Impulsive Noise Cancellation for Speech Enhancement using State Space Adaptive Algorithm

Rashmirekha Ram\textsuperscript{1}, Sarthak Panda\textsuperscript{2}, Hemanta Kumar Palo\textsuperscript{3}, Mihir Narayana Mohanty\textsuperscript{4}

\textsuperscript{1, 2, 3, 4}Department of Electronics and Communication Engineering, Siksha ‘O’ Anusandhan University, Bhubaneswar, Odisha, India
\textsuperscript{1}ram.rashmirekha14@gmail.com, \textsuperscript{2}sarthak216@rediffmail.com, \textsuperscript{3}hemantapalo@soauniversity.ac.in, \textsuperscript{4}mihir.n.mohanty@gmail.com

Abstract— The occurrence of noise in almost all types of signals is natural. Though the noise variants are many, the impulsive noise in signal highly affects its quality. In this piece of work, speech signal is considered for enhancement that is contaminated with impulsive noise. Generally, hiccups create such type of noise due to tiredness or myoclonic problem of human subjects. Removal of this type of impulsive noise can enhance the speech signal and can be used in case of recognition, security and in the field of medicine. The popular recursive least mean square (RLS) algorithm has been used for this purpose. Also the state space variant of RLS (SSRLS) application enhances the result and can be used for real time applications. The result shows its performance in terms of signal to noise ratio (SNR) and the visualization of the speech signal.

Keywords— Speech enhancement; impulsive noise; recursive least mean square; state space variant; signal to noise ratio

I. INTRODUCTION

Speech is by nature inherits some form of noise even with most favored recording environment. The noises are in the form of white, colored, Gaussian, impulsive or other natured that makes the signal corrupted. This poses difficulty in recognition and processing of speech signal for effective analysis. Arguably, a suitable man-machine interface that can provide clean signal by minimizing these noises remains a challenge. Due to myoclonic jerk in the vocal tract diaphragm of human being hiccups occur involuntarily. This will act as impulsive noise (approximately of 0.25 seconds duration) and add to the speech signal as noise during conversation. These noisy signals may occur once or many times hence need to be removed for better signal analysis. For Gaussian noise, linear filtering can affect its power spectral density (PSD) though the probability density function (PDF) remains intact. Due to increase in power level of this noise with respect to the signal, the signal quality degrades slowly. However, impulsive noise such as hiccups resembles to a train of pulses having very wide PSD. The PDF of such noise changes on account of filtering due to un-correlated in-phase and quadrature components although these are dependent. The noise can jam the system even with considerable amount of signal to noise ratio (SNR) [1]. One such impulsive noise cancellation for enhancement of speech is treated in this piece of work using two efficient adaptive algorithm.

II. RELATED LITERATURE

Many speech enhancement algorithms have been used since several decades. Among them, spectral subtraction (SS), filtering, statistical-model-based and subspace algorithm are mostly discussed in literature [2-6]. In SS method of enhancement, the clean speech is obtained by subtracting the noise spectrum from the noisy speech signal. However, the presence of musical noise makes the method less efficient [3]. As speech is non-stationary in nature, use of adaptive algorithms will be more of an application based approach for obtaining the desired signal enhancement. Least mean square (LMS) is the simplest and easy to implement adaptive algorithm as compared to Normalized LMS (NLMS) and Recursive Least square (RLS) algorithms. Nevertheless, dependence of the input signals on chosen scale factor and instability problem associated with this method makes it unreliable. NLMS is faster with large PSNR than LMS where as quick convergence, robustness and dynamic adaptive framework makes RLS method more suitable for speech enhancement. The method is reliable particularly when the signal is corrupted with coloured or white noise although, it is computationally more complex [7].

Statistical-model-based methods have found their presence in this field with effective results. Among these, minimum-mean square error (MMSE) estimator, MMSE short-time spectral amplitude estimator (MMSE-STS) and its variants such as MMSE log-spectral amplitude estimator (MMSE-LSA), the optimally-modified log-spectral amplitude (OMLSA) estimator etc. are few algorithm discussed in literature [8-9]. Speech enhancement using subspace approach has been proposed as an efficient method to combat noise inherited in the signal [10-11]. The authors estimated the clean speech signal by removing the signal components from the noise subspace by projecting the source signal into two subspace as noisy (signal plus noise) and noise subspace. However, these authors assumed an orthogonal relationship between speech and noise subspace, which cannot be guaranteed always as speech is non-periodic in nature. Deng et. al. in their work proposed sparse auto regressive hidden Markov model (SARHMM) - based single-channel speech
The error between the desired signal and output signal. The steps of RLS algorithm is explained below

1. Calculate the adaptive filter output \( \hat{s}(n) \).
2. For a desired signal \( s(n) \), estimate the error signal using the relation
   \[
   g(n) = \hat{s}(n) - s(n)
   \]
3. The filter coefficients are updated using the relation
   \[
   \hat{v}(n + 1) = \hat{v}(n) + g(n)p(n)
   \]
   Where, \( \hat{v}(n) \) is the filter coefficient vector and \( p(n) \) is given by the equation [6]
   \[
   p(n) = \frac{M(n)\hat{w}(n)}{\beta + M(n)\hat{w}(n)M(n)\hat{w}(n)}
   \]
   where, \( \beta \) denotes the forgetting factor and \( M(n) \) represents the inverse correlation matrix of the signal input. Using an identity matrix \( I \) and \( \eta^{-1}I \), \( M(n) \) is initialized, where, \( \eta \) is the regularization coefficient.
   The value of \( M(n) \) can be determined by using the equation
   \[
   M^{-1}(n) = \beta^{-1}M^{-1}(n-1) - \beta^{-1}p(n)p^T(n)M^{-1}(n-1)
   \]

### B. State space RLS (SSRLS)

In this adaptive method, RLS is represented by the state space algorithm. The algorithm has been quite useful in case of aperiodic signal such as EEG and other deterministic signals when the signal is associated with impulsive noise [15-16]. Since speech is a non-stationary signal, the algorithm is expected to perform efficiently in such cases. As the speech signal used in this work is mixed up with hiccups that represent the impulsive signal, the algorithm has been used as a means to filter the noise. The algorithm is implemented using following steps.

Consider a noisy speech signal \( s'(n) \) having \( g(n) \) as the error signal and \( p(n) \) as the observer gain. For a predicted input signal \( \hat{n} \) and estimated state \( \hat{n} \), the predicted output \( \hat{s}(n) \) is given by

\[
y'(n) = Qs(n)
\]

Where,
\[
s'(n) = s'(n) + p(n)g(n)
\]
\[
s'(n) = Rs'(n-1)
\]
And the estimated error is given by
\[
g'(n) = y(n) - y'(n)
\]

In this case, \((R, Q)\) are described by the state transition matrix having eigen values that lies in the unit circle. Assuming \((R, Q)\) as \(j\)-step observable, the cross-correlation matrix \( M(n) \) of this algorithm is described by

\[
M(n) = \beta(R^TM(n-1))R^{-1} + Q^TQ
\]
This $M(n)$ can be used to estimate the observer gain $p(n)$ using the relation

$$p(n) = M^{-1}(n)Q^T$$  \hspace{1cm} (10)

IV. RESULTS AND DISCUSSION

The word “hello” has been used in addition of hiccups in between to generate the noisy speech signal. Hiccups generate the desired impulsive noise signal and need to be removed to obtain the clean speech using two of the adaptive algorithms as RLS and SSRLS. The plots using adaptive RLS algorithm to the noisy speech signal is shown in Figure 2.

A comparison on the power spectral density of the noisy signal before filtering and clean signal after adaptive filtering is shown in Figure 4 using RLS algorithm and SSRLS algorithm in Figure 5. The PSD has increased in both the cases as shown in these Figures.

Figure 2: Elimination of impulsive noise (hiccups) from speech signal using RLS algorithm

Figure 2 is having three plots as desired signal (subplot 1), the noisy speech signal (subplot 2) and the adaptive output (subplot 3). It is observed from subplots 2 and subplot 3, that the desired signal resembles to that of the adaptive output.

Figure 3 provides the waveforms of the desired, noisy and adaptive output speech signal using SSRLS algorithm. The resemblance of subplot 1 and subplot 3 in this Figure proves adequate enhancement of the noisy speech signal as shown in subplot 2.

Figure 3: Elimination of impulsive noise (hiccups) from speech signal using SSRLS algorithm

Table I gives a comparison of adaptive algorithms for noisy speech signal (speech signal and hiccups) before and after adaptive filtering is done. Before application of the adaptive methods, the PSD has been -23.5 as measured from the noisy signal having signal to noise ratio (SNR) of 8.5 dB. The change in PSD and SNR when different adaptive algorithms are applied is provided as a means of comparison of these values without filtering of the noisy signal.
On comparing the table, following conclusions can be drawn. The PSD has been reduced both in case of RLS and SSRLS due to removal of the noise components from the signals. However, the PSD is reduced more in case of SSRLS as compared to RLS indicates better speech enhancement using the later method. The range of reduction in PSD is about 13 unit using SSRLS as compared to only 3 units for RLS. It proves the superiority of SSRLS algorithm as compared to RLS method of adaptive filtering.

Similar results could be observed on comparing the SNR of these two methods. The SNR is increased to 22.5 using SSRLS as compared to 21.6 in case of RLS adaptive algorithm. The improved is more in case of SSRLS that indicates its versatility as compared to RLS method. However, the real-time computation factor of SSRLS is found to be more compared to that of RLS. Hence we can conclude that RLS method of speech enhancement is more faster as compared to SSRLS method in the proposed work.

V. CONCLUSION

Presence of hiccups makes the speech signal noisy due to inclusion of impulsive noise. Removal of these impulsive parts from the signal can infuse clarity in signal recognition and detection. This will help medical diagnosis, security organization including speech and signal processing application. Removal of other impulsive signals that occur naturally during speech signal recording can give a new direction in this field. Similarly, different enhancement methods and their comparison can provide a platform to improve the SNR by providing clean signal.

References


<table>
<thead>
<tr>
<th>Adaptive methods</th>
<th>PSD Before filtering</th>
<th>PSD After filtering</th>
<th>SNR(dB) Before filtering</th>
<th>SNR(dB) After filtering</th>
<th>Time elapsed</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLS</td>
<td>-23.5</td>
<td>-20.4</td>
<td>8.5</td>
<td>21.6</td>
<td>4.2 sec</td>
</tr>
<tr>
<td>SSRLS</td>
<td>-23.5</td>
<td>-10.5</td>
<td>8.5</td>
<td>22.5</td>
<td>6.5 sec</td>
</tr>
</tbody>
</table>