Generalized Variable Parameter HMMs for Noise Robust Speech Recognition

Ning Cheng [1,2], *Xunying Liu* [3] *& Lan Wang* [1,2]

[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[2]The Chinese University of Hong Kong, Shatin, Hong Kong
[3]Cambridge University Engineering Dept, Trumpington St., Cambridge, CB2 1PZ U.K.

Abstract

Handling variable ambient noise is a challenging task for automatic speech recognition (ASR) systems. To address this issue, multi-style, noise *condition independent* (CI) model training using speech data collected in diverse noise environments, or uncertainty decoding techniques can be used. An alternative approach is to explicitly approximate the continuous trajectory of Gaussian component mean and variance parameters against the varying noise level, for example, using variable parameter HMMs (VP-HMM). This paper investigates a more generalized form of variable parameter HMMs (GVP-HMM). In addition to Gaussian component means and variances, it can also provide a more compact trajectory modelling for tied linear transformations. An alternative noise *condition dependent* (CD) training algorithm is also proposed to handle the bias to training noise condition distribution. Consistent error rate gains were obtained over conventional VP-HMM mean and variance only trajectory modelling on a medium vocabulary Mandarin Chinese in-car navigation command recognition task.

Index Terms: variable noise, generalized variable parameter HMM, noise robust speech recognition

1. Introduction

The presence of environmental noise often leads to severe degradation of automatic speech recognition (ASR) performance. In particular, when the ambient noise is of variable, non-stationary nature, this problem becomes even more challenging. To handle this issue, three major categories of techniques are often used. The first involves training a multi-style, or noise *condition independent* (CI) system on speech data collected in a wide range of diverse noise environments [11]. This exploits the implicit modelling ability of the underlying statistical models to achieve a good generalization to unseen noise conditions. The second category is based on uncertainty decoding (UD) [1, 5, 12, 6, 10]. Rather than using a point estimate of the corrupted features, the uncertainty that varies with the noise represented by, for example, a conditional distribution of the corrupted speech, is propagated into the recognizer. The third category explicitly approximate the continuous trajectories of optimal model parameters with respect to noise condition [7, 3, 15, 16, 17]. Due to their well understood mathematical properties and stability, lower order polynomial functions are commonly, for example, in variable parameter HMMs (VP-HMM) [3, 15, 16, 17].

This work was supported in part by National Natural Science Foundation of China (No.90920002) and Natural Science Foundation of Guangdong Province, China (No.10478922035-X004936).

Under the VP-HMM framework, Gaussian means and variances are dynamically re-computed during recognition for each noise condition detected in the test data using their associated polynomial functions. Their coefficients are trained on in a multi-style, or CI fashion using speech data collected in different noise environments. There are two issues associated with this approach. First, a Gaussian component level polynomial modelling of mean and variance trajectories are very expensive to use in recognition time. Hence, more compact forms of trajectory modelling are preferred. Second, by implicitly weighting data of different noise level, CI training can not only learn, but also incur a potentially undue bias, to the noise condition distribution found in the training data. For example, when the training data is a dominated by one single noise condition, a good generalization to a variety of unknown noise conditions in the test set can be problematic. Hence, alternative estimation methods without introducing such a bias are preferred.

To address these issues, this paper investigates a more generalized form of variable parameter HMMs (GVP-HMM). In addition to Gaussian means and variances, it can also provide a more compact trajectory modelling for tied linear transformations. An alternative noise *condition dependent* (CD) training algorithm is also proposed to handle the bias to training data noise condition distribution. The rest of the paper is organized as follows. The GVP-HMM framework is proposed in section 2. Noise condition independent and dependent training of GVP-HMMs are presented in sections 3 and 4 In section 5 GVP-HMM based noise compensation schemes are evaluated on a medium vocabulary Mandarin Chinese speech recognition task. Section 6 is the conclusion and future research.

2. Generalized Variable Parameter HMMs

Generalized variable parameter HMMs (GVP-HMMs) explicitly model the trajectory of optimal acoustic parameters that can vary with respect to the underlying noise condition. The type of parameter trajectories are not restricted to those of means and covariances of conventional tied mixture HMMs. Other more compact forms of parameters, such as model or feature space linear transformations [9, 8], may also be considered. In this paper, trajectories of Gaussian mean MLLR transforms are modelled. For a D dimensional observation $\boldsymbol{o}_t$ emitted from Gaussian mixture component m, assuming P^{th} order polynomials are used, this is given by

$$\boldsymbol{o}^{(t)} \sim \mathcal{N}\left(\boldsymbol{o}^{(t)}; \boldsymbol{\mu}^{(m)}(\mathbf{v}_t), \boldsymbol{\Sigma}^{(m)}(\mathbf{v}_t), \boldsymbol{W}^{(r_m)}(\mathbf{v}_t)\right) \quad (1)$$

where $\mathbf{v}_t^\top$ is a $(P+1)$ dimensional Vandermonde vector [2], such that $\mathbf{v}_{t,p} = v_t^{p-1}$. v_t is an auxiliary feature, and in this paper, the speech-noise-ratio (SNR) condition at frame t. $\boldsymbol{W}^{(r_m)}$ the $(D+1)\times D$ mean transform that component m is assigned to. $\boldsymbol{\mu}^{(m)}(\cdot)$, $\boldsymbol{\Sigma}^{(m)}(\cdot)$ and $\boldsymbol{W}^{(r_m)}(\cdot)$ are the P^{th} order mean, covariance and MLLR transform trajectory polynomials respectively. Assuming diagonal covariances are used, then the trajectories of the i^{th} dimension of the mean and variance, as well the transform element on i row and column j, are

$$
\begin{aligned}
\mu_i^{(m)}(\mathbf{v}_t) &= \mathbf{v}_t \cdot \mathbf{c}^{(\mu_i^{(m)})} \\
\sigma_{i,i}^{(m)}(\mathbf{v}_t) &= \breve{\sigma}_{i,i}^{(m)}\mathbf{v}_t \cdot \mathbf{c}^{(\sigma_{i,i}^{(m)})} \\
w_{i,j}^{(r_m)}(\mathbf{v}_t) &= \mathbf{v}_t \cdot \mathbf{c}^{(w_{i,j}^{(r_m)})}
\end{aligned} \tag{2}
$$

where $\mathbf{c}^{(\cdot)}$ is a $(P+1)$ dimensional polynomial coefficient vector such that $\mathbf{c}_p^{(\cdot)} = c_{p-1}^{(\cdot)}$, and $c_{p-1}^{(\cdot)}$ the $(p-1)^{th}$ order polynomial coefficient of the parameter trajectory being considered. $\breve{\sigma}_{i,i}^{(m)}$ is the clean speech based variance estimate.

If the training set has no data in a given SNR, standard HMMs will not know the optimal distribution parameters, whereas GVP-HMMs can instantly obtain both Gaussian component and linear transformation parameters for the matching condition by-design without requiring any multi-pass noise adaptation process, in common with VP-HMMs. Another major advantage of the GVP-HMM framework is that a more compact, and flexible form of parameter trajectory modelling is possible. For example, when only limited amounts of noisy training data is available, to ensure all polynomial coefficients are robustly estimated, only the trajectories associated with the elements of a globally tied mean MLLR transform are be considered. On the other hand, when large amounts of noisy training data is used, a more refined modelling resolution can also be obtained by increasing the number of tied transformations, or modelling the trajectories of multiple parameter types simultaneously.

3. CI Training of GVP-HMMs

Condition independent (CI) training aims to find the optimal parameter trajectory polynomial functions using a mixed set of training data that contains a range of observed noise conditions in a multi-style fashion [11]. The associated maximum likelihood (ML) auxiliary function is given by [4],

$$
\begin{aligned}
\mathcal{Q}_{\mathsf{CI}}(\lambda,\tilde{\lambda}) = &\sum_{m,t}\gamma_m(t)\log\mathcal{N}\Big(\boldsymbol{o}^{(t)};\boldsymbol{\mu}^{(m)}(\mathbf{v}_t), \\
&\boldsymbol{\Sigma}^{(m)}(\mathbf{v}_t),\boldsymbol{W}^{(r_m)}(\mathbf{v}_t)\Big)
\end{aligned} \tag{3}
$$

where $\gamma_m(t)$ is the posterior probability of frame $\boldsymbol{o}_t$ at component m. Combining the above with equations (1) and (2) and setting the gradient against the polynomial coefficient vectors associated with the mean, variance and MLLR transform element trajectories respectively to zero, the following solutions of their coefficients can be derived.

$$
\begin{aligned}
\mathbf{c}^{(\mu_i^{(m)})} &= \mathbf{U}^{(\mu_i^{(m)})-1}\mathbf{k}^{(\mu_i^{(m)})} \\
\mathbf{c}^{(\sigma_{i,i}^{(m)})} &= \mathbf{U}^{(\sigma_{i,i}^{(m)})-1}\mathbf{k}^{(\sigma_{i,i}^{(m)})} \\
\mathbf{c}^{(w_i^{(r_m)})} &= \mathbf{U}^{(w_i^{(r_m)})-1}\mathbf{k}^{(w_i^{(r_m)})}
\end{aligned} \tag{4}
$$

where $\mathbf{c}^{(w_i^{(r_m)})}$ is a $(D+1)\times(P+1)$ dimensional meta polynomial coefficient vector spanning across all elements of row i of transform $\boldsymbol{W}^{(r_m)}$, and the sufficient statistics are

$$
\begin{aligned}
\mathbf{U}^{(\mu_i^{(m)})} &= \sum_t \gamma_m(t)\sigma_{i,i}^{(m)-1}(\mathbf{v}_t)\mathbf{v}_t^\top\mathbf{v}_t \\
\mathbf{k}^{(\mu_i^{(m)})} &= \sum_t \gamma_m(t)\sigma_{i,i}^{(m)-1}(\mathbf{v}_t)o_i^{(t)}\mathbf{v}_t^\top \\
\mathbf{U}^{(\sigma_{i,i}^{(m)})} &= \sum_t \gamma_m(t)\breve{\sigma}_{i,i}^{(m)}\mathbf{v}_t^\top\mathbf{v}_t \\
\mathbf{k}^{(\sigma_{i,i}^{(m)})} &= \sum_t \gamma_m(t)\left(o_i^{(t)}-\mu_i^{(m)}(\mathbf{v}_t)\right)^2\mathbf{v}_t^\top
\end{aligned} \tag{5}
$$

$\mathbf{U}^{(w_i^{(r_m)})}$ is a $[(D+1)\times(P+1)]\times[(D+1)\times(P+1)]$ meta Vandermonde matrix, and $\mathbf{k}^{(w_i^{(r_m)})}$ a $(D+1)\times(P+1)$ dimensional meta regression target vector. The sub-matrices and sub-vectors associated with transform element $w_{i,j}^{(r_m)}$ are

$$
\begin{aligned}
\mathbf{U}^{(w_{i,j}^{(r_m)})} = &\Bigg[\sum_{m\in r_m,t}\gamma_m(t)\sigma_{i,i}^{(m)-1}(\mathbf{v}_t)\zeta_{t,j}^{(m)}\zeta_{t,1}^{(m)}\mathbf{v}_t^\top\mathbf{v}_t\ , \\
&\ldots, \sum_{m\in r_m,t}\gamma_m(t)\sigma_{i,i}^{(m)-1}(\mathbf{v}_t)\zeta_{t,j}^{(m)}\zeta_{t,i}^{(m)}\mathbf{v}_t^\top\mathbf{v}_t, \\
&\ldots, \sum_{m\in r_m,t}\gamma_m(t)\sigma_{i,i}^{(m)-1}(\mathbf{v}_t)\zeta_{t,j}^{(m)}\zeta_{t,D+1}^{(m)}\mathbf{v}_t^\top\mathbf{v}_t\Bigg] \\
\mathbf{k}^{(w_{i,j}^{(r_m)})} = &\sum_{m\in r_m,t}\gamma_m(t)\sigma_{i,i}^{(m)-1}(\mathbf{v}_t)\zeta_{t,j}^{(m)}o_i^{(t)}\mathbf{v}_t^\top
\end{aligned} \tag{6}
$$

where the $(D+1)$ dimensional extended mean vector trajectory is given by $\boldsymbol{\zeta}_t^{(m)} = [\boldsymbol{\mu}^{(m)}(\mathbf{v}_t),1]^\top$.

4. CD Training of GVP-HMMs

As discussed in section 1, CI estimation may introduce an undue bias to the underlying noise condition distribution in the training data. To handle this issue, an alternative noise *condition dependent* (CD) approach may be used. Statistics are separately accumulated over each observed noise condition in the training data, before being used to derive the coefficients of various trajectory polynomials via a regression estimation. This approach explicitly uses the fitting ability of polynomial functions to model the underlying parameter trajectory that vary against noise level. When sufficient amounts of training data is available for each noise condition, this approach may provide a more impartial and accurate approximation of the "true" model parameter trajectory than CI training. For a total of $Q, Q > P$ discretely quantized noise conditions found in the training data, $\{v_1,...,v_q,...,v_Q\}$, the associated ML auxiliary function for the q^{th} condition is,

$$
\begin{aligned}
\mathcal{Q}_{\mathsf{CD}}^{(q)}(\lambda,\tilde{\lambda}) = &\sum_{m,t,v_t=v_q}\gamma_m(t)\log\mathcal{N}\Big(\boldsymbol{o}^{(t)};\boldsymbol{\mu}^{(m)}(\mathbf{v}_t), \\
&\boldsymbol{\Sigma}^{(m)}(\mathbf{v}_t),\boldsymbol{W}^{(r_m)}(\mathbf{v}_t)\Big)
\end{aligned} \tag{7}
$$

Combining the above with equations (1) and (2) and for each noise condition separately setting the gradient against the polynomial coefficient vectors associated with the mean, variance and MLLR transform element trajectories respectively to zero, the following solutions of their coefficients in a similar form to those in equation (4) can be derived,

$$\begin{aligned}\mathbf{c}^{(\mu_i^{(m)})} &= \mathbf{U}^{(\mu_i^{(m)})-1}\mathbf{k}^{(\mu_i^{(m)})} \\ \mathbf{c}^{(\sigma_{i,i}^{(m)})} &= \mathbf{U}^{(\sigma_{i,i}^{(m)})-1}\mathbf{k}^{(\sigma_{i,i}^{(m)})} \\ \mathbf{c}^{(w_i^{(rm)})} &= \mathbf{U}^{(w_i^{(rm)})-1}\mathbf{k}^{(w_i^{(rm)})} \end{aligned} \quad (8)$$

where $\mathbf{U}^{(\mu_i^{(m)})}$ and $\mathbf{U}^{(\sigma_{i,i}^{(m)})}$ are $Q \times (P+1)$ Vandermonde matrices. $\mathbf{k}^{(\mu_i^{(m)})}$ and $\mathbf{k}^{(\sigma_{i,i}^{(m)})}$ are Q dimensional regression target vectors. The sufficient statistics associated with a particular noise condition at row q are

$$\begin{aligned}\mathbf{u}_q^{(\mu_i^{(m)})} &= \sum_{t, v_t = v_q} \gamma_m(t)\sigma_{i,i}^{(m)-1}(\mathbf{v}_t)\mathbf{v}_t \\ k_q^{(\mu_i^{(m)})} &= \sum_{t, v_t = v_q} \gamma_m(t)\sigma_{i,i}^{(m)-1}(\mathbf{v}_t)o_i^{(t)} \\ \mathbf{u}_q^{(\sigma_{i,i}^{(m)})} &= \sum_{t, v_t = v_q} \gamma_m(t)\check{\sigma}_{i,i}^{(m)}\mathbf{v}_t \\ k_q^{(\sigma_{i,i}^{(m)})} &= \sum_{t, v_t = v_q} \gamma_m(t)\left(o_i^{(t)} - \mu_i^{(m)}(\mathbf{v}_t)\right)^2 \end{aligned} \quad (9)$$

$\mathbf{U}^{(w_i^{(rm)})}$ is a $[Q \times (D+1)] \times [(D+1) \times (P+1)]$ meta Vandermonde matrix, and $\mathbf{k}^{(w_i^{(rm)})}$ is a $Q \times (D+1)$ dimensional meta regression target vector. Their respective rows associated with condition q and transform matrix element $w_{i,j}^{(rm)}$ are

$$\begin{aligned}\mathbf{u}_q^{(w_{i,j}^{(rm)})} &= \Bigg[\sum_{m \in r_m, t, v_t = v_q} \gamma_m(t)\sigma_{i,i}^{(m)-1}(\mathbf{v}_t)\zeta_{t,j}^{(m)}\zeta_{t,1}^{(m)}\mathbf{v}_t \,, \\ &\quad \ldots, \\ &\quad \sum_{m \in r_m, t, v_t = v_q} \gamma_m(t)\sigma_{i,i}^{(m)-1}(\mathbf{v}_t)\zeta_{t,j}^{(m)}\zeta_{t,i}^{(m)}\mathbf{v}_t, \\ &\quad \ldots, \\ &\quad \sum_{m \in r_m, t, v_t = v_q} \gamma_m(t)\sigma_{i,i}^{(m)-1}(\mathbf{v}_t)\zeta_{t,j}^{(m)}\zeta_{t,D+1}^{(m)}\mathbf{v}_t \Bigg] \\ k_q^{(w_{i,j}^{(rm)})} &= \sum_{m \in r_m, t, v_t = v_q} \gamma_m(t)\sigma_{i,i}^{(m)-1}(\mathbf{v}_t)\zeta_{t,j}^{(m)}o_i^{(t)} \end{aligned} \quad (10)$$

where again the $(D+1)$ dimensional extended mean vector trajectory $\boldsymbol{\zeta}_t^{(m)} = [\boldsymbol{\mu}^{(m)}(\mathbf{v}_t), 1]^\top$.

5. Experiments and Results

GVP-HMM based noise compensation schemes are evaluated on a medium vocabulary Mandarin Chinese in-car navigation command recognition task designed at CAS-SIAT. A total of 25 hours of clean training data and 1 hour of clean test data were used. A multi-style training data set was constructed by artificially corrupting the clean speech data with added car engine noise. Noise corrupted speech data generated under six sentence level SNR conditions: 0dB, 4dB, 8dB, 12dB, 16dB and 20dB, were used in training, while the corrupted test data set consists of five sentence level SNR conditions: 2dB, 6dB, 10dB, 14dB, and 18dB [13]. The baseline HMM based acoustic models were ML trained using HTK [14] on 42-dimensional HLDA projected PLP features further augmented with smoothed pitch parameters. Phonetic decision tree clustered cross-word tonal triphones HMMs with a left-to-right topology and three emission states were used. A total of 2416 distinct state level tied Gaussian mixture models with 12 components per state were used. The baseline character error rate (CER) performance of clean speech data, and multi-style trained conventional HMM systems on various SNR levels are shown as "clean" and "mcond" respectively in the first two rows of table 1. A separate set of HMM models trained using speech data corrupted at each matching SNR condition in the test set were also evaluated, and shown as "match" in the 3rd line of table 1.

System	2db	6db	10db	14db	18db	Avg
clean	72.27	49.12	28.83	17.57	12.36	36.03
mcond	55.77	31.04	17.84	12.16	9.89	25.34
match	41.42	28.88	19.51	14.04	11.57	23.08

Table 1: CER performance of baseline clean speech, multi-style and matching condition trained systems on various SNR levels.

As discussed in section 1, an important issue when using VP-HMM and GVP-HMM based noise compensation techniques is the appropriate order of polynomials to choose. As higher order polynomials are prone to Runge effect related instability, lower order polynomials are often used. In practice, 1st, 2nd and 3rd order polynomials are found to give comparable recognition performance. For example, the recognition performance of three different systems using CI trained polynomials of 1st, 2nd and 3rd order respectively for Gaussian component mean trajectories are shown in table 2. As is shown in the table, these three gave very close CER performance. As 2nd order polynomials also marginally outperform the simpler 1st order, and the more complicated 3rd order ones, they are used for all experiments in the rest of this paper.

Order	2db	6db	10db	14db	18db	Avg
1	48.81	29.40	18.16	12.74	10.52	23.93
2	49.29	29.08	18.01	12.64	10.35	23.87
3	49.24	29.02	18.01	12.72	10.40	23.88

Table 2: CER performance of CI trained polynomials of varying order for Gaussian component mean parameter trajectories.

A range of GVP-HMM based models were then evaluated using CI and CD training methods presented in sections 3 and 4. A detailed description of their configurations and the number of polynomial coefficients to train are shown in table 3. Both standard VP-HMMs and GVP-HMMs allow trajectory modelling of either, or both, of Gaussian component means and variances, shown as "mean", "var" and "mv" in the table. Two GVP-HMM systems modelling the trajectories of 2 or 256 MLLR mean transforms are shown as "tran2" and "tran256". Two GVP-HMMs systems that model the trajectories of all three parameter types using 2 or 256 transforms are shown as "mvt2" and "mvt256" in the bottom section of the table.

Performance of CI trained GVP-HMM systems are shown in table 4. Modelling both Gaussian mean and variance trajectories gave the best performance for standard VP-HMMs, as shown in the first three lines of table 4. Using this CI trained "mv" system, average CER gains of 5.35% (21.1% rel.) and 3.88% (16.3% rel.) absolute across all SNR conditions were obtained over the "mcond" HMM baseline shown in table 1, and the mean only "mean" system of table 4. The 2 transform

System	Trajectory Modelling				#Poly Coef
	mean	var	tran2	tran256	
mean	√	×	×	×	3.66M
var	×	√	×	×	3.66M
mv	√	√	×	×	7.32M
tran2	×	×	√	×	10.8K
tran256	×	×	×	√	1.39M
mvt2	√	√	√	×	7.32M
mvt256	√	√	×	√	8.71M

Table 3: Description of baseline and 2nd order GVP-HMMs: parameter types and the number of polynomial coefficients.

"tran2" system with only 10.8k coefficients to estimate, gave an average error rate of 27.17%. Increasing the number of mean transforms to 256, the "tran256" system with 62% fewer coefficients than the "mean" system, gave a comparable average CER performance of 24.95%. This suggests transform based GVP-HMMs can provide a more compact form of trajectory modelling than VP-HMMs. When all three types of parameter trajectories are modelled, the "mvt2" system gave CER gains of 0.01%-0.49% absolute over the "mv" system on all SRN levels except 2db. The "mvt256" outperformed the "mv" system by 0.33%-1.07% absolute on 2, 6 and 10db data. It gave the best average CER performance among all systems in table 4.

CI Sys	2db	6db	10db	14db	18db	Avg
mean	49.29	29.08	18.01	12.64	10.35	23.87
var	54.12	33.31	20.32	13.31	11.01	26.41
mv	39.63	23.67	15.43	11.31	9.93	19.99
tran2	57.79	33.91	19.95	13.52	10.66	27.17
tran256	51.06	30.36	18.86	13.36	11.12	24.95
mvt2	41.13	23.18	15.15	11.14	9.92	20.16
mvt256	39.05	22.60	15.10	12.06	10.68	19.90

Table 4: CER performance of CI trained polynomials modelling Gaussian mean, variance and transforms trajectories.

Performance of CD trained GVP-HMM systems are shown in table 5. The "mean", "tran2" and "tran256" systems outperformed their respective CI baselines in table 4 by 0.21%-0.50% absolute on average. Again performance of the more compact "tran256" system is close to the "mean" system in table 5. When modelling all three types of parameter trajectories, the "mvt2" system outperformed the "mv" system by 0.19% absolute (1.68% on 2db and 0.51% on 18db). However, increasing the number of transforms to 256 gave no further gains. It is possible that insufficient condition dependent training data was used in equations (8) and (10) to estimate transform trajectories.

CD Sys	2db	6db	10db	14db	18db	Avg
mean	48.81	29.28	17.96	11.96	10.27	23.66
var	60.37	37.54	23.03	15.09	12.23	29.65
mv	44.81	25.70	15.20	11.08	10.22	21.40
tran2	57.05	33.31	19.59	13.20	10.71	26.77
tran256	49.08	30.22	18.34	13.48	11.15	24.45
mvt2	43.13	26.06	15.71	11.46	9.71	21.21
mvt256	43.01	25.90	16.29	12.61	12.61	21.82

Table 5: CER performance of CD trained polynomials modelling Gaussian mean, variance and transform trajectories.

6. Conclusion

Generalized variable parameter HMMs (GVP-HMM) is investigated in this paper. In addition to Gaussian means and variances, it can also compactly model the optimal trajectories of tied linear transforms that can vary with respect to ambient noise level. Experimental results on a medium vocabulary recognition task suggest the proposed method may be useful for noise robust speech recognition. Future research will focus on modelling the trajectories of feature space transforms, discriminative and adaptive training of polynomial coefficients.

7. References

[1] J. A. Arrowood & M. A. Clements, "Using observation uncertainty in HMM decoding", in Proc. ICSLP, 1561-1564, 2002.

[2] A. Bjorck & V. Pereyra (1970). "Solution of Vandermonde Systems of Equations", Mathematics of Computation (American Mathematical Society) 24(112): pp. 893-903.

[3] X. Cui & Y. Gong, "A study of variable-parameter Gaussian mixture hidden Markov modeling for noisy speech recognition", IEEE Transactions on Audio, Speech and Language Processing, 15(4):1366-1376, 2007.

[4] A. P. Dempster, N. M. Laird & D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", Journal of the Royal Statistical Society, 39(1):1-39,1977.

[5] L. Deng, J. Droppo & A. Acero,"Dynamic compensation of HMM variances using the feature enhancement uncertainty computed from a parametric model of speech distortion", IEEE Trans. Speech and Audio, 13(3):412-421, 2005.

[6] J. Droppo, A. Acero & L. Deng, "Uncertainty decoding with SPLICE for noise robust speech recognition", in Proc. ICASSP2002 pp. 57-60.

[7] K. Fujinaga, M. Nakai, H. Shimodaira & S. Sagayama, "Multiple-Regression Hidden Markov Model", In Proc. IEEE ICASSP 2001, Vol 1, pp. 513-516.

[8] M. J. F. Gales, "Maximum likelihood linear transformations for HMM-based speech recognition", Computer Speech and Language, 12(2):75-98, 1998.

[9] C.J. Leggetter & P.C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density HMMs", Computer Speech and Language, 9:171-186,1995.

[10] H. Liao & M. J. F. Gales, "Issues with uncertainty decoding for noise robust speech recognition", Speech Communication, 50:265-277, 2008.

[11] R. Lippmann, E. Martin & D. Paul, "Multi-style training for robust isolated-word speech recognition", in Proc. IEEE ICASSP1987, pp. 705-708.

[12] T. T. Kristjansson & B. J. Frey, "Accounting for uncertainty in observations: A new paradigm for robust speech recognition", in Proc. ICASSP2002, pp. 61-64.

[13] R. Martin, "An efficient algorithm to estimate the instantaneous SNR speech signals", in Proc. Eurospeech, 1093-1096, 1993.

[14] S. Young et al., "The HTK Book Version 3.4.1 ", 2009.

[15] D. Yu, L. Deng, Y. Gong & A. Acero, "Discriminative training of variable-parameter HMMs for noise robust speech recognition," in Proc. Interspeech2008, pp. 285-288.

[16] D. Yu, L. Deng, Y. Gong & A. Acero, "Parameter clustering and sharing in variable-parameter HMMs for noise robust speech recognition," in Proc. Interspeech2008, pp. 1253-1256.

[17] D. Yu, L. Deng, Y. Gong & A. Acero, "A Novel Framework and Training Algorithm for Variable-Parameter Hidden Markov Models", IEEE Transactions on Audio, Speech and Language Processing, Vol 17(7), pp. 1348-1360.