

논문 2013-50-12-24

Least Squares 기반의 Volterra Filter를 이용한 비선형 반향신호 억제기

(Nonlinear Acoustic Echo Suppressor based on Volterra Filter using
Least Squares)

박 지 환*, 이 봉 기*, 장 준 혁**

(Jihwan Park, Bong-Ki Lee, and Joon-Hyuk Chang[Ⓢ])

요 약

기존의 반향신호 억제기는 스피커와 마이크 사이의 선형 관계만을 고려하여, 마이크로 입력된 신호로부터 반향신호를 억제한다. 하지만 실제로 스피커는 비선형성을 가지고 있으며, 이 때문에 기존의 반향신호 억제기는 비선형 반향신호 환경에서 그 성능이 저하된다. 본 논문에서는 스피커의 비선형성을 모델링하기에 적합한 주파수영역상의 Least square 방식의 Volterra filter를 적용한 비선형 반향신호 억제기를 제안하였다. 객관적 성능평가 방법인 Echo Return Loss Enhancement (ERLE)와 Speech Attenuation (SA)를 도입하여 제안된 알고리즘의 성능 검증에 사용하였다. 제안된 알고리즘이 기존의 반향신호 억제기보다 선형 및 비선형 반향신호 환경에서 우수한 성능을 보이는 것을 확인하였다.

Abstract

A conventional acoustic echo suppressor (AES) considering only room impulse response between a loudspeaker and a microphone eliminates acoustic echo from the microphone input. However, in a nonlinear acoustic echo environment, the conventional AES degraded because of a nonlinearity of the loudspeaker. In this paper, we adopt AES based on the frequency-domain second-order Volterra filter using Least Square method. For comparing performances, we conduct objective tests including Echo Return Loss Enhancement (ERLE) and Speech Attenuation (SA). The proposed algorithm shows better performance than the conventional in both linear and nonlinear acoustic echo environments.

Keywords : Nonlinear acoustic echo suppressor, Least squares, Volterra filter

* 학생회원, ** 평생회원, 한양대학교 전자컴퓨터통신공학과

(Department of Electronics and Computer Engineering, Hanyang University)

Ⓢ Corresponding Author(E-mail: jchang@hanyang.ac.kr)

※ 본 연구는 미래창조과학부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음(NIPA-2013-H0301-13-4005). 그리고 이 논문은 2011년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2011-0009182). 또한 본 논문은 2012년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2012R1A2A2A01004895).

접수일자: 2013년10월18일, 수정완료일: 2013년11월22일

I. 서 론

일반적으로 핸즈프리 통신환경에서, 음향학적 반향신호는 스피커로 출력된 신호가 다양한 반향경로를 통해 마이크로 다시 유입되는 것을 말한다. 이러한 음향학적 반향신호는 화자간의 통화를 방해하는 요인으로 작용하여 통화품질 저하의 주요인으로 꼽힌다. 그래서 핸즈프리 통신환경에서도 우수한 통화품질을 보장하기 위해 시간영역 상에서 동작하는 반향신호 제거기 (Acoustic Echo Canceller)나 주파수영역 상에서 동작하는 반향신호 억제기 (Acoustic Echo Suppressor)를 사용해 반향

신호를 제거한다. 음향학적 반향신호는 반향경로 (Echo Path)에 영향을 많이 받는데, 계속적으로 변화하는 반향경로를 모델링하기 위해 적응형 필터의 사용이 필수적이다. 그중에서도 Least Mean Square (LMS), Normalized Least Mean Square (NLMS), Recursive Least Mean Square (RLS), Proportionate Least Mean Square (PLMS), Affine Projection Algorithm (APA) 등과 같은 다양한 적응형 필터를 사용한 반향신호 제거기, 반향신호 억제기가 연구·개발되었다^[1]. 그 중에서도 주파수영역에서 동작하는 반향신호 억제기는 시간영역에서 동작하는 알고리즘에 비해 상대적으로 간단한 구조이며, 연산량이 적고, 우수한 성능을 보였다^[2]. 하지만 지금까지의 적응형 필터를 이용한 반향신호 제거기·억제기는 스피커와 마이크 사이의 반향경로, 즉 선형관계만을 고려한 것이 대부분이며, 스피커와 마이크 사이의 반향경로가 선형적이지 않을 때는 기존의 알고리즘의 성능이 크게 저하된다. 이는 실제적으로 스피커가 가지는 비선형성을 고려할 때 성능개선의 한계점으로 작용한다^[3].

본 논문에서는 스피커의 비선형성을 모델링하기 위해 Second-order Volterra filter를 사용하였다. Second-order Volterra filter를 주파수영역상에서 구현하고 Least square 업데이트 방식을 사용하여 스피커의 비선형성으로 인해 발생된 비선형 반향 신호 환경에서도 우수한 성능을 보이는 반향신호 억제기를 제안하였다.

객관적인 성능 검증을 위해 Echo Return Loss Enhancement (ERLE)를 도입해 반향신호 억제율을 측정하였고, Speech Attenuation (SA)를 도입해 반향신호가 아닌 신호의 손실 여부를 측정하여 제안된 알고리즘과 기존의 알고리즘을 비교하였다.

II. 본 론

1. 기존의 반향신호 억제기

기존의 반향신호 억제기는 그림 1과 같은 구조로 동작한다^[2]. 원단화자 신호 $x(n)$ 이 스피커를 통해 출력되면 해당 공간의 임펄스 응답에 따라 다음과 같이 반향신호가 결정된다.

$$d(n) = x(n) * h(n) \quad (1)$$

여기서 $d(n)$ 은 반향신호이다.

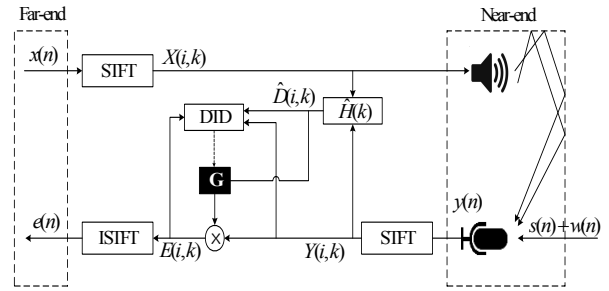


그림 1. 기존의 반향 억제기의 블록다이어그램

Fig. 1. Block Diagram of conventional acoustic echo suppressor.

따라서 실제 마이크로 입력되는 신호 $y(n)$ 은 스피커로 출력된 반향신호 $d(n)$, 근단화자 신호 $s(n)$, 잡음 $w(n)$ 의 합으로 정의할 수 있다.

$$y(n) = d(n) + s(n) + w(n) \quad (2)$$

이렇게 입력된 마이크 입력 신호 $y(n)$ 과 원단화자 신호 $x(n)$ 은 Short-Time Fourier Transform (STFT)을 이용해 주파수영역상에서 다음과 같이 표현된다.

$$X(i,k) = STFT\{x(n)\} \quad (3)$$

$$Y(i,k) = STFT\{y(n)\}$$

여기서 i 는 프레임 인덱스를 나타내고, k 는 주파수 채널 인덱스를 나타낸다.

STFT을 이용해 주파수 영역에서 표현된 원단화자 신호 $X(i,k)$ 와 근단화자 신호 $Y(i,k)$ 를 알면 스피커와 마이크 사이의 주파수 응답, 즉 선형변환 관계를 Least Square를 이용해 다음과 같이 예측할 수 있다^[3].

$$\hat{H}(k) = \frac{|Y(i,k)X^*(i,k)|}{|X(i,k)X^*(i,k)|} \quad (4)$$

여기서 $\hat{H}(k)$ 는 실제 임펄스 응답인 $h(n)$ 를 주파수영역에서 나타낸 $H(k)$ 의 추정치이며, *는 Complex Conjugator를 의미한다.

식 (4)를 통해 구한 주파수 응답 추정치를 이용해 주파수 영역상의 반향신호를 추정하면,

$$\hat{D}(i,k) = \hat{H}(i,k)X(i,k) \quad (5)$$

이다. 여기서 $\hat{D}(i,k)$ 는 주파수영역에서의 반향신호를 의미한다.

식 (5)로부터 얻을 수 있는 주파수 영역상의 반향신

호를 이용해 Wiener filter의 개념을 적용한 최종적인 반향신호 억제 이득 함수를 다음과 같이 계산 할 수 있는데,

$$G(i, k) = \left[\frac{\max(|Y(i, k)| - |\hat{D}(i, k)|, 0)}{|Y(i, k)|} \right] \quad (6)$$

이로부터 얻은 식 (6)을 이용해 마이크 입력에 적용해 주면,

$$\hat{S}(i, k) = G(i, k) Y(i, k) \quad (7)$$

여기서 $\hat{S}(i, k)$ 는 주파수영역상에서 표현된 근단화자 신호 추정치이다. 결과적으로 반향신호가 존재하는 환경에서 주파수응답을 모델링하고 주파수영역상의 반향신호를 추정하여, 마이크로 입력된 신호 중에서 반향신호만을 억제할 수 있다.

2. 제안된 비선형 반향신호 억제기

기존의 반향신호 억제기는 스피커와 마이크 사이의 선형 관계만을 고려하기 때문에, 실제적으로 스피커에 존재하는 비선형성으로 인해 발생하는 비선형 반향신호 환경에서 그 성능이 크게 떨어진다. 따라서 이러한 스피커의 비선형성을 모델링하기에 적합한 Second-order Volterra filter를 도입하여 비선형 잔향신호 환경에서도 반향신호 억제율이 우수한 알고리즘을 제안하였다.

시간영역과 주파수영역상에서 Second-order Volterra filter는 다음과 같이 정의한다^[4].

$$\hat{d}(n) = \sum_{i=0}^{N-1} \hat{h}_1(n) x(n-i) + \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \hat{h}_2(i, j) x(n-i) x(n-j) \quad (8)$$

$$\hat{D}(i, k) = H_1(k) X(i, k) + \sum_{p=0}^{K-1} \sum_{q=0}^{K-1} H_2(p, q) X(i, p) X(i, q) \delta_K(k-p-q) \quad (9)$$

여기서

$$\delta_K(k) = \begin{cases} 1, & (k \text{ modulo } K) = 0 \\ 0, & (k \text{ modulo } K) \neq 0 \end{cases} \quad (10)$$

이다. 주파수영역에서 표현된 Second-order Volterra filter는 Least Square 방식을 이용해 최적의 filter 계수를

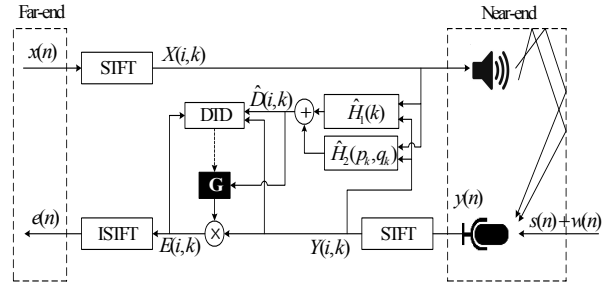


그림 2. 제안된 비선형 반향신호 억제기의 블록 다이어그램
Fig. 2. Block Diagram of the proposed nonlinear acoustic echo suppressor.

찾을 수 있다. Volterra filter에 Least Square를 적용하면^[3],

$$\hat{\underline{H}}_k = R_{i,k}^{-1} r_{i,k} \quad (11)$$

여기서

$$R_{i,k} = \underline{X}_{i,k}^T \underline{X}_{i,k} \quad (12)$$

$$r_{i,k} = Y(i, k) \underline{X}_{i,k} \quad (13)$$

$$\underline{X}_{i,k} = [X(i, k), X(i, p_1) X(i, q_1), \dots, X(i, p_K) X(i, q_K)]$$

$$\hat{\underline{H}}_k = [H_1(k), H_2(p_1, q_1), \dots, H_2(p_K, q_K)] \quad (14)$$

이다. 식 (11)을 이용해 얻은 최적의 filter 계수로부터 더욱 정확한 주파수영역상의 반향신호 $\hat{D}(i, k)$ 를 얻을 수 있으며, 이는 아래와 같이 표현할 수 있다.

$$\hat{D}(i, k) = \hat{\underline{H}}_k^T \underline{X}_{i,k} \quad (15)$$

식 (15)로부터 구한 주파수영역상의 반향신호는 기존의 알고리즘과 같이 Wiener filter의 개념을 적용한 반향신호 억제 이득 함수를 구하는데 사용되며, 식 (6)과 식 (7)을 이용해 주파수영역상에서 표현된 근단화자 신호 추정치를 계산해 낼 수 있다.

III. 실험 및 결과

본 논문에서는 컴퓨터 시뮬레이션을 이용한 성능 평가를 수행하였다. 실험에 이용된 데이터 파일은 남자 4명과 여자 4명으로부터 각각 8 kHz의 샘플링주파수로 입력받은 NTT 한국어 DB를 사용하였다. 남자 2명과 여자 2명을 근단 화자 음성으로, 나머지를 원단 화자

음성으로 사용하였다. 원단 화자 음성으로 분류된 데이터는 비선형화, 잔향효과, 감쇄효과를 주어 비선형 잔향 신호를 생성하였다. 먼저 스피커의 비선형성을 모델링하는 식을 사용하여 스피커에서 출력되는 신호를 생성하였다^[5].

$$x_{nl}(n) = \gamma \left(\frac{1}{1 + \exp(-pq(n))} - \frac{1}{2} \right) \quad (16)$$

여기서

$$q(n) = \frac{3}{2}x(n) - \frac{3}{10}x^2(n) \quad (17)$$

이다. 여기서 γ 은 Sigmoid 함수의 이득 값이며, 시뮬레이션 상에서는 $\gamma = 2$ 로 사용하였다. ρ 는 Sigmoid의 기울기 값이며, 이는

$$\rho = \begin{cases} 4, & q(n) > 0 \\ \frac{1}{2}, & q(n) \leq 0 \end{cases} \quad (18)$$

으로 정의한다. 실제 통화 환경을 고려하여 비선형화를 통해 얻은 x_{nl} 에 시뮬레이션을 통해 $5 \times 4 \times 3m^3$ 크기의 일반적인 사무실에서의 임펄스 응답을 생성하여 적용하였다^[6]. 각 프레임은 80개의 샘플로 이루어져있으며 Zero-padding 이후에 Window를 적용하고 128-point DFT를 이용하여 주파수 영역으로 변환 하였다. 실험은

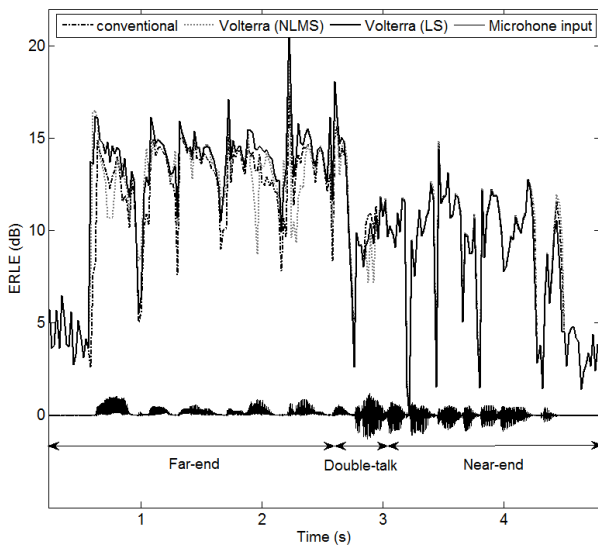


그림 3. 기존의 알고리즘과 제안된 알고리즘의 ERLE 결과 비교

Fig. 3. Comparing the ERLE results of conventional and proposed algorithm.

표 1. 기존의 알고리즘과 제안된 알고리즘의 ERLE 및 SA 결과 비교

Table 1. Comparing the ERLE and SA results of conventional and proposed algorithm.

	ERLE (dB)	SA (dB)
conventional	13.20 ± 0.59	8.03 ± 0.34
Volterra(NLMS)	13.63 ± 0.66	8.16 ± 0.85
proposed	14.12 ± 0.26	7.59 ± 0.23

(95% confidence interval)

근단화자 신호 또는 원단화자 신호만 존재하는 Single-talk, 근단화자 신호와 원단화자 신호 모두 존재하는 Double-talk을 모두 고려하여 진행하였다. Double-talk 발생시 반향경로 추정 알고리즘의 발산을 막기 위해 주파수영역상에서의 Double-talk detection 알고리즘을 사용하였다^[7].

반향 신호 제거와 음성왜곡도 측정에 대한 객관적인 성능평가를 위해 ERLE와 SA를 도입하였다. ERLE는 출력에서의 음향학적 반향 신호 감쇠 정도를 수치적으로 나타낼 수 있고, SA는 출력에서 근단화자 신호의 왜곡 정도를 알 수 있다. ERLE와 SA는 다음과 같이 정의한다^[8].

$$ERLE(n) = 10 \log_{10} \left[\frac{E[y^2(n)]}{E[e^2(n)]} \right] \quad (19)$$

$$SA(n) = \frac{1}{N} \sum 10 \log_{10} \left[\frac{E[s^2(n)]}{E[\tilde{s}^2(n)]} \right] \quad (20)$$

여기서 N 은 Double-talk 구간의 샘플 개수를 의미한다.

기존의 알고리즘, 주파수영역상에서의 NLMS 방식을 사용한 Volterra filter 기반의 알고리즘, 제안된 주파수영역상에서의 Least Square 방식을 사용한 Volterra filter 기반의 알고리즘의 ERLE 비교 결과를 그림 3에 나타내었다. 원단화자 신호만 존재하는 Single-talk 구간에서 제안된 알고리즘의 ERLE가 다른 알고리즘의 결과보다 더 큰 것을 보아 반향 억제 성능이 뛰어남을 알 수 있었다.

기존의 알고리즘과 제안된 알고리즘의 모든 실험 데이터에 대한 ERLE 및 SA 결과를 표 1에 나타내었다. 결과를 보아 알 수 있듯이, 제안된 알고리즘이 기존의 알고리즘보다 비선형 반향신호 억제율이 높으며, 근단화자 신호 왜곡이 작은 것을 확인 할 수 있다. 이를 통해 제안된 알고리즘의 비선형 반향신호 억제 성능이 기

존의 알고리즘보다 우수하며, 제안된 알고리즘이 기존의 알고리즘만으로는 제대로 추정하지 못했던 반향경로를 정확히 추정함으로써 근단화자 신호 손실도를 줄여 왜곡현상 또한 개선되었음을 확인하였다.

IV. 결 론

본 논문에서는 실제 비선형성을 포함하는 스피커 출력 환경에서 기존의 스피커와 마이크 사이의 선형관계만을 고려한 알고리즘의 문제점을 지적하고, 이를 개선하기 위해 주파수영역상에서의 Least Square 기반의 Volterra filter를 이용한 비선형 반향신호 억제기를 제안하였다. 스피커의 비선형성으로 인해 생성된 비선형 반향신호 환경에서, 제안된 알고리즘은 기존의 알고리즘보다 더 높은 반향신호 억제 성능을 보여주고 더 낮은 근단화자 신호 왜곡도를 보여주었다. 따라서 비선형성이 포함된 실제적인 스피커를 사용하는 환경에서, 제안된 알고리즘은 기존의 알고리즘보다 반향신호 억제 성능이 뛰어난 것을 확인하였다.

REFERENCES

[1] P.S.R. Diniz, *Adaptive Filtering: Algorithm and Practical Implementation*, Kluwer Academic Publishers, 1997.

[2] C. Faller and J. Chenm "Suppressing acoustic echo in a spectral envelope space," *IEEE Trans. On Speech and Audio Process.*, Vol. 13, No. 5, pp. 1048-1062, Sep. 2005.

[3] A. Guerin, G. Faucon, and R.L. Bouquin-Jeannes, "Nonlinear acoustic echo cancellation based on Volterra filters," *IEEE Trans. on Speech and Audio Process.*, Vol. 11, No. 6, pp. 672-683, Nov. 2003.

[4] F. Kuech and W. Kellermann, "Partitionaed block frequency-domain adaptive second-order Volterra filter," *IEEE Trans. on Signal Process.*, Vol. 53, No. 2, pp. 564-575, Feb. 2005.

[5] D. Comminiello, M. Scarpiniti, L.A. Azpicueta-Ruiz, J. Arenas-Garcia, and A. Uncini, "Functional link adaptive filters for nonlinear acoustic echo cancellation," *IEEE Trans. on Audio, Speech, and Langua. Process.*, Vol. 21, No. 7, pp. 1502-1512, Jul. 2013.

[6] S.G. McGovern, "Fast image method for impulse

response calculations of box-shaped rooms," *Applied Acoustics*, Vol. 70, No. 1, pp. 182-189, Jan. 2009.

[7] K.-H. Lee, J.-H. Chang, N. S. Kim, S. Kang, and Y. Kim, "Frequency-domain double-talk detection based on the Gaussian mixture model," *IEEE Signal Processing Letters*, Vol. 17, No. 5, pp. 453-456 May. 2010.

[8] S.Y. Lee and N.S. Kim, "A statistical model-based residual echo suppression," *IEEE Signal Processing Letters*, Vol. 14, No. 10, pp. 758-761, Oct. 2007.

저 자 소 개



박 지 환(학생회원)
2012년 부경대학교 전자정보통신 공학과 공학사 졸업.
2012년 한양대학교 전자컴퓨터 통신공학과 석박사통합 과정 재학중.

<주관심분야 : 음향신호처리, 음성신호처리>



이 봉 기(학생회원)
2010년 한양대학교 전자및통신 공학과 공학사 졸업.
2012년 한양대학교 전자컴퓨터 통신공학과 공학석사 졸업.
2012년 한양대학교 전자컴퓨터 통신공학과 박사과정 재학중.

<주관심분야 : 음성신호처리, 음성코덱>



장 준 혁(평생회원)
1998년 경북대학교 전자공학과 공학사 졸업.
2000년 서울대학교 전기공학부 공학석사 졸업.
2004년 서울대학교 전기컴퓨터 공학부 공학박사 졸업.

2004년 캘리포니아 주립대학, 산타바바라 (UCSB) 박사후 연구원.
2005년 한국과학기술연구원(KIST) 연구원.
2005년 인하대학교 전자공학부 조교수.
2011년 한양대학교 융합전자공학부 부교수.
<주관심분야 : 음성신호처리, 오디오신호처리, 통신신호처리, 휴먼/컴퓨터 인터페이스>