

머리 전달 함수를 이용한 합성 스테레오 음향 반향 제거기

박 장 식, 백 주 순, 손 경 식
동의공업대학 영상정보과, 부산대학교 전자공학과

Acoustic Echo Canceller for Synthetic Stereo Using HRTF

Jang Sik Park, Ju Soon Baek Kyung Sik Son
Dept. of Visual Technologies, Dong-Eui Institute of Technology
Dept. of Electronics Eng., Pusan National University

E-mail : jsipark@dit.ac.kr

요 약

In this brief, Acoustic echo cancellation scheme is proposed to enhance the presence of multiple participants of hands-free voice and video conference. Synthetic stereo using head related transfer function and the stereo echo cancellation scheme are proposed. It is shown that the proposed synthetic stereo echo cancellation scheme is well performed by computer simulation.

1. 서론

통신 시스템의 발전으로 다양한 음성 및 영상서비스가 제공되고 있다. 특히 음성 및 음향 분야에서는 몰입형 음향 시스템(immersive audio system)이 현장감 있는 음성 통신 뿐만아니라 가상현실, 오락, 텔레비전, 영화 등에 활용되어 가고 있다[1]. 그리고 편리하고 안전한 통화를 위하여 핸드프리(hands-free) 단말기를 이용한 음성 통신이 원거리 회의, 차량용 핸드프리 등에 활용되고 있다[2].

현장감을 제공하기 위한 스테레오 방식의 원격통신 시스템에서는 스피커의 출력이 마이크로 입력되는 음향 반향 신호가 발생한다. 따라서 스테레오 원격통신 시스템에는 스테레오 음향 반향 제거기가 갖추어져야 원활한 음성통신이 가능하다. 그러나 일반적인 스테레오 신호는 단일화자로부터 발생한 신호이므로 두 신호 간의 상호상관(cross-correlation)이 크다. 스테레오 신호간의 상호상관이 크기 때문에 적응필터가 반향경로(echo path)를 정확히 추정하지 못하여 스테레오 음

향 반향 제거기의 성능이 저하된다[2,3]. 또한 원단화자의 위치가 변하거나 다른 화자가 말을 이어서 할 때와 같이 원단화자의 환경변화에 적응필터가 오동작하게 된다.

스테레오 음향 반향 제거를 위하여 다양한 방법들이 제안되고 있다[3-6]. 제안된 스테레오 음향 반향 제거기는 크게 적응 알고리즘을 개선하는 방법과 전처리를 하는 방법으로 나눌 수 있다. 전처리를 하는 방법으로는 Benesty 등은 각 채널 신호에 원신호와 유사한 신호를 더하는 전처리를 함으로써 상호 상관을 줄이는 방법을 제안하였다[3]. 적응 알고리즘을 개선하는 대표적인 방법으로는 Shimauchi 등은 투영 알고리즘(projection algorithm)을 이용한 스테레오 음향 반향 제거기를 제안하였고[5] Benesty 등은 다중 채널 AP(multi-channel affine projection) 알고리즘으로 확장하였다[6].

합성 스테레오는 다자가 참여하는 화상회의에서 각 참여자의 음성을 가상의 음원 위치에 설정함으로써

회의의 효과를 높일 수 있다[7]. 단일 채널로 전송하더라도 수신측에서 임의의 가상 음원의 위치에 재생될 수 있도록 하는 것이 합성 스테레오이다. 합성 스테레오 음성 통신에서도 음향 반향에 의한 음질 저하는 여전히 남아 있다. J. Benesty 등은 합성 스테레오 반향 제거를 위하여 기존의 스테레오 반향 제거 기법을 그대로 적용하여 해결하고자 하였고[7] T. Yensen 등은 스테레오 반향 제거기의 구조를 바꾸어 그 성능을 향상시켰다[8,9].

반향 제거를 위한 적응 알고리즘으로는 NLMS 알고리즘, RLS(recursive least square) 알고리즘 그리고 AP(affine projection) 알고리즘[10-11]이 주로 사용되고 있다. NLMS 알고리즘은 유색의 음성신호의 자기상관에 의하여 수렴 속도가 느려진다. 유색 음성신호에 대하여 RLS 알고리즘의 수렴 특성이 우수하지만 계산량이 다소 많다.

본 논문에서는 다자가 참여하는 화상회의 시스템을 위한 합성 스테레오를 HRTF(head related transfer function)을 이용하는 각 화자의 가상적인 위치를 설정하고 이를 위한 반향 제거 알고리즘을 제안한다.

2. 합성 스테레오

단일 채널 전송신호를 이용하여 합성 스테레오로 화상회의 등의 참여자의 가상 음원의 위치를 재설정함으로써 보다 효율적으로 회의를 진행할 수 있다. 그림 1은 네 개의 스피커를 이용하여 5 영역의 가상 음원을 설정하는 예이다. 근단화자를 중심으로 회의의 참여자들의 음성신호를 가상 음원의 위치 설정을 위한 G_x 함수를 통과시킴으로써 5 영역의 가상 음원의 위치에 지정하여 분리함으로써 현장감 있는 회의가 가능해진다.

일반적인 스테레오 음향통신 시스템은 두 개의 스피커와 두 개의 마이크로 구성된다. 원단화자의 음성신호가 두 채널로 전송되어 스피커로 출력되고 근단화자의 마이크로 입력되어 재전송됨으로써 음향 반향이 발생한다. 일반적인 스테레오 음향통신 시스템에서 원단화자의 음성신호가 두 채널로 분리되기 때문에 두 채널 신호의 상호 상관도는 상당히 높다. 따라서 적응필터가 근단화자 룸에서의 반향경로를 정확하게 추정 못하게 되는 원인이 된다.

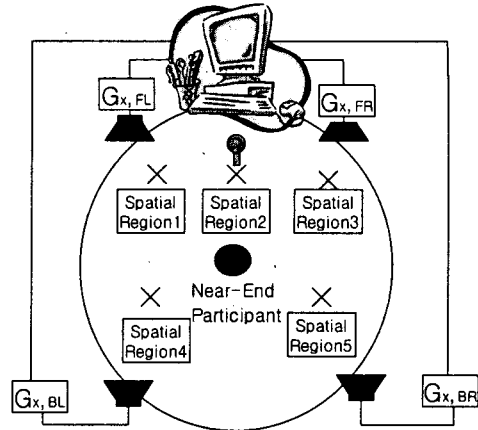


그림 1. 합성 스테레오를 이용한 원거리 화상회의 구성

이러한 입력신호의 상호 상관을 줄이기 위하여 스테레오 음향 반향 제거 방법은 두 입력신호의 상호상관을 줄여 적응필터에 입력함으로써 성능을 개선하였다[7].

T. Yensen 등[8,9]은 그림 2와 같이 가상 음원을 설정하는 G_x 함수의 입력신호, $V_M(n)$ 를 적응필터의 입력으로 활용함으로써 입력신호의 상호상관에 의한 오조정을 개선하였다.

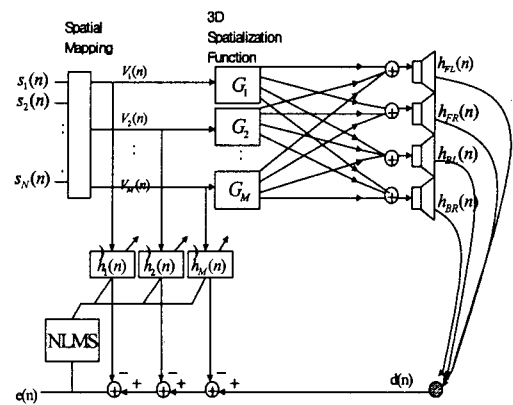


그림 2. 4채널 재생 시스템에서의 합성 스테레오 반향 제거기

3. 제안하는 합성 스테레오 반향 제거기

청취자가 가상음원을 설정하기 위하여 네 개의 스피커를 설치함에 따라 생기는 번거러움을 없애고 헤드폰 혹은 이어폰을 착용한 상태에서도 가상음원이 적용할 수 있도록 본 논문에서 HRTF를 이용하여 두 개의 재생 스피커를 이용하여 가상음원을 설정하고 합성 스테레오 반향을 제거하는 방법을 제안한다. 제안하는 방법은 기본적으로 T. Yensen 등이 제안한 합성 스테레오 반향 제거기 구조를 이용한다. 그림 3은 제안하는 합성 스테레오 반향제거기의 구조이며 두 영역에 대하여 가상 음원을 설정한다.

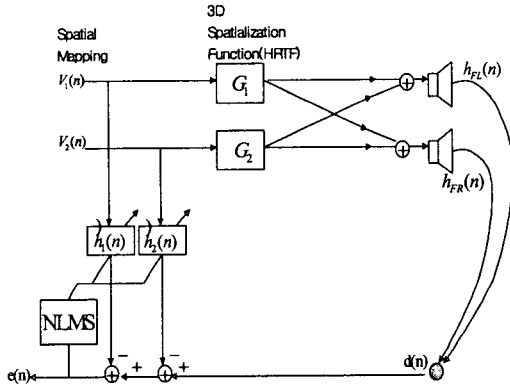


그림 3. 제안하는 HRTF를 이용한 합성 스테레오 반향 제거기

$v_1(n)$, $v_2(n)$ 는 두 가상 음원영역의 신호이며 G_1 과 G_2 는 가상 음원 설정을 위한 HRTF 이다. $h_{FL}(n)$, $h_{FR}(n)$ 은 스테레오 반향경로이며 $h_1(n)$, $h_2(n)$ 은 각 스테레오 반향경로를 추정하는 적응필터 계수벡터이다. $d(n)$ 와 $e(n)$ 은 각각 마이크 입력신호와 잔여 반향신호이다. 적응필터는 가상 음원영역 신호 $v_1(n)$, $v_2(n)$ 와 잔여 반향신호 $e(n)$ 을 이용하여 NLMS 알고리즘으로 갱신된다.

스테레오 음향 반향 제거기에서 두 입력 신호에 의한 상호상관에 의해서 성능이 저하되는 것은 아래의 식과 같이 적응 필터의 최적해가 반향경로를 정확하게 추정하지 못하는 데 있다.

$$\hat{h}_{1opt}(n) = R_1^{-1}(n) p_1(n) = h_1(n) + R_1^{-1}(n) R_{12}(n) h_2(n) \quad (1)$$

$$\hat{h}_{2opt}(n) = R_2^{-1}(n) p_2(n) = h_2(n) + R_2^{-1}(n) R_{21}(n) h_1(n) \quad (2)$$

$$R_1(n) = E[v_1(n) v_1^T(n)], \quad R_2(n) = E[v_2(n) v_2^T(n)]$$

$$R_{12}(n) = E[v_1(n) v_2^T(n)], \quad R_{21}(n) = E[v_2(n) v_1^T(n)]$$

$$p_1(n) = E[v_1(n) d(n)], \quad p_2(n) = E[v_2(n) d(n)]$$

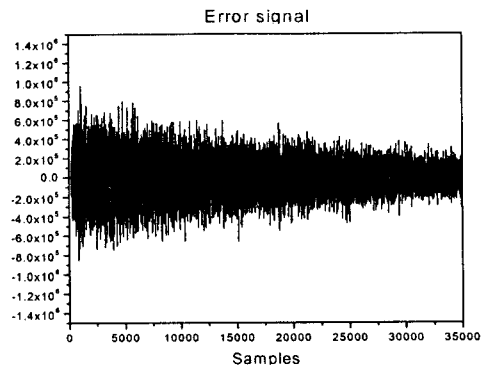
$$v_i(n) = [v_i(n) v_i(n-1) \dots v_i(n-L+1)]^T$$

식 (1)과 (2)에서 그림 3과 같이 가상음원 설정 함수를 통과하기 전 신호를 적응필터에 입력함으로써 적응필터의 두 입력신호가 서로 상관이 없기 때문에 일반적인 스테레오 음성통신에서 발생하는 상호상관에 의한 성능 저하가 생기지 않는다.

4. 시뮬레이션 결과 및 검토

제안하는 합성 스테레오 음향 반향 제거기의 성능을 확인하기 위하여 적응필터의 입력신호 즉 원단화자 신호를 각각 백색잡음과 음성신호를 입력하여 시뮬레이션한다.

그림 4는 백색잡음을 입력하여 시뮬레이션한 결과이다. 백색잡음은 Gauss 잡음을 사용한다. 반향경로와 적응필터 탭수는 1024로 하고 적응상수는 0.5로 두고 시뮬레이션한다. 그림 4의 (a)는 잔여 오차신호이며 (b)와 (c)는 각각 적응필터 계수 오조정지수와 ERLE(echo return loss enhancement)이다. 지속적으로 잔향이 줄어드는 것을 확인할 수 있다.



(a)

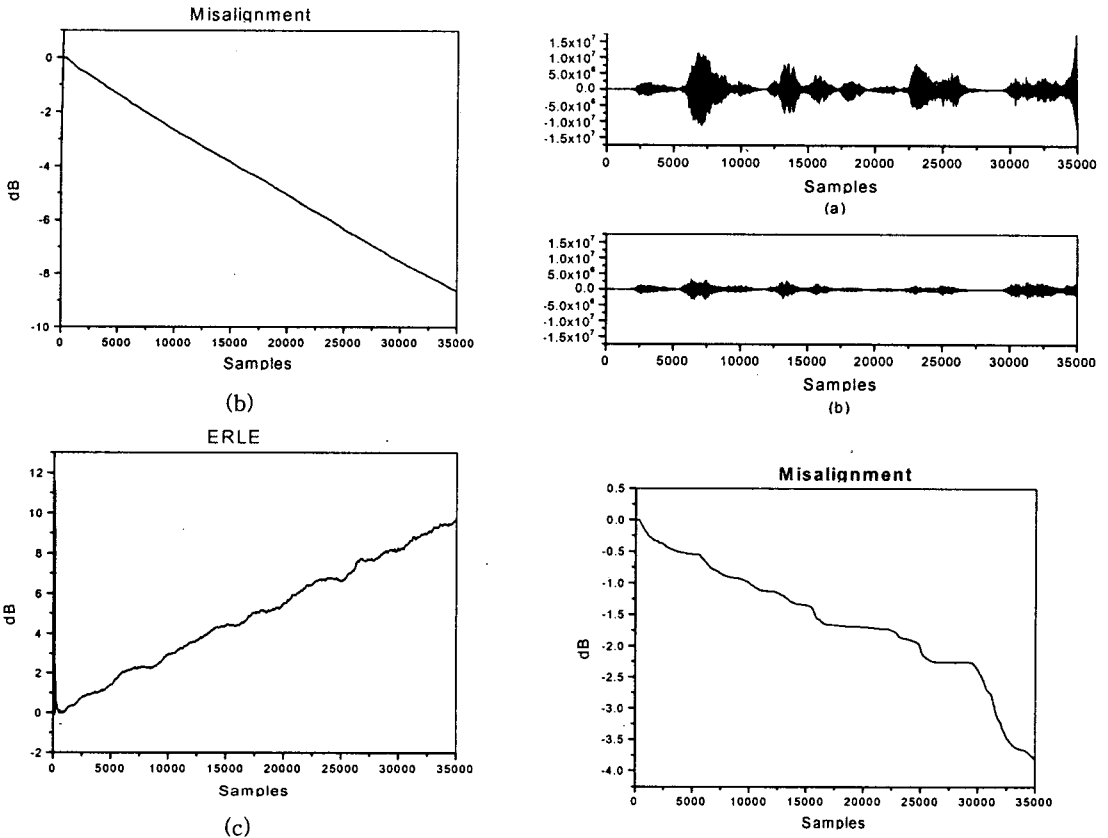


그림 4. 백색잡음을 이용한 시뮬레이션 결과
(a) 잔여신호 (b) 오조정 지수 (c) ERLE

음성 신호에 대한 합성 스테레오 반향 제거 성능을 확인하기 위하여 원단화자 신호는 16 kHz 로 샘플링 한 두 여성 음성 신호 35000 샘플을 사용한다. HRTF 를 이용한 가상 음원의 위치를 근단화자를 중심으로 각각 30°, 330° 위치에 설정한다. 근단화자 룸 임펄스 응답은 1024 탭으로 모델링하고 적응필터의 차수를 같이 한다. NLMS 알고리즘으로 적응 필터를 적용시키고 적응상수는 0.5로 한다.

그림 5는 위의 조건에 따라 시뮬레이션한 결과이다. (a)는 가상 음원 위치에서의 원단화자 음성신호이며 적응필터의 입력신호이다. (b)는 잔여 반향 신호이다. (c) 입력이 (a)의 음성신호일 때 적응필터의 계수 오조정 지수이며 (d)는 ERLE를 나타내고 있다. 완전히 반향이 제거되지 않았지만 약 10 dB 정도 반향이 제거되는 것을 확인할 수 있다.

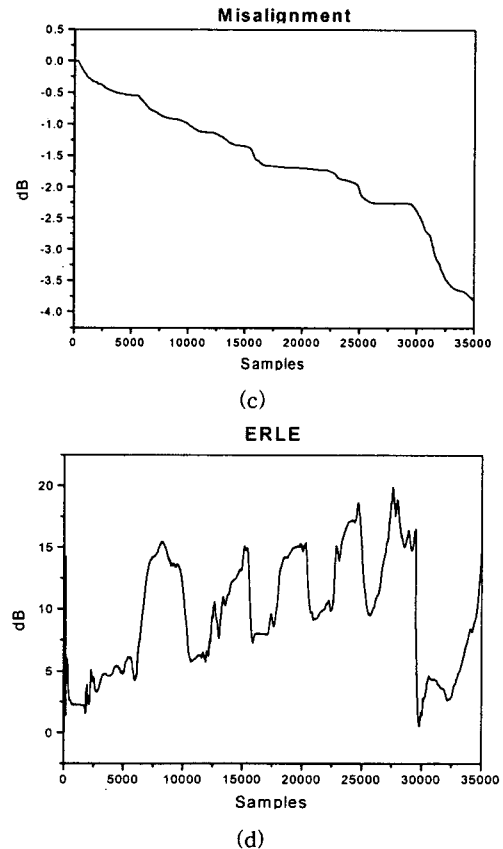


그림 5. 음성 신호를 이용한 합성 음향 반향 제거 시뮬레이션 결과

(a) 입력 음성신호 (b) 잔여 음성신호
(c) 계수 오조정 지수 (d) ERLE

5. 결론

본 논문에서는 현장감있는 VoIP 구현을 위한 합성 스테레오를 HRTF를 이용하여 구현하고 합성 스테레오 반향 제거하는 방법을 제안하였다. NLMS 알고리즘을 이용한 반향 제거 시뮬레이션 결과 반향을 평균 10 dB 정도 제거할 수 있음을 보였다. 시뮬레이션 결과를 바탕으로 DSP를 이용하여 하드웨어 구현을 하고자 한다.

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] S. Makino, "The past, present, and future of audio signal processing-acoustic echo cancellation," IEEE Signal Processing Magazine, pp.39-41, Sep., 1997.
- [3] J. Benesty, D. R. Morgan and M. M. Sondhi, "A better understanding and an improved solution to the problems of stereophonic acoustic echo cancellation," proc. on International Conference on Acoustics, Speech and Signal Processing 97, Munich Germany, pp.303-306, 1997.
- [4] Y. H. Lee, J. S. Park, and K. S. Son, "Performance improvement of stereophonic acoustic echo canceller using non-linear pre-processing," Proc. on International Workshop on Hands-Free Speech Comm., pp. 75-78, April. 2001.
- [5] S. Shimauchi and S. Makino, "Stereo projection echo canceller with true echo path estimation," Proc. on International Conference on Acoustics, Speech and Signal Processing 95, Detroit Michigan, pp.3059-3062, 1995.
- [6] J. Benesty, P. Duhamel and Y. Grenier, "A multichannel affine projection algorithm with applications to multichannel acoustic echo cancellation," IEEE Signal Processing Letters, Vol. 3, No. 2, pp 35-37, Feb. 1996.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its property," Electron. Comm. Japan, VOL. J67-A, NO. 2, pp.126-132, 1984.(Japanese)
- [11] M. Tanaka, Y. Kaneda, S. Makino and J. Kojima, "Fast projection algorithm and its step size control," Proc. International Conference on Acoustics, Speech and signal Processing 95, Detroit Michigan, pp. 945-948, 1995.