Jacobian Adaptation of Noisy Speech Models

S. Sagayama, Y. Yamaguchi and S. Takahashi
NTT Human Interface Laboratories
Musashino-shi, Tokyo 180 Japan

Abstract - A Jacobian approach to fast adaptation of acoustic models is described. Acoustic models of speech under assumed noise and channel condition A are compensated by Jacobian matrices with the difference between condition A and actual condition B. Compared with existing model composition approaches for noisy speech recognition, this approach drastically reduces the computational cost while providing equivalent recognition performance. Extension of this analytic approach to acoustic model adaptation is also extensively discussed.

1 Introduction

In practical applications of speech recognition, environmental conditions often vary from time to time (e.g., mobile applications) or with each usage (e.g., telephone applications). This often causes acoustic model mismatch resulting in a serious degradation of performance. Retraining acoustic models using the actual noise environment is not feasible in most cases. One compromise is to compose mathematically a noisy speech model from acoustic models of clean speech and noise. In existing methods including Parallel Model Composition (PMC)[3] and Noise-Voice Composition (NOVO)[4], however, too much computation is required to follow in real-time the instantaneous changes in noise spectrum and level.

Another idea is to adapt initial acoustic models under the initial condition, A, to the target condition, B. This framework is motivated by speaker adaptation from the initial speaker, A (or speaker-independent), to the target speaker, B. Based on this idea, we introduced a new approach that uses a Jacobian matrix[6]. In this paper, we extend the formulation to include the channel problem. We also discuss the future directions of this approach.

2 Jacobian Adaptation of Acoustic Models

2.1 Jacobian Adaptation of Cepstrum Vectors

In the linear-scale power spectrum domain, noisy speech observed through a transmission channel is represented by $\boldsymbol{Y} = \boldsymbol{H}(\boldsymbol{S}+\boldsymbol{N})$ where n-dimensional vectors $\boldsymbol{S}$, $\boldsymbol{N}$, $\boldsymbol{H}$, and $\boldsymbol{Y}$ denote the clean speech spectrum, additive noise

spectrum, the multiplicative channel characteristics (transfer function in the power spectral domain), and the resulted composite speech spectrum, respectively. Here, vector operations are applied component by component[1]. The signal-to-noise ratio is not treated separately; the noise power is included in $\boldsymbol{N}$. In the cepstral domain, the composite speech cepstrum, $\boldsymbol{C}_Y$, is a nonlinear function of the speech cepstrum, $\boldsymbol{C}_S$, the noise cepstrum, $\boldsymbol{C}_N$, and the channel cepstrum, $\boldsymbol{C}_H$, as follows:

$$\boldsymbol{C}_Y = \boldsymbol{\Psi}(\boldsymbol{C}_S, \boldsymbol{C}_N, \boldsymbol{C}_H) \equiv \boldsymbol{F}^* \Big[\log \big\{ \exp(\boldsymbol{F}\boldsymbol{C}_S) + \exp(\boldsymbol{F}\boldsymbol{C}_N) \big\} \Big] + \boldsymbol{C}_H \tag{1}$$

since these cepstra are related to the respective spectra by

$$\log \boldsymbol{Y} = \boldsymbol{F}\boldsymbol{C}_Y,\ \log \boldsymbol{S} = \boldsymbol{F}\boldsymbol{C}_S,\ \log \boldsymbol{N} = \boldsymbol{F}\boldsymbol{C}_N,\ \log \boldsymbol{H} = \boldsymbol{F}\boldsymbol{C}_H \tag{2}$$

where $\boldsymbol{F}$ is the $(n \times n)$ Fourier transform matrix and $\boldsymbol{F}^*$ is the transposed complex conjugate of $\boldsymbol{F}$ that $\boldsymbol{F}^*\boldsymbol{F} = \mathbf{1}$.

For arbitrary speech, noise, and channel cepstra, when $\boldsymbol{C}_N$ and $\boldsymbol{C}_H$ ("Condition A") change into $\boldsymbol{C}_N + \Delta\boldsymbol{C}_N$ and $\boldsymbol{C}_H + \Delta\boldsymbol{C}_H$ ("Condition B") with the speech spectrum fixed, the composite cepstrum also changes into:

$$\boldsymbol{C}_Y + \Delta\boldsymbol{C}_Y = \boldsymbol{\Psi}(\boldsymbol{C}_S,\ \boldsymbol{C}_N + \Delta\boldsymbol{C}_N,\ \boldsymbol{C}_H + \Delta\boldsymbol{C}_H) \tag{3}$$

If these changes are small, the resulting change, $\Delta\boldsymbol{C}_Y$, is

$$\Delta\boldsymbol{C}_Y = \frac{\partial \boldsymbol{C}_Y}{\partial \boldsymbol{C}_N} \Delta\boldsymbol{C}_N + \Delta\boldsymbol{C}_H \tag{4}$$

as is well known in the basics of calculus[2]. The Jacobian matrix is given by:

$$\begin{aligned} \boldsymbol{J}_C \equiv \frac{\partial \boldsymbol{C}_Y}{\partial \boldsymbol{C}_N} &= \frac{\partial \boldsymbol{C}_Y}{\partial \log \boldsymbol{Y}} \frac{\partial \log \boldsymbol{Y}}{\partial \boldsymbol{Y}} \frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{N}} \frac{\partial \boldsymbol{N}}{\partial \log \boldsymbol{N}} \frac{\partial \log \boldsymbol{N}}{\partial \boldsymbol{C}_N} = \boldsymbol{F}^* \frac{\mathbf{1}}{\boldsymbol{H}(\boldsymbol{S}+\boldsymbol{N})} \boldsymbol{H}\boldsymbol{N}\boldsymbol{F} \\ &= \boldsymbol{F}^* \frac{\boldsymbol{N}}{\boldsymbol{S}+\boldsymbol{N}} \boldsymbol{F} \end{aligned} \tag{5}$$

2.2 Jacobian Adaptation of Time Derivatives

Consider that the cepstrum is a continuous function of time from which we observe a sample sequence for discrete time points. Because the time derivative of spectrum, $\dot{\boldsymbol{S}}$, is related to the time-derivative of cepstrum, $\dot{\boldsymbol{C}}_S$, by

$$\dot{\boldsymbol{S}} = \frac{\partial}{\partial t} \big\{ \exp(\log \boldsymbol{S}) \big\} = \exp(\log \boldsymbol{S}) \frac{\partial}{\partial t} (\log \boldsymbol{S}) = S \frac{\partial}{\partial t} (F\boldsymbol{C}_S) = \boldsymbol{S}\boldsymbol{F}\dot{\boldsymbol{C}}_S \tag{6}$$

we obtain from Eq. (5) the Jacobian matrix between the time derivative of the composite cepstrum and the noise cepstrum as follows:

$$\boldsymbol{J}_{\dot{C}} \equiv \frac{\partial \dot{\boldsymbol{C}}_Y}{\partial \boldsymbol{C}_N} = \frac{\partial}{\partial t} \frac{\partial \boldsymbol{C}_Y}{\partial \boldsymbol{C}_N} = \boldsymbol{F}^* \frac{\partial}{\partial t} \left(\frac{\boldsymbol{N}}{\boldsymbol{S}+\boldsymbol{N}} \right) \boldsymbol{F} = \boldsymbol{F}^* \left(\frac{\dot{\boldsymbol{N}}\boldsymbol{S} - \boldsymbol{N}\dot{\boldsymbol{S}}}{(\boldsymbol{S}+\boldsymbol{N})^2} \right) \boldsymbol{F} \tag{7}$$

---

[1] For vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{ab} = (a_1 b_1, \cdots, a_n b_n)^T$, $\mathbf{a}/\mathbf{b} = (a_1/b_1, \cdots, a_n/b_n)^T$, $\exp \mathbf{a} = (\exp a_1, \cdots, \exp a_n)^T$, and $\log \mathbf{a} = (\log a_1, \cdots, \log a_n)^T$. One can regard vectors as diagonal matrices for consistency with matrix arithmetic.

[2] $df(x,y) = f_x(x,y)dx + f_y(x,y)dy$ where f is an analytic function of x and y.

Note that the noise and channel spectra are usually assumed to contain no time trend (i.e., the time derivatives, $\dot{\boldsymbol{N}}$ and $\dot{\boldsymbol{H}}$, are 0) in practical applications. On the same basis, we ignore the contribution of the time derivative of noise cepstrum to the composite cepstrum.

In practical calculation, the time derivative of cepstrum is substituted by a weighted least mean squares fit of a linear model to the vector sequence (i.e., "delta-cepstrum" as in its first appearance[1]), since it converges to the time derivative as the frame rate increases and the fitting window length shrinks.

2.3 Jacobian Adaptation of Means and Variances

We have discussed the point-to-point mapping from noise and channel cepstra to composite (speech with noise and channel) cepstrum vectors and derived its Jacobian matrix.

Assuming that the variance of the distribution of $\boldsymbol{C}_Y$ is sufficiently small and stays within the effective range of linear (Jacobian) approximation, we can extend the above relationship to mean vectors of statistical distributions. In other words, if the Jacobian matrix can be regarded as a constant within the distribution range, small changes of mean values of $\boldsymbol{C}_Y$, $\boldsymbol{C}_N$, and $\boldsymbol{C}_H$ are related to each other by:

$$\Delta Mean[\boldsymbol{C}_Y] = \boldsymbol{J}_C \Delta Mean[\boldsymbol{C}_N] + \Delta Mean[\boldsymbol{C}_H] \tag{8}$$

Similarly[3], if $\boldsymbol{C}_Y$, $\boldsymbol{C}_N$, $\boldsymbol{C}_H$, $\Delta\boldsymbol{C}_N$, and $\Delta\boldsymbol{C}_H$ are statistically independent,

$$\Delta Cov[\boldsymbol{C}_Y] = \boldsymbol{J}_C \; \Delta Cov[\boldsymbol{C}_N] \; \boldsymbol{J}_C^T + \Delta Cov[\boldsymbol{C}_H] \tag{9}$$

Moreover, for the time derivative, we obtain:

$$\Delta Mean[\dot{\boldsymbol{C}}_Y] = \boldsymbol{J}_{\dot{C}} \; \Delta Mean[\boldsymbol{C}_N] \tag{10}$$

$$\Delta Cov[\dot{\boldsymbol{C}}_Y] = \boldsymbol{J}_{\dot{C}} \; \Delta Cov[\boldsymbol{C}_N] \; \boldsymbol{J}_{\dot{C}}^T \tag{11}$$

since $\Delta Mean[\dot{\boldsymbol{C}}_H]$, $\Delta Cov[\dot{\boldsymbol{C}}_H]$ are considered to be 0 here.

2.4 Jacobian Adaptation of CM-HMMs

The above relation can be applied to adaptation of continuous mixture-density hidden Markov models (CM-HMMs) to the environmental noise and the transmission channel as follows.

Training Phase:

Step 1 - Train the model under the initial condition A.

Assume an initial noise and channel condition as the reference (Condition A). Train the initial speech model (CM-HMMs) under the initial

[3] Eq. (9) is derived as follows: $Cov[C_Y+\Delta C_Y] = Cov[C_Y+J_C\,\Delta C_N+\Delta C_H] = Cov[C_Y] + J_C\,Cov[\Delta C_N]J_C^T + Cov[\Delta C_H] = Cov[C_Y] + J_C\{Cov[C_N+\Delta C_N] - Cov[C_N]\}J_C^T + Cov[C_H+\Delta C_H] - Cov[C_H]$

condition with real or simulated (e.g., noise-added) speech data, or use PMC[3] (or NOVO[4]) to compose HMMs from clean speech and noise models. Also from the initial condition, obtain initial means and covariances of noise and channel cepstra, i.e., $Mean[\boldsymbol{C}_N]$, $Cov[\boldsymbol{C}_N]$, $Mean[\boldsymbol{C}_H]$, and $Cov[\boldsymbol{C}_H]$.

Step 2 - Calculate Jacobian matrices.
For each mean vector of all the mixture components in the CM-HMMs, calculate the corresponding linear spectrum $\boldsymbol{S}$ [with Eq. (2)], its time derivative $\dot{\boldsymbol{S}}$ [Eq. (6)], and Jacobian matrices $\boldsymbol{J}_C$ for the cepstrum [Eq. (5)] and $\boldsymbol{J}_{\dot{C}}$ for the delta-cepstrum [Eq. (7)], i.e., for practical calculation,

$$[\boldsymbol{J}_C]_{ij} = \sum_{\nu} F^*_{i\nu} \frac{N_\nu}{S_\nu + N_\nu} F_{\nu j}, \qquad [\boldsymbol{J}_{\dot{C}}]_{ij} = \sum_{\nu} F^*_{i\nu} \frac{-N_\nu \dot{S}_\nu}{(S_\nu + N_\nu)^2} F_{\nu j} \tag{12}$$

Recognition Phase:

Step 3 - Observe noise and channel under the target condition B.
Obtain the noise and channel cepstra and find the differences of the mean vectors, $\Delta Mean[\boldsymbol{C}_N]$ and $\Delta Mean[\boldsymbol{C}_H]$, and those of the covariance matrices, $\Delta Cov[\boldsymbol{C}_N]$ and $\Delta Cov[\boldsymbol{C}_H]$, between the initial and target conditions (i.e., Conditions A and B).

Step 4 - Update mean vectors and variance matrices.
Update all cepstrum and delta-cepstrum mean vectors and variances of mixture components in the HMMs by Eqs.(9) and (11), namely,

$$Mean[\boldsymbol{C}_Y] \leftarrow Mean[\boldsymbol{C}_Y] + \boldsymbol{J}_C \Delta Mean[\boldsymbol{C}_N] + \Delta Mean[\boldsymbol{C}_H] \tag{13}$$

$$Mean[\dot{\boldsymbol{C}}_Y] \leftarrow Mean[\dot{\boldsymbol{C}}_Y] + \boldsymbol{J}_{\dot{C}} \Delta Mean[\boldsymbol{C}_N] \tag{14}$$

$$Cov[\boldsymbol{C}_Y] \leftarrow Cov[\boldsymbol{C}_Y] + \boldsymbol{J}_C \Delta Cov[\boldsymbol{C}_N] \boldsymbol{J}_C^T + \Delta Cov[\boldsymbol{C}_H] \tag{15}$$

$$Cov[\dot{\boldsymbol{C}}_Y] \leftarrow Cov[\dot{\boldsymbol{C}}_Y] + \boldsymbol{J}_{\dot{C}} \Delta Cov[\boldsymbol{C}_N] \boldsymbol{J}_{\dot{C}}^T \tag{16}$$

The following items should be noted concerning the above procedure.

MFCC — The entire discussion holds also for frequency-warped spectral parameters including MFCC, band-limited cepstrum, and other parameters in the cepstrum domain.

Computation — The dimension, n, of symmetric spectrum and cepstrum vectors can be almost halved, and the Fourier transform is substantially the cosine transform. To further reduce the computation, only cepstral mean vectors are adapted using Eq. (13), other parts being left unchanged; in the recognition phase, the procedure requires only one matrix-vector multiplication and vector addition for each mixture component distribution.

Variances — Negative variances may be yielded by Eqs. (15) and (16) because of inaccurate short noise observation. To preclude this, adapt the means only, or carefully estimate $\Delta Cov[\boldsymbol{C}_N]$.

Separate Noise and Channel Cepstra — Obtaining noise and channel cepstra separately is not always feasible. If the channel can be considered fixed, the above procedure provides an adaptation to noisy speech[6]. Otherwise, see the following.

Adaptation to Unknown Channel and Noise — In some cases, noise and channel cepstral changes may be unknown separately but acoustic models in condition A and speech samples in condition B may be given. In such cases, Eq. (13) (and Eq. (15)) gives a set of simultaneous equations for the distinct model states Viterbi time-aligned to the speech samples. Thus, $\Delta Mean[\boldsymbol{C}_H]$ and $\Delta Mean[\boldsymbol{C}_N]$ are obtained as the solution (in the least mean square sense, for example) and applied to all models. This provides an algorithm for supervised adaptation to channel and noise.

3 Experimental Evaluation

To demonstrate the advantage of the present algorithm, we cite the results of experimental evaluation in speaker-independent isolated-word speech recognition[6, 7]. This is a case of noise adaptation only where $\boldsymbol{C}_H$ is considered as constant.

In the training phase, one minute of traffic noise at a crossroads was chosen as the "initial noise" (noise A) and the noise means and variances were also calculated. The initial noisy speech HMnet (context-dependent HMM phone models in a network form)[2] with a SNR of 10 dB was composed using the NOVO[4] (roughly equivalent to PMC) method from a speaker-independent HMnet, trained with a large clean speech database. In the recognition phase, a noise signal collected at an exhibition hall was chosen as the "target noise" (noise B) and added to clean test speech data of 100 city names from 13 speakers. The difference between the assumed and observed noise signals was

Table 1. Speech recognition rates (%) for various noise observation durations.

	clean speech	noise	noise observation length (sec)					
algorithm	model	mismatched	0.2	0.3	0.4	0.5	1.0	2.0
NOVO	45.8	50.0	52.6	62.9	67.3	69.4	75.4	77.4
JA			71.4	71.1	71.8	73.2	74.5	75.0
SS+NOVO	54.6	76.1	77.0	78.5	79.5	80.1	81.3	82.2
SS+JA			78.2	78.3	78.9	79.6	80.9	81.6

Table 2. Computational complexities (measured on SPARCstation20).

phase	JA (cep)	NOVO (cep)	ratio J/N	JA (cep & Δcep)
training	2216 msec	4416 msec	1/2	8033 msec
recognition	149 msec	5066 msec	1/34	349 msec

(JA adapts mean vectors only; covariances are left unchanged)

also calculated in the cepstrum domain and was multiplied with the Jacobian matrices to update only Gaussian mean vectors of cepstra for all Gaussian distributions with the vector dimension, $n = 16$. Recognition rates were evaluated at a 10-dB SNR for four different durations of noise observation and averaged over 13 (9 male + 4 female) testing speakers. Combination with noise spectrum subtraction (SS) was also tested.

Table 1 compares speech recognition rates: with clean HMM and with noise-mismatched HMM (using NOVO with a 60-sec noise sample), both cases with and without spectrum subtraction. These results also show that JA (Jacobian adaptation) yields even better performance with short noise observations than does NOVO. From Table 2, it is evident that JA has an outstanding advantage in computation. The combination with noise spectrum subtraction significantly enhanced the performance of JA.

In additional experiments with JA involving variances and delta-cepstra, no significant improvement was found compared with the adaptation of cepstral mean vectors only. One possible reason is that the delta-cepstrum is not sensitive to additive noise. A noise observation duration of less than 2 seconds seems too short to accurately obtain $Cov[\boldsymbol{C}_N]$ covering the temporal fluctuations. Adaptation of covariances sometimes yielded negative variances with short noise observation.

4 Discussion and Future Directions

To improve the Jacobian Adaptation (JA) for noisy speech adaptation, possible enhancements are:

Application Scenario — To use Jacobian adaptation in real time, consider a sophisticated scenario for obtaining the noise statistics. For example, in the following typical scenario, Jacobian adaptation of acoustic models to the environmental noise can be finished well before the user's utterance by collecting the environmental noise over telephone during [pause]:

> System: "Hello, this is XXX. [pause] How may I help you?"
> User: "Well, collect call, please."

as short (e.g., 0.2-sec) noise observation has been shown experimentally to be quite effective and the JA algorithm runs very quickly.

Initial Models — One of the advantages of JA is that it makes best use of the initial models. To obtain a better initial model of noisy speech (noise A), (1) train a model using actual noisy speech data, or (2) train a model using simulated noisy speech (speech + noise) data, rather than to (3) compose the model using NOVO/PMC.

Multiple Initial Models — Prepare multiple sets of noisy speech models (noises $A_1, A_2, \cdots, A_m$) in the training phase. In the recognition phase, select the best-matching model (closest in terms of noise spectrum and SNR) and use it as the initial model in the Jacobian adaptation.

Multi-state Noise Models — Model the observed noise with a multi-state (m-state ergodic) HMM and generate all possible combinations of speech and noise. This is understood as adaptation from noise A to noises $B_1, \cdots, B_m$ with weights. This is similar to the multi-state case of model composition.

Iterative Adaptation — Iteratively use Jacobian adaptation, i.e., update the Jacobian matrix based on the actually observed noise and prepare for the next speech input if the application scenario timing allows.

Higher-order Approximation — Use the second and higher-order terms in addition to the Jacobian (i.e., first-order) term. Using the Taylor series[4], we have the m-th component of $\boldsymbol{C}_Y$ (replacing Eq. (4)):

$$\Delta c_m^Y = \sum_i \frac{\partial c_m^Y}{\partial c_i^N} \Delta c_i^N + \sum_{i,j} \frac{\partial^2 c_m^Y}{\partial c_i^N \partial c_j^N} \Delta c_i^N \Delta c_j^N + \cdots \tag{17}$$

where higher-order partial differentials are derived from Eq. (12) as:

$$\frac{\partial^2 c_m^Y}{\partial c_i^N \partial c_j^N} = \sum_\nu F_{m\nu}^* \frac{S_\nu N_\nu}{(S_\nu + N_\nu)^2} F_{\nu i} F_{\nu j}, \tag{18}$$

$$\frac{\partial^3 c_m^Y}{\partial c_i^N \partial c_j^N \partial c_k^N} = \sum_\nu F_{m\nu}^* \frac{S_\nu N_\nu (S_\nu - N_\nu)}{(S_\nu + N_\nu)^3} F_{\nu i} F_{\nu j} F_{\nu k}, \tag{19}$$

... etc. As this point-to-point relationship is no longer linear, the distribution parameters updating formulae are not simple as in Eqs.(13) through (16). In a preliminary experiment, the second-order approximation simply applied to the mean vectors gave some improvementwhile it required 11 times the computational cost for the (first-order) Jacobian adaptation. A different formulation exists[5], also based on the Taylor series, but is computationally expensive.

Speaker Adaptation — Similarly, speaker adaptation can be one of the applications if speaker differences are modeled by an analytic function (e.g., vocal-tract length normalization).

Jacobian PMC — PMC[3] composes a Gaussian distribution of a noisy speech cepstrum from separate Gaussian distributions of speech and noise cepstra. The resulting mean and covariance are functions of speech and noise cepstral means and covariances. Thus, the Jacobian approach can also be applied to this non-linear analytic function instead of Eq. (1). Once Jacobian matrices are prepared, fast adaptation is done by Eqs.(13) through (16). In comparison, PMC (or NOVO) assumes that the combination of Gaussian distributions of clean speech and noise cepstra can be represented by another Gaussian distribution. That is, the sum of two variables with log-normal pdfs is approximated by another

[4]Taylor series of function $f(x)$ of vector x is given by: $f(x+h) = \sum_{k=0}^{\infty} \frac{1}{k!} (h^T \nabla)^k f(x)$ where $\nabla = (\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_n})^T$ denote the Laplacian operator.

single log-normal pdf. Jacobian adaptation, on the other hand, assumes that the target noise is close to the expected noise and that the HMM mixture component variances are small (both within the linearity range [Eq. (4)]). Both approaches involve approximations. A combination such as Jacobian PMC is one approach that might possibly relax the limitations of these two.

5 Conclusion

We have discussed Jacobian Adaptation (JA) of acoustic models to noise and channel for speech recognition. In noisy speech recognition at a 10-dB SNR, JA gave even better experimental performance with a 0.2-sec noise observation than did NOVO (equivalent to PMC) with a 60-sec noise observation, while having only 1/34 the computational complexity.

We have discussed several possible enhancements to JA aimed at improving its performance. Extensive theoretical discussions treated simultaneous adaptation to channel and noise and combination with PMC to show that JA provides a general framework for modeling problems.

References

[1] S. Sagayama and F. Itakura, "On Individuality in a Dynamic Measure of Speech," *Proc. ASJ Spring Conf. 1979*, 3-2-7, pp. 589–590, 1979 (in Japanese).

[2] J. Takami, S. Sagayama, "A Successive State Splitting Algorithm for Efficient Allophone Modeling," *Proc. ICASSP92*, pp. 573-576, 1992.

[3] M. J. F. Gales and S. J. Young, "An Improved Approach to the Hidden Markov Model Decomposition of Speech and Noise," *Proc. ICASSP92*, pp. 233–236, 1992.

[4] F. Martin, et al., "Recognition of Noisy Speech by Using the Composition of Hidden Markov Models," *Proc. 1992 Fall ASJ Conf.,* 1-7-10, 1992.

[5] P. J. Moreno, B. Raj, and R. Stern, "A Vector Taylor Series Approach for Environment-Independent Speech Recognition," *Proc. ICASSP96*, pp. 733–736, 1996.

[6] S. Sagayama, Y. Yamaguchi, S. Takahashi, and J. Takahashi: "Jacobian Approach to Fast Acoustic Model Adaptation," *Proc. ICASSP97*, Vol. 2, pp. 835–838, 1997.

[7] Y. Yamaguchi, S. Takahashi, and S. Sagayama, "Fast Adaptation of Acoustic Models to Environmental Noise Using Jacobian Adaptation Algorithm," *Proc. Eurospeech 97*, pp. 2051–2054, 1997.