

원거리 음성 인식을 위한 효율적인 에코제거 시스템

Efficient Acoustic Echo Cancellation System for Distant-Talking Automatic Speech Recognition

김기범* · 김상윤* · 이우정* · 권민석* · 고병섭*

Ki-Beom Kim, Sang-Yoon Kim, Woo-Jung Lee, Min-Seok Kwon and Byeong-Seob Ko

Key Words : subband acoustic echo canceller (대역분할 에코제거기), filterbank (필터뱅크), adaptive filter (적응필터)

ABSTRACT

본 논문에서는, 원거리 음성인식을 위한 서브밴드 필터링 기반의 빠르고 효율적인 에코제거 시스템을 제안한다. 제안하는 에코제거 시스템은 우선 채널간 유사도 (correlation) 가 높을 경우 적응필터가 오작동하는 것을 방지하기 위해 spatial decorrelation 을 적용하게 된다. 그리고 tree 형태를 가지는 IIR filterbank 기반의 subband 구조를 채택함으로써, 적은 차수로도 효과적인 analysis, synthesis 필터링을 수행할 수 있도록 한다. 이 과정에서 불가피하게 발생하는 서브 밴드간 spectral aliasing은 notch filter를 적용해 해결할 수 있다. 또한 적응 필터로는 improved proportionate normalized least-mean-square (IP-NLMS) 알고리즘을 사용해 수렴속도 및 에코제거 성능에서 우수함을 확인하였다. 마지막으로 decision-directed estimation 기반의 residual echo suppressor를 적용해 잔여 에코를 제거하게 된다. 본 논문에서는 각 단계를 구성하게 된 이론적인 배경을 소개하고, 실제 에코가 존재하는 환경에서 ERLE, 원거리 음성 인식률, computational complexity를 통해 제안하는 에코제거 시스템의 효과를 입증하도록 한다.

기 호 설 명

- x : 레퍼런스 (TV) 신호
- y : 마이크 신호
- s : 타겟 음성 신호
- e : error 신호 (desired 신호)
- AEC : Acoustic Echo Canceller

1. 서 론

telecommunication infrastructure의 특성과 연관된 어쿠스틱 에코를 효과적으로 제거하는 것은 음성인식을 위한 음질 향상의 중요한 요소이다. 에코는 스피커의 출력신호가 다시 마이크로 입력되는 피드백 구조에 기인한다. Fullband (전대역) AEC가 이론상 간결한 구조 덕분에 에코 제거를 위한 기본적인 시

스템으로 인식될 수 있지만, subband (부대역) AEC는 빠른 수렴속도와 적은 연산량 때문에 전대역 AEC 보다 많은 응용분야에서 사용되고 있다^[1]. 본 논문은 스테레오 에코 제거를 위한 저연산량 기반의 효율적이고 강인한 AEC 시스템 구성을 목표로 하고, 이를 위해 제안하는 전체 AEC 블록도가 그림 1에서 보여 진다. 이 논문은 AEC를 구성하는 각 모듈의 필요성과 이론적인 내용을 검토하며 다음과 같이 구성되어 있다. 섹션 2에서 AEC를 위한 전처리 단계로서의 spatial decorrelation 기법을 설명하고, 그 다음 실제 에코를 제거하기 위한 부대역 AEC 가 섹션 3에서 보여 진다. 그리고 사용된 AEC 의 성능 보완과 향상을 위한 notch filter와 residual echo suppressor를 각각 섹션 4와 5에서 설명한다. 섹션 6에서는 제안하는 시스템의 성능을 평가하고 섹션 7에서 제안하는 시스템에 대한 결론을 내리도록 한다.

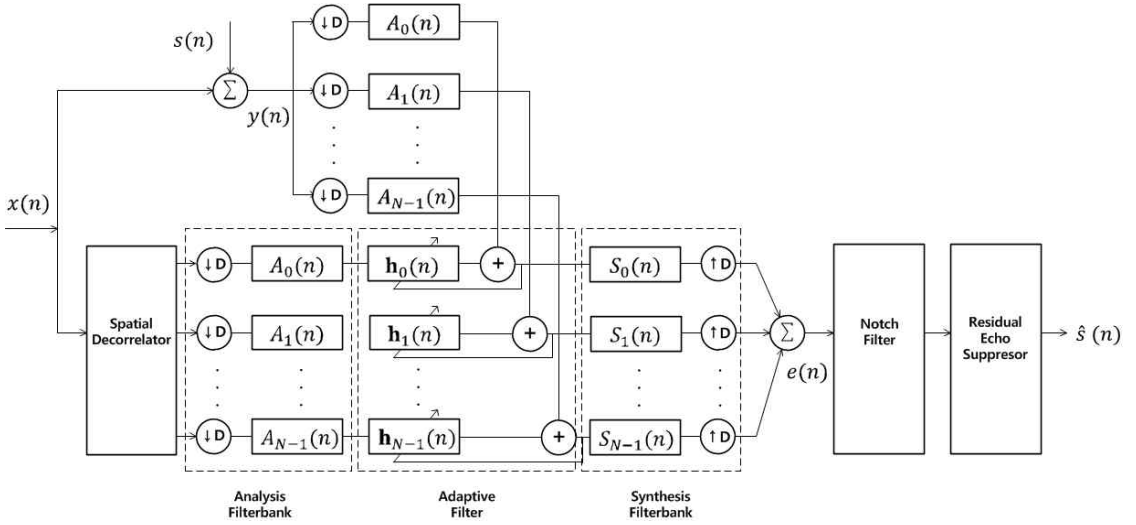


Figure 1. Proposed AEC system

2. Spatial Decorrelation

스테레오 출력을 지원하는 멀티미디어 기기의 스피커에서 발생하는 신호 간에 유사도가 높을 경우 AEC 내의 적응 필터는 쉽게 발산하고 이는 AEC의 오동작으로 이어질 수 있다. 이런 현상을 극복하기 위해 대부분의 AEC 시스템은 방송 신호에 다양한 방식으로 변형을 가해서 스피커 출력하게 된다.

본 논문에서는 방송 채널간의 유사도를 줄이기 위해 입력신호의 covariance를 identity matrix로 변환하여 입력되는 신호를 "well-posed" 로 만들어 주는 spatial decorrelation 기법^[2]을 적용해 AEC를 위한 전처리 작업을 수행하도록 한다. 채널간 신호의 유사도가 높은 입력신호의 covariance matrix는 ill-conditioned이며 이것은 AEC의 적응필터가 발산하는 직접적인 이유가 된다^[3].

spatial decorrelation의 목적은 레퍼런스 신호에 대하여 수식 (1) 을 만족하는 decorrelation matrix, W 을 구하는 것이다.

$$R_{xx} = E[xx^T] = I \quad (1)$$

이를 위해 우선 eigen-value decomposition을 수행해서 수식 (2)와 같이 입력 신호의 orthogonal matrix, V 와 diagonal matrix, Λ 을 구한다.

$$R_{xx} = V_x \Lambda_x V_x^T = V_x \Lambda_x^{1/2} \Lambda_x^{1/2} V_x^T \quad (2)$$

그 다음으로 decorrelation matrix, W 을 아래와 같이 설계한다.

$$W = \Lambda_x^{-1/2} V_x^T \quad (3)$$

또는, 임의의 orthogonal matrix, U 을 사용해서 수식 (4)와 같이 계산되어 질 수 있다.

$$W = U \Lambda_x^{-1/2} V_x^T \quad (4)$$

따라서 수식 (5) 과 같이 출력 신호는 mutually un-correlated 된다.

$$\begin{aligned} R_{xx} &= E[xx^T] \\ &= U \Lambda_x^{-1/2} V_x^T V_x \Lambda_x V_x^T V_x \Lambda_x^{-1/2} U^T \\ &= I_n \end{aligned} \quad (5)$$

3. Subband Adaptive Filtering

3.1 Filterbank

일반적으로 echo path 가 길 경우 전대역 AEC가 대응하려면 FIR 적응 필터의 차수가 매우 높아야 한다. 이것은 높은 연산량과 낮은 수렴속도를 야기하므로 적응필터로서 FIR을 사용하는 단점으로 부각된다. 대역분할 구조는 이러한 문제를 해결하기 위해 제안되었다^[4]. 특히 분할하고자 하는 대역의 수가 32 대역 이상일 경우, 적응 필터의 충격 응답의 길이가 증가할수록 계산량이 감소한다고 알려져 있으며, 128 대역의 경우 전대역 AEC 대비 50%에 가까운 계산량 감소를 기대할 수 있다^[5].

그림 2와 3은 대역 분할 구조를 위한 2밴드 버전의 analysis, synthesis 시스템을 보여준다. 실제적인 대역 분할 시스템에서는 2^n , ($n = 1, 2, 3, 4 \dots$) 개의 밴드를 사용하며, 전대역 입력을 1/N 다운샘플링 후 N 개의 band-pass 필터를 통과시켜 N 개의 대역 분할 신호를 얻게 된다.

이 대역분할 시스템을 구성하는 개별필터로서 IIR 필터를 사용하는 것은 FIR 필터나 DFT 필터를 사용하는 경우와 달리 매우 적은 계수만으로도 cut-off 주파수 부근에서 높은 감쇠 성능을 내도록 설계 수 있다^[6]. 개별 IIR 필터는 수식 (6) 과 같은 전달함수를 갖게 된다^[7].

$$H_0(z) = \frac{A_0(z^2) + z^{-1}A_1(z)}{2} \quad (6)$$

$$A_i(z^2) = \prod_{j=0}^{P_i-1} \frac{\alpha_{i,j} + z^{-2}}{1 + \alpha_{i,j}z^{-2}}, \quad i = 0, 1 \quad (7)$$

수식 (7)에서 $\alpha_{i,j}$ 는 all-pass filter 의 계수이며, 2차 all-pass filter를 대역 별로 연결하게 된다. $\alpha_{i,j}$ 는 lowpass, power-symmetric elliptic filter를 설계하는 절차^[8]에 따라 stopband edge 값과 peak stopband ripple 값을 지정 자동적으로 구하는 방식과 지정차수에 해당하는 계수를 사용하는 방식의 두 가지 기법을 통해 구할 수 있다. 설계 절차에 의해 자동적으로 $\alpha_{i,j}$ 을 구하는 전자의 경우 필터의 전달 함수가 과추정 되는 위험성이 있고, 주어진 스펙에 정확히 일치하는 필터뱅크를 만들기 위해 100 차 이상의 높은 차수의 IIR 필터를 생성하기도 한다. 본 논문에서는 연산량 및 인접 밴드와의 분리성과

군지연 등을 고려해 17차 IIR 필터를 사용하기로 한다^[9]. 표 1 은 IIR 필터의 계수를 보여준다.

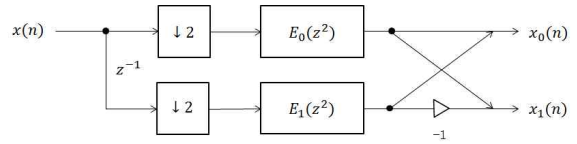


Figure 2. Two band analysis filter

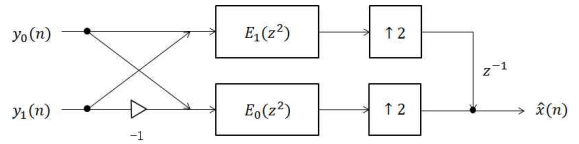


Figure 3. Two band synthesis filter

Table 1. IIR filter coefficients

Filter Coefficients	
Top branch	Bottom branch
0.06420239969799	0.22691819065413
0.42467376298912	0.60517810579077
0.74596983516706	0.84726220095035
0.91928513358894	0.97462589167035

3.2 adaptive filter

전통적으로 AEC의 적응필터로는 구현의 용이성 때문에 수식 (8) 의 Normalized Least Mean Square (NLMS) 알고리즘이 사용되었다^[10].

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + 2\mu \frac{\mathbf{x}(n+1)e(n+1)}{\|\mathbf{x}(n+1)\|^2 + \delta_{NLMS}} \quad (8)$$

NLMS는 적은 연산량으로도 mean-square error를 최소화하는 본래의 목적을 만족시키지만, 다양한 환경에서 성능의 안정성을 보장하지 못한다. 다음으로 빠른 수렴속도와 적응하고자 하는 데이터가 적은 환경에서의 에코 제거 성능 향상을 위해, 적응 가중치의 크기에 비례하는 업데이트 량을 반영하는 Proportionate NLMS (P-NLMS) 알고리즘이 개발되었으며^[11], 충격 응답이 분산성 (dispersive)을 가지게 되어 P-NLMS가 제대로 동작하지 못하게 되는

경우를 방지하고자, NLMS와 P-NLMS의 장점을 결합한 IP-NLMS 알고리즘^[12]이 개발되었다. 본 논문에서는 연산량 증가 대비 성능 향상을 고려해 부대역 적응필터로 IP-NLMS를 사용하였다. IP-NLMS는 수식 (9), (10), (11)의 P-NLMS 알고리즘의 동작 원리에 기반하며, 업데이트를 위한 가중치가 클 경우 그에 비례하는 증분을 반영하고, NLMS 기법으로부터 가중치와 무관한 부분을 동시에 고려해 수렴속도와 에코제거 성능을 향상시킨다. 또한 수식 (12)의 gain distributor는 가중치에 비례하는 양을 반영하므로 double-talk이 발생해 가중치 업데이트를 중단해야 할 경우에 작은 업데이트량을 보장할 수 있다. 2장에서 설명한 spatial decorrelator와 IP-NLMS 알고리즘을 적용하게 되면 일반적으로 복잡한 연산량을 가지는 double-talk detector (DTD)를 사용하지 않아도 double-talk 발생 시 적응필터의 발산을 억제할 수 있는 효과를 얻을 수 있다.

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu \frac{\mathbf{Q}(n) \mathbf{x}(n+1)e(n+1)}{\mathbf{x}^T(n+1)\mathbf{Q}(n)\mathbf{x}(n+1) + \delta} \quad (9)$$

$$\mathbf{Q}(n) = \begin{bmatrix} q_0(n) & 0 & \dots & 0 \\ 0 & q_1(n) & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & q_{L-1}(n) \end{bmatrix} \quad (10)$$

$$q_l(n) = \frac{k_l(n)}{\frac{1}{L} \sum_{i=0}^{L-1} k_i(n)} \quad (11)$$

$$k_l(n) = \frac{1-\alpha}{2L} + (1+\alpha) \frac{|h_l(n)|}{2 \|\hat{\mathbf{h}}(n)\|_1 + \epsilon} \quad (12)$$

4. Notch Filter

Analysis 필터링 시 저차수의 IIR lowpass 필터로 필터뱅크를 생성하는 것은 synthesis 필터링 후 밴드 간 중첩현상 (aliasing)을 야기하게 된다. 또한 연산량을 이유로 분할 대역 개수가 증가할 경우 발생하는 aliasing의 발생 횟수는 증가하게 된다.

일반적으로 aliasing으로부터 자유로운 대역 분할 시스템을 설계하는 것은 설계상의 제약 사항을 동반하거나, 매우 높은 차수의 bandpass 필터를 필요로 하지만, 이 밴드 간 aliasing은 comb 필터 기반의 노치 필터를 적용함으로써 일정부분 제거할 수 있다. 그림 4는 인접 밴드 사이의 aliasing을 제거하기 위한 노치 필터로서 부대역 사이에서 발생하는 주파수 성분들에 대해 -20 dB 감쇠 효과를 보인다.

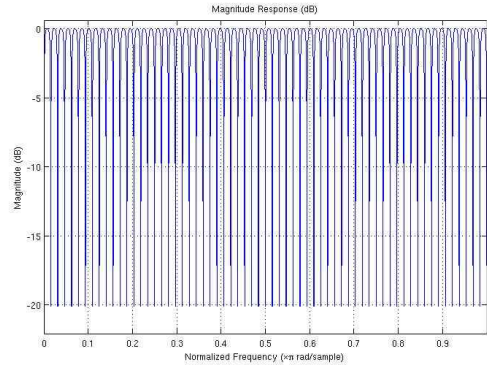


Figure 4. Magnitude response of notch filter

5. Residual Echo Suppression

AEC와 notch filter에서 제거하지 못한 잔여에코를 줄이기 위한 residual echo suppressor (RES)는 DTD의 도움 및 영향을 받지 않도록 하기 위해 일반적인 잡음제거기의 형태로 설계된다. 본 시스템에 적용된 RES는 decision-directed estimation^[13]기반의 spectral subtraction 구조를 가진다. 단 double-talk 구간에서 잔여에코가 과도하게 추정되는 것을 방지하기 위해^[14]에코의 파워 스펙트럼 추정 시 late reverberant spectral variance 추정에 일반적으로 사용되는 Polack's model^[15]을 변형해 수식 (13)와 같이 잔여에코의 파워 스펙트럼, $\hat{\lambda}_e$ 을 추정하도록 한다.

$$\hat{\lambda}_e(l, k) = e^{-2\rho(k) \frac{R}{f_s}} \hat{\lambda}_e(l-1, k) + c\lambda_x(l - \frac{N}{R}, k) \quad (13)$$

여기서 레퍼런스 신호의 파워스펙트럼, λ_x 은 수식 (14) 와 같이 first-order prediction을 따른다.

$$\lambda_x(l, k) = \max \left\{ \begin{array}{l} \eta \lambda_x(l-1, k) + (1-\eta) |X(l, k)|^2, \\ \lambda_{\min} \end{array} \right\} \quad (14)$$

6. 실험 결과

테스트를 위한 실험환경이 그림 5에 나타난다. 테스트 환경은 RT₆₀ 이 400~500 ms 인 일반 거실이며, 두 마이크 사이의 간격은 8 cm, 화자는 TV 정면을 기준으로 4m 에 위치해 있다. 화자 레벨은 60dB, 방송 레벨은 볼륨 20 으로 각각 세팅하였다.

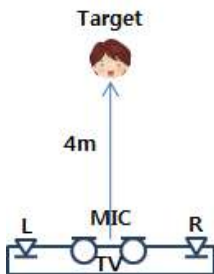


Figure 5. Test Environment

그림 6 의 상단과 하단은 각각 알고리즘 처리 전과 후를 나타낸다. 제안하는 AEC 시스템은 방송 신호의 고출력 상황에서도 1초 내의 수렴 속도를 보이며 평균 35.7 dB 의 ERLE (Echo Return Loss Enhancement) 향상을 보였다.

표 2는 2명의 화자가 뉴스, 예능 등의 정규 방송 하에서 총 50 회의 “Hi TV” 를 발성했을 때, Nuance VoCon 3200을 이용한 음성인식률을 나타낸다. 음성 인식률을 성능 지표로써 사용하기 위해 다양하고 많은 DB에 대한 테스트가 필요하다 할지라도, 본 논문에서 제안하는 AEC 시스템이 원음의 성분들을 크게 손상시키지 않음을 표 2를 통해 알 수 있다.

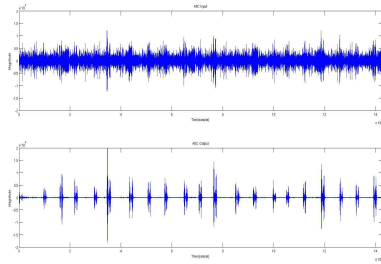


Figure 6. Waveform Before/After AEC

Table 2. Recognition Rate

	speaker 1	speaker 2
Recognition Rate (%)	95.24	96.55

마지막으로 본 논문에서 제안하는 시스템의 각 실시간 동작 모듈에 대한 복잡도 (computational complexity) 가 표 3에서 보여 진다. 본 시스템의 복잡도는 1채널 출력을 기준으로 ARM9EJ-S DSP 코어 기반의 CodeWarrior 상에서 MCPS (Million Clock Per Second) 단위로 측정하였다. 제안하는 시스템은 ARM9 코어 상에서 168.6 MCPS 의 저연산량으로도 효과적인 에코 제거 성능을 보임을 나타낸다.

Table 3. Computational Complexity

	MCPS
Filterbank	29.2
Adaptive Filter	87.0
Notch Filter	3.15
Residual Echo Suppressor	49.2
TOTAL	168.6

7. 결 론

본 논문에서는 IIR 필터뱅크 기반의 고성능 경량 AEC 시스템이 제안되었다. DTD를 필요로 하지 않는 부대역 적응 필터와 RES, 그리고 저연산량 필터뱅크의 단점을 보완할 수 있는 노치 필터로 이루어진 시스템을 제안하였으며, ARM9 기준 168.6 MCPS 의 연산량으로 35.7 dB 의 ERLE 향상과 96% 의 음성인식률을 얻을 수 있었다.

참 고 문 헌

- (1) Naylor P. A., O. Tanrıku and A. G. Constantinides, "Subband adaptive filtering for acoustic echo control using allpass polyphase IIR filterbanks", IEEE Trans. on S&P, vol. 6, no. 2, 1998.
- (2) N. G. Cho, J. Y. Cho, J. W. Lee, and Y. J. Kim, "Stereophonic acoustic echo cancellation using spatial decorrelation", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 93-96, 2011.
- (3) Benesty, J., Morgan, D. R., and Sondhi, M. M. . "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation", Speech and Audio Processing, IEEE Transactions on, vol. 6, no. 2, pp. 156-165, 1998.
- (4) K. A. Lee, W. S. Gan, and S. M. Kuo, "Subband Adaptive Filtering : Theory and Implementation", Hoboken, NJ: Wiley, 2009.
- (5) M. C. Shin, and S. M. Kim, "a comparative study of full-band and sub-band approaches to acoustic echo cancellation", Korean Society for Noise and Vibration Engineering, Autumn Conference, pp. 645-651, 2003
- (6) P. P. Vaidyanathan, "'Multirate digital filters, filter banks, polyphase networks, and applications : A tutorial,'" Proc. IEEE, vol. 78, pp. 56-93, Jan. 1990.
- (7) O. Tanrikulu, and A. G. Constantinides, "Cascaded power symmetric IIR filter banks and continuity constrained adaptive algorithms for acoustic echo cancellation in subbands." Signal Processing, IEEE Transactions on vol. 46, no. 8, pp. 2220-2225, 1998.
- (8) P.P. Vaidyanathan, "Multirate Systems and Filter Banks", Englewood Cliffs, NJ : Prentice Hall, 1993.
- (9) A. Krukowski, and I. Kale. "Polyphase IIR filter banks for subband adaptive echo cancellation applications." Circuits and Systems, 2003. ISCAS'03. Proceedings of the 2003 International Symposium on, vol. 4, pp. 405-408, IEEE, 2003.
- (10) Widrow. B, et al. "Adaptive noise cancelling : Principles and applications." Proceedings of the IEEE, vol. 63, no. 12, pp. 1692-1716, 1975.
- (11) Duttweiler, D. L., "Proportionate normalized least-mean-squares adaptation in echo cancelers", Speech and Audio Processing, IEEE Transactions on, vol. 8, no. 5, 508-518, 2000.
- (12) Benesty. J., and Steven L. Gay. "An improved PNLMS algorithm." Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on. Vol. 2, pp. 1881-1884, 2002.
- (13) Ephraim, Y. and Malah, D., "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 32, no. 6, pp. 1109-1121, 1984.
- (14) Habets, E. A., Gannot, S., & Cohen, I., "Late reverberant spectral variance estimation based on a statistical model", Signal Processing Letters, IEEE, vol. 16, no. 9, pp. 770-773, 2009.
- (15) Polack. J., "La transmission de lenergie sonore dans les salles", Ph.D., Univ. Maine, Le Mans, France, 1988.