A New Approach to Reducing Alarm Noise in Speech

Yilmaz Gül, Aladdin M. Ariyaeeinia and Oliver Dewhirst*

Fulcrum Voice Technologies, High Wycombe, Buckinghamshire, HP12 4HJ, UK
*University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, UK
[yilmaz.gul,oliver.dewhirst]@fulcrumvoicetech.com, a.m.ariyaeeinia@herts.ac.uk

Abstract

This paper presents a new single channel noise reduction method for suppressing periodic alarm noise in telephony speech. The presence of background alarm noise can significantly detract from the intelligibility of telephony speech received by emergency services, and in particular, by the fire brigade control rooms. The attraction of the proposed approach is that it targets the alarm noise without affecting the speech signal. This is achieved through discriminating the alarm noise by appropriately modelling the contaminated speech. The effectiveness of this method is confirmed experimentally using a set of real speech data collected by the Kent Fire Brigade HQ (UK).

1. Introduction

Operators working in the Fire Brigade Control rooms have to deal with numerous phone calls on a daily basis. Some of these, however, are heavily corrupted by the background alarm noise. In general, alarm systems are designed to be loud so as to ensure of their effectiveness. Frequently, the high-level tones produced by such systems are picked up by the phone used for communicating with the fire brigade. An inevitable consequence of this is a reduction in the intelligibility of telephony speech, which in turn may delay the response by the emergency services.

Additionally, the background alarm noise can cause undue disturbances to the hearing of the operators at emergency call centres. It has been shown that [1], [2] both a high acoustic level and unexpected noise can cause the small muscles within the inner ear to contract strongly. This muscle contraction can produce a sharp pain in some people and, in certain cases, cause face and neck pain. Moreover, this may result in hearing problems and associated medical problems such as nausea, giddiness and headaches.

The characteristics of the additive alarm noise vary considerably and depend highly on the manufacturer and the type of alarm system. Such additive noise could consist of a single or multiple frequency sinusoidal signals. It could be in the form of a constant tone or alternating tones. The contamination energy is also unpredictable and depends on a variety of factors. As a result, the conventional means of noise cancellation [3], [4] such as single-channel adaptive filtering, spectral subtraction and quantile methods may not be able to reduce this type of additive noise effectively without significantly degrading the speech signal. To tackle this problem, an alternative method is proposed and investigated in this paper. The proposed method can be viewed as processing the contaminated speech using a noise suppressing filter which is dynamically adapted according to the characteristics of the additive alarm.

The organisation of the paper is as follows. The next section provides a detailed description of the proposed approach. The experimental investigations and results are given in Section 3. The quality evaluation of the algorithm is presented in Section 4, and the overall conclusions are detailed in Section 5.

2. Alarm Noise Suppression

Alarm noise entering telephone systems are, in general, periodic sounds consisting mainly of sinusoidal signals. The frequencies of the alarms noise present in the telephone are limited by the bandwidth of the telephony system. The tones generated by fire alarms are grouped into four different categories. These are constant alarms which are the most common type; intermittent alarms occurring at regular intervals; alternate alarms which change repeatedly between two states; and sweeping (saw-tooth) alarms.

2.1 Proposed Method

Concentrating on constant alarms, which are the most commonly encountered type, it can be argued that an effective approach to removing this type of noise is that of spectral subtraction. This is regardless of whether the noise consists of a single or multiple sinusoidal signals. The important requirement in this case is that of obtaining a good estimate of the noise spectra. Theoretically, such an estimate can be obtained by averaging together several frame-wide spectra of the contaminated speech. In practice, however, the problems associated with the short-term Fourier transform-based spectra estimation are so significant that result in the failure of the approach. Examples of the said problems are the effects of windowing and the large variances of the estimates.

To tackle the above problem, the short-term spectral representation can be based on frame-wide linear predictive coefficients (LPC) obtained through the standard autoregressive signal modelling [5]. In this case, since the alarm noise is additive in both the time and frequency domains, the linear prediction vectors for the contaminated speech signal can be expressed as

$$\tilde{\mathbf{a}}_i = \mathbf{a}_i + \boldsymbol{\zeta}, \qquad (1)$$

where $\tilde{\mathbf{a}}_i$ is the i^{th} LP vector of the contaminated signal, $\mathbf{a}_i$ is the i^{th} LP vector of the speech part the received signal, and $\boldsymbol{\zeta}$ is the LP vector representing the continuous sinusoidal signal(s). The linear prediction analysis applied to the available contaminated speech yields $\tilde{\mathbf{a}}_i$. These vectors can then be decontaminated by removing their bias content as follows.

$$\hat{\mathbf{a}}_i = \tilde{\mathbf{a}}_i - \hat{\zeta} \qquad (2)$$

where $\hat{\zeta}$, representing an estimate of ζ, is the LP mean vector obtained using N contaminated vectors as

$$\hat{\zeta} = \frac{1}{N}\sum_i \tilde{\mathbf{a}}_i \qquad (3)$$

The above approach help suppress the representation of alarm signal in LP coefficients. In order to decontaminate the speech signal, it will need to be regenerated using the new LP coefficients. In general, the approach to this involves applying the LP residual signal to the input of a LP synthesis filter formed using the refined LP coefficients. A major difficulty with this general approach is that the residual signal is also contaminated with the additive sinusoidal signals. Unlike in the case for LP coefficients, the removal of the contaminating components in this case poses difficult and challenging problems. This difficulty, however, can be avoided by re-estimating the speech signal through applying the contaminated speech to a LP analysis system formed with the refined LP coefficients. In other words, for each given frame, the predicted speech samples are obtained as

$$\hat{S}_n = \sum_{k=1}^{p} \hat{a}_k S_{n-k} \qquad (4)$$

where $\hat{S}_n$ is the n^{th} sample of the re-estimated speech signal, S_{n-k} represent the past p samples of the contaminated speech, and $\hat{a}_k$'s are the refined predictor coefficients.

The expression in (4) is very similar to the standard linear prediction equation. However the use of the refined predictor coefficients means that each contaminated speech frame is processed by an FIR filter designed specifically to suppress the undesired sinusoidal components present in that particular frame.

3. Experiments

The speech data used for the purpose of this study was a subset of database collected by Kent Fire Brigade (HQ), UK. The complete database consists of 200 recordings of telephone calls made by different male and female callers from various locations. Each recording contains a conversation between the caller and the brigade operator, degraded by background alarm noise.

These analogue recordings were digitised by sampling them at 8 kHz and using 16 bits to represent each sample. The LPC parameters were obtained by segmenting the data using a Hamming window of length 25 ms at intervals of 12.5 ms.

The first part of the experiments involved the use of two telephone conversations between control room operators and members of the public reporting a fire. These were contaminated by continuous background alarms. The duration of each of these recordings and the nature of the contaminating noise signals are presented in Table 1. The estimation of the sinusoidal components (noise contamination) was based on a moving average of 20 LP vectors. The contaminated speech signal in Recording 1, and its enhanced version obtained using the proposed method are illustrated in Fig. 1 and Fig. 2 respectively. A visual inspection of these figures clearly confirm the extent of noise suppression offered by the method. The enhancement achieved in this case can also be observed by comparing the spectra of the noisy and refined speech presented in Fig. 3 and Fig. 4 respectively. As observed, the 2.5 kHz single sinusoidal present in the original speech has been effectively attenuated.

Speech Data	Embedded Alarm Noise Characteristics	Lengths [sec]
Recording 1	Constant tone at 2.5 kHz	30
Recording 2	Constant 2 tones at 1 & 3 kHz	25

Table 1: Characteristics of speech data used for test.

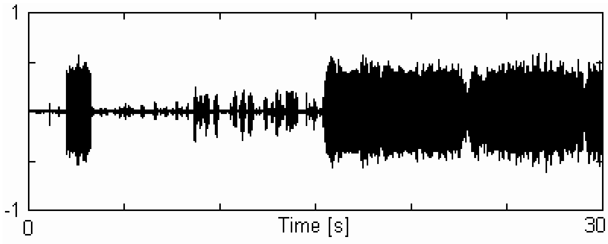

Figure 1: Contaminated speech in *Recording 1*.

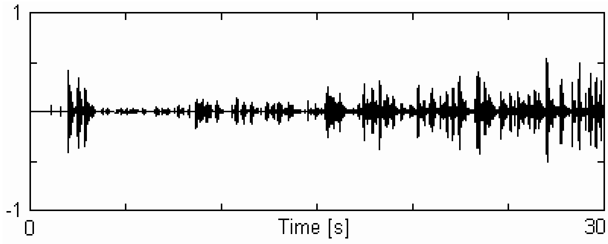

Figure 2: Enhancement achieved using the proposed method.

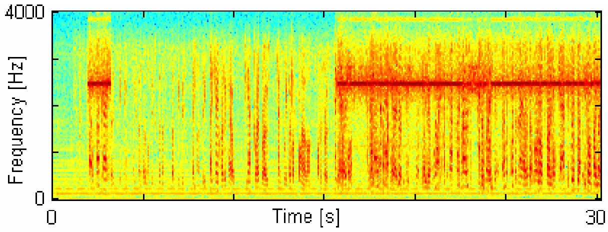

Figure 3: Spectrogram of the contaminated speech signal in *Recording 1*.

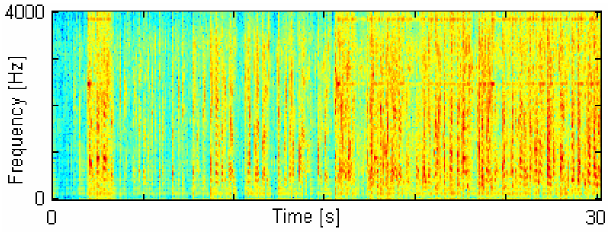

Figure 4: Spectrogram of the enhanced speech signal in *Recording 1*.

As indicated in Table 1, the contaminating part of the second recording consisted of two continuous sinusoidal signals of 1 and 3 kHz. Fig. 5 and Fig. 6 illustrate this signal before and after the refinement process respectively. The spectra of the degraded and enhanced speech in this case are given by Fig. 7 and Fig. 8 respectively. As shown in these figures, the proposed method is equally effective when the contaminating continuous alarm consists of a single or multiple sinusoidal signals.

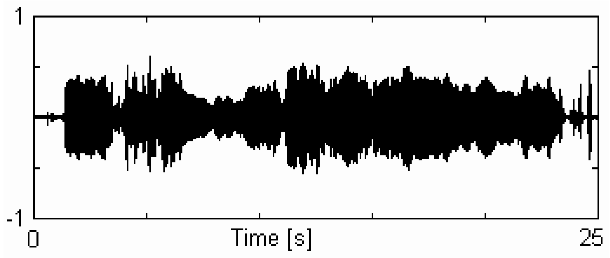

Figure 5: Contaminated speech signal in *Recording 2*.

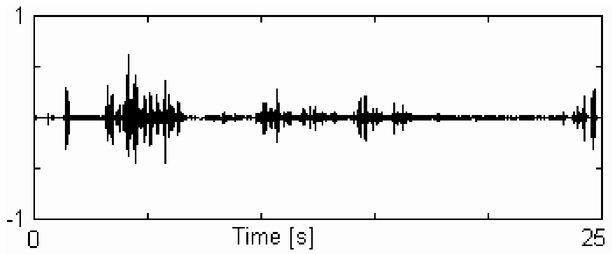

Figure 6: Enhancement achieved using the proposed method.

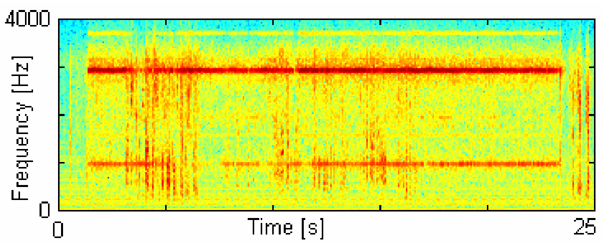

Figure 7: Spectrogram of noisy speech.

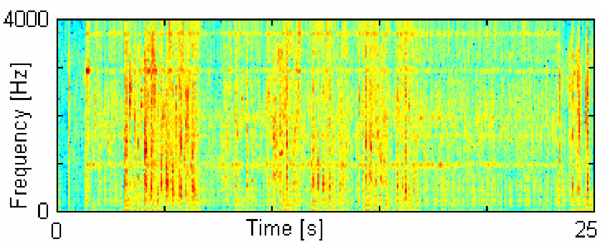

Figure 8: Spectrogram of decontaminated speech.

In order to obtain a measure of the relative performance of the proposed method, it was decided to compare its effectiveness with two other methods. These were the standard single channel adaptive filtering [6] and an approach currently used in a commercial system termed AFLU [7]. This experimental comparison was conducted using signals in Recording 1 and Recording 2 detailed earlier. The results are presented in Table 2.

Speech Data	**Tone**	**Alarm Reduction**		
		Adaptive Filtering	**Proposed Method**	**AFLU**
Recording 1	2.5 Hz	13 dB	34 dB	Cut-off
Recording 2	1 kHz	4 dB	16 dB	2 dB
	3 kHz	14 dB	27 dB	Cut-off

Table 2: Alarm reduction achieved using proposed method, ALFU and single-channel adaptive filtering.

As observed, the proposed method is considerably more effective than the standard single channel adaptive filtering method in attenuating both types of alarms. The method is also found to be more effective than AFLU in attenuating the 1 kHz sinusoidal component in Recording 2. On the other hand, AFLU is seen to completely remove the 2.5 and 3 kHz components. This is due to the fact that AFLU includes a low-pass filter with a cut-off frequency of 2.4 kHz. As a result, unlike the proposed method which aims to preserve speech, AFLU removes all speech components as well as contaminating noise with frequencies above 2.4 kHz.

The above experiments were concerned with the removal of the most common type of alarms consisting of continuous sinusoidal signals. However, as indicated in Section 2, there are other types of alarms which do not consist of continuous sinusoidal signals only. Examples are intermittent and alternate alarms. In order to remove such classes of alarms with equal effectiveness, it is essential that the start and end of each period of continuous contamination is determined a priori. Although this is forming part of further work, it was decided to examine the effectiveness of the current method for removing a range of alarm types. The results of this study are presented in Table 3. As expected, these result show that the method is most effective when the alarm is continuous. However, Table 3 also shows that the method is capable of attenuating other types of alarms as well, but not to the same extent. In this case, of course, the speech contents of the signals are also affected.

Alarm type	**Attenuation dB**
Continuous	20-35
Sweeping	6-14
Intermittent	6-12
Alternate	6-14

Table 3: Effectiveness of the proposed method in attenuating different alarm types.

4. Objective Quality Assessment

In order to conduct an objective quality assessment of the proposed method, two protocols described in [8] were adopted. These were Itakura-Saito (IS) measure and weighted spectral slope (WSS) measure. For this purpose a clean speech signal was employed and then contaminated in two different ways to generate two degraded signals. The contaminating noise in the first signal was a continuous single sinusoidal of 1 kHz. The contamination of the second signal involved the use of two sinusoidal signals of 1 and 2 kHz.
In order to obtain more meaningful results, it was decided to apply the same evaluation to AFLU and the single-channel adaptive filtering method. Table 4 shows the results of this evaluation. These results are in terms of the distances

(dissimilarity) between the original clean speech and each of the signals obtained by processing the degraded speech in different cases. As observed, in the cases of both distance measures, and for both degraded signals, the proposed method outperforms both AFLU and the adaptive filtering approach. In order to assess the level of enhancement achieved, Table 4 also gives the distance between the clean speech and degraded speech in each case.

Noise:1 kHz		
	IS	**WSS**
Degraded	3.17	123.5
AFLU	4.33	214.85
Proposed Method	1.19	57.13
Adaptive	1.61	71.95

Noise:2 kHz		
	IS	**WSS**
Degraded	3.83	257.9
AFLU	4.48	222.3
Proposed Method	0.96	55.23
Adaptive	2.86	186.8

Table 4: performance of the proposed method in objective quality tests.

5. Conclusion

An algorithm for the reduction of alarm noise in speech signals has been presented and evaluated. The main application of this algorithm is in removing alarm noise from telephone/voice communications in fire brigade call centres and control rooms. It has been argued that the removal of alarm noise not only improves the speech intelligibility, but also prevents personnel hearing loss and helps reduce the response time in emergency situations.

The speech enhancement in this method is based on using a sinusoidal suppressing filter developed by appropriately modifying the LPC parameters of the contaminated speech. The effectiveness of the method has been demonstrated using a set of experimental investigations. Moreover, through a set of objective quality assessments, it is shown that this method considerably improves the intelligibility of speech signals contaminated by high-energy background alarms.

To examine the practicality of the proposed method, it was also implemented in real-time using TMS320C6711 (floating point) digital processor. In this case, it was noted that only 6% of the processor power was used. This indicates that the algorithm has very low computation and can be suitable for implementation on hearing aids and telephone handsets.

6. Acknowledgements

This work was part of a collaborative project between Fulcrum Voice Technologies and the University of Hertfordshire. The project was funded by the British Government Department of Trade and Industry through the TCS programme 3092.

7. References

[1] Lower, M.C. and Davis, R.A., "Consultancy Report: Sound levels from headsets in the Control Room at Kent Fire Brigade HQ", *University of Southampton., ISVR Consultancy Services.*, Southampton., May 2001.

[2] Dillon, H. and Fisher, M., *National Acoustic Laboratories CRC for Cochlear Implant and Hearing Aid Innovation*, ISA, Melbourne, March 2002.

[3] Widrow, B. and Goodlin, R.C et al., "Adaptive Noise Cancelling: Principles and Applications"., *Proc. IEEE,* vol. 63, Pages(s): 1692-1716, Dec. 1975.

[4] Stahl, V. Fischer, A. and Bippus, R., "Quantile based noise estimation for spectral subtraction and Wiener filtering"., *IEEE., Proc. Acoustic, Speech and Signal Processing,* Vol 1., Page(s): 1875-1878.

[5] Markel, J.D. and Gray, A.H., *Linear Prediction of Speech,* Springer-Verlag Publishers, ISBN 0-387-07563-1, New York, 1976.

[6] Gahng, H.D. and Bae, K.S., "Single channel adaptive noise cancellation for enhancing noisy speech", *Proc. ISSIPNN* '94., Int. Symposium, Vol 1, Page(s): 343-346, 13-16 Apr 1994.

[7] Fulcrum Voice Technologies., Alarm Frequency Limitation Unit (AFLU), http://www.fulcrumvoicetech.com/aflu.html.

[8] Hansel, J.H.L. and Pellom, B., "An Effective Quality Evaluation Protocol for Speech enhancement algorithms", *ICSLP-98 Inter. Conf. On Spoken Language Processing,* Vol. 7, Page(s):2819-2822, Sydney, Australia, Dec. 1998.