A Hybrid Method for Automatic Speech Recognition Performance Improvement in Real World Noisy Environment

[1]Urmila Shrawankar and [2]Vilas Thakare

[1]Department of Computer Science and Engineering, G H Raisoni College of Engineering, Nagpur, India
[2]Department of Computer Science and Engineering, SGB Amravati University, Amravati, India

ABSTRACT

It is a well known fact that, speech recognition systems perform well when the system is used in conditions similar to the one used to train the acoustic models. However, mismatches degrade the performance. In adverse environment, it is very difficult to predict the category of noise in advance in case of real world environmental noise and difficult to achieve environmental robustness. After doing rigorous experimental study it is observed that, a unique method is not available that will clean the noisy speech as well as preserve the quality which have been corrupted by real natural environmental (mixed) noise. It is also observed that only back-end techniques are not sufficient to improve the performance of a speech recognition system. It is necessary to implement performance improvement techniques at every step of back-end as well as front-end of the Automatic Speech Recognition (ASR) model. Current recognition systems solve this problem using a technique called adaptation. This study presents an experimental study that aims two points, first is to implement the hybrid method that will take care of clarifying the speech signal as much as possible with all combinations of filters and enhancement techniques. The second point is to develop a method for training all categories of noise that can adapt the acoustic models for a new environment that will help to improve the performance of the speech recognizer under real world environmental mismatched conditions. This experiment confirms that hybrid adaptation methods improve the ASR performance on both levels, (Signal-to-Noise Ratio) SNR improvement as well as word recognition accuracy in real world noisy environment.

Keywords: Real World Noisy Environment, ASR Performance, Speech Processing Hybrid Methods, Environment Adaptation, ASR Back-end Techniques, ASR Front-end Techniques, SNR Improvement, Word Recognition Accuracy Improvement

1. INTRODUCTION

The performance of Automatic Speech Recognition (ASR) system based on acoustic model totally depends on the environment of training and testing data (Ding *et al*., 2010).

In a speech recognition system, many parameters affect the accuracy of the system. These parameters are speaker, isolated or continuous word recognition, size of vocabulary, language, environment conditions.

Robustness to additive noise remains a largely unsolved problem in automatic speech recognition research today. Various approaches to combat degradation of recognition performance due to noise distortion have been suggested (Laska *et al*., 2010), with some level of success. Many of the approaches to build noise-robust recognition systems can be classified into one of the three primary categories: back-end adaptation techniques, front-end enhancement algorithms and alternative feature approaches. This class of technique focuses on adapting acoustic model parameters to better match the environmental conditions present.

The experiment was performed by implementing the following category of filters and enhancement techniques independently as well as in combination.

Corresponding Author: Urmila Shrawankar, Department of Computer Science and Engineering, G H Raisoni College of Engineering, Nagpur, India

Basic fundamental filters: Low-pass, High-pass, Band-pass and Band-shop.

Adaptive filters (Mossi *et al*., 2010; Gorriz *et al*., 2009): Least Mean Squares (LMS) (Gorriz *et al*., 2009), Adaptive Least Mean Squares (ALMS), Normalized Least Mean Squares filter (NLMS), Echo Return Loss Enhancement (ERLE).

Normalization techniques (Veth and Boves, 1996; Hermansky and Morgan, 1994): Relative Spectral (RASTA) (Hermansky and Morgan, 1994) and Histogram Equalization (HE) (Veth and Boves, 1996).

Enhancement methods: Spectral Subtraction-Boll (1979), Spectral Subtraction-Berouti *et al*. (1979), The Generalized Spectral Subtraction-Sim *et al*. (1998), Multi-band Spectral subtraction *posterior*-Kamath and Loizou (2002), Wiener Filter a priori SNR-Wiener-Scalart (Breithaupt and Martin, 2010), MMSESTSA-Ephraim and Malah (1984; 1985), *posteriori* SNR-Ephraim and Malah (Ding *et al*., 2010; Ephraim and Malah, 1984), MMSE log-spectral Estimator-Cohen (2004).

This study uses all categories of environment to train the system so that word recognition accuracy will be increased. The categories of environment include different types of noise and SNR level like, clean, known SNR level noise, unknown SNR level noise in prior and real world environment. Artificial noise are taken from standard databases.

1.1. Speech Processing Models

The usability and popularity of any Speech User Interface (SUI) based application depends on the word accuracy rate shown by the application. It is very challenging to get good performance in adverse conditions like real world noisy environment. Therefore this work focuses on cleaning the speech signal and achieves the intangibility as much as possible at the back-end processing before passing the signal for further processing i.e., feature extraction model and followed by training and recognition.

1.2. Back-End Modeling

The speech processing starts from collecting the speech signals for the word samples. In the real world noisy environment, it is very difficult to get clean speech signals and we assume that these are corrupted with additive background noise and distorted due to adverse environmental conditions.

The speech processing back-end model helps to estimate the noise and distortion and tries to remove and enhance the signal quality with the help of filters and enhancement techniques (Shrawankar and Thakare, 2010a; Loizou and Kim, 2011).

1.3. Front-End Modeling

The signal processing front-end techniques play an important role in Speech Recognition Systems. These techniques convert the speech waveform to some type of parametric representation. This parametric representation is then used for further analysis and processing. Further these parameters are used for training and testing or recognition.

1.4. Front-end Techniques

There are many more methods available for extracting the characteristics of the signal, training the system and finally recognition. This system considers Mel Frequency Cepstral Coefficients (MFCC) (Motlicek, 2001; Indrebo *et al*., 2008; Yu *et al*., 2008) for the feature extraction, Bakis model of Hidden Markov Model (HMM) (Rabiner and Schafer, 1979; Rabiner, 1989; Sameti *et al*., 1998) for training and decoding and Viterbi algorithm for recognition.

1.5. Feature Extraction: MFCC

The use of Mel Frequency Cepstral Coefficients (MFCC) is considered for feature extraction (Motlicek, 2001). It is more suitable in speech recognition under a non-noise condition (Shrawankar and Thakare, 2010c). This method reduces the frequency information of the speech signal into small number of coefficients that emulate the separate critical bands in the basilar membrane of the ear, additionally, the logarithmic operation attempts to model loudness perception in the human auditory system.

1.6. Training: Hidden Markov Models (HMM)

Almost all Modern speech recognition systems are based on Hidden Markov Models (HMMs) (Rabiner, 1989; Sameti *et al*., 1998). This is a statistical model which gives us the *posterior* probability of an observed sequence of acoustic data given one or another word (or word sequence) enables to work out the most likely word sequence by the application of Bayes' rule (Rabiner and Schafer, 1979); this is called the Maximum a *Posterior* (MAP) criterion:

$$\Pr(word \mid acoustics) = \frac{\Pr(acoustics \mid word)\Pr(word)}{\Pr(acoustics)}$$

2. MATERIALS AND METHODS

This study is divided into two main stages.

The first stage is speech signal enhancement at back-end level and tested with SNR improvement test. It is completed and published (Shrawankar and Thakare, 2012). The brief summary is given in this section 3.1.

Second is word recognition accuracy improvement. The main aim is to clean the speech sample as much as possible with the help of hybrid enhancement method and train the system for all categories of environment.

Stage I:

This stage work is divided into following steps.

2.1. Step I: Samples Collection: Recording

The speech samples (words) are recorded with sampling rate of 8 kHz and time duration is set to 3 seconds. Samples are recorded from different speakers (male and female) and multiple utterances of ten isolated words (number digits 0-9) for each word from every speaker.

2.2. Step II: Adding Noise

Four categories of noise are considered:

- NOIZEUS database (www.utdallas.edu/~loizou/speech/noizeus)
- NOISEX database (www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html)
- Random Gaussian White Noise
- Real Environment (Natural) unknown noise (mixed environment)

2.3. Step III: Voice Activity Detector (VAD)

2.3.1. Windowing and Framing

The signal was divided into n-ms (milliseconds) frames with m% overlap between frames. The samples in the n-ms frame were used to construct a Toeplitz covariance matrix (Hu and Loizou, 2006).

2.4. Voice/Unvoiced/Silence Separation

The detection of speech presence is calculated by detecting the beginning and end-point of an utterance using VAD (Ramirez *et al*., 2007; Shrawankar and Thakare, 2010b). This two point detection algorithm is based on measures of the signal, zero crossing rate and short-time energy.

2.5. Step IV: Noise Cancellation

Under this Pre-emphasis step, filters are implemented to estimate and reduce or filter the noise.

In this experiment four fundamental traditional filters FIR like high-pass, low-pass, band-pass and band-shop filters are tested.

2.6. Step V: Enhancement

Speech enhancement algorithms attempt to recover a clean speech signal from a degraded signal containing additive noise. The evaluation of performance measures (Hu and Loizou, 2008; Ma *et al*., 2009) are performed using four category of noisy environments and eight speech enhancement algorithms encompassing different classes such as spectral subtractive, statistical-model-based (MMSE, log-MMSE and log-MMSE under signal presence uncertainty) and Wiener-filtering type algorithms (the a priori SNR estimation based method). The audible-noise suppression method is considered and tested for performance.

2.7. Step VI: Performance Evaluation

The performance measure is done using two performance measures: objective and subjective.

2.8. Objective Analysis: SNR Improvement Test

The Signal-to-Noise Ratio (SNR) improvement test is considered as an objective measure (Hu and Loizou, 2008; ITU-T P.832, 2000; Emiya *et al*., 2011; Ma *et al*., 2009). The SNR before and after implementing the filters are compared.

2.9. Subjective Analysis: Listening Test

The subjective quality evaluation is done by using a listening test (ITU-T P.832, 2000; Emiya *et al*., 2011; Hu and Loizou, 2006; Etame *et al*., 2011). The listening test was performed by normal hearing persons for observation of Overall quality (Intelligibility, Fidelity, Suppression), Musical noise salience, musical noise or other artifacts and Preference

2.10. Conclusion of Stage I

- In terms of overall quality and speech distortion, Priori SNR Wiener-Scalart 96, SS Berouti 79, SS Boll 79 and MBSS *posterior* Kamath algorithms performed the best
- Since the aim of this study is focused on environmental Unknown Natural Noise (Mixed Noise), Priori SNR Wiener, SS Berouti and MBSS *posterior* Kamath, proved to be the best solutions

Stage II:

This stage is further divided into the following steps.

2.11. Step I: Signal Enhancement Using Hybrid Methods

In the Stage I, all filters and enhancement methods are implemented independently. They were found effective

and showed performance improvement. Further, to enhance the performance of the system all enhancement methods are implemented with the combination of adaptive filters and Normalization methods.

Again the performance is observed using objective (SNR Improvement test) and subjective (informal listening test) parameters.

The Complete results are given in Results and Discussion Section.

It is observed that the combination of Basic filter + Adaptive Filter+Normalization method and Enhancement method is capable of removing additive background noise and noticeable distortion.

Now this improved enhanced signal (almost clean) is sent for front-end processing.

2.12. Step II: Feature Extraction

Next important task is feature extraction. Signal is windowed with a hamming window using a variable window length and the word is partitioned into small frames. The dimension of the frame is of variable size from 10 ms to 30 ms, with 40% overlap. Feature extraction is executed for each frame independently. The spectrum is calculated for each window using the FFT. The spectrum is then filtered with a special Mel-scaled filter bank to get corresponding Mel-coefficients. Single bands within the bank are typically triangular in form and overlapping one another. The logarithm of Mel-coefficients is then computed. The discrete cosine transform is used to transform into the cepstrum-space. Unnecessary (high-frequency) MFCC-coefficients are discarded and finally 20 MFCC coefficients are considered.

2.13. Step III: Training and Decoding

For training models, the method applied is based on Hidden Markov Models (HMM). A simple Markov model consists of a couple of states, with transition probabilities between states, which models discrete stochastic processes. In addition to simple Markov Models, each state of HMMs emits vectors with a specific distribution. During training phase, the number of states is defined (default equal to the number of letters of the word) and based on the training recordings, a model is decided for the emission of acoustic vectors and transition probabilities i.e., the discrete for isolated words. In the latter case, a Probability Density Function (PDF) for the distribution of emission vectors is found and is a mixture of Gaussians.

Some simplifications are made on the topology of the HMM. A HMMs Bakis model is considered for isolated word recognition, which simplifies the number of transition between states.

Training procedure is completed iteratively.

The next step is decoding. Viterbi algorithm is applied and the best path i.e., the path which has the highest probability is efficiently obtained. The probability is obtained from both emission and transition probabilities of the model. The value represents the probability of that model which corresponds to the observations. During this phase (training), the model is adjusted so that this probability increases. Considering the best path, the correspondence between each frame and each state gets modified. First consequence is the modification of transition probabilities. Second consequence is the modification of the input vectors. The following iterations begin with the new values for probabilities.

2.14. Step IV: Recognition

After the system is trained, actual recognition begins. Given an unknown observation, determine which model generated it with more probability. Front-end analysis is applied and the coefficients are extracted. Then the probabilities of correspondence between each model and the observations are computed. This is done using Viterbi algorithm. The model with the highest probability of compatibility is recognized.

Word recognition accuracy is calculated using formula:

$$\text{Word recognition rccuracy} = \frac{\text{Total Number of Words Recognised}}{\text{Total Number of Words tested}}\%$$

3. RESULTS AND DISCUSSION

Results are obtained in both the stages

The first is speech signal enhancement by using filters and enhancement methods independently on artificially added noise from the noise corpus.

SNR improvement test results are graphically shown below in **Fig. 1**:

- A Total of eleven (11) categories of noise is tested. As before SNR test results shows that White noise is most noisy and Volvo noise is less noisy
- Adaptive filtering algorithms alone are not able to clean the signal. Out of four filters LMS is better; therefore it is used further with combinations of enhancement methods
- In normalization technique RASTA gives good results and it is further implemented in the hybrid method

- In terms of overall quality and speech distortion as per SNR improvement and listening test, Priori SNR Wiener-Scalart 96, SS Berouti 79, SS Boll 79 and MBSS *posterior* Kamath algorithms performed well. These are selected for further implementation of hybrid methods
- Out of the four algorithms Wiener, Berouti, Boll and Kamath, Wiener is found to be the best

The second group consists of results of speech signal enhancement using independent filters.

SNR improvement test results are given shown in graph, **Fig. 2**. Samples are collected in real world noisy environment. Total of ten (10) types of different environmental noise is tested.

The following are the observations:

- Out of four Adaptive filtering algorithms LMS is better therefore it is used further with the combinations of enhancement methods
- In normalization technique RASTA has given a very good performance in the first two cases and average performance in rest of the cases
- In terms of overall quality and speech distortion as per SNR improvement and listening test, all spectral subtraction methods and wiener filtering methods show good performance
- Out of eight enhancement methods Berouti, Kamath, Wiener and Malah are giving good performance and Wiener is found to be the best

SNR improvement test is obtained using hybrid methods as shown in graph, **Fig. 3 and 4**.

First LMS is applied as an adaptive filter then filtered signal is passed to the RASTA for normalization and finally this processed signal is passed to all eight enhancement algorithms.

Following observations are noticed based on the results obtained from SNR test and listening test:

- Very slight improvement is observed with LMS. Better results are obtained in case of Volvo type of noise
- Consistent improvement is shown by RASTA in all cases
- All enhancement algorithms (Boh, Boll, Berouti, Kamath, Wiener, Malah) have performed well except Emhraim and Cohen
- Almost complete clean signal is obtained by Wiener.
- Wiener gave the best performance

Results for SNR improvement test obtained using hybrid enhancement methods for real world natural (mixed noise) environment.

Like result set in section 4.3, this method uses all combination of algorithms like LMS as an adaptive filter, RASTA as a Normalization technique and then eight enhancement algorithms. The following are the observations:

- In case of LMS very less improvement is noticed
- RASTA shows constant improvement in all cases
- Boh showed improvement only in two cases, in rest of the cases it is not effective
- Out of ten cases, Berouti showed good improvement in two cases, better improvement in another two cases and average improvement in the rest of the cases
- Kamth performs well in almost all cases except very few cases
- Malah's performance was good in almost all cases except three
- Wiener performed the best for all cases
- Ephraim and Cohen showed poor performance

To check the performance of hybrid method for artificially added noise samples, results are compared for SNR improvement with only Wiener enhancement method and SNR improvement with hybrid method. Results are shown in graphically in **Fig. 3**:

- Almost in all the cases hybrid method performs better
- Only in case of Bubble and Volvo type of noise wiener performs better

To check the performance of hybrid method for Real world environment, results are compared for SNR improvement with only Wiener enhancement method and SNR improvement with hybrid method. Results are shown graphically in **Fig. 4**.

In all the cases hybrid method performs better.

After testing SNR improvement for all categories of noise including real world environment (mixed) noise with the help of all independent filters and enhancement techniques, features are extracted and the system is trained.
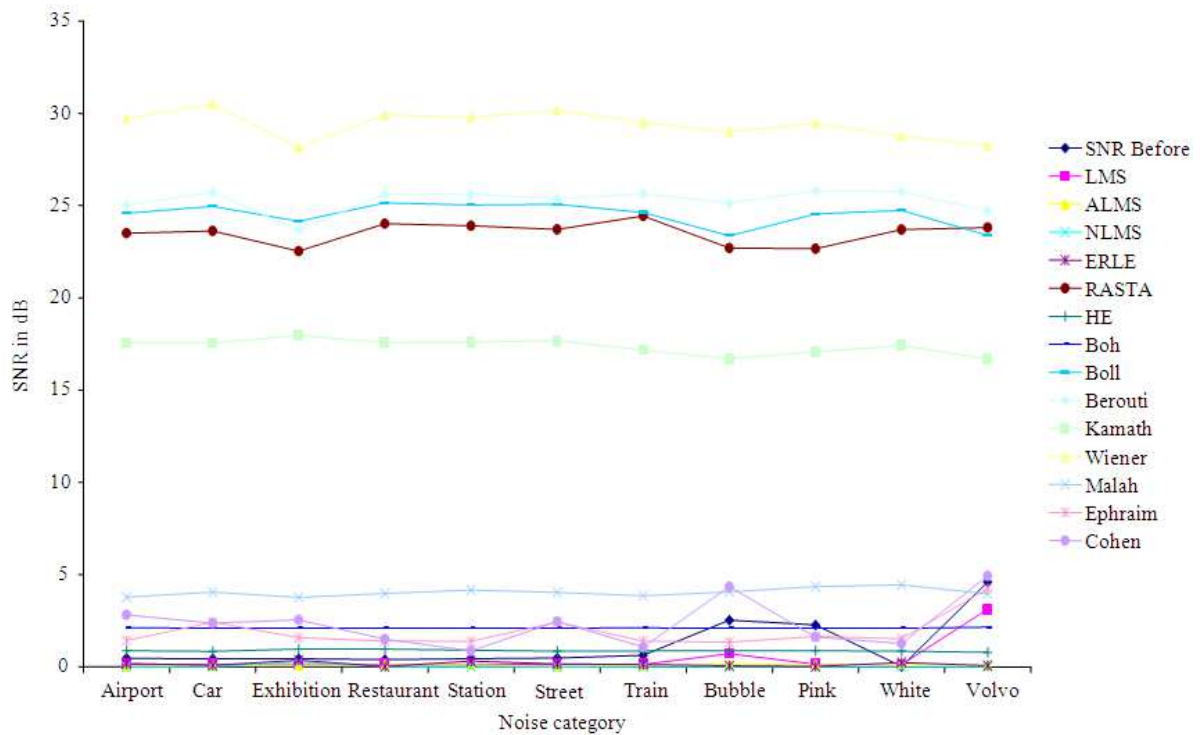
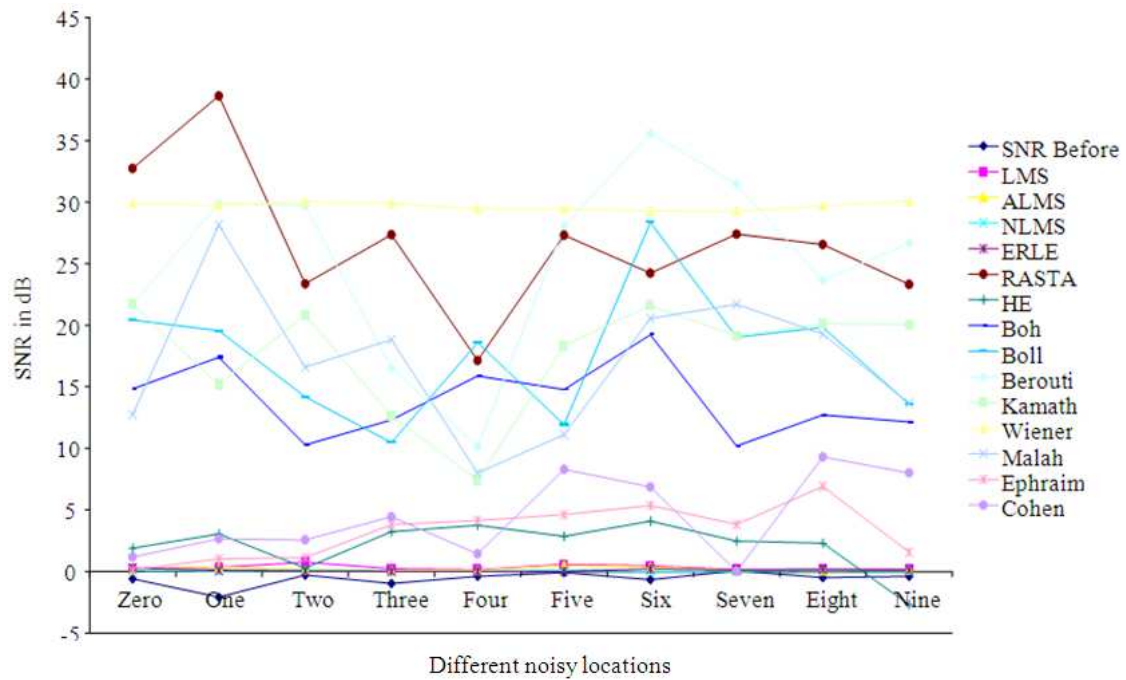Fig. 1. SNR improvement for artificial added noise using different filters

Fig. 2. SNR improvement for real world noisy environment using different filters
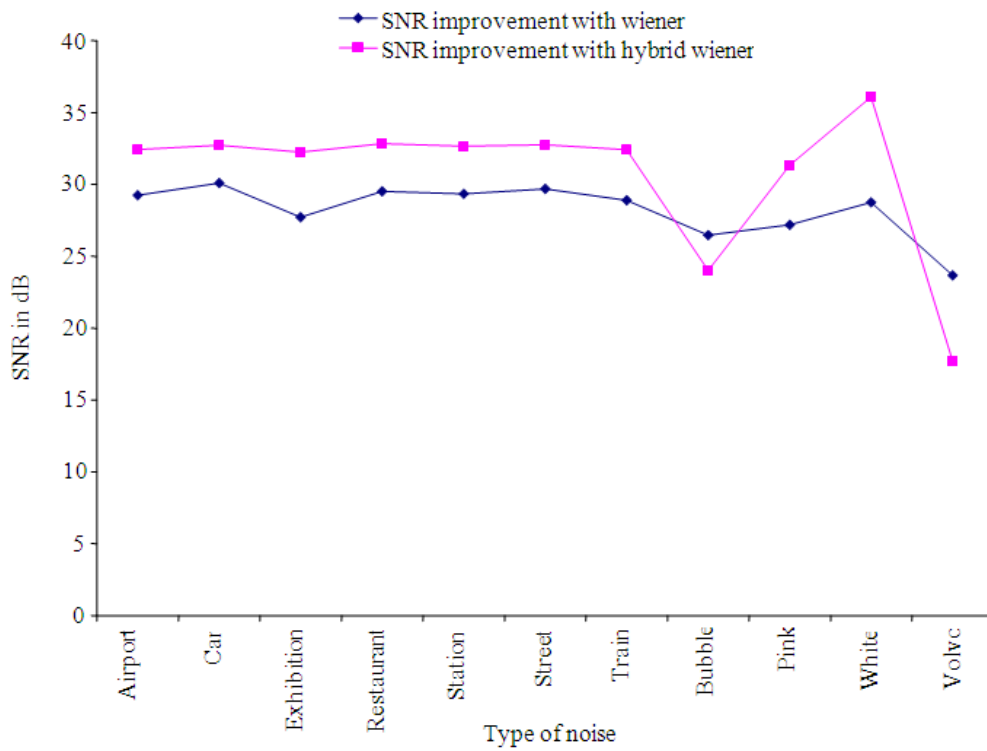
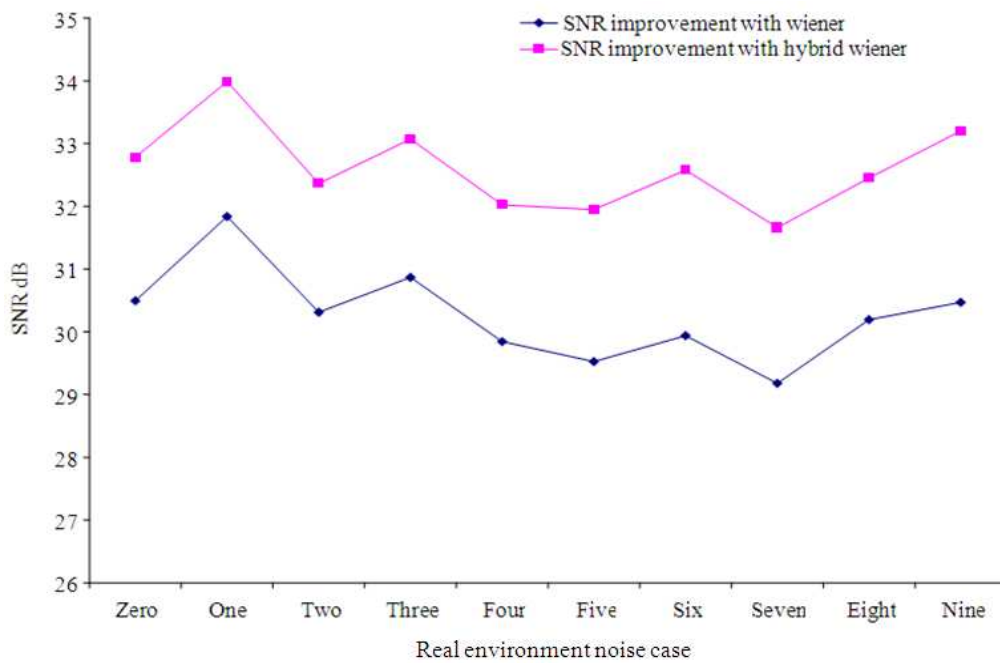**Fig. 3.** SNR improvement test results for artificial added noise

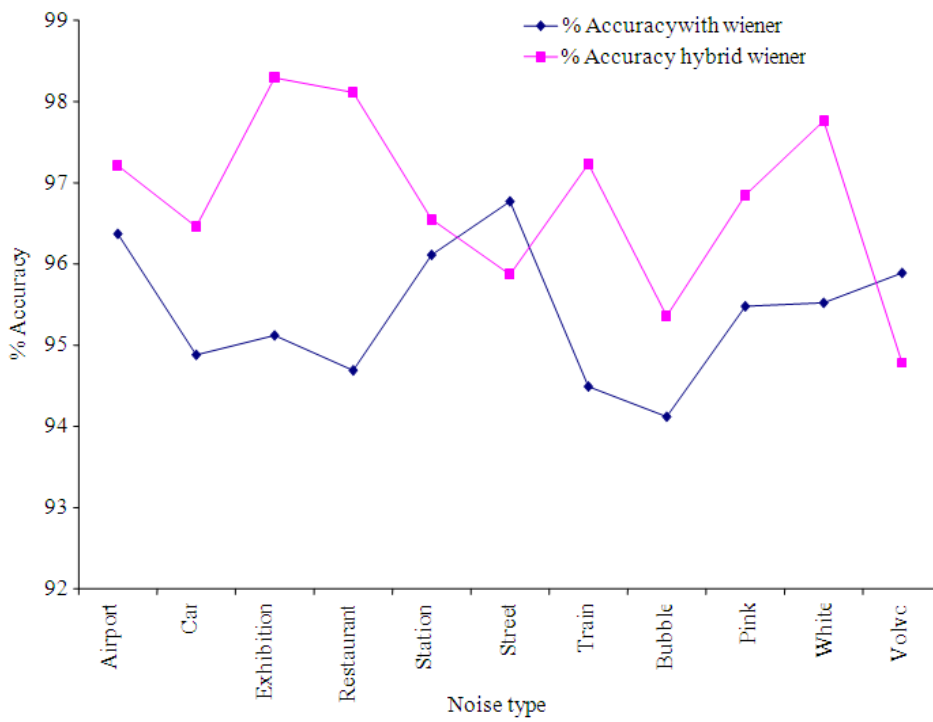**Fig. 4.** SNR improvement test results for real world environment noise

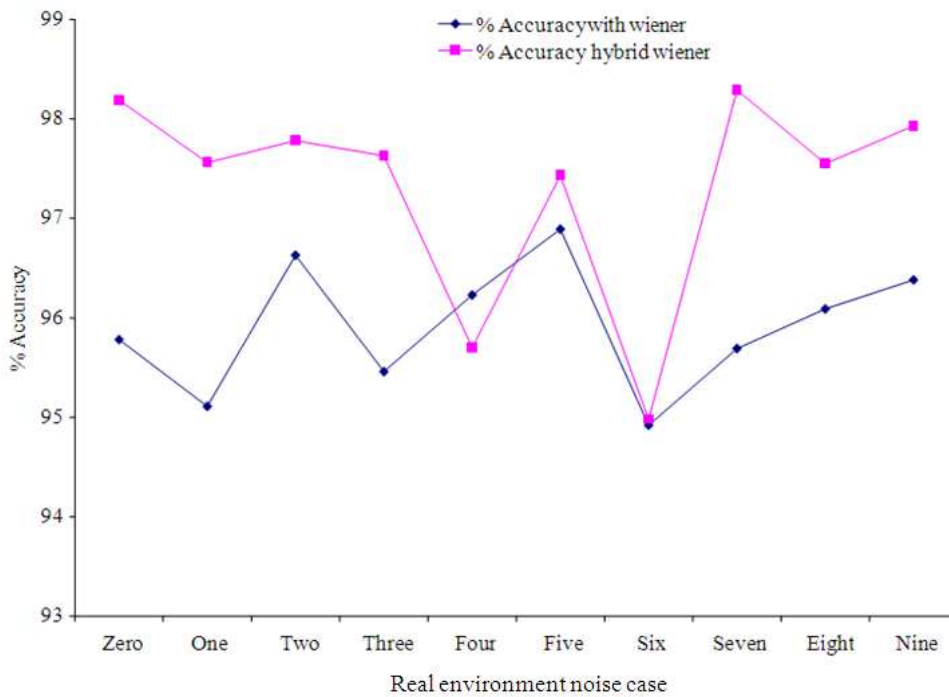**Fig. 5.** % Accuracy for artificial added noise

Fig. 6. % Accuracy for real world environment noise

Finally, word recognition accuracy is calculated for artificial noise and mixed noise separately; results are given graphically in **Fig. 5 and 6**:

- In case of artificial noise, in all cases hybrid method accuracy is better except street and Volvo type of noise
- In case of mixed noise, out of ten cases, only in one case hybrid method failed to improve accuracy, rest in all the cases hybrid method has given improved performance

4. CONCLUSION

- None of the noise filters or enhancement techniques can independently clean the signal with their intangibility. This was our assumption and it is confirmed
- Out of eight enhancement algorithms, Priori SNR Wiener-Scalart 96, SS Berouti 79, SS Boll 79 and MBSS *posterior* Kamath algorithms has shown good performance independently but could not totally clean the signal
- With the assumption that, noise category is unknown in case of mixed noise a prior, therefore LMS was used as an adaptive filter and RASTA as a normalization technique was implemented before implementing enhancement algorithm and this is found to be effective
- The hybrid method i.e., combination of one adaptive filter, one normalization technique and one enhancement method gives good performance in almost all the cases in presence of artificially added noise or mixed environment noise
- Hybrid method with Wiener filter showed consistently improved performance for both the tests i.e., SNR improvement test and accuracy improvement test
- Experimental results show significant improvement with hybrid method in word recognition accuracy in real world natural noisy (mixed noise) environment.
- Subjective evaluation listening test was found very helpful to confirm objective test results. Listeners noticed improvement in speech quality after implementing hybrid method
- Proposed hybrid method with artificially added noise case improves average word recognition accuracy by 1.3664%
- The average word recognition accuracy is improved by 1.3678% with the help of proposed hybrid method in real world noisy environment

- Thus hybrid method is one of the solutions to improve ASR performance in Real World Noisy Environment

5. REFERENCES

Berouti, M., R. Schwartz and J. Makhoul, 1979. Enhancement of speech corrupted by acoustic noise. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Apr. 2-4, IEEE Xplore Press, pp: 208-211. DOI: 10.1109/ICASSP.1979.1170788

Boll, S., 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process., 27: 113-120. DOI: 10.1109/TASSP.1979.1163209

Breithaupt, C. and R. Martin, 2010. Analysis of the decision-directed SNR estimator for speech enhancement with respect to low-SNR and transient conditions. IEEE Trans. Audio Speech Lang. Process., 19: 277-289. DOI: 10.1109/TASL.2010.2047681

Cohen, I., 2004. Modeling speech signals in the time-frequency domain using GARCH. Signal Process., 84: 2453-2459. DOI: 10.1016/j.sigpro.2004.09.001

Ding, H., I.Y. Soon and C.K. Yeo, 2010. Over-attenuated components regeneration for speech enhancement. IEEE Trans. Audio Speech Language Process., 18: 2004-2014. DOI: 10.1109/TASL.2010.2040792

Emiya, V., E. Vincent, N. Harlander and V. Hohmann, 2011. Subjective and objective quality assessment of audio source separation. IEEE Trans. Audio Speech Language Process., 19: 2046-2057. DOI: 10.1109/TASL.2011.2109381

Ephraim, Y. and D. Malah, 1984. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process., 32: 1109-1121. DOI: 10.1109/TASSP.1984.1164453

Ephraim, Y. and D. Malah, 1985. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process., 33: 443-445. DOI: 10.1109/TASSP.1985.1164550

Etame, T., R.L.B. Jeannes, C. Quinquis, L. Gros and G. Faucon, 2011. Towards a new reference impairment system in the subjective evaluation of speech codecs. IEEE Trans. Audio Speech Language Process., 19: 1301-1315. DOI: 10.1109/TASL.2010.2090145

Gorriz, J.M., J. Ramirez, S. Cruces-Alvarez, C.G. Puntonet and E.W. Lang *et al*., 2009. A novel LMS algorithm applied to adaptive noise cancellation. IEEE Signal Process. Lett., 16: 34-37. DOI: 10.1109/LSP.2008.2008584

Hermansky, H. and N. Morgan, 1994. RASTA processing of speech. IEEE Trans. Spech Audio Process., 2: 578-589. DOI: 10.1109/89.326616

Hu, Y. and P.C. Loizou, 2006. Subjective comparison of speech enhancement algorithms. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 14-19, IEEE Xplore Press, Toulouse. DOI: 10.1109/ICASSP.2006.1659980

Hu, Y. and P.C. Loizou, 2008. Evaluation of objective quality measures for speech enhancement. IEEE Trans. Audio Speech Language Process., 16: 229-238. DOI: 10.1109/TASL.2007.911054

Indrebo, K.M., R.J. Povinelli and M.T. Johnson, 2008. Minimum mean-squared error estimation of mel-frequency cepstral coefficients using a novel distortion model. IEEE Trans. Audio, Speech Language Process., 16: 1654-1661. DOI: 10.1109/TASL.2008.2002083

ITU-T P.832, 2000. Methods for objective and subjective assessment of quality.

Kamath, S. and P. Loizou, 2002. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. Preceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 13-17, IEEE Xplore Press, Orlando, FL, USA., pp: IV-4164-IV-4164. DOI: 10.1109/ICASSP.2002.5745591

Laska, B.N.M., M. Bolic and R.A. Goubran, 2010. Particle filter enhancement of speech spectral amplitudes. IEEE Trans. Audio Speech Language Process., 18: 2155-2167. DOI: 10.1109/TASL.2010.2042127

Loizou, P.C. and G. Kim, 2011. Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. IEEE Trans. Acoust. Speech Signal Process., 19: 47-56. DOI: 10.1109/TASL.2010.2045180

Ma, J., Y. Hu and P.C. Loizou, 2009. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. J. Acoust. Soc. Am., 125: 3387-3405. DOI: 10.1121/1.3097493

Mossi, M., C. Yemdji, N.W.D. Evans and C. Beaugeant, 2010. Acoustic echo cancellation in non-linear and noisy environment. Department of Communications Multimedia.

Motlicek, P., 2001. Feature extraction in speech coding and recognition. Oregon Graduate Institute of Science and Technology.

Rabiner, L.R. and R.W. Schafer, 1979. Digital Processing of Speech Signals. 1st Edn., Pearson Education India, Englewood Cliffs, ISBN-10: 8131705137, pp: 528.

Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. IEEE Proc., 77: 257-286. DOI: 10.1109/5.18626

Ramirez, J., J.M. Gorriz and J.C. Segura, 2007. Voice activity detection. Fundamentals and speech recognition system robustness. University of Granada, Spain.

Sameti, H., H. Sheikhzadeh, L. Deng and R. Brennan, 1998. HMM-based strategies for enhancement of speech signals embedded in nonstationary noise. IEEE Trans. Speech Audio Process., 6: 445-455. DOI: 10.1109/89.709670

Shrawankar, U. and V.M. Thakare, 2010a. Noise estimation and noise removal techniques for speech recognition in adverse environment. Proceedings of the 6th International Conference on Intelligent Information Processing, Oct. 13-16, Springer Berlin Heidelberg, Manchester, UK., pp: 336-342. DOI: 10.1007/978-3-642-16327-2_40

Shrawankar, U. and V.M. Thakare, 2010b. Voice activity detector and noise trackers for speech recognition system in noisy environment. Int. J. Adv. Comput. Technol., 2: 107-114.

Shrawankar, U. and V. Thakare, 2010c. Feature extraction for a speech recognition system in noisy environment: A study. Proceedings of the 2nd International Conference on Computer Engineering and Applications, Mar. 19-21, IEEE Xplore Press, Bali Island, pp: 358-361. DOI: 10.1109/ICCEA.2010.76

Shrawankar, U. and V.M. Thakare, 2012, Performance Analysis of Noise Filters and Speech Enhancement Techniques in Adverse Mixed Noisy Environment for HCI. Int. J. Res. Rev. Comput. Sci., 3: 1817-1825.

Sim, B.L., Y.C. Tong, J.S. Chang and C.T. Tan, 1998. A parametric formulation of the generalized spectral subtraction method. IEEE Trans. Speech Audio Process., 6: 328-337. DOI: 10.1109/89.701361

Veth, J.D. and L. Boves, 1996. Comparison of channel normalisation techniques for automatic speech recognition over the phone. Proceedings of the 4th International Conference on Spoken Language, Oct. 3-6, IEEE Xplore Press, Philadelphia, PA., pp: 2332-2335. DOI: 10.1109/ICSLP.1996.607275

Yu, D., L. Deng, J. Droppo, J. Wu and Y. Gong *et al.*, 2008. A minimum-mean-square-error noise reduction algorithm on Mel-frequency cepstra for robust speech recognition. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Mar. 31-Apr. 4, IEEE Xplore Press, Las Vegas, NV., pp: 4041-4044. DOI: 10.1109/ICASSP.2008.4518541