# DYNAMIC ADJUSTMENT OF THE FORGETTING FACTOR IN ADAPTIVE FILTERS FOR NON-STATIONARY NOISE CANCELLATION IN SPEECH

*R. Martínez, P. Gómez, A. Álvarez, V. Nieto, V. Rodellar, M. Rubio and M. Pérez*
Dept. Arquitectura ya Tecnol. de Sist. Inf., Facultad de Informática
Universidad Politécnica de Madrid, Campus de Montegancedo, s/n
Boadilla del Monte, 28660, Madrid, SPAIN

## ABSTRACT

The field of Speech Recognition in Noisy Environments is of vital importance for the success of certain applications in the domain of Communications, Automotion, Avionics, and other harsh situations where common recognizers fail to meet acceptable standards due to the negative influence of environmental noise collected during the recording of the Speech Trace [1]. To overcome such difficulties, Adaptive Filtering has been traditionally used with acceptable success in noise cancellation using two-microphone schemes. Some problems found in such situations are due to the non-stationary behavior of Speech and Noise, although the sudden changes in the energy levels of both Speech and Noise are the cause of frequent misadjustments and un-locking in the Adaptive Algorithms, rendering the task of noise cancellation a difficult one. Through the present paper, a method based on the continuous tracking of the Energy Differences between Speech and Noise is presented, which may be used for the dynamic adaptation of the Forgetting Factor in the Adaptive Filters, producing an effective control of the sudden changes in the Signal/Noise Ratio. The paper discusses the proposed methods and shows some results from practical conditions. These results show good stability and a special ability for Word-Boundary Detection under Highly Non-Stationary Conditions, especially useful in Isolated-Word Speech Recognition.

## 1. INTRODUCTION

Speech Recognition under high noisy conditions is a difficult task. Most recognition methods, which have shown to be highly efficient under noise-free conditions fail dramatically with S/N ratios around or below 10 dB as the one shown in Fig. 1. This is a set of isolated One-Command Words, corresponding to the sequence */left/, /right/, /up/, /down/, /go/,* and */stop/*, recorded by a primary microphone within a field of about 85-90 dB SPL. During the first part of the speech sequence (first four speech bursts), the noise level is around 85 dB. The energy of the speech signal is 10-15 dB above the noise level in the peaks. The last two words are immersed in a noise level above 90 dB. In this case the energy level of the speech trace is around 6-10 dB. One of the consequences of these high noise levels is that most Begin-End Point Detectors fail to separate properly the speech segments of the noise ones. Therefore, the speech recognition mechanisms will not have a clear boundary to start the processing of the signal, and as a consequence, speech segments will be lost, and noisy segments will be used in recognition. The overall reliability of the Speech Recognition System will dramatically experience the consequences of this impairment in its results. At this point we have to take into account that the S/N ratio changes rapidly when there is a sudden burst of energy due to a speech utterance (as the speaker forces loudness of speech to make himself understandable), and that the noise level is not uniform in most common environments. Abrupt changes have to be faced with new adjustments in the values of the adaptation factors of the Noise Cancellers, as otherwise, unlocking and instability would be present, thus reducing the efficiency of the canceling mechanisms. What is being proposed in the following sections is to use the side information provided by an Adaptive Noise Canceller to improve the adaptation methods to significantly extend the overall reliability of the recognition process under non-stationary changes in the S/N ratio.

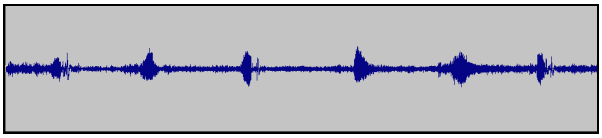

**Figure 1.a.** Noisy Speech Trace (Primary Microphone)

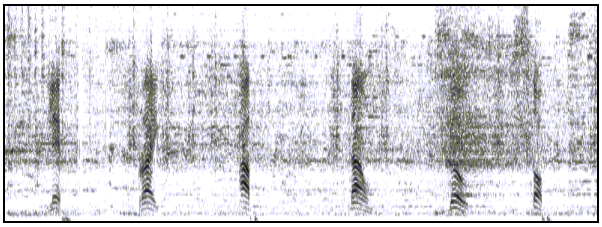

**Figure 1.b**. Power spectrum of the Noisy Speech Trace.

## 2. METHODOLOGY

The Noise Canceling Scheme, which may be seen in Fig. 2 is based in an Adaptive Lattice-Ladder Filter [2-4] to process two signal channels recorded by a two-microphone array (Speech Source or Primary and Noise Source or Reference). According with the strategies classically used to combine the different estimators of Noise and the *backward residuals*, several algorithms may be devised to update the weights of the *ladder filter*. These give place to different implementations of the basic canceling scheme. The ones initially checked in the present research [5] were:

a) Gradient Lattice Ladder [3, 4]. The adaptation step used for this filter was $\alpha$=0.9998.
b) RLS Lattice Ladder algorithm (Recursive LSL algorithm using *a posteriori* estimation errors) [3, 4]. The adaptation step used for this filter was $\alpha$=0.9999.
c) Direct Update (error feed-back) form of the RLS Lattice Ladder algorithm (Recursive LSL algorithm using *a priori* estimation errors with error feedback)

[3, 4]. The adaptation step used for this filter was $\alpha$=0.9997.

In all the cases the factor of *lattice initial residual error energy* was taken as $\varepsilon=5.10^8$, this value ensuring a high lock-up performance and an acceptable degree of stability during lock-up for relatively stationary conditions regarding the S/N ratio. The best microphone separation found was in the range from 15-30 cm., for which the best canceling properties were observed. The filter dimensions ranged correspondingly from 5+5 to 10+10 delay and processing stages, these values corresponding to a sampling rate of 11050 Hz. The two input channels were transformed by the Lattice-Ladder system in a new trace of Clean Speech. The average amount of cancellation obtained with these settings resulted in a S/N enhancement ranging from 9 to 12 dB.

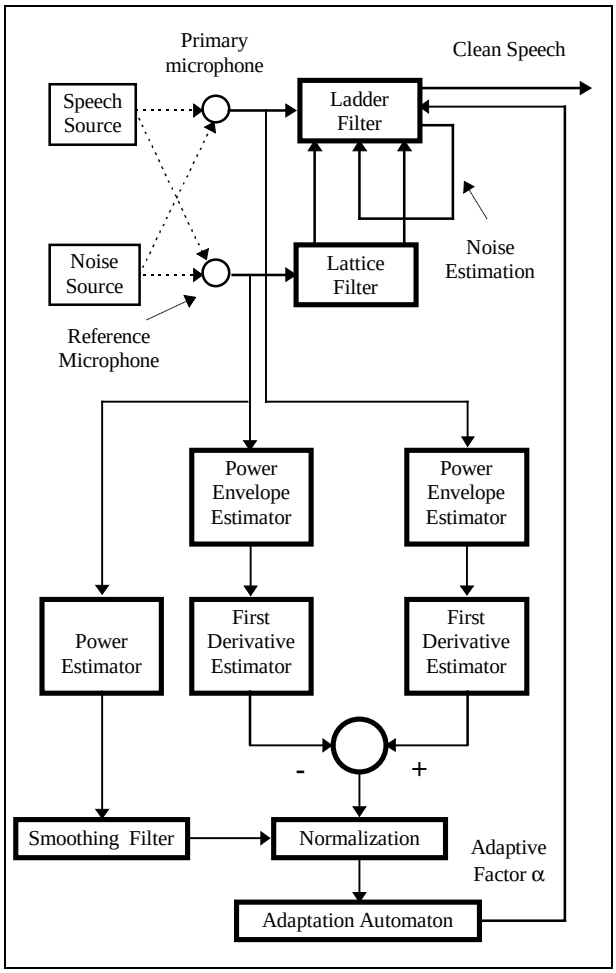


**Figure 2.** General framework for the automatic adjustment of the adaptation parameter in a Lattice-Ladder Noise Canceller.

Nevertheless, it was observed that the optimum values for the adapting factor $\alpha$ were not independent from the S/N ratio, especially during the insertion and decay of the speech bursts. During these periods the adaptive filters tend to lose the tracking of the signal (unlocking) In general, a value for $\alpha$ close to the unity produces a slower adaptation rate but keeps the cancellation ability within reasonable limits. On the other hand, smaller values for $\alpha$ would result in a faster re-locking process at the cost of decreasing the overall accuracy of the noise canceller. Therefore, a certain method to dynamically correct the value of the forgetting factor should be conceived to better control the presence of sudden and loud speech bursts over the noise level.

The proposed method relies on the continuous tracking and comparison of the power envelopes of both the primary and the reference signals, given as follows:

$$V_s(m) = max\ \{s^2((m-1)N+k))\};\qquad 0<k<N \quad (1)$$

$$V_r(m) = max\ \{r^2((m-1)N+k))\};\qquad 0<k<N \quad (1')$$

where $s(n)$ and $r(n)$ are respectively the primary (noisy speech) and reference (noise) signals, $n$ is the time index, and $m$ is the frame index. The envelopes of both signals $V_s(m)$ and $V_r(m)$ are evaluated on a frame-by-frame basis in non-overlapping frame windows of $N=128$ samples each. An estimator of the first derivative of each envelope is then evaluated, following the expressions:

$$V'_s(m) = -2\ V_s(m-2) - V_s(m-1) + V_s(m+1) + 2\ V_s(m+2) \quad (2)$$

$$V'_r(m) = -2\ V_r(m-2) - V_r(m-1) + V_r(m+1) + 2\ V_r(m+2) \quad (2')$$

The difference between both derivatives will remark the segments where speech is explicitly present above the noise level, as the effect of sudden variations in the level of noise are smoothed:

$$D(m) = V'_s(m) - V'_r(m) \quad (3)$$

This magnitude is then normalized against the average power of the reference signal $\overline{R}_0(m)$ to compensate differences in the gain of the signal acquisition chain:

$$d(m) = \frac{D(m)}{\overline{R}_0(m)} \quad (4)$$

where $\overline{R}_0(m)$ has been averaged over a window of $L=50$ non-overlapping frames:

$$\overline{R}_0(m) = \sum_{k=L-1}^{0} R_0(m-k) \quad (5)$$

with $R_0(m)$ evaluated within each frame as:

$$R_0(m) = \sum_{k=M}^{M-1} r^2((m-1)N+k) \quad (6)$$

In what follows we will refer to $d(m)$ as the *Normalized Average Differential Power* (NADP). This amount is a good tracer of the changes in the S/N ratio, and will be used to induce the transitions in a three-state automaton, which will

decide on which kind of adaptation rate will be better used (fast-inaccurate or slow-accurate). The behavior of the NADP may be summarized in Fig. 3. Traces a) and b) represent the energy envelopes of the noisy speech *s(n)* and reference noise *r(n)*. Their respective *smoothed* derivatives are given in c) and d). Trace e) would then represent the NADP. This last trace is thresholded by two empirically established values $\mu_1$ and $\mu_2$, which determine the decision-taking limits for the adjusting automaton, accordingly with environmental facts, as microphone behavior and quality, noise level and others.

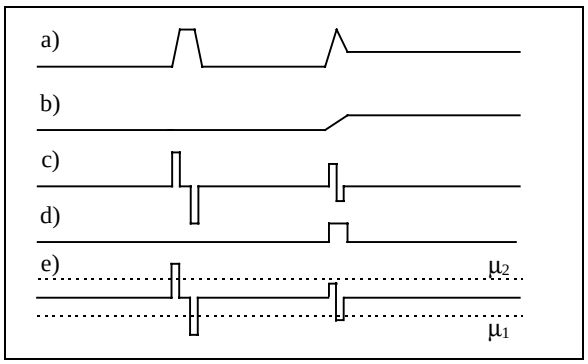


**Figure 3**. Decision-taking thresholding to use fast vs. slow convergence adaptation factors.

The adaptation automaton is composed of three states: $S_0$ or *neutral*, in which the S/N ratio is stationary; $S_1$ or *onset*, in which a sudden increase in the S/N ratio has been detected, and $S_2$ or *decay*, in which the adaptation step has to be modified to rapidly react to the new conditions in the S/N for better results. The flow diagram of the adaptation automaton may be seen in Fig. 4.

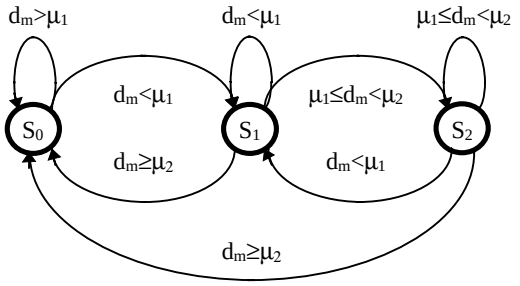


**Figure 4**. Adaptation Automaton to detect the changes in the S/N ratio and dynamically adjust the adaptive filter

The values of the adaptation steps in the experiments presented here are assigned depending on the state in which we are for a given S/N situation. States $S_0$ and $S_1$ will force $\alpha = 0.9999$, imposing a slow but accurate adaptation process. During state $S_2$ $\alpha = 0.992$, thus imposing a faster but rather less accurate adaptation process.

## 3. RESULTS AND DISCUSSION

To test the performance of the proposed thresholding method for the dynamic adjustment of the forgetting factor $\alpha$ a set of experiments was conducted which is summarized in what follows. In a first step an LSL (direct update) Lattice-Ladder Filter was used with the conditions exposed in Section 1. The output signal, and its corresponding spectrogram are given in Figs. 5.a and b. Figure 5.c gives the average energy difference between the noisy speech input and the clean speech output. It may be seen that the inadequate value of the forgetting factor, when is kept at a constant value, does not grant a fast recovery of the canceling ability of the filter, thus resulting in an impairment of the noise reduction.

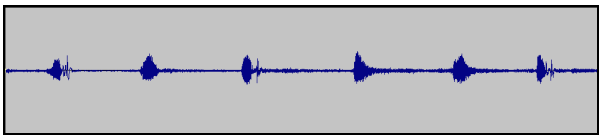

**Figure 5.a.** Clean speech trace at the output of an LSL Lattice-Ladder Filter. No adaptation of the forgetting factor.

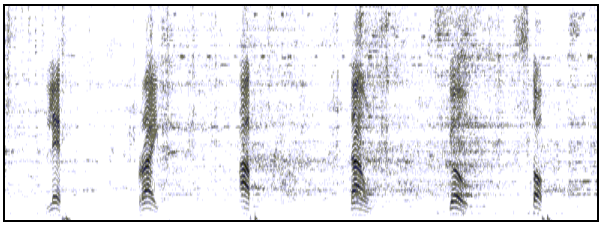

**Figure 5.b.** Spectrogram corresponding to the speech trace in Figure 5.a. Compare against that in Figure 1.b.

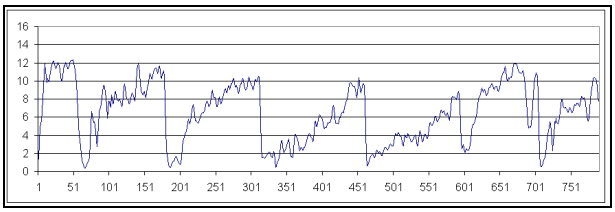


**Figure 5.c.** Average Energy Difference between the clean and noisy speech. The vertical axis is in dB, the horizontal axis is the frame index *m*.

It should be expected that during speech fragments the difference in energy between the noisy and clean speech traces would reduce to almost 0 dB, as most of the energy is due to speech, but during the silent segments, noise reduction could be within 10 an 12 dB. Due to the recovery time needed for the filter to *forget* the high dynamic range sweeps present during the speech segments, this recovery adopts the form of a decay, seen as a saw tooth in the average energy difference. In a second step the dynamic assignment method of the forgetting factor was used as explained before. The clean speech output in this case is given in Fig. 6.a, its spectrogram being plotted in Fig. 6.b.

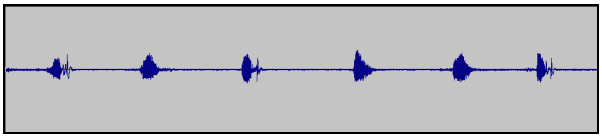

**Figure 6.a.** Clean speech trace at the output of an LSL Lattice-Ladder Filter with adaptation of the forgetting factor.

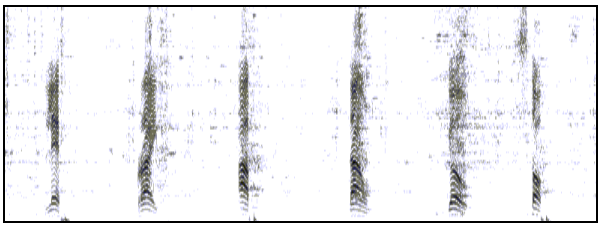

**Figure 6.b.** Spectrogram corresponding to the speech trace in Figure 5.a. Compare against those in Figs 1.b. and 5.b.

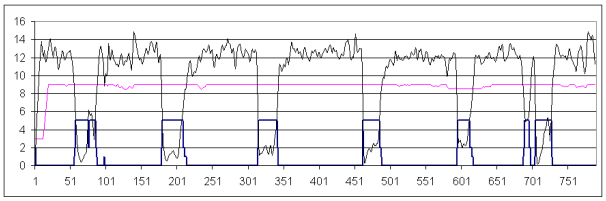


**Figure 6.c.** Average Energy Difference between the clean and noisy speech when dynamic assignment of the forgetting factor is used (upper trace). Proposed segmentation of speech using the Average Energy Difference (lower trace) The vertical axis given in dB, the horizontal axis gives the frame index *m*.

In this second case, it may be seen that the Average Energy Difference (Fig. 6.c) shows a much faster recovery to the dynamic changes in the signal, and the average level of noise cancellation is well above 12 dB during the noisy fragments of speech. This same trace may be used to detect the speech boundaries from the noisy background [6, 7], as pointed out by the lower trace in this same figure.

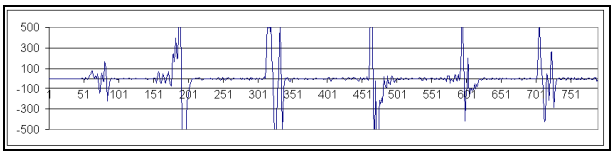


**Figure 7.a.** Normalized Average Differential Power corresponding to the experiment shown in Figs. 6.a-c.

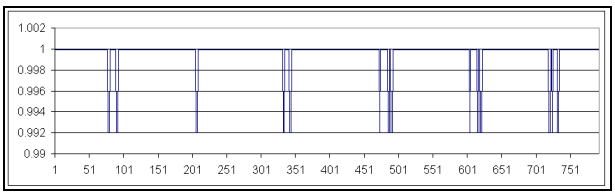


**Figure 7.b.** Value assignments for the forgetting factor of the Adaptive LSL filter ($\alpha$).

The NADP trace corresponding to the case shown may be seen in Fig. 7.a. This time series is signaling clearly the speech bursts where the adaptation of the forgetting factor should be carried out accordingly with the automaton given in section 2. The practical values used for both decision-taking thresholds were respectively $m_1$=-180 and $m_2$=180. The corresponding value assignments for the forgetting factor are shown in Fig. 7.b. As a conclusion, it must be pointed out that the dynamic assignment allows a faster recovery of the adaptation process (fast re-locking), produces a cleaner speech trace, keeps the average cancellation behavior as constant as possible, and helps in deciding on the begin-end point detection process. Applications of the method shown can be found in Robust Isolated-Word Speech Recognition, Clean Speech Communications, and others [8].

## 4. ACKNOWLEDGMENTS

The present research is being carried out under grants TIC95-0122, TIC96-1889-CE, ESPRIT IVORY nº 20277 and TIC97-1011.

## 5. REFERENCES

[1] Furui, "Recent Advances in Robust Speech Recognition", *Proc. of the ESCA-NATO Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels*, Pont-à-Mousson, France, 17-18 April 1997, pp. 11-20.

[2] Deller, J. R., Proakis, J. G. and Hansen, J. H. L., *Discrete Time Processing of Speech Signals*, MacMillan, 1993.

[3] Haykin, S., *Adaptive Filter Theory*, 3rd Ed., Prentice-Hall, Englewood Cliffs, N.J., 1996.

[4] Proakis, J. G., *Digital Communications*, 2nd. Ed, McGraw Hill, 1989.

[5] Martínez, A. Álvarez, V. Nieto, V. Rodellar and P. Gómez, "ASR in Highly Non-Stationary Environments using Adaptive Noise Canceling Techniques", *Proceedings of the ESCA-NATO Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels*, Pont-à-Mousson, France, 17-18 April, 1997, pp. 181-184.

[6] Martínez, A. Alvarez, V. Nieto, V. Rodellar and P. Gómez, "Implementation of an Adaptive Noise Canceller on the TMS320C31-50 for Non-Stationary Environments", *Proc. of the 13th International Conference on Digital Signal Processing*, Santorini, Greece, 2-4 July, 1997 pp. 49-52.

[7] Martínez, A. Alvarez, P. Gómez, M. Pérez, V. Nieto, V. Rodellar, "A Speech Pre-Processing Technique for End-Point Detection in Highly Non-Stationary Environments", *Proc. of EUROSPEECH'97*, Rhodes, Greece, 22-25 September, 1997, pp. 1111-1114.

[8] IVORY project: http://moral.datsi.fi.upm.es/projects/IVORY/IVORY.html.