ICA for Acoustic Echo Control

M. Magadán, R. Niemistö, U. Ruotsalainen, J. Möttönen
Institute of Signal Processing
Tampere University of Technology
Finland, FIN 33101.
e-mail: {margarita.magadan, riitta.niemisto} @ tut.fi

ABSTRACT

In this paper we propose a new paradigm for acoustic echo control and find a powerful echo cancellation method. We apply independent component analysis for separation of near-end signal from the echo. During double talk the near-end and far-end signals are uncorrelated and the underlying echo is a weighted sum of far-end signals with different delays (echo components). The simulation results obtained are giving a preliminary idea of the results where the near-end signal is separated and, moreover, it is not distorted (no disturbing artefacts appear). The level of echo attenuation is very high, 30–50 dB.

1 Introduction

Acoustic echo cancellation has received growing interest in recent years since many governments have forbidden or will forbid the use of cellular phone while driving a car unless a hands-free set is used. The problem is the most difficult to solve during the small periods of double talk. Then the adaptive algorithms that have been used traditionally tend to get confused and they produce annoying artefacts; therefore, adaptation of their coefficients is usually slowed down or completely halted ([1]).

In this paper we will consider a very different approach from the usage of adaptive filtering algorithms, namely independent component analysis (ICA) for the problem of acoustic echo cancellation during double talk. ICA has recently received growing attention ([2]) due to its potential of finding underlying factors or components from multidimensional statistical data. It offers the possibility to solve a large number of diverse problems based on the idea that: 1) Independent sources generate statistically independent signals and 2) if at most one of the sources is Gaussian and there exists at least an equal number of independent mixtures of the sources then it is possible to separate the sources.

In ICA we assume that we have n linear mixtures $x_1, \ldots, x_n$ of n independent sources $s_1, s_2, \ldots, s_n$

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \ldots + a_{jn}s_n, \text{ for all } j.$$

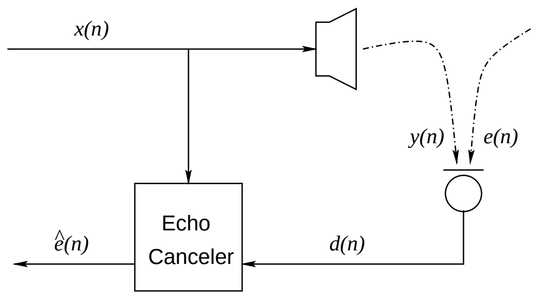

Figure 1: General configuration of an acoustic echo canceler. Signals: $x(n)$ – far-end, $y(n)$ – echo, $e(n)$ – near-end, $d(n)$ – microphone and $\hat{e}(n)$ – echo canceler output.

In the original ICA [2] model it is assumed that each mixture x_j and each independent component s_k is a random variable. It is convenient to use vector–matrix notation instead of the sums like in the previous equation. Let us denote $\mathbf{x} = (x_1, ..., x_n)^T$, $\mathbf{s} = (s_1, ..., s_n)^T$, and let

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

be the mixing matrix. Using the vector-matrix notation, the above mixing model is written as

$$\mathbf{x} = \mathbf{A}\mathbf{s}.$$

Now ICA of the random vector $\mathbf{x}$ consists of an iterative search for a mixing matrix $\mathbf{A}$ and a separating matrix $\mathbf{B}$ that minimize the statistical dependence between the components of

$$\hat{\mathbf{s}} = \mathbf{B}\mathbf{x} = \mathbf{B}\mathbf{A}\mathbf{s}.$$

Remark that even under ideal conditions there is no way for determining either the order of the separated sources $\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_n$ or their magnitude.

General configuration of an echo canceler is shown in Figure 1. Here the echo signal $y(n)$ and the near-end signal $e(n)$ are the independent sources to be separated from the microphone signal $d(n)$. We do not have another mixture available; therefore, we cannot use ICA directly as it has been described.

The organization of the paper is as follows. In the next section we propose a method for applying ICA for acoustic echo cancellation during double talk. In section 3 we present experimental results. Some conclusive remarks are found in the last section.

2 ICA posing for echo cancellation

We will now represent the problem of echo cancellation in terms of ICA. We assume linear echo, where the echo

$$y(n) = \sum_{m=0}^{M-1} w_m x(n-m) \tag{1}$$

is a weighted sum of echo components $x(n-m)$, $m = 0, 1, \ldots, M-1$. The echo components are delayed versions of the far-end signal and the microphone signal $d(n)$ is a sum of near-end signal $e(n)$ and echo $y(n)$ (cf. Figure 1). Now we consider the problem of echo cancellation as an ICA problem with the sources

$$\left\{ \begin{array}{rcl} s_1 & = & e(n) \\ s_2 & = & x(n) \\ & \vdots & \\ s_{M+1} & = & x(n-M+1) \end{array} \right. \tag{2}$$

and the mixtures

$$\left\{ \begin{array}{rcl} x_1 & = & d(n) \\ x_2 & = & x(n) \\ & \vdots & \\ x_{M+1} & = & x(n-M+1). \end{array} \right. \tag{3}$$

The underlying mixing matrix $\mathbf{A}$ is sparse with ones on the main diagonal, the weights w_m ($m = 0, 1, \ldots, M-1$) on the first row (elements $2, 3, \ldots, M+1$) and zeros elsewhere. The echo components are correlated at each time instant because the echo components are, as speech, heavily autocorrelated and they are delayed versions of each other. Therefore, the problem is not directly applicable for ICA. However, our task is only to separate the near-end signal from the echo components and the near-end signal is independent of any echo component.

In acoustic echo cancellation the length of the acoustic echo path in a car cabin is typically 32 ms. Therefore, the number of parameters in the linear model (1) is 256 samples (sampling frequency being 8 kHz). In other words, there are 257 independent sources to be separated in real time. Therefore, the problem must be simplified. Firstly, we only need to separate one signal, the near-end; secondly, knowing that the near-end appears in the microphone signal, but not in the echo components, we can give a good initial value to ICA.

The ICA method [2] is applied to any input $\mathbf{x}$ such that first a candidate $\mathbf{a}_0$ of the first column of the mixing matrix is chosen either randomly or using an initial guess. Then the first column $\mathbf{a}$ of the mixing matrix and the first row $\mathbf{b}$ of the separating matrix are found iteratively, using whitening and dewhitening matrices of the input and a nonlinear transformation. From (2) and (3) we know that the first column $\mathbf{a}$ of the mixing matrix must be

$$\mathbf{a} = (1, 0, 0, \ldots, 0)^T.$$

Therefore, the first row of the $\mathbf{b}$ of the separation matrix can be computed in one step. The resulting algorithm is summarized in Table 1. The parameters of the linear model (1) are a side product of the algorithm as the first row of the separation matrix must be of the form

$$\mathbf{b} = (1, -\hat{w}_0, -\hat{w}_1, \ldots, -\hat{w}_{M-1}).$$

Acoustic echo cancelation is an application for system identification. The method proposed does not differ too much from the methods that are derived using the Wiener or the least squares approach to the same problem [4]. However, in Wiener filtering it is assumed that the underlying signals consist of white Gaussian noise and near-end signal is considered as an estimation error. Conversely, in ICA our task is to separate the near-end, not to minimize its energy. Also, in the ICA

Table 1: ICA for echo cancellation

1. Compose the input matrix (3)

   $$\mathbf{X} = (\mathbf{d}(n), \mathbf{x}(n), \mathbf{x}(n-1), \ldots, \mathbf{x}(n-M+1))$$

   where

   $$\mathbf{x}(n) = (x(n), x(n-1), \ldots, x(n-N+1))^T,$$

   $$\mathbf{d}(n) = (d(n), x(d-1), \ldots, x(d-N+1))^T$$

   and N is the length of the segment of input being considered.

2. Find the covariance matrix $\mathbf{R} = \mathbf{X}^T\mathbf{X}$ of the input and its eigenvalue–eigenvector decomposition ($\mathbf{R} = \mathbf{EDE}^T$).

3. Build the whitening matrix $\mathbf{R}_w = \mathbf{D}^{-1/2}\mathbf{E}^T$ and the dewhitening matrix $\mathbf{R}_{dw}\mathbf{D}^{1/2}\mathbf{E}$.

4. Build the mixing vector $\mathbf{a} = (1, 0, \ldots, 0)^T$.

5. Solve the separation vector $\mathbf{b}'$ from

   $$\begin{array}{rcl} \mathbf{R}_{dw}\mathbf{z} & = & \mathbf{a} \\ \mathbf{b}' & = & \mathbf{z}^T\mathbf{R}_w \end{array}$$

   and normalize with $\mathbf{b} = \mathbf{b}'/b'_0$, (the scalar b'_0 being the first element of $\mathbf{b}'$).

6. Separate near-end from echo computing $\hat{\mathbf{e}} = \mathbf{b}\mathbf{X}^T$ (where $\hat{\mathbf{e}} = (\hat{e}(0), \hat{e}(1), \ldots, \hat{e}(N))^T$).

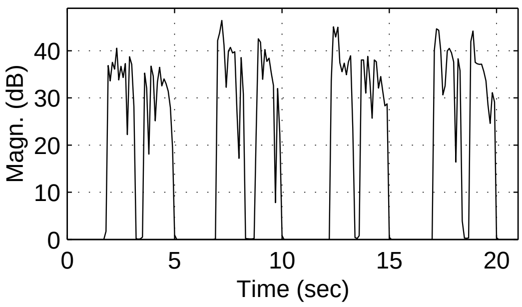

Figure 2: Far-end signal

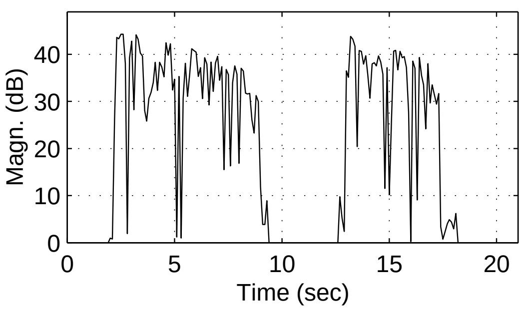

Figure 3: Near-end signal

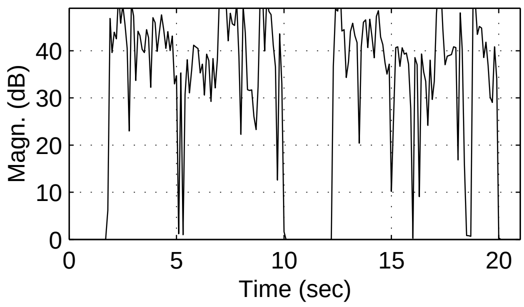

Figure 4: Microphone signal ($M = 50$)

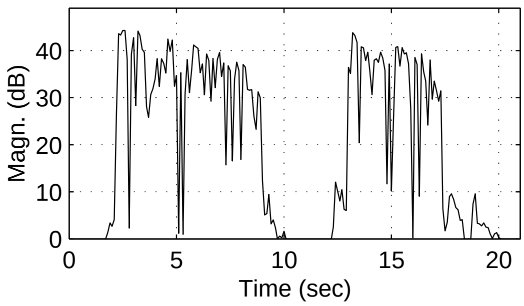

Figure 5: Estimate of near-end ($M = 50$)

approach we do not assume that the input is Gaussian, which gives us more flexibility.

3 Simulations

The performance of the echo cancelers have been traditionally measured in terms of echo return loss enhancement

$$\mathrm{ERLE} = 10 \log_{10} \frac{\mathrm{E}(y^2(n))}{\mathrm{E}(e_r^2(n))} \tag{4}$$

where $e_r(n) = \hat{e}(n) - e(n)$ is the residual echo. Since the echo $y(n)$ and the near-end $e(n)$ are generally not available, usually ERLE can only be estimated using the microphone signal $d(n)$ and echo canceler output $\hat{e}(n)$. ERLE is usually computed in time windows of $20 - 100$ ms and the estimate is reliable during single talk. However, during double talk the cancelers performance can only be measured if the near-end signal is available. For hands-free telephones ERLE should be 25 dB during double talk, according to ITU-T Recommendation G.167.

We simulated the performance of the ICA method using real speech signals in an artificially built environment. The length of the speech signals was 21 seconds and the sampling frequency was 8 kHz. The near-end signal is a set of two female spoken sentences *'Tampereen uusmediakeskuksessa työstetään tulevaisuutta eli kymmeniä teknologian alan tutkimus- ja markkinointihankkeita.'* and *'Uuteen ajatteluun kannustetaan muun muassa multimediamessuin ja -kilpailuin.'* The far-end consists of the following two pairs of short sequences *'Hän ei pelkää mitään. Lapsi opettelee puhumaan.'* and *'Pöydällä on sanomalehtiä. Elokuva oli jännittävä.'* that are first spoken by a male speaker and then by a female speaker. Envelopes of the far-end and near-end signals are shown in Figures 2 and 3, respectively. We cancelled the echo separately during the four periods of far-end talk and the ICA method was also applied during far-end single talk.

We considered four echo paths of different length, 20, 30, 40 and 50 samples (or 2.5 ms, 3.75 ms, 5 ms and 6.25 respectively). Even the longest echo path, 50 samples, is not a realistic length, but experiments using longer test signals required too much time and memory. The echo path was generated using the linear model (1) where the parameters $w_0, w_1, \ldots, w_{M-1}$ correspond to the M most significant successive parameters of a true echo path obtained using the well known normalized LMS algorithm (cf. e.g. [1]) and real measurement signals recorded in a car cabin. Envelope of the microphone signal in the case $M = 50$ is shown in Figure 4 and envelope of the estimate of the near-end resulted in the case shown in Figure 5.

The simulation results were also listened and it was found out that the level of the resulting echo is pretty low; moreover, there appeared no disturbing artefacts to either near-end signal or to the residual echo. The results were also measured in terms of ERLE (4). Since we have the near-end signal available it is possible to subtract the near-end signal from echo canceler output and microphone signal and measure ERLE (4) from

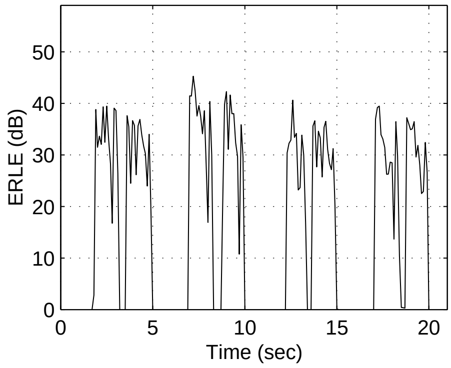

Figure 6: Estimate of ERLE ($M = 20$)

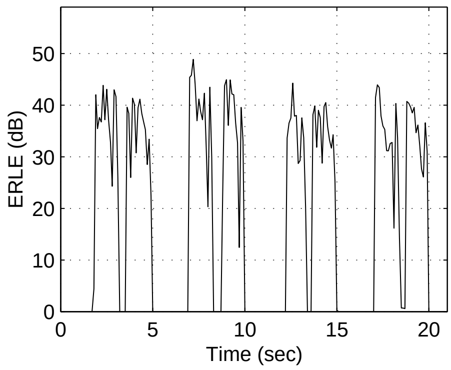

Figure 7: Estimate of ERLE ($M = 30$)

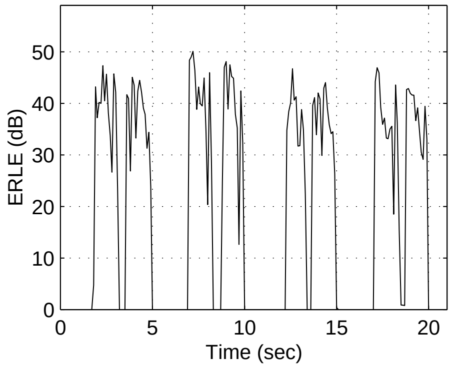

Figure 8: Estimate of ERLE ($M = 40$)

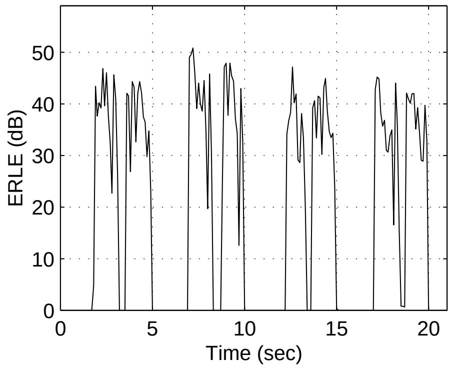

Figure 9: Estimate of ERLE ($M = 50$)

there, as shown in Figures 6–9, and we can see that the ERLE is 30–50 dB also during double talk. Furthermore, the level of the echo signal was different when different lengths of echo path were considered but the level of the residual echo was always the same, approximately 35–40 dB below the level of the near-end.

We must keep in mind that the results were obtained using clean speech signals and a rather short artificial echo path. However, as shown in the figures, the level of ERLE does not drop as the length of the echo path increases. This suggests that the ICA method should also be applicable in real environment.

4 Conclusions and future work

We have proposed and simulated a new method for acoustic echo control that is based on the theory of independent component analysis. The near-end signal is separated from the echo components; furthermore, it is not distorted such that its subjectively measured quality suffers.

This work was done to show that ICA can be used as an effective echo controlling method. The method requires computation of eigenvalue-eigenvector decomposition of the covariance matrix of the inputs. Its implementation into a real echo cancellation tool needs more work to be done. A future work should include more testing with realistic echo path lengths and noisy signals. Also recursive algorithms should be studied for developing the method to be functioning in real time during double talk.

References

[1] Steven L. Gay and Jacob Benesty (editors) *Acoustic Signal Processing for Telecommunications* Kluwer Academic Publisher, Boston/Dordrecht/London, 2000

[2] Aapo Hyvärinen , Juha Karhunen and Erkki Oja, *Independent Component Analysis*, Wiley-Interscience Publication, John Wiley & Sons, New York, 2001

[3] Jean Francois Cardoso, *The invariant approach to source separation*, Proc. NOLTA, pp. 55–60, 1995

[4] Simon Haykin, *Adaptive Filter Theory 3rd edition*, Prentice Hall, Englewood Cliffs, New Jersey, 1996.