Adaptive crosstalk-resistant noise-cancellation using Kalman filters.

T.J. Moir — School of Engineering, AUT University, Auckland, New Zealand
Tom.Moir@aut.ac.nz

Introduction

The problem of suppression of a non-stationary noise from a non-stationary signal has been studied for decades with varying degrees of success. For example, the simplest method of minimizing mean-square error for stationary signals and noise was solved by Wiener[1] (continuous time) and Kolmogorov[2] (discrete-time). The non-stationary case was first solved by Kalman[3]. These approaches all required an accurate model of the signal and noise generating processes. This can be quite appropriate for problems in say rocket research, but in acoustic problems where the signal and noise are both non-stationary speech, it is quite impractical. Indeed, other approaches were taken much later using the least-mean-squares (LMS) algorithm as an estimator of the statistical models of the signals and noise. This work had its foundations with the work of Widrow et al[4]. The idea was to use more than one sensor. One sensor could measure the noise in an area which was signal-free. This “noise-alone” measurement could then be phase and amplitude shifted using an adaptive LMS algorithm (as part of an adaptive filter) and subsequently an estimate of the noise obtained which could be subtracted from the signal plus noise mixture. The phase and amplitude shifting is done with respect to signal plus noise so as to minimize mean-square error. Such an approach requires a degree of isolation so that the noise measurement is free from signal, otherwise both signal and noise will be cancelled or at least the performance will be severely impaired. In the case of stationary signal and noise, the converged filter becomes the optimal Wiener filter as approximated by a finite-impulse response (FIR) adaptive filter. However, making one sensor signal-free necessitates moving the sensors far apart, which in turn leads to large filter orders.

As an alternative, a few authors use a similar approach which puts both sensors close together rather than far apart[5,6]. By switching two LMS algorithm in turn, one during signal and noise and one during noise-alone, the signal can again be estimated as minimum mean-square error. Although this idea has elegance, it requires an extra algorithm to detect when signal is present without the signal. When dealing with speech signals such an algorithm has become known as a voice-activity detector (VAD). The estimate of the noise on its own via the VAD determines the performance of the overall adaptive filter. If fact it is a hard problem to discriminate what is noise and what is crosstalk signal when both are non-stationary, as with the case of two speech signals. A good approach which is immune to absolute threshold values of energy and instead works on the geometry of the positioning of desired signal is given in[7]. Generally speaking, even the best VAD is not good enough to distinguish one signal from another and any errors in distinguishing the noise (or undesired speech) will result in cancellation of the desired speech, as is the case with the original Widrow approach. A method therefore had to be found which was immune to crosstalk and eliminate VAD’s altogether.

Such an approach was used in[8,9], and involved a cross-coupling of the error signals in two parallel running LMS identification algorithms

This method was much later extended to microphone arrays[10]. The technique does not guarantee convergence to the optimal values, but is nevertheless a successful practical method in that the microphones (or sensors) can be placed closely together and no VAD is required. Although other approaches exist (for example Independent Component Analysis (ICA)[11]) they are in general far more computationally demanding. The newer methods have become known as “Blind-Sourced separation (BSS)” since no a priori knowledge is assumed about the noise or signal statistics other than that they are statistically independent.

Methodology

The method below is known in the literature as backwards separation[8,13,14]. This method is more of a deconvolution method than a direct signal estimation method. However, there remains the problem of estimating the unknown coupling transfer functions.

We can write two cross-coupled Kalman filters that can perform blind-source separation of two random signals. We use the model of Figure 1 as the mixing and un-mixing process.

\[ P_{11} = P_{12} + \frac{P_{11} \left( P_{11} + P_{22} \right)}{2} \]

\[ P_{22} = P_{12} + \frac{P_{22} \left( P_{11} + P_{22} \right)}{2} \]

Finally, the equal gain mixers are updated according to

\[ K_1 = \frac{P_{11}}{P_{11} + P_{22}} \]

\[ K_2 = \frac{P_{22}}{P_{11} + P_{22}} \]

Example: Consider the mixture of helicopter noise and pure speech signal through an AR-3 vector process given by:

\[ y_t = c_1 x_1 + c_2 x_2 + \epsilon_t \]

\[ x_1 = x_2 + \epsilon_{12} \]

The n-square positive definite error-covariance matrices are found iteratively from

\[ P_{11} = P_{12} + \frac{P_{11} \left( P_{11} + P_{22} \right)}{2} \]

\[ P_{22} = P_{12} + \frac{P_{22} \left( P_{11} + P_{22} \right)}{2} \]

Table 1. Comparison of NRR’s for various methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>LMS</th>
<th>RLS</th>
<th>Kalman</th>
<th>Not Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRR (dB)</td>
<td>4.78</td>
<td>-158</td>
<td>115.8</td>
<td>115.8</td>
</tr>
</tbody>
</table>

Conclusion

An extension to the cross-coupled LMS approach to adaptive filtering has been shown. The new method which uses cross-coupled Kalman filters, has a similar performance to the Natural Gradient method and outperforms LMS and RLS.

References


Fig 1. Signal Generation by mixing (left) and the separation process (right)