DIFFERENTIAL APPROACH TO ACOUSTIC MODEL ADAPTATION

Shigeki Sagayama

Japan Adavanced Institute of Science and Technology
Tatsu-no-kuchi, Ishikawa, 923-1292, Japan
E-mail: sagayama@jaist.ac.jp

ABSTRACT

This paper discusses a 'differential approach' to acoustic model adaptation to different channel, noise and speaker conditions for speech recognition with a concept of vector field adaptation. Adaptation of acoustic model, such as HMM output probability densities, is modeled as a vector field in the acoustic feature vector space which effects on the models to move along the local vector directions. Specifically, if the feature vector is based on cepstrum, multiplicative channel characteristic is interpreted as a uniform vector field where the channel cepstrum vector is added to all feature distributions. Additive noise forms a vector field where 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. Experimentally compared with existing model composition approaches for noisy speech recognition, this approach drastically reduced the computational cost while providing equivalent recognition performance. Difference between speakers is also formulated as a vector field where only assumption is that the vector field is smooth from a physiological view. Vector Field Smoothing (VFS) technique for speaker adaptation is then discussed in this context.

1 INTRODUCTION

One of most important and practical issues in speech recognition is the mismatch between acoustic model and speech input. 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 and channel environment is not feasible in most cases. Speaker is another factor that causes acoustic mismatch.

Acoustic model adaptation is one of effective way to cope with this problem along with other possible approaches in all stages of acoustic feature manipulation such as input speech acquisition effort with less noise and channel distortion, signal processing for noise removal, and noise-robust acoustic features.

Suppose reasonably good acoustic models under a particular noise, channel and speaker condition (*condition "A"*) are present. If a relatively small amount of data from a new acoustic condition (*condition "B"*), i.e., new noise, channel and speaker combination, is given, the models can be improved through acoustic model adaptation. In this paper, we discuss such acoustic model adaptation based on the differential characteristics of the acoustic feature space.

An early introduction of differential approach to speech recognition dates back to 1979 with the so-called delta-cepstrum which was proposed for describing speaker characteristics and later proved effective for speech recognition. This new parameter was essentially a time-differentiation of cepstrum used as the acoustic feature vector. From the same point of view of differential characteristics of speech, acoustic differences between different speakers (speakers A and B) were modeled by the space-differentials in the feature space. In this work, only assumption in speaker adaptation was that the differential space should be smooth from a physiological insight. Vector Field Smoothing (VFS) for speaker adaptation was thus proposed as a simple and quite effective algorithm for speaker adaptation. The same basic idea was further applied to a noise-adaptive speech recognition. The same idea led to an approach of adapting models under noise condition A to condition B citeSagayama97ICASSP using a differential (Jacobian) characteristics of a cepstrum space.

The common basic idea here is vector field in the acoustic feature space. Given a change of environmental conditions including noise, channel and speaker, acoustic models are updated or modified according to the differential characteristics of the nature of the feature space. This idea is loosely related to the main concept of differential geometry.

The purpose of this paper is to discuss the idea of adaptation vector field as a general paradigm of treating the problem of acoustic model mismatch which is quite different from popular Bayesian approaches.

2 CHANNEL VECTOR FIELD

2.1 Multiplicative Channel

If the acoustic feature parameters used in speech recognition are cepstrum coefficients (or cepstrum-based features) such as MFCCs, the multiplicative effect of linear characteristics of input channel is simply added

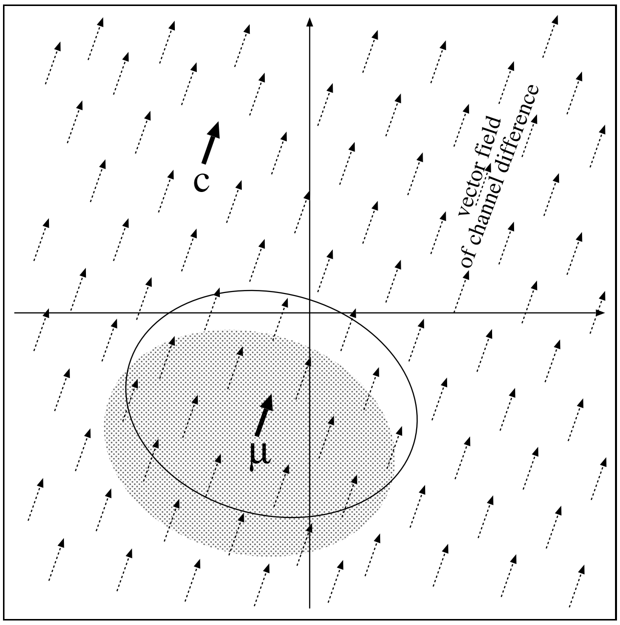

Figure 1: Uniform vector field of multiplicative channel in the cepstral domain (conceptual illustration). – The Jacobian matrix is only a unit matrix here in the entire feature space.

to speech in the cepstrum domain. Namely,

$$\boldsymbol{C}_Y = \boldsymbol{C}_S + \boldsymbol{C}_H \tag{1}$$

where $\boldsymbol{C}_S$, $\boldsymbol{C}_H$, and $\boldsymbol{C}_Y$ denotes cepstrum vectors corresponding to the speech power spectrum, $\boldsymbol{S}$, the channel transfer function, $\boldsymbol{H}$ (in the power spectrum dimension), and the speech spectrum affected by the channel, $\boldsymbol{Y}$, namely

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

where the operator $\boldsymbol{F}$ represents the Fourier transform.[1]

Suppose a speech cepstrum vector is given with an assumed channel characteristics A. The 'channel' may be a telephone channel, a particular microphone, or any other linear characteristics and their composites that affect the speech spectrum passing though them. If a real application condition has slightly different channel characteristics, B, the difference between channels A and B in the cepstrum domain is represented by a difference vector

$$\Delta \boldsymbol{C}_H = \boldsymbol{C}_H^B - \boldsymbol{C}_H^A \tag{3}$$

which can be added to the cepstrum vector under the assumed condition A. Acoustic models thus can be adapted to the new condition B by adding $\Delta \boldsymbol{C}_H$ to their mean vectors if $\Delta \boldsymbol{C}_H$ is known.

[1]Hereinafter, vector operations for vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are defined as follows: $\boldsymbol{ab} = (a_1b_1, \cdots, a_nb_n)^T$, $\boldsymbol{a}/\boldsymbol{b} = (a_1/b_1, \cdots, a_n/b_n)^T$, $\exp \boldsymbol{a} = (\exp a_1, \cdots, \exp a_n)^T$, $\log \boldsymbol{a} = (\log a_1, \cdots, \log a_n)^T$.

2.2 Vector Field Interpretation

The adaptation vector $\Delta \boldsymbol{C}_H$ forms a uniform vector field. Every cepstrum vector in the cepstrum space receives a force to move by $\Delta \boldsymbol{C}_H$ to fit the new condition from the field. Figure 1 conceptually illustrates the uniform vector field caused by a change in the channel characteristics. This simple interpretation leads to a class of algorithms of acoustic model adaptation to different noise and speaker with non-uniform vector fields.

One of simple applications of the channel vector field is cepstral mean normalization where the channel vector field is estimated from the speech input and subtract it from the input cepstrum sequence.

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:

$$\Delta \boldsymbol{V} = \frac{\partial \boldsymbol{V}}{\partial \boldsymbol{U}} \Delta \boldsymbol{U} \tag{4}$$

where $\Delta\boldsymbol{U}$=$(\Delta u_1, \Delta u_2, \cdots, \Delta u_n)^T$, $\Delta\boldsymbol{V}$=$(\Delta v_1,\ \Delta v_2,\ \cdots,\ \Delta v_n)^T$.

$$\frac{\partial \boldsymbol{V}}{\partial \boldsymbol{U}} = \begin{array}{c} i \\ \left(\begin{array}{ccc} & \vdots & \\ \cdots & \dfrac{\partial v_j}{\partial u_i} & \cdots \\ & \vdots & \end{array} \right) j \end{array} \tag{5}$$

where the $(n \times n)$-matrix $\left[\dfrac{\partial \boldsymbol{V}}{\partial \boldsymbol{U}}\right]_{ij} = \dfrac{\partial v_j}{\partial u_i}$ is called the Jacobian matrix between $\boldsymbol{U}$ and $\boldsymbol{V}$.

Extending the consideration stated in the above subsection, a (small) change in the channel cepstrum $\boldsymbol{C}_H$ causes the same amount of change in the resulted composite cepstrum $\boldsymbol{C}_Y$ because the Jacobian relationship between the cause and the result is a unit matrix:

$$\boldsymbol{J} = \frac{\partial \boldsymbol{C}_Y}{\partial \boldsymbol{C}_H} = \boldsymbol{I} \tag{6}$$

Namely, the feature space is associated with a unit matrix everywhere as for the channel changes, i.e., the entire feature space is characterized by a unit Jacobian matrix concerning channel issues.

3 NOISE VECTOR FIELD

3.1 Additive Noise

In the case of additive noise, the vector field is no more uniform but may be dependent on the location in the feature vector space. In the linear spectrum domain, the composite spectrum is simply the sum of speech and noise spectra, $\boldsymbol{S}$ and $\boldsymbol{N}$, i.e., $\boldsymbol{S} + \boldsymbol{N}$ where $\boldsymbol{N}$ denotes the spectrum of the additive noise signal. The composite cepstrum, $\boldsymbol{C}_Y$, of speech and

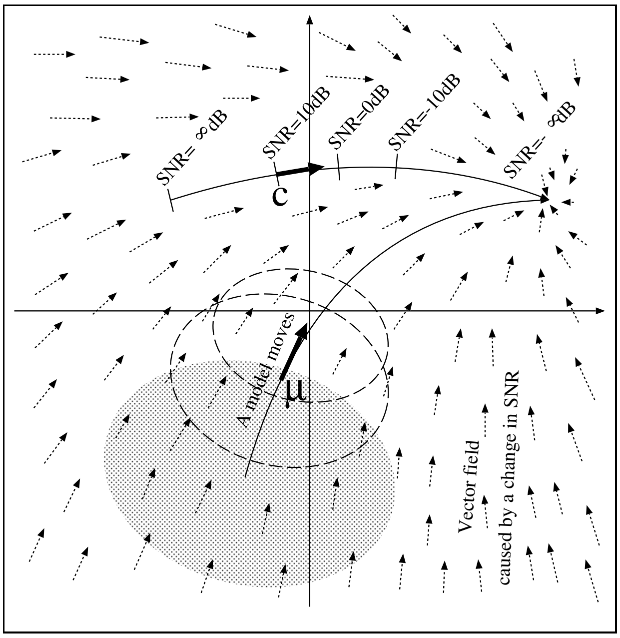

Figure 2: Vector field caused by a change in the signal-to-noise ratio of the additive noise (cconceptual illustration). – All points in the space is pushed toward the noise cepstrum.

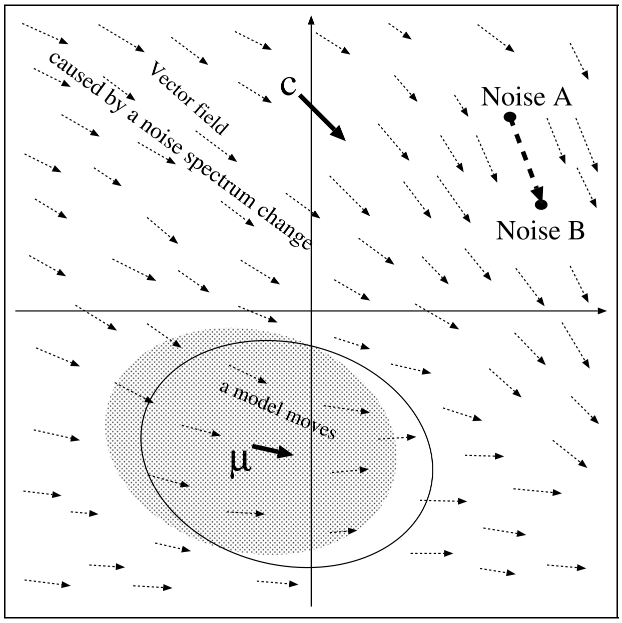

Figure 3: Vector field caused by a change in the additive noise spectrum (conceptual illusstration). – All points in the space is forced to move accordingly.

noise is represented by a non-linear function of speech cepstrum $\boldsymbol{C}_S$ and noise cepstrum $\boldsymbol{C}_N$ as follows:

$$\boldsymbol{C}_Y = \boldsymbol{F}^* \Big[\log \big\{ \exp(\boldsymbol{F}\boldsymbol{C}_S) + \exp(\boldsymbol{F}\boldsymbol{C}_N) \big\} \Big] \quad (7)$$

Note that the cepstrum vector must include the 0-th coefficient so as to take the spectrum energy into account in adding speech and noise spectra, i.e., the signal-to-noise ratio is not treated separately; the noise power is included in $\boldsymbol{N}$. With $\boldsymbol{C}_S$ fixed, the vector field caused by a small change in $\boldsymbol{C}_N$ is given by

$$\Delta \boldsymbol{C}_Y = \frac{\partial \boldsymbol{C}_Y}{\partial \boldsymbol{C}_N} \Delta \boldsymbol{C}_N \quad (8)$$

where the Jacobian matrix is dependent on $\boldsymbol{C}_S$ and $\boldsymbol{C}_N$, as is well known in the basics of calculus.

Since these cepstra are related to the respective spectra by

$$\log \boldsymbol{Y} = \boldsymbol{F}\boldsymbol{C}_Y,\ \log \boldsymbol{S} = \boldsymbol{F}\boldsymbol{C}_S,\ \log \boldsymbol{N} = \boldsymbol{F}\boldsymbol{C}_N \quad (9)$$

The Jacobian matrix $\frac{\partial \boldsymbol{C}_Y}{\partial \boldsymbol{C}_N}$ is given by:

$$\begin{aligned} \boldsymbol{J}_N &\equiv \frac{\partial \boldsymbol{C}_Y}{\partial \boldsymbol{C}_N} = \frac{\partial \boldsymbol{C}_Y}{\partial \log \boldsymbol{Y}} \frac{\partial \log \boldsymbol{Y}}{\partial \boldsymbol{Y}} \frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{N}} \frac{\partial \boldsymbol{N}}{\partial \log \boldsymbol{N}} \frac{\partial \log \boldsymbol{N}}{\partial \boldsymbol{C}_N} \\ &= \boldsymbol{F}^* \frac{\boldsymbol{N}}{\boldsymbol{S} + \boldsymbol{N}} \boldsymbol{F} \end{aligned} \quad (10)$$

Note also that this differential property is completely inherent from the nature of the cepstrum vector space. (The same property holds for MFCC). Once the noise cepstrum vector is given, the vector field is then automatically determined in the entire space. This property of the space is meaningful in acoustic model adaptation to noise. Models in the feature space are moved according to the local vector field values. Thus, the entire feature space is characterized by a Jacobian matrix of the form (Eq. 10) concerning additive noise issues.

Figure 2 illustrates the vector field concept for a small change in the signal-to-noise ratio. Along a trajectory from clean speech (SNR= $-\infty$dB) to noise only (SNR= ∞dB), a noisy environment condition (A) has a vector field tangent to the trajectory. Every point in the vector space is pushed toward the noise cepstrum as the SNR decreases.

The case of slight noise changes from condition A to condition B is conceptually illustrated in Figure 3.

3.2 Jacobian Adaptation

We have discussed the point-to-point mapping from the noise cepstrum to the composite (speech with noise) cepstrum vector 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 $\boldsymbol{C}_Y$ and $\boldsymbol{C}_N$ are related to each other by:

$$\Delta Mean[\boldsymbol{C}_Y] = \boldsymbol{J}_N \Delta Mean[\boldsymbol{C}_N] \quad (11)$$

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

$$\Delta Cov[\boldsymbol{C}_Y] = \boldsymbol{J}_N \ \Delta Cov[\boldsymbol{C}_N] \ \boldsymbol{J}_N^T \tag{12}$$

As for time-derivatives (delta-cepstrum) of the features, see [3].

Based on this idea, a new approach that uses a Jacobian matrix was introduced[3]. 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 condition with real or simulated (e.g., noise-added) speech data, or use PMC[2] (or NOVO[?]) 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., $Mean[\boldsymbol{C}_N]$ and $Cov[\boldsymbol{C}_N]$.

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. (9)] Jacobian matrices $\boldsymbol{J}_N$ for the cepstrum [Eq. (10)], i.e., for practical calculation,

$$[\boldsymbol{J}_N]_{ij} = \sum_{\nu} F^*_{i\nu} \frac{N_\nu}{S_\nu + N_\nu} F_{\nu j}, \tag{13}$$

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 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.(11) and (12), namely,

$$Mean[\boldsymbol{C}_Y] \leftarrow Mean[\boldsymbol{C}_Y] + \boldsymbol{J}_N \Delta Mean[\boldsymbol{C}_N] \tag{14}$$
$$Cov[\boldsymbol{C}_Y] \leftarrow Cov[\boldsymbol{C}_Y] + \boldsymbol{J}_N \Delta Cov[\boldsymbol{C}_N] \boldsymbol{J}_N^T \tag{15}$$

$Cov[\boldsymbol{C}_Y]$ should be taken care not to fall negative due to statistical errors included in the Jacobian matrices.

According to an experimental evaluation in speaker-independent isolated-word speech recognition[3] with noise adaptation, it was shown that Jacobian adaptation to additive noise performed equivalently (or even slightly better) to NOVO method (considered equivalent to PMC) only using a noise data of 0.2-sec duration. The speech recognition rates of 45.8% with a clean speech model and 50.0% with a noise mismatched model raised to 71.4% by JA with a 0.2-sec noise sample and to 75% with 2.0-sec noise. In comparison of the computational complexity in the recognition phase, JA required only 1/34 the computation required in NOVO (or PMC).

In the training phase of the experiment, 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) with a SNR of 10 dB was composed using the NOVO 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 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.

3.3 Related Issues

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

Initial Models — Train the initial model using a data set of realistic noisy speech. In such cases, JA is advantageous compared with model composing methods such as PMC. JA makes the best use of the initial models if it is close to the target environment. The relatively small noise mismatch is efficiently compensated by JA.

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.

Channel and Noise Joint Adaptation — Apply least mean square error estimation to decompose the observed vector field into channel and noise components. According to their Jacobian properties, the joint compansation vector, ΔC_Y, can be obtained at any vector point in the feature space.

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.

4 SPEAKER VECTOR FIELD

4.1 Underlying Smooth Vector Field

The basic idea of vector field is also applicable to speaker changes (from *speaker A* to *speaker B*). The corresponding feature vectors, models, and distribution centroids move from those of *speaker A* to those of *speaker B*. Such a move is considered as a result of an underlying vector field of speaker differences.

As opposed to the multiplicative channel and additive noise cases, the speaker change can not be analytically formulated using a Jacobian matrix. One possible assumption on this formulation is that the vector field can be considered as smooth, i.e., acoustically close points may move similarly.

From the articulatory point of view, a smooth motion of the articulatory organs produces a smooth cepstral vector trajectory which corresponds to a smooth sequence of HMM state outputs. It is thus likely that the underlying vector field is smooth which means that states acoustically close to each other in the initial model are mapped to states acoustically close to each other in the target speaker model.

4.2 Vector Field Smoothing

A speaker adaptation method named "Vector Field Smoothing (VFS)" was proposed based on the above idea in 1991 (first in Japanese) [4, 5]. A typical VFS algorithm for speaker adaptation of continuous mixture density subword HMMs consists of the following three steps. It is assumed that fairly good initial models, either speaker-dependent or - independent, and a limited amount of phonetically transcribed training data are given.

Step 1 - Obtain samples of the vector field.

Viterbi alignment of the training data with the model gives a difference vector which is regarded as a sample of the underlying vector field of speaker difference, ΔC_S, at the model location in the feature space. This can be also obtained through Baum-Welch retraining (with variances unchanged) of the model using the data to find the difference vector, ΔC_S, between the model centroids before and after retraining. They are considered as "probed samples" of the underlying vector field with observation errors (i.e., statistical fluctuations).

This is often done for a limited amount of training data, e.g., 10 words. For some of HMM state output probability densities (or mixture components), the vector field is left 'unprobed'.

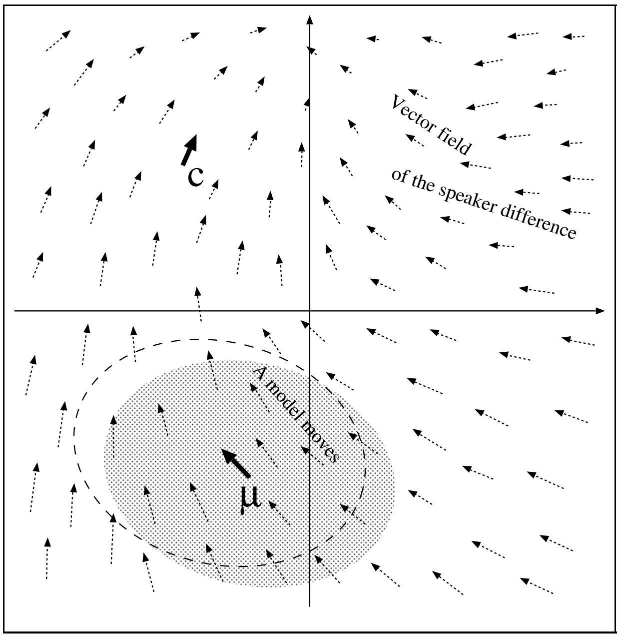

Figure 4: A vector field representing the difference between two speakers (conceptual illustration). – The mismatch between the model and the speaker forms a vector field which can be assumed to be smooth from an articulatory view.

Step 2 - Smooth the sampled vectors to obtain a smooth vector field.

The underlying smooth vector field is retrieved. Namely, the sampled vectors are smoothed to reduce the statistical observation errors and interpolated to retrieve underlying vector field values at model centroids. This is usually done by spatial filtering using Gaussian moving average or fuzzy membership functions applied to k-nearest neighbors. The smoothing parameter (the variance of Gaussian smoothing or the fuzzy membership function) controls the degree of smoothing. Stronger smoothing corresponds to a smoother vector field and is effective in the case of a small amount of adaptation data where fewer samples of the vector field with larger estimation errors are expected.

Step 3 - Apply the adaptation vector field to

adapt the models.

The retrieved smooth vector field $\Delta \boldsymbol{C}_S$ is applied to the initial model to obtain an adapted model, $\boldsymbol{C}_S + \Delta \boldsymbol{C}_S$.

In a simple implementation, the interpolation and smoothing in Step 2 are often done separately, interpolating the missing data followed by overall smoothing. One further enhancement of VFS is to adjust the Gaussian variances according to the vector field spreading/shrinkage.

It has been experimentally found that the mismatch between a speaker-dependent phone models and the target speaker is significantly improved. In an early study in 1991, when phrase-by-phrase utterances were given from 11 sentences for adaptation, the recognition rate of Japanese 23 phonemes was improved from 47.3% to 87.4% and the phrase recognition rate by HMM-LR continuous speech recognition was improved from 21.5% to 85.1% [4].

VFS has a wide applicability to various acoustic modeling adaptation problems. For instance, the VQ codebook of a specific speaker (reference) was adapted to a new speaker (target) by this approach. If pairs of utterances by both speakers are given, they are time-aligned to each other by DTW (Dynamic Time Warping) and the differences between the time-aligned frame vectors are considered as samples from the underlying speaker adaptation vector field. They are smoothed and interpolated by VFS to estimate the VQ codewords for the target speaker. This is directly applicable to VQ-based discrete HMMs [5]. Since VFS is considered as a generalization of microphone/line characteristics normalization, VFS is applicable to the mixed problem of new speaker, new microphone, and unknown line, as well as to noisy environments.

VFS is also applicable to MAP (maximum a priori probability) estimated models. Since MAP estimation is done separately for each subword category, it is considered as "intra-class training" and is not free from the missing data problem. On the contrary, VFS is "inter-class smoothing". Their combination, "MAP/VFS", yields better adaptation speed and accuracy. An error reduction rate of 21.6% (averaged over 4 speakers) was experimentally obtained when 10-word utterances were incrementally given.

5 DISCUSSION

As for a simultaneous adaptation problem to channel, noise, and speakers, the differential approach based on vector field modeling gives an interesting possible solution. The joint vector field is probed through matching between the initial model and the given speech data (with channel, noise, and speaker differences). The mismatch between them is modeled as an overlapped sum of three vector fields, i.e., the channel vector field $\Delta \boldsymbol{C}_H$, the noise vector field $\Delta \boldsymbol{C}_N$, and the speaker vector field $\Delta \boldsymbol{C}_S$. These distinct vector fields have different differential properties. They may be decomposed into individual components by minimizing the total modeling error between the data and their differential properties. This suggests an optimal solution search problem for maximum uniformity of $\Delta \boldsymbol{C}_H$, maximum fitness of $\Delta \boldsymbol{C}_N$ formulated by the Jacobian matrix, and maximum smoothness of the speaker vector field $\Delta \boldsymbol{C}_S$. Further formulation of speaker differences such as vocal tract length as a constrained model of describing speaker differences may be effective in decomposition of vector field components. Exeperimental studies should be yet to be done.

6 CONCLUSION

We have discussed a vector field model for differential approach to acoustic model adaptation. This idea discusses the differential property (e.g., Jacobian) of the space the model is placed instead of discussing the model itself. It gives an interesting class of algorithms for channel, noise, and speaker adaptation, including joint estimation and adaptation to these components.

As for a specific application of this idea to adaptation to noisy speech, Jacobian Adaptation (JA) to noise gives a fast and effcient algorithm. 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.

Another specific application to speaker adaptation provides a simple and significantly effective algorithm. One adavantage is that this algorithm can be combined with other algorithms such as MAP estimation.

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

[4] K. Ohkura, M. Sugiyama and S. Sagayama: "Speaker Adaptation Based on Transfer Vector Field Smoothing with Continuous Mixture Density HMMs," Proc. ICSLP92, pp. 369–372, Oct 1992.

[5] H. Hattori and S. Sagayama: "Vector Field Smoothing Principle for Speaker Adaptation," Proc. ICSLP92, pp. 381–384, Oct 1992.