Noise cancellation with adaptive filtering

Alexandre Trilla, Xavier Sevillano

Departament de Tecnologies Mèdia
LA SALLE – UNIVERSITAT RAMON LLULL
Quatre Camins 2, 08022 Barcelona (Spain)
atrilla@salle.url.edu, xavis@salle.url.edu

2011

Abstract

This tutorial presents a practical application of adaptive filtering to cancel noise.

1 Introduction

Noise cancellation is a classical example of the practical application of adaptive filtering. It applies signal processing techniques to remove (or at least minimise) the disturbing effect of a noise polluting a signal of interest.

In order to apply the noise cancellation procedure it is required to have some knowledge about the relation among the signals involved. In this work, an isolated measure of the interfering noise is available, taken in a place different from where the signal of interest (plus modified interfering noise) is taken, see Figure 1.

As it can be observed in the figure, two signals are involved in the process: $x[n]$ represents the signal of interest in addition to a modified version of the interfering noise, and the pure interfering noise itself $s[n]$. The interfering noise variation ($s'(t)$ wrt $s(t)$) is caused by the spatial difference between the places where the signals are taken, as it is proposed in this scenario. The system S represents the acoustic transfer function between these two measure points.

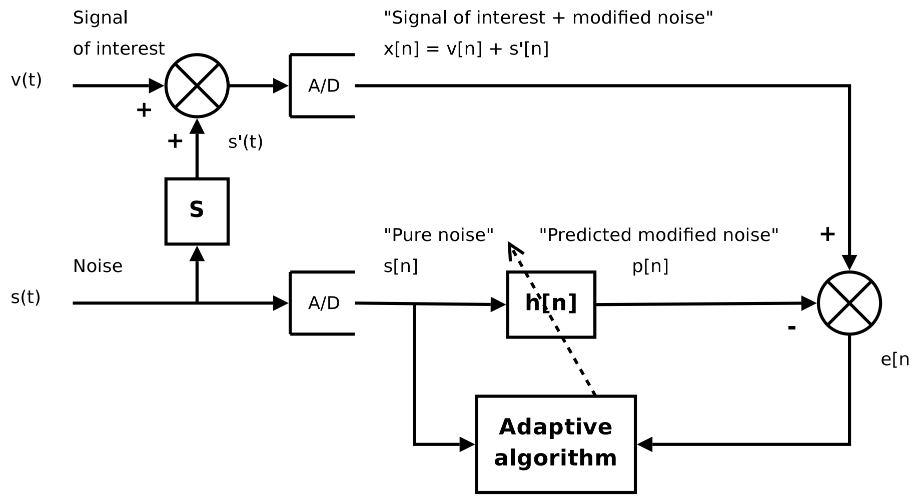

Figure 1: Noise cancellation diagram.

2 Adaptive filter

In order to perform the successful removal of interfering noise from the "polluted" signal of interest $x[n]$, the filter $h[n]$ needs to be *dynamically adapted* in order to filter out the noise. $h[n]$ may well be an N-point Finite Impulse Response (FIR) filter with real values, given the real nature of audio signals.

Overall, the adaptive filter $h[n]$ identifies the system S. Therefore, by estimating the modified interfering noise $p[n] \sim s'[n]$, the error signal $e[n] = v[n]+s'[n]-p[n]$ is minimised, leaving $v[n]$, i.e., the signal of interest without noise interference.

3 Objective function

The correct statement of an adaptive procedure first defines an objective function, aka cost function or error function, to optimise under a certain criterion. Such cost function J_e is determined to be the (e.g.) energy of the *error* signal $e[n] = x[n] - p[n]$ (over a span of M points), see Eq.(1).

$$J_e = \mathscr{E}_e = \|e[n]\|_2^2 = \sum_{n=0}^{M-1} \mid e[n] \mid^2 = \sum_{n=0}^{M-1} \mid x[n] - p[n] \mid^2 \tag{1}$$

Therefore, the optimisation criterion is the minimisation of the squared error defined by such objective function J_e. Since the square function J_e is a well-defined convex function, some extremum is determined to exist.

4 Gradient descent procedure

The gradient descent (GD) is an iterative procedure to minimise the criterion function J_e, thus defining a minimum squared-error procedure. Overall, it is an adaptive method to implement Wiener's filter (i.e., the optimum filter), which determines $h[n]$ to be a solution vector that minimises J_e. Note that J_e is a function of $h[n]$, given that the predicted noise is $p[n] = s[n] * h[n]$. For practical purposes, $p[n]$ is implemented as the scalar product $\mathbf{h}^T\mathbf{s}$.

The gist of this procedure is that the adaptation is performed in the direction of the *inverse gradient* of the objective function as the procedure iterates over the data, see Eq.(2).

$$\mathbf{h} \leftarrow \mathbf{h} - \eta[n] \cdot \nabla J_e \tag{2}$$

Note that Eq.(2) introduces the $\eta[n]$ parameter, called the *learning rate*, which states how fast the algorithm converges to the optimum solution. If $\eta[n]$ is too small, convergence is needlessly slow, whereas if it is too large, the adaptation process will overshoot and may diverge. Refer to [Duda et al., 2000] for further details.

4.1 Widrow-Hoff or Least Mean Square (LMS) rule

This is a particular GD procedure that reduces the storage requirements of the iterative method by considering the samples *sequentially*. This implies that the analysis window of the error is reduced from M samples to 1 sample. Therefore, it is a sample-by-sample approximation of the error estimate. Its online parameter update rule is given in Eq.(3).

$$\mathbf{h} \leftarrow \mathbf{h} + \eta[n] \cdot e[n] \cdot \mathbf{s} \tag{3}$$

where $\eta[n]$ absorbs the 2 constant of the gradient.

Considering one sample at a time, the adaptation never ceases: the filter parameters orbit around the optimum when the noise $s[n]$ remains stationary. If the statistical properties of $s[n]$ changed with time, the adaptive filter $h[n]$ would be able to evolve and readapt to the new scenario.

5 Training inhibition

The presented noise cancellation technique displays a problem when the signal of interest $v[n]$ is present, because then the error should maintain such signal ($e[n] \sim v[n]$) instead of readapting the coefficients trying to force its minimum with a signal that is totally unrelated, i.e., the pure noise reference $s[n]$. Therefore, a method to inhibit the adaptation of the coefficients is needed.

One way to determine when to inhibit the adaptation could be controlling that the energy of $x[n]$ remains lower that a fraction of the energy of $s[n]$. Therefore, if the signal of interest $v[n]$ is predominant in $x[n]$, then the filter should not continue readapting.

Another way to proceed could be applying a variable learning rate (note that $\eta[n]$ allows the evolution with time) as is shown in Eq.(4).

$$\eta[n] = \frac{\eta[0]}{\mathscr{E}_s + \mathscr{E}_e} \tag{4}$$

where $\mathscr{E}_s$ and $\mathscr{E}_e$ are online estimations of the energy of the pure noise and error signals, respectively.

Hence, when the signal of interest $v[n]$ is present, the energy of the error signal $\mathscr{E}_e$ is bound to grow, thus shrinking the learning rate $\eta[n]$, causing the adaptation temporally to cease.

References

[Duda et al., 2000] Duda, R. O., Hart, P. E., and Stork, D. G. (2000). *Pattern Classification*. Wiley-Interscience, New York, NY, USA.