STEREOPHONIC ACOUSTIC ECHO CANCELLATION USING BLIND SOURCE SEPARATION AS POST-PROCESSING

Takatoshi Okuno and M. O. Tokhi

Department of Automatic Control and Systems Engineering,
The University of Sheffield,
Mappin Street, Sheffield, S1 3JD, UK.
{cop98to, o.tokhi}@sheffield.ac.uk

ABSTRACT

Stereophonic or multi-channel acoustic echo cancellation systems suffer from cross-talk [1, 2]. This problem is also common to other multi-channel sound systems. Almost all attempts made to solve this problem are based on pre-processing techniques, which decorrelate the multi-channel input signals [3, 4]. However, pre-processing of these signals leads to a degradation of the audio quality, and further attempts need to be made to improve the audio quality [5]. Conversely, post-processing has the potential of avoiding these problems and producing a natural sound.

This paper investigates post-processing in stereophonic acoustic echo cancellation (SAEC) using the blind source separation (BSS) technique. It is demonstrated through computer simulation that reasonable results are obtained with this proposed method.

1. INTRODUCTION

This paper presents a new concept for SAEC using post-processing techniques. It is noted in the literature that almost all previously reported research has used pre-processing to decorrelate stereo input signals. This means that the stereo sound emitted from loudspeakers in the near-end is degraded by this pre-processing. Post-processing is a more natural approach than pre-processing. Considering post-processing, the problem can be regarded as a signal separation problem, such as BSS.

Figure 1 shows the configuration of the proposed SAEC system, where $\boldsymbol{h}_{11}$, $\boldsymbol{h}_{21}$, $\boldsymbol{h}_{12}$ and $\boldsymbol{h}_{22}$ represent the four acoustic paths of the stereophonic communication system. Using these notations, the relationships between the input signal vectors $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and the desired signal vectors $\boldsymbol{d}_1$, $\boldsymbol{d}_2$ can be obtained as

$$\boldsymbol{d}_1 = \boldsymbol{h}_{11} * \boldsymbol{u}_1 + \boldsymbol{h}_{21} * \boldsymbol{u}_2 \quad (1)$$

$$\boldsymbol{d}_2 = \boldsymbol{h}_{12} * \boldsymbol{u}_1 + \boldsymbol{h}_{22} * \boldsymbol{u}_2 \quad (2)$$

The authors thank Mr. E. Begic for fruitful comments in this research.

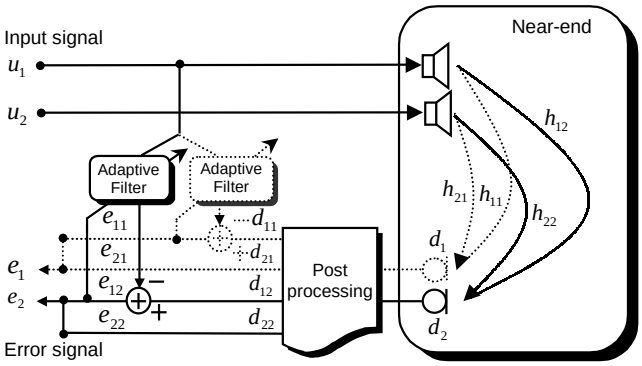

Figure 1: The configuration of a proposed SAEC system. (adaptive filters are depicted only for input signal $\boldsymbol{u}_1$.)

where $*$ denotes the convolution operation between two vectors. Equations (1) and (2) can be collected into a single equation in the frequency domain using a matrix expression as

$$\begin{bmatrix} D_1 \\ D_2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{21} \\ H_{12} & H_{22} \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}. \quad (3)$$

It is easily noted that there are four unknown valiables, namely H_{11}, H_{21}, H_{12} and H_{22}, and only one deficient set of simultaneous equations.

2. THE PROPOSED POST-PROCESSING

Let the four ideal desired signal vectors $\boldsymbol{d}_{11}$, $\boldsymbol{d}_{21}$, $\boldsymbol{d}_{12}$ and $\boldsymbol{d}_{22}$ in Figure 1 be expressed, using equations (1) and (2), as

$$\boldsymbol{d}_{11} = \boldsymbol{h}_{11} * \boldsymbol{u}_1 + \alpha \quad (4)$$

$$\boldsymbol{d}_{21} = \boldsymbol{h}_{21} * \boldsymbol{u}_2 + \beta \quad (5)$$

$$\boldsymbol{d}_{12} = \boldsymbol{h}_{12} * \boldsymbol{u}_1 + \gamma \quad (6)$$

$$\boldsymbol{d}_{22} = \boldsymbol{h}_{22} * \boldsymbol{u}_2 + \delta \quad (7)$$

where α, β, γ and δ are defined as the corresponding separation errors. In this case, conventional adaptive

filtering (for example, LMS, RLS) could be utilised to estimate the impulse responses $\boldsymbol{h}_{11}$, $\boldsymbol{h}_{21}$, $\boldsymbol{h}_{12}$ and $\boldsymbol{h}_{22}$.

To obtain these ideal desired signals separately, consider a BSS system as shown in Figure 2. This has been reported by Torkkola [6, 7] for the blind separation of convolved sources and is based on the concept of information maximisation (Infomax) [8].

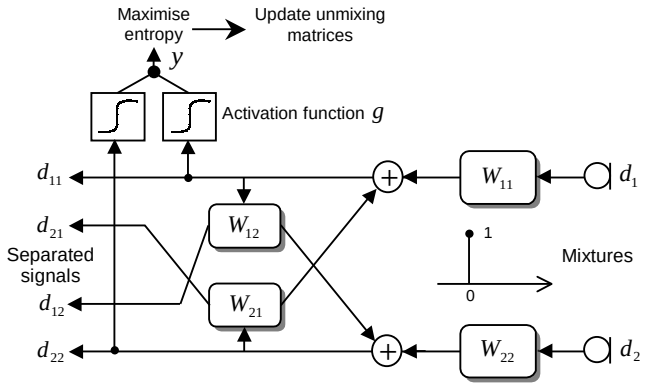

Figure 2: The configuration of the feedback architecture of BSS using Infomax as post-processing.

As shown in Figure 2, the relationships between the input and output are described through the feedback architecture as

$$d_{11}(n) = \sum_{k=0}^{N-1} w_{1k1} d_1(n-k) + \sum_{k=1}^{N-1} w_{2k1} d_{22}(n-k), \quad (8)$$

$$d_{21}(n) = -\sum_{k=1}^{N-1} w_{2k1} d_{22}(n-k), \quad (9)$$

$$d_{12}(n) = -\sum_{k=1}^{N-1} w_{1k2} d_{11}(n-k), \quad (10)$$

$$d_{22}(n) = \sum_{k=0}^{N-1} w_{2k2} d_2(n-k) + \sum_{k=1}^{N-1} w_{1k2} d_{11}(n-k), \quad (11)$$

where N denotes the number of filter coefficients and w_{ikj} is the k-th filter coefficient of the filter W_{ij}. Although d_{21} and d_{12} are not in general utilised in applications employing BSS, these signals are used here as output signals as well as to cancel cross-talk.

The stochastic adaptation rule for coefficients of filters W_{ij} on the basis of the Infomax criterion [6], can be derived as follows

$$w_{i0i}(n+1) = w_{i0i}(n) + \eta \left(\hat{y}_i d_i + \frac{1}{w_{i0i}(n)} \right), \quad (12)$$

$$w_{iki}(n+1) = w_{iki}(n) + \eta\, \hat{y}_i d_i(t-k), \quad (13)$$

$$w_{ikj}(n+1) = w_{ikj}(n) + \eta\, \hat{y}_i d_{jj}(t-k), \quad (14)$$

where η is defined as the learning rate and $\hat{y}_i$ is defined as

$$\hat{y}_i = \frac{\partial}{\partial y_i} \frac{\partial y_i}{\partial d_{ii}}$$

where $y_i = g(d_{ii})$, with g denoting an activation function, such as a sigmoid function, as

$$y_i = g(d_{ii}) = \frac{1}{1 + \exp(-d_{ii})}. \quad (15)$$

The Infomax criterion has a side effect; it temporarily whitens the outputs, which should be removed in this application [7]. If W_{11} and W_{22} are forced to be scaling coefficients, instead of using equations (12) and (13), this effect can be avoided by

$$W_{11}(z) = W_{22}(z) = 1. \quad (16)$$

Hence, W_{12} and W_{21} will ideally converge to

$$W_{21} = -H_{21}(z) H_{22}(z)^{-1} \quad (17)$$

$$W_{12} = -H_{12}(z) H_{11}(z)^{-1}. \quad (18)$$

Finally, the outputs of the system can be obtained as in equations (4)-(7), and then conventional adaptive filtering algorithm can be used to cancel acoustic echoes.

After adaptive filtering is performed, two error signals that are transmitted to the far-end are calculated as

$$e_1(n) = e_{11}(n) + e_{21}(n), \quad (19)$$

$$e_2(n) = e_{12}(n) + e_{22}(n). \quad (20)$$

3. COMPUTER SIMULATION

Computer simulations were performed on the basis of the proposed methodology, using the parameters in Table 1.

Table 1: Parameters used in the computer simulation of BSS and adaptive filtering.

Parameters	
Sampling frequency	8kHz
Step size μ in N-LMS	0.01
Learning rate η in BSS	0.001
Filter length for adaptive filters	16 samples
Filter length for BSS	16 samples

Two uncorrelated speech signals discretised at 8 kHz sampling frequency, were utilised as stereo input signals (female voice and male voice). Although using these uncorrelated speech signals is not realistic, it seems worthy to use these as a first step, as such input

signals have not been used previously to perform BSS. It means that there would be possibilities to use these input signals for modifying the separation algorithm, such as cross correlation function, even if these input signals are highly correlated with each other. Four room impulse responses, which will be estimated, were artificially produced as [9]

$$\begin{aligned} h_{11}(k) &= 0.9 + 0.5k^{-1} + 0.3k^{-2}, \\ h_{21}(k) &= -0.7k^{-5} - 0.3k^{-6} - 0.2k^{-7}, \\ h_{12}(k) &= 0.5k^{-5} + 0.3k^{-6} + 0.2k^{-7}, \\ h_{22}(k) &= 0.8 - 0.1k^{-1}. \end{aligned}$$

These room impulse responses characterise a minimum phase response. Note that due to the inverse matrix in equations (17) and (18), precise estimation of the room impulse response with non-minimum phase characteristic is not easy.

Figures 3 and 4 show examples of the BSS simulation. It is remarkable to note that almost perfect separation has been achieved, even at periods when there is no signal on one of the channels.

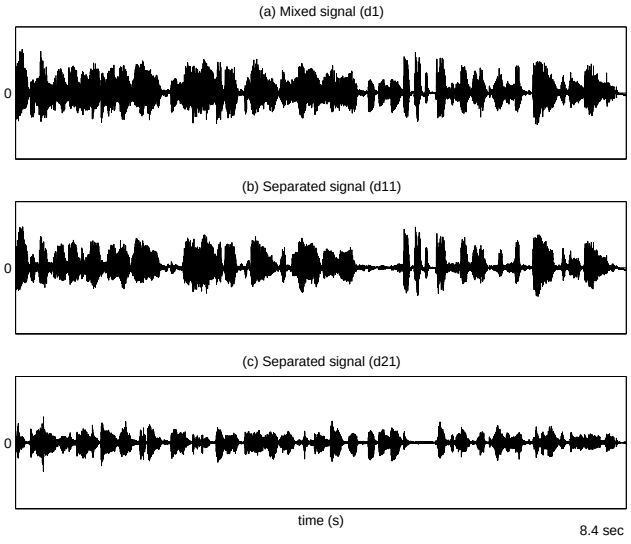

Figure 3: The simulation result of BSS using Infomax criterion, (a): mixed signal d_1, (b): separated signal d_{11} and (c): separated signal d_{21}.

Figure 5 indicates the convergence behaviour of the cross filter w_{21}, using the coefficient error between true and estimated values as

$$10\log_{10}\left(\frac{\sum_{k=0}^{N-1}|g(k)-\hat{g}(k)|^2}{|g(0)-\hat{g}(0)|^2}\right) \qquad (21)$$

where $g(k)$ is true value of the coefficients, $\hat{g}(k)$ is the estimate of the filter coefficient. The convergence is quite fast because the length of the room impulse response is very short. However, a steady-state value of

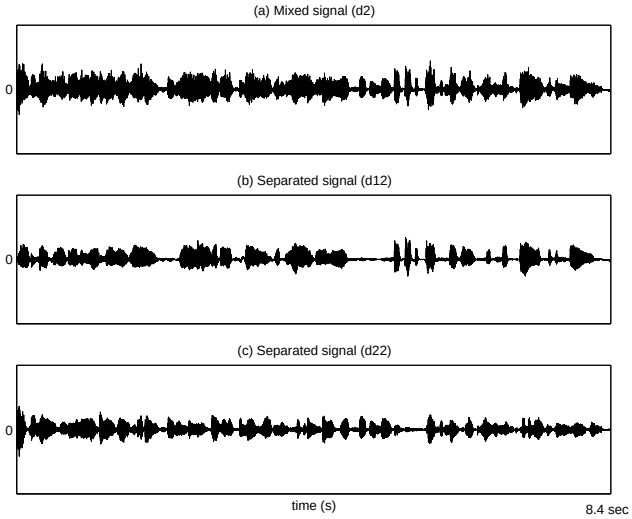

Figure 4: The simulation result of BSS using Infomax criterion, (a): mixed signal d_2, (b): separated signal d_{12} and (c): separated signal d_{22}.

about -15 dB would be evaluated as a good result.

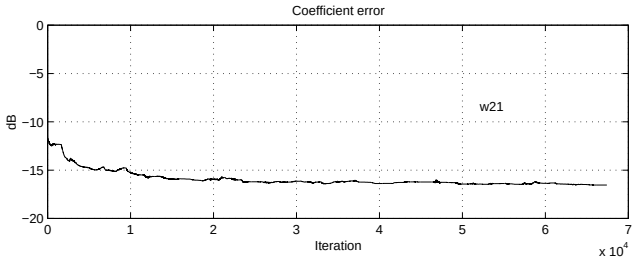

Figure 5: Coefficient error between real and estimated values for w_{21}.

After applying BSS, the output signals were calculated using converged cross filters w_{21}, w_{12} and forced filters w_{11}, w_{22} to evaluate equations (4)-(7). Then, normalised LMS algorithm was used to identify individually each room impulse response, $\boldsymbol{h}_{11}$, $\boldsymbol{h}_{21}$, $\boldsymbol{h}_{12}$ and $\boldsymbol{h}_{22}$ [10]. Figure 6 shows an example of the acoustic echo cancellation with normalised LMS algorithm for d_{11} only. The coefficient error converged to around -15 dB only because there was a separation error as in equation (4). Although this separation error might be recognised as double-talk signal, it must be surpressed. Unless a perfect solution by BSS is obtained, this separation error can not be neglected.

Figure 7 shows two error signals e_{11}, e_{21} and the final error e_1 in Figure 1, which is sent to the far-end after adaptive filtering.

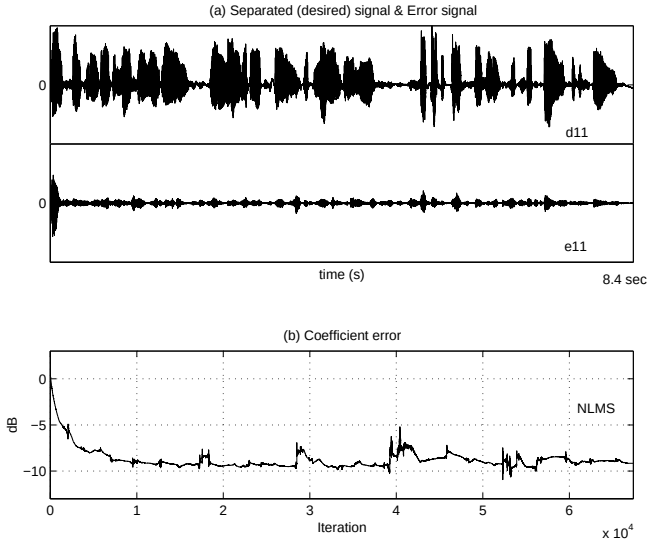

Figure 6: Normalised LMS adaptive filtering to cancel out the acoustic echo d_{11} after applying BSS, (a): separated signal d_{11} and error signal e_{11}, (b): coefficient error.

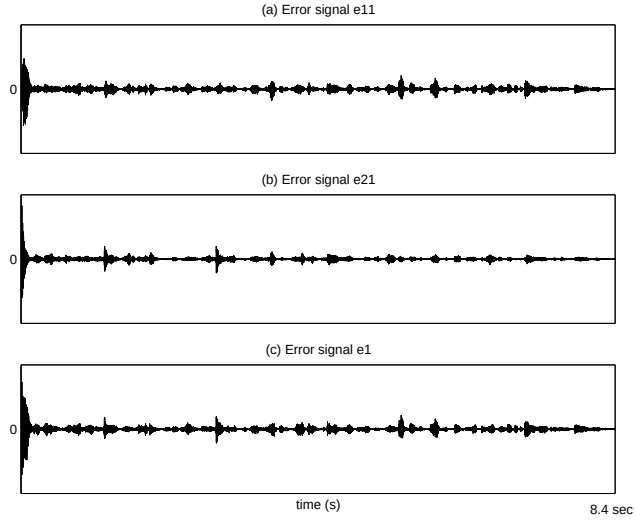

Figure 7: Two error signals (a): e_{11}, (b): e_{21} and (c): the final output signal e_1.

4. CONCLUSIONS

SAEC using BSS has been investigated from the view of post-processing. It has been demonstrated with simulated exercises that the proposed method has a great deal of potential with BSS for SAEC. It has been shown that BSS using the Infomax criterion is a powerful method for separating mixed signals. The next step in this research would be to consider room impulse responses with non-minimum phase characteristics and highly correlated input signals. Moreover, this research will investigate the convergence speeds of BSS and adaptive filters and the effect of BSS on the adaptive filters.

5. REFERENCES

[1] S. L. Gay and J. Benesty, Ed.: *Acoustic Signal Processing for Telecommunication*. Kluwer Academic Publishers, Boston 2000.

[2] M. M. Sondhi, D. R. Morgan, and J. L. Hall: *Stereophonic acoustic echo cancellation - an overview of the fundamental problem*. IEEE Signal Processing Letter, **2**, pp. 148-151. 1995.

[3] S. Shimauchi, Y. Haneda, S. Makino, and Y. Kaneda: *New configuration for a stereo echo canceller with nonlinear pre-processing*. IEEE Proceedings of ICASSP 98, pp. 3685-3688. 1998.

[4] J. Benesty, D. R. Morgan, and M. M. Sondhi: *A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation*. IEEE Transactions on Speech Audio Processing, **6**, pp. 156-165. 1998.

[5] A. Gilloire, and V. Turbin: *Using auditory properties to improve the behaviour of stereophonic acoustic echo cancellers*. IEEE Proceedings of ICASSP 98, pp. 3681-3684. 1998.

[6] K. Torkkola: *Blind separation of convolved sources based information maximization*. IEEE Workshop on Neural Networks for Signal Processing, pp. 423-432. 1996.

[7] S. Haykin, Ed.: *Unsupervised adaptive filtering, Vol. I*. Chapter 8, pp. 321-375, John Wiley & Sons Inc. 2000.

[8] A. J. Bell and T. J. Sejnowski: *An information-maximisation approach to blind separation and blind deconvolution*. Neural Computation, **7**(6), pp. 1129-1159. 1995.

[9] T. W. Lee: *Independent Component Analysis*. Kluwer Academic Publishers, Boston 1998.

[10] S. Haykin: *Adaptive filter theory*. Prentice-Hall, NJ 1996.