FREQUENCY-DOMAIN MULTICHANNEL SIGNAL ENHANCEMENT: MINIMUM VARIANCE VS. MINIMUM CORRELATION

Julien Bourgeois and Klaus Linhard

DaimlerChrysler Research and Technology
P.O. Box 23 60 – 89013 Ulm – Germany
{julien.bourgeois,klaus.linhard}@daimlerchrysler.com

ABSTRACT
In this paper we compare two optimization criteria that can be used in multichannel signal enhancement. The first criterion, Minimization of the output Variance (MV), is related to conventional adaptive beamforming whereas the second one, minimization of the output cross-correlation (or Symmetric Adaptive Decorrelation – SAD), is related to blind source separation techniques. We develop an implementation-independent comparison framework, where the quality of the MV adaption control is identified with the power of the target. We account for the correlation of the source of finite length. Our analysis shows that SAD is attractive in cases where target and interferer have about the same power level at each sensor. We evidence the fact that SAD requires longer signals, not only to maintain a low source correlation coefficient, but also because SAD learning rules suffer from larger estimation errors.

1. INTRODUCTION
Conventional multichannel signal enhancement systems are based on the availability of a low Signal-to-Interference Ratio (SIR) noise reference. The noise canceller is adapted to minimize the output power, as illustrated in Fig.2 (MV). The noise reference can be obtained for example by placing a sensor close to the interference source. In the context of a multi-speaker audio input device, a practical scenario arises when each speaker has a close-talk microphone. In this case, cross-talk can be removed using an adaptive noise canceller structure, creating individual speech-input cocoons, as depicted in Fig. 1 (left). We refer to this situation as coocooning. Beside coocooning, this method is typically - but not only [4] - applied in the Noise Canceller block in a GSC beamformer architecture, for example as on Fig.1 (right). In many practical scenarios however, the noise reference is contaminated by more or less significant components of the target signal, which causes target cancellation at the output if no special care is taken. An alternative technique lies on a symmetric architecture that is adapted to provide independent outputs, as illustrated in Fig. 2 (SAD). In this paper we want to compare these two techniques without being dependent on a particular implementation or experimental setting.

Araki et al. [1, 2] found out an equivalence of both algorithms that converge to the same solution if their underlying assumptions are fulfilled [1]. They also reported that SAD suffers from the strong source independence hypothesis [2].

In the next section, we recall formal expressions of the MV and SAD criteria and give the distance between their respective optima and the ideal solution, when the assumptions on which they lie collapse. In section 3 structural differences between the two approaches are discussed and their convergence behavior in a stochastic gradient descent is described.

2. FREQUENCY DOMAIN CRITERIA

2.1 Problem statement
Let us introduce a few notations with the problem statement. For the sake of simplicity, we limit ourselves to the case of two non-stationary zero-mean sources \( s_1(t) \) (target) and \( s_2(t) \) (interferer) of power \( \sigma_1^2(t) \) and \( \sigma_2^2(t) \) received at two microphones \( x_1(t) \) and \( x_2(t) \). Without loss of generality, we absorb the propagation coefficients from \( s_i \) to \( x_i \) into the definition of \( s_i, \) \( i = 1, 2, \) so that the acoustic mixing can be written in the frequency domain as

\[
x_1(t, \omega) = s_1(t, \omega) + h_1(\omega)s_2(t, \omega) + n_1(t, \omega)
\]

\[
x_2(t, \omega) = s_2(t, \omega) + h_1(\omega)s_1(t, \omega) + n_2(t, \omega)
\]
where \( h_i(\omega) \) are the cross-talk (or coupling) coefficients describing the acoustic propagation of \( s_i \) to \( x_j \), and \( n_i(t, \omega) \) denotes the background noise of power \( \sigma^2_\omega \), for \( i, j = 1, 2 \). In the remainder of this paper, we may abbreviate the notation by dropping the time-frequency arguments \( t, \omega \). If \( h_i h_j = 1 \), which corresponds to the case where target and interferer are both located at the same place with respect to the microphone, then it is not possible to recover the sources \( s_i \). Therefore, the quantity \( h_i h_j - 1 \) is called conditioning of the problem.

The first approach – called Minimum Variance (MV) – considers the observations \( x_1 \) and \( x_2 \) as signal and noise references. It consists in searching \( w_2 \) that minimizes the contribution of the noise reference \( x_2 \) in the source estimate \( y_1 \)

\[
y_1(t, \omega) = x_1(t, \omega) + w_2(o)x_2(t, \omega)
\]

To mitigate the leakage problem, it is usual in speech applications to stop the adaptation during periods of speech, i.e. when \( \sigma^2_\omega \) is larger than zero. This requires a Voice Activity Detector (VAD) that estimates the signal power \( \sigma^2_t \). During adaption, the estimated \( \sigma^2_t \) vanishes, such that the variance of the VAD

\[
E[|\sigma^2_t - \sigma^2_\omega|^2] = \sigma^4_t
\]

is the square of the target power.

The second approach – referred to as Symmetric Adaptive Decorrelation (SAD) [6] – is based on decorrelation principles. This approach was investigated in the context of blind source separation [5]. It is based on the assumption that the sources are statistically independent, so that in particular the magnitude of the correlation coefficient

\[
\delta(o) = E[x_1(t, \omega)x_2^*(t, \omega)]
\]

is negligible. This hypothesis may be asymptotically true, but the correlation estimate of finite length signals is non-zero. Note that \( |\delta| < \sigma_t \). SAD consists in searching for \( w_1 \) and \( w_2 \) that minimize a dependence measure of the outputs \( y_1, y_2 \) given by (1) and

\[
y_2(t, \omega) = x_2(t, \omega) + w_1(o)x_1(t, \omega).
\]

Clearly, cross-talk components are cancelled if we converge to the ideal solution \( (w_1, w_2) = -(h_1, h_2) \).

### 2.2 Minimum Variance

MV determines \( w_2 \) as

\[
\arg \min_{w_2} E[|y_1|^2].
\]

The cost function \( E[|y_1|^2] \) is minimized with a gradient descent. The gradient is formally obtained by taking derivatives formally with respect to the conjugate quantity [3] \( w_2^* \), i.e.

\[
\frac{\partial E[|y_1|^2]}{\partial w_2^*} = E[y_1 x_2^*] = E[x_1 x_2^*] + w_2 E[|x_2|^2].
\]

Setting the gradient to zero provides the unique solution \( w_2 = -E[x_1 x_2^*] / E[|x_2|^2] \) that is given in the noise-free case by

\[
w_2 = -\frac{\sigma_1^2 h_1^* + \sigma_2^2 h_2 + \delta + h_1^* h_2 \delta^*}{\sigma_2^2 + |h_1|^2 \sigma_1^2 + h_1^* \delta + h_1 \delta^*}.
\]

We can consider the distance \( d_{sv} \) between this MV-optimum and the solution point \(-h_2\)

\[
d_{sv} = |w_2 + h_2| = \frac{|\delta + \sigma_1^2 h_1^* (1 - h_1 h_2)|}{\sigma_2^2 + |h_1|^2 \sigma_1^2 + h_1^* \delta + h_1 \delta^*}.
\]

For a non-zero conditioning \(|1 - h_1 h_2|\), we can observe that \( d_{sv} = 0 \) if and only if \( \sigma_2^2 = 0 \), which means that a perfect VAD is necessary to cancel the cross-talk.

### 2.3 Symmetric Adaptive Decorrelation

SAD determines \( w_1 \) and \( w_2 \) jointly as

\[
\arg \min_{w_1, w_2} E[|y_1 y_2|^2].
\]

Its gradient is given with

\[
\frac{\partial E[|y_1 y_2|^2]}{\partial w_1^*} = E[y_1 x_2^*] E[y_1 x_1^*] \quad \text{and} \quad \frac{\partial E[|y_1 y_2|^2]}{\partial w_2^*} = E[y_2 x_1^*] E[y_2 x_2^*]
\]

Setting the gradient to zero provides a set of four equations. Only two of these equations provide minima of the cost-function, namely \( E[y_1 y_2] = E[y_1 y_2^*] = 0 \). Clearly, these two equations are equivalent and do not provide enough constrain to determine \( w_1 \) and \( w_2 \), showing that a second-order criterion does not suffice to separate independent sources. \( E[y_1 y_2^*] = 0 \) defines \( w_2 \) as a rational function of \( w_1 \). Specifically, we have in the noise-free case \( \sigma_2^2 = 0 \)

\[
E[y_1 y_2^*] = \sigma_1^2 (w_1^* + w_1)^2 (1 + w_2 h_1) + \sigma_2^2 (w_2 h_2 + w_1 h_1^*) + \delta^2 (1 + w_2 h_1^*) (1 + w_1 h_2^*) + \delta (h_2^* + w_2 h_1^*) (h_1^* + w_1^*).
\]

Thus, the set of all pairs \((w_1, w_2)\) minimizing the cost function defines an hyperbola \( \Omega \) in the parameter space. If the mixing well-conditioned, i.e. \(|1 - h_1 h_2| > 0 \), this hyperbola depends on the statistics of the sources. Non-stationary sources draw time-dependent hyperbolas \( \Omega(t) \) so that the SAD optimum is determined as the intersection \( \Omega \cap \Omega(t) \), hence the separability of non-stationary sources with a second-order criterion [5].

If \( \delta = 0 \), this intersection consists in the ideal solution \(-h_1, h_2\) and at the permuted solution \(-(1/h_2, 1/h_1)\) that corresponds to \((y_1, y_2) = (s_1, s_1) (h_0 h_2 - 1) / h_1\). This is the well-known permutation problem of blind source separation. Therefore, the separation problem becomes closer to each other. It means that the permutation problem will occur more often when sources are close to each other. If \( \delta \neq 0 \), then \( E[y_1 y_2^*] = |\delta| |1 - h_1 h_2| \) at the separation point \(-h_1, h_2\) that does therefore not belong to \( \Omega \). We evaluate the minimal distance \( d_{sv} \) between \(-h_1, h_2\) and the hyperbola

\[
d_{sv} = \min\{|w_2 + h_2| \text{ s.t. } (w_1, w_2) \in \Omega\}
\]
with the approximation
\[ d_{\text{sad}} = \left| w_2 + h_2 \right| \text{ with } w_2 \text{ s.t. } (-h_1, w_2) \in \Omega \]
Note that this is an overestimated approximation for \( |w_2 + h_2| \). We have
\[ d_{\text{sad}} = \left| \delta (1 - h_1 h_2) \right| \frac{1}{\sigma_2^2 + \delta h_1} \]

3. COMPARISON

We compare now the asymptotic performances of MV and SAD, their complexities as well as their stochastic behavior.

3.1 Error at the optimum and Complexity

It can be shown that \( d_{\text{MV}} = d_{\text{sad}} \) only if \( |\delta| = \sigma_1 \sigma_2 \), which represent a worst case scenario for SAD. For that reason, we can state
\[ d_{\text{MV}} \geq d_{\text{sad}}. \]

The SAD optimum is always closer to the ideal solution \( w_2 = -h_2 \) than the MV optimum. This stresses the fact that SAD is less sensitive to collapsing hypothesis than MV. In an adaptive framework, it means that SAD requires less adaption control than MV.

SAD parameter space is 2-dimensional (2D search), whereas MV parameter space has dimension 1 (1D search). Therefore MV has a faster convergence to the optimum than SAD and its complexity is lower. Moreover, the SAD criterion has two minima at \(- (h_1, h_2) \) and \(- (1/h_2, 1/h_1) \).

3.2 Stochastic behavior

Beyond these structural differences, SAD saturates before MV: if the coupling coefficients \( h \) are small, which is the case in the cocooning scenario, SAD will not be able to separate the sources because the correlation of the observations is drown out in noise. Let us enlighten this phenomenon. We consider the case were only one source is present and of unit power: \( h_1 = \sigma_1^2 = 0 \) and \( \sigma_2^2 = 1 \). We force \( w_1 = 0 \) so that MV and SAD consists solely in the adaption of \( w_2 \) with the gradients expressions
\[ G_{\text{mv}} = \frac{\partial \text{E} \{ x_1 x_2^2 \}}{\partial w_2} = \text{E} \{ x_1 x_2^2 \} + w_2 \text{E} \{ x_2^4 \} \]
and
\[ G_{\text{sad}} = \frac{\partial \text{E} \{ x_1 x_2^2 \}}{\partial w_2} = \left( \text{E} \{ x_1 x_2^2 \} + w_2 \text{E} \{ x_2^4 \} \right) \text{E} \{ x_2^4 \}. \]

In practice, we have no access to the mathematical expectations, but only to their estimations, whose preciseness depends on the number of available samples and on the SNR. It is usual to assume ergodicity of the source signals and to estimate the mathematical expectation by time average. This leads to (random) estimates \( \hat{G}_{\text{mv}} \) and \( \hat{G}_{\text{sad}} \). In order to observe the effect of this approximation, a numerical experiment was carried out with a gaussian source signal \( x_2 \) of unit variance and of length \( T = 15 \) samples. We look at the variance \( \sigma_\varepsilon^2 \) of the relative error on the gradient
\[ e = \left| \frac{G_m - \hat{G}_m}{G_m} \right|, \ m \in \{ \text{MV}, \text{SAD} \}. \]

at the initial point \( w_1 = 0 \). Fig. 3 and 4 show how the statistical estimation error varies for \( h_1 = h \in [-40, 0] \) dB along with the Cramer-Rao lower bound (CRB) on \( G \):
\[ \text{CRB}(\hat{G}) = \frac{\sigma_\varepsilon^2}{2T} \left( 1 + \left| h \right|^2 + \sigma_\varepsilon^2 \right). \]

Note that the CRB remains the same independently of whether \( \sigma_\varepsilon^2 \) is known or unknown. For each value of \( h_2 \), \( \sigma_\varepsilon^2 \) was averaged on \( 10^4 \) trials.

![Figure 3: Statistical estimation error variance \( \sigma_\varepsilon^2 \) of the gradients at the initial point without background noise (hence CRB(\( \hat{G} \)) = 0) for MV (dotted line) and SAD (solid line) adaption criteria.](image)

![Figure 4: Statistical estimation error variance \( \sigma_\varepsilon^2 \) of the gradients at the initial point with background noise (SNR = 15 dB) for MV (dotted line), SAD (solid line) adaption criteria and the Cramer-Rao Bound.](image)

One can see that the estimation variance on \( \hat{G}_{\text{sad}} \) is always smaller than the estimation variance on \( \hat{G}_{\text{mv}} \), even if there is no noise. At SNR = 15 dB and for \( h < -22 \) dB, the adaption with SAD makes no progress since the error on the gradient
is in the same order of magnitude as its value. Conversely, MV still makes progress until \( h = -30 \text{ dB} \).

### 3.3 Behaviour near the separation point

To observe the behaviour of each method near the optimum, we propose to consider the value of the gradient at the separation point \((w_1, w_2) = \left(-h_1, h_2\right)\). Since this gradient should vanish at this point, its actual value offers a asymptotic goodness measure of the algorithm near convergence. We consider the equivalent MV cost function \( E\{v_1^2\}^2 \) that is homogeneous to the SAD cost function, so that the comparison is fair. Taking into account the statistics of the sources, we get

\[
e_{\text{MV}} = \frac{\partial \{E\{v_1^2\}\}^2}{\partial w_2^2} \biggm|_{w_2 = -h_2} = 2\sigma_1^2 \left| 1 - h_1 h_2 \right|^2 (1 - h_1 h_2) (\delta + h_1^2 \sigma_1^2)
\]

and

\[
e_{\text{SAD}} = \frac{\partial \{E\{v_1, v_2\}\}^2}{\partial w_2^2} \biggm|_{(w_1, w_2) = (-h_1, h_2)} = 2\delta \left| 1 - h_1 h_2 \right|^2 (1 - h_1 h_2) (\sigma_2^2 + h_1^2 \delta^2).
\]

We consider their ratio

\[
\rho = \frac{e_{\text{MV}}}{e_{\text{SAD}}} = \frac{\sigma_1^2}{\delta \sigma_2^2 + h_1^2 \delta^2}.
\]

\( \rho \) is depicted on Fig. 5 for varying spatial coupling \( h_1 \), target power \( \sigma_1^2 \) and source correlation \( \delta = \delta_0 \sigma_1 \sigma_2 \) and \( \sigma_2^2 = 1 \). We can clearly notice that SAD has no advantage when the power of \( s_i \) stays under a certain threshold and is most beneficial when the coupling \( |h_1| \) is close to 0 dB. When the coupling is small, then MV is profitable even if the VAD has a large variance. As expected, when \( \delta_0 \) gets smaller, the region where SAD is better than MV enlarges.

#### Figure 5: \( \rho \) level curves for \( |\delta_0| = 0.01 \) (left) and \( \{\rho = 1\} \) curve for \( \delta_0 \in \{0.1, 0.05, 0.01\} \).

### 4. SUMMARY AND CONCLUSION

The results of Akari et al. [1, 2] showing the superiority of beamforming (MV) over blind source separation (SAD) reflect two things: in a controlled environment, MV converges faster and provides better SIR improvement than SAD because of saturation of the latter. However, their conclusion has to be mitigated since it omits the leakage problem.

Our analysis showed that the asymptotic performance of SAD is better than that of MV particularly if the sources and coupling levels \( |h_1| \) are close to 0 dB. This performance is bounded by the correlation of the sources which is not negligible on a short time scale. SAD needs more samples than MV

1. in order to maintain an acceptable independency,
2. and because its estimation variance on the separation parameters is larger than that of MV.

In particular in noisy conditions, if the coupling level is smaller than a certain threshold, SAD will not converge at all. This threshold is smaller for MV. At last, SAD is structurally more complex (2D/1D search) than MV.

As a conclusion, MV is preferable to SAD when it is applicable, but SAD becomes necessary when the coupling level is close to 0 dB and when it is difficult to detect whether the target/interferer is active or not.

#### 5. ACKNOWLEDGEMENT

Julien Bourgeois acknowledges the financial support provided through the European Community’s Human Potential Programme under contract HPRN-CT-2002-00276 HOARSE.

#### REFERENCES


