A Study of Variable-Parameter Gaussian Mixture Hidden Markov Modeling for Noisy Speech Recognition

Author: Xiadong Cui, Yifan Gong

Professor:陳嘉平

Reporter:葉佳璋

Abstract—**To improve recognition performance in noisy environments, multicondition training is usually applied in which speech signals corrupted by a variety of noise are used in acoustic model training. Published hidden Markov modeling of speech uses multiple Gaussian distributions to cover the spread of the speech distribution caused by noise, which distracts the modeling of speech event itself and possibly sacrifices the performance on clean speech. In this paper, we propose a novel approach which extends the conventional Gaussian mixture hidden Markov model (GMHMM) by modeling state emission parameters (mean and variance) as a polynomial function of a continuous environment-dependent variable. At the recognition time, a set of HMMs specific to the given value of the environment variable is instantiated and used for recognition. The maximum-likelihood (ML) estimation of the polynomial functions of the proposed variable-parameter GMHMM is given within the expectation-maximization (EM) framework. Experiments on the Aurora 2 database show significant improvements of the variable-parameter Gaussian mixture HMMs compared to the conventional GMHMMs.**

Introduction

- Typical Gaussian mixture hidden Markov modeling of speech uses multiple Gaussian distributions to ***cover the spread of feature distributions due to events and noises***.

- The variable parameter Gaussian mixture HMM (VP-GMHMM) proposed in this paper extends existing HMM by allowing conventional Gaussian HMM (CV-GMHMM) parameter **to *change as function of a continuous variable that depends on the environment***

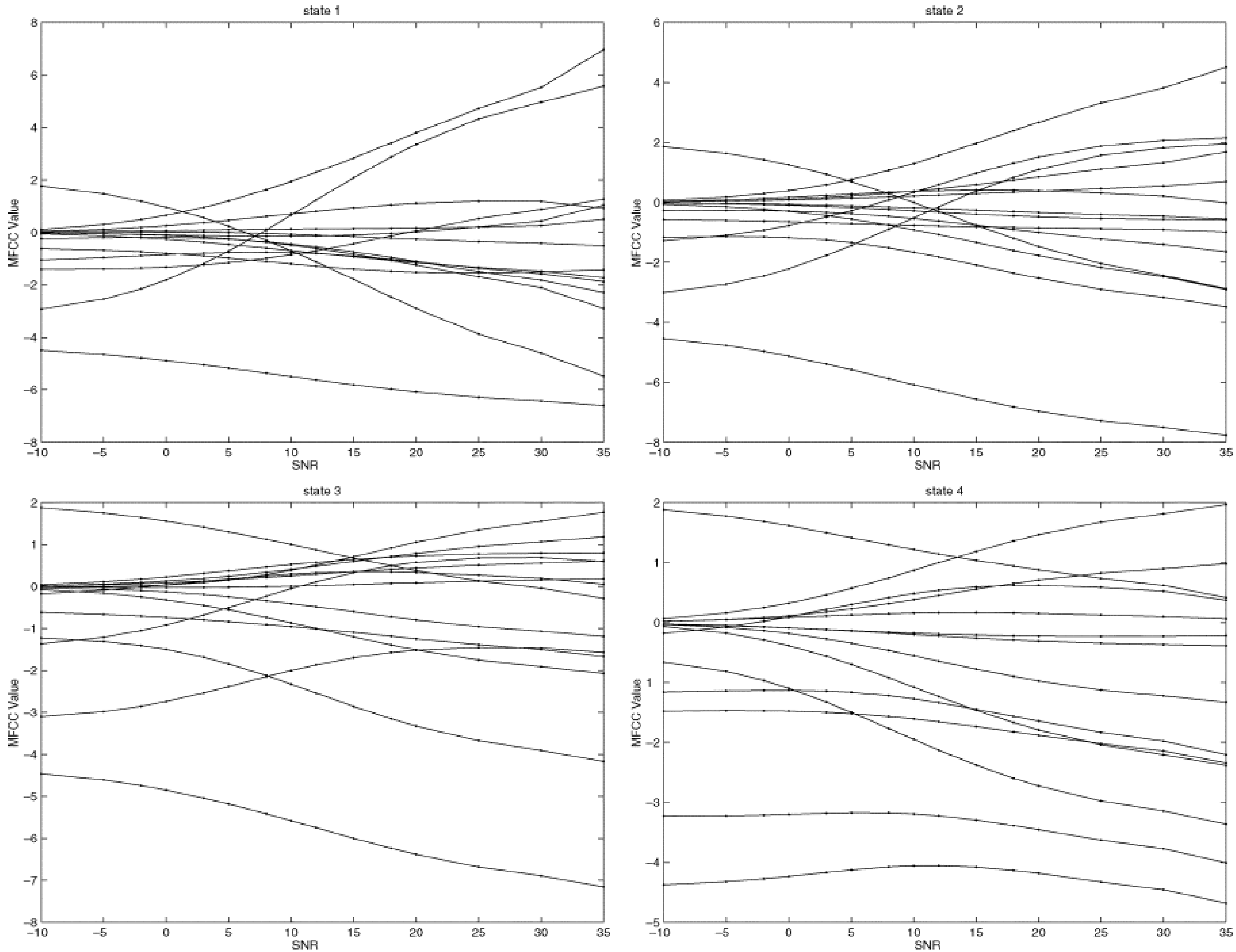


Fig. 1. Variation of elements of the mel frequency cepstral coefficient (MFCC) mean vector $(C_1 \cdots C_7; \triangle C_1 \cdots \triangle C_7)$ against SNR for the four states in the /ah/ sound of female speakers.

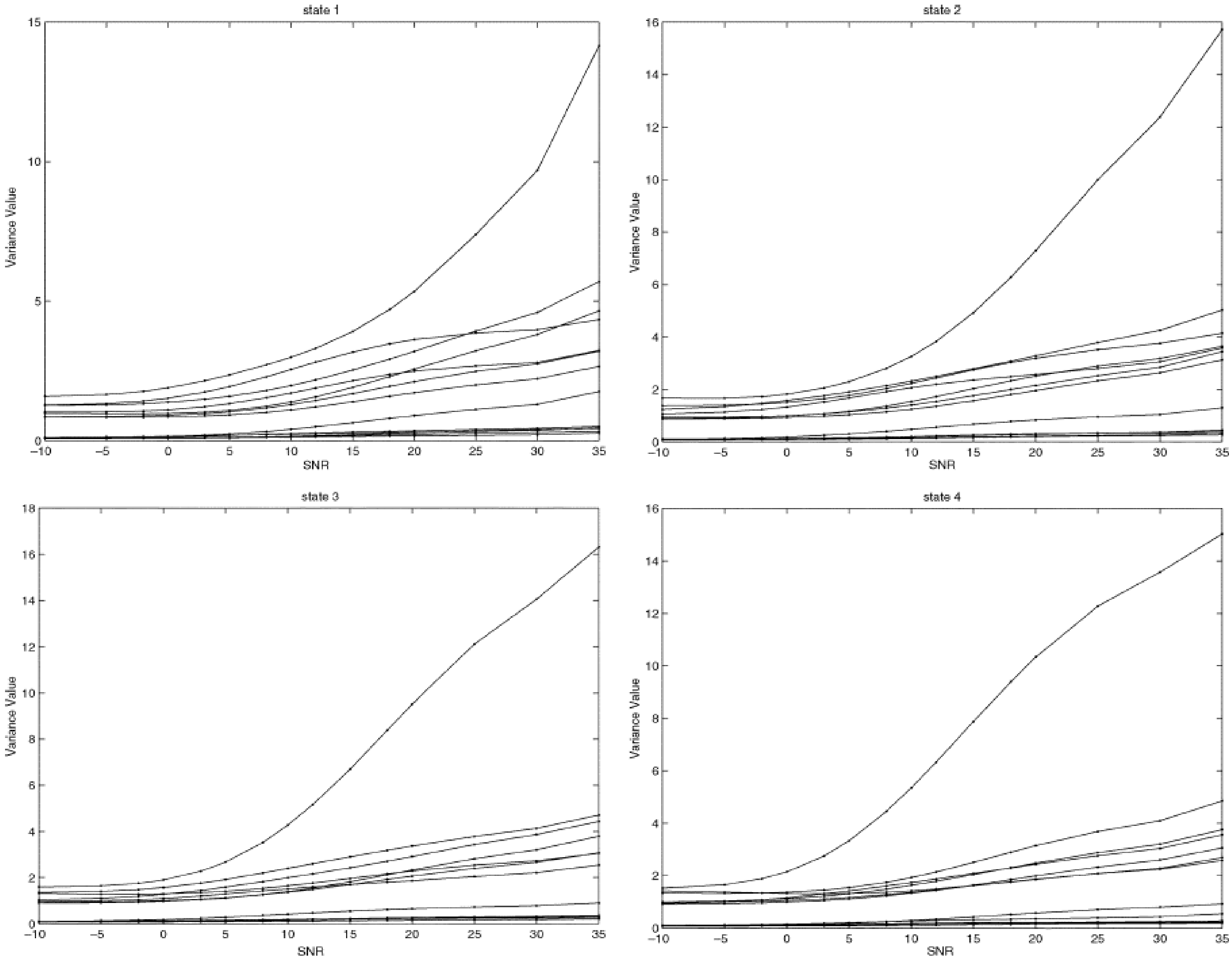


Fig. 2. Variation of elements of the MFCC variance vector $(C_1 \cdots C_7; \triangle C_1 \cdots \triangle C_7)$ against SNR for the four states in the /ah/ sound of female speakers.

Motivation

- Motivation of VP-GMHMM which models CV-GMHMM state emission parameter as a function of a continuous environmental variable.
  - ✓ CV-GMHMM is imperfect and inadequate to deal with the phenomena illustrated in Figures.
    - ➢ The mean of Gaussian distribution of the observed feature is a function of SNR.
    - ➢ The variances tend to be much similar in value when the environment is very noisy.

Reasons

- In VP-GMHMM, means and variances are described by polynomials of environment v
  - With high degree, polynomials can approximate any continuous function.
  - The derivatives of polynomial function are easy to obtain.
  - The change of means and covariance in terms of the environment is smooth and can be modeled by low-order polynomials.

Formulation

- We assume observations in each HMM state i has a multivariate Gaussian mixtures distribution

$$p(o_t \mid s_t = i) = \sum_k \alpha_{ik} b_{ik}(o_t)$$

where $b_{ik}(o_t) \sim N(o_t; \mu_{ik}; \Sigma_{ik})$ is the kth multivariate Gaussian mixtures distribution in state i with weight α_{ik}

  - N and M: number of HMM states and Gaussian mixtures per state;
  - Ω_s and Ω_m : indices of the set of states and Gaussian mixtures;
  - $O = \{O^1, O^2, \cdots O^R\}$: R observation sequences;
  - $O^r = \{O_1^r, O_2^r, \cdots O_{T_r}^r\}$: rth observation sequence with T_r frame vectors;
  - $s^r = \{s_1^r, s_2^r, \cdots s_{T_r}^r\}$: sate sequence for the rth observation;
  - $\Xi^r = \{\xi_1^r, \xi_2^r, \cdots \xi_{T_r}^r\}$: Gaussian mixture sequence for the rth observation;

Form of mean and covariance

- Mean vector and covariance matrix

$$\mu_{ik}(v) = \sum_{j=0}^{P} c_{ikj}^{(m)} v^j$$

$$\Sigma_{ik}(v) = \Lambda(v)\Sigma_{ik}^{0}$$

$$\Lambda(v) = \begin{pmatrix} e^{\sum_{j=0}^{P} c_{ik0j}^{(v)} v^j} & & \\ & \ddots & \\ & & e^{\sum_{j=0}^{P} c_{ikDj}^{(v)} v^j} \end{pmatrix}$$

  - Environmental variable v is a scalar which can be regarded as SNR
  - Coefficient $c_{ikj}^{(m)} = \left[c_{ik0j}^{(m)} \cdots c_{ikDj}^{(m)} \right]^T$ is a vector with the same dimension as feature vectors
  - Σ_{ik}^{0} : original covariance matrix
  - Λ : scaling matrix which is a diagonal matrix
  - $\cdot^{(m)}$ and $\cdot^{(v)}$: the polynomial coefficients of mean and variance

Estimation of Mean Polynomial

- According to the EM algorithm, the auxiliary function

$$Q_b\left(\lambda,\overline{\lambda}\right)=\sum_{r=1}^{R}\sum_{i\in\Omega_s}\sum_{k\in\Omega_m}\sum_{t=1}^{T_r}\gamma_t^r(i,k)\cdot\log b_{ik}\left(o_t^r\right)$$

$$\gamma_t^r(i,k)=p\left(s_t^r=i,\xi_t^r=k\mid O^r,\overline{\lambda}\right)$$

- Taking the derivative with respect to $c_{ikj}^{(m)}$ and setting to zero

$$\sum_{p=0}^{P}\sum_{r=1}^{R}\sum_{t=1}^{T_r}\gamma_t^r(i,k)\cdot\Sigma_{ik}^{-1}(v_r)\cdot v_r^{p+j}\cdot c_{ikp}^{(m)}$$

$$=\sum_{p=0}^{P}\sum_{r=1}^{R}\sum_{t=1}^{T_r}\gamma_t^r(i,k)\cdot\Sigma_{ik}^{-1}(v_r)\cdot v_r^{p+j}\cdot o_t^r,\ j=0,1,\cdots,P$$

- Define for each state i and mixture component k

$$l_{ik}(\zeta,\eta,\alpha,\beta)=\sum_{r=1}^{R}\sum_{t=1}^{T_r}\gamma_t^r(i,k)\cdot\Sigma_{ik}^{-1}(v_r)\cdot\zeta^{\alpha}\cdot\eta^{\beta}$$

$$\sum_{p=0}^{P}l_{ik}(v_r,v_r,j,p)\cdot c_{ikp}^{(m)}=l_{ik}\left(v_r,o_t^r,j,1\right),\ j=0,\cdots,P$$

Estimation of Mean Polynomial

- A linear equation system with P+1 variable defined in the vector space, which can be written a compact form

$$A_{ik} c_{ik}^{(m)} = b_{ik}$$

- A_{ik} is a $(P+1) \times (P+1)$ block matrix, and $u_{ik}(j,p)$ itself is a $D \times D$ matrix

$$A_{ik} = \begin{pmatrix} u_{ik}(0,0) & \cdots & u_{ik}(0,P) \\ \vdots & u_{ik}(j,p) & \vdots \\ u_{ik}(P,0) & \cdots & u_{ik}(P,P) \end{pmatrix}, \; u_{ik}(j,p) = l_{ik}(v_r, v_r, j, p)$$

- b_{ik} is a P+1 block vector, where v_{ikj} itself is a D dimension vector

$$b_{ik} = \left[v_{iko}^{T} \cdots v_{ikj}^{T} \cdots v_{ikP}^{T} \right]^{T}, \; \mathrm{v}_{ikj}(j,p) = l_{ik}\left(v_r, o_t^r, j, 1\right)$$

- c_{ik} is a P+1 block vector , where c_{ikj} itself is a D dimension vector

$$c_{ik}^{(m)} = \left[c_{ik0}^{(m)T} \cdots c_{ikj}^{(m)T} \cdots c_{ikP}^{(m)T} \right]^{T}$$

Estimation of Mean Polynomial

- For the diagonal covariance matrix case, as the dimensions of μ_{ik} are not correlated, c_{ik} can be solved dimension-by-dimension.
  - ✓ Each element of $A_{ik}, u_{ik}(j,p)$ is in this case a diagonal matrix
    - ➢ $A_{ik}^{(d)}$ be the $(P+1)\times(P+1)$ matrix made of the dth diagonal element of each element of A_{ik}
    - ➢ $b_{ik}^{(d)}$ be the vector made of the dth dimension of each element of b_{ik}
    - ➢ $c_{ik}^{(m)(d)}$ be the vector made of the dth dimension of each element of $c_{ik}^{(m)}$
- We can easily obtain $c_{ik}^{(m)(d)}$ by solving the linear system

$$A_{ik}^{(d)} c_{ik}^{(m)(d)} = b_{ik}^{(d)}, \qquad d = 1, 2, \cdots, D$$

Estimation of Variance Scaling Polynomial

- No variance scaling is being applied (special case)

$$\Sigma_{ik} = \frac{\sum_{r=1}^{R} \sum_{t=1}^{T_r} \gamma_t^r(i,k) \left(o_t^r - c_{ikj}^{(m)} v_r^j \right) \left(o_t^r - c_{ikj}^{(m)} v_r^j \right)^T}{\sum_{r=1}^{R} \sum_{t=1}^{T_r} \gamma_t^r(i,k)}$$

- If the scaling matrix Λ is applied to the origin covariance matrix, the ML estimation is performed over the diagonal element of Λ
- Rewrite the Gaussian mixture pdf as

$$b_{ik}\left(o_t^r\right) = \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi \cdot \sigma_{ikd}^2 \cdot e^{\Sigma_{j=0}^{P} c_{ikdj}^{(v)} v_r^j}}} \cdot e^{-\frac{\left(o_{td}^r - \mu_{ikd}(v_r)\right)^2}{2 \cdot \sigma_{ikd}^2 \cdot e^{\Sigma_{j=0}^{P} c_{ikdj}^{(v)} v_r^j}}}$$

Estimation of Variance Scaling Polynomial

- Taking the derivative with respect to $c_{ikdj}^{(v)}$ and setting to zero

$$\sum_{r=1}^{R} e^{-\Sigma_{p=0}^{P} c_{ikdj}^{(v)} v_r^p} \cdot v_r^j \sum_{t=1}^{T_r} \gamma_t^r(i,k) \cdot \frac{\left(o_{td}^r - \mu_{ikd}(v_r)\right)^2}{\sigma_{ikd}^2} = \sum_{r=1}^{R} \sum_{t=1}^{T_r} \gamma_t^r(i,k) \cdot v_r^j,\ j = 0,1,\cdots,P$$

$$F_j\left(c_{ikd0}^{(v)} \cdots c_{ikdP}^{(v)}\right) \triangleq \sum_{r=1}^{R} e^{-\Sigma_{p=0}^{P} c_{ikdj}^{(v)} v_r^p} \cdot v_r^j \sum_{t=1}^{T_r} \gamma_t^r(i,k) \cdot \frac{\left(o_{td}^r - \mu_{ikd}(v_r)\right)^2}{\sigma_{ikd}^2} - \sum_{r=1}^{R} \sum_{t=1}^{T_r} \gamma_t^r(i,k) \cdot v_r^j,\ j = 0,1,\cdots,P$$

- Solve the P+1 simultaneous nonlinear equations for the P+1 unknowns $\left\{c_{ikd0}^{(v)} \cdots c_{ikdP}^{(v)}\right\}$, and obtain the re-estimation of the Pth-order variance scaling polynomial for dth dimension

$$\begin{cases} F_0\left(c_{ikd0}^{(v)} \cdots c_{ikdP}^{(v)}\right) = 0 \\ \vdots \\ F_P\left(c_{ikd0}^{(v)} \cdots c_{ikdP}^{(v)}\right) = 0 \end{cases}$$

Estimation of Variance Scaling Polynomial

- The number of summation in F_j is proportional to the number of utterance in the training set due to the real-valued environment variable v_r .
- To reduce the number of summation terms, the SNR range is quantized into P+1 values in this paper

$$\left(v_j, v_{j+1}\right] \to v_j^*,\ j = 0, \cdots, P$$

where $v_0 = -\infty$ and $\mathrm{v}_{\mathrm{p+1}} = +\infty$

- Can be written as

$$F_j\left(c_{ikd0}^{(v)} \cdots c_{ikdP}^{(v)}\right) \triangleq \sum_{r=1}^{R} \sum_{v_r \in \left[v_q, v_{q+1}\right)} e^{-\Sigma_{p=0}^{P} c_{ikdj}^{(v)} v_q^{*p}} \cdot v_q^{*j} \sum_{t=1}^{T_r} \gamma_t^r(i,k) \cdot \frac{\left(o_{td}^r - \mu_{ikd}\left(v_q^*\right)\right)^2}{\sigma_{ikd}^2} - \sum_{r=1}^{R} \sum_{t=1}^{T_r} \gamma_t^r(i,k) \cdot v_q^*,\ j = 0, 1, \cdots, P$$

Estimation of Variance Scaling Polynomial

- Be converted into a linear system form

$$\Psi_{ikd} \cdot k_{ikd} = \Phi_{ikd}$$

- Where $\Psi_{ikd} = \begin{pmatrix} \Psi_{ikd}(0,0) & \cdots & \Psi_{ikd}(0,P) \\ \vdots & \Psi_{ikd}(j,p) & \vdots \\ \Psi_{ikd}(P,0) & \cdots & \Psi_{ikd}(P,P) \end{pmatrix}$

with $\Psi_{ikd}(j,p) \triangleq \sum_{v_r \in [v_q, v_{q+1})} \sum_{t=1}^{T_r} \gamma_t^r(i,k) \cdot \frac{\left(o_{td}^r - \mu_{ikd}(v_q^*)\right)^2}{\sigma_{ikd}^2} \cdot v_q^{*j}$ and

$$\Phi_{ikd} = \left[\Phi_{ikd}(0) \cdots \Phi_{ikd}(P)\right]^T, \Phi_{ikd}(j,p) \triangleq \sum_{v_r \in [v_q, v_{q+1})} \sum_{t=1}^{T_r} \gamma_t^r(i,k) \cdot v_q^{*j}$$

$$k_{ikd} = \left[k_{ikd}(0) \cdots k_{ikd}(P)\right]^T, k_{ikd}(q) \triangleq e^{-\sum_{P=0}^{P} c_{ikdp}^{(v)} \cdot v_q^{*P}}$$

- k_{ikd} can be obtain as

$$k_{ikd} = \Psi_{ikd}^{-1} \Phi_{ikd}$$

Estimation of Variance Scaling Polynomial

- Based on k_{ikd} and its definition, the coefficient of the variance scaling polynomial $c_{ik}^{(v)} = \{c_{kid0} \cdots c_{kidP}\}$ can be readily obtained by another linear system equation

$$\begin{pmatrix} 1 & v_0^{*1} & \cdots & v_0^{*P} \\ 1 & v_1^{*1} & \cdots & v_1^{*P} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & v_P^{*1} & \cdots & v_P^{*P} \end{pmatrix} \cdot \begin{bmatrix} c_{ikd0} \\ c_{ikd1} \\ \vdots \\ c_{ikdP} \end{bmatrix} = \begin{bmatrix} -\ln k_{ikd0}(0) \\ -\ln k_{ikd1}(1) \\ \vdots \\ -\ln k_{ikdP}(P) \end{bmatrix}$$

Polynomial Selection for Unknown Noise

- An ML approach is used to polynomial selection for unknown noise.
- A polynomial against the environmental variable is estimated from only the noise frames at *the beginning of training utterance for each type of noise in the training set*
  - Assume there are K candidate polynomials $\Lambda = \{\lambda^1, \cdots, \lambda^K\}$
  - Assume that the first T_0 frames have only noise that they are independent
  - The environment v is first measured for the utterance and the best matched polynomial is chosen as the one which maximize the first T_0 noise frames

$$\lambda_0 = \arg\max_{\lambda^k \in \Lambda} P\left(O^1 \cdots O^{T_0} \mid \lambda^k(v)\right)$$

Experiment Results

- Aurora2 database
  - Experiments are Performed on Set A and B
  - Four types of noise in the training set which include subway, babble, car and exhibition
- MFCC feature extracted and utterance SNR is estimated from the speech signals
  - 12 static MFCCs plus log energy
  - There first- and second-order derivatives
- We estimate utterance SNR by the minimum statistic tracking algorithm

Performance Versus Order of Polynomial

- Experiments are conducted with first, second, and third order of mean and variance scaling polynomials.
- In terms of variance scaling, the SNR quantization is performed as
  - First-order polynomial: $(-\infty, 20] \rightarrow 10$
    $(20, +\infty) \rightarrow 25$
  - Second-order polynomial: $(-\infty, 10] \rightarrow 5$
    $(10, 25] \rightarrow 15$
    $(25, +\infty) \rightarrow 35$
  - Third-order polynomial: $(-\infty, 10] \rightarrow 5$
    $(10, 20] \rightarrow 15$
    $(20, 30] \rightarrow 25$
    $(30, +\infty) \rightarrow 35$

Performance on Set A Four Types of Background Noise With Different Orders of Polynomials Without (VP-GMHMM1) and With (VP-GMHMM2) Variance Scaling. There Are Three Gaussian Mixtures in Each State

	VP-GMHMM1			VP-GMHMM2		
	1st	2nd	3rd	1st	2nd	3rd
Subway	93.78	93.86	92.94	93.87	93.77	93.22
Babble	90.98	91.07	90.25	91.16	91.24	91.44
Car	92.09	92.92	91.83	92.83	93.10	92.15
Exhibition	91.89	92.12	91.52	92.51	92.75	91.79

Performed on Matched Noise

- To quantitatively compare VP-HMM and CV-HMM
  - HMMs with one, two, three, six Gaussian mixtures per state
  - Four types of background noise from Set A
  - These four types of noise in the test set match the noise in the training set
  - VP-GMHMM1: the VP-GMHMM with no variance scaling
  - VP-GMHMM2: the VP-GMHMM with variance scaling

Subway and Car

PERFORMANCE ON SUBWAY NOISE WITH ONE (1M), TWO (2M), THREE (3M), AND SIX (6M) GAUSSIAN MIXTURES

	Experimental Conditions					
	Clean	20 dB	15 dB	10 dB	5 dB	0 dB
1M (CV-GMHMM)	97.08	96.65	95.67	93.40	86.31	56.13
1M (VP-GMHMM1)	97.57	97.45	96.50	93.80	88.89	68.59
1M (VP-GMHMM2)	97.86	97.67	96.99	94.44	90.51	76.39
2M (CV-GMHMM)	97.91	97.76	96.39	95.58	90.67	68.96
2M (VP-GMHMM1)	98.60	98.42	97.85	96.40	91.50	76.48
2M (VP-GMHMM2)	98.64	98.45	97.84	96.48	92.45	78.51
3M (CV-GMHMM)	98.31	98.13	97.45	95.86	92.26	73.90
3M (VP-GMHMM1)	98.78	98.59	97.91	96.96	92.82	78.11
3M (VP-GMHMM2)	98.83	98.71	97.94	97.06	92.80	77.25
6M (CV-GMHMM)	98.74	98.37	97.85	96.62	93.55	77.22
6M (VP-GMHMM1)	98.86	98.86	98.13	97.30	93.76	78.08
6M (VP-GMHMM2)	98.93	98.93	98.19	97.24	93.94	78.20

PERFORMANCE ON CAR NOISE WITH ONE (1M), TWO (2M), THREE (3M), AND SIX (6M) GAUSSIAN MIXTURES

	Experimental Conditions					
	Clean	20 dB	15 dB	10 dB	5 dB	0 dB
1M (CV-GMHMM)	95.97	96.03	95.49	92.21	81.01	47.40
1M (VP-GMHMM1)	97.70	96.84	96.30	93.43	84.48	56.39
1M (VP-GMHMM2)	97.96	97.25	96.93	94.87	88.24	70.24
2M (CV-GMHMM)	98.00	97.19	97.25	95.31	88.99	61.76
2M (VP-GMHMM1)	98.39	97.55	97.49	95.37	90.18	70.84
2M (VP-GMHMM2)	98.36	98.06	97.85	96.24	91.19	72.63
3M (CV-GMHMM)	98.27	97.85	97.73	96.51	91.61	68.24
3M (VP-GMHMM1)	98.75	98.36	98.18	96.81	92.01	73.34
3M (VP-GMHMM2)	98.70	98.48	98.30	96.87	91.88	74.37
6M (CV-GMHMM)	98.69	98.33	98.12	97.01	92.48	73.31
6M (VP-GMHMM1)	98.93	98.72	98.48	97.28	92.60	75.22
6M (VP-GMHMM2)	98.87	98.90	98.60	97.28	92.81	74.91

Babble and Exhibition

PERFORMANCE ON BABBLE NOISE WITH ONE (1M), TWO (2M), THREE (3M), AND SIX (6M) GAUSSIAN MIXTURES

	Experimental Conditions					
	Clean	20 dB	15 dB	10 dB	5 dB	0 dB
1M (CV-GMHMM)	96.43	96.55	95.68	92.93	82.95	52.39
1M (VP-GMHMM1)	97.97	97.35	96.26	93.32	84.49	55.59
1M (VP-GMHMM2)	98.40	97.64	96.55	94.62	86.40	62.00
2M (CV-GMHMM)	98.13	97.64	97.10	95.19	87.27	59.95
2M (VP-GMHMM1)	98.79	97.85	97.19	95.86	89.12	64.96
2M (VP-GMHMM2)	98.74	98.13	97.85	96.01	89.57	66.51
3M (CV-GMHMM)	98.37	97.70	97.58	95.68	88.81	62.70
3M (VP-GMHMM1)	98.88	97.97	97.49	96.16	89.84	66.09
3M (VP-GMHMM2)	98.88	98.25	97.88	96.37	89.99	66.08
6M (CV-GMHMM)	98.88	98.13	97.52	96.04	89.75	65.11
6M (VP-GMHMM1)	98.88	98.28	97.76	96.49	90.42	66.90
6M (VP-GMHMM2)	98.79	98.43	98.04	96.58	90.55	66.45

PERFORMANCE ON EXHIBITION NOISE WITH ONE (1M), TWO (2M), THREE (3M), AND SIX (6M) GAUSSIAN MIXTURES

	Experimental Conditions					
	Clean	20 dB	15 dB	10 dB	5 dB	0 dB
1M (CV-GMHMM)	96.36	95.59	94.66	90.90	82.84	56.80
1M (VP-GMHMM1)	98.40	97.25	96.11	92.38	85.93	66.49
1M (VP-GMHMM2)	98.24	97.62	96.98	93.98	88.65	72.88
2M (CV-GMHMM)	97.50	96.39	95.96	93.15	87.10	66.24
2M (VP-GMHMM1)	98.83	97.93	97.41	94.88	88.68	71.61
2M (VP-GMHMM2)	98.86	97.84	97.59	95.16	90.04	75.59
3M (CV-GMHMM)	98.46	97.69	97.04	94.85	89.76	71.43
3M (VP-GMHMM1)	99.14	98.27	97.38	95.40	89.73	72.82
3M (VP-GMHMM2)	98.95	98.33	97.59	95.47	90.74	75.04
6M (CV-GMHMM)	98.55	98.15	97.22	95.22	90.19	75.10
6M (VP-GMHMM1)	99.17	98.55	97.96	95.56	90.69	75.93
6M (VP-GMHMM2)	99.23	98.52	98.09	95.99	90.77	76.03

Performed on Unknown Noise

- Performance on Set B with four types of background noise that do not appear in the training data.
- The CV-GMM are trained by pooling all the data in the training set with subway, babble, car, exhibition noise.
- VP-GMHMM1 is used since its variance, which are constant, are trained across all the SNR condition rather than tuned for the particular type of noise VP-GMHMM2.

PERFORMANCE ON RESTAURANT NOISE WITH ONE (1M), TWO (2M), THREE (3M), AND SIX (6M) GAUSSIAN MIXTURES

	Experimental Conditions					
	Clean	20 dB	15 dB	10 dB	5 dB	0 dB
1M (CV-GMHMM)	96.90	94.20	92.02	87.75	78.23	53.09
1M (VP-GMHMM1)	98.13	96.84	95.61	90.51	80.29	54.07
2M (CV-GMHMM)	98.04	96.35	95.00	91.22	82.78	57.26
2M (VP-GMHMM1)	98.56	97.61	96.53	92.29	84.36	60.33
3M (CV-GMHMM)	98.37	96.59	95.33	91.68	83.54	58.89
3M (VP-GMHMM1)	98.86	98.04	96.59	92.85	84.00	59.16
6M (CV-GMHMM)	98.93	97.33	95.49	92.05	82.71	58.67
6M (VP-GMHMM1)	99.08	98.19	96.75	93.12	84.96	60.76

PERFORMANCE ON AIRPORT NOISE WITH ONE (1M), TWO (2M), THREE (3M), AND SIX (6M) GAUSSIAN MIXTURES

	Experimental Conditions					
	Clean	20 dB	15 dB	10 dB	5 dB	0 dB
1M (CV-GMHMM)	96.75	95.73	94.03	90.21	80.63	54.87
1M (VP-GMHMM1)	97.94	96.96	95.40	91.22	81.22	59.04
2M (CV-GMHMM)	97.85	97.04	95.91	92.93	85.28	63.16
2M (VP-GMHMM1)	98.42	97.73	96.84	93.69	83.97	61.49
3M (CV-GMHMM)	98.42	97.49	96.39	93.43	86.78	65.28
3M (VP-GMHMM1)	98.66	98.00	97.13	93.37	88.13	65.30
6M (CV-GMHMM)	98.84	98.00	96.69	93.73	86.45	66.54
6M (VP-GMHMM1)	98.99	98.48	97.67	94.09	88.34	66.16

PERFORMANCE ON STREET NOISE WITH ONE (1M), TWO (2M), THREE (3M), AND SIX (6M) GAUSSIAN MIXTURES

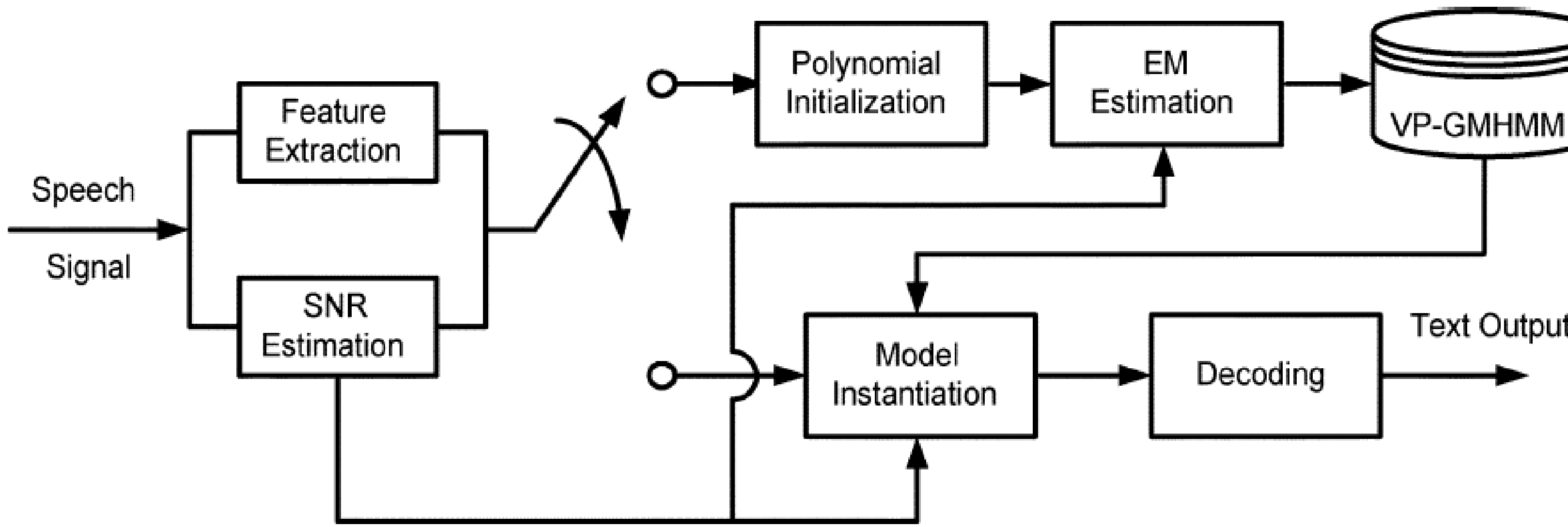
	Experimental Conditions					
	Clean	20 dB	15 dB	10 dB	5 dB	0 dB
1M (CV-GMHMM)	96.98	95.53	94.20	90.27	77.90	49.85
1M (VP-GMHMM1)	97.85	96.86	95.31	91.24	77.69	58.51
2M (CV-GMHMM)	97.94	97.13	95.50	93.68	84.07	59.31
2M (VP-GMHMM1)	98.58	97.82	96.89	94.35	86.38	60.80
3M (CV-GMHMM)	98.43	97.58	96.22	94.38	86.22	62.09
3M (VP-GMHMM1)	98.61	98.22	97.19	94.95	88.13	58.01
6M (CV-GMHMM)	98.85	97.85	96.83	94.95	86.88	64.48
6M (VP-GMHMM1)	98.96	98.61	97.40	94.71	88.28	59.49

PERFORMANCE ON STATION NOISE WITH ONE (1M), TWO (2M), THREE (3M), AND SIX (6M) GAUSSIAN MIXTURES

	Experimental Conditions					
	Clean	20 dB	15 dB	10 dB	5 dB	0 dB
1M (CV-GMHMM)	96.64	95.16	92.84	88.46	75.44	40.05
1M (VP-GMHMM1)	98.43	97.04	94.45	89.76	78.96	52.11
2M (CV-GMHMM)	97.87	96.42	94.79	91.79	81.52	51.16
2M (VP-GMHMM1)	98.70	97.81	95.56	92.32	84.84	57.74
3M (CV-GMHMM)	98.33	96.85	95.25	92.56	83.15	56.06
3M (VP-GMHMM1)	98.92	97.90	96.30	93.15	83.74	59.77
6M (CV-GMHMM)	98.73	97.41	95.80	92.97	84.29	60.32
6M (VP-GMHMM1)	99.11	98.49	96.85	93.95	84.88	59.98

Training and Recognition Procedures

- In the training stage
  - Training data are categorized into several subsets in terms of their SNRs. (e.g. 20, 10 db, etc)
  - CV-GMHMMs are trained for each subset
  - The Gaussian means of the same mixture are regressed with respect SNR to obtain the initial mean polynomials.
- In the recognition stage
  - Utterance SNR is estimated for the speech signal
  - One set of environment-dependent HMM parameters is instantiated based on the SNR estimate to decode the speech signal.


Speech
Signal
Feature Extraction
SNR Estimation
Polynomial Initialization
EM Estimation
VP-GMHMM
Model Instantiation
Decoding
Text Output