Speech Enhancement Based On Noise Reduction

Kundan Kumar Singh
Electrical Engineering Department
University Of Rochester
ksingh11@z.rochester.edu

Yating Gu
Electrical Engineering Department
University Of Rochester
ygu9@ur.rochester.edu

ABSTRACT

This paper addresses the problem of signal distortion often faced in real life scenarios like teleconferencing, VoIP, cellular calls, speech recognition due to presence of unwanted background noise. This acoustic noise gets automatically added to the signal and is picked up by microphone causing a reduction in the perceived quality or intelligibility of the audio signal at the receiver end.
Consequently the techniques for enhancing speech degraded by uncorrelated additive noise, when only noisy speech is available, have been widely studied in the past and are still an active field of research.
Some of the feasible solutions include implementation of certain speech enhancement algorithms at the receiver side to enhance perceived sound quality or by using hearing aids which have built in noise reduction hardware.
In this paper, we present various adaptive algorithms developed for noise cancellation in past few years namely LMS (Least Mean Square), NLMS (Normalized Least Mean Square) and RLS (Recursive Least Square) and LPC (Linear Predictive Coding). We then compare the noise cancellation performance of these algorithms in terms of percentage of noise removal in the restored signal.
The simulation results on MATLAB confirmed that RLS algorithm leads to best subjective results and objective results but at the cost of large computational complexity and memory requirement.

Keywords: ***Adaptive filter; noise reduction; LMS; Mean Squared Error; NLMS; RLS; LPC.***

1. INTRODUCTION

In the process of transmission of voice signal from the source to receiver side, noise from the various sources gets added and corrupts the signal. These sources of noise can be due to the co-channel interference over the wireless communication channel, shot and thermal noise due to circuit elements present in the handsets, environmental sources like background sounds of winds, traffic, people etc. Several noise reduction techniques aim to suppress the effect of noise without introducing any perceptible distortion in the signal with an underlying assumption that the system itself is ideal (i.e. system is not adding any noise to the signal by itself) and that the only environmental sources of noise are responsible for signal distortion.

Thus, techniques for effective removal or reduction of noise are an active area of current research. The usage of adaptive filters is one of the most popular proposed solutions to reduce the signal corruption caused by predictable and unpredictable noise.
An adaptive filter has the property of self-modifying its frequency response to change the behavior in time, allowing the filter to adapt the response to the input signal characteristics change. Due to this capability the overall performance and the construction flexibility, the adaptive filters have been employed in many different applications, some of the most important are: telephonic echo cancellation, radar signal processing, navigation systems, communications channel equalization and biometrics signals processing.

The basic adaptive algorithms widely used for performing weight up gradation in an adaptive filter are: the LMS (Least Mean Square), NLMS (Normalized Least Mean Square) and the RLS (Recursive Least Square) algorithm.

Performances of all these algorithms depend upon the number of microphones available at the receiver unit. Typically, the larger the number of microphones, the easier the speech enhancement task becomes. For Adaptive cancellation at least one microphone is required near the noise source. The microphones need to be separated in order to prevent the clean speech signal from being included in the noise reference. Using two or more microphone inputs, coefficients of the adaptive filter are adaptively adjusted to remove the noise from the noisy signal.

All the real time applications of these algorithms require them to have a low computation and convergence time. In this paper, we have tried to simulate the real time environmental conditions by testing each of these algorithms against a set of 11 different noise samples with widely varying characteristics and then tried to evaluate their performance through a comparative analysis using MATLAB.

2. ADAPTIVE ALGORITHMS

The purpose of an adaptive filter in noise cancellation is to remove the noise from a signal adaptively to improve the signal to noise ratio. Figure 1 shows the diagram of a typical Adaptive Noise Cancellation (ANC) system.
The algorithm relies on two separate sensors, with one known as primary sensor (senses the mixed signal 'd(n)') and other as secondary or reference sensor (senses the background noise). Signal s(n) is generated by the source or speaker and is transmitted through the channel between source and the primary sensor, and $n_0(n)$ is the noise sensed at the primary sensor. The secondary sensor or

reference sensor senses background noise as $n_1(n)$ which is correlated to $n_0(n)$ in some sense but uncorrelated with s(n). The reference noise input $n_1(n)$ is filtered by the adaptive filter (by performing convolution with filter weights h(n)) to produce and adjust the output y(n) as close as possible to $n_0(n)$ for effective noise cancellation [8].

The filter output y(n) is then subtracted from d(n) to obtain an estimation error e(n). The objective here is to minimize the error signal e(n) by using it to incrementally adjust the filter's weights for the next time instant [8].

The estimation error signal is also known as de-noised signal or noise cancelled speech signal.

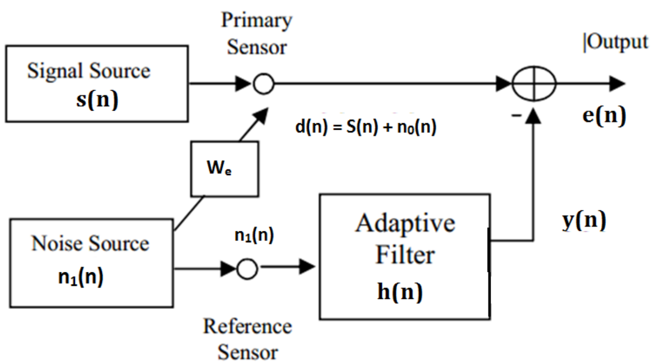


Fig 1: Adaptive noise cancellation setup

2.1 LMS ALGORITHM

The LMS uses the error signal to calculate the filter coefficients and hence one of the simplest algorithms [9].

The output y(n) of the adaptive filter is calculated from the equation (1)

$$y(n)=\sum_{m=0}^{N-1} w(m)\,x(n-m) \qquad \text{- Eq. (1)}$$

The estimation error is calculated by equation (2).

$$e(n)=d(n)-y(n) \qquad \text{- Eq. (2)}$$

The filter weights are updated as per the equation (3)

$$w(n+1)=w(n)+\mu e(n)x(n) \qquad \text{- Eq. (3)}$$

Where:

w(n) is the current weight value vector,
w(n+1) is the next weight value vector and
μ is the convergence factor which determines the convergence time of the filter.

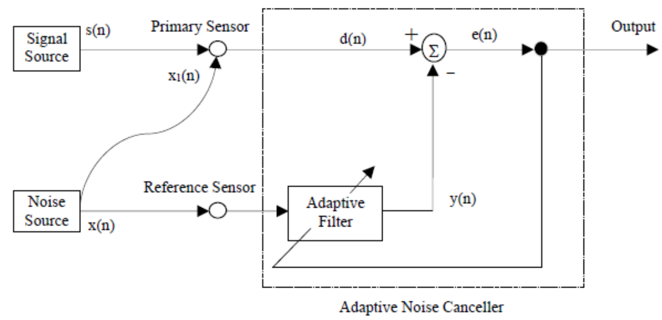


Fig. 2. Adaptive Noise Cancellation

2.2 NLMS ALGORITHM

The LMS algorithm experiences large convergence time for greater values of μ. NLMS offers to solve this problem by normalizing the weight vector w(n) at instant (n+1) with respect to the squared Euclidean norm of input vector x(n) at instant n. Thus, the step size in this algorithm varies with time [7].
The convergence factor is calculated as equation (4).

$$\mu(n)=\frac{\alpha}{c+\|x(n)\|^2} \qquad \text{- Eq. (4)}$$

Where:
α is the NLMS adaption constant aimed to optimize the convergence rate of the algorithm.
Preferably,

$0< \alpha <2$

C is the constant and is always less than 1.

The Filter weights are updated as shown in equation (5).

$$w(n+1)=w(n)+\frac{\alpha}{c+\|x(n)\|^2}e(n)x(n) \qquad \text{- Eq. (5)}$$

2.3 RLS algorithm

The RLS algorithm performs the best in time varying environments but at the cost of an increased computational complexity and some stability problems.
In this algorithm the filter tap weight vector is updated using equation (6).

$$w(n)=\overline{w}^T(n-1)+k(n)\overline{e}_{n-1}(n) \qquad \text{- Eq. (6)}$$

where, intermediate gain vectors used to compute tap weights are calculated as equation (7) & (8),

$$k(n)=u(n)/(\lambda+x^T(n)u(n)) \qquad \text{Eq. (7)}$$
$$u(n)=w_{\lambda}^{-1}(n-1)x(n) \qquad \text{Eq. (8)}$$

Where:
λ is a small positive constant < 1.
The filter output is calculated using the filter tap weights of previous iteration and the current input vector as shown by the

$$\overline{y}_{n-1}(n)=\overline{w}^T(n-1)x(n) \qquad \text{Eq. (9)}$$
$$\overline{e}_{n-1}(n)=d(n)-\overline{y}_{n-1}(n) \qquad \text{Eq. (10)}$$

2.4 LPC Algorithm

LPC algorithm views speech y(z) as an excitation signal e(z) filters with vocal track h(z). Where LPC coefficient can be expressed as p(z),the relationships between e(z), h(z), p(z), y(z) are shown in the figure (3).

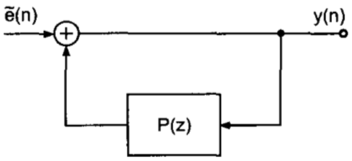


Fig 3. Excitation signal filter with LPC

Where p(z) can be written as:

$$P(z) = \sum_{k=1}^{p} a_k z^{-k}.$$

Eq. (11)

$$A(z) = 1 - P(z) = 1 - \sum_{k=1}^{p} a_k z^{-k}$$

Eq. (12)

$$E(z) = X(z)A(z).$$

Eq. (13)

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - P(z)}$$

Eq. (14)

$$Y(z) = \tilde{E}(z) \cdot H(z)$$

Eq. (15)

Here, we replace the excitation signal with our mixed signal, i.e. s(n)+x(n) (shown in Fig.2).
P(z),A(z), and H(z) are obtained from the clean speech s(n). The LPC algorithm was adopted with STFT synthesis and inverse STFT re-synthesis, thus the coefficients were updated frame by frame.

3. Experimental Results

To establish the fact that 11 sources of noise used as the database for evaluating the performance of our algorithms are highly diverse and highly uncorrelated, we analyzed their spectrograms (shown in Fig.4).

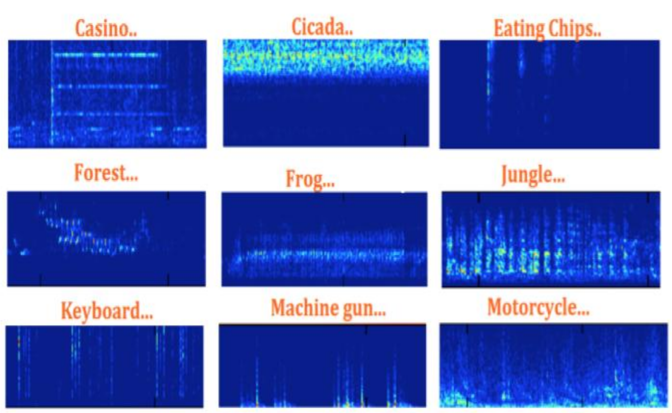


Fig.4. spectrogram of 9 experimented noise environment

Shown below are the performance results of all 4 algorithms against a mixed signal wherein ocean sound was used as the background noise for the clean speech signal s(n).

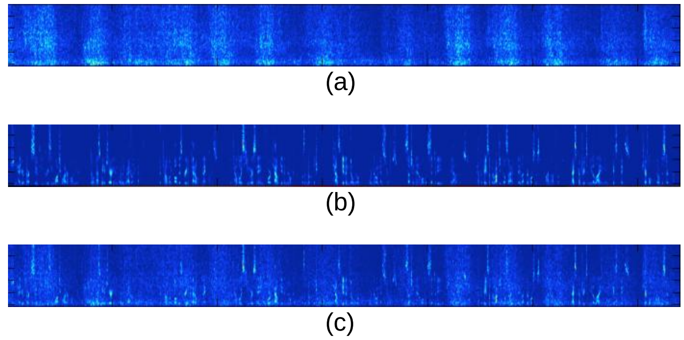


Fig.5. Spectrogram of three inputs (a) ocean noise (b) clean speech (c) mixed speech

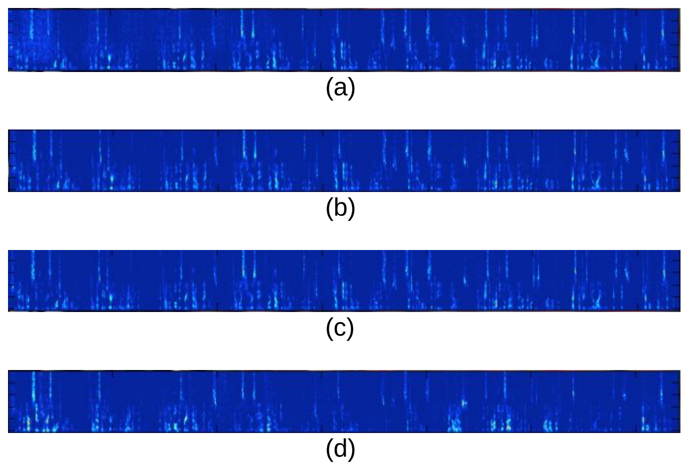


Fig.6. Spectrogram of noise reduced mixed speech signal (a) LMS (b) NLMS (c) RLS (d) LPC

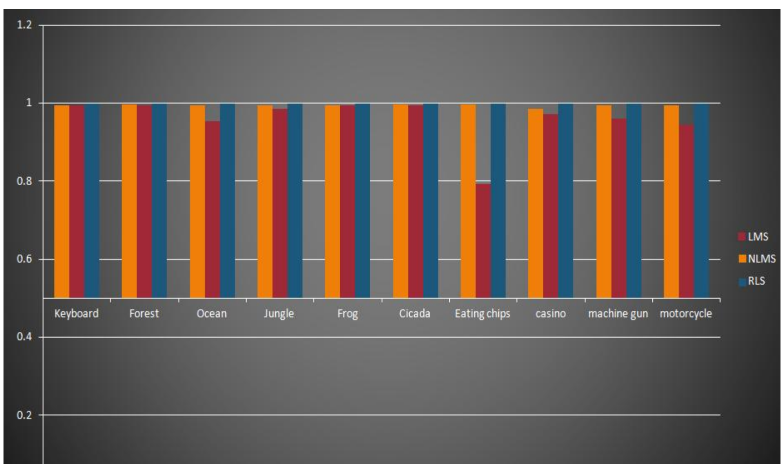
The comparison between run time performances of the 4 algorithms for the above mentioned mixed signal showed LPC to have the highest computation time of 9.693 sec followed by RLS at 3.087 sec, NLMS at 2.056 sec and LMS with lowest of 2.033 sec. Analysis of spectrograms of the noise reduced mixed signals (Fig.6) indicated another drawback of LPC i.e. though it was able reduce noise in silent portions of the speech, it distorted the perceived audio quality of the noise reduced mixed signal overall.

A thorough comparison between LMS, RLS and NLMS algorithms was made for 11 different sources of noise at different values of input SNR and the percentage of correlation between the noise cancelled signal and original clean speech signal was computed (Fig 7). As evident from the data below, RLS restored signal had the highest level of correlation with the original clean speech signal and hence proved to be the best noise cancellation technique.
The data for LPC hasn't been shown as it proved to be very inferior in performance and hence its performance statistics were not comparable to the rest 3 algorithms.

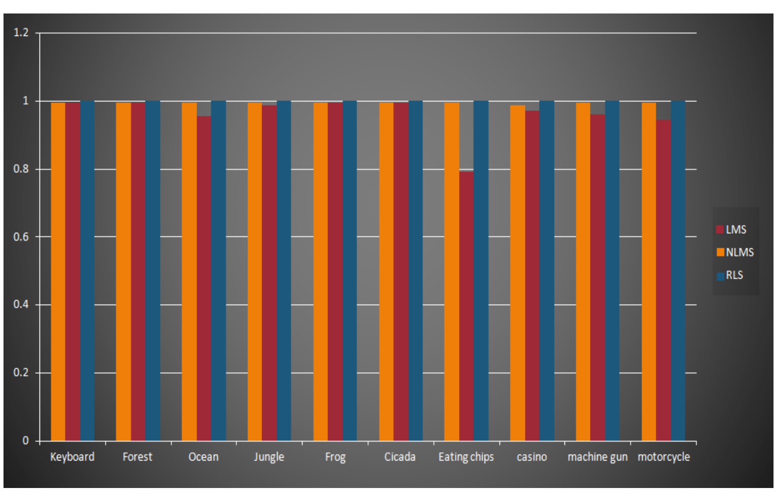
Noise type	Input SNR (dB)	LMS	NLMS	RLS
Keyboard	15	0.8726	0.9861	0.9992
	10	0.96	0.9890	0.9998
	5	0.9876	0.9799	0.9999
	0	0.9957	-28.21	1.000
Ocean	15	0.9152	0.9400	0.9967
	10	0.9639	0.9431	0.9989
	5	0.9859	0.9480	0.9996
	0	0.9943	0.9550	0.9999
Forest	15	0.8862	0.9935	0.9997
	10	0.9630	0.9951	0.9999
	5	0.9882	0.9954	1.0
	0	0.9961	0.9951	1.0
Jungle	15	0.8871	0.9805	0.9995
	10	0.9587	0.9846	0.9998
	5	0.9871	0.9865	0.9999
	0	0.9955	0.9861	1.0
Frog	15	0.8839	0.9925	0.9998
	10	0.9610	0.9950	0.9999
	5	0.9879	0.9955	1.0
	0	0.9951	0.9952	1.0
Cicada	15	0.8740	0.9938	1.0
	10	0.9587	0.9965	1.0
	5	0.9873	0.9967	1.0
	0	0.9960	0.9957	1.0
Eating chips	15	0.8694	0.9894	0.9994
	10	0.9524	0.9916	0.9998
	5	0.9856	0.9929	0.9999
	0	0.9961	0.7925	1.0
Casino	15	0.8967	0.9705	0.9994
	10	0.9600	0.9717	0.9998
	5	0.9867	0.9717	0.9999
	0	0.9948	0.9718	1.0
Machine gun	15	0.9242	0.9448	0.9977
	10	0.9762	0.9446	0.9994
	5	0.9903	0.9497	0.9998
	0	0.9948	0.9601	0.9999
Motorcycle	15	0.9139	0.9231	0.9955
	10	0.9668	0.9326	0.9986
	5	0.9862	0.9416	0.9996
	0	0.9939	0.9453	0.9999

Fig 7: As seen from table above, the restoration results for noise distorted mixed signal are the best for RLS.

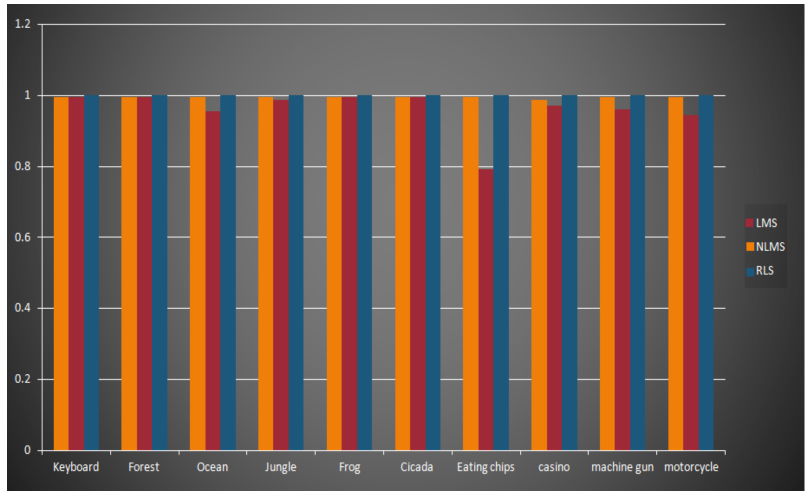
Case (a): Input SNR = 15 dB



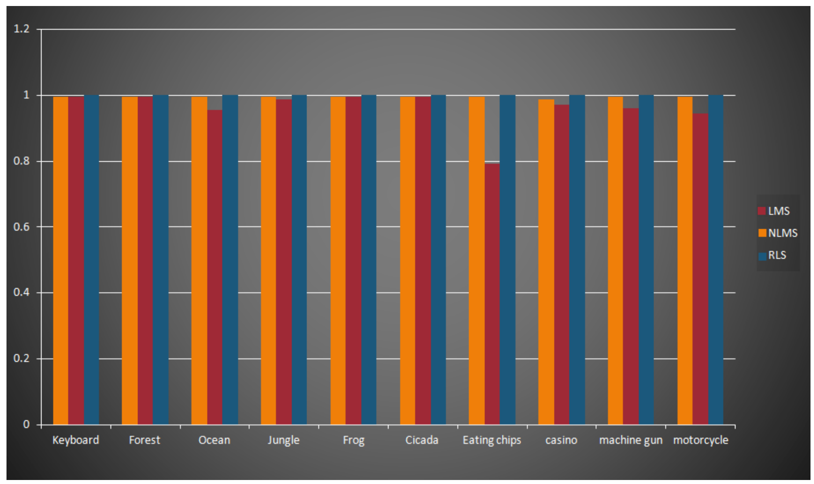
Case (b): Input SNR = 10 dB



Case (c): Input SNR = 5 dB



Case (d): Input SNR = 0 dB



4. Conclusion

We observed that for a particular noise source and algorithm, as the SNR decreases the perceived audio quality of the restored signal is better.

Further, based on the simulation results, we can deduce the following order of performance for the 4 algorithms,

RLS> NLMS > LMS> LPC

There is a slight deterioration in performance of LMS for near 0 SNR because of the estimation errors introduced by the noise PSD estimator. However both NLMS and LMS are the most frequently used algorithms for noise reduction and this can be attributed to their low complexity and robustness. Their performance wasn't remarkable for non-stationary environment (i.e. when the noise input exhibits widely varying characteristics with respect to time).The performance can be improved to some extent in such situations by choosing the step size properly.

In general, LMS suffered from slow convergence time.

The RLS on the other hand, demonstrated the best in non-stationary environments with high convergence time but at the cost of higher complexity.

The NLMS algorithm changes the step-size according to the energy of input signals hence it is suitable for both stationary as well as non-stationary environment and its performance lies between LMS and RLS. It provides a trade-off between convergence time and computational complexity.

LPC algorithm considers the noise reduction problem from a perceptual and intuitive perspective. However aside from its poor computational performance, the algorithm only cleans out the noise between silent intervals. However, the voiced moment gets distortion effect after filtering the speech from vocal track LPC filter.

5. Future work

It has to be noted that since all the algorithms implemented in this paper are basically adaptive in the sense that they need time to analyze noise characteristics in order to filter it out. Consequently they take a few milliseconds to converge before they actually remove the effect of noise from the mixed output signal. This convergence time can pose a serious limitation to these algorithms when the noise in the background is intermittent and has duration shorter than the convergence time of the algorithm.

Further, the performance of these algorithms differ with varying sampling rates and for the scenarios where the mixture signal contains more of noise and less of original clean speech signal which motivates for further research in this respect.

6. Acknowledgements

The authors would like to thank Professor Zhiyao Duan and also the teaching assistant Xuchen for their extended guidance throughout the research, analysis and documentation process of this paper.

7. References

[1]Pascal Scalart et al. " Speech Enhancement Based on a priori signal to noise estimation".

[2] Kris Hermus et al. "A Review of Signal Subspace Speech Enhancement and Its Application to Noise Robust Speech Recognition".

[3]Raj Kumar Thenua "Simulation And Performance Analysis Of Adaptive Filter In Noise Cancellation".

[4] J. Benesty et. al. , "Adaptive Filtering Algorithms for Stereophonic Acoustic Echo Cancellation".

[5] Slock, D.T.M "On the convergence behavior of the LMS and the normalized LMS algorithms".

[6] Sanaullah khan "Comparison of LMS, RLS and Notch Based Adaptive Algorithms for Noise Cancellation of a typical Industrial Workroom".

[7] Yuu Seng lau et. al. "Performance of Adaptive Filtering Algorithms: A Comparative Study".

[8] Bernard Widrow "Adaptive Noise Cancelling: Principles and Applications".

[9] Udo Zolzer. "Digital Audio Effects".

[10] James Brain Richardson. "LPC-Synthesis Mixture: A Low Computational Cost Speech Enhancement Algorithm".

[11] Philipos C. Loizou "Speech Enhancement Theory and Practice".