NOISE REDUCTION FOR DRIVER-TO-PIT-CREW COMMUNICATION IN MOTOR RACING

Mark Hadley, Ben Milner and Richard Harvey

School of Computing Sciences, University of East Anglia, Norwich, UK

{mlh@cmp.uea.ac.uk, bpm@cmp.uea.ac.uk, rwh@cmp.uea.ac.uk}

ABSTRACT

This work proposes a noise reduction system for use in high noise environments such as motor racing. First, an analysis of the audio data received from the driver reveals that the main noise sources are from the engine, airflow and tyres. These are found to relate to engine speed, road speed and throttle position information that is received in a data stream from the car's on-board computer. A twostage noise compensation strategy is proposed which first suppresses engine harmonics using adaptive filtering with engine speed reference information taken from the data stream. Second, a maximum a posteriori (MAP) prediction of the tyre and airflow noise is made from data stream values and this is combined with spectral subtraction for noise suppression. Human listening tests reveal that both noise reduction stages lead to good improvements in the intelligibility of the speech with a comparative mean opinion score (CMOS) of +1.57 being obtained.

1. INTRODUCTION

The aim of this work is to reduce the amount of noise present on speech that exists during radio driver-to-pit-crew communications in motor racing applications. Many of the noise sources encountered in motor racing are similar to those found in conventional road cars but the higher speeds and more powerful engines cause the resulting signal-to-noise ratios (SNRs) to be as low as -20dB in some instances. This is considerably lower than that found in road cars where SNRs may be down to -5dB. Work has been reported on speech enhancement in car environments, where techniques using spectral subtraction or Wiener filtering have been successful [6, 3, 5, 2, 1]. Typical noises encountered are from sources such as the engine, tyres, airflow, radio, air-ducts and indicators. These noise compensation methods typically make estimates of the contaminating noise during non-speech periods and then use the estimates to suppress the noise.

In motor racing applications fewer noise sources are present but those that do exist tend to be powerful. The three dominant noise sources in motor racing environments come from the engine, tyres and air flowing past the car. However, in motor racing, detailed information is available from the car's on-board computer which provides a data stream that gives information on road speed, engine speed and throttle position. This paper demonstrates this information is related to the primary contaminating noises and shows how this explicit information can be used to aid noise reduction. An analysis is made in section 2 of this paper into the contaminating noises and how they relate to the data stream information for the purpose of noise modelling.

This leads to a two-stage noise reduction strategy. The first stage aims to remove engine noise using adaptive filtering and is discussed in section 3. The second stage removes tyre and airflow noise and is discussed in section 4. Experimental results are presented in section 5 which demonstrates the effectiveness of noise reduction through a series of human listening tests.

2. ANALYSIS OF NOISE SOURCES

In addition to the driver's speech, the car generates a data stream which comprises road speed (measured in kmh⁻¹), r_i , engine speed (measured in rpm), e_i , and throttle position (measured as a percentage of full throttle), t_i . These values are updated every 10ms and can be considered as a sequence of vectors, $\mathbf{d}_i = [r_i, e_i, t_i]$. To illustrate the relationship between the received audio signal and data stream parameters, Fig 1(a) shows the spectrogram of a 10s frame of audio while the car accelerates from 100 kmh⁻¹ to 300 kmh⁻¹. Accompanying this, Figs 1(b) to 1(d) show the corresponding engine speed, throttle position and road speed. Two dominant noises can be observed in the spectrogram – the narrow bands of high energy corresponding to the engine harmonics and the low frequency energy which comes from the tyres and airflow. The graphs of engine harmonics and the engine speed data parameter have similar shape. Most noticeable in Figs 1(a) and 1(b) are the sudden drops in frequency of the engine harmonics when a gear change occurs. Also noticeable is the increase in energy of these harmonics as throttle is increased. The low frequency tyre and airflow noise shows relation with the road speed. For example an increase in low frequency energy can be observed from 1.2s onwards in Fig 1(a) which relates to the increase in road speed in Fig 1(d), also reported in [7].

A simple model of the received noisy speech can therefore be made as:

$$x(n) = s(n) + d_E(n) + d_{TA}(n)$$
(1)

where x(n) is the received noisy speech signal at sample n, s(n) is the original clean speech and $d_E(n)$ and $d_{TA}(n)$ are the noise components from the engine, tyres and airflow.

Engine noise can be modelled as a function, g_E , of engine speed and throttle position:

$$d_E(n) \approx g_E(e_i, t_i) \tag{2}$$

It is more appropriate to consider this function in the magnitude spectral domain, $|D_E(f)|$, where engine noise can be modelled as a series of impulses:

$$|D_E(f)| \approx \sum_{m=1}^{2M-1} h_m \delta\left(f - \frac{mf_0}{2}\right) \tag{3}$$

where h_m is the amplitude of the m^{th} engine sub-harmonic, M is the number of harmonics in the spectrum and f_0 is the fundamental frequency of engine rotation.



Fig. 1. a) Spectogram of 10 seconds of audio recorded under acceleration, b), c) and d) are the corresponding engine speed, throttle position and road speed parameters.

The tyre and airflow noise can similarly be modelled by the following function, g_{TA} of the road speed,

$$d_{TA}(n) \approx g_{TA}(r_i) \tag{4}$$

Based on this breakdown of the contaminating noise into a harmonically structured engine component and a wideband tyre and airflow noise component, noise reduction is implemented as a two-stage process.

3. ENGINE NOISE REDUCTION

The first stage of noise reduction is the cancellation of the rapidly time-varying engine harmonics. This is achieved through the use of a multiple adaptive notch filter using the LMS algorithm [4]. The restored speech signal at the output of this first stage of noise reduction, $\hat{s}_1(n)$, is given as,

$$\hat{s}_1(n) = x(n) - \sum_{m=1}^M w(n)_{m,1} \cos(\phi_m(n)) + w(n)_{m,2} \sin(\phi_m(n))$$
(5)

Since the noise ensemble is known, the filter can be configured in a spatial configuration as discussed in [10] and illustrated in Fig 2. The reference signal to the adaptive filter comprises a series of cosine and sine inputs, $\cos(\phi_m(n))$ and $\sin(\phi_m(n))$, where *m* indicates the harmonic number. The phase $\phi_m(n)$ is derived from the normalised angular engine speed $\Omega(n)$ and is computed as $\phi_m(n+1) = \phi_m(n) + m\Omega(n+1)$. The normalised angular engine speed, $\Omega(n)$, is computed from the engine speed data parameter, e_i . First the engine speed parameter is up-sampled to the sampling rate of the audio, f_s , to give $e_{up}(n)$, and the normalised angular engine speed is computed as, $\Omega(n) = \frac{2\pi e_{up}(n)}{60f_s}$. The filter coefficients, $w_{m,k}(n)$, are updated using the LMS algorithm which is described as

$$w(n+1)_{m,k} = w(n)_{m,k} + \alpha \hat{s}_1(n) x(n)$$
(6)

Where α is the convergence factor. For a filter with a fixed number of notches, the number of engine harmonics that can be filtered depends on the maximum possible engine speed, e_{\max} , and the sampling frequency, f_s and given as number of notches $= 2 \frac{f_s}{e_{\max}/60} - 1$. Doubling the number of integer engine harmonics gives the number of half harmonics including the sub-harmonic at half the frequency of the fundamental. Since a notch is not required at $\frac{f_0}{2}$ the total number of notches is reduced by 1. Good results were obtained when $f_s = 8$ kHz with 52 notches.

4. TYRE AND AIRFLOW NOISE REDUCTION

Tyre and airflow noise reduction is implemented by first making an estimate of the magnitude spectrum of the noise and then applying spectral subtraction to suppress the noise. In most spectral subtraction implementations a voice activity detector (VAD) is employed to identify non-speech periods from which to estimate the noise. However the high level of noise power encountered makes the use of a VAD unrealistic. Instead, the relation observed between road speed



Fig. 2. Configuration of LMS adaptive notch filter applied to engine noise cancellation

and the tyre and airflow noise is exploited to enable a maximum a posteriori (MAP) prediction of the noise spectrum to be made from a road speed measurement.

4.1. Noise prediction

Prediction of the tyre and airflow noise is achieved using a Gaussian mixture model (GMM) which models the joint density of the tyre and airflow noise magnitude spectrum and the road speed. The GMM is trained using a set of augmented feature vectors, \mathbf{z}_i , defined as,

$$\mathbf{z}_i = [\mathbf{D}_{TA_i}, r_i] \tag{7}$$

Vector \mathbf{D}_{TA_i} represents the magnitude spectrum of the i^{th} frame of audio and comprises a set of N = 128 magnitude spectral bins, given as $\mathbf{D}_{TA} = [|\mathbf{D}_{TA_i}(1)|, ..., |\mathbf{D}_{TA_i}(f)|, ..., |\mathbf{D}_{TA_i}(N)|]$ and r_i is the corresponding road speed measurement taken from the data stream.

From a set of training data vectors that have passed through the engine noise removal stage, unsupervised clustering is implemented using the expectation-maximisation (EM) algorithm to produce a GMM comprising a set of K clusters,

$$p(\mathbf{z}) = \sum_{k=1}^{K} \beta_k f(\mathbf{z}, \boldsymbol{\mu}_k^z, \boldsymbol{\Sigma}_k^z)$$
(8)

Associated with each cluster is a prior probability β_k which reflects the proportion of training data vectors allocated to the k^{th} cluster. Within the GMM, each of the K clusters is represented by a probability density function (PDF), with mean vector and covariance matrix given as:

$$\boldsymbol{\mu}_{k}^{z} = \begin{bmatrix} \boldsymbol{\mu}_{k}^{\mathbf{D}_{TA}} \\ \boldsymbol{\mu}_{k}^{r} \end{bmatrix} \quad \boldsymbol{\Sigma}_{k}^{z} = \begin{bmatrix} \boldsymbol{\Sigma}_{k}^{\mathbf{D}_{TA},\mathbf{D}_{TA}} & \boldsymbol{\Sigma}_{k}^{\mathbf{D}_{TA},r} \\ \boldsymbol{\Sigma}_{k}^{r,\mathbf{D}_{TA}} & \boldsymbol{\Sigma}_{k}^{r,r} \end{bmatrix} \quad (9)$$

Prediction of the noise magnitude spectrum, D_{TA} , given the input road speed, r_i , is made from the closest cluster, k^* , to the input road speed, r_i , given as:

$$k^* = \underset{k}{\operatorname{argmax}} \quad \{ p(r_i | \Psi_k^r) \beta_k \}$$
(10)

where $p(r_i|\Psi_k^r)$ is the marginal distribution of road speed for the k^{th} cluster, Ψ_k , in the GMM.

Using the joint density of road speed and noise magnitude spectrum from the cluster k^* , together with the current road speed, r_i , a maximum a posteriori (MAP) prediction of the noise magnitude spectrum, $\hat{\mathbf{D}}_{TA_i}$, can be made,

$$\mathbf{D}_{TA_i} = \operatorname*{argmax}_{\mathbf{D}_{TA}} \{ p(\mathbf{D}_{TA} | r_i, \Psi_{k^*}) \}$$
(11)

This leads to the prediction of the noise magnitude spectrum from cluster k^* as,

$$\hat{\mathbf{D}}_{TA_{i}} = \boldsymbol{\mu}_{k^{*}}^{\mathbf{D}_{TA}} + \boldsymbol{\Sigma}_{k^{*}}^{\mathbf{D}_{TA},r} \left(\boldsymbol{\Sigma}_{k^{*}}^{r,r}\right)^{-1} \left(r_{i} - \boldsymbol{\mu}_{k^{*}}^{r}\right)$$
(12)

4.2. Spectral subtraction

Using the predicted tyre and airflow noise magnitude spectrum, spectral subtraction is applied to produce a clean magnitude spectral estimate of the speech, $|\hat{S}_2(f)|$, from the magnitude spectral output of the engine noise removal stage, $|\hat{S}_1(f)|$,

$$|\hat{S}_2(f)| = |\hat{S}_1(f)| - \gamma |\hat{D}_{TA}(f)|$$
(13)



Fig. 3. a) Spectogram of 4 seconds of audio before processing, (b) Spectogram of the audio after noise reduction.

where γ is the over-subtraction factor. Post-processing is then applied to reduce musical noise distortions by placing constraints on the minimum permissible duration of isolated tones in the output signal [9]. The clean speech magnitude estimate, $|\hat{S}_2(f)|$, is then combined with the original noisy phase, $\angle X(f)$, to give the complex frequency-domain estimate of the clean speech, $\hat{S}_2(f)$. This is converted back into a time-domain frame of speech using an inverse Fourier transform. Overlap-and-add is then applied to successive frames of speech to produce the restored speech signal, $\hat{s}_2(n)$.

5. RESULTS

This section analyses the effectiveness of the two noise reduction stages in improving the intelligibility of the speech. First, spectrogram analysis is made which compares the audio signal before and after processing. Secondly, the results of a series of human listening tests are presented.

5.1. Spectogram analysis

Figure 3(a) shows a 4s duration spectrogram of noisy audio taken while the car is accelerating. Figure 3(b) shows a spectrogram of the same audio signal but after the application of the 2-stage noise reduction system. Labels are shown on both spectrograms to indicate where speech events occur in the audio. Figure 3(b) shows clearly that most of the engine harmonics have been removed by the adaptive filtering although some residual engine harmonic noise is present at low frequencies. Figure 3(b) also shows that spectral subtraction has removed most of the low frequency tyre and airflow noise. However as a consequence of spectral subtraction, a number of short duration tones have been introduced into the audio. Their occurrence is governed by the amount of over-subtraction and application of post-processing. Experimentation has determined that a trade-off exists between the amount of noise surpression possible against the amount of musical noise introduced. In practice it is simple for a listener to adjust the musical noise for maximum intelligibility.

5.2. Human listening tests

Many studies into speech enhancement measure their success on SNR comparisons before and after noise reduction. In this work it was found that the very low SNRs encountered made such analysis less conclusive. Instead the success of the noise compensation methods used in this work have been evaluated using a series of human listening tests. These subjective listening tests were conducted in accordance to ITU guidelines [8], using a Comparative Mean Opinion Score (CMOS). The CMOS scale is used to compare speech quality before and after noise reduction. Ratings are made on a scale of -3 to +3, where -3 indicates that the quality of the second sample played is much worse than the first, +3 indicates quality of the second sample is much improved and 0 indicates no discernible difference. A total of 30 listeners were employed in the tests and subjects were instructed to rate the samples according to speech intelligibility. In each test, the listener was played 20 pairs of audio files and was asked to rate the second audio sample in relation to the first sample heard. The pairs of audio files arranged so that comparisons could be made between

- Unfiltered vs. engine noise removal
- Unfiltered vs. engine noise removal and airflow/tyre noise removal

In each listening test, two different sets of these conditions were played to the listener, with pairs selected at different SNRs. The pairs of audio files were also arranged so that the unfiltered audio occurred first in half of the pairs and second in the remaining pairs. Table 1 shows the CMOS obtained from the 30 listeners when comparing the original audio with that after engine noise removal and also when comparing the original audio with that after both engine noise and tyre and airflow noise removal. The results show that a significant improvement in the audio was obtained after the application of adaptive filtering to remove the engine noise. A further increase in audio quality was then heard when applying spectral subtraction to remove tyre and airflow noise ¹.

Filtering	CMOS score
Unfiltered vs. Adaptive filtering	1.32
Unfiltered vs. Adaptive filtering and spectral subtraction	1.57

 Table 1. Results of the listening tests using the Comparative Mean
 Opinion Score(CMOS)

6. CONCLUSION

This work has examined the characteristics of noise present in the very high noise environment of motor racing and established that relation exists between the various noise sources and data stream parameters measured by the car's on-board computer. A two-stage noise reduction system has been developed which first removes engine harmonics using adaptive filtering by forming a reference signal directly from the engine speed data parameter. Secondly, a GMM has been employed to model the joint density of tyre and airflow noise and road speed, which enables a MAP prediction of noise to be made from road speed measurements. A series of human listening tests then confirms the success of both stages of noise reduction, giving an overall CMOS of +1.57 in comparison to the unfiltered audio.

7. REFERENCES

- P.Gomez A. Alvarez, V. Nieto and R. Martinez. Speech enhancement for a car environment using lp residual signal and spectral subtraction. *EUROSPEECH*, pages 1373–1376, 2003.
- [2] S. Vaseghi A. Chen and P. McCourt. State based sub-band lp wiener filters for speech enhancement in car environments. *Proc. ICASSP'00*, pages 213–216, June 2000.
- [3] S. F. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust., Speech, Signal Processing*, 27:113–120, April 1979.
- [4] John R. Glover. Adaptive noise canceling applied to sinusoidal interferences. *IEEE Trans. Acoust., Speech, Signal Processing*, 25:484–491, December 1977.
- [5] J.Boudy I. Lecomte, M.Lever and A. Tassy. Car noise processing for speech input. *Proc. ICASSP'89*, pages 512–515, May 1989.
- [6] Joerg Meyer and Klaus Uwe Simmer. Mult-channel speech enhancement in a car environment using wiener filtering and spectral subtraction. *Proc. ICASSP*'97, pages 1167–1170, April 1997.
- [7] Henning Puder and Frank Steffens. Improved noise reduction for hands-free car phones utilizing information on vehicle and engine speeds. Proc. EUSIPCO - tenth European Conference on Signal Processing, 3:1851–1854, 2000.
- [8] International Telecommunication Union. Methods for subjective determination of transmission quality. P.800, October 1996.
- [9] Saeed V. Vaseghi. Advanced Digital Signal Processing and Noise Reduction. Wiley, second edition, 2000.
- [10] Bernard Widrow and Samuel D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, 1985.

¹Note: audio samples of the results discussed here are available from http://www.cmp.uea.ac.uk/~bpm/ICASSP2006/motorracing/motorracing.htm