

음향학적 반향 제거를 위한 IP-INLMS 알고리즘

김민승*, 신종원^o

An IP-INLMS Algorithm for Acoustic Echo Cancellation

Minseung Kim*, Jong Won Shin^o

요약

본 논문은 음향학적 반향 제거를 위한 IP-INLMS 알고리즘을 제안하였다. 제안된 알고리즘은 원단 신호인 반향을 제외한 근단 신호가 존재하는 환경에서도 강인하게 동작하도록 스텝사이즈를 조절하는 INLMS 알고리즘에 필터의 수렴을 빠르게 하는 IP-NLMS 알고리즘의 개념을 결합하였다. 제안된 방법은 기존의 INLMS 알고리즘 대비 하여 음향학적 반향 제거 실험에서 부정합과 필터의 수렴속도 측면에서 모두 향상된 성능을 보여주었다.

Key Words : Acoustic echo cancellation, IP-NLMS, INLMS, gradient-adaptive algorithm, variable step-size algorithm

ABSTRACT

In this paper, the IP-INLMS algorithm for acoustic echo cancellation is proposed. The proposed algorithm introduces the concept of the IP-NLMS algorithm, which has fast initial convergence, to the INLMS algorithm, which has the robustness to the near-end interference by adjusting the step-size. In the experiment, the proposed method show the improved performance in terms of misalignment and convergence speed of the filter comparing to INLMS algorithm.

1. 서론

음향학적 반향은 마이크로폰과 라우드스피커 간의 음향학적 커플링(coupling)에 의해 발생한다. 이러한 음향학적 반향은 사용자에게 듣기 불편한 소리를 유발하며 음성 인식과 같은 음성 신호 후처리 모듈의 성능을 저해하는 요소이다. 이러한 음향학적 반향을 제거하기 위한 접근으로 음향학적 반향 제거 기술과 음향학적 반향 억제 기술이 있다. 음향학적 반향 제거 기술은 스피커와 마이크 간의 반향 경로를 적응형 알고리즘을 통해 추정하고, 마이크 신호에서 추정된 반

향 신호를 감하는 방식으로 설계된다¹⁾. 한편 음향학적 반향 억제 기술은 기존의 반향 경로를 추정하는 방법에서 벗어나 반향 스펙트럼을 추정하는 접근이다²⁾.

반향 경로를 추정하기 위한 적응형 알고리즘 중에서 대표적인 것으로는 정규화된최소평균제곱(NLMS)이 있다. NLMS 알고리즘은 레퍼런스 신호의 파워 크기 변화에 덜 민감하다는 장점이 있고, 주로 그 안정성이 스텝사이즈에 영향을 받는다. 통상적으로 스텝사이즈는 값이 클수록 빠른 수렴속도를 갖지만 높은 부정합(misalignment)을 갖고, 그 값이 작을수록 느린 수렴속도를 가지면서 낮은 부정합을 가진다. 그러므로

※ 이 연구는 2019년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임(10076583).

※ 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2019R1H1A1079910).

• First Author : Gwangju Institute of Science and Technology, School of Electrical Engineering and Computer Science, kms0603@gist.ac.kr, 학생회원

◦ Corresponding Author : Gwangju Institute of Science and Technology, School of Electrical Engineering and Computer Science, jwshin@gist.ac.kr, 정회원

논문번호 : 201911-264-A-RE, Received October 31, 2019; Revised December 3, 2019; Accepted December 3, 2019

알고리즘이 빠른 수렴속도와 낮은 부정합을 갖기 위해서는 상황에 따라 적절한 스텝사이즈를 추정하는 기술이 필요하다.

본 논문에서는 잡음이 존재하는 환경에서도 강인하게 동작하는 간섭-정규화된최소평균제곱(INLMS) 알고리즘^[3]에 빠른 수렴속도를 장점으로 갖는 개선된비례-정규화된최소평균제곱(IP-NLMS)^[4]를 도입함으로써 향상된 음향학적 반향 제거 알고리즘을 제안하였다.

II. 본 론

음향학적 반향 제거 시스템에서 시간 영역에서 관측된 마이크로폰 신호 $d(n)$ 은 다음과 같이 모델링 된다.

$$d(n) = \mathbf{x}^T(n)\mathbf{h}(n) + v(n) \quad (1)$$

여기에서 $\mathbf{h}(n) = [h_0(n), \dots, h_{L-1}(n)]^T$ 은 L 개의 길이로 모델링 된 시스템 벡터이고, $\mathbf{x}^T(n) = [x(n), \dots, x(n-L+1)]$ 은 원단 신호(far-end signal) L 개로 이루어진 벡터이고, $v(n)$ 은 근단 음성 신호(near-end speech)와 배경 잡음으로 이루어진 성분이다. 여기서 $v(n)$ 은 모델링 된 필터의 길이보다 실제 시스템의 길이가 길 때 발생한 반향 성분과 비선형 반향 신호 등 모델링된 선형 반향 성분을 제외한 성분들을 모두 포함한다. 에러 신호 $e(n)$ 은 다음과 같이 마이크로폰 신호에서 추정된 반향 신호를 감하여 얻어진다.

$$e(n) = d(n) - \mathbf{x}^T \hat{\mathbf{h}}(n) \quad (2)$$

추정되는 시스템 벡터 $\hat{\mathbf{h}}(n)$ 는 IP-NLMS 알고리즘 [4]에 따라 다음과 같이 업데이트 된다.

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu(n) \frac{G(n)\mathbf{x}(n)e(n)}{\mathbf{x}^T(n)G(n)\mathbf{x}(n)} \quad (3)$$

$$G(n) = \text{diag} [g_0(n), \dots, g_{L-1}(n)] \quad (4)$$

$$g_l(n) \equiv \frac{1-\kappa}{2L} + (1+\kappa) \frac{|\hat{h}_l(n)|}{2\|\hat{\mathbf{h}}(n)\|_1} \quad (5)$$

여기에서 $\mu(n)$ 은 스텝사이즈이고, $\hat{h}_l(n)$ 은 추정된 시스템의 l 번째 탭(tap), $\|\cdot\|_1$ 은 L1-norm,

$-1 \leq \kappa \leq 1$ 이다. $x(n)$ 과 $v(n)$ 의 평균이 각각 0 이고 서로 상관관계가 없다는 가정 하에서 최적의 스텝사이즈는 다음과 같다^[5].

$$\mu_{opt}(n) = \frac{E[r^2(n)]}{E[e^2(n)]} \quad (6)$$

여기에서 $r(n) = y(n) - \hat{y}(n)$ 은 잔여 반향 신호이다. INLMS 알고리즘^[3]에서는 잔여 반향 신호의 파워를 다음과 같이 추정한다.

$$E[r^2(n)] \approx \eta(n-1)E[\hat{y}^2(n)] \quad (7)$$

여기에서 $\eta(n)$ 은 정규화된 필터 부정합의 형태를 가지며 잔여 반향 신호의 파워를 직접 추정하는 것보다 상대적으로 추정이 쉽다. 단, 필터의 적응 초기에는 $\hat{y}(0) = 0$ 이므로, 적응 초기의 짧은 시간동안 고정된 스텝사이즈를 사용하여 적응을 진행시켜주는 과정이 필요하다^[3]. 스텝사이즈의 추정은 식 (7)을 식 (6)에 대입하고 스텝사이즈를 1이하로 제한하여 다음과 같이 표현된다.

$$\hat{\mu}(n) = \min(\eta(n-1) \frac{\sigma_y^2}{\sigma_e^2}, 1) \quad (8)$$

여기에서 σ_y^2 와 σ_e^2 은 각각 $E[\hat{y}^2(n)]$ 와 $E[e^2(n)]$ 의 추정값이며, 주로 회귀스무딩(recursive smoothing) 방법으로 구해진다. 에러 신호의 파워의 $\eta(n)$ 에 관한 그레디언트(gradient)는 다음과 같이 표현된다.

$$\frac{\partial e^2(n)}{\partial \eta(n-1)} = -2 \frac{\mathbf{x}^T(n)e(n)\Psi(n-1)}{\mathbf{x}^T(n)G(n)\mathbf{x}(n)} \times \frac{\sigma_y^2}{\sigma_e^2} \quad (9)$$

$$\Psi(n) = [\mathbf{I} - \frac{\mu(n)G(n)\mathbf{x}(n)\mathbf{x}^T(n)}{\mathbf{x}^T(n)G(n)\mathbf{x}(n)} + G(n)\mathbf{x}(n)e(n)]\Psi(n-1) \quad (10)$$

식 (9)에서 도출된 그레디언트를 기반으로 한 $\eta(n)$ 의 지수적(exponential) 업데이트 식은 다음과 같이 표현된다.

$$\eta(n) = \eta(n-1) \exp(\frac{\rho}{\sigma_e^2} \cdot \frac{\partial e^2(n)}{\partial \eta(n-1)}) \quad (11)$$

여기에서 ρ 는 학습 파라미터이고, σ_e^2 으로 나누어 주는 것은 차원이 없는 값을 얻기위함이다³⁾.

III. 실험

제안된 기술과 기존 기술들의 음향학적 반향 제거 성능을 비교하기 위한 실험 환경은 다음과 같다. 시스템의 임펄스응답과 필터의 길이(L)는 512를 사용하였고, 샘플링레이트(sampling rate)는 8 kHz이다.

입력 신호는 통계적으로 비정상적(non-stationary)인 음성 신호를 사용하였다. 사용된 음성 신호는 하버드문장(Harvard sentences)⁶⁾으로부터 임의로 추출된 12개의 문장으로 이루어져 있다. 입력 신호를 구성하기 위해 약 38초의 지속시간을 갖는 12개의 문장을 사용하였고 문장과 문장사이에는 약 0.5초의 짧은 사일런스 구간을 포함하도록 구성하였다. 비교적 긴 입력신호를 사용한 것은 적응알고리즘의 수렴 속도와 수렴 이후 정상상태(steady state)에서의 알고리즘의 행동을 모두 관찰하기 위함이다.

반향 경로의 갑작스러운 변화 상황에서 알고리즘의 재수렴 성능을 관찰하기 위해서 약 60초 이후에는 이전과 같은 문장이지만 초기와는 다른 반향 경로를 사용하여 반향 신호를 합성하였다. 반향 경로의 변화를 모사하기 위해 약 60초 이후에는 시스템의 임펄스응답을 오른쪽으로 12 샘플 시프트(shift) 하였다. 이렇게 실험적으로 모의된 음향학적 반향 신호 $y(n)$ 은 그림 1과 같다.

알고리즘이 배경 잡음에 얼마나 강인하게 동작하는지 관찰하기 위해서 합성된 음향학적 반향 신호에 추

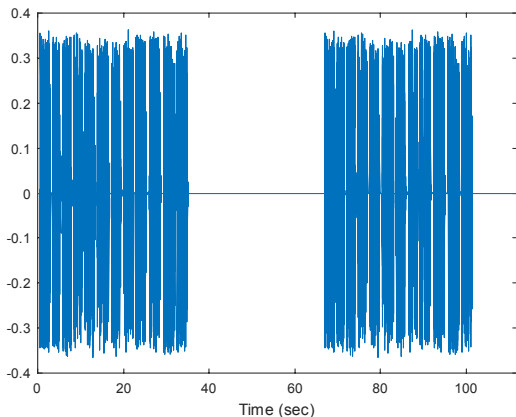


그림 1. 실험 모의된 시간 영역에서의 음향학적 반향 신호.
Fig. 1. Simulated acoustic echo signal in time domain. At 60 sec, echo path is changed.

가적으로 백색 잡음 신호를 섞어서 최종적으로 마이크로폰 신호 $d(n)$ 을 모의하였다. 이 때 신호대잡음비(SNR)는 10, 20, 그리고 30dB로 섞었다. 알고리즘의 성능 평가는 dB 스케일의 정규화된 부정합을 사용하였고 수식은 다음과 같다.

$$\text{misalignment}(n) = 20 \log_{10} \left[\frac{\|\mathbf{h} - \hat{\mathbf{h}}(n)\|_2}{\|\mathbf{h}\|_2} \right] \quad (12)$$

여기에서 $\|\cdot\|_2$ 은 L2-norm이다. 부정합은 추정된 시스템이 실제 시스템과 가까울수록 그 값이 낮아짐을 알 수 있다.

실험에서는 각각 0.05의 낮은 고정된 스텝사이즈(이하 기준방법1)와 1.00의 높은 고정된 스텝사이즈(이하 기준방법2)를 사용하는 기존 NLMS 알고리즘, INLMS[3] 알고리즘(이하 기준방법3)과 제안된 IP-INLMS 알고리즘(이하 제안방법)의 성능을 비교하였다.

그림 2-4에서는 시간에 따른 부정합을 비교하였는데, 부정합이 초기에는 빠르게 감소하다가 일정 시점 이후에는 그 변화량이 미미해지는 모습을 관찰할 수 있다. 여기서 부정합이 빠르게 감소하여 정상상태에 이르기까지 걸리는 시간으로 알고리즘의 수렴 속도가 얼마나 빠르는지 해석할 수 있고, 정상상태에서의 부정합 수치로 반향의 제거 성능을 확인할 수 있다. 또한 약 60초 이후의 부정합이 급격히 증가한 것은 반향 경로가 변화한 모습이고, 이후 알고리즘이 재수렴하면서 부정합이 다시 감소하는 모습을 관찰할 수 있다. 여기에서 알고리즘의 재수렴 성능을 확인할 수 있다. 배경 잡음의 강도에 따른 알고리즘의 성능은 각 그림을 개별적으로 관찰함으로써 확인할 수 있다.

그림 2에서는 SNR 30dB에서의 음향학적 반향 제거 성능을 비교하였는데 초기 수렴속도는 기준방법2, 기준방법3, 그리고 제안방법이 비슷하게 빠르고 수렴 후 부정합 성능은 제안방법의 성능이 가장 우수하다. 재수렴의 경우 제안방법, 기준방법2, 기준방법3, 기준방법1 순으로 빠르며 수렴 후 부정합 성능은 제안방법이 가장 우수하다. 하지만 재수렴 초기에 제안방법이 기준방법3에 비해 다소 불안정한 모습을 관찰할 수 있다.

그림 3에서는 SNR 20dB에서의 음향학적 반향 제거 성능을 비교하였는데 그림 2에 비해 기준방법2가 훨씬 불안정해진 모습을 관찰할 수 있다. 기준방법3은 부정합 측면에서 기준방법1과 기준방법2보다 우수하

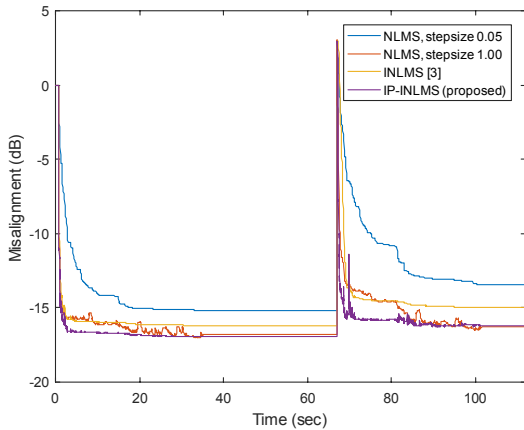


그림 2. 신호대잡음비 30dB에서의 제안한 방법과 베이스라인의 시간에 따른 부정합 성능 비교

Fig. 2. Comparison of misalignment performance between the proposed method and the baseline methods, SNR 30dB

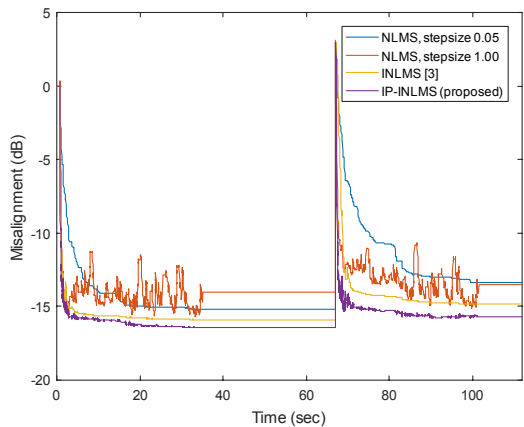


그림 3. 신호대잡음비 20dB에서의 제안한 방법과 베이스라인의 시간에 따른 부정합 성능 비교

Fig. 3. Comparison of misalignment performance between the proposed method and the baseline methods, SNR 20dB

지만 기존방법2에 비해 수렴속도가 다소 느리다. 제안 방법이 필터의 수렴속도와 부정합 측면에서 가장 우수한 성능을 보여준다.

그림 4에서는 SNR 10dB에서의 음향학적 반향 제거 성능을 비교하였는데 그림 2와 그림 3과 비교했을 때 전체적으로 필터들이 조금 더 불안정해진 모습을 관찰할 수 있다. 기존방법들의 경우 모두 필터의 수렴속도와 재수렴속도 또한 느려진 것을 관찰할 수 있는데 제안방법은 여전히 빠른 수렴속도와 재수렴속도를 보여준다. 제안방법이 부정합 성능 측면에서도 기존방법들에 비해 가장 우수하다. 기존의 잡음에 강인하게 동작하는 INLMS 알고리즘에 필터의 수렴속도의 측

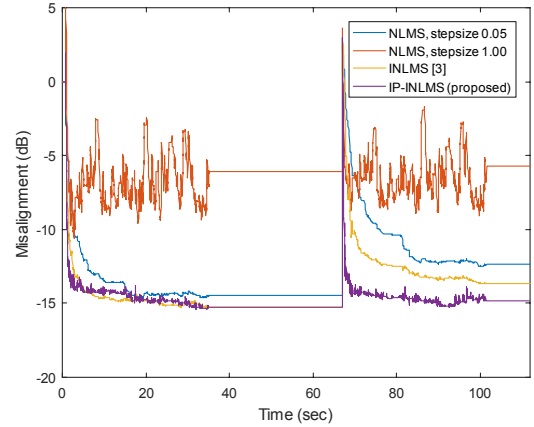


그림 4. 신호대잡음비 10dB에서의 제안한 방법과 베이스라인의 시간에 따른 부정합 성능 비교

Fig. 4. Comparison of misalignment performance between the proposed method and the baseline methods, SNR 10dB

면에서 우수한 IP-NLMS의 개념이 합쳐지면서 종합적으로 성능향상에 기여했다고 생각된다.

IV. 결론

본 논문에서는 INLMS 알고리즘과 IP-INLMS 알고리즘을 결합한 IP-INLMS 알고리즘을 제안하였다. IP-INLMS 알고리즘은 스텝사이즈를 잡음에 강인하게 조절하는 INLMS의 특성과 NLMS에 비해 수렴속도가 빠르다고 알려진 IP-NLMS의 개념을 융합함으로써 각 알고리즘의 장점을 가질 것으로 기대하였다. 제안한 알고리즘의 성능을 평가하기 위해 진행한 다양한 잡음의 SNR 환경의 음향학적 반향 제거 실험에서 제안한 방법은 기존 방법들에 비해 가장 우수한 음향학적 반향 제거 성능을 보여주었다.

References

- [1] S. Haykin, *Adaptive Filter Theory*, 4th Ed., Upper Saddle River, NJ: Prentice-Hall, 2002.
- [2] C. M. Lee, S. H. Bae, J. H. Kim, and N. S. Kim, "Spectro-temporal filtering based on soft decision for stereophonic acoustic echo suppression," *J. Commun. Networks (JCN)*, vol. 39, no. 12, pp. 1346-1351, Dec. 2014.
- [3] J. -M. Valin and I. B. Collings, "Interference-normalized least mean square algorithm," *IEEE Sign. Process. Lett.*, vol. 14, no. 12, 2007.
- [4] J. Benesty and S. L. Gay, "An improved

PNLMS algorithm,” *IEEE Int. Conf. Acoustics, Speech, and Sign. Process.*, pp. II-1881-II-1884, 2002.

- [5] C. Paleologu, S. Ciochina, J. Benesty, and S. L. Grant, “An overview on optimized NLMS algorithms for acoustic echo cancellation,” *EURASIP J. Advances in Sign. Process.*, vol. 2015:97, pp. 1-19, 2015.
- [6] *IEEE Recommended Practice for Speech Quality Measurements*, IEEE Standard 297-1969, pp. 1-24, Jun. 1969.

신 종 원 (Jong Won Shin)



2002년 2월 : 서울대학교 전기공학부 학사

2008년 8월 : 서울대학교 전기컴퓨터공학부 박사

2008년 12월~2012년 8월 : Qualcomm Inc., Senior Engineer

2012년 9월~현재 : 광주과학기술원 전기전자컴퓨터공학부 부교수

<관심분야> 음성/음향/오디오처리

[ORCID:0000-0002-8910-0264]

김 민 승 (Minseung Kim)



2016년 2월 : 광주과학기술원 전기전자컴퓨터공학부 학사

2018년 2월 : 광주과학기술원 전기전자컴퓨터공학부 석사

2018년 3월~현재 : 광주과학기술원 전기전자컴퓨터공학부 박사과정

<관심분야> 음성신호처리

[ORCID:0000-0002-2270-9382]