A ROBUST ADAPTIVE FILTERING ALGORITHM AGAINST IMPULSIVE NOISE

Leonardo Rey Vega[†], *Hernán Rey*[†], *Jacob Benesty*[‡], *Sara Tressens*[†]

[†]Facultad de Ingeniería, Universidad de Buenos Aires,
Paseo Colón 850, Buenos Aires, Argentina (e-mail: lrey,hrey,stres@fi.uba.ar).
[‡]INRS-EMT, Université du Québec,
800 de la Gauchetiere Ouest, Montréal, Québec, Canada (e-mail: benesty@emt.inrs.ca).

ABSTRACT

A new framework for designing robust adaptive filters is introduced. It is based on the optimization of a certain cost function subject to a time-dependent constraint on the norm of the filter update. Particularly, we will derive a robust variable step-size NLMS algorithm which optimizes the square norm of the *a posteriori* error subject to the constraint on the norm of the filter change. We also show the link between the proposed algorithm and another one derived using a robust statistics approach. The algorithm is then tested in different environments for system identification and acoustic echo cancelation applications.

Index Terms— Adaptive filters, NLMS algorithm, impulsive noise, robust filtering, acoustic echo cancellation

1. INTRODUCTION

In real-world adaptive filtering applications, severe impairments may occur. Perturbations such as background noise and impulsive noise can deteriorate the performance of many adaptive filters under a system identification setup. In echo cancellation, double-talk situations can also be viewed as impulsive noise sources.

Many different approaches have been proposed in the literature to deal with this problem [1]– [7]. Most of them are directly or indirectly related with the optimization of a combination of L_1 and L_2 norms as the objective function. The former presents a low sensitivity against perturbations and the latter improves the convergence speed of the adaptive filter. In this work we introduce a new framework for the construction of robust adaptive filters.

First, we introduce the new framework that is based on the optimization of a certain cost function subject to a constraint on the norm of the adaptive filter update. Particularly, when the square of the *a posteriori* error is used as the cost function, the result is another algorithm that provides an automatic mechanism for switching between the normalized least-mean-square (NLMS) and normalized sign algorithm (NSA). In principle, the algorithms derived with the new approach and with the robust statistics ideas given in [8] are equivalent. However, important differences exist with respect to the assumptions used on each framework. Finally, the performance of the algorithm is tested under several scenarios in system identification and acoustic echo cancellation applications. These results with others important ones regarding the convergence behaviour of the algorithm can be found in [9].

We present certain definitions and notation that are used in the paper. Let $\mathbf{w}_i = (w_{i,0}, w_{i,1}, \ldots, w_{i,M-1})^T$ be an unknown $M \times 1$ linear finite-impulse response system. The $M \times 1$ input vector at time i, $\mathbf{x}_i = (x_i, x_{i-1}, \ldots, x_{i-M+1})^T$, passes through the system giving an output $y_i = \mathbf{x}_i^T \mathbf{w}_i$. This output is observed, but it is usually corrupted by a noise, v_i, which will be considered additive. In many practical situations, $v_i = b_i + \eta_i$, where b_i stands for the background measurement noise and η_i is an impulsive noise or an undetected near-end signal in echo cancellation applications. Thus, each input $\mathbf{x}_i$ gives an output $d_i = \mathbf{x}_i^T \mathbf{w}_i + v_i$. We want to find $\hat{\mathbf{w}}_i$ to estimate $\mathbf{w}_i$. This adaptive filter receives the same input, leading to an output error $e_i = d_i - \mathbf{x}_i^T \hat{\mathbf{w}}_{i-1}$. We also define the misalignment vector $\tilde{\mathbf{w}}_i = \mathbf{w}_i - \hat{\mathbf{w}}_i$ and the *a posteriori* error signal $e_{\mathrm{p},i} = \mathbf{x}_i^T \tilde{\mathbf{w}}_i + v_i$.

This work was supported in part by Universidad de Buenos Aires, project UBACYT I005.

2. NEW FRAMEWORK FOR DERIVATION OF ROBUST ADAPTIVE FILTERS

Suppose an adaptive filter has a given estimate of the true system at a certain time-step. Now, if a large noise sample perturbs it, the result will be a large change in the system estimate, degrading the performance of the adaptive filter. To prevent these situations, the proposed approach is to constrain the energy of the filter update at each iteration. This can be formally stated as:

$$||\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}||^2 \leq \delta_{i-1}, \tag{1}$$

where $\{\delta_i\}$ is some positive sequence. Its choice will influence the dynamics of the algorithm. Nevertheless, (1) guarantees that any noise sample can perturb the square norm of the filter update by at most the amount δ_{i-1}.

Next, a cost function is required and the adaptive filter will be the result of optimizing this cost function subject to the constraint (1). Different choices of the cost function and the $\{\delta_i\}$ sequence will lead to different algorithms.

2.1. Special Case With the A Posteriori Error Cost Function

Now, we propose to find the updating strategy as:

$$\hat{\mathbf{w}}_i = \arg \min_{\hat{\mathbf{w}}_i \in \mathbb{R}^M} e_{\mathrm{p},i}^2, \tag{2}$$

subject to the constraint (1).

To perform this optimization, we divide the problem in two cases: (a) the hypersphere defined by (1) has a *non-empty* intersection with the hyperplane defined by $e_{\mathrm{p},i} = 0$. (b) the hypersphere defined by (1) has an *empty* intersection with the hyperplane defined by $e_{\mathrm{p},i} = 0$. These situations are graphically shown in Fig. 1 for the case $M = 2$.

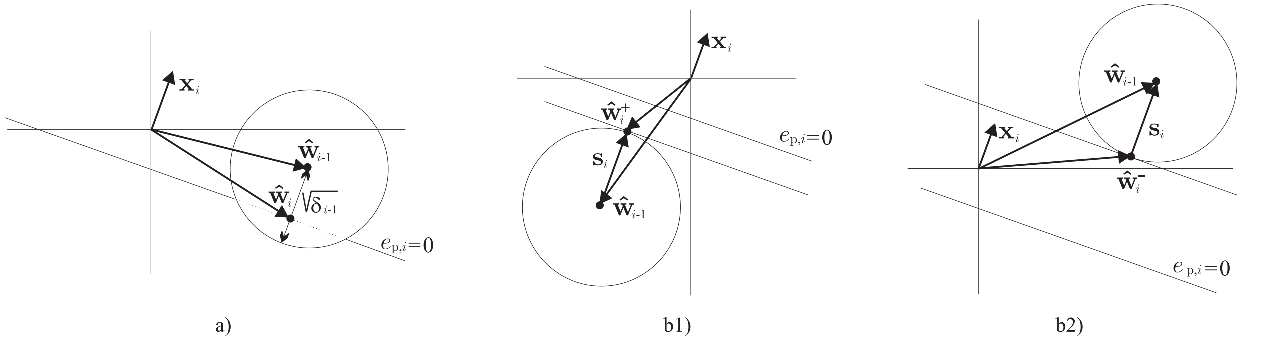

Fig. 1: a) In this case, an infinite number of solutions exists (represented by the dotted line). Out of those solutions, we choose that which provides the lowest energy in the update. b1) and b2) Two updates, $\hat{\mathbf{w}}_i^+$ and $\hat{\mathbf{w}}_i^-$, are possible depending on the relative position between the hyperplane and the hypersphere.

In the first case, there is an infinite number of valid solutions. Among them, we particularly choose that which provides the lowest energy in the update, i.e., $\min ||\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}||^2$. As in this case the *a posteriori* error is zero, the solution is the popular NLMS algorithm, i.e.,

$$\hat{\mathbf{w}}_i = \hat{\mathbf{w}}_{i-1} + e_i \frac{\mathbf{x}_i}{||\mathbf{x}_i||^2}. \tag{3}$$

This update has to be used at each time-step when the distance between the point $\hat{\mathbf{w}}_{i-1}$ and the hyperplane defined by $e_{\mathrm{p},i} = 0$ is smaller than $\sqrt{\delta_{i-1}}$. It can be easily shown that this is satisfied when:

$$\frac{|e_i|}{||\mathbf{x}_i||} \leq \sqrt{\delta_{i-1}}. \tag{4}$$

In the second case, when the hyphersphere defined by (1) has an empty intersection with the hyperplane defined by $e_{\mathrm{p},i} = 0$, two different possibilities can be considered as shown in Fig. 1. The one that leads to the minimum $e_{\mathrm{p},i}^2$ should be used as the update for a certain time-step, as long as (4) is not fulfilled. It can be easily shown that this could be checked with the sign of the estimation error (the details are given in [9]). Thus, with the initial condition $\hat{\mathbf{w}}_0$ and for some $\{\delta_i\}$, the new algorithm can be put in a compact way such as:

$$e_i = d_i - \mathbf{x}_i^T \hat{\mathbf{w}}_{i-1},$$
$$\hat{\mathbf{w}}_i = \hat{\mathbf{w}}_{i-1} + \min\left[\frac{|e_i|}{||\mathbf{x}_i||}, \sqrt{\delta_{i-1}}\right] \mathrm{sign}(e_i) \frac{\mathbf{x}_i}{||\mathbf{x}_i||}. \tag{5}$$

2.2. Link With the Robust Statistics Approach

In [1] an adaptive algortihm is derived using a robust statistic approach [8]. In that work a non-linear function of the error is optimized using a stochastic gradient approach. That function depends on $p(v)$, the assumed PDF of the disturbing noise. The algorithm derived is similar to that given in (5), but with a time-independent value of δ and a general normalized factor instead of $\|\mathbf{x}_i\|$. Although the two algorithm seem to be very similar they are derived under very different frameworks. Moreover the meaning of δ is quite different in both cases. In the algorithm given in [1], it is the cutoff of the noise PDF (we could consider the error PDF, which will give a time-dependent δ, but this would require the knowledge of the evolution of the error statistic which could be difficult to have in practice). Thus, the dynamics (if we consider the error PDF) of the delta sequence will be dependent on the PDF assumed in the model. In the new algorithm, δ is the square of the radius of a certain hypersphere centered at $\hat{\mathbf{w}}_{i-1}$. At every time-step, the new adaptive filter is allowed to evolve only to a new point inside this hypersphere. As a consequence, the dynamics of the delta sequence could be absolutely arbitrary.

It should be noted that the new framework introduced allows the design of robust adaptive filters without requiring any statistical information of the noise v_i nor the error signal e_i. This is probably the most important difference with the robust statistics approach.

2.3. Choice of the Delta Sequence

The new algorithm has two operation modes: an NLMS with $\mu = 1$ or the NSA with step-size $\sqrt{\delta_{i-1}}$. However, this last update is only used when $\sqrt{\delta_{i-1}} < |e_i|/||\mathbf{x}_i||$. So we can view the update of the new algorithm as another NLMS with step-size:

$$\mu' = \min\left[\frac{\sqrt{\delta_{i-1}}}{|e_i|/||\mathbf{x}_i||}, 1\right]. \tag{6}$$

This fact allows us to interpret the new algorithm as a variable step-size NLMS algorithm. Its step-size belongs to $(0, 1]$. For this reason, we name the new algorithm as *robust variable step-size NLMS* (RVSS-NLMS). It is very clear how the algorithm operates. When the normalized absolute of the error is smaller than $\sqrt{\delta_{i-1}}$ the adaptive filter acts like a NLMS algorithm with unit step-size. On the other hand, when $\sqrt{\delta_{i-1}}$ is exceeded, a large sample noise might be present and the algorithm acts like a NSA with step-size equal to $\sqrt{\delta_{i-1}}$. It should be very clear why a time-independent δ is not a good option. If δ is too small, the algorithm will be robust but extremely slow. On the contrary, if δ is large, the robust behavior is not possible. This is a consequence of the transmission of large amplitude noise samples to the adaptive filter through the error signal in the NLMS update. Then we look for a time-dependent delta sequence.

In principle, one would desire δ_i to have values as large as possible at the beginning of the adaptation. This will lead to a good initial speed of convergence because the NLMS will act in this situation. Still, it should not be too large, so that the robust performance against large noise samples is not lost. On the other hand, when the algorithm is close to its steady-state, lower values of δ_i will lead to a lower final error. This behavior can not be achieved using a fixed parameter δ.

A natural selection should make the sequence $\{\delta_i\}$ dependent on the convergence dynamics of the adaptive filter. Thus, we propose:

$$\begin{aligned}\delta_i &= \alpha\delta_{i-1} + (1-\alpha)||\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}||^2 \\ &= \alpha\delta_{i-1} + (1-\alpha)\min\left[\frac{e_i^2}{||\mathbf{x}_i||^2}, \delta_{i-1}\right],\end{aligned} \tag{7}$$

where $0 < \alpha < 1$ is a memory factor. In [9] we proved under very mild conditions that this sequence converges to zero almost everywhere, and that under the assumption of a stationary environment the adaptive filter converges with probability one to the true system, i.e., the adaptive filter estimates the true stationary system with zero variance if we wait a sufficiently long time.

Based on the dynamics of (7), δ_i is decreased only when $\sqrt{\delta_{i-1}} > |e_i|/||\mathbf{x}_i||$. In this case, the performed update is the NLMS with $\mu = 1$. So if the sequence $\{\delta_i\}$goes to zero as $i \to \infty$, the above condition should be satisfied an infinite number of times. The reason why this NLMS algorithm can still have a robust performance against noise is that these (infinite number of) updates take place only at the iterations where the error is small enough. Thus, the performance of the NLMS is not compromised even when the update is done using $\mu = 1$. This property allows the algorithm to outperform the NLMS with fixed $\mu = 1$ in both impulsively and non-impulsively perturbed environments.

3. PRACTICAL CONSIDERATIONS

In this section we discuss certain algorithm implementation issues. First of all, when the NLMS update is used, a regularization constant $\beta = 20\sigma_x^2$ is added to the denominator. Also, every time a division by $\|\mathbf{x}_i\|$ is computed, a small constant ϵ is added to the denominator to avoid division by zero. These changes still guarantee that the constraint (1) is satisfied. The forgetting factor α is chosen according to the rule of thumb:

$$\alpha = 1 - \frac{1}{\kappa M}, \tag{8}$$

where κ is a parameter that depends on the color of the input signal and tipically ranges in [1,6].

Finally, a major issue should be considered carefully. As the proposed delta sequence has the decreasing property shown mentioned in Section 2, although the algorithm becomes more robust against perturbations, it also loses its tracking capacity. For this reason, if there is a chance of being in a non-stationary environment, an *ad hoc* control should be included. The objective is to detect changes in the true system. We cannot include the description of this control here for lack of space. They are presented with detail in [9].

4. SIMULATION RESULTS

The system is taken from a measured acoustic impulse response and it was truncated to $M = 512$. Its gain is normalized so that the input and output powers are equal, i.e., $\sigma_x^2 = \sigma_y^2$. The adaptive filter length is set to M in each case. We use the *mismatch* as a measure of performance:

$$10 \log_{10} \left[\frac{||\tilde{\mathbf{w}}_i||^2}{||\mathbf{w}_i||^2} \right]. \tag{9}$$

The plots are the result of single realizations of all the algorithms without any additional smoothing. A zero-mean Gaussian white noise b_i is added to the system output such that a certain SBNR is achieved:

$$\mathrm{SBNR} = 10 \log_{10} \left[\frac{\sigma_y^2}{\sigma_b^2} \right], \tag{10}$$

where σ_b^2 is the power of the background noise.

The behavior of the proposed algorithm is compared with other strategies. We simulate an NLMS algorithm with $\mu_1 = 1$ and a regularization factor $\beta = 20\sigma_x^2$ and a NSA algorithm:

$$\hat{\mathbf{w}}_i^{\mathrm{sign}} = \hat{\mathbf{w}}_{i-1}^{\mathrm{sign}} + \mu_2 \ \mathrm{sign}(e_i) \frac{\mathbf{x}_i}{||\mathbf{x}_i|| + \epsilon}, \tag{11}$$

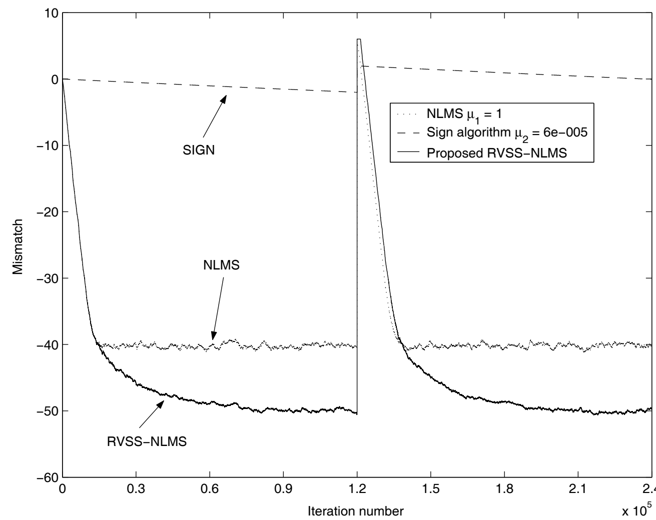

Fig. 2: Mismatch (in dB). AR1(0.8) input. No impulsive noise. SBNR=40 dB. The curves are the result of a single realization of the algorithms.

with the μ_2 value that gives the same steady-state mismatch as that of the proposed RVSS-NLMS.

We also tried different schemes proposed in the literature [3]–[7] but as the results were not good under the experimental setup chosen here, we decided not to show them. Another possibility is to include the algorithm derived in Section 2 with a fixed time-independent δ. However, it should be clear that this will lead to a slow robust performance or a fast non-robust performance (this was already discussed in Section 2).

4.1. System Identification Under Impulsive Noise

The input process is an AR1 with pole in 0.8. In addition to the background noise b_i, an impulsive noise η_i could also added to the output signal y_i. The impulsive noise is generated as $\eta_i = \omega_i N_i$, where ω_i is a Bernoulli process with probability of success $P[\omega_i = 1] = p_{\mathrm{imp}}$ and N_i is a zero-mean Gaussian with power $\sigma_N^2 = 1000\,\sigma_y^2$. A sudden change in the true system is introduced at a certain time-step by multiplying the system coefficients by -1.

In Fig. 2, no impulsive noise is present. As can be seen, the NLMS has a good initial convergence while the NSA has a dramatically slow performance. However, the RVSS-NLMS can extract the good properties of both. It shows the same speed of convergence of the NLMS but with a 10 dB lower steady-state error. This is possible because when δ_i is small enough, the algorithm employs the NLMS update with $\mu' < 1$, which allows it to reach a lower steady-state level.

Then we include the impulsive noise with $p_{\mathrm{imp}} = 0.1$. In Fig. 3, the NLMS with $\mu = 1$ has a positive mismatch (in dB), while the performance of the other two algorithm remains barely unchanged. The interesting thing is that the NLMS update with $\mu = 1$ is performed by the proposed algorithm only at the iterations when the error is small enough. On the other hand, as can be seen from (6), the larger the error (probably due to the appearance of a large noise sample), the smaller the value of μ' used in the update. This allows an NLMS algorithm to perform robustly against impulsive noise. In [9] the algorithm was tested in others identification scenarios showing a very good performance.

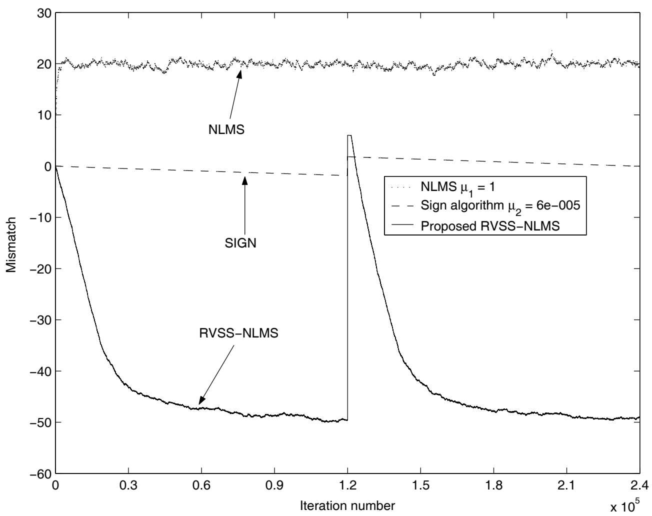

Fig. 3: Mismatch (in dB). AR1(0.8) input. $p_{\mathrm{imp}} = 0.1$. The other parameters are the same as in Fig. 2.

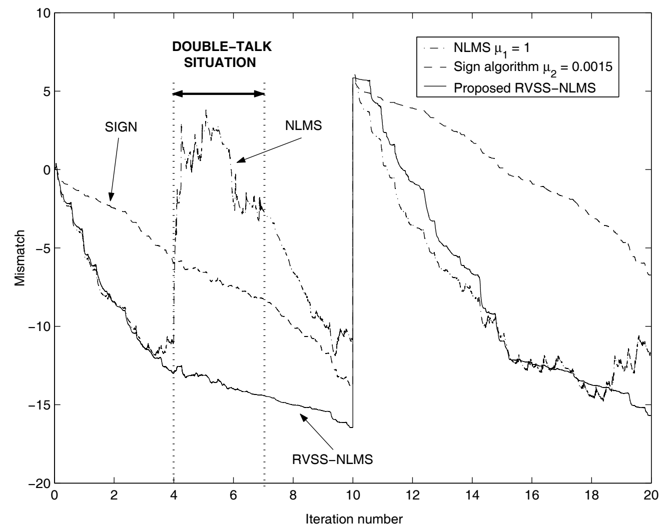

Fig. 4: Mismatch (in dB) for speech input. SBNR=20 dB. STNR=0 dB.

4.2. Acoustic Echo Cancellation With Double-Talk Situations

In echo cancellation applications, a double-talk detector (DTD) is used to suppress adaptation during periods of simultaneous far- and near-end signals. A simple and efficient way of detecting double-talk is to compare the magnitude of the input and output signals and declare double-talk if the output magnitude is larger than a value set by the input signal. A proven algorithm that has been in commercial use for many years is the Geigel DTD [10]. Although the adaptation is usually inhibited for a certain period of time when double-talk is declared, we choose to stop it just for a single time-step. If the algorithm is robust enough to deal with the undetected near-end signal, this decision will avoid an unnecessary decrease in the speed of convergence.

The far-end and near-end signals are speech sampled at 8 kHz. The SBNR is 20 dB while the *signal to total noise ratio* (STNR), i.e.,

$$\mathrm{STNR} = 10 \log_{10} \left[\frac{\sigma_y^2}{\sigma_b^2 + \sigma_h^2} \right],$$

is set to 0 dB, where σ_h^2 is the power of the near-end signal before passing through the DTD. As shown in Fig. 4, the NLMS with $\mu = 1$ is highly affected by the double-talk situation. This is not the case for the other two algorithms but the RVSS-NLMS performs faster than the NSA. When the sudden change occurs, the RVSS-NLMS can track the system with a small delay with respect to the NLMS while the NSA shows a very poor performance.

5. CONCLUSIONS

A new framework for designing robust adaptive filters was introduced. It is based on the optimization of a certain cost function subject to a time-dependent constraint ($\{\delta_i\}$) on the norm of the filter update. Particularly, we derived the RVSS-NLMS algorithm by optimizing the square of the *a posteriori* error. We showed that the new algorithm is equivalent to another one derived using a robust statistics approach. However, the new framework requires no statistical information about the probability density functions of the noise nor the error signal. Then, we proposed a certain dynamics for $\{\delta_i\}$ which provides the algorithm with fast initial convergence as the NLMS with $\mu = 1$ but also a performance against noise as robust as the NSA. As shown in the simulations, even under severe conditions, the performance of the algorithm was very good.

6. REFERENCES

[1] T. Gänsler, S.L. Gay, M.M. Sondhi, and J. Benesty, “Double-talk robust fast converging algorithms for network echo cancellation,” *IEEE Trans. Speech Audio Processing*, vol. 8, no. 6, pp. 656–663, Nov. 2000.

[2] T. Gänsler and J. Benesty, “The fast normalized cross-correlation double-talk detector,” *Signal Processing*, vol. 86, no. 6, pp. 1124–1139, Jun. 2006.

[3] P. Petrus, “Robust Huber adaptive filter,” *IEEE Trans. Signal Processing*, vol. 47, no. 4, pp. 1129–1133, Apr. 1999.

[4] Y. Zou, S.C. Chan, and T.S. Ng, “Least mean M-estimate algorithms for robust adaptive filtering in impulse noise,” *IEEE Trans. Circuits Syst. II*, vol. 47, no. 12, pp. 1564–1569, Dec. 2000.

[5] E.V. Papoulis and T. Stathaki, “A normalized robust mixed-norm adaptive algorithm for system identification,” *IEEE Signal Processing Lett.*, vol. 11, no. 1, pp. 56–59, Jan. 2004.

[6] E.V. Papoulis and T. Stathaki, “Improved performance adaptive algorithm for system identification under impulsive noise,” *Electronics Letters*, vol. 39, no. 11, pp. 878–880, May 2003.

[7] J. Arenas-García and A.R. Figueiras-Vidal, “Adaptive combination of normalized filters for robust system identification,” *Electronics Letters*, vol. 41, no. 15, pp. 874–875, Jul. 2005.

[8] P.J. Huber, *Robust Statistics*, New York: John Wiley and Sons, 1981.

[9] L. Rey Vega, H. Rey, J. Benesty, S. Tressens, “A new robust variable step-size NLMS algorithm,” *IEEE Trans. Signal Processing*, submitted.

[10] D.L. Duttweiler, “A twelve channel digital echo canceler,” *IEEE Trans. Comm.*, vol. 26, no. 5, pp. 647–653, May. 1978.