VOL. 7, NO. 7, JULY 2012 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

A TAXONOMY-ORIENTED OVERVIEW OF NOISE COMPENSATION TECHNIQUES FOR SPEECH RECOGNITION

Syed Abbas Ali[1], Najmi Ghani Haider[2] and Mahmood Khan Pathan[2]
[1]Computer and Information Systems Engineering, N.E.D University of Engineering and Technology, Pakistan
[2]Computer Science and Information Technology, N.E.D University of Engineering and Technology, Pakistan
E-Mail: saaj@neduet.edu.pk

ABSTRACT

Designing a machine that is capable for understanding human speech and responds properly to speech utterance or spoken language has intrigued speech research community for centuries. Among others, one of the fundamental problems to building speech recognition system is acoustic noise. The performance of speech recognition system significantly degrades in the presence of ambient noise. Background noise not only causes high level mismatch between training and testing conditions due to unseen environment but also decreases the discriminating ability of the acoustic model between speech utterances by increasing the associated uncertainty of speech. This paper presents a brief survey on different approaches to robust speech recognition. The objective of this review paper is to analyze the effect of noise on speech recognition, provide quantitative analysis of well-known noise compensation techniques used in the various approaches to robust speech recognition and present a taxonomy-oriented overview of noise compensation techniques.

Keywords: noise compensation techniques, speech recognition, noise robustness.

1. INTRODUCTION

In most practical applications of automatic speech recognition, the input speech is contaminated by background noise. This strongly degrades the performance of speech recognizers [1]. Noise is unpredictable, time-varying and has temporal characteristics in nature. One of the significant issue in robust speech recognition is to develop an accurate noise model, whereas noise estimation itself a difficult problem. The non-linear interaction between noise and clean speech in producing noise corrupted speech generate high degree of imperfection and complexity in decoding speech. To become an integral part of real world applications, speech recognition system maintain a significant level of recognition accuracy in difficult and time varying acoustic environment. Different approaches have been applied in reducing the effects of noise on the acoustic speech signal captured through microphone, outcomes of several studies have established that the speaker independent speech recognition system performance degrades dramatically, when training condition differ from the testing environment [2-4]. Application of speech recognition in different environments such as telephonic conversation, automobile, industries, cocktail party, or in office setup needs higher degree of environmental robustness. This paper presents an overview of the commonly used noise compensation techniques in the various approaches to robust speech recognition and identifies some research issues that need to be addressed. Rest of the paper is organized as follows. The subsequent section briefly review model of environmental distortion including the additive noise. This is followed by a brief discussion of noise effect on speech recognition. Section 3 then summarize the well-known noise compensation techniques used in the various approaches to robust speech recognition in qualitative manner. Section 4 discusses taxonomy-oriented overview of noise compensation techniques using two different axes and some research activities in the field of robust speech recognition.

2. A MODEL OF THE ENVIRONMENT

One major concern in the design of the speech recognition systems is its performance in real environment. In such conditions, different sources could generally be classified as additive noise and channel distortion. Noise is inherently unpredictable. Fortunately, noise may be approximately characterized by an environmental acoustic model. This is summarized in a model form [5] shown in Figure-1.

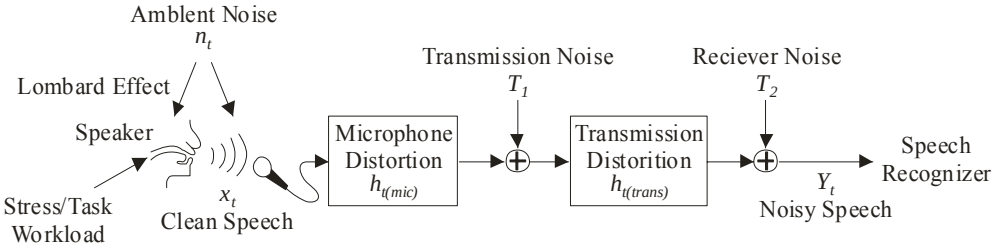

Figure-1. Sources of noise and distortion.

In the model given in Figure-1, a major source of corrupted noise is the additive, ambient environment noise, n_t, present when the user is speaking. The combined noise and speech signal is then captured and filtered by microphone impulse response, $h_{t,(mic)}$ which can be another large source of distortion. Transmission may also add noise, represented by T_1 and $h_{t\ (trans)}$, although it is expected to be small. Noise at the receiver side $T_2$is also expected to be minimal. Figure-1 may be simplified by combining the various additive and convolute noise sources into a single ambient noise, n_t, and linear channel noise, h_t, variables. A simplified oft-used model [6, 7, 8] of the noisy acoustic environment in the time domain is shown in Figure-2. A speech signal x_t in the time domain is contaminated by ambient noise n_t and a stationary convolution channel impulse response h_t to give noisy signal Y_t, the resulting noisy signal will become:

$$Y_t = x_t * h_t + n_t \quad (1)$$

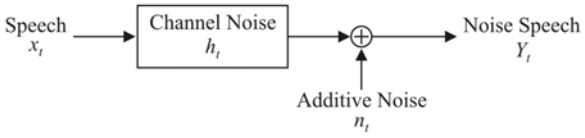

Figure-2. Model of the noisy acoustic environment.

The effect of noise on the input signal can be represented in the form of log-spectral or cepstral domain. In the log-spectral or cepstral domain Eq. (1) leads us to the following relationship between speech, noise and noisy speech.

$$Y_t = x_t + h + \mathrm{C}\log(1 + \exp(\mathrm{C}^{-1}(n_t - x_t - h)) \quad (2)$$

$$Y_t = x_t + \mathrm{f}(x_t, n_t, h),$$

Where f (x_t, n_t, h) is referred to as environment function, log and exponential functions indicate element-wise operations that yield a vector of the same dimensionality as he input vector. Eq. (2) indicates that the noisy speech is a composite non-linear function of the channel noise, additive noise and speech. There are following advantages to performing noise compensation in log-spectral or cepstral domain are smaller numbers of parameters need to be estimated, log-spectral or cesptral based features are used in existing speech recognition system, statistical models are more accurate and easily developed in the cepstral or log-spectral domain.

2.1 Effects of noise on speech

There are number of sources responsible for acoustic contamination that can reduce the accuracy of speech recognizer. The main sources of speech variation can be classified into three main categories [9]. In the first category, the recorded speech is the sum of the speech produced by user and the acoustic ambient noise. It is generally a colored noise and the noise structures can vary according to sources such as babble noise, office environment, industrial noises, etc. Second category of speech variation related to the convolution of the speech signal. They can be produced by mounting position and types of microphone, make use of different microphone for testing and training and room reverberation, etc. In the third category of speech variation, the user can be affected by factors like stress, mental state and emotions in his speaking style. When user speaks under high noise conditions, they change their utterance like pitch, sound intensity, sound duration and formant frequencies, etc. This Lombard effect degrades the performance of speech recognizer.

Most of the research in robust speech recognition has been focused toward compensation for the effects of ambient noise. Above mentioned discussion leads us toward the development of noise compensation strategies to cope with ambient noise. Before going into the development of the noise compensation techniques, there is a need to address some of the important issues related to speech and noise [10].

- Speech patterns are represented by a sequence of feature vectors rather than single feature vector. The complicated HMM is used to model the temporal dynamics of speech.
- Noise is unpredictable, time -varying and has temporal characteristics in nature.
- In speech recognition, with N classes, the features are projected into N-dimensional vectors in log-likelihood domain. It is difficult to carry study unless we consider the correct and competing classes.
- The assumption that the noise term is additive and independent from speech is not true in real feature extraction of speech recognition systems, such as (MFCC) Cepstral domain. The relation between noise and speech is highly nonlinear [6].

With the above challenges among many others, it is mathematically difficult to study the noise effect for practical speech recognition system. As the SNR decreases, the noise component in SNR starts to dominate and the mean values of the noisy speech shifted toward noise. As noise dominates in the speech component the value of SNR suppressed significantly. The net effect is that the training of speech data set in clean environment doesn't match with the testing environment and recognition rates decrease rapidly [11].

3. APPROACHES TO NOISE ROBUSTNESS

The acoustic model of a speech recognition system is trained on clean speech data set and the decision boundary of the model fits well to the distribution of the clean training data. When the speech corrupted by noise, the statistical characteristics of the noisy speech will be different from clean speech. The decision boundary that fit to clean speech may not fit for noisy speech and recognition rate will degrade. To overcome this problem and improve the robustness of speech recognition against noise distortion, many methods have been proposed to

decreasing the acoustic mismatch between testing and training conditions. These techniques can be classified into two distinct approaches shown in Figure-3. Feature compensation methods and Model compensation methods.

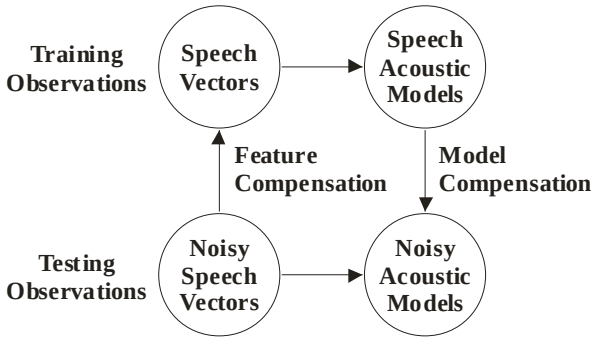

Figure-3. Approaches to noise robust recognition.

3.1 Robust feature extraction techniques

The main goal of the feature extraction is to find a set of parameters to represent speech signal in the ASR system which are robust against the variation in the speech signal due to ambient noise or channel distortion. The extensive research has resulted in such well known techniques, this section provide the possible robust feature extraction techniques describe in literature like RASTA filtering [12], Cepstral mean normalization [13], Dynamic spectral features [14], Short time modified coherence [15], One-sided autocorrelation LPC [16], Differential Power spectrum [17], Relative autocorrelation Sequence [18] as shown in Table-1.

Table-1. Feature extraction techniques.

Relative auto-correlation sequence (RAS)	Yuo and Wang, 1998
Cepstral mean normalization (CMN)	Kermorvant, 1999
One-sided auto-correlation LPC (OSALPC)	Hernando and Nadeu, 1997
Short time modified coherence (SMC)	Mansour and Juang, 1989
Dynamic spectral features	Furui, 1986
Differential power spectrum (DPS)	Chen *et al*., 2003
RASTA filtering	Morgan and Hermansky, 1994

3.2. Feature compensation techniques

The primary goal of the feature compensation technique is to suppress the effect of noise in the extracted features which contaminates speech signal. The widely used speech enhancement techniques in the presented work by researchers are Spectral Subtraction [19, 20], Wiener Filtering (for stereo type data, SPLICE Algorithm) [21], Approach-II (Non-stereo data, VTS expansion) [8], Cepstral mean normalization (CMN and MVN) [22, 23], Histogram Equalization (HEQ) [24] and RASTA Temporal filter [12] summarized in Table-2.

Table-2. Feature compensation techniques.

Spectral subtraction	This technique provides magnitude estimation of speech by explicitly subtracting the noise magnitude spectrum from the noisy magnitude spectrum.
Wiener filtering (Approach -I) for stereo type data. SPLICE algorithm	In this approach, SPLICE algorithm computes minimum mean square error (MMSE) estimate for stereo data type.
Approach -II (Non- stereo data). First-Order VTS expansion	This type of approach dealing with the non-stereo data and make use of first - order vector Taylor series (VTS) to establish the relationship between noise, speech and noisy speech.
Cepstral mean normalization (CMN) and (MVN)	CMN and CVN normalize first and second moment of probability distribution of speech feature. Recent research enhance this basic idea for higher order moment of probability distribution of speech feature s.
Histogram equalization (HEQ)	This approach make use of transformation mechanism in which the distribution of test speech map on the pre-defined distribution, utilizing the relation-ship between the CDF of test speech and those of referenced speech (training). Histogram Equalization normalizes all the moments of probability distribution of test speech feature to those of the training ones.
RASTA temporal filter	The basic working of the RASTA Filter is to suppress the spectral components that change more quickly or slowly than the rate of change of speech.

3.3 Model-based noise compensation

The main objective of the model compensation is to compensate the acoustic model to match noisy environment. Modify the recognition models parameters such as means and variance to improve the recognition rates between training and testing conditions. The model compensation has great potential to improve robustness and it makes use of detailed knowledge of the underlying clean speech encoded in the acoustic models. Model-based noise compensation is merging the clean speech model with the noise model including single Gaussian noise and multi Gaussian mode for highly varying noise conditions. In the frame work of model-based noise compensation, statistical models such as Hidden Markov Model (HMMs) are consider to remove the mismatch between the training model and the noisy speech to improve the performance of ASR systems. Model based compensation performs adjustment of model parameters in order to obtain a model appropriate for recognition in the noisy environment. Due to unpredictable nature of noise, it is not possible to account for all conditions that may be encountered by including them in the training data. Thus other acoustic model compensation methods that updates the model parameters may be categorized as either; 1) Predictive: In the predictive method the speech model is merge with the noise model to generate noisy speech model using acoustic environment model. 2) Adaptive: In adaptive method enough noisy speech data are available to update the acoustic model to match the noisy speech observations. MAP [25] and MLLR-style [26] transforms can be measured as adaptive forms; on the other hand PMC [27] and VTS [28] are predictive techniques. The widely used approaches for model-based noise compensation are as follows:

- **Matched-style training:** In matched-style training a new sample of the waveform taking from the new environment, and merge it to all the utterance in the existing training databases without making any change in the system. If noise characteristics are known beforehand, using this method, we can adopt the new environment with fewer amounts of data from the new environment.

- **Multi-style training:** Multi-style compensation crates an artificial acoustic environment by contaminating the speech training databases with noise samples of varying levels (05 dB, 10 dB, etc.) and types (babble, car, etc.).
- **Model adaptation:** In the real world scenario, there will always be new speaker and unseen environment. The solution to this problem is adaptation. The purpose of adaptation is to permits a less amount of data from user to transform an acoustic model set in to compact recognition parameters set and this recognition parameters set an be used to reduce mismatch between training and testing conditions. Different adaptation methodologies have been suggested in the literature. Many of these approaches can be used with in a speaker adaptive training (SAT) frame work. Some of the commonly used schemes are summarized in Table-3.

Table-3. Model adaptation schemes.

Feature-based schemes	Feature-based method based on the acoustic features. There are three basic approaches define in this scheme: Mean and Variance normalization; Gaussianisation; Vocal tract length normalization (VTLN)
Linear transform-based schemes	One of the most effective forms of adaptation is based on linear transformation methods. Linear transformation makes use of model parameters and requires a transcription of the adaptation data. Most commonly use approaches are MLLR. MLLR works under the assumption of likelihood maximization of adapted data and linear transforms-based schemes are used to map an existing model parameters set into a new adapted model parameters set.
Gender/cluster -dependent model (Cluster adaptive training CAT)	Make use of multiple clusters instead of the single cluster to represent the user. There are two approaches are used to combining the cluster 1) Likelihood combination and 2) parameters combination. PCA used to determine the cluster mean and EM-algorithm implemented for the model training.
MAP adaptation	The basic concept of Model adaptation is to use the standard statistical methods to find robust parameters estimation instead of assuming a concept of transformation to signify the difference among users.

- **Parallel Model Compensation (PMC):** ("The aim of the model compensation can be viewed as obtaining the parameters of speech plus noise distribution from the clean speech mode and the noise model. One approach to model-based compensation is parallel model compensation [27]. Standard PMC assume that the feature vectors are linear transforms of the log spectrum, such as MFCC, and that the noise is primarily additive. The basic idea of the algorithm is to map the Gaussian means and variance in the cepstral domain, back into the linear domain, where the noise is additive, compute the means and variances of the new combined speech plus noise distribution, and then map back to cepstral domain") [11]. The parameters of the noisy speech distribution, $N(\mu_n, \sum_n)$ can be found from

$$\mu_n = E\{Y_t\}$$

$$\sum_n = E\{Y_t Y_t^T\} - \mu_Y \mu_Y^T$$

Where Y_t is the noisy speech acquired from a clean speech component combined with noise from the noise model. Various approximations have been proposed [11], because there is no direct solution of these equations is available. Many speech recognizer are based on the HMM and use

MFCC Features. The calculation of MFCC involves the log operation, the relation between noise, speech and noise-corrupted speech becomes non-linear. The basic objective of the model based compensation is to find speech model in noisy conditions and approximation of this non-linear relationship. Commonly used approaches to solving this problem [29] are shown in Table-4.

Table-4. Approximation techniques.

Log-normal approximation and log-add Approximation.	("The distribution in the linear domain will be log- normal but the sum of two log normal distribution is not log normal. Standard PMC ignores this and assume that the combined distribution is log-normal. This is referred as log- normal distribution/ A simpler, faster, PMC approximation is to ignore the variance of the speech and noise distribution. This is referred as log-add approximation") [11].
Schwartz-Yeh approximation	This method works under the assumption that resulting distribution is normal but should be more accurate than the log -normal approximation. Calculations are performed in log-spectral or cepstral domain and there is no transformation to the linear domain is needed. Method has high computational complexity than log-normal approximation.
Lagrange polynomial Approximation	Lagrange polynomials were used to approximate mean parameters for noisy speech.
Monte-Carlo Techniques	A high mapping accuracy can be attained using Monte-Carlo techniques to estimate new distribution by combining the sample of noise and speech distribution. Data Driven PMC obtain mean and co-variances matrix through Monte-Carlo simulation

- **Vector Taylor Series (VTS):** PMC and VTS are the main methods for acoustic model compensation and adaptation. Because the non-linear nature of the corrupted model for cepstral parameters under the condition of additive noise; some approximation need to be made to facilitate efficient computation. PMC uses Log-normal approximation for this non-linear model.

VTS approximates the non-linear model with its first -order vector Taylor series expansion and transform it into linear one. Like PMC, the noisy speech model is produced by merging of speech HMM and the noise HMM. Unlike PMC, the VTS approach combines the parameters of speech HMM and noise HMM linearly in the cesptral domain [8, 28, 30].

4. A TAXONOMY-ORIENTED OVERVIEW

Noise compensation techniques have been developed by speech community during last two decades. Model compensation or model-domain approaches are used to update the HMM model parameters such as means and variances, while the feature enhancement approaches are used to reduce the noise distortion from the speech feature vectors. Model techniques can be categorized into two main classes based on two different approaches. In the first category, Unstructured or linear transformation are used to convert model parameters. These techniques are applicable to noise compensation as well as speaker adaptation. They required number of parameters and large amount of data sample to estimate model parameters. Commonly used algorithms in this category can be summarized as Maximum Likelihood Linear Regression (MLLR) [26], Maximum a Posteriori (MAP) [25], Constrained MLLR [31], Noisy constrained MLLR [32], Multi-style Training [33]. In the second category, structured or non-linear transformation which takes into account the way the noisy speech features (Cepstra) are produced from the mixing speech and noise. Techniques of this category not used for other types of acoustic variation and physical knowledge are used to approximate the mixing process of noise and speech. Common techniques in this category include Parallel model combination (PMC) [27], Vector Taylor series (VTS) [28] and Phase sensitive model compensation [34, 35].

The feature compensation based on the same schemes as it counterpart. In the first category, structured feature compensation use similar structured transformation as describe for model compensation. Some of the commonly used techniques in this category include Vector Taylor Series (VTS) [8], Algonquin [36] and Phase-Sensitive model for feature enhancement [37]. Second category is unstructured feature domain compensation, in which techniques are developed and don't provide any structured knowledge of how noise and speech mix expressed in the log domain. This category includes

several feature normalization methods. Commonly used techniques in this category are SPLICE [21], Spectral subtraction [19, 20], CMN [22], MVN [23], HEQ [24] and RASTA [12]. A grouping of noise compensation techniques is shown in Table-5.

Table-5. A grouping of noise compensation techniques.

	Model-based	**Feature-based**
Linear transformation	MAP, C-MLLR, MLLR, N-CMLLR, multi-style training	Spectral subtraction, SPLICE, HEQ, RASTA wiener filter, MMSE, MMSE-Cepstra, MVN, CMN
Non-linear transformation	VTS, PMC, phase-sensitive model	VTS, Algonquin, phase-sensitive model

In spite of major progress in robust speech recognition during last two decades, most of the well-known noise compensation techniques (Feature-domain and Model-domain) have been quantitatively summarized in this paper but the problem is not being solved. Speech research community intensively put their effort in this field of research to improve the accuracy of the real-world speech recognition problem and applications under different acoustic conditions. Here, a brief discussion from the authors prospective on the research activities in the field of robust speech recognition.

First, despite of benefit performing noise compensation in cepstral domain such as smaller numbers of parameters estimation, Log-spectral based features are widely used in the existing speech recognition systems and statistical models are more accurate and easily developed in the log-spectral or cepstral domain, the relation between speech, noise and noisy speech is a complex non-linear function of the channel, clean speech and noise in log-spectral or cepstral domain. Two aspects of the mentioned issue are needed to be addressed. i) How to approximate this relationship and ii) how to improve the exciting approaches in order to achieve good recognition performance.

Second, acoustic models have critical importance in speech technology. Speech research communities are trying to focus two major aspect of acoustic modeling: 1) how to develop the statistical models and derive their structures and 2) how to learn these structures automatically from data. We know that as the level of ambient noise increases, the associated uncertainty of speech increases accordingly. One of the promising research activities in this direction is to learn the structure of the acoustic sounds to enhance the generalization capability of the acoustic model being used for speech, noise and their interaction.

Third, with reference to above discussion on improved acoustic modeling for speech and noise interaction, two aspect of acoustic modeling 1) speed of parameter estimation and 2) improving discriminative power are gaining significant important among speech research community. Improving the discriminative power for speech recognition system, one of the prominent issues with speaker independent speech recognition system is that the acoustic model trained on large data set has to waste a large number of parameters for recognizing the inconsistency among users rather than desire words. On other hand discriminative and margin based training methods can be used to improve speed of parameter estimation by reducing the empirical risk and enhancing the generalization capability of acoustic model.

Fourth, front end features may not always be strict log-spectra. Whereas, most of the speech recognizers make use of cepstra based on which all nonlinear structured acoustic distortion models are developed and approximated. PLP features, discriminatively derived feature, Mean and variance normalization often performs better than plain MFCC. There is a need to rethinking of new emerging technology from deep machine learning, which can provide the opportunity to automatically derive features from the raw data.

5. CONCLUSIONS

The noise effect is reduced by either feature compensation methods or model compensation methods. In this review paper, we revisited the well-known noise compensation techniques used in the various approaches to noise robust speech recognition and analyzed the effects of noise on speech recognition. In feature based noise compensation, the noisy features are compensated to eliminate the effects of noise, where as in model based noise compensation the clean acoustic model are compensated to match the noisy environment. Feature compensation methods are simpler, easy to implement and computationally efficient. In contrast, Model-compensation methods have potential for greater robustness but they are computationally very expensive. Model-based noise compensation methods that update the model parameters may be categorized as either Adaptive forms such as MAP or MLLR may be used to compensate for noise and predictive forms such as VTS and PMC have the advantage that only a noise model of environment is necessary to compensate the systems. Noise is inherently unpredictable, it may be characterized as additive and convolutional noise components, leading to noise

compensation techniques that can help Automatic Speech Recognition handle adverse environments.

In this review paper, we presented taxonomy-oriented overview of noise compensation techniques for speech recognition using two different axes, feature domain vs. model domain and linear vs. non-linear transformation and discussed some research activities in the field of robust speech recognition where difficulty of the task and variety of acoustic conditions surges.

REFERENCES

[1] Y. Gong. 1995. Speech recognition in noisy environment: A survey. Speech Communication. 16(3): 261-291.

[2] A. Acero. 1993. Acoustic and Environmental Robustness in Automatic Speech Recognition. Kluwer Academic Publishers, Boston, M.A.

[3] B. H. Juang. Speech Recognition in Adverse Environments. Computer Speech and Language. 5: 275-294.

[4] R.M. Stern, A. Acero, F. H. Liu and Y. Ohshima. Signal Processing for Robust Speech Recognition. In: Speech Recognition, C. H. Lee and F. Soong, Eds., Boston: Kluwer Academic Publishers.

[5] J.H.L. Hansen. 1996. Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition. Speech Communication. 20(2): 151-170.

[6] A. Acero. 1990. Acoustical and Environmental Robustness in Automatic Speech Recognition. PhD thesis, Carnegie Mellon University, USA.

[7] M.J.F. Gales. 1995. Model-Based Techniques for Noise Robust Speech Recognition. PhD thesis, Cambridge University, U.K.

[8] P.J. Moreno. 1996. Speech Recognition in Noisy Environments. PhD thesis, Carnegie Mellon University, USA.

[9] Jean-Pal Haton. 2004. Automatic Speech Recognition: A Review. Camp *et al* (Eds), Enterprise Information Systems. Kluwer Academic Publishers. 5: 6-11.

[10] Xiong Xiao, J. Li, Eng Siong Chng, H. Li and Chin-Hui Lee. 2010. A study on the generalization capability of Acoustic modes for Robust Speech Recognition. IEEE Trans on Audio, Speech and Language Processing. 18(6).

[11] J.F. Gales and S. Young. 2008. The Application of Hidden Markov Models in speech Recognition. Foundation and trends in Signal Processing. 1(3) (2007): 195-304.

[12] H. Hermansky and N. Morgan. 1994. RASTA processing of speech. IEEE Trans. Speech Audio processing. 2(4): 578-589.

[13] C. Kermorvant. 1999. A comparison of noise reduction techniques for robust speech recognition. IDIAP-RR99-10.

[14] S. Furui. 1986. Speaker- independent isolated word recognition using dynamic features of speech spectrum. IEEE Trans. on Acoustics, Speech and Signal Processing. 34(1): 52-59.

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

[16] J. Hernando and C. Nadeu. 1994. Linear prediction of one-sided autocorrelation sequence for noisy speech recognition. IEEE Trans. Speech Audio Processing. 5(1): 80-84.

[17] J. Chen, K.K. Paliwal and S. Nakamura. 2003. Cepstrum derived from differentiated power spectrum for robust speech recognition. Speech Communication. 41(2-3): 469-484.

[18] K.-H. You, H.-C. Wang. 1998. Robust feature derived from temporal trajectory filtering for speech recognition under the corruption of additive and convolutional noise. Proc. ICASSP. pp. 577-580.

[19] J. Beh and H. Ko. 2003. A novel spectral subtraction scheme for robust speech recognition: spectral subtraction using spectral harmonics of speech. Proc. ICASSP. 1: 648-651.

[20] S. F. Boll. 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. on Acoustic, Speech and Signal Processing. 27(2): 113-120.

[21] J. Droppo, L. Deng and A. Acero. 2001. Evaluation of the SPLICE Algorithm on the aurora 2 database. In: proceeding of Eurospeech, Aalberg, Denmark. pp. 217-220.

[22] S. Furui. 1981. Cepstral analysis technique for automatic speaker verification. IEEE Trans. Acoustics, Speech, Signal Processing. ASSP-29(2): 254-272.

[23] O. Viikki and K. Laurila. 1998. Cepstral domain segmental feature vector normalization for noise robust speech recognition. Speech Comm. 25: 133-147.

[24] Y. Suh, M. Ji and H. Kim. 2007. Probabilistic class histrogram equalization for robust speech recognition. IEEE Signal Process. Let. 14(4): 287-290.

[25] J.L. Gauvain and C.H. Lee. 1994. MAP estimation for multivariate Gaussian mixture observation of markov chains. IEEE Trans. Speech Audio Process. 2(2): 291-298.

[26] C.J. Leggetter and P.C. Woodland. 1995. MLLR for speaker adaptation of continuous density hidden markov models. Compt. Speech Lang. 9: 171-185.

[27] M. J.F. Gales and S.J. Youngm. 1995. Robust speech recognition in additive and convolutional noise using PMC. Computer Speech and Language. 9; 289-307.

[28] [25]A. Acero, Li. Deng, T. Kristjansson and J. Zhang. 2000. HMM adaptation using Vector Taylor series for noisy speech recognition. In: Proc. ICSLP. 3: 869-872.

[29] S. G. Petterson, M.H. Johnsen and T.A. Myrvoll. A. Comparative Study of Model Compensation Methods for Robust Speech Recognition in Noisy Conditions.

[30] D.Y. Kim, C.K. Un and N.S. Kim. 1998. Speech recognition in noisy environments using first-order vector Taylor series. Speech Communication. 24(1): 39-49.

[31] M.J. Gales. 1998. Maximun likelihood linear transformation for HMM-based speech recognition. Computer Speech and Language, 12 (January).

[32] D. Kim and J.F. Gales. 2010. Noisy constrained MLLR for noise robust speech recognition. IEEE Trans. Audio Speech and language processing.

[33] J. Droppo and A. Acero. 2007. Environmental Robustness. In: Handbook of Speech Processing, Springer.

[34] J. Li, L. Deng, D. Yu, Y. Gong and A. Acero. 2008. HMM adaptation using Phase-Sensitive acoustic distortion model for environment-robust speech recognition. In: Proc. ICASSP, Las Vegas, USA.

[35] M. Seltzer, K. Kalgaonkar and A. Acero. 2010. Acoustic model adaptation via linear spline interpolation for robust speech recognition. In: Proc. ICASSP.

[36] B. Fery, L. Deng, A. Acero and T.T. Kristjansson. 2001. Algonquin: Iterating Laplace method to remove multiple types of acoustic distortion for robust speech recognition. In: Proc. Eurospeech, Aalborg, Denmark.

[37] V. Stouten, H. Vanhamme and P. Wambacq. 2005. Effect of Phase-Sensitive environment model and higher order VTS on noisy speech feature enhancement. In: Proc. ICASSP. pp. 433-436.