3차원 입체음향 환경에서의 음향반향제거기

성창숙, 김현태*, 박장식**, 손경식
부산대학교 전자공학과, 동의대학교 멀티미디어공학과, 동의공업대학 영상정보과

Acoustic Echo Canceler under 3-Dimensional Synthetic Stereo Environments

Chang-Sook Sung, Hyun-Tae Kim, Jang-Sik Park, Kyung-Sik Son
Dept. of Electronics Eng., Pusan National University
Dept. of Multimedia Eng., Dongeui University
Dept. of Visual Technologies, Dongeui Institute of Technology

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 재현하고 합성 입체 음향반향제거를 위하여 AP(affine projection) 알고리즘을 이용하여 3차원 입체음향 반향 제거 방법을 제안한다. 컴퓨터 시뮬레이션 결과 제안하는 합성 입체 음향반향제거기가 효과적으로 반향을 제거할 수 있음을 보인다.

1. 서론

통신 시스템의 발전으로 다양한 음성 및 영상 서비스가 제공되고 있다. 음성 및 음향 분야에서는 몰입형 음향 시스템(immersive audio system)이 현장감 있는 음성 통신 뿐만 아니라 가상현실, 오락, 텔레비전, 영화 등에 활용되어 가고 있다[1]. 그리고 편리하고 안전한 통화를 위하여 핸즈프리(hands-free) 단말기를 이용한 음성 통신이 원거리 회의, 차량용 핸즈프리 등에 활용되고 있다.

다수가 참여하는 원격 가상회의에서 청취자는 음성만으로는 참여자 중에서는 누가 발언을 하고 있는지 구별하기 어려운 경우가 있다. 특히 청취자가 화자의 목소리에 익숙하지 않다면 구별하기가 더욱 어려워진다. 임의의 가상 음원의 위치에서 재생될 수 있도록 하는 3차원 합성 입체음향은 다자가 참여하는 원격 가상회의에서 청취자가 참여자를 구별하거나 공간감을 부여함으로 회의의 효과를 높일 수 있다[2-4].

VoIP를 포함한 핸즈프리 음성 통신에서는 스피커로 전달되는 상대방의 음성이 마이크를 통해 상대방에게 재전송되어 상대방은 일정 시간 뒤에 자신의 음성을 다시 듣게 된다. 이러한 현상을 음향 반향(acoustic echo)이라 하며 통화의 불편과 통화 품질을 저하시켜 고품질의 음성서비스를 위해서는 음향 반향을 제거하여야한다. 음향 반향을 제거하기 위한 연구가 다양하게 진행되고 있으며 최근에는 합성 입체음향 환경에서의 반향 제거를 위한 연구가 진행되고 있다[2-4].

본 논문에서는 다자간 원격회의를 위하여 HRTF(head-related transfer function)를 이용하여 스테레오 채널만으로 입체음향을 실현하고 합성 스테레오 음향 반향을 제거하기 위한 적응 알고리즘을 제안한다[5,6]. HRTF로 입체음향을 재생하므로써 스피커와 마이크 시스템으로 구성된 통화 환경 뿐만아니라 헤드폰 혹은 이어폰을 착용한 상태에서도 가상음원을 적용할 수 있다. HRTF를 이용한 합성 입체음향의 반향은 원단화자의 음성신호로부터 HRTF 공간화함수에 의한 변형 및 반향 경로를 거쳐 마이크로 유입되면서 발생한다. 따라서 합성 입체음향 환경에서 요구되어지는 음향반향제거기는 HRTF 및 반향 경로를 동시에 추정하여야 반향을 적절히 제거할 수 있다.

반향제거 알고리즘으로 안정한 수렴을 하는 NLMS 알고리즘(normalized least mean square algorithm)이 음향 반향 제거 알고리즘으로 널리 사용되고 있다. NLMS 알고리즘은 입력 신호가 음성 신호와 같이 유색 신호(colored signal)인 경우에는 수렴 속도가 현저히 저하된다. 유색신호에 대하여 수렴이 빠른 RLS(recursive least square) 알고리즘은 계산량이 많

다. 계산량이 크게 증가하지 않으면서 음성신호에 대해서 수렴 속도가 크게 저하되지 않는 AP 알고리즘(affine projection algorithm)[7,8]이 음향반향를 위한 적응 알고리즘으로 제안되고 있다.

본 논문에서는 HRTF를 이용하여 입체음향을 재생하고 AP 알고리즘으로 반향을 제거하는 합성입체음향 반향 제거 방법을 제안한다. 컴퓨터 시뮬레이션을 통하여 합성 입체음향 환경에서 제안하는 음향 반향제거기의 성능이 기존의 방법에 비하여 우수함을 보인다.

2. 합성 스테레오 음향 반향 제거기

단일 채널의 전송신호를 합성 입체음향신호로 변형하여 가상회의에서 지정한 공간에 참여자의 음원을 정위시키면 보다 효율적으로 회의를 진행할 수 있다. 그림 3은 네 개의 스피커를 이용하여 5 개 영역의 가상음원을 설정한다. 회의의 참여자들의 음성신호를 가상음원의 위치 설정하는 G_x 를 통과시킴으로써 근단화자를 중심으로 5 영역에 가상음원이 지정되어 분리됨으로써 현장감 있는 가상회의가 이루어질 수 있도록 한다. 그림 3과 같은 핸즈프리 음성통신에서는 스피커를 통하여 출력된 원단화자의 음성신호가 근단화자의 마이크로 입력되어 재전송됨으로써 음향 반향이 발생한다.

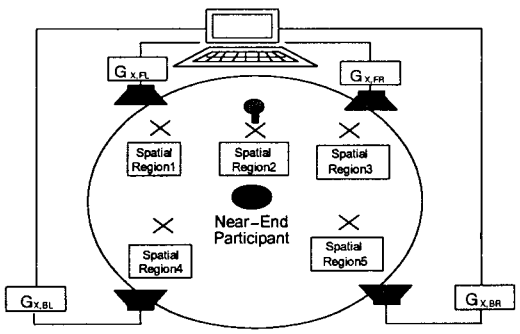


그림 3. 합성입체음향을 이용한 원거리 화상회의 구성

그런데 다중채널 환경에서는 단일 채널의 원단화자 음성신호가 여러 채널로 분리되기 때문에 채널 신호간의 상호 상관도는 상당히 높다. 따라서 적응필터가 반향경로를 정확하게 추정 못하게 되는 원인이 된다.

다중채널 환경에서 음향 반향을 원활히 제거하기 위하여 그림 4와 같이 적응필터 $\widehat{h_x}(n)$으로 반향경로 $h_x(n)$을 추정하여 제거한다. 적응필터의 입력신호를 공간함수 G_x 를 통과하기 전 신호 $v_x(n)$을 사용함으로써 적응필터 입력신호간의 상호 상관도가 낮아져서 반향경로를 정확하게 추정할 수 있다[2-4].

가상음원을 정위시키기 위하여 다중 스피커를 이용하여야 하기 때문에 사용자들이 설치하기 어려운 문제가 있으며 헤드셋을 이용하여 입체음향 환경에서 원격화상회의를 진행하는 상황에서는 가상음원으로 분리할 수 없다.

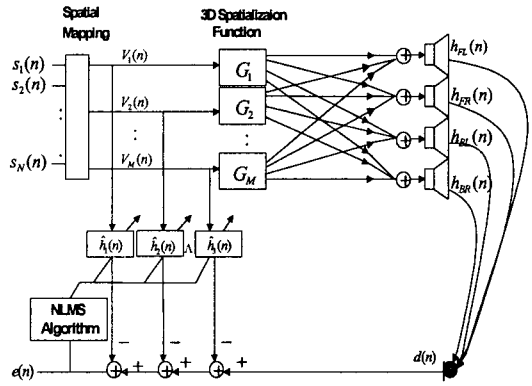


그림 4. 4채널 재생시스템에서의 합성입체 음향반향제거기

3. 합성 입체음향 반향 제거 방법

다중채널 방식의 경우 비용과 설치의 어려움이 있고 전 후 스피커 사이의 음원이나 움직이는 음원에 대한 공간 이미지를 정확한 위치에 재생하기 힘들다. 본 논문에서는 HRTF를 이용하여 두 개의 재생 스피커만을 이용하여 가상음원을 설정하여 가상공간에서의 다양한 위치에 대한 입체음향 환경을 구축하고 합성 스테레오 반향을 제거하는 방법을 제안한다. 2채널 스피커를 구비한 PC 가 대부분을 차지하고 있음을 고려하면 일반 사용자들이 손쉽게 입체음향을 구현하기 위해서는 2채널 스피커 기반 기술을 사용하는 것이 현명한 방법이다.

제안하는 방법은 기본 적으로 T. Yensen 등이 제안한 합성 스테레오 반향 제거기 구조를 이용한다. 그림 8은 제안하는 합성 스테레오 반향제거기의 구조이며 두 영역에 대하여 가상 음원을 설정한다. $v_1(n)$, $v_2(n)$는 두 가상 음원영역의 신호이다. G_1과 G_2는 가상음원 설정을 위한 HRTF 이며 공간영역 1을 나타내기 위한 G_1은 $g_{1L}(n)$, $g_{1R}(n)$로 구성되고

공간영역 2를 나타내기 위한 G_2는 $g_{2L}(n)$, $g_{2R}(n)$로 구성된다.

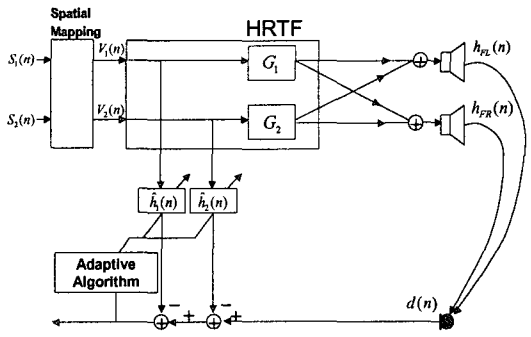


그림 4. 다자간 화상회의를 위한 제안하는 입체음향 구현 및 반향제거기

스피커 출력 $x_L(n)$, $x_R(n)$ 는 입력 신호 $v_1(n)$, $v_2(n)$ 과 공간화 함수 HRTF와 컨벌루션되어 형성된다.

$$\boldsymbol{x}_L(n)=[x_i(n)\ x_i(n-1)\ \cdots\ x_i(n-L+1)]^T$$

$$\boldsymbol{x}_R(n)=[x_R(n)\ x_R(n-1)\ \cdots\ x_R(n-L+1)]^T$$

$$x_{FL}(n)= v_1(n)* g_{1L}(n)+ v_2(n)* g_{2L}(n)$$

$$x_{FR}(n)= v_1(n)* g_{1R}(n)+ v_2(n)* g_{2R}(n)$$

수신룸에서 마이크에 받아들여진 두 개의 채널의 입력 신호 $d(n)$ 은

$$d(n)= x_{FL}(n)* h_1(n)+ x_{FR}(n)* h_2(n)$$

으로 표현된다. 오차 신호는,

$$e(n)=d(n)- \hat{h}_1{}^T(n)\ v_1(n)- \hat{h}_2{}^T(n)\ v_2(n)$$

가 된다. 여기서 T는 전치를 의미하고 $\hat{h}_1(n)$ 과 $\hat{h}_2(n)$ 은 공간영역 1과 공간영역2에 대한 음향반향 경로 모델이며 공간화 함수와 수신룸의 임펄스 함수의 컨벌루션의 합이다.

$$\hat{h}_1(n)= g_{1L}(n)* h_1(n)+ g_{1R}(n)* h_2(n)$$

$$\hat{h}_2(n)= g_{2L}(n)* h_1(n)+ g_{2R}(n)* h_2(n)$$

적응필터는 가상 음원영역 신호 $v_1(n)$, $v_2(n)$와 잔여 반향신호 $e(n)$을 이용하여 AP 알고리즘으로 갱신된다.

4. 컴퓨터 시뮬레이션 결과

제안하는 합성 스테레오 음향 반향 제거기의 성능을 확인하기 위하여 적응필터의 입력신호 즉 원단화자 신호는 16 비트로 양자화하고 8 kHz로 샘플링된 여성의 음성을 사용하였다. 수신룸의 임펄스 응답은 2 채널 데이터 레코더로 실제 회의실에서 측정 하였으며 임펄스 응답의 길이는 1024 탭으로 하였다. 적응필터의 적응상수는 0.5로 두고 시뮬레이션 하였다. HRTF를 이용한 가상 음원의 위치는 근단화자를 중심으로 각각 30°, 330°위치에 설정하였다. 임펄스 응답은 MIT의 Media lab에서 측정한 것으로 KEMAR 더미 헤드 마이크를 이용하여 측정된 것이다.

시뮬레이션에 AP 알고리즘의 입력신호의 전력은 모두 running power estimate 로 추정하였으며 망각지수(forgetting factor)는 0.998로 두었다.

그림 5과 6는 위의 조건에 따라 시뮬레이션한 결과이다. 그림 5의 (a)는 가상 음원 위치에서의 원단화자 음성신호이며 적응필터의 입력신호이다. (b)는 잔여 반향 신호이다. 그림 6의 (a) 입력이 그림 5의 (a)인 음성신호일 때 ERLE이며 (b)는 계수오조정 지수를 나타내고 있다. 완전히 반향이 제거되지 않았지만 약 10 dB 정도 반향이 제거되는 것을 확인할 수 있다.

5. 결론

합성 입체음향은 현장감 또는 공간감 등이 요구되는 가상현실 분야 뿐 아니라 VoIP를 통한 다자가 참여하는 가상회의장에서 각 토론자의 음성을 가상의 음원 위치에 설정함으로써 회의의 효과를 높일 수 있다. 본 논문에서는 현장감 있는 VoIP 환경의 구현을 위해 HRTF를 이용한 합성 스테레오 기법을 사용하였으며, 그에 따른 음향 반향 제거 알고리즘을 제안하였다.

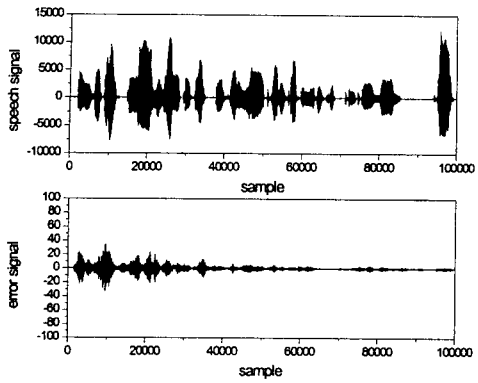
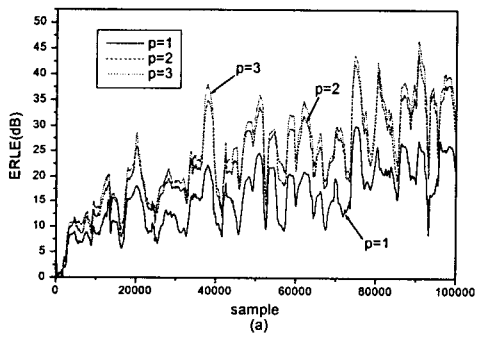


그림 5. 음성 신호를 이용한 합성 음향 반향 제거 시뮬레이션 결과 (a) 입력 음성신호 (b) 잔여 음성신호



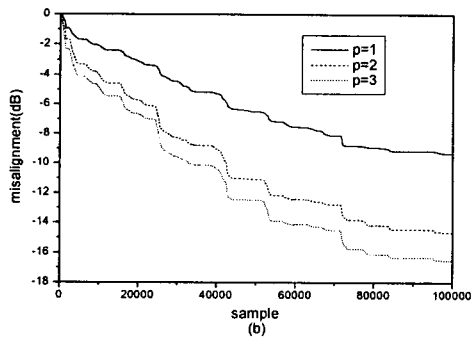


그림 6. 음성 신호를 이용한 합성 음향 반향 제거 시뮬레이션 결과 (a) 계수 오조정 지수 (b) ERLE

[참고문헌]

[1] 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.

[2] 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.

[3] 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.

[4] 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.

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

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

[7] 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.

[8] M. Tanaka, Y. Kaneda, S. Makino, and J. Kojima, "Fast projection algorithm and its step size control," in Proc. Int. Conf. Acoustics, Speech, and Signal Processing, Detroit, MI, 1995, pp. 945-948.