Subspace Modeling and Selection for Noisy Speech Recognition

Jen-Tzung Chien and Chuan-Wei Ting

Department of Computer Science and Information Engineering,
National Cheng Kung University, Tainan, Taiwan, ROC
{chien, cwting}@chien.csie.ncku.edu.tw

ABSTRACT

This paper presents a new subspace modeling and selection approach for noisy speech recognition. In subspace modeling, we develop *factor analysis* (FA) for representing noisy speech. FA is a data generation model where the common factors are extracted with factor loading matrix and specific factors. We bridge the connection of FA to signal subspace (SS) approach. Interestingly, FA partitions noisy speech space into a *principal subspace* containing speech and noise and a *minor subspace* containing residual speech and residual noise. To estimate clean speech, we minimize the energies of speech distortion in principal subspace as well as minor subspace. More importantly, in subspace selection, we explore *optimal subspace partition* via solving *hypothesis test* problems. We test the equivalence of eigenvalues in minor subspace so as to determine subspace dimension. To fulfill FA spirit, we further examine the hypothesis of uncorrelated residual speech. Optimal solutions are realized through *likelihood ratio test* with the approximated chi-square distributions as test statistics. Subspace partition is performed according to the confidence towards rejecting null hypotheses. In the experiments on Aurora2 database, FA outperforms SS in subspace modeling. New selection algorithms effectively determine subspace dimension for noisy speech recognition.

Index Terms: subspace modeling, subspace selection, factor analysis, speech recognition

1. INTRODUCTION

Automatic speech recognition (ASR) is crucial for building human computer interaction systems. However, in noisy environments, system performance is deteriorated. To achieve robustness in noisy speech recognition, one popular approach is to perform front-end denoising process. The enhanced speech can be recognized by matching with clean speech models. Basically, front-end processing is advantageous because the computation cost is low and no adaptation data is required beforehand. In the literature, spectral subtraction and signal subspace were acted as popular speech enhancement approaches. Spectral subtraction [3] was exploited to estimate clean speech by subtracting additive noise from noisy speech in spectral domain. This approach suffered from the annoying tonal "musical noise" in the produced residual noise. Accordingly, Ephraim and Van Trees [7] presented signal subspace approach where a signal subspace and a noise subspace were partitioned via eigen-decomposition of noisy speech. Clean signal was estimated from signal subspace. The complementary noise subspace was removed. Signal estimation was done by minimizing signal distortion while limiting energy of residual noise below a threshold. Nevertheless, the denoising performance highly relied on the partition of signal/noise subspaces. In previous studies, subspace dimension was empirically determined.

In this paper, we are motivated to generalize subspace paradigm for noisy speech analysis and recognition. The generalization is twofold. First, we extend the representation of noisy speech using factor analysis (FA) which is a powerful data analysis mechanism originated in societies of social science and machine learning [2]. FA is closely related to principal component analysis developed for feature dimension reduction. Typically, FA extracts common factors which are useful to represent correlation between different features. This property is crucial for modeling full covariance matrix which is essential in subspace modeling. FA was applied for estimating covariance matrices and building hidden Markov model (HMM) based ASR systems [9]. Here, we present FA subspace modeling for noisy speech recognition. Noisy speech is decomposed into principal factors and minor factors, or correspondingly projected onto two subspaces. The first subspace represents the majority of clean speech with little noise information. The other subspace is a residual subspace containing noise and residual speech. FA can recover clean signal from principal subspace as well as minor subspace instead of using SS where the noise subspace was entirely discarded [7]. In the second generalization, the subspace approach is activated with adaptive selection of subspace dimension. We exploit two selection solutions to testing whether the decomposition of covariance matrices satisfy FA modeling. Experiments on Aurora2 database show that proposed subspace modeling and selection attain desirable speech recognition performance in presence of different signal-to-noise ratios.

2. SUBSPACE MODELING

2.1. Modeling of Noisy Signal

Signal subspace (SS) [7] is well-known for modeling and enhancing noisy signal $\mathbf{z}$. Assuming *K*-dimensional clean signal $\mathbf{y}$ is corrupted by additive noise $\mathbf{n}$, SS expresses noisy signal by

$$\mathbf{z} = \mathbf{y} + \mathbf{n} = \mathrm{W}\mathbf{x} + \mathbf{n}\,. \tag{1}$$

In (1), clean signal is assumed to be in a *M*-dimensional subspace ($M < K$) of Euclidean space $\Re^K$ with spanning set consisting of columns of W. $\mathbf{x}$ denotes the coordinate vector. Through eigendecomposition of covariance matrix of $\mathbf{z}$, we can construct a *signal subspace* and a complementary *noise subspace* using eigenvectors corresponding to M positive eigenvalues and $K - M$ zero eigenvalues, respectively. Signal subspace coincides with the subspace span W for constructing clean signal. Clean signal can be estimated by minimizing signal distortion and simultaneously limiting permissible level of residual noise [7].

More attractively, we present FA modeling of noisy signal. The basic idea of FA is to use a $K \times M$ factor loading matrix Φ, a *M*-dimensional common factor vector $\mathbf{f}$ and a *K*-dimensional specific factor vector $\mathbf{r}$ to represent signal $\mathbf{z}$ as

$$\mathbf{z} = \Phi\mathbf{f} + \mathbf{r}\,. \tag{2}$$

Specific factors $\mathbf{r}$ are viewed as residual signal or modeling error. In analysis of noisy speech, these factors carry information of residual speech as well as residual noise. FA modeling should possess the following properties [2]. Common factors and residual signal are uncorrelated $E[\mathbf{fr}^T]=0$ and Gaussian distributed with zero mean vectors $E[\mathbf{f}]=E[\mathbf{r}]=0$ and diagonal covariance matrices $E[\mathbf{ff}^T]=I_M$ and $E[\mathbf{rr}^T]=\Psi$. Then, In general, the dependencies among features can be properly modeled by common factors. The covariance matrix of residual signal Ψ should be diagonal. Namely, residual factors are specific and uncorrelated. Observation vectors are then Gaussian distributed with $\mathbf{z} \sim N(0, \Phi\Phi^T + \Psi)$.

Different from principal component analysis (PCA) [8] developed for dimension reduction, FA aims to extract common factors for data modeling. PCA finds principal components for representing majority of data variability while FA characterizes data dependencies using a small number of common factors. Using FA, factor loadings Φ should be estimated. One approach of finding Φ was derived from probabilistic PCA model [6][10] using maximum likelihood estimation. Also, parameter Φ can be estimated via eigendecomposition of covariance matrix [8]

$$R_z = \Phi\Phi^T + \Psi = W\Lambda W^T = W_p\Lambda_p^{1/2}\Lambda_p^{1/2}W_p^T + W_m\Lambda_m W_m^T, \quad (3)$$

where $W=[W_p \; W_m]$ and $\Lambda = \text{diag}[\Lambda_p \; \Lambda_m]$ are partitioned eigenvector matrix and eigenvalue matrix, respectively. Factor loadings are obtained by $\Phi = W_p\Lambda_p^{1/2}$ using *principal* submatrices W_p, Λ_p corresponding to the preceding M eigenvalues. Covariance matrix of specific factors Ψ is generated using minor submatrices W_m, Λ_m corresponding to the remaining $K-M$ eigenvalues. To connect the relation of FA to PCA in realization of R_z, we can formulate noisy signal in PCA form by $\mathbf{z} = W\Lambda^{1/2}\mathbf{c}$. The whitened sample $\mathbf{c}$ with $\mathbf{cc}^T = I_K$ can be obtained by $\mathbf{c} = \Lambda^{-1/2}W^T\mathbf{z}$ and partitioned by $\mathbf{c} = [\mathbf{c}_p^T \; \mathbf{c}_m^T]^T$. Then, common factors $\mathbf{f}$ and specific factors $\mathbf{r}$ can be obtained from $\mathbf{z} = W_p\Lambda_p^{1/2}\mathbf{c}_p + W_m\Lambda_m^{1/2}\mathbf{c}_m = \Phi\mathbf{f} + \mathbf{r}$. This is a PCA oriented approach for estimating FA parameters. In this FA approximation, we have $E[\mathbf{ff}^T] = E[\mathbf{c}_p\mathbf{c}_p^T] = I_M$ but non-diagonal Ψ. We use this approach to realize FA.

It is interesting that FA can be viewed as subspace approach because parameters $\Phi, \mathbf{f}, \Psi$ are derived from principal subspace $V_p = \text{span}\, W_p$ and complimentary minor subspace $V_m = \text{span}\, W_m$. Using SS approach, clean signal and noise signal individually constitute the signal subspace and the noise subspace, respectively. Differently, common factors $\mathbf{f}$ in FA come from the sources of clean signal $\mathbf{f}_y$ and noise signal $\mathbf{f}_n$, $\mathbf{f} = \mathbf{f}_y + \mathbf{f}_n$. Also, specific factors $\mathbf{r}$ can be expressed as the sum of factors associated with residual clean signal $\mathbf{r}_y$ and residual noise $\mathbf{r}_n$, $\mathbf{r} = \mathbf{r}_y + \mathbf{r}_n$. Assuming these factors are independent, covariance matrix of noisy signal is yielded by

$$\begin{aligned} R_z &= E[(\Phi\mathbf{f}_y + \mathbf{r}_y + \Phi\mathbf{f}_n + \mathbf{r}_n)(\Phi\mathbf{f}_y + \mathbf{r}_y + \Phi\mathbf{f}_n + \mathbf{r}_n)^T] \\ &= \Phi R_{fy}\Phi^T + R_{ry} + \Phi R_{fn}\Phi^T + R_{rn} = R_y + R_n \end{aligned}, \quad (4)$$

where R_{fy}, R_{fn}, R_{ry} and R_{rn} are covariance matrices corresponding to $\mathbf{f}_y$, $\mathbf{f}_n$, $\mathbf{r}_y$ and $\mathbf{r}_n$, respectively. In subspace modeling, covariance matrix plays a critical role. In this study, we are motivated to estimate complete clean signal $\mathbf{y}$ not only from principal subspace but also from minor subspace. Such FA approach provides better generalization and deeper insight into data correlation compared to SS where clean signal is recovered only from M dimensional signal subspace. SS disregards the residual signal existing in noise/minor subspace.

2.2 Signal Estimation

Similar to SS estimation of clean signal, we are deriving a linear estimator $\hat{\mathbf{y}} = H\mathbf{z}$ using $K \times K$ matrix H through minimizing the signal distortion while keeping the energy of residual noise. Using FA approach, clean signal is estimated from principal subspace and minor subspace

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}_p + \hat{\mathbf{y}}_m = H_p\mathbf{z} + H_m\mathbf{z}. \quad (5)$$

For the case of principal subspace, the estimation error is given by

$$\mathbf{e}_p = \hat{\mathbf{y}}_p - \mathbf{y}_p = (H_p - \mathbf{I}_K)\mathbf{y}_p + H_p\mathbf{n}_p = \mathbf{e}_{py} + \mathbf{e}_{pn}, \quad (6)$$

where $\mathbf{e}_{py}$ and $\mathbf{e}_{pn}$ indicate signal distortion and noise distortion, respectively. Signal estimator $\hat{H}_p$ is then derived through the constrained optimization

$$\min_{H_p} \bar{\varepsilon}_{py}^2 \quad \text{subject to}: \; \bar{\varepsilon}_{pn}^2 \le \gamma\sigma_{pn}^2, \quad (7)$$

with $\bar{\varepsilon}_{py}^2 = \text{tr}E[\mathbf{e}_{py}\mathbf{e}_{py}^T] = \text{tr}[(H_p - \mathbf{I}_K)\Phi R_{fy}\Phi^T(H_p - \mathbf{I}_K)^T]$ and $\bar{\varepsilon}_{pn}^2 = \text{tr}E[\mathbf{e}_{pn}\mathbf{e}_{pn}^T] = \text{tr}[H_p\Phi R_{fn}\Phi^T H_p^T]$ being energies of signal and noise distortion, respectively. The permissible masking level is proportional to variance of noise signal σ_{pn}^2 with $0 \le \gamma \le 1$. After Lagrange optimization, we obtain FA solution to signal estimation in principal subspace as

$$\hat{H}_p = (\Phi R_{fy}\Phi^T)(\Phi R_{fy}\Phi^T + \mu_p\Phi R_{fn}\Phi^T)^{-1}, \quad (8)$$

where μ_p is a Lagrange multiplier. SS solution in [7] is referable to obtain (8). Here, we adopt covariance matrices $\Phi R_{fy}\Phi^T$ and $\Phi R_{fn}\Phi^T$ due to signal $\mathbf{y}_p$ and noise $\mathbf{n}_p$ in principal subspace. Similarly, considering the estimation of residual signal $\mathbf{y}_m$ in minor subspace, we execute Lagrange optimization and derive residual signal estimator as

$$\hat{H}_m = R_{ry}(R_{ry} + \mu_m R_{rn})^{-1}. \quad (9)$$

The covariance matrices R_{ry}, R_{rn} of residual signal/noise $\mathbf{y}_m, \mathbf{n}_m$ are incorporated. In (8)(9), different Lagrange parameters μ_p, μ_m are used. Typically, larger eigenvalues possessing higher energies are gathered in principal subspace especially in speech frames. We can retain more signal energy in this subspace because energy of speech signal will mask that of noise signal. But, in minor subspace, the energies of speech signal and noise signal are closer or even the energy of noise signal is dominated. Tuning parameter of masking level μ_m should be larger than that μ_p for signal estimation in principal subspace.

3. SUBSPACE SELECTION

Previously, we explored model selection of HMMs for speech recognition [5]. In this paper, we concern selection of common factors for FA subspace approach. In subspace modeling, it is critical to determine the partition of principal/signal subspace and minor/noise subspace or equivalently their dimensions $\dim V_{\mathrm{p}} = M$ and $\dim V_{\mathrm{m}} = K - M$. This partition also corresponds to choose number of factors for FA. Using SS [7], subspace selection was empirically controlled by the estimated noise variance σ_{n}^2 . The smaller the variance was, the larger the signal subspace was specified for modeling noisy signal.

3.1 Selection via Testing Equivalence of Eigenvalues

To significantly perform subspace decomposition, we employ hypothesis test principle to determine dimension M . With decomposition of covariance matrix R_z , the problem of subspace selection turns out to *evaluate the equivalence of the last $K-M$ eigenvalues*. Conceptually, decision boundary between principal and minor subspaces can be determined when the last $K-M$ eigenvalues are relative small. We can test the null hypothesis that the last $K-M$ eigenvalues are equal against alternative hypothesis that at least two of them are different [1].

$H_0 : \lambda_{M+1} = \lambda_{M+2} = \cdots = \lambda_K$

H_1 : At least two of the last $K-M$ eigenvalues are different

Equivalently, we are testing isotropic [10] eigenvalues in minor subspace. Let R_z be calculated using training samples $\mathbf{Z} = \{\mathbf{z}_1 \cdots \mathbf{z}_N\}$. Eigenvalues of R_z represent the variances of decorrelated samples $\{\mathbf{d}_1 \cdots \mathbf{d}_N\}$ transformed by $\mathbf{z} = W_{\mathrm{m}} \mathbf{d}$. Assuming that eigenvalues are Gaussian distributed, we can represent the likelihood under null hypothesis as

$$L(H_0) = (2\pi)^{-\frac{N(K-M)}{2}} \left|\Lambda_{\mathrm{m}}\right|^{-\frac{N}{2}} \cdot \exp\left\{-\frac{1}{2}\sum_{n=1}^{N} \Delta\mathbf{d}_n \Lambda_{\mathrm{m}}^{-1} \Delta\mathbf{d}_n^T\right\}, \quad (10)$$

where $\Delta\mathbf{d}_i = \mathbf{d}_i - \overline{\mathbf{d}}$ and $\Lambda_{\mathrm{m}} = \mathrm{diag}[\lambda_{M+1}, \cdots, \lambda_K]$ is eigenvalue matrix associated with minor subspace. $L(H_0)$ can be further arranged as

$$(2\pi)^{-\frac{N(K-M)}{2}} \left[\left(\frac{1}{K-M}\sum_{k=M+1}^{K}\lambda_k\right)\right]^{-\frac{N}{2}} \cdot \exp\left\{-\frac{N}{2}\mathrm{tr}[\Lambda_{\mathrm{m}}\Lambda_{\mathrm{m}}^{-1}]\right\} \quad (11)$$

Also, likelihood under alternative hypothesis is derived by

$$L(H_1) = (2\pi)^{-\frac{N(K-M)}{2}} \left(\prod_{k=M+1}^{K}\lambda_k\right)^{-\frac{N}{2}} \cdot \exp\left\{-\frac{N}{2}\mathrm{tr}[\Lambda_{\mathrm{m}}\Lambda_{\mathrm{m}}^{-1}]\right\}. \quad (12)$$

Optimal solution is carried out by evaluating likelihood ratio $q = L(H_0)/L(H_1)$. The test statistic $-2\log q$ has the form

$$-N\log\prod_{k=M+1}^{K}\lambda_k + N(K-M)\log\left(\frac{1}{K-M}\sum_{k=M+1}^{K}\lambda_k\right). \quad (13)$$

This statistic can be approximated as a chi-square density χ_ν^2 with degree of freedom being $\nu = 0.5(K-M)(K-M+1) - 1$ [1]. Finally, null hypothesis H_0 is rejected at a significance level α if $-2\log q \geq \chi_{\nu;\alpha}^2$.

3.2 Selection via Testing Diagonal Covariance Matrix

However, selection via testing equivalence of eigenvalues is feasible to different subspace approaches concerning evaluation of eigenvalues. We are highlighting on developing subspace selection approach for fulfilling FA paradigm. In FA modeling, the covariance matrix of specific factors Ψ should be diagonal because the principal common factors in FA have fully captured correlation between features. The residual specific factors $\mathbf{r}$ should be decorrelated. This property is essential for modeling noisy speech in minor subspace. To realize a truly FA framework, we are motivated to *evaluate whether the covariance matrix of the estimated residual signal $\mathbf{r}_{\mathrm{y}}$ is diagonal or not so as to determine subspace dimension M*. Estimation of $\mathbf{r}_{\mathrm{y}}$ depends on the choice of M . Let $\sigma_{\mathrm{ry}}^2(i,j)$ denote the (i,j) entry of reconstructed covariance matrix R_{ry} . Null hypothesis and alternative hypothesis are stated as

$$H_0 : \sigma_{\mathrm{ry}}^2(i,j) = 0 \ \text{ for all } \ i \neq j,$$

$$H_1 : \sigma_{\mathrm{ry}}^2(i,j) \neq 0 \ \text{ for all } \ i \neq j.$$

Assuming that the residual speech samples $\{\mathbf{r}_{\mathrm{y},1} \cdots \mathbf{r}_{\mathrm{y},N}\}$ corresponding to observations $\{\mathbf{z}_1 \cdots \mathbf{z}_N\}$ are Gaussian distributed, the likelihood function is written by

$$L(H_1) = (2\pi)^{-\frac{NK}{2}} \left|R_{\mathrm{ry}}\right|^{-\frac{N}{2}} \cdot \exp\left\{-\frac{1}{2}\mathrm{tr}\left[R_{\mathrm{ry}}^{-1}\sum_{n=1}^{N}(\mathbf{r}_{\mathrm{y},n} - \overline{\mathbf{r}}_y)(\mathbf{r}_{\mathrm{y},n} - \overline{\mathbf{r}}_y)^T\right]\right\} \quad (14)$$

Under null hypothesis, likelihood function becomes

$$L(H_0) = \prod_{i=1}^{K}(2\pi)^{-\frac{NK}{2}}(\sigma_{\mathrm{ry}}^2(i,i))^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma_{\mathrm{ry}}^2(i,i)}\sum_{n=1}^{N}(r_{\mathrm{y},n,i} - \overline{r}_{\mathrm{y},n,i})^2\right\} \quad (15)$$

Therefore, we can arrange likelihood ratio of $L(H_0)$ to $L(H_1)$ as

$$\frac{\prod_{i=1}^{K}\left\{(2\pi)^{-\frac{NK}{2}}(\sigma_{\mathrm{ry}}^2(i,i))^{-\frac{N}{2}}\exp\left[-\frac{1}{2}NK\right]\right\}}{(2\pi)^{\frac{-NK}{2}}\left|R_{\mathrm{ry}}\right|^{\frac{-N}{2}}\exp\left\{-\frac{1}{2}NK\right\}} = \left[\frac{\left|R_{\mathrm{ry}}\right|}{\prod_{i=1}^{K}\sigma_{\mathrm{ry}}^2(i,i)}\right]^{\frac{N}{2}}. \quad (16)$$

In (16), sample covariance matrix R_{ry} is a Wishart density. After careful derivation, the test statistic is asymptotically χ_ν^2 with degree of freedom $\nu = 0.5K(K+1)$ [4]. In the implementation, we evaluate different M in a descending order and determine optimal M until null hypothesis is rejected. Figure 1 shows the procedure of FA modeling and selection for estimation of clean speech.

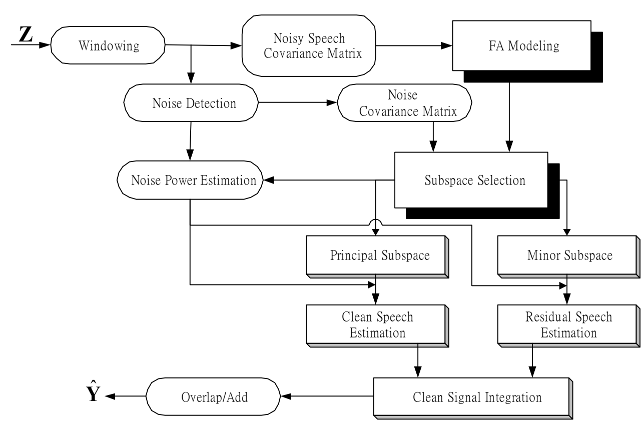

Figure 1. FA modeling and selection for clean speech estimation.

4. EXPERIMENTS

4.1 Speech Database and Experimental Setup

We evaluated SS and FA subspace approaches for noisy speech recognition using Aurora2 database. Aurora2 consisted of English digits in the presence of additive noise and linear convolutional distortion. There were three test sets in the corpus. Set A had four noise types (subway, babble, car and exhibition hall) that were similar to those in the training data, and set B contained four noise types (restaurant, street, airport and station noises) different from those in the training data. An additional convolutional channel was used in set C. All these three test sets consisted of six SNR conditions (-5dB, 0dB, 5dB, 10dB, 15dB and 20dB) and clean condition. Acoustic models in clean training and multi-condition training were investigated. There were 8,440 clean training utterances.

Speech features consisted of 13 MFCC coefficients and energy along with the delta and acceleration coefficients. We estimated continuous-density HMM parameters and built speech recognizer using HTK toolkit. We specified 16 states per word and three Gaussian mixture components per state. Subspace decomposition and selection were performed frame by frame. In signal estimation procedure, we used 40 sampling points as a frame and shifted every 20 points. Time-domain filters with 40×40 matrices were estimated. When computing covariance matrix R_z, a window of nine frames was considered. The control parameters μ_p and μ_m were tuned for different environments and SNRs. Larger μ_p is applied to produce smaller residual noise and larger signal distortion in principal subspace. On the contrary, smaller μ_m extracts residual speech with larger noise. In the experiment, we tuned two multipliers in ranges of $0 \le \mu_p \le 3$, $0 \le \mu_m \le 7$. Significance level was set as $\alpha = 0.95$ in two FA selection methods.

4.2 Experimental Results

The effectiveness of FA subspace modeling and selection is illustrated in Table 1. We report recognition rates averaged by three test sets in clean training for cases of baseline system, SS and FA approaches. FA subspace selection methods via testing eigenvalues and variances are labeled by FA I and FA II, respectively. We find that subspace denoising procedures do improve baseline speech recognition rates. FA I and FA II outperform SS in presence of different SNR conditions. The lower the SNR, the better the improvement is obtained. FA II achieved higher recognition rates than FA I. We have confirmed the statistical significance of recognition improvement of using FA compared to SS via matched-pairs test. Also, similar results are obtained when evaluating different methods in multi-condition training as shown in Table 2. Improvement is moderate. From two sets of experiments, we assure the effectiveness of proposed subspace model with selection algorithms for noisy speech recognition.

5. CONCLUSIONS

In this paper, we have presented a new FA subspace modeling of noisy speech for robust speech recognition. This model was an extension of SS which has been widely applied for speech enhancement. Using FA, signal estimators in principal subspace and minor subspace were derived for sophisticated recovery of clean signal. Importantly, we determine subspace dimension via testing equivalence of eigenvalues and decorrelation of covariance matrix. Optimal solutions were formulated for subspace selection. Experiments on Aurora2 confirmed the performance of FA modeling and selection for noisy speech recognition. In the future, we are investigating FA subspace approach in frequency domain and developing real-time estimation of μ_p and μ_m robust to different SNRs. Also, we are exploring subspace modeling for other speech applications, e.g. speaker adaptation, acoustic modeling and language modeling.

Table 1. Speech recognition rates (%) of different methods and noise conditions in case of clean training

	Baseline	SS	FA I	FA II
Clean	99.1	99.3	99.3	99.3
20 dB	97.4	97.7	97.7	97.9
15 dB	93.8	94.5	94.7	94.9
10 dB	81.7	85.0	86.0	87.2
5 dB	56.8	64.0	66.3	70.1
0 dB	30.3	38.3	41.8	44.7
-5 dB	15.2	19.8	21.3	22.7

Table 2. Speech recognition rates (%) of different methods and noise conditions in case of multi-condition training

	Baseline	SS	FA I	FA II
Clean	98.9	98.9	99.0	99.0
20 dB	98.3	98.6	98.6	98.6
15 dB	97.6	97.9	98.0	98.1
10 dB	95.6	96.3	96.4	96.5
5 dB	88.3	90.9	91.3	91.5
0 dB	63.0	75.3	75.8	76.6
-5 dB	27.5	45.1	45.7	46.5

6. REFERENCES

[1] T. W. Anderson, "Asymptotic theory for principal component analysis", *Annals of Mathematical Statistics*, vol. 34, pp.122-148, 1963.

[2] A. Basilevsky, *Statistical Factor Analysis and Related Methods - Theory and Applications*, John Wiley & Sons, 1994.

[3] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. Acoustic, Speech and Signal Processing*, vol. 27, pp. 113–120, 1979.

[4] G. E. P. Box, "A general distribution theory for a class of likelihood criteria", *Biometrika*, vol. 36, pp.317-346, 1949.

[5] J.-T. Chien and S. Furui, "Predictive hidden Markov model selection for speech recognition", *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 3, pp. 377-387, 2005.

[6] J.-T. Chien and C.-W. Ting, "Speaker identification using probabilistic PCA model selection", *ICSLP*, vol. 3, pp. 1785-1788, 2004.

[7] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement", *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 4, pp. 251-266, 1995.

[8] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, 1986.

[9] L. K. Saul and M. G. Rahim, "Maximum likelihood and minimum classification error factor analysis for automatic speech recognition", *IEEE Trans. Speech and Audio Processing*, vol. 8, no. 2, pp. 115-125, 2000.

[10] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers", *Neural Computation*, vol. 11, pp. 443-482, 1999.