Noise Compensation Using Interacting Multiple Kalman Filters

Jianping Deng, Martin Bouchard, and Tet Hin Yeap

School of Information Technology and Engineering
University of Ottawa, Ottawa (Ontario), Canada
Jdeng,bouchard,tet@site.uottawa.ca

Abstract

This paper presents an approach to compensate the effects of noise with an Interacting Multiple Model algorithm using Unscented Kalman Filters (IMM-UKF) in log-spectral domain. The performance of this approach is studied experimentally on a continuous digits recognition task with additive noise conditions and compared with results previously obtained by the implementation of the Interacting Multiple Model algorithm using Extended Kalman Filters (IMM-EKF) in log-spectral domain. Simulation results show that a better performance in terms of word recognition rates can be obtained with the suggested approach.

1. Introduction

The performance degradation of a speech recognizer in the presence of additive noise is one of the major problems that still remain unsolved in the real-field applications of speech recognition technology. Towards solving the noise robustness problem, in the past few years, a variety of noise compensation techniques have been developed either in the time domain (as in many speech enhancement models), in the spectral domain, or in the real feature domain. In the last approach, the speech features to be enhanced are usually those used by most speech recognizers, such as the log-spectral, the Mel-Frequency Ceptrum Coefficients (MFCC), etc. It has been shown that methods applied to the speech recognition feature domains usually yield a better performance in terms of speech recognition rates than methods applied in the other domains [1]-[5].

One difficulty with speech enhancement in the feature domain is that the distortion caused by additive environment noise is highly non-linear in the log-spectral domain or in the cepstral domain. Most of the current approaches are based on piecewise linear approximation of the nonlinear function using a Taylor series expansion or a higher order Taylor series expansion [1]-[3]. Using nonlinear models to compensate the noise has been proposed in [4][5]. The sequential expectation maximization (EM) algorithm or statistical minimum mean square error (MMSE) matching is used in those methods to estimate the clean speech features. Further improvement on the sequential EM algorithm has been achieved with the application of the interacting multiple model (IMM) algorithm using a bank of Kalman filters in [6][7]. While the Kalman filter method has also been applied to address the speech enhancement problem in the time domain as in [8][9], its application in the feature domain is complicated by the fact that the speech contamination rule is expressed in a nonlinear way. A way to solve this problem is to use the Extended Kalman Filter (EKF) instead as in [6][7]. However, there are some drawbacks with the IMM algorithm using Extended Kalman Filters: the EKF approximates a non-Gaussian density by a Gaussian density [10], while the IMM algorithm approximates the Gaussian mixture by a single Gaussian density [8]. If these assumptions break down, the IMM algorithm using an EKF may diverge. As an alternative to the EKF, the Unscented Kalman Filter (UKF) was proposed in [11][12]. Compared with the EKF, the UKF can handle nonlinearities without using numerical derivatives and provides higher order approximation for both Gaussian and non-Gaussian distributions.

In this paper, an IMM method with UKF to replace the EKF, is studied for noisy speech feature compensation in the log-spectral domain. Clean feature vectors are statistically represented by a mixture of Gaussian distributions and each mixture component forms an Unscented Kalman Filter. The rest of this paper is organized as follows: Section 2 describes the dynamical model for speech feature enhancement. The Unscented Kalman Filter algorithm is described in Section 3. Experimental results for a continuous digits recognition task in the presence of additive noise are presented in Section 4. Conclusions are drawn in Section 5.

2. Noise Compensation Model

Assume that the clean speech signal $z(t)$ is corrupted by an independent ambient noise $x(t)$ in the time domain, the noisy speech $y(t)$ can be presented as:

$$y(t) = z(t) + x(t) \tag{1}$$

If we frame the speech signals and convert them into log-spectral domain on a Mel-scale, the above relation becomes a complex nonlinear function [3] as:

$$\mathbf{y}(k) = \mathbf{z}(k) + \log(\mathbf{l} + \exp(\mathbf{x}(k) - \mathbf{z}(k))) \tag{2}$$

where $\mathbf{y}(k)$, $\mathbf{x}(k)$ and $\mathbf{z}(k)$ represent the noisy speech vector, the noise vector and the hypothetical clean speech vector at k-th time frame respectively.

As in [4], we model the sequence of log-spectral vectors of noise as the output of a first-order auto-regressive (AR) system excited by a zero mean Gaussian process $\mathbf{v}$ with covariance matrix $\mathbf{Q}$ as follows:

$$\mathbf{x}(k) = \mathbf{A}\mathbf{x}(k\text{-}1) + \mathbf{v}(k) \tag{3}$$

Equations (1) and (2) form the state and observation equations of a dynamical system. It is found from our experiments that equation (3) could provide a better description of the log-spectral dynamics than the following equation used in [6][7]:

$$\mathbf{x}(k) = \mathbf{x}(k\text{-}1) + \mathbf{v}(k) \tag{4}$$

In the experiments for this paper, the AR predictor for noise was trained from 30 seconds of training samples of the

noise. Different from the conventional approach where the speech signal is modeled as the output of the dynamic system, the noise compensation problem here is treated inversely by assuming that the speech is the nonlinear obscuring influence that prevents us from observing the noise.

Without losing generality, we could assume that the probability density function (pdf) of the clean speech feature vector $\mathbf{z}$ can be represented by a multivariate Gaussian mixture model (GMM) as:

$$p(\mathbf{z}) = \sum_{j=1}^{q} p(j) N(\mathbf{z}; \mu_{\mathbf{z},j}, \Sigma_{\mathbf{z},j}) \tag{5}$$

where q is the total number of mixture components and $p(j)$, $\mu_{\mathbf{z},j}$ and $\Sigma_{\mathbf{z},j}$ represent the given a priori probability, the mean and the covariance of the jth Gaussian distribution, respectively. Let each mixture component form a Kalman filter, based on this state-space model, the interacting multiple model (IMM) approach can be applied to tract the noise feature vector $\mathbf{x}(k)$. The noise sequence $\mathbf{x}(k)$ will then be removed from the noisy signal sequence $\mathbf{y}(k)$ by using the approximated minimum mean squared estimation (MMSE) procedure developed in [13] as :

$$\hat{\mathbf{z}}(k) = \mathbf{y}(k) - \sum_{j=1}^{q} p(j \mid \mathbf{y}(k), \mathbf{x}(k)) \log(\mathbf{1} + \exp(\mathbf{x}(k) - \mu_{\mathbf{z},j})) \tag{6}$$

where q is the number of mixture components and $p(j \mid \mathbf{y}(k), \mathbf{x}(k))$ is given by:

$$p(j \mid \mathbf{y}(k), \mathbf{x}(k)) = \frac{p(j) N(\mathbf{y}(k); \mu_{\mathbf{y},j}, \Sigma_{\mathbf{y},j})}{\sum_{i=1}^{q} p(i) N(\mathbf{y}(k); \mu_{\mathbf{y},i}, \Sigma_{\mathbf{y},i})} \tag{7}$$

where $\mu_{\mathbf{y},i}, \Sigma_{\mathbf{y},i}$ denote the mean vector and the covariance matrix of $\mathbf{y}$ which are compensated by a first order Vector Taylor Series (VTS) based approach[13] with parameters $\mu_{\mathbf{z},i}, \Sigma_{\mathbf{z},i}$.

Finally, the estimated clean speech feature vector will be mapped from the log-spectral domain into the MFCC domain using a DCT transform, and passed on to a speech recognizer for the recognition task.

3. Unscented Kalman Filter

The Unscented Kalman Filter is a straightforward extension of the unscented transformation (UT) to the recursive estimation. The UT is a method for calculating the statistics of a random variable which undergoes a nonlinear transformation [12]. It is described below in summary [10].

Given an L-dimension random variable $\mathbf{x}$ with mean $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{P}_{\mathbf{x}}$, the statistics of a random variable $\mathbf{y} = g(\mathbf{x})$ can be calculated by the following procedure: Firstly, form a matrix χ of $2L+1$ *sigma* vectors χ_i with corresponding weights W_i, according to the following rules:

$$\begin{cases} \chi_0 = \bar{\mathbf{x}} \\ \chi_i = \bar{\mathbf{x}} + \left(\sqrt{(L+\lambda)\mathbf{P}_{\mathbf{x}}}\right)_i & i = 1, \cdots, L \\ \chi_i = \bar{\mathbf{x}} - \left(\sqrt{(L+\lambda)\mathbf{P}_{\mathbf{x}}}\right)_{i-L} & i = L+1, \cdots, 2L \\ W_0^{(m)} = \lambda/(L+\lambda) \\ W_0^{(c)} = \lambda/(L+\lambda) + (1-\alpha^2+\beta) \\ W_i^{(m)} = W_i^{(c)} = 1/\{2(L+\lambda)\} & i = 1, \cdots .2L \end{cases} \tag{8}$$

where $\lambda = \alpha^2(L+\gamma) - L$ is a scaling parameter, α determines the spread of the sigma points around $\bar{\mathbf{x}}$ and is usually set to a small positive value. γ is a secondary scaling parameter which is usually set to 0, and β is used to incorporate a priori knowledge of the distribution of $\mathbf{x}$ (for Gaussian distributions, $\beta = 2$ is optimal). $\left(\sqrt{(L+\lambda)\mathbf{P}_{\mathbf{x}}}\right)_i$ is the ith row of the matrix square root. These sigma vectors are propagated through the nonlinear function,

$$\mathrm{Y}_i = g(\chi_i) \qquad i = 0, \cdots, 2L \tag{9}$$

and the mean and covariance for $\mathbf{y}$ are approximated by:

$$\begin{aligned} \bar{\mathbf{y}} &\approx \sum_{i=0}^{2L} W_i^{(m)} \mathrm{Y}_i \\ \mathbf{P}_{\mathbf{y}} &\approx \sum_{i=0}^{2L} W_i^{(c)} \{\mathrm{Y}_i - \bar{\mathbf{y}}\}\{\mathrm{Y}_i - \bar{\mathbf{y}}\}^T \end{aligned} \tag{10}$$

For Gaussian inputs, the UT results can achieve the accuracy of the third order. For non-Gaussian inputs, approximations are accurate to at least the second-order. The accuracy of third and higher order moments is determined by the choice of α and β [12]. Hence the UKF is performed by applying the UT sigma point selection scheme, as shown in equation (8), to the state vector to calculate the corresponding sigma matrix.

For frame k, some notations of the noise feature vector conditioned on the state $S_k = j$ are first defined:

$$\begin{aligned} \hat{\mathbf{x}}^j(k \mid k) &= \mathrm{E}[\hat{\mathbf{x}}(k \mid k) \mid \mathbf{y}(1:k), S_k = j] \\ \mathbf{P}_{\hat{\mathbf{x}}}^j(k \mid k) &= \mathrm{cov}[\hat{\mathbf{x}}(k \mid k) \mid \mathbf{y}(1:k), S_k = j] \\ L^j(k) &= \mathrm{P}r(\mathbf{y}(k) \mid \mathbf{y}(1:k-1), S_k = j) \end{aligned} \tag{12}$$

Initialize with:

$$\begin{aligned} \hat{\mathbf{x}}(0 \mid 0) &= E[\mathbf{x}(0)] \\ \mathbf{P}_{\hat{\mathbf{x}}}(0 \mid 0) &= E\{[\mathbf{x}(0) - \hat{\mathbf{x}}(0 \mid 0)][\mathbf{x}(0) - \hat{\mathbf{x}}(0 \mid 0)]^T\} \end{aligned} \tag{13}$$

Given a noisy speech feature vector $\mathbf{y}(k)$ (k=1,…,K) and parameters $(\mu_{\mathbf{z},j}, \Sigma_{\mathbf{z},j}, j = 1 \ldots q)$, $\mathbf{x}(k)$ is found from the interacting multiple model algorithm using a bank of Unscented Kalman Filters with the following steps performed in sequence:

$$\begin{aligned} &\left(\hat{\mathbf{x}}^j(k \mid k), \mathbf{P}_{\hat{\mathbf{x}}}^j(k \mid k), L^j(k)\right) \\ &= UKF(\hat{\mathbf{x}}(k-1 \mid k-1), \mathbf{P}_{\hat{\mathbf{x}}}(k-1 \mid k-1), \mathbf{y}(k); \mu_{\mathbf{z},j}, \Sigma_{\mathbf{z},j}) \end{aligned}$$

$$M^{j}(k \mid k) = \Pr\left(S_k = j \mid \mathbf{y}(1:k)\right) = \frac{L^{j}(k)\, p(S_k = j)}{\sum_{j} L^{j}(k)\, p(S_k = j)} \tag{14}$$

$$\left(\hat{\mathbf{x}}(k \mid k), \mathbf{P}_{\hat{\mathbf{x}}}(k \mid k)\right) = Collapse\left(\hat{\mathbf{x}}^{j}(k \mid k), \mathbf{P}_{\hat{\mathbf{x}}}^{j}(k \mid k), M^{j}(k \mid k)\right)$$

where $p(S_k = j) = p(j)$ is the a priori probability of the jth GMM mixture component and is assumed to be independent of the previous observation.

The definition of *UKF* is as follows: Firstly, form $\hat{\mathbf{x}}^{a}(k-1 \mid k-1)$, $\mathbf{P}_{\tilde{x}}^{a}(k-1 \mid k-1)$, χ^{a} as in [10]:

$$\begin{aligned}
&\hat{\mathbf{x}}^{a}(k-1|k-1) \\
&= [\hat{\mathbf{x}}^{T}(k-1|k-1) \quad \mathbf{0} \quad \mu_{\mathbf{z},j}{}^{T}]^{T} \\
&\mathbf{P}_{\tilde{x}}^{a}(k-1|k-1) \\
&= \begin{bmatrix} \mathbf{P}_{\hat{\mathbf{x}}}(k-1|k-1) & \mathbf{0} & \mathbf{0}; & \mathbf{0} & \mathbf{Q} & \mathbf{0}; & \mathbf{0} & \mathbf{0} & \Sigma_{\mathbf{z},j} \end{bmatrix} \\
&\chi^{a} = [(\chi^{x})^{T} \quad (\chi^{y})^{T} \quad (\chi^{z})^{T}]^{T}
\end{aligned} \tag{15}$$

Sigma matrices: $i = 1, \cdots, L$

$$\begin{aligned}
&\chi_0^{a}(k-1|k-1) \\
&= \hat{\mathbf{x}}^{a}(k-1|k-1) \\
&\chi_i^{a}(k-1|k-1) \\
&= \hat{\mathbf{x}}^{a}(k-1|k-1) + \left(\sqrt{(L+\lambda)\mathbf{P}_{\tilde{x}}^{a}(k-1|k-1)}\right)_i \\
&\chi_{i+L}^{a}(k-1|k-1) \\
&= \hat{\mathbf{x}}^{a}(k-1|k-1) - \left(\sqrt{(L+\lambda)\mathbf{P}_{\tilde{x}}^{a}(k-1|k-1)}\right)_i
\end{aligned} \tag{16}$$

Time update equations:

$$\begin{aligned}
&\chi_i^{x}(k|k-1) \\
&= \mathbf{A}\chi_i^{x}(k-1|k-1) + \chi_i^{v}(k-1|k-1) \qquad i = 0, \cdots, 2L \\
&\hat{\mathbf{x}}^{j}(k|k-1) = \sum_{i=0}^{2L} W_i^{(m)} \chi_i^{x}(k|k-1) \\
&\mathbf{P}_{\hat{\mathbf{x}}}^{j}(k|k-1) \\
&= \sum_{i=0}^{2L} W_i^{(c)} [\chi_i^{x}(k|k-1) - \hat{\mathbf{x}}^{j}(k|k-1)][\chi_i^{x}(k|k-1) - \hat{\mathbf{x}}^{j}(k|k-1)]^{T} \\
&\Upsilon_i(k|k-1) = \chi_i^{z}(k|k-1) \\
&\quad + \log\left(\mathbf{I} + \exp(\chi_i^{x}(k|k-1) - \chi_i^{z}(k|k-1))\right) \quad i = 0, \cdots, 2L \\
&\hat{\mathbf{y}}(k|k-1) = \sum_{i=0}^{2L} W_i^{(m)} \Upsilon_i(k|k-1)
\end{aligned} \tag{17}$$

Measurement update equations:

$$\begin{aligned}
&\mathbf{P}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}(k) \\
&= \sum_{i=0}^{2L} W_i^{(c)} [\Upsilon_i(k|k-1) - \hat{\mathbf{y}}(k|k-1)][\Upsilon_i(k|k-1) - \hat{\mathbf{y}}(k|k-1)]^{T}
\end{aligned} \tag{18}$$

$$\begin{aligned}
&\mathbf{P}_{\hat{\mathbf{x}}\hat{\mathbf{y}}}(k) \\
&= \sum_{i=0}^{2L} W_i^{(c)} [\chi_i(k|k-1) - \hat{\mathbf{x}}^{j}(k|k-1)][\Upsilon_i(k|k-1) - \hat{\mathbf{y}}(k|k-1)]^{T}
\end{aligned} \tag{19}$$

$$\mathbf{K}(k) = \mathbf{P}_{\hat{\mathbf{x}}\hat{\mathbf{y}}}(k)\mathbf{P}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}(k) \tag{20}$$

Filtered estimate of the state vector:

$$\hat{\mathbf{x}}^{j}(k \mid k) = \hat{\mathbf{x}}^{j}(k \mid k-1) + \mathbf{K}(k)(\mathbf{y}(k) - \hat{\mathbf{y}}(k \mid k-1)) \tag{21}$$

Filtered state-error covariance matrix:

$$\mathbf{P}_{\hat{\mathbf{x}}}^{j}(k \mid k) = \mathbf{P}_{\hat{\mathbf{x}}}^{j}(k \mid k-1) - \mathbf{K}(k)\mathbf{P}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}(k)\mathbf{K}^{T}(k) \tag{22}$$

$$L^{j}(k) = N\left(\mathbf{y}(k); \hat{\mathbf{y}}(k \mid k-1), \mathbf{P}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}(k)\right) \tag{23}$$

Collapse is a moment matching function, defined as in [9]:

$$\hat{\mathbf{x}}(k \mid k) = \sum_{j} M^{j}(k)\hat{\mathbf{x}}^{j}(k \mid k) \tag{24}$$

$$\begin{aligned}
&\mathbf{P}_{\hat{\mathbf{x}}}(k|k) = \sum_{j} M^{j}(k)\mathbf{P}_{\hat{\mathbf{x}}}^{j}(k|k) \\
&+ \sum_{j} M^{j}(k)(\hat{\mathbf{x}}^{j}(k|k) - \hat{\mathbf{x}}(k|k))(\hat{\mathbf{x}}^{j}(k|k) - \hat{\mathbf{x}}(k|k))^{T}
\end{aligned} \tag{25}$$

In [6][7], to make the IMM-EKF method stable, a shrink factor was introduced to reduce the Kalman gain in a manner that is not mathematically justified. From a number of experiments, it is found that the shrink factor will bring negative effects to the results of our method. So no shrink factor was used in our experiments.

4. Speech Recognition Experiments

Performances of the proposed method were evaluated with speaker independent continuous digits recognition experiments. The speech data used for the experiments is from the Numbers v1.3 corpus provided by Oregon Health & Science University (OGI). The corpus is a collection of 8kHz telephone speech data, including both isolated digits and continuous digit strings [14]. In our experiments, speech files with fixed-length 5 connected digits are used. To construct the dynamic state space model for feature enhancement, utterances of 100 digit strings form the training data. For each frame of speech signal, the feature vector was represented by 19 mel-scaled filterbank log-spectral features. The distribution of the log-spectral features of these training data was modeled by a mixture of 18 Gaussian distributions (GMM) with full covariance matrices. In the testing stage, an independent set of 40 utterances was used for evaluation. Two kinds of typical noise sources were artificially added to the speech signal by a computer, with SNR varying from 0 dB to 20dB, to simulate the noisy environment. They are the white and street noise signals from the ITU-T Supplement P.23 speech database. Environment compensation procedures were carried out to transform the noisy feature vectors into clean ones.

By applying a discrete cosine transform (DCT) on the log-spectral feature vectors, a 12th order cepstral coefficient (MFCC) vector can be derived for each frame. Derived

cepstrum vectors and their first and second order derivatives were used for recognition. The recognition system used in the experiments is based on continuous HMM model trained by 527 utterances. Each phoneme is represented by a left-to-right monophone HMM containing 5 states (3 observation states, an entry and an exit state) and 8 mixtures for each state. Both training and recognition phases are performed using the HTK toolbox [15].

Figure 1, 2 shows the average word error rate (WER) for white noise and street noise mixtures. As a comparison, recognition with uncompensated noisy speech (No NR), with conventional spectral subtraction (SS) and with cepstra derived by the IMM-EKF method [7] are also shown.

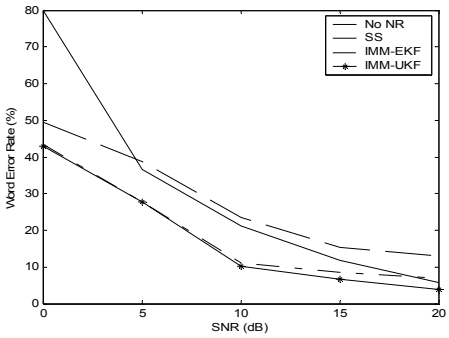

Figure 1: Results for speech mixed with white noise

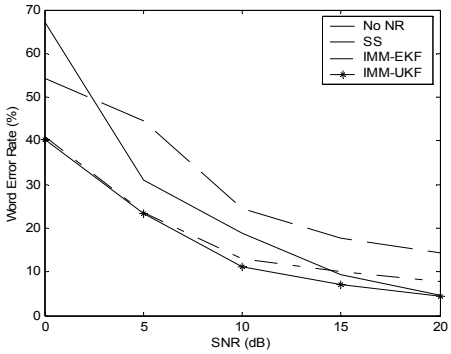

Figure 2: Results for speech mixed with street noise

The results show that the interacting multiple model approach with either Extended Kalman Filtering or Unscented Kalman Filtering is effective at improving the recognition performance. At low SNRs, the performance with the IMM-EKF method is similar to that with the IMM-UKF. For both types of noise, the IMM-UKF method showed its superiority to the EKF counterpart at higher SNRs. At SNR higher than 10 dB where the speech is fairly clean, the recognition performance obtained with the IMM-EKF compensated speech is actually poorer than that obtained with the uncompensated noisy speech, while the algorithm continues to provide performance improvement with the IMM-UKF approach.

5. Conclusions

In this paper, an interacting multiple model algorithm with Unscented Kalman Filters (IMM-UKF) was applied to compensate the environment noise in the log-spectral domain because of the drawbacks of the Extended Kalman Filter in nonlinear systems. The Unscented Kalman Filter captures the nonlinearities of the dynamic system without an approximation by a Taylor series expansion. The simulation results show that the IMM-UKF approach provides a better performance than the IMM-EKF in higher SNRs.

6. References

[1] Moreno, P.J., Raj, B., Stern, R.M., "A vector Taylor series approach for environment-independent speech recognition", *IEEE ICASSP'96*, Volume 2, Pages:733 – 736, May 1996

[2] Kim, N.S, "Nonstationary environment compensation based on sequential estimation", *IEEE Signal Processing Letters*, Volume 5, Pages:57 – 59, March 1998

[3] Deng, L., Droppo, J., Acero, A., "A Bayesian approach to speech feature enhancement using the dynamic cepstral prior", *IEEE ICASSP'02*, Volume: 1, Pages:829-1 - 829-32, May 2002

[4] Singh, R., Raj, B., "Tracking noise via dynamical systems with a continuum of states", *IEEE ICASSP'03*, Volume: 1, Pages:I-396 - I-399, April 2003

[5] Jiang H., Wang Q., "Nonlinear noise compensation in feature domain for speech recognition with numerical methods", *IEEE ICASSP'04*, Volume: 1, Pages: I-985-8, May 2004

[6] Kim, N.S., "Time-varying noise compensation using multiple Kalman filters", *IEEE ICASSP'99*, Volume: 1, Pages:429–432, March 1999

[7] Kim, N.S., "IMM-based estimation for slowly evolving environments", *IEEE Signal Processing Letters*, Volume: 5, Pages:146–149, June 1998

[8] Kim, J. B., Lee, K.Y., and Lee, C.W., "On the application of the Interacting Multiple Model Algorithm for Enhancing Noisy Speech," *IEEE Trans. Speech and Audio Processing*, Vol. 8, No 3, May 2000

[9] Deng, J., Bouchard, M. and Yeap, T., "Speech Enhancement using a switching Kalman filter with a perceptual post-filter", *IEEE ICASSP'05*, March, 2005

[10] Wan, E.A., Van Der Merwe, R.; "The unscented Kalman filter for nonlinear estimation", *The IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium AS-SPCC.*, Pages:153 – 158, Oct. 2000

[11] Julier, S.; Uhlmann, J.; Durrant-Whyte, H.F.; "A new method for the nonlinear transformation of means and covariances in filters and estimators", *IEEE Transactions on Automatic Control*, Volume: 45, Pages:477 - 482, March 2000

[12] Julier, S. J. and Uhlmann, J. K., "A new extension of the Kalman Filter to Nonlinear Systems." *The 11th Int. Symp. On Aerospace/Defence Sensing, Simulation and Controls*, 1997.

[13] Moreno, P.J., *Speech recognition in noisy Environments*, Ph.D Thesis, Carnegie Mellon University, 1996

[14] http://www.cslu.ogi.edu/corpora/

[15] Young, S., Kershaw, D., Odell, J., Ollason, D., Valtchev V., woodland, P., *The HTK book – version 2.2*, *Entropic*, 1999