Least Squares Filtering of Speech Signals for Robust ASR

Vivek Tyagi[1,2] and Christian Wellekens[1,2]

[1] Institute Eurecom, Sophia-Antipolis, France.
[2] Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland
vivek.tyagi@eurecom.fr, christian.welleken@eurecom.fr

Abstract. The behavior of the least squares filter (LeSF) is analyzed for a class of non-stationary signals that are composed of multiple sinusoids whose frequencies, phases and the amplitudes may vary from block to block and which are embedded in white noise. Analytic expressions for the weights and the output of the LeSF are derived as a function of the block length and the signal SNR computed over the corresponding block. Recognizing that such a sinusoidal model is a valid approximation to the speech signals, we have used LeSF filter estimated on each block to enhance the speech signals embedded in white noise. Automatic speech recognition (ASR) experiments on a connected numbers task, OGI Numbers95[20] show that the proposed LeSF based features yield an increase in speech recognition performance in various non-stationary noise conditions when compared directly to the un-enhanced speech and noise robust JRASTA-PLP features.

1 Introduction

Speech enhancement, amongst other signal de-noising techniques, has been a topic of great interest for past several decades. The importance of such techniques in speech coding and automatic speech recognition systems can only be understated. Towards this end, adaptive filtering techniques have been shown to be quite effective in various signal de-noising applications. Some representative examples are echo cancellation[9], data equalization[10–12], narrow-band signal enhancement[8, 13], beamforming[14, 15], spectral estimation[3], radar clutter rejection, system identification[16] and speech processing[8].

Most of the above mentioned representative examples require an explicit external noise reference to remove additive noise from the desired signal as discussed in [8]. In situations where an external noise reference for the additive noise is not available, the interfering noise may be suppressed using a Wiener linear prediction filter (for stationary input signal and stationary noise) if there is a significant difference in the bandwidth of the signal and the additive noise [8, 3, 2]. One of the earliest use of the least mean square filtering for speech enhancement is due to Sambur[5]. In his work, the step size of the LMS filter was chosen to be one percent of the reciprocal of the largest eigenvalue of the correlation matrix of the first voiced frame. However, speech being a non-stationary

signal, the estimation of the step size based on the correlation matrix of just single frame of the speech signal, may lead to divergence of the LMS filter output. Nevertheless, the exposition in [5] helped to illustrate the efficacy of the LMS algorithm for enhancing naturally occurring signals such as speech. In [3], Zeidler et. al. have analyzed the steady state behavior of the adaptive line enhancer (ALE), an implementation of least mean square algorithm that has applications in detecting and tracking stationary sinusoidal signals in white noise.

In [2], Anderson et al extended the above mentioned analysis for a stationary input consisting of finite band-width signals in white noise. These signals consist of white Gaussian noise (WGN) passed through a filter whose band-width α is quite small relative to the Nyquist frequency, but generally comparable to the bin width $1/L$. They have derived analytic expressions for the weights and the output of the LMS adaptive filter as function of input signal band-width and SNR, as well as the LMS filter length and bulk delay 'z^{-P}' (please refer to Fig. 1).

In this paper, we extend the previous work in [2, 3] for enhancing a class of non-stationary signals that are composed of multiple sinusoids whose frequencies and the amplitudes may vary from block to block and which are embedded in white noise. The key difference in the approach proposed in this paper is that we relax the assumption of the input signal being stationary. Therefore the input signal is blocked into frames and we analyze a L-weight least squares filter (LeSF), estimated on each frame which consists of N samples of the input signal.

We have derived the analytical expressions for the impulse response of the L-weight least squares filter (LesF) as a function of the input SNR (computed over the current frame), effective band-width of the signal (due to finite frame length), filter length 'L' and frame length 'N'. Recognizing that such a time-varying sinusoidal model[7] is a reasonable approximation to the speech waveforms, we have applied the block estimated LeSF filter for de-noising speech signals embedded in broad-band noise. Sinusoidal model is particularly suitable for voiced speech which consists of sinusoids with frequencies at the multiple of the fundamental frequency (pitch). The ASR experiments were performed on the OGI Numbers95[19] database which consists of free-format connected numbers. The clean utterances were corrupted by realistic additive noise from the Noisex[20] database. The usual Mel-frequency cepstral coefficient (MFCC) [17] features are derived from the LeSF enhanced speech signal for automatic speech recognition (ASR) application. The experimental results indicate a significant improvement in the ASR performance.

2 Least Squares filter (LeSF) for signal enhancement

The basic operation of the LeSF is illustrated in figure (1) and it can be understood intuitively as follows. The autocorrelation sequence of the additive noise $u(n)$ that is broad-band decays much faster for higher lags than that of the speech signal. Therefore the use of a large filter length ('L') and the bulk delay P causes de-correlation between the noise components of the input signal,

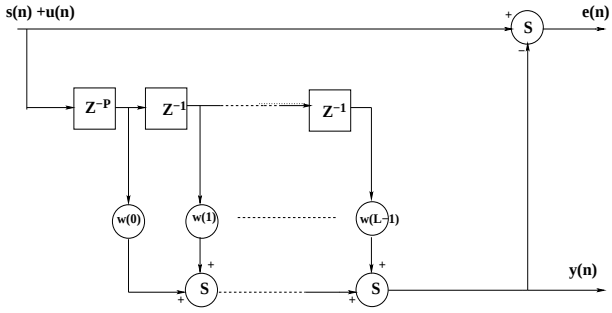

Fig. 1. The basic operation of the LeSF. The input to the filter is noisy speech, ($x(n) = s(n) + u(n)$), delayed by bulk delay $=P$. The filter weights w_k are estimated using the least squares algorithm based on the samples in the current frame. The output of the filter $y(n)$ is the enhanced signal.

namely ($u(n-L-P+1)$, $u(n-L-P+2)$, ..., $u(n-P)$) and the noise component of the reference signal, namely ($u(n)$). The LeSF filter responds by adaptively forming a frequency response which has pass-bands centered at the frequencies of the formants of the speech signal while rejecting as much of broad-band noise (whose spectrum lies away from the formant positions). Denoting the clean and the additive noise signals by $s(n)$ and $u(n)$ respectively, we obtain the noisy signal $x(n)$.

$$x(n) = s(n) + u(n) \tag{1}$$

The LeSF filter consists of L weights and the filter coefficients w_k for k $\in [0, 1, 2..L-1]$ are estimated by minimizing the energy of the error signal $e(n)$ over the current frame, $n \in [0, N-1]$.

$$e(n) = x(n) - y(n) \tag{2}$$

$$\text{where } y(n) = \sum_{i=0}^{L-1} w(i)x(n-P-i) \tag{3}$$

Let $\mathbf{A}$ denote the $(N+L) \times L$ data matrix[6] of the input frame $\mathbf{x} = [x(0), x(1), ...x(N-1)]$ and $\mathbf{d}$ denote the $(N+L) \times 1$ desired signal vector which in this case is just a delayed version of signal $\mathbf{x}$. The LeSF weight vector $\mathbf{w}$ is then given by

$$\mathbf{w} = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H\mathbf{d} \tag{4}$$

As is well known, $\mathbf{A}^H\mathbf{A}$ is a symmetric $L \times L$ Toeplitz matrix whose (i,j) element is the temporal autocorrelation of the signal vector $\mathbf{x}$ estimated over the frame length [6].

$$\left[\mathbf{A}^H\mathbf{A}\right]_{i,j} = r(|i-j|) \tag{5}$$

$$= \sum_{n=0}^{N-|i-j|} x(n)x(n+|i-j|) \tag{6}$$

In practice, $\mathbf{A}^H\mathbf{A}$ can always be assumed to be non-singular due to presence of additive noise[6] for filter length $L < N$. The weight vector $\mathbf{w}$ in (4) can be obtained using Levinson Durbin algorithm[6] without incurring a significant computational cost.

3 LeSF applied to Sinusoidal model of Speech

As proposed in [7], speech signals can be modeled as a sum of multiple sinusoids whose amplitudes, phases and frequencies can vary from frame to frame. Let us assume that a given frame of speech signal $\mathbf{s(n)}$ can be approximated as a sum of M sinusoids. Then the noisy signal $x(n)$ can be expressed as

$$x(n) = \sum_{i=1}^{M} A_i \cos(\omega_i n + \phi_i) + u(n) \tag{7}$$

where $n \in [0, N-1]$ and $u(n)$ is a realization of white noise. Then the k^{th} lag autocorrelation can be shown to be,

$$\begin{array}{c} r(k) = \sum_{n=0}^{N-k-1} x(n)x(n+k) \\ \simeq \sum_{i=1}^{M}(N-k)A_i^2 cos(2\pi f_i k) + N\sigma^2\delta(k) \end{array} \tag{8}$$

where it is assumed that $N \gg 1/(f_i - f_j)$ for all frequency pairs (i, j) and the noise $u(n)$ is white, ergodic and uncorrelated with the signal s(n). The LeSF weight vector $w(k)$ is then obtained as the solution of the Normal equations,

$$\begin{array}{c} \sum_{k=0}^{L-1} r(l-k)w(k) = r(l+P) \\ l \in [0, 1, 2..L-1] \end{array} \tag{9}$$

The set of L linear equations described in (9) can be solved by elementary methods if the z-transform $(S_{xx}(z))$ of the symmetric autocorrelation sequence $(r(k))$ is a rational function of 'z' [1].

$$S_{xx}(z) = \sum_{k=-\infty}^{\infty} r(k)z^{-k} \tag{10}$$

Consider then, a real symmetric rational z transform with M pairs of zeros and M pairs of poles.

$$S_{xx}(z) = G\frac{\prod_{m=1}^{M}(z - e^{-\beta_m + j\Psi_m})(z^{-1} - e^{-\beta_m - j\Psi_m})}{\prod_{m=1}^{M}(z - e^{-\alpha_m + j\omega_m})(z^{-1} - e^{-\alpha_m - j\omega_m})} \tag{11}$$

If the signal $\mathbf{x}$ is real, then so is its autocorrelation sequence, $r(k)$. In this case the power spectrum, $S_{xx}(z)$, has quadruplet sets of poles and zeros because of the presence of conjugate pairs at $z = exp(\pm\alpha_m \pm \omega_m)$ and $z = exp(\pm\beta_m \pm \Psi_m)$.

Anderson et. al.[2] have derived the general form of the solution to (9) for input signal with rational power spectra such as that described by (11). In this case, the LeSF weights are given by,

$$w(k) = \sum_{m=1}^{M} \left(B_m e^{-\beta_m k + j\Psi_m k} + C_m e^{+\beta_m k + j\Psi_m k} \right) \tag{12}$$

As can be seen, LeSF consists of an exponentially decaying term and an exponentially growing term attributed to reflection [8], that occurs due to finite filter length L. The value of the coefficients B_m and C_m can be determined by solving the set of coupled equations obtained by substituting the expression for $w(k)$ given in (12) into (9).

To be able to use the general form of the solution of the LeSF filter as in (12), we need a pole-zero model of the input autocorrelation in the form as described in (11). For sufficiently large frame length N, such that filter length $L \ll N$, we can make the following approximation.

$$\begin{aligned} (N-k) &\simeq N e^{-k/N} \\ k &\in [0, 1, 2, \ldots, L] \text{ and } L \ll N \end{aligned} \tag{13}$$

Using this approximation in (8), we get,

$$r(k) = N e^{-k/N} \sum_{i=1}^{M} A_i^2 cos(\omega_i k) + N\sigma^2 \delta(k) \tag{14}$$

In this form, $r(k)$ corresponds to a sum of multiple decaying exponential sequences and its z transform takes up the form,

$$\begin{aligned} S_{xx}(z) = \sum_{m=1}^{M} & \frac{N A_i^2 (1 - e^{-2\alpha})}{2} \times \\ & \left(\frac{1}{(z - e^{-\alpha_m + j\omega_m})(z^{-1} - e^{-\alpha_m - j\omega_m})} \right. \\ & \left. + \frac{1}{(z - e^{-\alpha_m - j\omega_m})(z^{-1} - e^{-\alpha_m + j\omega_m})} \right) + N\sigma^2 \\ \text{where } \alpha_m &= 1/N \ \forall\, m \in [1.M] \end{aligned} \tag{15}$$

Under the approximation that the decaying exponentials are widely spaced along the unit circle, the power spectrum $S_{xx}(z)$ in (15) that consists of sum of certain terms can be approximated by a ratio of the product of terms (of the form $(z - e^{\rho + j\theta})$), leading to a rational 'z' transform. Specifically, as explained in [1, 2] and making the following assumptions,

– The pole pairs in (15) lie sufficiently close to the unit circle (easily satisfied as $\alpha \simeq 0$.)
– All the frequency pairs (ω_i, ω_j) in (15) are sufficiently separated from each other such that their contribution to the total power spectrum do not overlap significantly.

the z transform of the total input can be expressed as,

$$S_{xx}(z) = \sigma^2 \frac{\prod_{m=1}^{M}(z - e^{-\beta_m + j\omega_m})(z - e^{+\beta_m + j\omega_m})}{\prod_{m=1}^{M}(z - e^{-\alpha_m + j\omega_m})(z - e^{+\alpha_m + j\omega_m})} \times \frac{(z - e^{+\beta_m - j\omega_m})(z - e^{-\beta_m - j\omega_m})}{(z - e^{+\alpha_m - j\omega_m})(z - e^{-\alpha_m - j\omega_m})} \quad (16)$$

where $\alpha_m = 1/N$

Corresponding to each of the sinusoidal component in the input signal there are four poles at locations $z = e^{\pm\alpha \pm \omega_m}$ and there are four zeros on the same radial lines as the signal poles but at different distances away from the unit circle. Using the general solution described in (12), which has been derived at length in [2], the solution of the LeSF weight vector to the present problem is,

$$w(n) = \sum_{m=1}^{M} \left(B_m e^{-\beta_m n} + C_m e^{+\beta_m n}\right) \cos \omega_m (n + P) \quad (17)$$

The values of β_m, B_m and C_m can be determined by substituting (17) and (14) in (9). The l^{th} equation in the linear-system described in (9) has terms with coefficients $exp(-\beta_m l)$, $exp(+\beta_m l)$, $exp(-\alpha l)\cos(\omega_m(l + P))$ and $exp(\alpha l)\cos(\omega_m(l + P))$. Besides these, there are two other kind of terms that can be neglected.

– "Non-stationary" terms that are modulated by a sinusoid at frequency $2\omega_m$ where $m \in [1, M]$. For $\omega_m \neq 0$, $\omega_m \neq \pi$, their total contribution is approximately zero.[3]
– Interference terms that are modulated by a sinusoid at frequency $\Delta\omega = (\omega_i - \omega_j)$ where $(i, j) \in [1, \ldots, M]$. If filter length $L \gg 2\pi/\Delta\omega$, these interference terms approximately sum up to zero and hence can be neglected.

The coefficients of the terms $exp(-\beta_m l)$, $exp(+\beta_m l)$ are the same for each of the L equations and setting them to zero leads to just one equation which relates β_m to α and the SNR. Let ρ_i denote the "partial" SNR of the sinusoid at frequency ω_i i.e $\rho_i = A_i^2/\sigma^2$ and the complementary signal SNR be denoted as $\gamma_i = (\sum_{m=1, m\neq i}^{M} A_i^2)/\sigma^2$. Then we have the following relation,

$$\cosh \beta_i = \cosh \alpha + \frac{\rho_i}{2\gamma_i + \rho_i + 2} \sinh \alpha \quad (18)$$

[3] due to self cancelling positive and negative half periods of a sinusoid.

There are two interesting cases. First case is when the sinusoid at frequency ω_i is significantly stronger than other sinusoids such that γ_i is quite low. This is illustrated in figure (2), where we plot the bandwidth β_i of the LeSF's pass-band that is centered around ω_i as a function of the partial SNR of the i^{th} sinusoid, ρ_i. The complementary signal's SNR is quite low at $\gamma_i = -6.99db$. We plot curves for different "effective" input sinusoid's bandwidth α. From (15), we note that α is reciprocal of frame length N. The vertical line in figure (2) corresponds to the case when $\rho_i = \gamma_i$. We note that for a given partial SNR ρ_i, the LeSF bandwidth becomes narrower as the frame length N increases, indicating a better selectivity of the LeSF filter.

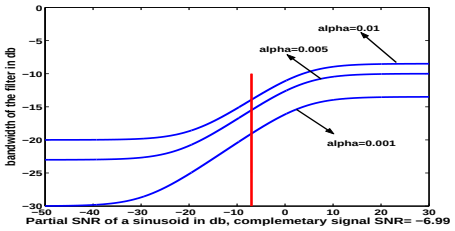

Fig. 2. *Plot of the filter bandwidth β_i centered around frequency ω_i as a function of partial sinusoid SNR ρ_i for a given complementary signal SNR $\gamma_i = -6.99db$ and "effective" input bandwidth $\alpha(alpha) = 0.01, 0.005, 0.001$ respectively. The vertical line meets the three curves when $\rho_i = \gamma_i$.*

In figure (3), we plot the bandwidth β_i as a function of ρ_i for the cases when complementary signal SNR is high at $\gamma_i = 10db$ and is low at $\gamma_i = -6.99db$. The two dots correspond to the case when $\rho_i = \gamma_i$. We note that $\gamma_i = 10db$ corresponds to a signal with high overall SNR[4]. Therefore the cross-over point $(\gamma_i = \rho_i)$ for low γ_i occurs at narrower bandwidth as compared to high γ_i case. This is so because in the former case the overall signal SNR is low and thus the LeSF filter has to have narrower pass-bands to reject as much of noise as possible.

B_i and C_i in (17) are determined by equating their respective coefficients. The "non-stationary" interference terms between all of the pairs of the frequency (ω_i, ω_j), can be neglected if $(\omega_i - \omega_j) >> 2\pi/L$. This requires that LeSF's frequency resolution $(2\pi/L)$ should be able to resolve the constituent sinusoids.

$$B_i = \frac{2e^{-\beta_i} e^{-\alpha P}(\alpha + \beta_i)^2(\beta_i - \alpha)}{((\alpha + \beta_i)^2 - e^{-2\beta_i L}(\beta_i - \alpha)^2)}$$
$$C_i = \frac{2e^{-\beta_i(2L+1)+1} e^{-\alpha P}(\alpha + \beta_i)(\beta_i - \alpha)^2}{((\alpha + \beta_i)^2 - e^{-2\beta_i L}(\beta_i - \alpha)^2)} \qquad (19)$$

[4] As overall SNR of the signal $= 10\log_{10}(10^{10\gamma_i} + 10^{10\rho_i})$

We note from (18) that the various sinusoids are coupled with each other through the dependence of their bandwidth β_i on the complementary signal SNR γ_i. As a consequence of that B_i, C_i are also indirectly dependent on the powers of the other sinusoids through β_i.

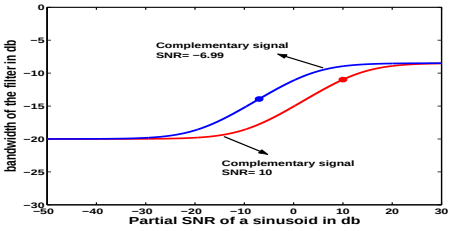

Fig. 3. *Plot of the filter bandwidth β_i centered around frequency ω_i as a function of partial sinusoid SNR ρ_i for given complementary signal SNRs $\gamma_i = -6.99db, 10db$ respectively. The "effective" input bandwidth $\alpha(alpha) = 0.01$ for both the curves. The two dots correspond to the cases when the partial SNR ρ_i is equal to complementary signal SNR γ_i.*

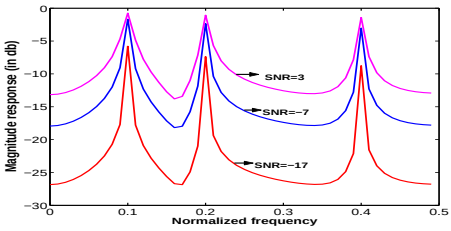

Fig. 4. *Plot of the magnitude response of the LeSF filter as a function of the input SNR. The input consists of three sinusoids at normalized frequencies (0.1, 0.2, 0.4) with relative strength* $(1 : 0.6 : 0.4)$ *respectively.*

In Fig.4, the magnitude response of the LeSF filter is plotted for various SNR. The input in this case consist of three sinusoids at normalized frequencies (0.1, 0.2, 0.4). The frame length is $N = 500$ and filter length is $(L = 100)$. As the signal SNR decreases, the bandwidth of the LeSF filter starts to decrease in order to reject as much of noise as possible. The LESF filter's gain decreases with decreasing SNR. Similar results were reported in [2, 3] for the case of stationary inputs.

In Fig. 5, we plot the spectrograms of a clean speech utterance. Fig. 6 and Fig. 7 display the same utterance embedded in F16-cockpit noise at SNR 6dB and its LeSF enhanced version respectively. As can be seen from the spectrograms, the

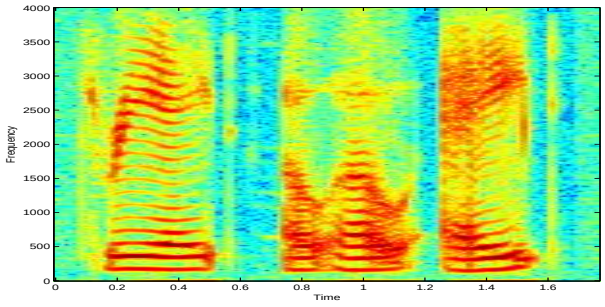

Fig. 5. *Clean spectrogram of an utterance from the OGI Numbers95 database*

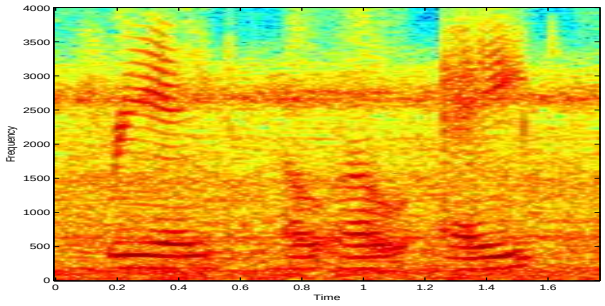

Fig. 6. *Spectrogram of the utterance corrupted by F16-cockpit noise at 6dB SNR.*

LeSF filter has been able to reject significant amount of additive F-16 cockpit noise [20] from the speech signal.

4 Experiments and Results

In order to assess the effectiveness of the proposed algorithm, speech recognition experiments were conducted on the OGI Numbers[19] corpus. This database contains spontaneously spoken free-format connected numbers over a telephone channel. The lexicon consists of 31 words[5]. The train-set and the test-set consist of 3233 and 1206 utterances respectively. Speech signals were blocked into frames of 500 samples (62.5ms) each and a 100 tap LeSF filter was derived using (4) for each frame. Noting that the autocorrelation coefficients of a periodic signal are themselves periodic with the same period (hence they do not decay with the increasing lag), Sambur[5] has used a bulk delay equal to the pitch period of the voiced speech for its enhancement. However, for the un-voiced speech a high bulk delay will result in a significant distortion by the LeSF filter as its autocorrelation coefficients decay quite rapidly for the higher lags. Therefore, we kept the bulk delay at '$P = 1$' as a good choice for enhancing both the voiced

[5] With confusable numbers like 'nine', 'ninety', 'ninteen' and so on.

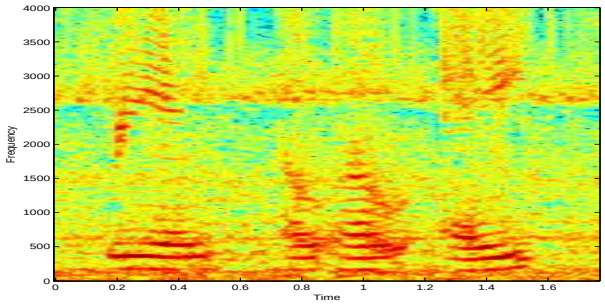

Fig. 7. *Spectrogram of the noisy utterance enhanced by a* ($L = 100$) *tap LeSF filter that has been estimated over blocks of length* ($N = 500$).

and un-voiced speech frames. However, we note that the LeSF filter is inherently more suitable for enhancing the voiced speech as in this case we can represent the speech frame as a sum of small number of sinusoids as in (7). The relatively high order ($L = 100$) of the LeSF filter is required to be able to have sufficiently high frequency resolution ($2\pi/L$) to resolve the constituent sinusoids. Each speech frame was then filtered through its corresponding LeSF filter to derive an enhanced speech frame. Finally MFCC feature vector was computed from the enhanced speech frame. These enhanced LeSF-MFCC were compared to the baseline MFCC features and noise robust JRASTA-PLP[4] features with the same window size (62.5ms). The MFCC feature vector computation is the same for the baseline and the LeSF-MFCC features. The only difference is that the MFCC baseline features are computed directly from the noisy speech while the LeSF-MFCC features are computed from LeSF enhanced speech signal. Hidden Markov Model and Gaussian Mixture Model (HMM-GMM) based speech recognition systems were trained using public domain software HTK[18] on the clean training set from the original Numbers corpus. The system consisted of 80 tied-state triphone HMM's with 3 emitting states per triphone and 12 mixtures per state.

To verify the robustness of the features to noise, the clean test utterances were corrupted using Factory and F-16 cockpit noise from the Noisex92 [20] database. The speech recognition results for the baseline MFCC, RASTA-PLP and the proposed LeSF-MFCC, in various levels of noise are given in Tables 1 and 2. Cepstral mean subtraction was performed on all the reported features. The proposed LeSF processed MFCC performs significantly better than others in all noise conditions. The slight performance degradation of the LeSF-MFCC in the clean is due to the fact that the LeSF filter being an all-pole filter does not model the valleys of the clean speech spectrum well. As a result, the LeSF filter sometimes amplifies the low spectral energy regions of the clean spectrum.

Table 1. Word error rate results for factory noise. Parameters of the LeSF filter, L=100 and N=500.

SNR	MFCC	PLP-JRASTA	LeSF MFCC
Clean	5.7	7.8	6.8
12 dB	12.3	12.2	11.9
6 dB	27.1	23.8	21.0
0 db	71.0	59.8	42.6

Table 2. Word error rate results for F16-cockpit noise. Parameters of the LeSF filter, L=100 and N=500.

SNR	MFCC	PLP-JRASTA	LeSF MFCC
Clean	5.7	7.8	6.8
12 dB	13.6	14.2	12.4
6 dB	28.4	25.3	20.6
0 db	72.3	59.2	41.2

5 Conclusion

We consider a class of non-stationary signals as input that are composed of multiple sinusoids whose frequencies and the amplitudes may vary from block to block and which are embedded in the white noise. We have derived the analytical expressions for the impulse response of the L-weight least squares filter (LesF) as a function of the input SNR (computed over the current frame), effective bandwidth of the signal (due to finite frame length), filter length 'L' and frame length 'N'. Recognizing that such a time-varying sinusoidal model[7] is a reasonable approximation to the speech waveforms, we have applied the block estimated LeSF filter for de-noising speech signals embedded in the realistic[20] broadband noise as commonly encountered on a factory floor and an aircraft cockpit. The proposed technique leads to a significant improvement in ASR performance as compared to noise robust JRASTA-PLP[4] and the MFCC features computed from the unprocessed noisy signal.

6 Acknoledgements

This work has been supported by the EU 6th Framework Programme, under contract number IST-2002-002034 (DIVINES project).

References

1. E. Satorius, J. Zeidler and S. Alexander, " Linear predictive digital filtering of narrowband processes in additive broad-band noise, " Naval Ocean Systems Center, San Diego, CA, Tech. Rep. 331, Nov. 1978.

2. C. M. Anderson, E. H. Satorius and J. R. Zeidler, " Adaptive Enhancement of Finite Bandwidth Signals in White Gaussian Noise, " In IEEE Trans. on ASSP, Vol. ASSP-31, No.1, February 1983.
3. J. R. Zeidler, E. H. Satorius, D. M. Chabries and H. T. Wexler, " Adaptive Enhancement of Multiple Sinusoids in Uncorrelated Noise, " In IEEE Trans. on ASSP, Vol. ASSP-26, No. 3, June 1978.
4. H. Hermansky, N. Morgan, " Rasta Processing of Speech," IEEE Trans. on SAP, vol.2, no.4, October 1994.
5. M. R. Sambur, " Adaptive noise canceling for Speech signals," In IEEE Trans. on ASSP, vol. ASSP-26, No.5, October 1978.
6. S. Haykin, Adaptive Filter Theory, Prentice-Hall Publishers, N.J., USA, 1993.
7. R. J. McAulay and T. F. Quatieri, " Speech Analysis/Synthesis Based on a Sinusoidal Representation, " In IEEE Trans. on ASSP, Vol. ASSP-34, No. 4, August 1986.
8. B Widrow et. al., " Adaptive noise cancelling: Principles and applications, " Proc. IEEE, vol.65, pp 1692-1716, Dec 1975.
9. M. Sondhi and D. Berkley, " Silencing echoes on the telephone network," Proc. of IEEE, vol.68, pp948-963, Aug. 1980.
10. A Gersho, " Adaptive equalization of highly dispersive channels for data transmission, " Bell Syst. Tech. J., vol.48, pp.55-70, Jan. 1969.
11. E. Satorius and S. T. Alexander, " Channel equalization using adaptive lattice algorithms, " IEEE Trans. Commun. vol. 27, pp.899-905, June 1979.
12. E. Satorius and J. Pack, " Application of least squares lattice algorithms for adaptive equalization, " IEEE Trans. on Commun. vol. COM-29, pp.136-142, Feb. 1981.
13. N. Bershad, P. Feintuch, F. Reed and B. Fisher, " Tracking characteristics of the LMS adaptive line-enhancer -Response to a linear chirp signal in noise, " IEEE Trans. on ASSP, vol. ASSP-28, pp504-517, Oct. 1980
14. L. J. Griffiths, " A simple adaptive algorithm for real time processsing in antenna arrays, " Proc. of IEEE, vol. 57, pp.1696-1704, Oct. 1969.
15. O.L. Frost, " An algorithm for linearly constrained adaptive array processing , " Proc. of IEEE, vol. 60, pp.926-935, Aug. 1972.
16. L. Marple, " Efficient least squares FIR system identification, " IEEE Trans. on ASSP, vol.ASSP-29, pp.62-73, Feb. 1981.
17. S. B. Davis and P. Mermelstein, "Comparison of Parametric Representation for Monosyllabic Word Recognition in Continuously Spoken Sentences, " IEEE Trans. on ASSP, Vol. ASSP-28, No. 4, August 1980.
18. S. Young, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, The HTK Book, Cambridge University, 1995.
19. R. A. Cole, M. Fanty, and T. Lander, "Telephone speech corpus at CSLU," Proc. of ICSLP, Yokohama, Japan, 1994.
20. A. Varga, H. Steeneken, M. Tomlinson and D. Jones, " The NOISEX-92 study on the effect of additive noise on automatic speech recognition, " Technical report, DRA Speech Research Unit, Malvern, England, 1992.