TR－H－084

A New discriminant generalised cross correlator

David Rainton

1994. 7. 11

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2 ☎07749-5-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1011
Facsimile: +81-7749-5-1008

A new discriminant generalised cross correlator

D. Rainton
ATR Human Information Processing Laboratories
2-2 Hikaridai
Seika-cho Soraku-gun
Kyoto 619-02
Japan
E-mail rainton@hip.atr.co.jp

Abstract

A new discriminant generalised cross correlator(DGCC) for passive time delay estimation, and noise reduction, is presented. Its interpretation is that of a generalised cross correlator(GCC) where the filter coefficients are adjusted in a supervised fashion, to minimise a suitably chosen delay estimation error function.

Contents

1 Introduction

The most common method of determining the time delay between two inputs $x_1(t)$ and $x_2(t)$ is to compute the cross correlation function

$$R_{x_1x_2}(\tau) = \frac{1}{T-\tau}\int_{\tau}^{T} x_1(t)x_2(t-\tau)dt. \qquad (1)$$

The argument τ that maximises equation 1 provides an estimate of the delay. In order to improve this estimate it is desirable to pre-filter $x_1(t)$ and $x_2(t)$ prior to cross correlation. For want of a better name, this simple, but very important process is known as *generalised cross correlation* [1]. The generalised cross correlator, implemented as a pre-processor on the input waveforms, is shown in figure 1.

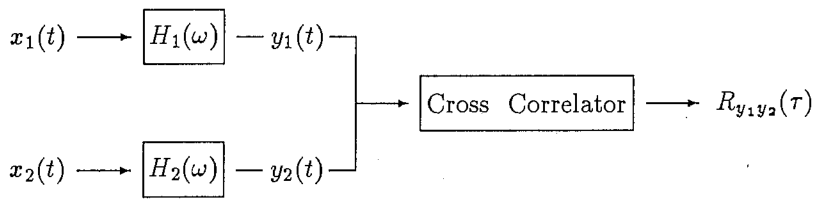


Figure 1: The generalised cross correlator.

It is known that when properly selected, the filters H_1 and H_2 can significantly enhance the estimation of time delay. The filters are designed to accentuate the signal passed to the correlator at those frequencies at which the coherence or signal-to-noise ratio(SNR) is highest. For example, it is well known how to choose the filters to produce a minimum variance time delay estimate(TDE) under Gaussian assumptions. In addition to which a whole family of ad hoc filters have been proposed in the literature.

However, problems with this approach include those of non-Gaussian noise and errors in estimating the signal-to-noise ratio, or coherence, all of which produce sub-optimal TDE performance.

This paper proposes a minimum error approach to selecting the filters H_1 and H_2, i.e. H_1 and H_2 are adjusted to minimise the number of delay estimation errors. Each input pair $x_1(t)x_2(t)$ is classified by the generalised cross correlator as having a delay $\tau_{\text{estimated}}$, where

$$\tau_{\text{estimated}} = \underset{\tau}{\text{argmax}}\, R_{y_1y_2}(\tau). \qquad (2)$$

In classical pattern recognition language $R_{y_1y_2}(\tau)$ is the discriminant function for the input pair $x_1(t)x_2(t)$. Given the input pair

$$
\begin{aligned}
x_1(t) &= n_1(t) + s(t) \\
x_2(t) &= n_2(t) + s(t + \tau_{\text{true}})
\end{aligned} \tag{3}
$$

where $n_1(t)n_2(t)$ are noise, and $s(t)$ is the signal whose delay τ_{true} we are trying to estimate, then an estimation error occurs when $\tau_{\text{estimated}} \neq \tau_{\text{true}}$. A misclassification measure $d_{x_1,x_2}(H_1, H_2)$ is introduced to quantify these errors. The misclassification measure must be positive when the estimated and true delays differ and negative when they are the same. Although there are many possible choices of misclassification measure, the simplest to implement is

$$
\begin{aligned}
d_{x_1,x_2}(H_1, H_2) &= -R_{y_1y_2}(\tau_{\text{true}}) + R_{y_1y_2}(\tau_{\max}) \\
\tau_{\max} &= \underset{\tau \neq \tau_{\text{true}}}{\operatorname{argmax}} R_{y_1y_2}(\tau).
\end{aligned} \tag{4}
$$

To minimise the number of estimation errors the filters H_1 and H_2 are adjusted to minimise $d_{x_1,x_2}(H_1, H_2)$. This adjustment can be achieved via gradient descent, although any suitable optimisation technique (e.g. simulated annealing, etc.) could be used in theory.

Typically the cross correlation is implemented as a running cross correlation function of the general form

$$
R_{y_1y_2}(\tau, t) = \int_{\tau}^{t} y_1(t')y_2(t' - \tau)w(t - t')dt' \tag{5}
$$

where $w(\cdot)$ is a suitably chosen window function. For example, one possible choice for $w(\cdot)$ is the exponential function

$$
w(t) = \begin{cases} e^{-t/T_c} & t \geq 0 \\ 0 & t < 0 \end{cases} \qquad T_c > 0 \tag{6}
$$

where T_c is the window time constant. One simple way of implementing an exponentially decaying window function is

$$
R_{y_1y_2}(\tau, t) = (1 - \alpha)R_{y_1y_2}(\tau, t - 1) + \alpha y_1(t)y_2(t - \tau) \qquad 0 \leq \alpha \leq 1 \tag{7}
$$

where α is a forgetting factor, directly proportional to the inverse of the time constant. The time varying equivalent of the misclassification measure defined in equation 4 is

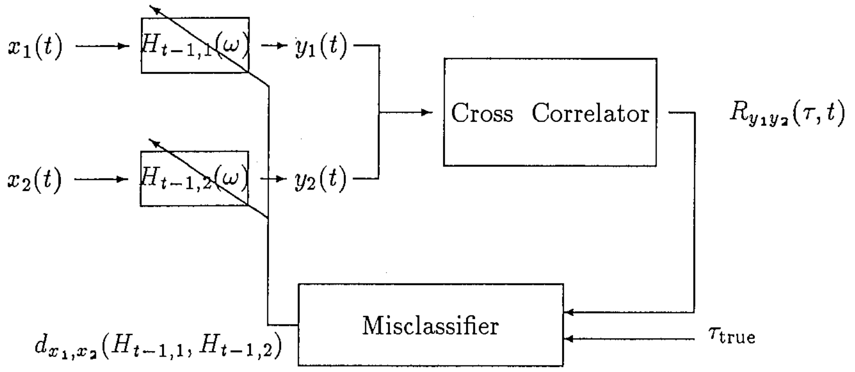
$$
\begin{aligned}
d_{x_1,x_2}(H_{t-1,1}, H_{t-1,2}) &= -R_{y_1y_2}(\tau_{\text{true}}, t) + R_{y_1y_2}(\tau_{\max}, t) \\
\tau_{\max} &= \underset{\tau \neq \tau_{\text{true}}}{\operatorname{argmax}} R_{y_1y_2}(\tau, t).
\end{aligned} \tag{8}
$$

where the filters $H_{t-1,1}$, $H_{t-1,2}$ are updated at each time t using gradient descent, i.e.

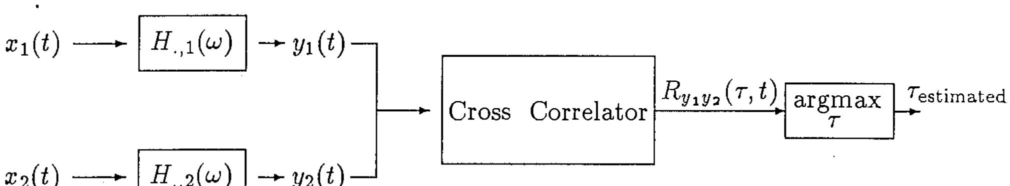
$$H_{t,j} = H_{t-1,j} - \eta \frac{\partial d_{x_1,x_2}(H_{t-1,1}, H_{t-1,2})}{\partial H_{t-1,j}} \qquad j = 1, 2 \tag{9}$$

where η is a suitably chosen learning constant.

The block diagram in figure 2 shows the resulting discriminant generalised cross correlator.



a) Training H_1, H_2 with know delays τ_{true}



b) Unknown delay estimation where H_1 and H_2 are now fixed having been previously trained in part a).

Figure 2: The discriminant generalised cross correlator; a) initial training, b) unknow delay estimation.

The only assumptions made concerning the signal and noise statistics are 1) both are long term stationary over the training and test periods and 2) both arrive from different spatial locations (the algorithm is unable to separate signal and noise arriving from the same spatial location). Unlike the conventional GCC, no error prone coherence, or signal-to-noise ratio estimates are required. Indeed no knowledge of either the signal or noise power spectra are required! The filters are simply adjusted until the cross correlation function peaks at the true delay. The idea, although embarrassingly simple, is deceptively powerful.

2 Implementation Details

Let H_1 and H_2 be causal finite impulse response(FIR) filters(see figure 3), characterised by the time-varying coefficient vectors $\vec{w}_{t,1}$ and $\vec{w}_{t,2}$ respectively.

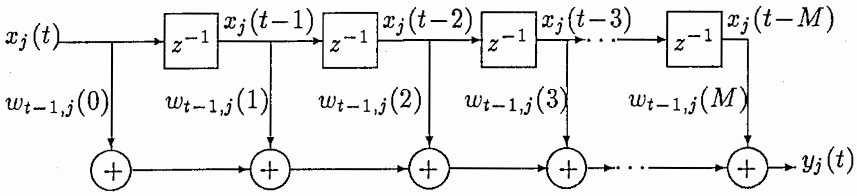


Figure 3: An adaptive FIR filter H_j at time t. The filter has $M+1$ coefficients.

The misclassification measure in equation 8 can be completely expressed in terms of the filter coefficients by the following series of equations

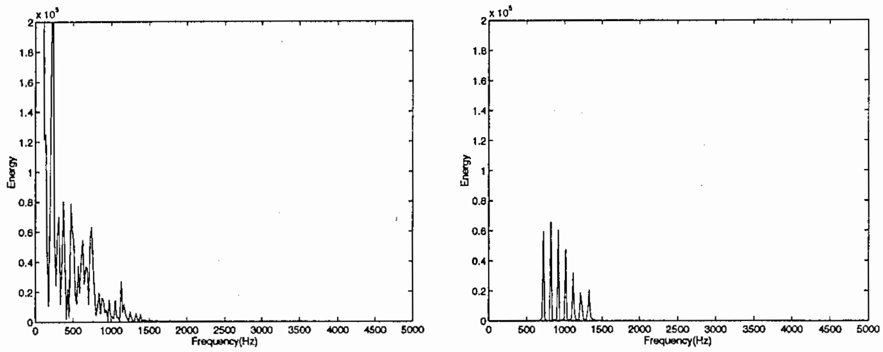
$$\begin{aligned}
d_{x_1,x_2}(\vec{w}_{t-1,1},\vec{w}_{t-1,2}) &= -R_{y_1y_2}(\tau_{\text{true}},t)+R_{y_1y_2}(\tau_{\max},t) \\
\tau_{\max} &= \operatorname*{argmax}_{\tau \neq \tau_{\text{true}}} R_{y_1y_2}(\tau,t) \\
R_{y_1y_2}(\tau,t) &= (1-\alpha)R_{y_1y_2}(\tau,t-1)+\alpha y_1(t)y_2(t-\tau) \qquad 0\le\alpha\le 1 \\
y_j(t) &= \sum_{i=0}^{M} x_j(t-i)w_{t-1,j}(i) \qquad j=1,2 \\
\vec{w}_{t,j} &= [w_{t,j}(0),w_{t,j}(1),\ldots,w_{t,j}(M)] \qquad j=1,2
\end{aligned} \qquad (10)$$

where, at time $t+1$, the $j'th$ FIR filter has $M+1$ coefficients $w_{t,j}(i)$. The filter weights are updated at each new input sample according to the equation

$$w_{t,j}(i)=w_{t-1,j}(i)-\eta\frac{\delta d_{x_1,x_2}(\vec{w}_{t-1,1},\vec{w}_{t-1,2})}{\delta w_{t-1,j}(i)} \qquad j=1,2 \quad 0\le i\le M.$$

3 Typical Application Scenario

For time delay estimation we envisage applying the filter in the following scenario. H_1 and H_2 are suitably initialised, either randomly, or to heuristically determined transfer functions. During the training phase an ensemble of signals are generated from known locations. It is important that both the range of locations and the spectral characteristics of the training ensemble are representative of the testing ensemble. The DGCC is slowly adapted until convergence, or the discriminant function drops below some predetermined threshold. The filters are then fixed.



a) Noise power spectrum   b) Clean signal power spectrum

Figure 4: Power spectra(0-5kHz)

During subsequent testing, sounds with the same overall spectral characteristics as those used during training are generated from the same range of locations, and their positions estimated using the DGCC.

4 Results

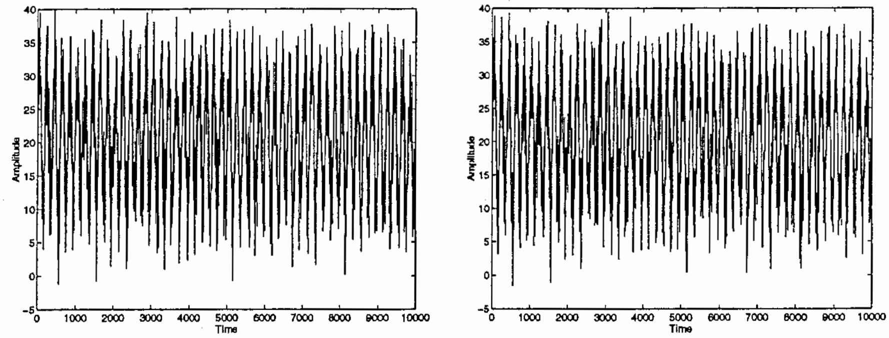
The following preliminary experiments demonstrate the ability of the DGCC to extract the spectral characteristics of a clean signal from a noisy background. Although normally we would expect to train the DGCC with an ensemble of noisy signals generated at various locations, here we use just a single noisy signal from a single location.

The clean signal and noise power spectra are shown in figure 4. The noisy input signals $x_1(n)$ and $x_2(n)$ were defined as

$$\begin{aligned} x_1(n) &= \text{noise}(n) + \text{signal}(n) \\ x_2(n) &= \text{noise}(n) + \text{signal}(n+4). \\ & \quad 0 \leq n \leq 10000 \end{aligned} \tag{11}$$

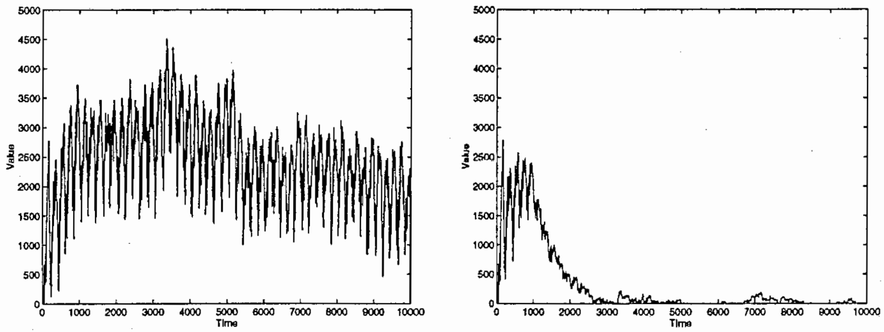
and are shown in figure 5.

The only information provided to the DGCC is that the clean signal has a delay of four (the noise happens to have a delay of zero, although the algorithm was not told that). Both H_1 and H_2 were arbitrarily chosen to be 21 coefficient symmetric FIR filters(although depending on the problem, symmetry is not always a requirement) and both initialised to identity filters. The learning constant, after a small amount of experimentation, was set to $1e^{-5}$, and the running cross correlation attenuation constant α set to 1/500. The range of possible delays was restricted to between -10 and 10.



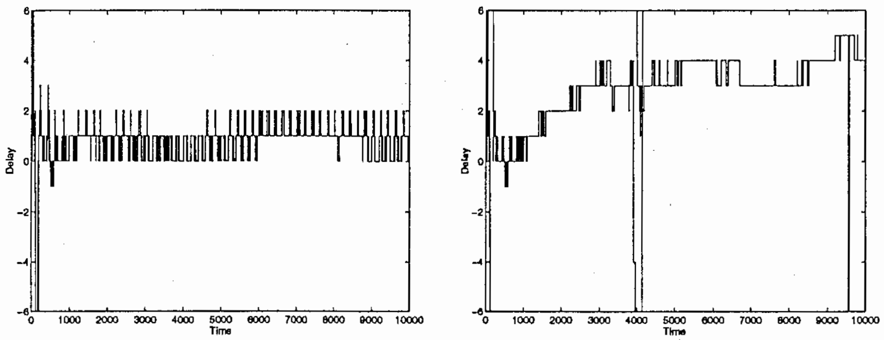
a) $x_1(t)$ b) $x_2(t)$

Figure 5: Noisy input signals



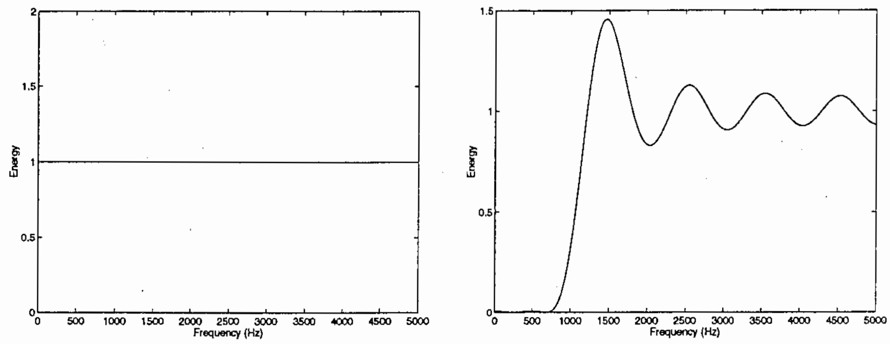
a) No adaptation b) With adaptation

Figure 6: $d(\vec{w}_{t,1}, \vec{w}_{t,2})$ v.s. t



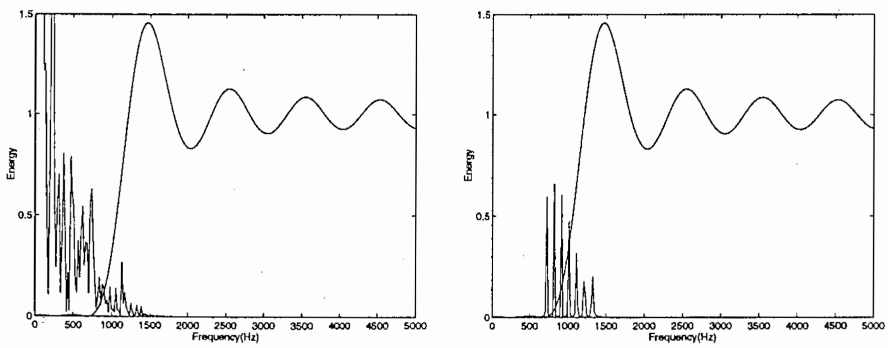
a) No adaptation b) With adaptation

Figure 7: Estimated delay($\tau_{\text{estimated}}$) v.s. t



a) Before adaptation   b) After adaptation

Figure 8: Filter transfer functions ($H_1 = H_2$)



a) Noise   b) Clean signal

Figure 9: Filter transfer functions after adaptation overlayed on the noise and signal power spectra

Two experiments were performed. In the first, no adaptation of H_1 and H_2 was allowed, and the delay estimated over the signal duration. As can be seen in figure 7 a) the estimated delay is that of the noise (i.e. ≈ 0), not the signal. The reason being that the noise energy is significantly larger than the signal energy. The value of the discriminant function with no adaptation of H_1 and H_2 is shown in figure 6 a). The second experiment was identical to the first, except that H_1 and H_2 were now allowed to adapt. The resulting delay estimates, and discriminant function are show in figures 6 b) and 7 b). It can be seen that after about 3000 samples, the filters H_1 and H_2 have successfully adapted to cancel the majority of the noise, and the resulting delay estimates are close to the true signal delay (i.e. ≈ 4). The after adaptation transfer functions for H_1 and H_2 were identical, which was expected since the noise was the same in both x_1 and x_2. Under more general conditions however, where the noise in x_1 and x_2 differs, then H_1 and H_2 would also differ.

The transfer functions before and after adaptation are shown in figures 8 a) and b) respectively. Figures 9 a) and b) show the H_1, H_2 transfer function after adaptation overlayed onto the signal and noise spectra respectively. These plots clearly show that both filters successfully adapted to attenuate the noisy part of the spectrum.

We believe that these and several similarly successful preliminary experiments are sufficient to suggest the merit of this approach.

5 Summary

We have described a new type of adaptive filter, based upon the concept of minimising a suitably defined discriminant function. Potential applications include time delay estimation (described in this paper) and speaker separation (the infamous cocktail party effect), to mention just a few.

References

[1] G. Clifford Carter. Coherence and time delay estimation. *Proc. IEEE*, 75(2):236–255, Feb. 1987.