ACADEMIC
World Academic Union

ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 6, No. 1, 2011, *pp*. 055-063

Normalized Autocorrelation based Features for Robust Speech Recognition in Context with Noisy Environment

Poonam Bansal[1], Amita Dev[2] and Shail Bala Jain[3]

[1]Department of Computer Science and Engineering, Amity School of Engineering and Technology, Guru Gobind Singh Indraprastha University, New Delhi, India

[2]Department of Computer Science and Engineering, Ambedkar Institute of Technology, New Delhi, India

[3]Department of Electronics and Communication Engineering, IGIT, Guru Gobind Singh Indraprastha University, New Delhi, India

(*Received December 3, 2009, accepted December 20, 2010*)

Abstract. This paper presents a robust approach for an automatic speech recognition system (ASR) when both additive and convolutional noises corrupt the speech signal. Robust features are derived by assuming that the corrupting noise is stationary and the channel effect is fixed during the utterance. In the proposed method the effect of additive and convolutional distortions are minimized by two stage filtering. The first filtering stage includes differential temporal filtering in the autocorrelation domain for reducing additive noise effects, followed by additional filtering in the logarithmic spectrum domain to reduce convolutional noise effects. Convolutional channel distortion is assumed to be linear and time invariant. A task of multispeaker isolated Hindi word recognition is conducted to demonstrate the effectiveness of using these robust features. The cases of channel filtered speech signal corrupted by white noise and different colored noises such as factory, babble and F16, which are further corrupted by channel distortion are tested. Experimental results show that the proposed method can significantly improve the performance of isolated Hindi word recognition system in noisy environment.

Keywords: ASR, Channel distortion, CMN, MFCC

1. Introduction

Today, a major concern in the design of speech recognition systems is their performance in noisy conditions. Therefore, a substantial amount of research is devoted to the development of noise-robust speech features. One of the domains attracted attention in this regard is the autocorrelation domain. A number of feature extraction algorithms have been devised using this domain as the initial domain of choice. Considering a noisy speech, if the speech signal and noise are considered uncorrelated, then the autocorrelation of their sum is equal to the sum of their autocorrelations. Most important property of the autocorrelation sequence is considered to be the preservation of the original signal poles. If the original signal can be modeled by an all pole sequence which has been excited by an impulse train and a white noise, the poles of the autocorrelation sequence would be the same as the poles of the original signal [1],[2]. Which means that the features extracted from the autocorrelation sequence could replace the features extracted from the original speech signal. Second property of the autocorrelation sequence is that as the autocorrelation of noise could in many cases be considered relatively constant over time, a high pass filtering of the autocorrelation sequence could lead to substantial reduction in its effect. Among the techniques used to exploit the autocorrelation properties are Short-time Modified Coherence [1], One-Sided Autocorrelation LPC (OSALPC) [3], Relative Autocorrelation Sequence (RAS) [4],[5] and Autocorrelation Me1 Frequency Cepstral Coefficient (AMFCC).

To overcome the problem of additive noise and the channel distortion, techniques proposed are parallel model combination (PMC) [6], Stochastic matching (SM) [7]-[9] and combining channel identification with power spectrum estimation [10],[11]. Recently, several novel techniques for handset and channel compensation are also proposed for speaker identification. [12] proposed a robust feature extraction using a

[1] Corresponding author. *E-mail address*: pbansal89@yahoo.co.in

non-linear artificial neural network to optimize the speaker recognition performance. Some approaches have used magnitude spectrum of higher lag autocorrelation coefficients [13] and others have stressed upon preservation of spectral peaks [14]. Furthermore, for normalization various techniques are developed such as cepstral mean normalization (CMN), RelAtive SpecTrAl (RASTA) [15] and Blind Equalization (BE) [16]. Each of the above-mentioned autocorrelation-based methods has their own disadvantages. The PMC method needs a prior knowledge of the noise and SM takes time for iterative estimation of noise statistics. The RAS method, works well in low SNRs but does not perform as well in high SNRs. RAS and AMFCC methods considers distortion only due to additive noises, the convolutional noise in the form of channel distortion has not been considered. The AMFCC method works well for car and subway noises, but in babble and exhibition noises does not work as well. The reason could be that for the two later noise types, the noise properties are very similar to speech properties. Authors have already proposed a double fold additive noise suppression method for reducing the effect of additive noise. It is based upon the Differentiated Relative Autocorrelation Sequence Spectrum Mel Frequency Cepstral Coefficients (DRASS-MFCC) [17]. In this paper a novel dual filtering method on autocorrelation sequence is proposed to nullify the effects of both additive as well as convolutional noise. A temporal filtering method on autocorrelation domain is done to nullify the effect of additive noise plus an additional filtering has been suggested in the logarithmic spectrum domain by mean subtraction method to remove the channel effect. New derived parameters are called Channel Adaptive Relative Autocorrelation Sequence (CARAS). From the magnitude of CARAS the mel-scale frequency cepstral coefficients (MFCCs) of CARAS are derived. These MFCC are denoted as CARAS-MFCC. Comparisons in the recognition rate are made with DRASS-MFCC, RAS-MFCC and standard MFCC. It is well known that cepstral mean normalization (CMN) and delta cepstral coefficients are two effective methods for removing bias in traditional MFCC feature vector set. Here we have used CMN for that. CARAS-MFCC shows remarkable results in recognition accuracy for the signals corrupted by both additive as well as convolutional noises.

The remainder of the paper is organized as follows. Mathematical fundamentals for extracting RAS, DRASS and CARAS are derived in section 2. Block diagram of the proposed front end is described in section 3. In section 4 experiments conducted in clean and noisy environment with the proposed method are discussed. Finally a conclusion is given in section 5.

2. Robust features extraction

2.1. Extraction of RAS

Let m be the frame index and n be the time index within a frame. The clean speech x(m,n) corrupted by the additive noise u(m,n) result in a noisy speech expressed by

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

Where M denotes the number of frames in an utterance and N denotes the number of samples in a frame. If the noise is uncorrelated with the speech, it follows that the autocorrelation of the noisy speech y(m,n) is the sum of autocorrelation of the clean speech x(m,n) and autocorrelation of the noise u(m,n), i.e.

$$r_{yy}(m,k) = r_{xx}(m,k) + r_{uu}(m,k),\ 0 \le m \le M-1,\ 0 \le k \le N-1 \tag{2}$$

where $r_{yy}(m, k)$, $r_{xx}(m, k)$ and $r_{uu}(m, k)$ are the one-sided autocorrelation sequences of noisy speech, clean speech and noise respectively, and k is the autocorrelation sequence index within each frame. If the additive noise is assumed to be stationary, the autocorrelation sequence of noise can be considered to be identical for all frames. Hence, the frame index m can be dropped out, and (2) becomes

$$r_{yy}(m,k) = r_{xx}(m,k) + r_{uu}(k),\ 0 \le m \le M-1,\ 0 \le k \le N-1 \tag{3}$$

Here the N-point ryy (m,k) is computed from N-point y(m,n) using the following equation,

$$r_{yy}(m,k) = \sum_{i=0}^{N-1-k} y(m,i)\, y(m, i+k),\ 0 \le k \le N-1 \tag{4}$$

Applying the temporal filtering on both sides of (3), it comes out

$$\Delta r_{yy}(m,k) = \Delta r_{xx}(m,k),\ 0 \le m \le M-1,\ 0 \le k \le N-1 \tag{5}$$

where

$$\Delta r_{yy}(m,k) = r_{yy}(m+1,k) - r_{yy}(m-1,k) \text{ and } \Delta r_{xx}(m,k) = r_{xx}(m+1,k) - r_{xx}(m-1,k)$$

The sequence, $\left\{\Delta r_{yy}(m,k)_{k=0}^{N-1}\right\}$ is named the Relative Autocorrelation Sequence (RAS) of noisy speech at the mth frame.

2.2. Extraction of DRASS

In order to preserve spectral peaks of speech signal, we use spectrum differentiation of the filtered signal, which we get from previous step (RAS). This further contributes to immunization against noise. By this approach the flat parts of the spectrum are almost removed while each spectral peak is split into two, one positive and one negative. If we call the output of the filtered signal in time domain as z(m,n), we can write

$$z(m,n) = x(m,n) + v(m,n),\ 0 \le m \le M-1, 0 \le n \le N-1 \tag{6}$$

Where x(n) and v(n) are clean speech and the remaining noise after filtering, respectively. Then we calculate the autocorrelation function as follows

$$r_{zz}(m,k) = r_{xx}(m,k) + r_{vv}(m,k),\ 0 \le k \le N-1, \tag{7}$$

Applying the Fourier transform to both sides of (7) yields

$$FT\{r_{zz}(m,k)\} = FT\{r_{xx}(m,k)\} + FT\{r_{vv}(k)\}, \text{ or}$$

$$Z(\omega) = X(\omega) + V(\omega) \tag{8}$$

where FT{ } represents the Fourier Transform and ω indicates radian frequency. The Differential Power Spectrum (DPS) will then be defined as

$$DZ(\omega) = \frac{dZ(\omega)}{d\omega} = \frac{dX(\omega)}{d\omega} + \frac{dV(\omega)}{d\omega} = DX(\omega) + DV(\omega) \tag{9}$$

where DX(ω) and DV(ω)are the differential power spectra of the given frame of noise free speech and noise signal.

2.3. Extraction of CARAS

Any speech signal x(m,n) , first corrupted by additive noise and then distorted by channel distortion can be written as

$$y(m,n) = [x(m,n) + u(m,n)] \otimes h(n), \quad 0 \le m \le M-1,\ 0 \le n \le N-1 \tag{10}$$

Where u(m, n) is the additive noise, and h(n) is the channel distortion and "⊗" represents the convolution operation. Note that we have assumed the channel effect is fixed in an utterance, and thus h(n) is independent of frame index m. If x(m,n), u(m, n) and h(n) are considered uncorrelated, the autocorrelation of the noisy speech can be expressed as

$$r_{yy}(m,k) = [r_{xx}(m,k) + r_{uu}(m,k)] \otimes h(k) \otimes h(-k),\ 0 \le m \le M-1,\ 0 \le k \le 2N-1 \tag{11}$$

where $r_{yy}(m, k)$, $r_{xx}(m, k)$ and $r_{uu}(m, k)$ are the two-sided autocorrelation sequences of the noisy speech, clean speech and additive noise, respectively, and k is the autocorrelation index within a frame. Since additive noise is assumed to be stationary, the frame index m, can be dropped out and (11) becomes

$$r_{yy}(m,k) = r_{xx}(m,k) \otimes h(k) \otimes h(-k) + r_{uu}(k) \otimes h(k) \otimes h(-k),\ 0 \le m \le M-1, 0 \le k \le 2N-1 \tag{12}$$

By applying the temporal filtering on both sides of (12), we obtain :

$$\Delta r_{yy}(m,k) = \Delta r_{xx}(m,k) \otimes h(k) \otimes h(-k) \quad 0 \le m \le M-1, 0 \le k \le 2N-1 \tag{13}$$

where $\Delta r_{yy}(m,k) = r_{yy}(m+1,k) - r_{yy}(m-1,k)$ and $\Delta r_{xx}(m,k) = r_{xx}(m+1,k) - r_{xx}(m-1,k)$. Note that while applying the temporal filtering we have removed the noise term ruu(k) ⊗ h(k)⊗h(-k)

Taking the 2N- Point DFT on both sides of (13) with respect to k, we get

$$S_{yy}^{\Delta}(m,f) = S_{xx}^{\Delta}(m,f).\ |H(f)|2,\ 0 \le f \le 2N-1 \tag{14}$$

where $S_{yy}^{\Delta}(m,f)$, $S_{xx}^{\Delta}(m,f)$ and H(f) denotes the spectra of $\Delta r_{yy}(m,k)$,$\Delta r_{xx}(m,k)$ and h (k), respectively

Taking the logarithm of (14) we obtain

$$\log S^{\Delta}_{yy}(m,f) = \log S^{\Delta}_{xx}(m,f) + 2\log|H(f)|,\ 0 \le f \le 2N-1 \quad (15)$$

As the channel noise becomes an additive term in the logarithmic power spectrum . Taking the 2N point inverse DFT with respect to f, we get CARAS

$$\overline{r_{yy}}(m,k) =$$

$$= \text{Inverse DFT}\left\{ \exp\left(\log S^{\Delta}_{yy}(m,f) - \frac{1}{M}\sum_{m=0}^{M-1} \log S^{\Delta}_{yy}(m,f) \right) \right\},\ 0 \le k \le 2N-1,\ 0 \le f \le 2N-1,\ 0 \le m \le M-1 \quad (16)$$

Channel Adaptive Relative Autocorrelation Sequence (CARAS). CARAS can be considered as an alternative representation of the original speech signal in the time domain that is robust to additive noise and the effects of channel distortion.

3. Description of proposed method

In this section, description of the proposed method to get new robust features for speech recognition is presented. First, the speech signal is split into frames of 256 samples each, and a pre-emphasis filter is applied on each frame. A Hamming window is applied and after that, the autocorrelation sequence of the framed signal is obtained using an unbiased estimator. A temporal filtering is then applied to the autocorrelation sequence to get the relative autocorrelation sequence (RAS). When a speech is corrupted by the additive noise, the noise component is additive to the speech not only in the autocorrelation domain but also in the power spectrum domain. By calculating RAS, we broadly remove the additive noise. Then differentiated relative autocorrelation sequence spectrum (DRASS) is calculated by taking the differential power spectrum of the relative autocorrelation coefficients (RAS). Further a set of cepstral coefficients (DRASS-MFCC) are derived from the magnitude of the relative autocorrelation power spectrum by applying it to a conventional mel-frequency filter-bank and finally passing its logarithm to the DCT block. For speech signals corrupted by the additive noise and channel distortion RAS is transformed to CARAS, to remove the channel effect by applying mean subtraction in the logarithmic

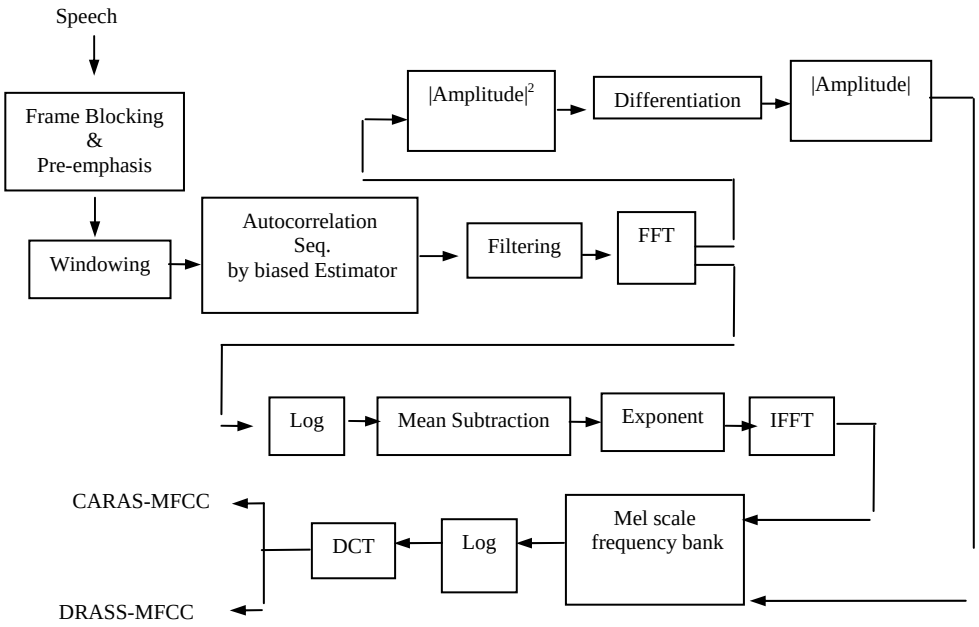

Figure 1. Block Diagram showing Feature Extraction by CARAS

spectrum domain. A set of cepstral coefficients (CARAS-MFCC) can be derived from the magnitude of the CARAS by applying it to a conventional mel-frequency filter-bank and finally passing its logarithm to the DCT block. CMN is further used for removing the bias in the traditional MFCCs. Block diagram in Figure 1. displays the proposed front-end diagram of our method.

4. Experiments

A digital database of 200 Hindi words spoken by 30 speakers (Table 1) was used for the experiments of speaker- independent isolated word recognition system. The spoken samples were recorded by 15 male, 10 female and 5 child speakers in a studio environment condition using Sennheiser microphone model MD421 and a tape recorder model Philips AF6121. Each speaker uttered 5 repetitions of words. The database was partitioned for the use of training and testing. The 1200 utterances from 20 speakers were used for training the HMM model. The test set contained similar utterances from 10 speakers that were not included in the training set. The features in all the cases of training and testing are computed using 16 msec. frame length with 8 msec. of frame shift. Pre-emphasis coefficient used is 0.9375. For each speech frame, a 20-channel Mel-scale filter bank is used. First, word models of training database are created by seven state left-right Hidden Markov model using MFCC. While testing, features are extracted using MFCC (for comparison purposes), RAS-MFCC, DRASS-MFCC and CARAS-MFCC. Word recognition rates for testing database are computed with RAS-MFCC, DRASS-MFCC and CARAS- MFCC under different noise conditions and compared with the traditional MFCC.

4.1. Clean Speech Testing

This experiment is to evaluate the performance of MFCC, RAS-MFCC, DRASS-MFCC and CARAS-MFCC when the training data & the testing data are in clean environment (S/N> 40dB). The results are shown in Table 2. These are the baseline results for comparison purposes. Performance on the basis of recognition rate is observed to be more or less same in the case of MFCC, RAS-MFCC or CARAS-MFCC. But DRASS-MFCC shows a prominent increase. The recognition rate gets improved to 99.64% as compared to above three MFCC and with CMN it reaches 99.84%. This is due to the peculiar property of DRASS, which combines the features of RAS and Differential Power Spectrum (DPS). By taking differentiation, the spectral peaks, which convey the most important information are preserved, hence show better recognition rate. Secondly it is found that CMN optimization works well with each type of feature vector set.

Table 1. Hindi speech Database for a vocabulary of 200 words used in the experiment

1. Language	Standard Hindi (Khari Boli)
2. Vocabulary Size	A set of 200 most frequently occurring Hindi words
3. Speakers	30 Speakers
4. Utterances	(15 male, 10 female and 5 children) 5 repetitions each
5. Audio Recording	Recording on a casette tape in studio S/N > 40
6. Digitization	16 kHz., Sampling 16 bit quantization.

Table 2. Recognition rates (%) under clean enrollment

	Recognition Rate (%)	Recognition Rate (%) with CMN
MFCC	98.24	99.27
RAS-MFCC	98.24	99.32
DRASS-MFCC	99.64	99.84
CARAS-MFCC	98.73	99.78

4.2. Noisy Speech Testing - Corrupted by Additive Noise

In this experiment the testing speech is polluted by the additive noise. RAS-MFCC and DRASS-MFCC are evaluated and compared with the traditional MFCC. The testing utterances are generated by adding the artificial noises in five SNR levels. The white noise is generated by using a random number generation program, and other colored noises, i.e., factory noise, F16 noise, and babble noise, are extracted from the NATO RSG-10 corpus [18]. MFCC, RAS-MFCC and DRASS-MFCC with Cepstral Mean Normalization (CMN) are also evaluated. The results are summarized in Table 3(a-d). For the case of white noise corruption the performance of MFCC degrades most significantly among all features. Although MFCC with CMN can make some improvement, its performance is still worse than RAS-MFCC and DRASS-MFCC Additionally DRASS with CMN shows the best performance. Table 3(b),(c) and (d), shows the recognition rates when the testing speech is corrupted by factory, babble and F16 noise, respectively. Figure 2. Plots the average

recognition rates shown in Table 3. It is observed from the figure that the performance of MFCC degrades significantly. The best performance comes from DRASS-MFCC combined with CMN. As the speech is corrupted by additive noise, the noise component is additive to the speech not only in the autocorrelation domain but also in the power spectral domain. Broadly the additive noise is removed in the autocorrelation domain while calculating RAS. By differentiating the Relative Autocorrelation Power Spectrum (in DRASS), the flat part of the spectrum is transformed into some values approximating to zero, and the spectral peaks which convey the most important information in speech signal are preserved, which gives the enhanced information of the speech signal. Further CMN also enhances the rate of recognition.

Table 3. Recognition rate(%) for testing speech corrupted by additive noise (all MFCCs compensated with CMN)

(a) White noise

SNR (dB)	40	20	15	10	5	0
MFCC	98.24	67.24	33.64	14.01	7.69	3.03
MFCC(CMN)	99.27	71.05	40.62	22.85	15.38	11.23
RAS-MFCC	98.24	91.66	85.18	58.42	32.31	12.70
RAS-MFCC(CMN)	99.32	92.10	87.50	64.81	32.91	13.19
DRASS-MFCC	99.64	97.01	89.00	71.10	39.13	16.12
DRASS-MFCC(CMN)	99.84	97.91	91.94	73.02	41.16	17.98

(b) Factory noise

SNR (dB)	40	20	15	10	5	0
MFCC	98.24	90.90	61.06	29.5	7.95	4.87
MFCC(CMN)	99.27	92.19	63.04	31.41	9.05	5.87
RAS-MFCC	98.24	94.09	83.09	57.76	29.05	5.87
RAS-MFCC(CMN)	99.32	95.89	84.97	59.00	30.95	7.06
DRASS-MFCC	99.64	98.65	97.00	84.83	46.86	12.59
DRASS-MFCC(CMN)	99.84	98.00	96.06	86.90	48.06	14.91

(c) Babble noise

SNR (dB)	40	20	15	10	5	0
MFCC	98.24	97.19	83.1	50.50	20.23	5.93
MFCC(CMN)	99.27	97.99	85.22	50.99	22.98	7.13
RAS-MFCC	98.24	95.03	90.25	74.05	41.98	15.80
RAS-MFCC(CMN)	99.32	96.04	91.99	76.09	44.00	17.24
DRASS-MFCC	99.64	99.00	98.00	87.99	60.70	16.99
DRASS-MFCC(CMN)	99.84	99.00	96.89	89.00	65.78	18.00

(d) F-16 noise

SNR (dB)	40	20	15	10	5	0
MFCC	98.24	80.21	45.67	23.54	6.22	1.10
MFCC(CMN)	99.27	82.00	46.99	25.98	8.79	1.37
RAS-MFCC	98.24	92.00	78.00	35.03	13.09	3.03
RAS-MFCC(CMN)	99.32	94.06	80.00	36.92	14.99	2.40
DRASS-MFCC	99.64	98.92	97.70	85.11	47.00	14.6
DRASS-MFCC(CMN)	99.84	98.94	98.00	89.00	49.07	16.9

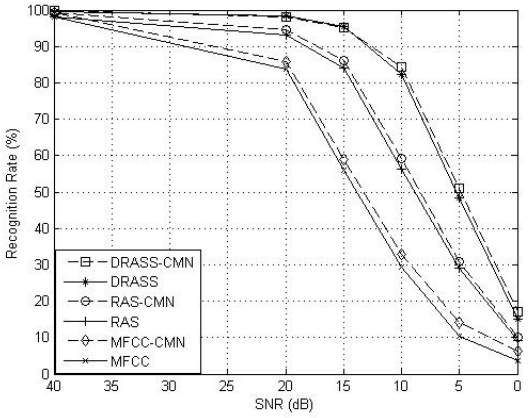

Figure 2 Average Recognition rates (%) for testing speech corrupted by Additive noise only

4.3. Noisy speech testing- Corrupted by Additive noise and channel Distortion

This is the experiment when the testing speech is polluted by additive noise and channel distortion simultaneously. First the speech data is corrupted with white and colored noises at five SNR levels and then distorted by channel distortion. The white noise is generated by using a random number generation program, and other colored noises, i.e., factory noise, F16 noise, and babble noise, are extracted from the NATO RSG-10 corpus [18]. The noises are added to the clean speech signal at 20, 15, 10 5 and 0 dB SNRs. RAS-MFCC, DRASS-MFCC and CARAS-MFCC are evaluated and compared with the traditional MFCC. Results are shown in Figure 3(a),(b),(c) and (d). It is observed that the performance of MFCC is worse than RAS-MFCC, DRASS-MFCC and CARAS-MFCC. For the clean data DRASS-MFCC and CARAS-MFCC shows comparable recognition rates, but as the noise level increases CARAS-MFCC becomes better, because it takes care of additive noise and channel effect effectively. Figure 3(b), (c) and (d), depicts the performance in terms of recognition rate when the testing speech is corrupted by factory, babble and F16 noises along with channel distortion respectively. The best performance comes from CARAS-MFCC. This is due to the mean normalization in frequency-domain during the derivation of CARAS. The normalization in frequency domain also provides the compensation to additive noise when the testing speech is corrupted by colored noise. The other finding in Figure 3(d) is that the assumption of corrupting by stationary noise during the derivation of RAS-MFCC and CARAS-MFCC does not hurt too much to the performance of our experiments. The babble noise is usually the case of non-stationary noise. Figure 3(d) shows that CARAS can still obtain relatively high accuracy rates for babble noise corruption even in low SNR's.

5. Conclusion

In this paper, a robust feature vector set CARAS-MFCC has been introduced for isolated Hindi word recognition. The CARAS- MFCC proves to be robust to additive noise as well as channel distortion. For the case of clean enrollment DRASS-CMN gives the best performance. Several types of noise corruption to the testing data with various levels of SNRs are evaluated. Experimental results show that proposed robust feature, CARAS-MFCC is effective for overcoming additive as well as channel distortion in case of low SNR environment. This method is effective for colored noise corruption also. As our method has used two step noise elimination approach, it outperforms the well known RAS approach. The limitation of the proposed method is that the derivation of RAS, DRASS and CARAS is based on the assumption of corrupting by stationary noise. This may limit the application of CARAS to a more diverse environment.

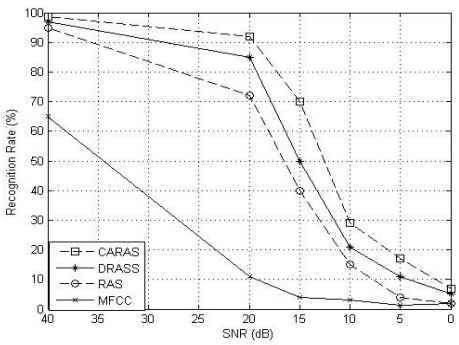

Figure 3(a) Recognition rate (%) for testing speech corrupted by White noise and Channel distortion

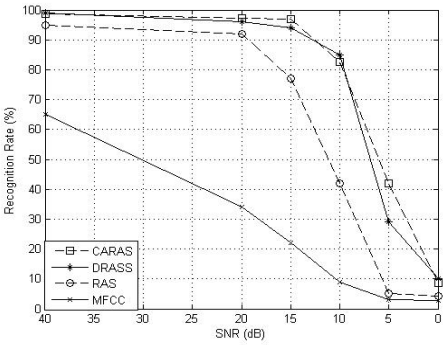

Figure 3(b) Recognition rate (%) for testing speech corrupted by Factory noise and Channel distortion

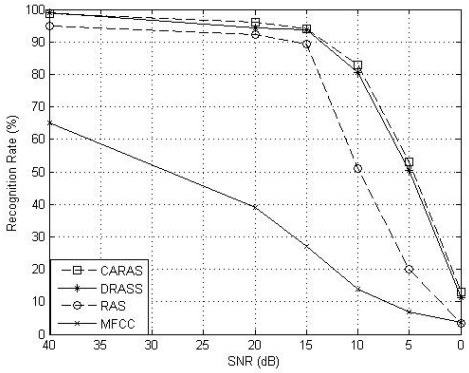

Figure 3(c) Recognition rate (%) for testing speech corrupted by Babble noise and Channel distortion

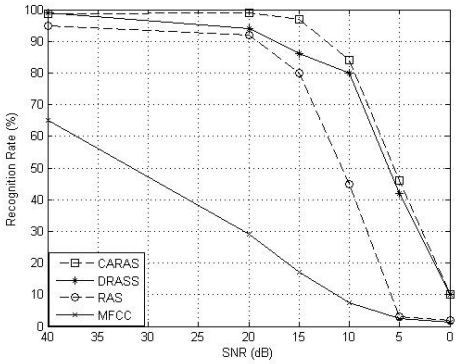

Figure 3(d) Recognition rate (%) for testing speech corrupted by F16 noise and Channel distortion

6. References

[1] D. Mansour, B.H. Juang. The short-time modified coherence representation and noisy speech recognition. *IEEE Transactions on Acoustics and Signal Processing*. 1989, **37**(6): 795-804.

[2] D.P. McGinn, D.H. Johnson. Estimation of all-pole model parameters from noise-corrupted sequence. *IEEE Trans on Acoustics Speech and Signal Processing*. 1989, **37**(3): 433-436.

[3] J. Hernando , C. Nadeu. Linear prediction of the one-sided autocorrelation sequence for noisy speech recognition. *IEEE Trans. Speech Audio Processing*. 1997, **5**(1): 80-84.

[4] K.H. You, H.C. Wang. Robust features derived from temporal trajectory filtering for speech recognition under the corruption of additive and convolutional noises. *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*. 1998, pp. 577-580.

[5] K.H. Yuo et al.. Robust features for noisy speech recognition based on temporal trajectory filtering of short-time autocorrelation sequences. *Speech Communication*. 1999, **28**: 13-24.

[6] M.J.F Gales, S.J. Young. Robust speech recognition in additive and convolutional noise using parallel model combination. *Comput. Speech Lang.* 1995, **9**: 289-307.

[7] A. Sankar, C.H. Lee. A maximum-likelihood approach to stochastic matching for robust speech recognition. *IEEE Trans. SpeechAudio Process.* 1996, **4**: 190-202.

[8] O. Siohan, C.H. Lee, Iterative noise and channel estimation under the stochastic matching algorithm framework. *IEEE Signal Process. Lett.* 1997, **4**: 304-306.

[9] Y. Zhao, Channel identification and signal spectrum estimation for robust automatic speech recognition. *IEEE Signal Process. Lett.* 1998, **5**(12): 305-308.

[10] M. Afify et al. A general joint additive and convolutive bias compensation approach applied to noisy Lombard speech recognition. *IEEE Trans. Speech Audio Process.* 1998, **6**(6): 524-538.

[11] H.A. Murthy et al. Robust text-independent speaker identification over telephone channels. *IEEE Trans. Speech*

Audio Process. 1999, **7**(5): 554-568.

[12] L.P. Heck et al. Robustness to telephone handset distortion in speaker recognition by discriminative feature design. *Speech Commun.* 2000, **31**: 181-192.

[13] B. Shannon, K.K. Paliwal, Feature extraction from higher-lag autocorrelation coefficients for robust speech recognition. *Speech Communication*. 2006, **48**(11): 1458-1485.

[14] B. Strope, A. Alwan. Robust word recognition using threaded spectral peaks. *Proceedings of ICASSP*. 1998, pp. 625- 628.

[15] Hermansky and Morgan. RASTA processing of speech. *IEEE Transactions Speech Audio Processing*. 1994, **2**(4): 578-589.

[16] L. Mauuary. Blind equalization for robust telephone based speech recognition. *Proceedings of European Signal Processing Conference*, 1998.

[17] Poonam Bansal et al. Role of spectral peaks in Autocorrelation domain for robust speech recognition. *Journal of Computing and Information Technology*. 2009, **17**(3): 295- 303.

[18] A. Varga et al, Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Commun.* 1993, **12**: 247-251.