JACOBIAN APPROACH TO FAST ACOUSTIC MODEL ADAPTATION

Shigeki Sagayama, Yoshikazu Yamaguchi, Satoshi Takahashi, and *Jun-ichi Takahashi*†

NTT Human Interface Laboratories
1-1, Hikari-no-Oka, Yokosuka-shi, Kanagawa, 239 Japan

ABSTRACT

This paper describes a Jacobian approach to fast adaptation of acoustic models to noisy environments. Acoustic models under a noise assumption are compensated by Jacobian matrices with the difference between assumed and observed noise cepstra. Detailed mathematical formulation and algorithm derivation are presented. Experiments showed that when a small amount of training data is given, this approach outperforms the existing approaches (such as PMC and NOVO) for composing a model from speech and noise models. It drastically reduces computational cost by replacing the complicated computation of model composition by simple matrix arithmetic and enables real-time environmental noise adaptation. Combination with spectrum subtraction is also discussed.

1. INTRODUCTION

Acoustic model mismatch often occurs in speech recognition, even if the model has been carefully trained in a particular environment, because environmental conditions may vary from time to time (e.g., mobile applications) or with each usage (e.g., telephone applications). This often results in a serious degradation of performance.

Existing methods for the adaptation of acoustic models for clean speech to environmental conditions include PMC[2, 3] and NOVO[1]. Though they can create an adapted acoustic model from a clean-speech model and a noise model, too much computation is required to follow in real-time the instantaneous changes in noise spectrum and level. Moreover, these methods tend to require a considerable amount of training noise sample data.

Considering that speaker adaptation is done from the initial speaker, A (or speaker-independent), to the target speaker, B, it should be reasonable to consider acoustic model adaptation from noise A to noise B. Based on this idea, the question here is what can be done for compensation of the mismatch if we are given the difference between the assumed (expected) conditions, A, when training and the observed (real) conditions, B, when recognizing, when provided with only a short sample of environmental noise. If the change is relatively small, there may be a simpler and more effective method for short samples.

We introduce a new approach from a mathematical point of view that uses a Jacobian matrix for fast adaptation. Acoustic models are considered non-linear functions of the conditions such as noise spectrum, speaker characteristics, microphone, etc. A small change in the condition domain is propagated to the model domain so that the relationship between the source and destination is connected by a Jacobian matrix. Though one previous work[4] on the use of Taylor series is related to this paper in the sense that the first-order coefficient of the vector Taylor series used there is equivalent to a Jacobian matrix, the formulation and application here are different.

†Currently with NTT System Electronics Laboratories, 3-1 Wakamiya, Atsugi-shi, Kanagawa, 243-01 Japan. Authors' E-mail addresses: {saga,yamaguch,taka}@nttspch.hil.ntt.co.jp and tak@aecl.ntt.co.jp, respectively.

2. JACOBIAN ADAPTATION (JA) OF ACOUSTIC MODELS

2.1. Jacobian Matrix

It is well known in the basics of calculus that the differential of an analytic function, $f(x,y)$, is represented in terms of its partial derivatives:

$$df(x,y) = f_x(x,y)dx + f_y(x,y)dy \tag{1}$$

which means that small changes, dx in the x-domain and dy in the y-domain, cause a change, $df(x,y)$, in the function domain.

This relation holds also in the vector domain. If an n-dimensional vector, $\boldsymbol{V} = (v_1, v_2, \cdots, v_n)^T$, is an analytic function of a vector, $\boldsymbol{U} = (u_1, u_2, \cdots, u_n)^T$, a small change, $\Delta\boldsymbol{U}$, causes another small change, $\Delta\boldsymbol{V}$, and they are related to each other by:

$$\Delta\boldsymbol{V} = \frac{\partial \boldsymbol{V}}{\partial \boldsymbol{U}}\Delta\boldsymbol{U} \tag{2}$$

where $\Delta\boldsymbol{U}=(\Delta u_1, \Delta u_2, \cdots, \Delta u_n)^T$, $\Delta\boldsymbol{V}=(\Delta v_1, \Delta v_2, \cdots, \Delta v_n)^T$. The $(n \times n)$-matrix $\left[\frac{\partial \boldsymbol{V}}{\partial \boldsymbol{U}}\right]_{ij} = \frac{\partial v_j}{\partial u_i}$ is called the Jacobian matrix between $\boldsymbol{U}$ and $\boldsymbol{V}$.

2.2. Jacobian Adaptation of Cepstrum Vectors

When the environmental noise spectrum is observed in the cepstrum domain, the relationship between cepstra of speech, noise, and speech + noise is a rather complicated non-linear function $\boldsymbol{\Psi}$: where $\boldsymbol{F}$ denotes the $(n \times n)$ Fourier transform matrix which is essentially the cosine transform in real symmetric spectrum cases, namely, $F_{ij} = \cos\frac{2ij\pi}{n}$. Suppose n is sufficiently large so that a Fourier matrix well approximates the Fourier integral. Vectors with smaller dimensions can

be extended to dimension n with zero-valued components.[1]

$$\boldsymbol{C}_{S+N} = \boldsymbol{\Psi}(\boldsymbol{C}_S, \boldsymbol{C}_N) \quad (3)$$

$$= \boldsymbol{F}^* \Big[\log \{ \exp(\boldsymbol{F}\boldsymbol{C}_S) + \exp(\boldsymbol{F}\boldsymbol{C}_N) \} \Big] \quad (4)$$

The relationship between small changes in "speech + noise" cepstrum $\boldsymbol{C}_{S+N}$ and noise cepstrum $\boldsymbol{C}_N$ is as follows when the speech cepstrum $\boldsymbol{C}_S$ is fixed:

$$\Delta \boldsymbol{C}_{S+N} = \frac{\partial \boldsymbol{C}_{S+N}}{\partial \boldsymbol{C}_N} \Delta \boldsymbol{C}_N \quad (5)$$

To obtain the Jacobian matrix, denote the speech and noise spectra by $\boldsymbol{S} = (S_1, S_2, \cdots, S_n)^T$ and $\boldsymbol{N} = (N_1, N_2, \cdots, N_n)^T$ in the linear-scaled spectrum domain. Also denote the cepstrum of "speech+noise" by $\boldsymbol{C}_{S+N}$. They are related to their cepstra by

$$\log \boldsymbol{S} = \boldsymbol{F}\boldsymbol{C}_S, \ \log \boldsymbol{N} = \boldsymbol{F}\boldsymbol{C}_N, \ \log(\boldsymbol{S}+\boldsymbol{N}) = \boldsymbol{F}\boldsymbol{C}_{S+N} \quad (6)$$

From Eq. (5), the Jacobian matrix is rewritten as follows:

$$\frac{\partial \boldsymbol{C}_{S+N}}{\partial \boldsymbol{C}_N} = \frac{\partial \boldsymbol{C}_{S+N}}{\partial \log(\boldsymbol{S}+\boldsymbol{N})} \frac{\partial \log(\boldsymbol{S}+\boldsymbol{N})}{\partial (\boldsymbol{S}+\boldsymbol{N})} \frac{\partial (\boldsymbol{S}+\boldsymbol{N})}{\partial \boldsymbol{N}} \frac{\partial \boldsymbol{N}}{\partial \log \boldsymbol{N}} \frac{\partial \log \boldsymbol{N}}{\partial \boldsymbol{C}_N}$$

$$= \boldsymbol{F}^* \frac{\boldsymbol{1}}{\boldsymbol{S}+\boldsymbol{N}} \boldsymbol{1} \boldsymbol{N} \boldsymbol{F} = \boldsymbol{F}^* \frac{\boldsymbol{N}}{\boldsymbol{S}+\boldsymbol{N}} \boldsymbol{F} \quad (7)$$

where $\boldsymbol{F}^*$ is the transposed complex conjugate of the Fourier transform matrix $\boldsymbol{F}$ that $\boldsymbol{F}^*\boldsymbol{F} = \boldsymbol{1}$. This gives a practical calculation of Jacobian components:

$$\left[\frac{\partial \boldsymbol{C}_{S+N}}{\partial \boldsymbol{C}_N} \right]_{ij} = \sum_k F_{ik}^{-1} \frac{N_k}{S_k + N_k} F_{kj} \quad (8)$$

To summarize, for arbitrary speech and noise cepstra, $\boldsymbol{C}_S$ and $\boldsymbol{C}_N$, when $\boldsymbol{C}_N$ slightly changes into $\tilde{\boldsymbol{C}}_N$, the composite cepstrum, $\boldsymbol{C}_{S+N} = \boldsymbol{\Psi}(\boldsymbol{C}_S, \boldsymbol{C}_N)$, changes into $\tilde{\boldsymbol{C}}_{S+N}$ given by:

$$\tilde{\boldsymbol{C}}_{S+N} = \boldsymbol{C}_{S+N} + \frac{\partial \boldsymbol{C}_{S+N}}{\partial \boldsymbol{C}_N} (\tilde{\boldsymbol{C}}_N - \boldsymbol{C}_N) \quad (9)$$

with good approximation.

2.3. Jacobian Adaptation of Time Derivatives

Consider that the cepstrum is a continuous function of time from which we usually observe a sample sequence for discrete time points. Denote by $\dot{\boldsymbol{C}}$ the time derivative of $\boldsymbol{C}$.

Because the time derivative of spectrum $\dot{\boldsymbol{S}}$ is related to the time-derivative of cepstrum $\dot{\boldsymbol{C}}$ by

$$\dot{\boldsymbol{S}} = \frac{\partial}{\partial t} \{ \exp(\log \boldsymbol{S}) \} = \exp(\log \boldsymbol{S}) \frac{\partial}{\partial t} (\log \boldsymbol{S}) = \boldsymbol{S}\boldsymbol{F}\dot{\boldsymbol{C}}_S \quad (10)$$

we obtain the Jacobian matrix of the time derivative of the composite cepstrum from Eq.(8) by using a relation between the time derivative of the linear spectrum and the cepstrum:

$$\frac{\partial \dot{\boldsymbol{C}}_{S+N}}{\partial \boldsymbol{C}_N} = \frac{\partial}{\partial t} \frac{\partial \boldsymbol{C}_{S+N}}{\partial \boldsymbol{C}_N} = \boldsymbol{F}^* \frac{\partial}{\partial t} \left(\frac{\boldsymbol{N}}{\boldsymbol{S}+\boldsymbol{N}} \right) \boldsymbol{F}$$

$$= \boldsymbol{F}^* \left(\frac{\dot{\boldsymbol{N}}\boldsymbol{S} - \boldsymbol{N}\dot{\boldsymbol{S}}}{(\boldsymbol{S}+\boldsymbol{N})^2} \right) \boldsymbol{F} \quad (11)$$

which leads to the practical calculation:

$$\left[\frac{\partial \dot{\boldsymbol{C}}_{S+N}}{\partial \boldsymbol{C}_N} \right]_{ij} = \sum_k F_{ik}^{-1} \frac{\dot{N}_k S_k - N_k \dot{S}_k}{(S_k + N_k)^2} F_{kj} \quad (12)$$

Furthermore, the Jacobian matrix between the time derivatives of the cepstra for speech and noise, and noise is given by

$$\frac{\partial \dot{\boldsymbol{C}}_{S+N}}{\partial \dot{\boldsymbol{C}}_N} = \frac{\partial \dot{\boldsymbol{C}}_{S+N}}{\partial \boldsymbol{C}_N} \frac{\partial \boldsymbol{C}_N}{\partial \dot{\boldsymbol{C}}_N} = \frac{\partial \dot{\boldsymbol{C}}_{S+N}}{\partial \boldsymbol{C}_N} \frac{\dot{\boldsymbol{C}}_N}{\ddot{\boldsymbol{C}}_N} \quad (13)$$

where the double dots denote a second-degree derivative. In practice, we can assume that the mean of the delta cepstrum of the noise signal is 0 and we can ignore the above formula.

2.4. Adaptation of Mean Vectors

We have discussed the point-to-point correspondence between a noise cepstrum vector and a composite (speech + noise) cepstrum vector and derived the Jacobian matrix between them. This relation, however, can not be simply applied to mean vectors of statistical distributions because of non-linearity.

We have to assume that the variance of the distribution of $\boldsymbol{C}_{S+N}$ is sufficiently small and stays within the effective range of linear (Jacobian) approximation. Then 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, when the mean vector $Mean[\boldsymbol{C}_N]$ of a statistical distribution changes by a small amount $\Delta Mean\ [\boldsymbol{C}_N]$, it causes another small change, $\Delta Mean[\boldsymbol{C}_N]$, in $Mean[\boldsymbol{C}_{S+N}]$. They are related to each other by

$$\Delta Mean[\boldsymbol{C}_{S+N}] = \frac{\partial \boldsymbol{C}_{S+N}}{\partial \boldsymbol{C}_N} \Delta Mean[\boldsymbol{C}_N] \quad (14)$$

2.5. Adaptation of Variance Matrices

To discuss the adaptation of a variance matrix of a statistical distribution, assume $\boldsymbol{C}_{S+N}$, $\boldsymbol{C}_N$, and $\Delta \boldsymbol{C}_N$ are statistically independent from each other. When $\boldsymbol{C}_{S+N}$ slightly changes by $\Delta \boldsymbol{C}_N$, the resulting change $\Delta Cov[\boldsymbol{C}_{S+N}]$ of the covariance matrix of $\boldsymbol{C}_{S+N}$ is well approximated by[2]

$$\begin{aligned} & Cov[\ \boldsymbol{C}_{S+N} + \Delta \boldsymbol{C}_{S+N}\ ] \\ &= Cov[\ \boldsymbol{C}_{S+N} + \boldsymbol{J}_C\ \Delta \boldsymbol{C}_N\ ] \\ &= Cov[\boldsymbol{C}_{S+N}] + \boldsymbol{J}_C\ Cov[\Delta \boldsymbol{C}_N]\ \boldsymbol{J}_C^T \\ &= Cov[\boldsymbol{C}_{S+N}] + \boldsymbol{J}_C\ \{ Cov[\boldsymbol{C}_N + \Delta \boldsymbol{C}_N] - Cov[\boldsymbol{C}_N]\ \}\ \boldsymbol{J}_C^T \end{aligned}$$

[1] Hereinafter, vector operations are defined as follows for vectors $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$\begin{aligned} \boldsymbol{a}\boldsymbol{b} &= (a_1 b_1, a_2 b_2, \cdots, a_n b_n)^T \\ \boldsymbol{a}/\boldsymbol{b} &= (a_1/b_1, a_2/b_2, \cdots, a_n/b_n)^T \\ \log \boldsymbol{a} &= (\log a_1, \log a_2, \cdots, \log a_n)^T \end{aligned}$$

Note that matrix arithmetic is different. It is possible to regard vectors as diagonal matrices for consistency with matrix arithmetic.

[2] Defined here is $Cov[\boldsymbol{x}] \equiv \mathcal{E}[(\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}])^T] = \mathcal{E}[\boldsymbol{x}\boldsymbol{x}^T] - \mathcal{E}[\boldsymbol{x}]\mathcal{E}[\boldsymbol{x}]^T$, where $\boldsymbol{x}$ represents a stochastic variable vector.

where J_C denotes the Jacobian matrix between C_N and C_{S+N}. In a more compact form:

$$\Delta Cov[C_{S+N}] = J_C \; \Delta Cov[C_N] \; J_C^T \qquad (15)$$

Similarly, we have:

$$\Delta Cov[\dot{C}_{S+N}] = J_{\dot{C}} \; \Delta Cov[\dot{C}_N] \; J_{\dot{C}}^T \qquad (16)$$

We now have Jacobian adaptation formulae for means and variances of statistical distributions of the cepstrum, and time derivatives of the cepstrum.

2.6. Jacobian Adaptation of CMHMMs

To adapt continuous mixture hidden Markov models (CMHMMs) to the environmental noise, we need further approximations. First, the time derivative of cepstrum C is approximated by a weighted least mean square fit of a linear model to the vector sequence, i.e., the so-called "delta-cepstrum" as formulated in the first appearance of delta-cepstrum[6].

The outline of the procedure for HMM adaptation to the change in noise is as follows.

Training Phase:

Step 1 - Assume the reference noise.[3]
Assume a particular noise condition as the reference. From the noise cepstrum, C_N, obtain its mean vector, $Mean[C_N]$, and the variance matrix, $Cov[C_N]$. Also obtain the spectrum N corresponding to the cepstral mean vector by using Eq. (6).

Step 2 - Train the model.
Train HMMs with noisy speech data or simulated noisy speech data. More practically, "speech + noise" HMMs can be composed from clean speech and noise models using PMC[2, 3] or NOVO[1].

Step 3 - Calculate Jacobian matrices.
For each mean vector of all the mixture components in the CMHMMs, calculate the corresponding linear spectrum S [with Eq. (6)], its time derivative $\dot{S}$ [Eq. (10)], and Jacobian matrices J_C for the cepstrum [Eq. (7) or (8)] and $J_{\dot{C}}$ for the delta-cepstrum [Eq. (11)].

Recognition Phase:

Step 4 - Observe the noise.
Observe the environmental noise cepstrum and find the differences of the mean vectors, $\Delta Mean[C_N]$, and those of the covariance matrices, $\Delta Cov[C_N]$, between the noise assumed in the training phase and that actually observed.

Step 5 - Update mean vectors.
Update all cepstrum and delta-cepstrum mean vectors of mixture components in the HMMs by Jacobian adaptation of means.

$$Mean(C_{S+N}) \leftarrow Mean(C_{S+N}) + J_C \Delta Mean(C_N)$$
$$Mean(\dot{C}_{S+N}) \leftarrow Mean(\dot{C}_{S+N}) + J_{\dot{C}} \Delta Mean(\dot{C}_N)$$

[3]The time derivative of cepstrum $\dot{N}$ [Eq. (10)] is assumed to be 0 as the noise cepstrum should not contain a time trend. The mean vector, $Mean[\dot{C}_N]$, and the variance matrix, $Cov[\dot{C}_N]$, of the noise time derivative $\dot{C}_N$ are also assumed to be 0.

Step 6 - Update variance matrices.
Update all cepstrum and delta-cepstrum covariance matrices in the HMMs by Jacobian adaptation of covariances.

$$Cov(C_{S+N}) \leftarrow Cov(C_{S+N}) + J_C \Delta Cov(C_N) J_C^T$$
$$Cov(\dot{C}_{S+N}) \leftarrow Cov(\dot{C}_{S+N}) + J_{\dot{C}} \Delta Cov(\dot{C}_N) J_{\dot{C}}^T$$

In the above procedure, the critical part is considered to be the adaptation of the cepstral mean vectors. Other parts can be removed to make the whole procedure simpler; in the recognition phase, the procedure requires only one matrix-vector multiplication and vector addition for each mixture component distribution.

2.7. Theoretical Limit of JA Approximation

To summarize the approximations included in the above procedure, it is assumed that

- The cepstral difference between the assumed and observed noises is within the linearity range. [Eq. (5)]
- The noise cepstrum contains no trend along time. (Thus, the mean vector of the time derivative of the noise cepstrum is zero.)
- Fourier transform is well approximated with a finite number, n, of frequency points.
- The distribution ranges (variances) of mixture components are small enough to stay within the linearity range.
- C_{S+N}, C_N, and ΔC_N are statistically independent.
- When covariance matrices are chosen to be diagonal, the resulted off-diagonal components from Eqs. (15) and (16) can be ignored. In this case, this causes another approximation while the required computational time is considerably less than for the full-covariance case.

These theoretical limits are relaxed where the linear compensation of a non-linear function is even better than no compensation, and the effective range of the present procedure may not be limited within the linearity range. Note also that the above formulation is not limited within Gaussian mixtures.

3. ADAPTATION IN SPECTRAL DOMAIN

In this paper, noise is observed in the cepstrum and cepstral time-derivative domains. Alternatively, it can be observed in the linear spectral domain from which we can directly derive the Jacobian matrix:

$$\frac{\partial C_{S+N}}{\partial N} = \frac{\partial C_{S+N}}{\partial(\log(S+N))} \frac{\partial(\log(S+N))}{\partial(S+N)} \frac{\partial(S+N)}{\partial N}$$
$$= F \cdot \frac{1}{S+N} \cdot 1 = \frac{F}{S+N} \qquad (17)$$

This leads to another fast noise adaptation procedure which is not treated here.

4. EXPERIMENTAL EVALUATION

4.1. Experimental Setup

The present algorithm was tested in a speaker-independent isolated-word speech recognition experiment as follows:

Table 1. Speech recognition rates (%) for various noise observation durations (JA - cepstral means only).

algorithm	clean speech model (no adaptation)	noise-mismatched initial model	noise observation length (sec)					
			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

* These results were averaged over 13 testing speakers (9 male + 4 female) at 10dB SNR.

Training – Step 1: In the training phase, one minute of traffic noise at a crossroads was chosen as the "assumed noise." Step 2: A noise-adapted HMnet (context-dependent HMM phone models in a network form)[5] with an SNR of 10 dB was composed using the NOVO[1] method from a speaker-independent HMnet, trained with a large clean speech database, and the noise means and variances. Step 3: The Jacobian matrices for all mean vectors and covariances were calculated.

Recognition – Step 4: In the recognition phase, a noise signal collected at an exhibition hall was chosen as the "observed noise" and added to clean test speech data of 100 city names from 13 speakers. The difference between the assumed and observed noise signals was also calculated in the cepstrum domain. Step 5 & 6: The Jacobian matrices were calculated and Gaussian mean vectors and covariance matrices were updated for all Gaussian distributions.

Evaluation – Recognition rates were evaluated for 4 different durations of noise observation and averaged over 13 testing speakers. Combination with noise spectrum subtraction (SS) was also tested.

4.2. Experimental Results and Discussion

Table 1 shows a comparison of 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. It also shows that JA (Jacobian Adaptation) yields even better performance with short noise observations than does NOVO (which is roughly equivalent to PMC). Table 2 shows that JA has an outstanding advantage in computation. These are results from JA of cepstral means only. The combination with noise spectrum subtraction significantly enhanced the performance of JA.

One possible application of this outstanding advantage is real-time environment adaptation where short periods of environmental noise are observed (e.g., between the guidance sentences) before a speaker's utterances.

In preliminary experiments, no performance improvement was seen with the matrix dimension, n, above the cepstrum dimension.

In additional experiments of JA of 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 $Cov[C_N]$ covering the temporal fluctuations. As might be expected, in adaptation of covariances, negative variances are sometimes yielded from Eq. (15) when the observed noise duration is short and the resulting difference of noise variances is a large negative value.

Table 2. Computational complexities (JA - cepstral mean vectors only; measured on SPARCstation20).

phase	JA	NOVO	ratio J/N
training	2216 msec	4416 msec	1/2
recognition	149 msec	5066 msec	1/34

5. CONCLUSION

Jacobian Adaptation (JA) of acoustic models for speech recognition has been presented in this paper. In noisy speech recognition at 10 dB SNR, JA experimentally gave even better performance with a 0.2-second noise observation than did NOVO (or, equivalently, PMC) with a 60-second noise observation, while having only 1/34 the computational complexity. The JA framework has a wide applicability to acoustic model adaptation.

6. REFERENCES

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

[2] 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.

[3] M. J. F. Gales and S. J. Young: "A Fast and Flexible Implementation of Parallel Model Combination," Proc. ICASSP95 (Detroit), pp. 133–136, 1995.

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

[5] J. Takami, S. Sagayama, "A Successive State Splitting Algorithm for Efficient Allophone Modeling," Proc. ICASSP92 (San Francisco), 66.6, 1992.

[6] Shigei Sagayama and Fumitada Itakura: "On Individuality in a Dynamic Measure of Speech," Proc. ASJ Spring Conf. 1979, 3-2-7, pp. 589–590, June 1979.