SIREN NOISE ATTENUATION BY NON-LINEAR PROCESSING OF TIME-FREQUENCY INFORMATION

HAMID SEPEHR*, MASOUD AHMADI ** PAUL BRENNAN*
*Electrical Engineering Department, UCL, London ** ElaraTek Ltd., London, UK

ABSTRACT

Siren signals are used to warn other cars and pedestrian about the presence of an emergency vehicle, this signal is leaked into the communication path between the emergency vehicle and the emergency call centre and substantially degrades the intelligibility of speech signal. In past, research has been mainly focused on siren cancellation on the source of siren signal by using extra microphones which requires complex acoustic and electrical consideration.

In this paper, two solutions and approach are presented for attenuation of the siren in an emergency call centre without having any knowledge about the emergency vehicle, type of siren and the presence of a siren.

Key Word: Siren reduction, Noise reduction, STFT, Wavelet, Notch Filter.

1. INTRODUCTION

In any emergency vehicle, during communication with the emergency call centre, the loud level of siren and alarm leaks into the communication path and reduce the intelligibility of communication. Figure 1. Shows such a scenario.

As the quality and intelligibility of speech is very important in an emergency call, research has been carried out to detect and remove different type of sirens while preserving the quality of the speech signal; such a system can be used to:

- Reduce the need for turning down or turning off the siren during communication to the control room and keep the traffic and pedestrians informed of the emergency vehicle.
- Increasing the quality of speech in hostile acoustic environments for fire brigades, police and health services.
- Improving the intelligibility and subjective quality of speech communication signal in a mobile radio application in an emergency condition; siren signal reduces the intelligibility of speech signal as well as adversely reduce the quality of low bit rate Codec used in an emergency communication system like TETRA (Terrestrial Truncated Radio).
- Potential integration of a VOX (Voice Operational Switch) system and siren canceller for a VOX operated digital mobile radio in an emergency vehicle.

Previous approaches and research for siren cancellation are mainly achieved by using an extra microphone mounted close to the source of siren signal or by direct access to the electrical siren signal at the output of siren generator circuit. Such a solution has the following pitfalls:

- Difficulty in installation of the siren canceller inside the emergency vehicle.
- Variety of emergency vehicles.
- Variety of different hardware for the radio systems.
- Ineffectiveness of the solution when a portable radio is used out of the vehicle.

If an effective siren reduction algorithm is deployed in the received side of the telecommunication link, following benefits can be achieved:

- Ease of implementation & integration into the existing call centre systems.
- It can also be provided as a standalone solution as a down-lead for each headset.
- Low cost & low maintenance and ease of upgradeability for any new sirens or any other signal processing algorithm in future.
The processed signal and original signal can be simultaneously accessed at the received end.

In following chapters, prior art and suggested solutions are explained in more details. Prior art for siren cancellation is explored in section 2. Our first solution and its mathematical background is presented in section 3 and finally in section 5, performance analysis and conclusion of our approach is presented.

2. PREVIOUS RESEARCH

There are different methods and approaches for siren canceller system. They can be classified into two main categories:

2.1. Siren removal inside the emergency vehicle.

In this method, the removal of siren noise is handled at the source. Usually two or more microphones are used to remove the unwanted siren noise from the communication path. One or more microphone are located closer to the siren or alarm generator and the output of this microphone is used as a reference in an adaptive system to remove the siren signal from the microphone which is close to the speaker.

Figure 2 shows an example of implementation of this approach with an extra microphone used as the reference signal [3], [4], [6]. There are also solutions based on multi-microphone using beam forming algorithm [1] and there are solutions that electrical siren signal generated by the siren generator is used as a siren reference signal [5].

2.2. Blind Siren Canceller:

There are other methods that blindly remove the siren signal from the communication path and they don’t need an extra reference signal.

The typical solutions proposed for this approach are:

- Wiener filters and spectral subtraction
- Adaptive siren cancellation by using the delayed contaminated signal as a siren reference signal.

This category of solution works well on specific siren types and they usually degrade the quality of speech while removing the siren signal. Figure (3) shows a model of such a solution that uses a delayed version of filtered contaminated speech signal as a reference signal, the filtering is done by using an anti-speech filter which is formed based on Yule–Walker Recursive filter, this filter is constructed according to characteristics of siren signal.

In addition of knowledge about the siren signal, according to author the approach doesn’t work well for all types of siren signal [3].

3. SIREN ATTENUATION BY PROCESSING OF VISUAL REPRESENTATION OF AUDIO SIGNAL

The speech signal is a quasi-stationary signal mixed with siren signal while siren is a narrow-band signal with varying main frequency by time.

To obtain information about frequency content of signal changing with time, the signal should be transformed from time domain into time-frequency plane. There are number of mathematical transformations that can be used for different application. In the siren attenuation scenario, as we are interested to detect the siren signals which are narrow-band signals in frequency domain, the short time Fourier transform (STFT) can be a good choice.

Short time Fourier transform is defined as equation (3.1).

\[ X_x(e^{j\omega}) = \sum_{n=-\infty}^{\infty} w(n-m) x(m)e^{-j\omega m} \]  

Equation (3.1) shows that \( w(n) \), the window, selectively determines the portion of \( x(n) \) which is being analyzed in frequency domain and \( X_x(e^{j\omega}) \) is the result of STFT transformation that are called STFT coefficients.

To have full reconstruction of the original signal with STFT coefficients, there are conditions for the windows to be met which are considered in this paper [8].

In our approach, the STFT coefficients are used to precisely locate the position of siren signal and then the STFT coefficients are modified in order to attenuate the siren signal. At final stage, the STFT modified coefficients are used to reconstruct the clean speech signal.

In Figure (4), the time-frequency representation of three common type of siren signal is shown.
Looking into Figure (4), the narrow-band frequency varying siren signal is appeared as a continuous contour in time-frequency visual representation of siren signal. If the location of contours can be identified and the corresponding STFT coefficients at continuous contour points can be attenuated, the reconstructed signal from the modified STFT coefficients is expected to be a clean speech signal.

The main challenge in this method is the identification of continues contour in the image formed by STFT coefficients. In our approach, this image is formed from magnitude of STFT coefficients and contour detection methods widely used in image processing are applied to the formed image to identify the time and frequency location of added siren signal.

Commonly in image processing, the canny edge detector is used for detection of edges [7]. In a canny edge detector, firstly, the image is blurred by a smoothing filter in order to reduce the effect of noise and then the gradient of smoothed filter is calculated. The magnitude and direction of gradient can be calculated as defined in equation (3.2) and (3.3) and the contour can be extracted with a simple threshold operation on the extracted gradient magnitude.

\[
M = \sqrt{(\frac{dl(t,f)}{dt})^2 + (\frac{df(t,f)}{df})^2} \tag{3.2}
\]

\[
\theta = \tan^{-1}\left(\frac{\frac{df(t,f)}{df}}{\frac{dl(t,f)}{dt}}\right) \tag{3.3}
\]

In Figure (5), a mixed of speech and siren signal and its time-frequency representation is shown.

By applying the canny edge detector on the image generated in duration of 500 millisecond (parameters of STFT in this algorithm were selected as follows: FFT length of 512, window size of 128 samples and 96 sample overlap in windowing) and attenuation of STFT coefficients at the location of edge detected by a canny edge detector, the siren signal is attenuated to a low-level signal which increases the quality and intelligibility of speech signal. The output result of suggested algorithm on noisy audio signal in figure (5) can be seen in following figure.

In Figure (6), a mixed of speech and siren signal and its time-frequency representation is shown.

To avoid detection of speech formants as an edge, pre-knowledge information of siren signal is used. For example, according to SAE (Society of Automotive Engineers) J1849 [10], siren signal in the US are above 650 Hz and therefore frequency content of the noisy signal below 650 Hz can be avoided for attenuation in a solution used in a contact centre in US.

The flow diagram of suggested algorithm is presented in figure (7) whereby the STFT coefficient are calculated and an image signal from magnitude of STFT coefficients is formed. Then, continues contours in the formed image are detected and accordingly the location of contours is attenuated in the STFT coefficients. At final stage, the modified STFT coefficients are synthesized to reconstruct the clean speech signal.

Although this solution provided promising results, the long processing delay and complexity of algorithm was not satisfactory to make this algorithm a practical solution, therefore an efficient method with lower complexity and delay was suggested that will be further investigated in next section.
4. EFFICIENT SIREN REDUCTION SOLUTION BY NOTCH FILTERING THE TIME FREQUENCY INFORMATION

The narrow-band characteristics of a siren signal imply that an adaptive notch filter is a suitable candidate for detection and removal of siren signals with unknown frequencies. Adaptive notch filters are widely used in communication systems, control process systems and signal detection applications. In our approach, a second order zero-pole adaptive notch filter is used; this is a lattice-based IIR filter with very low computation and a simplified robust adaptation algorithm [6]. Equation (4.1) shows the transfer function of such an adaptive filter whereby \( \rho \) controls the notch filter bandwidth.

\[
H(z) = \frac{1 - 2K[n]z^{-1} + z^{-2}}{1 - K[n](1 + \rho)z^{-1} + \rho z^{-2}} \quad K[n] = \cos(2\pi f_s)
\] (4.1)

This group of adaptive notch filter algorithms are robust with little complexity as their adaptation algorithms are not gradient based, which makes them a perfect choice for low complexity implementation in a DSP.

The input signal of an adaptive notch filter is a combination of speech and siren, and this could divert the adaptive notch filter to tune to formants of speech signal instead of narrow-band frequency of the siren signal, as a result, the speech quality could degrade significantly without even removing the siren i.e. worse than the original signal.

To overcome the divergence problem, a pre-processing algorithm needs to be applied to separate the speech and siren (noise) content before adaptive notch filtering is applied to the incoming signal.

The Wavelet transform has attracted tremendous interest since its discovery because of its excellent time-frequency localization and multi-resolution analysis (MRA). The time-frequency localization property of the wavelet transform was thought to help the notch filter to follow the unknown frequency of the added siren especially at the frequency bands with less presence of speech.

In this paper, a time-frequency decomposition algorithm followed by adaptive notch filtering algorithm is used to remove different siren signals from various frequency bands; experiments were carried out with different time-frequency representation algorithms to find the optimum transformation for siren reduction. Various wavelet based transforms including Daubechies, Symlets and Bi-Orthogonal were used. Also, STFT (Short Time Fourier Transform) and QMF (Quadrature Mirror Filter) were benchmarked in these experiments and some of the objective results of our comparison are reported in section 5.

Amongst different time-frequency algorithms, wavelet based algorithms provided the best signal representation for the notch filter, which consequently could remove the sirens. The bi-orthogonal wavelet of order 11 was identified to provide the optimum results.

The known key characteristics of both speech and sirens are utilized in order to detect siren signals precisely and consequently remove them with minimal or no degradation to the speech signal. The main pre-knowledge is the characteristics of the siren signal which is a set of moving frequency sinusoids in which the harmonics and frequency of the siren signal starts from 600 Hz [9].

In addition, no down-sampling was performed for the wavelet analyses since down-sampling reduced the performance of adaptive notch filter in the later stage.

To use the pre-knowledge information of the siren signal, two parallel wavelet analyses were employed, as figure 5 below illustrates. On the top path, a high-pass filter with cutoff frequency of 600 Hz [9] is used to identify the moving frequency of the added siren signal. The adaptive notch filters identify the notch frequencies in various frequency bands. This information is then passed to the parallel path that goes through a series of notch filter to remove the previously identified sirens (Figure 8 below). The notch filtered wavelet coefficients at these stages are then fed to the wavelet synthesis algorithm for reconstructing the clean speech signal.

Using notch filters at all times, could result in degradation of the speech signal when siren signal is not present; therefore an algorithm to detect the presence of the siren was critical in order to preserve the speech quality. This was achieved by measuring the NSR (Narrow-band to Speech Ratio) in the received audio signal within the frequency range of 600-1700 Hz (where most of siren signals' energy is present). If the NSR was over a pre-defined threshold, a flag would be generated which operates as a switch to enable/disable the fixed notch filters module (please see Figure 8 below).
5. RESULTS AND CONCLUSION

Two different objective analyses were performed on the algorithm in order to evaluate the quality of the processed audio output signals. In addition, for verification of the tests results, the audio output files were selectively listened.

Segmental SNR (SegSNR) and SNR improvement were measured for evaluation of the suggested siren attenuation algorithm; this was carried out for different signal/noise at different levels and with different siren types.

Segmental SNR is defined as the average of the SNR values over short segments of signal and defined as equation (5.1):

\[ \text{SNR}_{\text{seg}} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left[ \frac{\sum_{i=N_{m}}^{N_{m}+N-1} \left( \frac{x^2(i)}{x(i) - y(i)} \right)}{\sum_{i=N_{m}}^{N_{m}+N-1} \left( x(i) - y(i) \right)^2} \right] \]

where \( x(i) \) is the original audio signal and \( y(i) \) is the processed audio signal. The length of segments was selected to be 20 ms in our measurement. In order to avoid silence regions, the SegSNR is applied for frames with energy level above a specified threshold.

Signal-to-Noise Ratio (SNR), is a special case of SegSNR when \( M=1 \) in equation 5.1 and one segment includes the whole audio data.

Table 1 shows the average of SNR improvement in first suggested siren attenuation algorithm by processing of visual representation of audio signal. These results are obtained by running a test harness algorithm for two male and female speakers mixed with three different siren types (wail, yelp and Ho-Lo) at different SNR levels (6,0,-6,-12 and -18).

Table 2 shows the average segSNR improvement of our suggested siren attenuation algorithm for two male and female speakers in three different siren types (wail, yelp and Ho-Lo) and different SNR levels.

Table 9 shows the average of SNR and SegSNR improvement in the proposed siren attenuation algorithm for three different time-frequency decomposition algorithms. These results are obtained by testing the algorithms for two male and female speakers mixed with three different siren types (wail, yelp and Ho-Lo) at different SNR levels (6,0,-6,-12 and -18).
Our observations show that when the algorithm was applied to a noisy speech signal with a good SNR (input SNR > 6), the flag for the siren detection was disabled which consequently disabled the siren removal algorithm, therefore this solution is highly suitable for harsh environments with low level of SNR. However the algorithm performance declined in very low SNR (e.g. -18) which was mainly due to the siren saturation at the input stage. This reduces the capability of the notch filter to completely remove the siren noise.

6. REFERENCES


