Notice of Violation of IEEE Publication Principles

"A New Approach for Robust Speech Recognition using Minimum Variance Distortionless Response"
by V. Srinas, Ch. Santhi Rani
in the Proceedings of the 2nd International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), March 2015

After careful and considered review of the content and authorship of this paper by a duly constituted expert committee, this paper has been found to be in violation of IEEE's Publication Principles.

This paper duplicates the original text and figures from the paper cited below. The original text was copied with minor edits, without attribution (including appropriate references to the original author(s) and/or paper title) and without permission.

Due to the nature of this violation, reasonable effort should be made to remove all past references to this paper, and future references should be made to the following article:

"New Features Using Robust MVDR Spectrum of Filtered Autocorrelation Sequence for Robust Speech Recognition"
by Sanaz Seyedin, Seyed Mohammad Ahadi, Saeed Gazor
in The Scientific World Journal, Volume 2013 Hindawi

IEEE Sponsored 2nd International Conference on Innovations in Information, Embedded and Communication systems (ICIIECS) 2015

A New Approach for Robust Speech Recognition using minimum variance distortionless response

V. Srinivas
Associate Professor and Head
Department of ECE
SIET
Narasapur, A.P, India

Dr Ch. Santhi Rani
Professor
Department of ECE
D.M.S & S.V.H Engineering College
Machilipatnam, India

Abstract- In this paper, we proposed a new technique presents an extraction method for robust speech recognition using the MVDR (Minimum Variance Distortionless Response) spectrum of short time autocorrelation sequence which can reduce the effects of leftover of the nonstationary additive noise that remains after filtering the autocorrelation. To produce a further robust front-end, we present the customized robust distortionless constraint of the MVDR spectral estimation method through revised weighting of the subband power spectrum values based on the sub-band signal to noise ratios (SNRs), which adjust it to the new proposed technique. The new proposed functions allow the components of the input signal at the frequencies with minimum affected by noise to pass with better weights and attenuate more effectively the noisy and unwanted components. This revision results in decrease of the noise residuals of the projected spectrum from the filtered autocorrelation sequence, thus advancing to a more robust algorithm. Our proposed technique, when analyzed on Aurora 2 task for recognition applications, best performed all MFCC (Mel frequency cepstral coefficients) as the fundamental, respective autocorrelation sequence MFCC (RAS MFCC), and the proposed MVDR related features in numerous different noisy conditions.

Keywords— MFCC,MVDR,RAS MFCC

I. INTRODUCTION

The system of Speech recognition is usually prepared in good conditions and verified in different environments such as noisy and clean. This difference between the preparation and test conditions significantly degrades the effectiveness of the automatic speech recognition (ASR) systems in noisy environments.

Robust speech recognition is accounted as one of the most demanding area in speech processing technology since the form of the noise encountered in test conditions is generally not predictable. Robust speech recognition method may be categorized into four main types [1]:

1. Robust speech quality extraction
2. Speech augmentation for improved recognition
3. Model based rectification for noise
4. Model based characteristic enhancement

Finding and study a set of parameters that are robust towards the variations considered by different noises on speech signals is the fundamental aim of the first method. This type can be further divided into two main classifications:

1. Extracting additional robust features
2. Postprocessing of the characteristic features for robustness.

Several speech processing systems such as speech recognition, speech enhancement, and speech coding applies the magnitude information of speech signals in various sparse domain like STFT (short time Fourier transform) [2, 3]. For example, MFCC (Mel frequency cepstral coefficients) [2], and perceptual linear prediction (PLP) [3] are accounted as well-known features used. Hence, rectifying the speech signal power spectrum to make it robust towards convolution or additive distortions is further widely used in the previous type. From the useful methods in this type, here refer to differential power spectrum (DPS) [4], AMFCC (autocorrelation Mel frequency cepstral coefficients) [5], comparative autocorrelation sequence MFCC (RAS-MFCC) [6], DCT and differentiated autocorrelation sequence (DAS) [7], and MVDR based features [8–11]. Feature and Characteristic extraction algorithms based on auditory system like PNCC (power normalized cepstral coefficients) [12] and FBCC (Fourier-Bessel Cepstral Coefficients) [13] are furthermore among other methods in the previous group. Enhancing Mel-filtered log spectrum of noisy signals from estimated distribution of speech in this domain has been also proposed for extracting additional robust features [14]. In addition, feature normalization is counted as best significant group of types in the postprocessing of features. HEQ Histogram equalization [15], CSN (cepstral subband normalization) [18] and CMNM (cepstral moment normalization methods) [16, 17], and are the most excellent examples in this division.

The Purpose of proposed method in this paper is to modify the noisy speech signal power spectrum to get additional robust features. In this method, we categorized first division as robust feature extraction approach to extract robust spectrum of MVDR of filtered autocorrelation sequence.

Hence, the robust feature in proposed method is obtained from to the following strategies:

1. Filtering the STAS (short time autocorrelation sequence), sequentially to decrease the noise effects
2. Extracting the spectrum of MVDR, against the common periodogram technique
3. Improving the spectrum of MVDR and getting a additional robust to decrease the residual noise effects including nonstationary noises.

Estimation methods of spectral are either parametric [19] or nonparametric When the periodogram of FFT based is one of the extremely popular method of the previous strategy, particularly in the domain of speech recognition, MVDR and model identification methods are most popular of the latter [19]. The model identification methods are divided in three divisions as moving average (MA), auto regressive (AR) and ARMA [19]. The conventional speech features, like MFCC is extracted by FFT based periodogram power spectrum, whose estimation is, suffer from variance and large bias [10]. Mainly bias is caused from the leakage of power by the bandpass filter frequencies of surrounding, used to calculate the power [10]. Employing a single sample in the power calculation process is caused for large variance [10]. Hence, with this drawbacks have been acknowledged by the proposed spectrum of MVDR estimation method [8, 10].
Include the spectral of MVDR calculation method that decreases the variance and bias of the spectrum estimates promising the effectiveness in extracting robust speech features. This appears because;

1. The variance and bias of the spectrum estimate influence speech features that are applied to extract Gaussian parameters modeling the speech classes.
2. Improving the level of noise in noisy signals extends the variance of the power spectrum and therefore degrades the recognition accuracy.
3. The features used in ASR systems, likely LP (linear prediction) [2] and PLP, that are based on LP or AR or LP methods, which are not good one for exact calculation of the power spectrum of voiced speech, considerably high pitch voices.

Hence, the spectrum of LP based that also sensitive to noise, due to its feature that tends to follow the well structure of speech spectrum in such voices [10]. Hence, we proposed MVDR- based speech feature extraction as a suitable strategy for building ASR systems additional robust towards noise. Additionally, filtering the temporal trajectories of STAS, referred as RAS, is useful in removing the noise effects in stationary noise [6]. Thus, the additive noise appeared mainly in ASR systems is a nonstationary. So, this approach cannot eliminate complete distortions.

Hence, we presented to find the spectrum of MVDR of this filtered autocorrelation sequence to additional decrease the noise residuals. Furthermore, sequentially to build the presented method additionally robust even in SNR, we advise using a robust MVDR strategy related to [11]. The additional robust in [11] is obtained by modify the distortionless constraint of the spectral of MVDR calculation method by loading the values of subband power spectrum based on the subband SNRs.
We also proposed to modify the function proposed in [11] to not only correct it to the new proposed method but also develop the recognition efficiency in both low and high SNRs. We have recommended this revisions on the weighting function to get used to it to the proposed method on the temporally filtered perceptual spectrum autocorrelation sequence, that has higher subband SNR compare with nonfiltered. The higher subband SNR is caused by suppressing some parts of the noise by the mentioned temporal filter.
The new function that presents, a new robust distortionless constraint for the filtered spectrum of MVDR causes additional considerable components of input signal at the frequencies slightly affected from noise to pass with higher weights, when attenuating the noisy with less reliable components with lesser weights. Therefore, the noise effects still remain after apply the mention filter on the autocorrelation sequence should reduced, and is useful in nonstationary noise. The recognition results show that this strategy is very helpful in extracting more robust features.

II. RELATED WORK

Robust MVDR Spectral Estimation

The main aim of proposed spectral of MVDR calculation is decreasing the variance and bias of the estimated spectrum. This target is obtained by designing the FIR filter, $h(n)$, that minimizes the output power subject to the constraint which is response at the frequency, ω_l, has gain unity. This distortionless constraint validates that the input signal components with the frequency pass exclusive of any distortion throughout the filter. Furthermore, power minimization of the output avoids the leakage of power by surrounding frequencies, ensuring in reduced bias. The signal power at the frequency of refer will be the same of the filtered signal power [10, 19]. Therefore, computing the power spectrum by all output samples reduces the variance. The designed MVDR filter by solving the following constrained optimization problem [10]:

$$\min_{\mathbf{h}} \mathbf{h}^H \mathbf{R}_{L+1} \mathbf{h} \quad \text{subject to } \mathbf{v}^H(wl)\,\mathbf{h} = 1 \qquad (1)$$

which results in

$$\mathbf{h}_l = \frac{\mathbf{R}_{L+1}^{-1}\mathbf{v}(\omega_l)}{\mathbf{v}^H(\omega_l)\,\mathbf{R}_{L+1}^{-1}\mathbf{v}(\omega_l)}, \qquad (2)$$

Where k(ω) = [1, $ej\omega$, $ej2\omega$, . . . , $ejL\omega$] , $\mathbf{R}_{L+1}$ is the $(L + 1) \times (L + 1)$ Toeplitz autocorrelation matrix of the data, and $\mathbf{h} = [h0, h\ 1, \ldots, h_L]^T$. The MVDR spectrum for all frequencies is then computed by [10]

$$P_{\mathrm{MV}}(\omega) = \frac{1}{\mathbf{v}^H(\omega)\,\mathbf{R}_{L+1}^{-1}\mathbf{v}(\omega)}. \tag{3}$$

According to the distortionless constraint in (1), the filter responses at all frequencies contribute to the final result with the same weighting since all have unity gains. Therefore, noise generally affects the speech signal differently in different frequencies. As a result, if some frequencies are corrupted by noise, the resultant MVDR power spectrum at such frequencies will also be corrupted. With this problem, we proposed a robust distortionless constraint in [11] by modifying this constraint like that the response of the filter at the frequency of refer has a gain that is driven by the SNR at that frequency, alternatively of a unity gain. By this process will be the same as weighting the value of power spectrum at the frequency of refer based on the ratio of the signal energy to the noise energy of at that frequency which make the spectrum of MVDR robust towards noise. Hence, the robust spectrum of MVDR for all frequencies will be computed by [11]

$$P_{\mathrm{RMVDR}}(\omega) = \frac{w(\omega)^2}{\mathbf{v}^H(\omega)\,\mathbf{R}_{L+1}^{-1}\mathbf{v}(\omega)}, \tag{4}$$

Where

Where $S(\omega_l)$ and $N(\omega_l)$ are the clean signal and noise at the frequency of interest, ωl, respectively. Hence, we assign larger weights to the input signal components at the frequencies slight affected by noise, while the others obtain smaller weights. In [11], employing the experimental findings of psychoacoustics [20, 21], we proposed using the following weighting function with values between zero and one:

$$w_i^2 = 1 - \exp\left(-\frac{\mathrm{SNR}_i}{\gamma_i}\right), \tag{6}$$

Where SNR_i is computed from the ratio of the of noisy signal energy to noise in the ith mel frequency subband and γ_i, the gain that regulates the weighting function steepness. So, this weighting function has been recommended because using the raw subband SNR as the weighting factors did not ahead to adequate recognition efficiencies in low SNRs as per experimental results. The optimum function eq.(7), that prepared the difference between two sigmoidal functions, was proposed for γ_i in [11] based on recognition experiments.

$$\gamma_i = \frac{1}{1 + \exp\left(-3\left(\mathrm{SNR}_i - 0.5\right)\right)} - \frac{1}{1 + \exp\left(-3\left(\mathrm{SNR}_i - 3.5\right)\right)}. \tag{7}$$

In figure 1, shows the flow diagram for extracting the proposed RPMCC (robust perceptual MVDR based cepstral coefficients) according to the explained procedure. RPMCC features are extracted from the warped power spectrum by incorporating the PLP structure as in [11]. This gives better recognition results because exploiting the perceptual information always improves the speech recognition. The power law of hearing blocks and equal loudness curve are according to [3].We calculate the warped power spectrum by applying the conventional triangular Mel-based filter bank to the FFT based periodogram. Subsequently the warped power spectrum of MVDR is computed from the Mel warped spectrum after employing weighting to subbands. Followed by, the cepstral features are calculated by employing IFFT to the Mel scale MVDR log spectrum. Subband spectrum is referred for The Mel-warped spectrum in the area of speech recognition.

III. THE PROPOSED APPROACH

In this paper, we proposed a new Front End Based on Robust Spectrum of MVDR Filtered Autocorrelation Sequence.

We consider an additive noise model as follows:

$$y(m,n) = x(m,n) + v(m,n), \quad 0 \le m \le M-1,\ 0 \le n \le N-1, \tag{8}$$

Where $x(m, n)$ referred as clean speech, $V(m, n)$ represents a additive noise and $y(m, n)$ is a noisy speech waveform. In a frame, m is the frame index and n is the discrete time index and M represents the number of frames and N the represents frame samples.

A related autocorrelation additive equation for of cleans speech, noisy speech and noise on the condition that noise is considered to be uncorrelated with speech

$$r_{yy}(m,k) = r_{xx}(m,k) + r_{vv}(m,k), \quad 0 \le m \le M-1,\ 0 \le k \le N-1, \tag{9}$$

Where
$r_{xx}(m, k)$ referred as SAS of Clean Speech
$r_{yy}(m, k)$ referred as the Noisy Speech
$r_{VV}(m, k)$ referred as additive noise
k is the autocorrelation sequence index

If the additive noise is assumed to have the same values for all frames. Hence, we can avoid the frame index m from $r_{VV}(m,k)$ in (9).

$$r_{yy}(m,k) = r_{xx}(m,k) + r_{vv}(k), \quad 0 \le m \le M-1, \quad 0 \le k \le N-1. \tag{10}$$

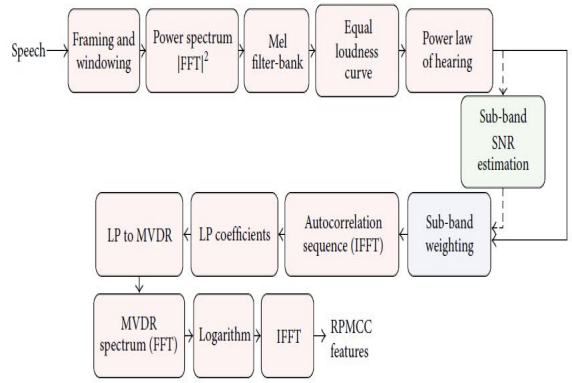


Figure 1: Diagram for extracting RPMCC features [11]. The subband weighting is applied according to (6) and (7).

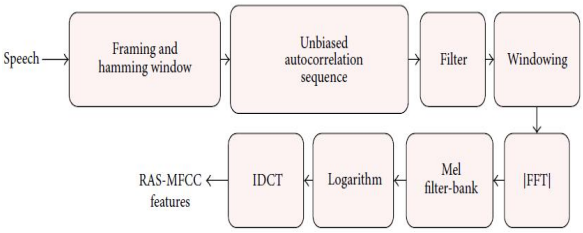


Figure 2: Diagram for extracting RAS-MFCC features similar to [6]

Therefore, we can compute the differentiation of both sides of (10) with respect to m, to calculate the RAS of noisy and clean speech, which yields [6]

$$\frac{\partial r_{yy}(m,k)}{\partial m} = \frac{\partial r_{xx}(m,k)}{\partial m}, \quad 0 \le m \le M-1, \quad 0 \le k \le N-1. \tag{11}$$

This differentiation can be acquired by FIR filter on the temporal autocorrelation trajectory. Then the transfer function is [6]:

$$H(z) = \frac{1}{T_Q} \sum_{t=-Q}^{Q} t z^t, \tag{12}$$

Where

$$T_Q = \sum_{t=-Q}^{Q} t^2, \tag{13}$$

Where $(2Q + 1)$ is the frame range of the applying filter.

In Figure 2, shows the block diagram for extracting the features of RAS MFCC. We have varied the approach proposed in [6] by including a hamming window at the beginning of the process to make it similar to other speech feature extraction methods. Therefore, for applying the RAS filter to find the correct power spectrum, a DDR (double dynamic range) hamming window is required. The reason for adding this block is that the power spectrum of an autocorrelation sequence has a dynamic range twice that of the corresponding power spectrum of the signal. Therefore, to build an N-length DDR hamming window, the following procedure is adopt that is similar to [5]:

1. Construct Hamming window with an $N/2$ length
2. Compute its $(N-1)$ length two sided autocorrelation sequence, has a maximum at zeroth lag in the center,
3. Pad one zero at the end to make a N-length desired window.

In other words, the RAS of clean speech can be calculated by applying the high-pass filter in (12) to the autocorrelation of noisy speech in the frame range specified by $(2Q + 1)$. According to (11), providing the additive noise is stationary, then RAS of noisy speech will be same to RAS of clean speech, and therefore the effect of noise is removed. Though, non-stationary additive noise is often used in ASR systems. Thus, this approach cannot eliminate the complete distortions and only suppresses the slowly varying noise or stationary noise and DC. Hence to decrease the noise residuals that remain after applying RAS filter. we proposed finding the spectrum of MVDR of this filtered autocorrelation sequence. Furthermore, to restrain the noise effects and then find additional robust features while in low SNRs also, we proposed using a robust MVDR strategy refer in [11]. Hence, we extract the proposed spectrum of MVDR comparative PMSR features from the power spectrum of subband MVDR of filtered SAS. The proposed PMSR features of extracting are shown in Figure 3(a). Figure 3(b) illustrates the proposed procedure for extracting RPMSR features. To compute RPMSR coefficients, we first pass the SAS during a RAS filter in (12). Subsequently we find the proposed features from the spectrum of RPMVDR of this filtered autocorrelation sequence.
The spectrum of RPMVDR is computed similar to the strategy explained in Section 2. Though, considering to our experiments shown in Section 4, the signal subband SNRs o passed during a RAS filter are improved compared with no filter to the autocorrelation sequence. This occurs due to suppression of noise effects by processing the signal with RAS approach. This causes the subband SNRs to be computed additional reliably. Hence, the propose modify

the subband weighting function recommended in [11]. Finally, added two free parameters to the steepness controlling gain in (7) to build it extra flexible to subband SNR variations

$$\gamma_i = \frac{1}{1+\exp\left(-3\left(\mathrm{SNR}_i-\gamma_{1_i}\right)\right)} - \frac{1}{1+\exp\left(-3\left(\mathrm{SNR}_i-\gamma_{2_i}\right)\right)}, \quad (14)$$

Where

$$\gamma_{1_i} = 0.4 + \frac{0.1}{1+\exp\left(-\left(\mathrm{SNR}_i-1\right)\right)}, \qquad \gamma_{2_i} = 3 + \frac{0.5}{1+\exp\left(4\left(\mathrm{SNR}_i-1\right)\right)}. \quad (15)$$

While the proposed γ_i in (14) is applied as the controlling gain in the subband weighting function in (6), higher weights are assigning to higher SNRs, when lower SNRs obtain smaller weights. Thus, this weighting function make the algorithm additional robust due to encounter less error when computing the subband SNRs of the warped power spectrum estimated from the RAS filtered autocorrelation sequence compared with nonfiltered. So, this robust weighting acknowledge the input signal components at the frequencies referred slightly affected by noise to pass with higher weights, when attenuating such components that are unwontedly more affected by noise through assigning lesser weights. In addition, the new proposed function for γ_i makes it more flexible to variations of SNR; that is, it can be more easily tuned to a desired environment. The fixed values used in the proposed functions for γ_{1i} and γ_{2i} have been selected to increase the speech recognition efficiencies. The experimental results prove the usefulness of this proposed algorithm. Figure 4 compares the proposed γ_i and the resulted weighting function with those used in [11] for signals not passed through a RAS filter.

Hence, to extract the RPMCC features based on the subband weighting in (6) and (7), when used (6) and (14) to estimate the subband weighting for RPMSR coefficients. Therefore to develop the performance of proposed algorithm towards nonstationary noises, we compute the power spectrum of noise by a easy updating algorithm, while the first some Nonspeech frames are calculated as the initial values of noise.

if $\mathrm{E}[y_l(i)] \le \beta \mathrm{E}[N_l(i-1)]$ then
$\mathrm{E}[N_l(i)] = \alpha \mathrm{E}[N_l(i-1)] + (1-\alpha)\,\mathrm{E}[y_l(i)]$ else
$\mathrm{E}[N_l(i)] = \mathrm{E}[N_l(i-1)]$,

Where $\mathrm{E}[N_l(i)]$ and $\mathrm{E}[y_l(i)]$ are the calculated noisy signal energies and the noise of the lth subband in frame i, respectively. Additionally, setting a α to 0.99 and β to 2. In addition, $\mathrm{SNR}_l(i)$, is the lth subband in frame i, is computed as:

$$\mathrm{SNR}_l(i) = \frac{\mathrm{E}[y_l(i)]}{\mathrm{E}[N_l(i)]}. \quad (17)$$

For calculation purposes, the Lth order spectrum of MVDR is calculated by LP coefficients a_k and error variance P_e [10, 19]

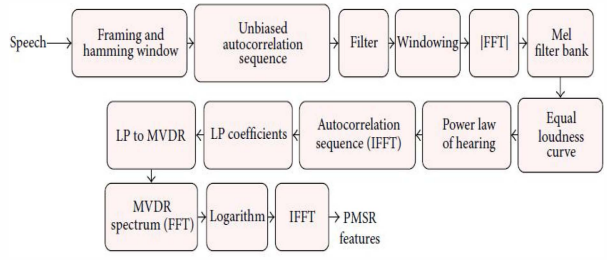
$$P_{\mathrm{MVDR}}(\omega) = \frac{1}{\sum_{-L}^{L}\mu(k)\,e^{-j\omega k}}, \quad (18)$$

$$\mu(k) = \begin{cases} \frac{1}{P_e}\sum_{i=0}^{L-k}(L+1-k-2i)\,a_i a_{i+k}^* & k=0,\ldots,L \\ \mu^*(-k) & k=-L,\ldots,-1, \end{cases} \quad (19)$$

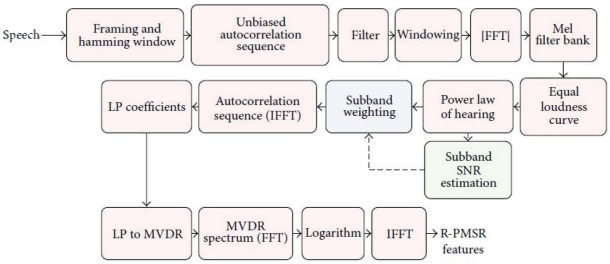
Where $(2L+1)$ coefficients of $\mu(k)$ are referred coefficients of MVDR and spectrum of MVDR can simply be computed by an FFT computation based on (18).

IV. EXPERIMENTAL RESULTS

We have conducted the recognition experiments on Aurora 2 task [23] with clean preparation scenario. A well known task, Aurora 2 is a often adopted for analyzing the robust speaker in dependent speech recognition. This has been developed from the database of TI Digits, composing of connected digits spoken by American English talkers, and is sampled to 8 kHz. It adds two training methods: multi condition and clean-condition training. we only use the clean condition training set, includes 8440 utterances consisting the recordings of 55 female and 55 male adults.
Where the all signals has been filtered by G.712 characteristic. Aurora 2 task test data having three sets, like, test set A, B, and C. Subsequently, 4004 utterances from test set data of TI Digits are dividing with four subsets and each as 1001 utterances. In addition the clean speech signals, one noise type is included to each subset at SNRs of 20 dB, 15 dB, 10 dB, 5 dB, 0 dB, and −5 dB. The test set A is a babble, suburban train, exhibition hall and car. Besides, street, restaurant, train station and airport noises are computed in set B test. Set C as 2 of the 4 subsets. Thus, when each of the tests sets of A and B having 28028 utterances and set C is 14014 utterances. In test set C, street and Suburban train are used as additive noises, then noises and speech are filtered with an MIRS characteristic before including the so known noises. Set C is used to analyze the presentation of ASR systems consider the additive distortions and convolution.

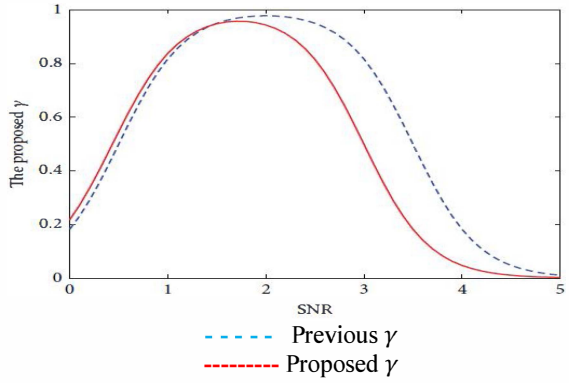


(a) PMSR features

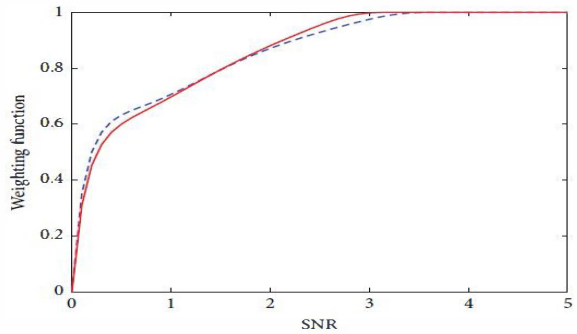


(b) R-PMSR features

Figure 3: The diagram for extracting the proposed new front end based on RMVDR spectrum



(a) The proposed γ function



(b) The proposed weighting function

Figure 4: Comparison of the proposed γ.

We used HMMs (Hidden Markov models) to model the pauses and digits and applying the topology in [23]. The robustness of the acquired features has been analyzed on Aurora 2 task by HTK software [25]. The principle uses the well known MFCC features. Speech has been segmented into 25ms frames for extracting all the features, with a frame shift of 10 ms. 23 triangular filters are accommodated in the Mel filter bank.

15 model orders has used for MVDR based coefficients that gives the excellent average recognition efficiency shown in previous experiments. We used the RAS filter for features based on filtering the autocorrelation sequence with an order of 2 to obtain the good results of recognition. Typically, by applying a Juang lifter [2] with a parameter of 22 to cepstral coefficients to additional improve the recognition for all features. Lastly, each frame was presented by a vector having 12 cepstral features increased by their first and second derivatives of order. We forwarded a set of beginning experiments to access the idea of modifying γ_i as the steepness controller of the weighting function. Finally, compared the average subband SNR estimated after applying the RAS filter to the unbiased autocorrelation sequence that computed no filtering.

Hence, created a compact corpus consisting of 110 Aurora 2 files extracted from the clean training utterances. We have chosen these files like that the ensuing compact database contains both single and connected digit utterances. We also included four noises of babble, subway, exhibition and car, at SNRs of 20 dB, 15 dB, 10 dB, 5 dB, 0 dB, and −5 dB with the procedure in [23]. Estimated subband SNRs obtained over this compact database for both cases of using the RAS filter and without it, as explained by figure 5. In Figure 3(b), shown that the subband SNRs are calculated at output of Mel filter banks by equal power law of hearing and loudness curve.

Thus, we proposed subband weighting based on (6) and (14), as discussed before in Section 3. The fixed values of the functions in (14) and (15) were also tuned using the recognition experiments results. Table 1 gives the average recognition accuracies compared with different noise types and test sets. Figures 6, 7, and 8, shows the recognition accuracies for the proposed and baseline features. The RPMSR features lead to a improvement of 35.3%, 31.1%, and 34.6% for test sets compared to MFCC in the average WER (word error rate). This relative improvement over RAS MFCC is equal to 26.77%, 24.6%, and 24.8%.

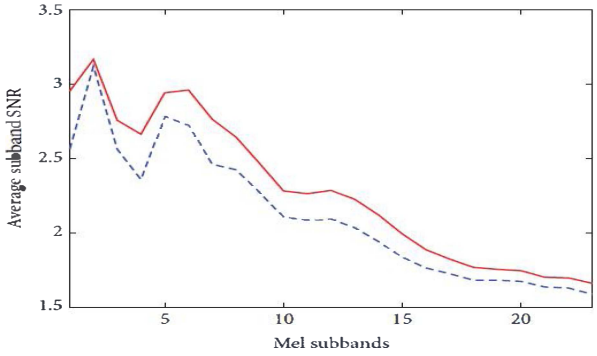
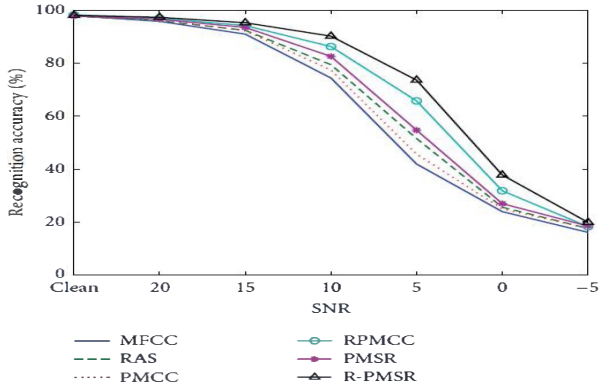


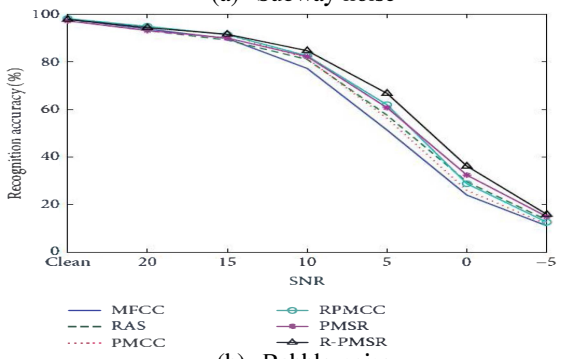
Figure 5: Comparison of the average estimated subband SNRs.

Table 1: Average recognition accuracies over different noise types and SNRs for test sets A, B, and C and different features.

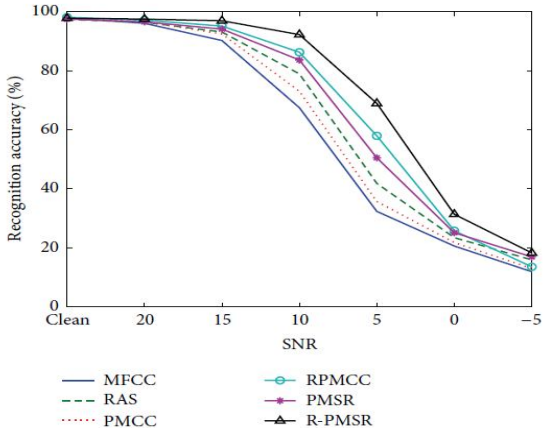
Feature	*Set A*	*Set B*	*Set C*
MFCC	63.90	66.15	58.21
RAS-MFCC	68.10	69.06	63.70
PMCC	66.25	68.73	60.87
RPMCC	72.36	72.99	67.17
PMSR	70.28	71.11	65.79
R-PMSR	76.64	76.68	72.69



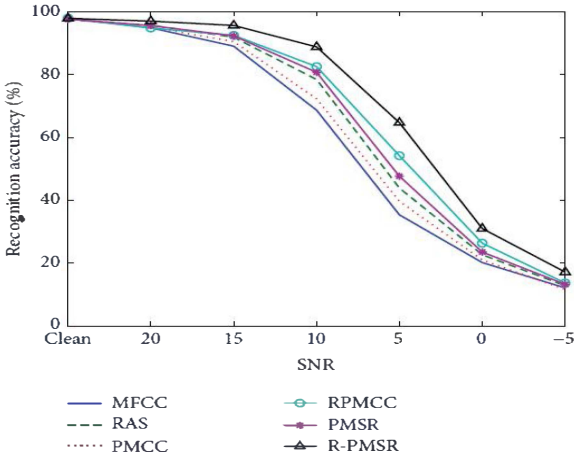
(a) Subway noise



(b) Babble noise

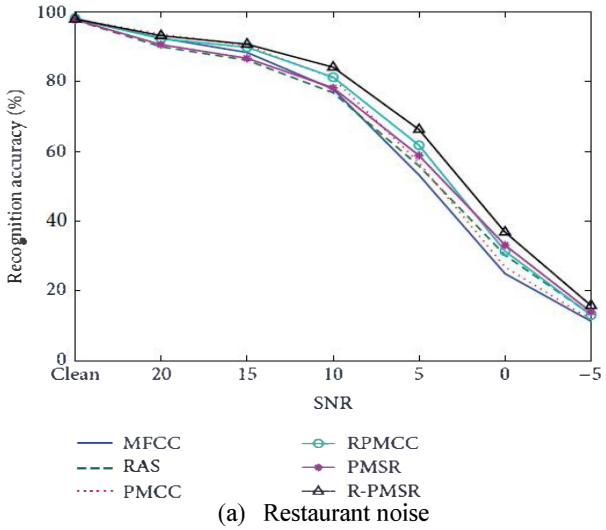


(c) Car noise

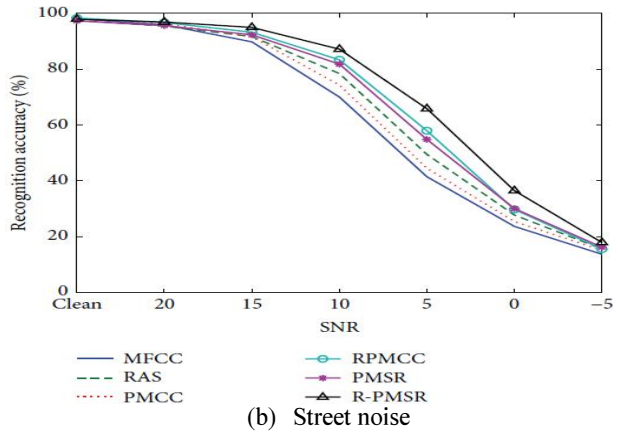


(d) Exhibition noise

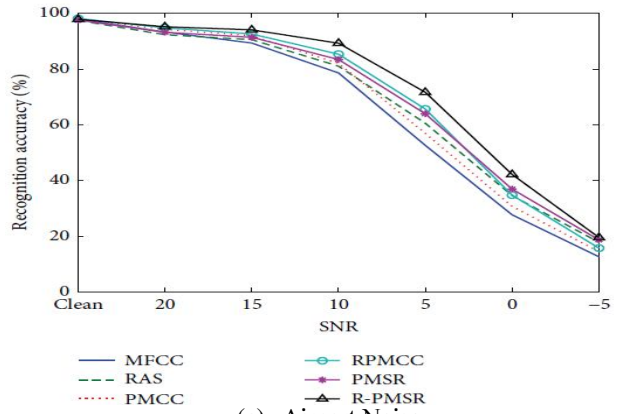
Figure 6: Recognition accuracies for different features in various noise types of test set A of Aurora 2 task.



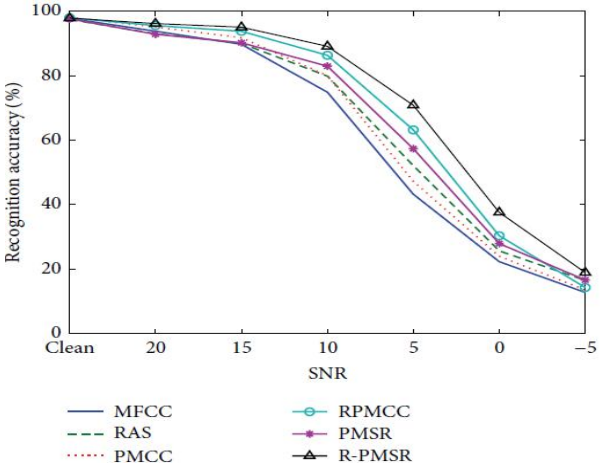
(a) Restaurant noise



(b) Street noise



(c) Airport Noise



(d) Train station Noise

Figure 7: Recognition accuracies for different features.

V. CONCLUSION

We proposed a new front end for robust speech recognition which is based on spectrum of RPMVDR of RAS filtered autocorrelation sequence. While we attend the nonstationary additive noise in ASR, temporal trajectories are filtered of SAS cannot eliminate completely in the distortions.

So, we proposed finding the spectrum of PMVDR of this filtered autocorrelation sequence to further decrease the noise residuals. This proposal led to PMSR features with a excellent performance than RAS MFCC in all noisy and clean cases. Furthermore, we modified our previously recommended weighting function for RPMCC, not only regulate with new proposed strategy and also improve the recognition efficiencies in both low and high SNRs.

References

[1] M. Su, P. Li, Z.Wang, P. Ding, and B. Xu, "A novel noise robust front- end using first order VTS in construction of mel-warped wiener filter," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, pp. I777–I780, Toulouse, France, May 2006.

[2] L. Rabiner and B. Juang, Fundamentals of Speech Recognition,Prentice Hall, 1993.

[3] H. Hermansky, "Perceptual linear predictive (PLP) analysis ofspeech," Journal of the Acoustical Society of America, vol. 87, no.4, pp. 1738–1752, 1990.

[4] J. Chen, K. K. Paliwal, and S. Nakamura, "Cepstrum derived from differentiated power spectrum for robust speech recognition," *Speech Communication*, vol. 41, no. 2-3, pp. 469–484, 2003.

[5] B. J. Shannon and K. K. Paliwal, "Feature extraction from higher-lag autocorrelation coefficients for robust speech recognition," *Speech Communication*, vol. 48, no. 11, pp. 1458–1485, 2006.

[6] K.-H. Yuo and H.-C. Wang, "Robust features for noisy speech recognition based on temporal trajectory filtering of short-time autocorrelation sequences," *Speech Communication*, vol. 28, no.1, pp. 13–24, 1999.

[7] G. Farahani, S.M. Ahadi, andM.M. Homayounpour, "Features based on filtering and spectral peaks in autocorrelation domain for robust speech recognition," *Computer Speech and Language*, vol. 21, no. 1, pp. 187–205, 2007.

[8] S. Seyedin and M. Ahadi, "Feature extraction based on DCT and MVDR spectral estimation for robust speech recognition," in Proceedings of the 9th International Conference on Signal *Processing (ICSP '08)*, pp. 605–608, Beijing, China, October 2008.

[9] S. Seyedin and S. M. Ahadi, "Robust MVDR-based feature extraction for speech recognition," in *Proceedings of 7th IEEE* International Conference on Information, Communications and *Signal Processing*, Macao, China, December 2009.

[10] S. Dharanipragada, U. H. Yapanel, and B. D. Rao, "Robust feature extraction for continuous speech recognition using the MVDR spectrum estimation method," *IEEE Transactions on* Audio, Speech and Language Processing, vol. 15, no. 1, pp. 224– 234, 2007.

[11] S. Seyedin and S.M. Ahadi, "A new subband-weighted MVDRbased front-end for robust speech recognition," *IEICE Transactions on Information and Systems*, vol. E93-D, no. 8, pp. 2252– 2261, 2010.

[12] C. Kim and R. M. Stern, "Power-normalized cepstral coefficients (PNCC) for robust speech recognition," in *Proceedings* of the IEEE International Conference on Acoustics, Speech and *Signal Processing (ICASSP '12)*, pp. 4101–4104, Kyoto, Japan, 2012.

[13] C. Prakash and S. V. Gangashetty, "Fourier-Bessel cepstral coefficients for robust speech recognition," in *Proceedings of the* IEEE International Conference on Signal Processing and Communications *(SPCOM '12)*, Bangalore, India, July 2012.

[14] S. Seyedin, S. Gazor, and S. M. Ahadi, "On the distribution of Mel-filtered log spectrum of speech in additive noise," under review, 2013.

[15] A. de la Torre, A. M. Peinado, J. C. Segura, J. L. P´erez-C´ordoba, M. C. Ben´ıtez, and A. J. Rubio, "Histogram equalization of speech representation for robust speech recognition," *IEEE* Transactions on Speech and Audio Processing, vol. 13, no. 3, pp. 355–366, 2005.

[16] R. Togneri, A. M. Toh, and S. Nordholm, "Evaluation and modification of cepstral moment normalization for speech recognition in additive babble ensemble," in *Proceedings of the 11th Australasian International Conference on Speech Science and Technology (SST '06)*, Auckland, New Zealand, December 2006.

[17] C.-W. Hsu and L.-S. Lee, "Higher order cepstral moment normalization for improved robust speech recognition," *IEEE* Transactions on Audio, Speech and Language Processing, vol. 17, no. 2, pp. 205–220, 2009.

[18] C.-W. Hsu and L.-S. Lee, "Higher order cepstral moment normalization for improved robust speech recognition," *IEEE* Transactions on Audio, Speech and Language Processing, vol. 17, o. 2, pp. 205–220, 2009.

[19] S. S. Wang, J. W. Hung, and Y. Tsao, "A study on cepstral subband normalization for robust ASR," in *Proceedings of the International* SymposiumonChinese SpokenLanguageProcessing, pp. 141–145, Hong Kong, 2012.

[20] S. L.Marple,Digital Spectral AnalysiswithApplications, Prentice Hall, 1987.

[21] K. K.Chu and S. H. Leung, "SNR-dependent non-uniformspectral compression for noisy speech recognition," in *Proceedings* of the IEEE International Conference on Acoustics, Speech, and *Signal Processing*, pp. 973–976, Montreal, Canada, May 2004.

[22] E. Zwicker and H. Fastl, Psychoacoustics, Facts and Models, chapter 8, Springer, 3rd edition, 2007.

[23] Z. Junhui, K. Jingming, X. Xiang, and H. Shilei, "Noise suppression based on teager energy operator for improving the robustness of ASR front-end," in Proceedings of the International Workshop on Acoustic Echo and Noise Control, pp. 135–138, Antibes, France, September 2003..

[24] H. G. Hirsch and D. Pearce, "The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in Proceedings of the International Workshop on Automatic Speech Recognition: Challenges for the *new Millenium (ASR '00)*, pp. 181–188, Paris, France, September 2000.

[25] "Transmission performance characteristics of pulse code modulation channels," ITU recommendation G.712, November 1996.

[26] HTK, "The hiddenMarkovmodel toolkit," 2002, http://htk.eng.cam.ac.uk/.

[27] www,hindwai.com