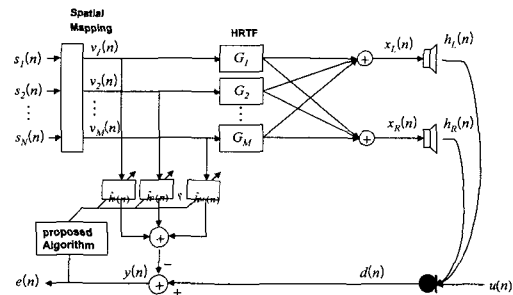


그림 3. 제안하는 HRTF를 이용한 합성 스테레오 반향 제거기

$$x_R(n) = \sum_{i=1}^M v_i^T(n) \mathbf{g}_{iR}(n) \quad (3)$$

$\mathbf{g}_{iL}(n)$, $\mathbf{g}_{iR}(n)$ 그리고 $v_i(n)$ 는 아래와 같다.

$$\mathbf{g}_{iL}(n) = [g_{iL}(n) \ g_{iL}(n-1) \ \dots \ g_{iL}(n-L+1)]^T \quad (4)$$

$$\mathbf{g}_{iR}(n) = [g_{iR}(n) \ g_{iR}(n-1) \ \dots \ g_{iR}(n-L+1)]^T \quad (5)$$

$$v_i(n) = [v_i(n) \ v_i(n-1) \ \dots \ v_i(n-L+1)]^T \quad (6)$$

L 은 임펄스응답의 길이이다. $d(n)$ 은 마이크 입력 신호로써 식(7)과 같이 구하여 진다.

$$d(n) = \mathbf{x}_L^T(n) \mathbf{h}_L(n) + \mathbf{x}_R^T(n) \mathbf{h}_R(n) + u(n) \quad (7)$$

여기서 $u(n)$ 은 근단화자를 포함한 주변잡음 신호이며 $\mathbf{x}_L(n)$, $\mathbf{x}_R(n)$ 은 아래 식과 같이 표현되는 왼쪽과 오른쪽의 스피커 출력신호 벡터이며

$$\mathbf{x}_L(n) = [x_L(n) \ x_L(n-1) \ \dots \ x_L(n-L+1)]^T \quad (8)$$

$$\mathbf{x}_R(n) = [x_R(n) \ x_R(n-1) \ \dots \ x_R(n-L+1)]^T \quad (9)$$

$\mathbf{h}_L(n)$ 과 $\mathbf{h}_R(n)$ 은 각각 왼쪽과 오른쪽 스피커로부터 마이크 사이의 음향반향경로를 의미한다. 오차 신호 $e(n)$ 은

$$e(n) = d(n) - \sum_{i=1}^M \hat{\mathbf{h}}_i^T(n) v_i(n) \quad (10)$$

가 된다. $\hat{\mathbf{h}}_i(n)$ 는 공간영역 i 에 대한 음향 반향 경로를 추정하는 적응필터의 계수벡터이다. 음향반

향제거기의 적응필터 계수벡터 $\hat{\mathbf{h}}_i(n)$ 는 사상된 입력신호 $v_i(n)$ 과 추정오차신호 $e(n)$ 을 이용하여 $\mathbf{h}_L(n)$, $\mathbf{h}_R(n)$, $\mathbf{g}_{iL}(n)$ 그리고 $\mathbf{g}_{iR}(n)$ 를 추정하여야 한다

일반적으로 스테레오 음향반향제거기에서 두 입력신호에 의한 상호상관에 의하여 적응필터가 최적해로 수렴하지 못하는 문제가 있지만 그림 3과 같은 구조에서는 $v_i(n)$ 을 적응필터에 입력함으로써 각 적응필터 입력신호 간의 상관관계가 없기 때문에 최적해로 수렴할 수 있다.

적응필터의 계수벡터 $\hat{\mathbf{h}}_i(n)$ 의 적응은 식(11)과 같이 SMAP 알고리즘으로 적용시킬 수 있다.

$$\hat{\mathbf{h}}_i(n+1) \mapsto \begin{cases} \hat{\mathbf{h}}_i(n) + \mu V_i(n) \mathbf{g}(n), & \text{if } |e(n)| > \lambda \\ \hat{\mathbf{h}}_i(n), & \text{otherwise} \end{cases} \quad (11)$$

여기서, μ 는 적응상수이며 p 는 투영차수이다. $V_i(n)$ 는 $L \times p$ 행렬로써 식(12), (13)과 같이 정의된다.

$$V_i(n) = [v_i(n) \ v_i(n-1) \ \dots \ v_i(n-p+1)] \quad (12)$$

$$v_i(n) = [v_i(n) \ v_i(n-1) \ \dots \ v_i(n-L+1)]^T \quad (13)$$

정규화 오차신호벡터 $\mathbf{g}(n)$ 은 다음 식과 같이 계산되고

$$\mathbf{g}(n) = (\mathbf{R}_i(n) + \delta \mathbf{I})^{-1} \mathbf{e}(n) \quad (14)$$

δ 는 특이해를 갖지 않도록 하기 위한 작은 양수이다. $\mathbf{R}_i(n)$ 과 $\mathbf{e}(n)$ 은 다음과 같이 정의된다.

$$\mathbf{R}_i(n) = \sum_{i=1}^M V_i(n) V_i^T(n) \quad (15)$$

$$\mathbf{e}(n) = [e(n) \ e(n-1) \ \dots \ e(n-p+1)]^T \quad (16)$$

식 (7), (10) 그리고 (14)를 식 (11)의 적응필터 계수벡터를 갱신하는 조건에 대입하면 하여 정리하면

$$\mathbf{w}(n+1) = [I - \mu \mathbf{V}_i(n) \mathbf{R}_i^{-1}(n) \mathbf{V}_i^T(n)] \mathbf{w}(n) + \mu \mathbf{V}_i(n) \mathbf{R}_i^{-1}(n) \mathbf{U}(n) \quad (17)$$

$$\mathbf{w}(n) = \hat{\mathbf{h}}_i(n) - \mathbf{h}_{i,o}$$

$$\mathbf{U}(n) = [u(n) \ u(n-1) \ \dots \ u(n-p+1)]^T$$

된다. $\mathbf{w}(n)$ 은 오조정 벡터(misalignment vector)이며 $\mathbf{h}_{i,o}$ 는 $\hat{\mathbf{h}}_i(n)$ 이 추정하여야 하는 최적 계수벡터이다. 식 (17)에서 두 번째 항, $\mu \mathbf{V}_i(n) \mathbf{R}_i^{-1}(n) \mathbf{U}(n)$ 이 계수를 오조정시키는 요소가 된다. 계수 오조정을 발생시키는 요소를

$$\mathbf{Q}(n) = \mu \mathbf{V}_i(n) \mathbf{R}_i^{-1}(n) \mathbf{U}(n) \quad (18)$$

로 정의하고 $u(n)$ 이 평균이 0 인 백색 Gaussian 신호라고 가정하면, $\mathbf{Q}(n)$ 의 분산은

$$\begin{aligned} E[\mathbf{Q}(n)^2] &= \mu^2 E[(\mathbf{V}_i(n) \mathbf{R}_i^{-1}(n) \mathbf{U}(n)) \\ &\quad * (\mathbf{V}_i(n) \mathbf{R}_i^{-1}(n) \mathbf{U}(n))^T] \\ &= \mu^2 p \mathbf{R}_i(n)^{-1} \sigma_u^2 \end{aligned} \quad (19)$$

이 된다. 따라서 SMAP 알고리즘의 과잉평균자승 오차(excess mean square error)은 투영차수와 주변잡음 신호의 전력에 비례하여 증가하는 것을 식 (19)에서 확인할 수 있다. 그림 4는 25000 샘플과 30000 샘플 사이 그리고 50000 샘플과 55000 샘플 사이에 각각 분산이 1, 2 인 백색잡음을 주변잡음 신호로 입력하여 시뮬레이션한 결과이다. 수렴 초기에는 투영차수가 높은 것이 빠르게 수렴하지만 25000 샘플에서 30000 샘플 그리고 50000 샘플과 55000 샘플 구간에서는 투영차수가 높을수록 계수 오조정이 크지는 것을 알 수 있다.

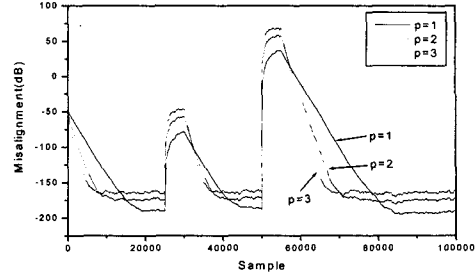


그림 4. SMAP 알고리즘의 투영차수에 따른 계수 오조정 지수

따라서 적응상수를 조절하거나 정규화하는 식 (14)에서 투영차수와 주변잡음을 고려하여 정규화하는 것이 필요하다. 식 (20)과 (21)로 적응 및 정규화를 하여 적응필터의 계수벡터가 오조정되는 것을 줄인다.

$$\hat{\mathbf{h}}_i(n+1) \mapsto \begin{cases} \hat{\mathbf{h}}_i(n) + \mu \mathbf{V}_i^T(n) \mathbf{g}(n), & \text{if } |e(n)| > \lambda \\ \hat{\mathbf{h}}_i(n), & \text{otherwise} \end{cases} \quad (20)$$

$$\mathbf{g}(n) = (\mathbf{R}_i(n) + pL\sigma_e^2 \mathbf{I})^{-1} \mathbf{e}(n) \quad (21)$$

추정오차신호의 분산인 σ_e^2 에는 식 (7)과 식(10)을 통하여 주변잡음신호 성분인 $u(n)$ 을 포함하고 있다. 특히 $u(n)$ 과 입력신호 $v_i(n)$ 이 독립이고 적응필터가 수렴하였다고 가정하면 주변잡음신호의 전력, σ_u^2 이 크면 정규화가 크게 되어 오조정을 줄일 수 있다. 그림 5는 같은 조건으로 제안하는 알고리즘으로 시뮬레이션한 결과이다. 제안하는 알고리즘은 기존의 SMAP 에 비하여 주변잡음신호에 의한 계수 오조정이 현저히 감소된 것을 확인할 수 있다.

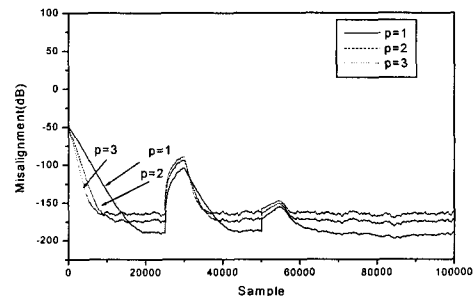


그림 5. 제안하는 적응알고리즘의 투영차수에 따른 계수오조정 지수

IV. 시뮬레이션 결과 및 검토

제안하는 합성 입체음향 반향제거기의 성능을 확인하기 위하여 다자간 토론 환경에서 주변잡음신호가 없는 상황과 큰 주변잡음신호, 근단화자의 음성신호가 있는 동시통화 상황에 대하여 시뮬레이션을 통하여 확인한다.

주변잡음신호가 없는 상황은 다자간 토론 환경을 서로 다른 가상음원에 할당된 여성과 남성의 음성신호를 다른 기간 동안 발생시킨다. 기본적인 반향제거 성능을 확인하고 하나의 가상음원에서 또 다른 가상음원으로 변환됨에 따라 반향제거기의 성능의 변화 확인하기 위한 것이다. 합성 입체음향 시스템의 입력신호는 모두 16 kHz로 표본화하고 16 비트로 양자화한 음성 신호로써 여성 음성신호는 “치료와 회복에 도움이 될 수도 있다.” 라고 문장이며 남성 음성신호는 “그 때만해도 엄마 등에 업혔을 때하고는 달리 서러움을 적당히 고조시키고 싶어 피까지 섰다.” 라는 문장이다.

수신실의 반향경로 h_L , h_R 은 2 채널 DAT 레코더로 회의실에서 측정하였으며 임펄스응답의 길이는 1024 이며 그림 6과 같다.

HRTF를 이용한 가상 음원의 위치는 근단화자를 중심으로 각각 30°, 330° 위치에 설정한다. 임펄스 응답은 MIT의 Media lab에서 측정된 것으로 KEMAR 더미헤드 마이크를 이용하여 측정된 것이다. 그림 7의 (a)~(d)는 각각 고도 0°에서 측정된 방위각 30°와 330°에서 측정된 길이 512 의 HRTF 응답으로 수신실의 왼쪽과 오른쪽 스피커로 입력되어 정위감을 준다. 투영차수, $p=2$ 로 두고 제안하는 적응알고리즘을 기존 AP와 SMAP 알고리즘과 비교한다.

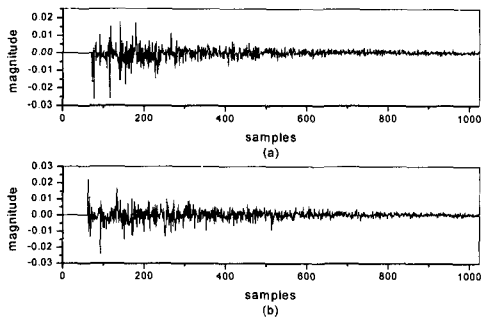


그림 6. 수신실의 스피커에서 마이크까지의 반향경로 임펄스 응답
 (a) h_L , 왼쪽 스피커에서 마이크까지의 반향경로 임펄스 응답
 (b) h_R , 오른쪽 스피커에서 마이크까지의 반향경로 임펄스 응답

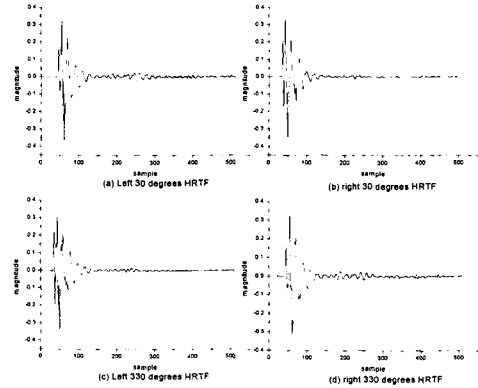


그림 7. 방위각 30°과 330° 에 가상음원을 정위시키기 위한 HRTF
 (a) 방위각 30° 에 정위시키기 위한 왼쪽 스피커 신호의 HRTF
 (b) 방위각 30° 에 정위시키기 위한 오른쪽 스피커 신호의 HRTF
 (c) 방위각 330°에 정위시키기 위한 왼쪽 스피커 신호의 HRTF
 (d) 방위각 330°에 정위시키기 위한 오른쪽 스피커 신호의 HRTF

입력신호의 전력은 모두 running power estimate 로 추정하였으며 망각지수(forgetting factor)는 0.998로 두었다. 알고리즘의 적응상수는 0.5 로 두었다. 적응필터의 길이는 HRTF와 수신실의 임펄스 응답을 콘벌루션한 것과 같이 두었다. 음향반향제거기의 성능 평가를 위하여 ERLE(echo return loss enhancement)와 계수 오조정을 사용한다. ERLE는 식(22)과 같이 계산한다.

$$ERLE(k) = 10 \log_{10} \frac{\sum_{i=0}^{L-1} d^2(k-i)}{\sum_{i=0}^{L-1} (d(k-i) - y(k-i))^2} \quad (22)$$

ERLE가 클수록 반향 제거 성능이 우수하다. 그리고 계수 오조정은 식(23)와 같이 계산한다.

$$misalignment = 10 \log_{10} \frac{\|h_{1,o}(k) - \hat{h}_1(k)\|^2 + \|h_{2,o}(k) - \hat{h}_2(k)\|^2}{\|h_{1,o}(k)\|^2 + \|h_{2,o}(k)\|^2} \quad (23)$$

계수 오조정은 HRTF 및 반향경로의 임펄스 응답과 적응필터의 계수와와의 오조정 정도를 나타낸다. 그림 7 과 8은 주변잡음신호가 없을 때 제안하는 알고리즘과 기존 적응알고리즘의 성능을 비교한 결과이다. 근단

화자를 중심으로 30° 방위각에 위치하도록 여성 참여자의 음성을 정위시키고 40000 샘플 이후부터 330° 방위각에 남성 참여자의 음성신호를 정위시킨다.

그림 8의 (a)는 ERLE 이며 (b)는 계수오조정 지수를 비교한 것이다. 제안하는 알고리즘과 기존의 적응알고리즘들이 주변잡음신호가 없는 경우에는 비슷한 성능을 나타낸다. 그림 8의 (b)에서 계수오조정 지수가 40000 샘플까지 크게 감소하지 않는 것은 40000 샘플이후부터 방위각 330°에 해당하는 HRTF 함수 및 반향경로가 추정되기 때문이다.

그림 9는 각 알고리즘을 적용하여 반향제거를 한 출력신호인 잔여반향을 나타낸 것이다. (a)는 마이크로 입력된 반향신호이며 (b), (c), (d)는 각각 AP, SMAP 그리고 제안하는 적응알고리즘의 결과로써 비슷한 크기의 잔여반향이 있다. 따라서 제안하는 알고리즘은 주변잡음 신호가 없을 때는 기존의 알고리즘과 비슷한 성능을 보임을 확인할 수 있다.

원격화상회의에 참여하고 있는 참여자 중에서 원단화자의 음성신호가 있을 때 근단화자의 음성신호 혹은 주변잡음신호가 있는 동시통화 환경에 대하여 시뮬레이션하고 그 결과를 비교한다. 시뮬레이션에

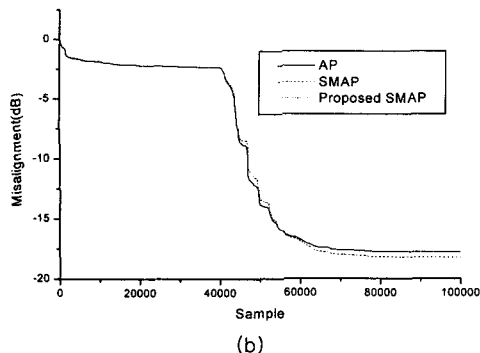
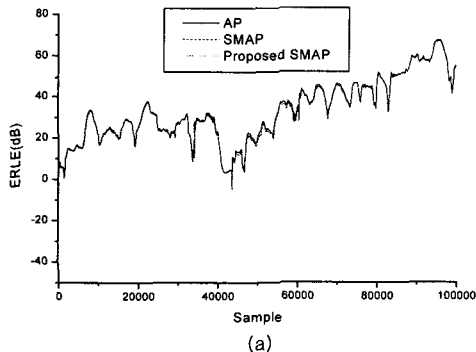


그림 8. 주변잡음 신호가 없을 때 ERLE와 계수 오조정 (P=2) (a) ERLE의 비교 (b) 계수오조정 지수의 비교

서 여성 음성신호는 동일하며 근단화자의 음성신호는 “운영 관리실” 이며 남성의 음성신호를 사용하였다.

그림 10와 11에서는 동시통화에 대하여 제안하는 적응알고리즘과 기존의 적응알고리즘의 성능을 비교한다. 각 알고리즘의 투영 차수 p 를 2로 두고 시뮬레이션하였다. 근단화자를 중심으로 30° 방위각에 여성 참여자의 음성신호를 정위시키고 근단화자의 음성신호는 7500 샘플부터 14000 샘플까지 발생하도록 하였다. 그림 9는 ERLE와 계수오조정 지수를 비교한 것이다. (a)는 ERLE를 비교한 것으로 동시통화구간에서는 근단화자의 음성신호가 크기 때문에 0 dB 가 되는 것이 정상적인 것이다. 근단화자의 음성신호가 있는 구간에서 AP와 SMAP 알고리즘의 ERLE가 높지만 실질적으로 근단화자의 음성신호가 왜곡되고 있다. 그리고 동시통화 이후에도 AP 알고리즘은 SMAP와 제안하는 알고리즘에 비하여 성능이 상당히 좋지 않다.

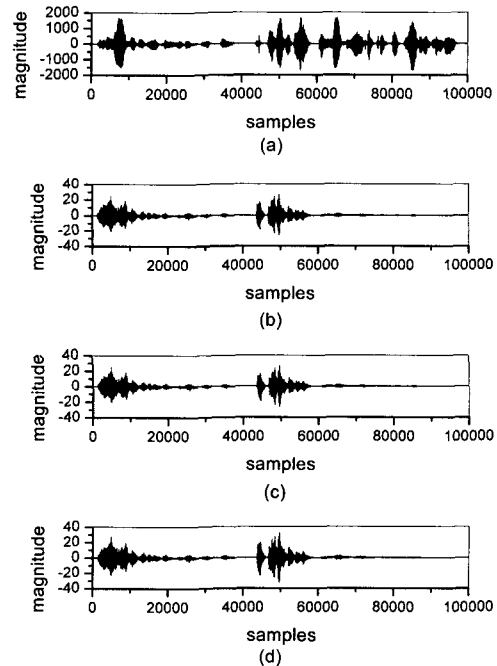


그림 9. 주변잡음신호가 없는 상황에서는 음향반향제거 시뮬레이션 결과 (a) 반향신호 (b) AP 알고리즘의 잔여반향 (c) SMAP 알고리즘의 잔여반향 (d) 제안하는 알고리즘의 잔여반향

(b)는 각 알고리즘의 계수오조정 지수를 나타낸 것으로 기존의 적응알고리즘은 근단화자의 음성신호가

있는 구간에서부터 계수 오조정이 커지만 제안하는 알고리즘은 계수오조정이 커지지 않았으며 제안하는 적응알고리즘은 동시통화가 끝난 뒤에도 적응이 안정적으로 이루어지고 있다.

그림 11은 각 알고리즘으로 반향제거를 한 출력신호이다. (a)는 마이크로 입력되어 혼합된 근단화자의 음성신호와 반향신호이다. (b), (c) 그리고 (d)는 각각 AP, SMAP 그리고 제안하는 적응알고리즘의 출력신호이다. 7500 샘플부터 14000 샘플 사이의 근단화자의 음성신호는 왜곡없이 그대로 출력되어야 한다. AP 와 SMAP 알고리즘의 출력신호에서는 근단화자 신호가 많이 왜곡되었으며 근단화자의 음성신호가 끝난 뒤에도 근단화자의 음성신호에 의하여 적응필터의 계수가 오조정이 되어 음향반향이 제거되지 않는 것을 확인할 수 있다. 그러나 제안하는 적응알고리즘의 출력신호에는 근단화자의 음성신호가 거의 왜곡되지 않았으며 동시통화가 지난 후에도 음향반향이 정상적으로 제거되고 있음을 확인할 수 있다. 따라서 제안하는 알고리즘은 주변잡음신호가 없을 때는 기존의 적응알고리즘과 비슷한 성능을 나타내며 동시통화 등의 주변잡음 신호가 있는 경우는 기존의 적응알고리즘에 비하여 계수오조정이 작고 반향제거 성능이 우수함을 알 수 있다.

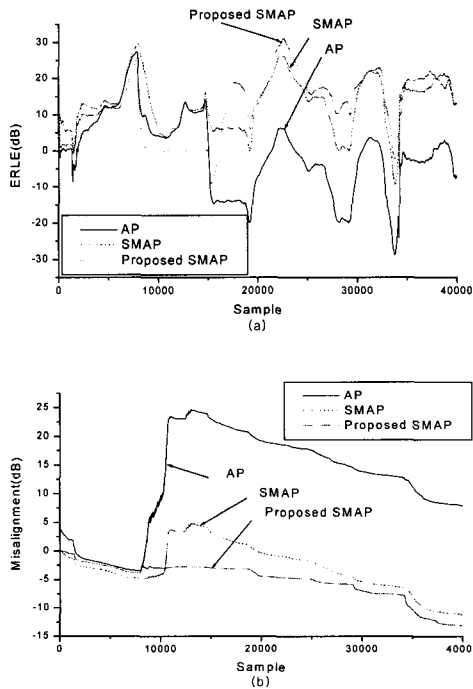


그림 10. 동시통화에서 ERLE와 계수 오조정(P=2)
(a) ERLE의 비교 (b) 계수오조정 지수의 비교

V. 결론

합성 입체음향은 현장감 또는 공간감이 요구되는 가상현실 분야 뿐 아니라 VoIP를 통한 다자간 원격화상회의에서 참여자들의 음성을 각각 다른 가상 음원에 설정함으로써 회의의 효과를 높일 수 있다. 본 논문에서는 현장감 있는 VoIP 환경의 구현을 위해 HRTF를 이용하여 입체음향을 구현하고 렌즈 프리 환경에서의 음향반향제거를 위한 적응알고리즘을 제안하였다.

제안하는 알고리즘은 기존의 SMAP 알고리즘을 변형하여 오차신호의 전력과 투영 차수의 곱을 입력신호의 전력에 더하여 입력신호와 추정오차신호 벡터의 상호상관을 정규화하여 계수를 갱신한다. 제안한 알고리즘은 기존 알고리즘과 비슷한 수렴 속도를 가지며 동시통화 등의 주변잡음신호가 있는 경우 기존의 알고리즘에 비해 월등히 향상된 성능이 우수함을 보였다.

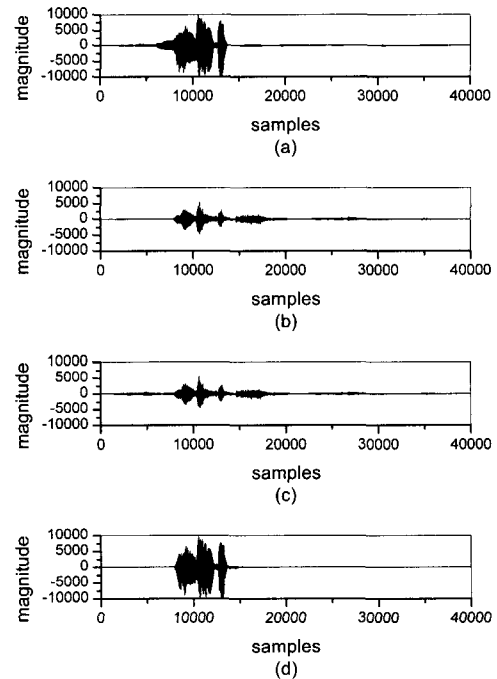


그림 11. 동시통화 상황에서 반향제거기 출력 신호 비교
(a) 반향신호
(b) AP 알고리즘의 출력신호
(c) SMAP 알고리즘의 출력신호
(d) 제안하는 알고리즘의 출력신호

참 고 문 헌

[1] J. Benesty, D. R. Morgan, J. L. Hall and M. M. Sondhi, "Synthesized stereo combined with acoustic echo cancellation for desktop conference," *Proc. on IEEE International Conference on Acoustics, Speech and Signal Processing* 1999, pp. 853-856, 1999

[2] T. Yensen, R. Gourbran, and I. Lambadaris, "Synthetic stereo acoustic echo cancellation structure for multiple participant VoIP conferences," *IEEE Trans. on Speech and Audio Processing*, VOL. 9, NO. 3, Feb. 2001.

[3] T. Yensen and R. Goubran, "An acoustic echo cancellation structure for synthetic surround sound," *Proc. International Conference on Acoustics, Speech and Signal Processing 2001*, Salt Lake, pp.312-315, May, 2001.

[4] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commum. Jpn. A*, VOL. 67, pp. 126-132, 1984.

[5] S. Makino, "The past, present, and future of audio signal processing-acoustic echo cancellation," *IEEE Signal Processing Magazine*, pp.39-41, Sep., 1997.

[6] M. Tanaka, Y. Kaneda S. Makino and J. Kojima, "Fast Projection Algorithm and Its Step Size Control," *Proc. on International Conference on Acoustics, Speech and Signal Processing 95*, Detroit Michigan, vol. 2, pp. 945-948, 1995.

[7] Stefan Werner, Paulo S. R. Diniz, "Set-Membership Affine Projection Algorithm," *IEEE Signal Processing Letters*, VOL. 8, NO. 8, Aug., 2001.

[8] C. Kyriakakis, P. Tsakalides and T. Holman, "Acquisition and rendering methods for immersive audio surrounded by sound," *IEEE Signal Processing Magazine*, pp. 55-66, Jan. 1999.

[9] Kraemer A., "Two speakers are better than 5.1 [surround sound] ", *IEEE Spectrum*, Volume: 38 Issue: 5, May 2001.

[10] H. Moller, et. al. "Head-Related Transfer Functions of Human Subjects", *J. Audio Eng. Soc.*, Vol.43, pp.300-321, 1995

[11] Chong-Jin Tan and Woon-Seng Gan, "User-defined spectral manipulation of HRTF for improved localisation in 3D sound systems", *ELECTRONIC LETTERS*, Vol. 34, No. 25, Dec. 10, 1998.

김 현 태(Hyun-Tae Kim)

정회원



1989년 2월 부산대학교 전자공학과(학사)

1995년 2월 부산대학교 전자공학과(석사)

2000년 2월 부산대학교 전자공학과(박사)

2002년 3월 ~ 현재 동의대학교

멀티미디어공학과 전임강사

<주관심분야> 음성 및 음향신호처리, 멀티미디어 신호처리

박 장 식(Jang-Sik Park)

정회원



1992년 2월 부산대학교 전자공학과(학사)

1994년 2월 부산대학교 전자공학과(석사)

1999년 2월 부산대학교 전자공학과(박사)

1997년 3월 ~ 현재 동의공업대학 영상정보과 부교수

<주관심분야> 음성 및 음향신호처리, 멀티미디어통신, 입체음향