Stereo Acoustic Echo Cancellation Based on Maximum Likelihood Estimation with Inter-channel-Correlated Echo Compensation

Byung Joon Cho and Hyung-Min Park (지능정보처리연구실)

SOGANG UNIVERSITY

Abstract

This paper presents batch and online algorithms of a stereo acoustic echo cancellation (SAEC) method. In SAEC, the non-uniqueness problem causes performance degradation, especially for highly coherent far-end signals. In our method, this problem can be avoided without an additional decorrelation preprocessor or multi-microphones by overestimating far-end echoes and compensating for the overestimated inter-channel-correlated echo to obtain a desired echo-canceled signal. In addition, our method is based on the maximum likelihood estimation (MLE) criterion of the echo-canceled signal under the assumption that the signal in the time-frequency domain follows a zero-mean complex Gaussian distribution with a time-varying variance. Furthermore, a variable forgetting factor based on the cross spectral density (CSD) between the echo-canceled signal and a far-end echo is presented in the online algorithm to improve the convergence of adaptive filters with a high cancellation performance when converged. Experimental results under various conditions demonstrate that the proposed method can successfully perform SAEC even in the presence of inter-channel correlation, double-talk, and abrupt echo path changes.

Problem Formulation

The microphone input signal

$$X[m,k] \approx S[m,k] + \sum_{v\in\nu}\sum_{l=0}^{L-1} H^{(v)^*}[l,k]U^{(v)}[m-l,k]$$

The echo-canceled signal

$$Y[m,k] \approx X[m,k] - \sum_{v\in\nu}\sum_{l=0}^{L-1} \hat{H}^{(v)^*}[l,k]U^{(v)}[m-l,k]$$

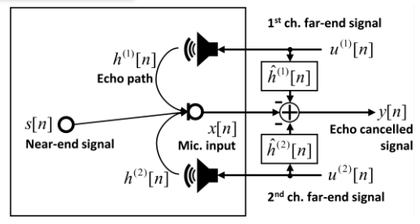

Fig. 1. Block diagram of the SAEC in the time domain

We estimate far-end echoes without considering the inter-channel correlation first, to deal with SAEC without the additional decorrelation. However, if the echo of each channel is estimated, the inter-channel-correlated echo is overestimated due to the overlap between the two estimated echoes. Therefore, we compensate for the overestimated inter-channel-correlated echo based on an overestimation ratio to avoid an oversubtraction of the inter-channel-correlated echo. The echo-canceled signal is calculated as follows:

$$Y[m,k] = X[m,k] - \sum_{v\in\nu}\mathbf{g}_k^{(v)^H}\mathbf{u}_{m,k}^{(v)} + \frac{\zeta_k}{2}\sum_{v\in\nu}\mathbf{g}_k^{(v)^H}\mathbf{u}_{m,k}^{(v)}.$$

- ζ_k : The overestimation ratio to compensate for the inter-channel-correlated echo

Proposed SAEC algorithm

We assume that the echo-canceled signal in each frequency bin follows a zero-mean complex Gaussian distribution with a time-varying variance to cope with the non-stationarity of the signal.

- The input signal for the *v*-th channel

$$X^{(v)}[m,k] = S[m,k] + D_u^{(v)}[m,k] + \left(\omega_k^{(1)} + \omega_k^{(2)}\right) D_c[m,k]$$

- The desired echo-canceled signal using $X^{(v)}[m,k]$

$$Y[m,k] = X^{(v)}[m,k] - \mathbf{g}_k^{(v)^H}\mathbf{u}_{m,k}^{(v)}$$

- The SAEC problem to find the unknown parameter sets

$$\hat{\Theta}_{m,k}^{(v)} = \arg\max_{\Theta_{m,k}^{(v)}} \mathcal{L}(\Theta_{m,k}^{(v)})$$

$$\mathcal{L}(\Theta_{m,k}^{(v)}) = \sum_m \log p(Y[m,k];\Theta_{m,k}^{(v)}) = -\sum_m\left(\log(\pi\lambda_Y[m,k]) + \frac{\left|X^{(v)}[m,k] - \mathbf{g}_k^{(v)^H}\mathbf{u}_{m,k}^{(v)}\right|^2}{\lambda_Y[m,k]}\right)$$

- The iterate update rule for $\mathbf{g}_k^{(v)}$

$$\mathbf{g}_k^{(v)} = \mathbf{R}_k^{(v)^{-1}}\mathbf{r}_k^{(v)} \quad \mathbf{R}_k^{(v)} = \sum_m \frac{\mathbf{u}_{m,k}^{(v)}\mathbf{u}_{m,k}^{(v)^H}}{\lambda_Y[m,k]} \quad \mathbf{r}_k^{(v)} = \sum_m \frac{\mathbf{u}_{m,k}^{(v)}X^{(v)^*}[m,k]}{\lambda_Y[m,k]} \quad \lambda_Y[m,k] = |Y[m,k]|^2$$

- The overestimation ratio

$$\zeta_k = \min\left(\frac{2\left|\Phi^{(1)(2)}[k]\right|}{\Phi^{(1)}[k] + \Phi^{(2)}[k]}, 1\right)$$

$$\Phi^{(v)}[k] = \frac{1}{N_m}\sum_m \left|\mathbf{g}_k^{(v)^H}\mathbf{u}_{m,k}^{(v)}\right|^2$$

$$\Phi^{(1)(2)}[k] = \frac{1}{N_m}\sum_m \left(\mathbf{g}_k^{(1)^H}\mathbf{u}_{m,k}^{(1)}\right)^*\left(\mathbf{g}_k^{(2)^H}\mathbf{u}_{m,k}^{(2)}\right)$$

Online SAEC based on frame-by-frame processing is required to cope with various environments in which the echo path may change. Similar to the batch parameter update rules, online parameter update rules can be derived based on RLS. The log-likelihood function to derive the online algorithm is defined as follows:

$$\mathcal{L}(\Theta_{m,k}^{(v)}) = \sum_{\mu=1}^{m}\gamma^{m-\mu}\log p(Y[\mu,k];\Theta_{\mu,k}^{(v)}) = -\sum_{\mu=1}^{m}\gamma^{m-\mu}\left(\log(\pi\lambda_Y[\mu,k]) + \frac{\left|X^{(v)}[\mu,k] - \mathbf{g}_{\mu,k}^{(v)^H}\mathbf{u}_{\mu,k}^{(v)}\right|^2}{\lambda_Y[\mu,k]}\right).$$

- Update of the adaptive filter for the first channel

1 : $\hat{\Phi}^{(v)}[m,k] = \alpha\Phi^{(v)}[m-1,k] + (1-\alpha)\left|\mathbf{g}_{m-1,k}^{(v)^H}\mathbf{u}_{m,k}^{(v)}\right|^2, v\in\{1,2\}.$

2 : $\hat{\Phi}^{(1)(2)}[m,k] = \alpha\Phi^{(1)(2)}[m-1,k] + (1-\alpha)\left(\mathbf{g}_{m-1,k}^{(1)^H}\mathbf{u}_{m,k}^{(1)}\right)^*\left(\mathbf{g}_{m-1,k}^{(2)^H}\mathbf{u}_{m,k}^{(2)}\right)$

3 : $\hat{\zeta}_{m,k} = \min\left(\frac{2\left|\hat{\Phi}^{(1)(2)}[m,k]\right|}{\hat{\Phi}^{(1)}[m,k] + \hat{\Phi}^{(2)}[m,k]}, 1\right)$

4 : $\hat{Y}[m,k] = X[m,k] - \sum_{v\in\nu}\mathbf{g}_{m-1,k}^{(v)^H}\mathbf{u}_{m,k}^{(v)} + \frac{\hat{\zeta}_{m,k}}{2}\sum_{v\in\nu}\mathbf{g}_{m-1,k}^{(v)^H}\mathbf{u}_{m,k}^{(v)}$

5 : $\hat{\lambda}_Y[m,k] = \left|\hat{Y}[m,k]\right|^2$

6 : $\mathbf{k}_{m,k}^{(1)} = \frac{\mathbf{\Psi}_{m-1,k}^{(1)^{-1}}\mathbf{u}_{m,k}^{(1)}}{\gamma\hat{\lambda}_Y[m,k] + \mathbf{u}_{m,k}^{(1)^H}\mathbf{\Psi}_{m-1,k}^{(1)^{-1}}\mathbf{u}_{m,k}^{(1)}}$

7 : $\mathbf{\Psi}_{m,k}^{(1)^{-1}} = \gamma^{-1}\left(\mathbf{\Psi}_{m-1,k}^{(1)^{-1}} - \mathbf{k}_{m,k}^{(1)}\mathbf{u}_{m,k}^{(1)^H}\mathbf{\Psi}_{m-1,k}^{(1)^{-1}}\right)$

8 : $\mathbf{g}_{m,k}^{(1)} = \mathbf{g}_{m-1,k}^{(1)} + \mathbf{k}_{m,k}^{(1)}\hat{Y}^*[m,k]$

- Update of the adaptive filter for the second channel

9 : $\Phi^{(1)}[m,k] = \alpha\Phi^{(1)}[m-1,k] + (1-\alpha)\left|\mathbf{g}_{m,k}^{(1)^H}\mathbf{u}_{m,k}^{(1)}\right|^2$

10 : $\tilde{\Phi}^{(1)(2)}[m,k] = \alpha\Phi^{(1)(2)}[m-1,k] + (1-\alpha)\left(\mathbf{g}_{m,k}^{(1)^H}\mathbf{u}_{m,k}^{(1)}\right)^*\left(\mathbf{g}_{m-1,k}^{(2)^H}\mathbf{u}_{m,k}^{(2)}\right)$

11 : $\tilde{\zeta}_{m,k} = \min\left(\frac{2\left|\tilde{\Phi}^{(1)(2)}[m,k]\right|}{\Phi^{(1)}[m,k] + \hat{\Phi}^{(2)}[m,k]}, 1\right)$

12 : $\tilde{Y}[m,k] = X[m,k] - \mathbf{g}_{m,k}^{(1)^H}\mathbf{u}_{m,k}^{(1)} - \mathbf{g}_{m-1,k}^{(2)^H}\mathbf{u}_{m,k}^{(2)} + \frac{\tilde{\zeta}_{m,k}}{2}\left(\mathbf{g}_{m,k}^{(1)^H}\mathbf{u}_{m,k}^{(1)} + \mathbf{g}_{m-1,k}^{(2)^H}\mathbf{u}_{m,k}^{(2)}\right)$

13 : Repeat Steps 5-8 with $\hat{Y}[m,k]$ and (1) replaced by $\tilde{Y}[m,k]$ and (2).

- Calculation of the echo canceled signal

14 : $\Phi^{(2)}[m,k] = \alpha\Phi^{(2)}[m-1,k] + (1-\alpha)\left|\mathbf{g}_{m,k}^{(2)^H}\mathbf{u}_{m,k}^{(2)}\right|^2$

15 : $\Phi^{(1)(2)}[m,k] = \alpha\Phi^{(1)(2)}[m-1,k] + (1-\alpha)\left(\mathbf{g}_{m,k}^{(1)^H}\mathbf{u}_{m,k}^{(1)}\right)^*\left(\mathbf{g}_{m,k}^{(2)^H}\mathbf{u}_{m,k}^{(2)}\right)$

16 : $\zeta_{m,k} = \min\left(\frac{2\left|\Phi^{(1)(2)}[m,k]\right|}{\Phi^{(1)}[m,k] + \Phi^{(2)}[m,k]}, 1\right)$ **17** : $Y[m,k] = X[m,k] - \sum_{v\in\nu}\mathbf{g}_{m,k}^{(v)^H}\mathbf{u}_{m,k}^{(v)} + \frac{\zeta_{m,k}}{2}\sum_{v\in\nu}\mathbf{g}_{m,k}^{(v)^H}\mathbf{u}_{m,k}^{(v)}$

Experimental Results

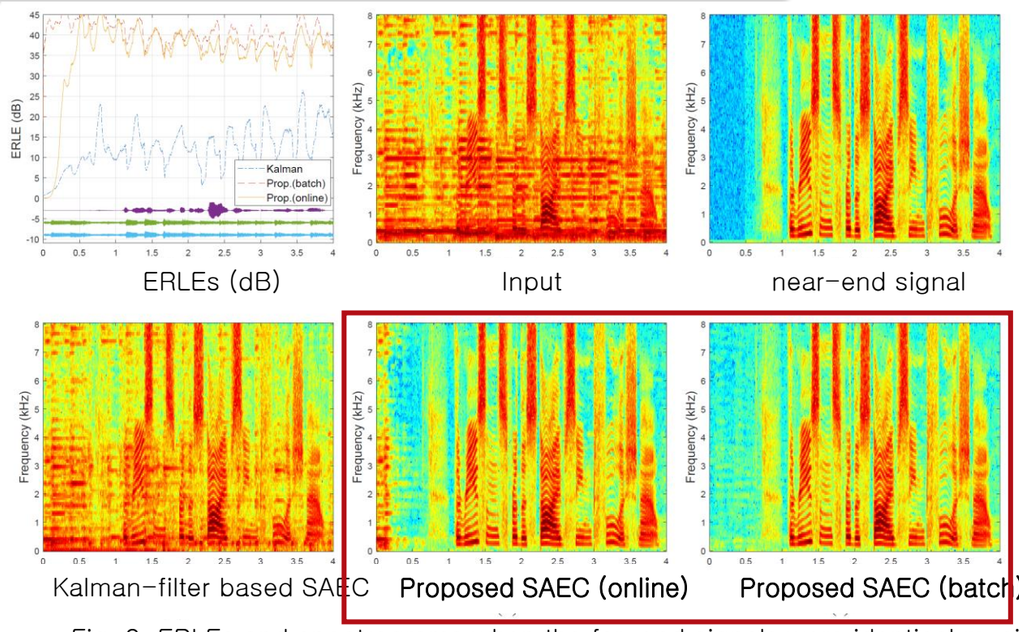

Fig. 2. ERLEs and spectrograms when the far-end signals were identical music signals.

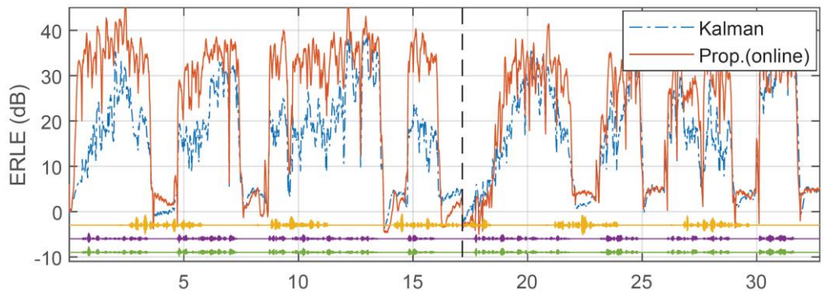

Fig. 2. ERLEs of the Kalman-filter-based and proposed online SAEC methods without the decorrelation preprocessor. Both echo paths changed in approximately 17.15 s, as denoted by the black dashed line.

Table 1. ERLE (dB) averaged over 370 utterances

Input SER	IR type	Far-end signals	Method		
			Kalman	Prop. (online)	Prop. (batch)
0 dB	Random IR	Speech	8.08	26.08	**32.62**
		Music	12.03	31.57	**39.02**
		Speech, Music	7.44	21.42	**33.07**
	RIR	Speech	16.66	26.12	**32.44**
		Music	19.50	32.10	**39.20**
		Speech, Music	17.03	22.18	**32.70**
5 dB	Random IR	Speech	7.00	24.07	**30.56**
		Music	10.98	29.91	**36.91**
		Speech, Music	6.70	20.29	**32.24**
	RIR	Speech	14.82	23.54	**30.30**
		Music	17.73	30.87	**37.10**
		Speech, Music	15.34	21.05	**31.89**
10 dB	Random IR	Speech	5.88	21.74	**27.55**
		Music	9.85	27.90	**33.67**
		Speech, Music	5.91	18.92	**29.49**
	RIR	Speech	12.88	20.67	**27.24**
		Music	16.03	29.27	**33.89**
		Speech, Music	13.71	19.61	**29.36**

Table 2. PESQ scores averaged over 370 utterances

Input SER	IR type	Far-end signals	Method			
			No proc.	Kalman	Prop. (online)	Prop. (batch)
0 dB	Random IR	Speech	1.87	2.33	3.86	**4.06**
		Music	1.91	2.52	4.05	**4.23**
		Speech, Music	1.80	2.08	3.31	**3.98**
	RIR	Speech	1.83	2.95	3.77	**4.06**
		Music	1.66	2.94	3.85	**4.16**
		Speech, Music	1.63	2.74	3.20	**3.90**
5 dB	Random IR	Speech	2.20	2.58	4.01	**4.19**
		Music	2.27	2.80	4.17	**4.30**
		Speech, Music	2.16	2.40	3.56	**4.15**
	RIR	Speech	2.15	3.16	3.91	**4.19**
		Music	2.00	3.15	4.04	**4.27**
		Speech, Music	1.98	2.97	3.47	**4.09**
10 dB	Random IR	Speech	2.53	2.84	4.11	**4.26**
		Music	2.61	3.07	4.25	**4.34**
		Speech, Music	2.51	2.69	3.78	**4.22**
	RIR	Speech	2.49	3.37	4.02	**4.26**
		Music	2.35	3.35	4.18	**4.32**
		Speech, Music	2.33	3.19	3.70	**4.20**

Conclusion

In this paper, we proposed and derived batch and online algorithms of an SAEC method based on the MLE of an echo-canceled signal by assuming that the signal in the time-frequency domain follows a zero-mean complex Gaussian distribution with a time-varying variance. To avoid the non-uniqueness problem without an additional decorrelation preprocessor or multi-microphones even for highly coherent far-end signals, the proposed method obtained an echo-canceled signal by overestimating far-end echoes and compensating for an overestimated inter-channel-correlated echo based on the overestimation ratio. In addition, a variable forgetting factor based on the CSD between the echo-canceled signal and a far-end echo in the online algorithm was presented to improve the convergence of the adaptive filters significantly with a high ERLE when converged. Experimental results under various conditions demonstrated that the proposed method achieved a superior performance compared with the Kalman-filter-based method even in the presence of inter-channel correlation, double-talk, and abrupt echo path changes.

서강대학교 SOGANG UNIVERSITY

2020퀄컴논문대전 최우수논문