HIDDEN MARKOV MODELS BASED ON MULTI-SPACE PROBABILITY DISTRIBUTION FOR PITCH PATTERN MODELING

Keiichi Tokuda [1], *Takashi Masuko* [2], *Noboru Miyazaki* [3], *Takao Kobayashi* [2]

[1] Department of Computer Science, Nagoya Institute of Technology, Nagoya, 466-8555 Japan
[2] Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama, 226-8502 Japan
[3] NTT Basic Research Laboratories, Atsugi, 243-0198 Japan
Email: tokuda@ics.nitech.ac.jp, {masuko,tkobayas}@ip.titech.ac.jp, nmiya@atom.brl.ntt.co.jp

ABSTRACT

This paper discusses a hidden Markov model (HMM) based on multi-space probability distribution (MSD). The HMMs are widely-used statistical models to characterize the sequence of speech spectra and have successfully been applied to speech recognition systems. From these facts, it is considered that the HMM is useful for modeling pitch patterns of speech. However, we cannot apply the conventional discrete or continuous HMMs to pitch pattern modeling since the observation sequence of pitch pattern is composed of one-dimensional continuous values and a discrete symbol which represents "unvoiced". MSD-HMM includes discrete HMM and continuous mixture HMM as special cases, and further can model the sequence of observation vectors with variable dimension including zero-dimensional observations, i.e., discrete symbols. As a result, MSD-HMMs can model pitch patterns without heuristic assumption. We derive a reestimation algorithm for the extended HMM and show that it can find a critical point of the likelihood function.

1. INTRODUCTION

The hidden Markov models (HMMs) are widely-used statistical models to characterize the sequence of speech spectra, and the performance of HMM-based speech recognition systems have been improved by techniques which utilize the flexibility of HMMs: context-dependent modeling, dynamic feature parameters, mixtures of Gaussian densities, tying techniques, speaker/environment adaptation techniques. From these facts, one can surmise that the HMM is useful for modeling pitch patterns of speech, and further modeling pitch patterns and speech spectra in a unified framework with feature vectors which consist of spectral and pitch parameters.

However, we cannot apply the conventional discrete or continuous HMMs to pitch pattern modeling since pitch values are not defined in the unvoiced region, i.e., the observation sequence of pitch pattern is composed of one-dimensional continuous values and discrete symbol which represents "unvoiced". Several methods have been investigated [1] for handling the unvoiced region: (i) replacing each "unvoiced" symbol by a random vector generated from a probability density function (pdf) with a large variance and then modeling the random vectors explicitly in the continuous HMMs [2], (ii) modeling the "unvoiced" symbols explicitly in the continuous HMMs by replacing "unvoiced" symbol by 0 and adding an extra pdf for the "unvoiced" symbol to each mixture, (iii) assuming that pitch values is always exist but they cannot observed in the unvoiced region and applying the EM algorithm [3].

This paper describes a new kind of HMM for pitch pattern modeling, in which the state output probabilities are defined by multi-space probability distributions (MSDs). Each space in the MSD has its weight and continuous probability density function whose dimension depends on the space. An observation consists of an n-dimensional continuous vector and a set of space indices which specify n-dimensional spaces. We assume that zero-dimensional space has only one sample point which corresponds to a discrete symbol. It is noted that the MSD is the same as the discrete probability distribution if all spaces are zero-dimensional. On the other hand, the MSD is the same as the continuous G-mixture density if all G spaces are n-dimensional and the set of space indices always contains all space indices. Accordingly, MSD-HMM includes the discrete and continuous mixture HMMs as special cases, and further can model the observation sequence composed of continuous vectors with variable dimension including zero-dimentional observations, i.e., discrete symbols. As a result, MSD-HMMs can model pitch patterns without heuristic assumption. Reestimation formulas for the extended HMM are derived, and it is shown that the reestimation algorithm can find a critical point of the likelihood function.

This paper is organized as follows. Multi-space probability distribution and MSD-HMM are defined in Sections 2 and 3, respectively. A reestimation algorithm for MSD-HMMs are derived in Section 4. The relation between the conventional and the proposed HMMs, and the application of MSD-HMM to pitch pattern modeling are discussed in Section 5. Concluding remarks and our plans for future work are presented in the final section.

2. MULTI-SPACE PROBABILITY DISTRIBUTION

We consider a sample space Ω shown in Fig. 1, which consists of G spaces:

$$\Omega = \bigcup_{g=1}^{G} \Omega_g \tag{1}$$

where Ω_g is an n_g-dimensional real space R^{n_g}, and specified by space index g. Each space Ω_g has its probability w_g, i.e., $P(\Omega_g) = w_g$, where $\sum_{g=1}^{G} w_g = 1$. If $n_g > 0$, each space has a probability density function $\mathcal{N}_g(\boldsymbol{x})$, $\boldsymbol{x} \in R^{n_g}$, where $\int_{R^{n_g}} \mathcal{N}_g(\boldsymbol{x}) d\boldsymbol{x} = 1$.

This work was partially supported by the Ministry of Education, Science and Culture of Japan, Grant-in-Aid for Encouragement of Young Scientists, 0780226, 1998.

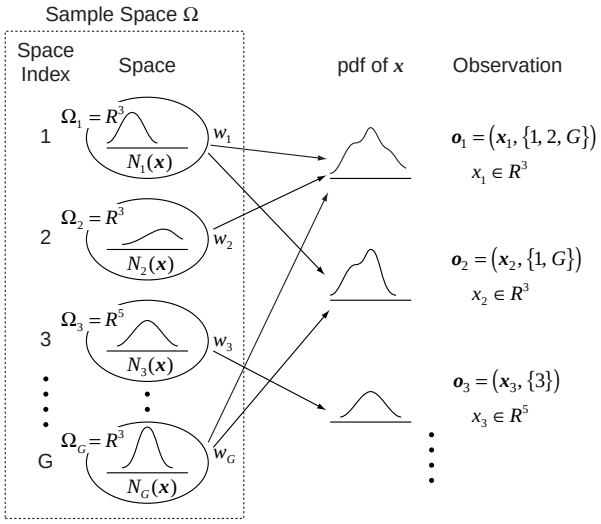

Figure 1: Multi-space probability distribution and observations.

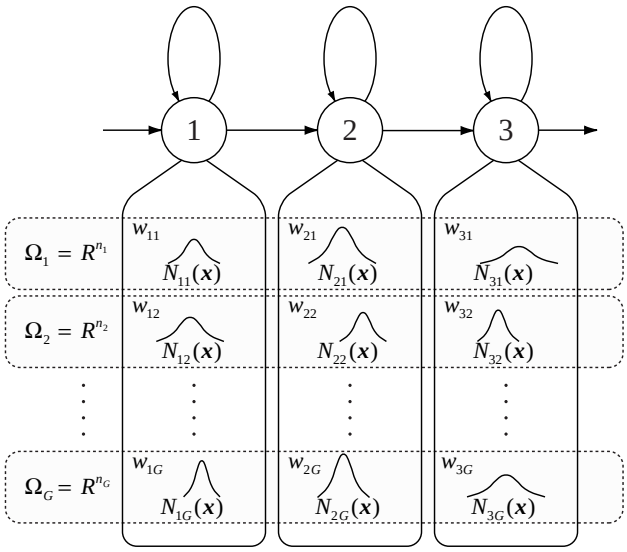

Figure 2: An HMM based on multi-space probability distribution.

We assume that Ω_g contains only one sample point if $n_g = 0$. Accordingly, letting $P(E)$ be the probability distribution, we have

$$P(\Omega) = \sum_{g=1}^{G} P(\Omega_g) = \sum_{g=1}^{G} w_g \int_{R^{n_g}} \mathcal{N}_g(\boldsymbol{x}) d\boldsymbol{x} = 1. \quad (2)$$

It is noted that, although $\mathcal{N}_g(\boldsymbol{x})$ does not exist for $n_g = 0$ since Ω_g contains only one sample point, for simplicity of notation, we define as $\mathcal{N}_g(\boldsymbol{x}) \equiv 1$ for $n_g = 0$.

Each event E, which will be considered in this paper, is represented by a random variable $\boldsymbol{o}$ which consists of a continuous random variable $\boldsymbol{x} \in R^n$ and a set of space indices X, that is,

$$\boldsymbol{o} = (\boldsymbol{x}, X) \quad (3)$$

where all spaces specified by X are n-dimensional. The observation probability of $\boldsymbol{o}$ is defined by

$$b(\boldsymbol{o}) = \sum_{g \in S(\boldsymbol{o})} w_g \mathcal{N}_g(V(\boldsymbol{o})) \quad (4)$$

where

$$V(\boldsymbol{o}) = \boldsymbol{x}, \qquad S(\boldsymbol{o}) = X. \quad (5)$$

Some examples of observations are shown in Fig. 1. An observation $\boldsymbol{o}_1$ consists of three-dimensional vector $\boldsymbol{x}_1 \in R^3$ and a set of space indices $X_1 = \{1, 2, G\}$. Thus the random variable $\boldsymbol{x}$ is drawn from one of three spaces Ω_1, Ω_2, $\Omega_G \in R^3$, and its probability density function is given by $w_1\mathcal{N}_1(\boldsymbol{x}) + w_2\mathcal{N}_2(\boldsymbol{x}) + w_G\mathcal{N}_G(\boldsymbol{x})$.

The probability distribution defined in the above, which will be refered to as *multi-space probability distribution* (MSD) in this paper, is the same as the discrete distribution and the continuous distribution when $n_g \equiv 0$ and $n_g \equiv m > 0$, respectively. Further, if $S(\boldsymbol{o}) \equiv \{1, 2, \ldots, G\}$, the countinuous distribution is represented by a G-mixture probability density function. Thus multi-space probability distribution is more general than either discrete or continuous distributions.

3. HMMS BASED ON MULTI-SPACE PROBABILITY DISTRIBUTION

The output probability in each state of MSD-HMM is given by the multi-space probability distribution defined in the previous section. An N-state MSD-HMM λ is specified by initial state probability distribution $\pi = \{\pi_j\}_{j=1}^N$, the state transition probability distribution $A = \{a_{ij}\}_{i,\,j=1}^N$, and state output probability distribution $B = \{b_i(\cdot)\}_{i=1}^N$, where

$$b_i(\boldsymbol{o}) = \sum_{g \in S(\boldsymbol{o})} w_{ig}\ \mathcal{N}_{ig}(V(\boldsymbol{o})), \quad i = 1, 2, \ldots, N. \quad (6)$$

As shown in Fig. 2, each state i has G probability density functions $\mathcal{N}_{i1}(\cdot), \mathcal{N}_{i2}(\cdot), \ldots, \mathcal{N}_{iG}(\cdot)$, and their weights $w_{i1}, w_{i2}, \ldots, w_{iG}$.

Observation probability of $\boldsymbol{O} = \{\boldsymbol{o}_1, \boldsymbol{o}_2, \ldots, \boldsymbol{o}_T\}$ is written as

$$\begin{aligned} P(\boldsymbol{O}|\lambda) &= \sum_{\text{all } \boldsymbol{q}} \prod_{t=1}^{T} a_{q_{t-1}q_t}\, b_{q_t}(\boldsymbol{o}_t) \\ &= \sum_{\text{all } \boldsymbol{q}, \boldsymbol{l}} \prod_{t=1}^{T} a_{q_{t-1}q_t}\, w_{q_t l_t} \mathcal{N}_{q_t l_t}(V(\boldsymbol{o}_t)) \end{aligned} \quad (7)$$

where $\boldsymbol{q} = \{q_1, q_2, \ldots, q_T\}$ is a possible state sequence, $\boldsymbol{l} = \{l_1, l_2, \ldots, l_T\} \in \{S(\boldsymbol{o}_1) \times S(\boldsymbol{o}_2) \times \ldots \times S(\boldsymbol{o}_T)\}$ is a sequence of space indices which is possible for the observation sequence $\boldsymbol{O}$, and $a_{q_0 j}$ denotes π_j.

The forward and backward variables:

$$\alpha_t(i) = P(\boldsymbol{o}_1, \boldsymbol{o}_2, \ldots, \boldsymbol{o}_t, q_t = i|\lambda) \quad (8)$$

$$\beta_t(i) = P(\boldsymbol{o}_{t+1}, \boldsymbol{o}_{t+2}, \ldots, \boldsymbol{o}_T|q_t = i, \lambda) \quad (9)$$

can be calculated with the forward-backward inductive procedure in a manner similar to the conventional HMMs. According to the definitions, (7) can be calculated as

$$P(\boldsymbol{O}|\lambda) = \sum_{i=1}^{N} \alpha_T(i) = \sum_{i=1}^{N} \beta_1(i). \quad (10)$$

The forward and backward variables are also used for calculating the reestimation formulas derived in the next section.

4. REESTIMATION ALGORITHM

For a given observation sequence $\boldsymbol{O}$ and a particular choice of MSD-HMM, the objective in maximum likelihood estimation is to maximize the observation likelihood $P(\boldsymbol{O}|\lambda)$ given by (7), over all parameters in λ. In a manner similar to [4], [5], we derive reestimation formulas for the maximum likelihood estimation of MSD-HMM.

4.1. Q-function

An auxiliary function $Q(\lambda', \lambda)$ of current parameters λ' and new parameter λ is defined as follows:

$$Q(\lambda', \lambda) = \sum_{\text{all } \boldsymbol{q}, \boldsymbol{l}} P(\boldsymbol{O}, \boldsymbol{q}, \boldsymbol{l}|\lambda') \log P(\boldsymbol{O}, \boldsymbol{q}, \boldsymbol{l}|\lambda) \tag{11}$$

In the following, we assume $\mathcal{N}_{ig}(\cdot)$ to be the Gaussian density with mean vector $\boldsymbol{\mu}_{ig}$ and covariance matrix $\boldsymbol{\Sigma}_{ig}$. However, extension to elliptically symmetric densities which satisfy the consistency conditions of Kolmogorov is straightforward. We will present the following three theorems without extensive proofs[1]:

Theorem 1

$$Q(\lambda', \lambda) \geq Q(\lambda', \lambda') \rightarrow P(\boldsymbol{O}, \lambda) \geq P(\boldsymbol{O}, \lambda') \tag{12}$$

Theorem 2 *If, for each space Ω_g, there are among $V(\boldsymbol{o}_1), V(\boldsymbol{o}_2), \ldots, V(\boldsymbol{o}_T)$, $n_g + 1$ observations $g \in S(o_t)$, any n_g of which are linearly independent, $Q(\lambda', \lambda)$ has a unique global maximum as a function of λ, and this maximum is the one and only critical point.*

Theorem 3 *A parameter set λ is a critical point of the likelihood $P(\boldsymbol{O}|\lambda)$ if and only if it is a critical point of the Q-function.*

We define the parameter reestimates to be those which maximize $Q(\lambda', \lambda)$ as a function of λ, λ' being the latest estimates. Because of the above theorems, the sequence of resetimates obtained in this way produce a monotonic increase in the likelihood unless λ is a critical point of the likelihood.

4.2. Maximization of Q-function

For given observation sequence $\boldsymbol{O}$ and model λ', we derive parameters of λ which maximize $Q(\lambda', \lambda)$. From (7), $\log P(\boldsymbol{O}, \boldsymbol{q}, \boldsymbol{l}|\lambda)$ can be written as

$$\begin{aligned}
&\log P(\boldsymbol{O}, \boldsymbol{q}, \boldsymbol{l}|\lambda) \\
&= \sum_{t=1}^{T} \left(\log a_{q_{t-1} q_t} + \log w_{q_t l_t} + \log \mathcal{N}_{q_t l_t}(V(\boldsymbol{o}_t)) \right). \qquad (13)
\end{aligned}$$

Hence Q-function (11) can be written as

$$\begin{aligned}
Q(\lambda', \lambda) &= \sum_{i=1}^{N} P(\boldsymbol{O}, q_1 = i|\lambda') \log \pi_i \\
&+ \sum_{i,j=1}^{N} \sum_{t=1}^{T-1} P(\boldsymbol{O}, q_t = i, q_{t+1} = j|\lambda') \log a_{ij} \\
&+ \sum_{i=1}^{N} \sum_{g=1}^{G} \sum_{t \in T(\boldsymbol{O}, g)} P(\boldsymbol{O}, q_t = i, l_t = g|\lambda') \log w_{ig} \\
&+ \sum_{i=1}^{N} \sum_{g=1}^{G} \sum_{t \in T(\boldsymbol{O}, g)} P(\boldsymbol{O}, q_t = i, l_t = g|\lambda') \log \mathcal{N}_{ig}(V(\boldsymbol{o}_t))
\end{aligned} \tag{14}$$

where

$$T(\boldsymbol{O}, g) = \{t \,|\, g \in S(\boldsymbol{o}_t)\}. \tag{15}$$

The parameter set $\lambda = (\pi, A, B)$ which maximizes (14), subject to the stochastic constraints $\sum_{i=1}^{N} \pi_i = 1$, $\sum_{j=1}^{N} a_{ij} = 1$ and $\sum_{g=1}^{G} w_g = 1$, can be derived as

$$\pi_i = \sum_{g \in S(\boldsymbol{o}_1)} \gamma'_1(i, g) \tag{16}$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi'_t(i, j)}{\sum_{t=1}^{T-1} \sum_{g \in S(\boldsymbol{o}_t)} \gamma'_t(i, g)} \tag{17}$$

$$w_{ig} = \frac{\sum_{t \in T(\boldsymbol{O}, g)} \gamma'_t(i, g)}{\sum_{h=1}^{G} \sum_{t \in T(\boldsymbol{O}, h)} \gamma'_t(i, h)} \tag{18}$$

$$\boldsymbol{\mu}_{ig} = \frac{\sum_{t \in T(\boldsymbol{O}, g)} \gamma'_t(i, g)\, V(\boldsymbol{o}_t)}{\sum_{t \in T(\boldsymbol{O}, g)} \gamma'_t(i, g)}, \quad n_g > 0 \tag{19}$$

$$\boldsymbol{\Sigma}_{ig} = \frac{\sum_{t \in T(\boldsymbol{O}, g)} \gamma'_t(i, g)(V(\boldsymbol{o}_t) - \boldsymbol{\mu}_{ig})(V(\boldsymbol{o}_t) - \boldsymbol{\mu}_{ig})^T}{\sum_{t \in T(\boldsymbol{O}, g)} \gamma'_t(i, g)}, \quad n_g > 0 \tag{20}$$

where $\gamma_t(i, h)$ and $\xi_t(i, j)$ can be calculated by using the forward variable $\alpha_t(i)$ and backward variable $\beta_t(i)$ as follows:

$$\begin{aligned}
\gamma_t(i, h) &= P(q_t = i, l_t = h|\boldsymbol{O}, \lambda) \\
&= \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)} \cdot \frac{w_{ih}\mathcal{N}_{ih}(V(\boldsymbol{o}_t))}{\sum_{g \in S(\boldsymbol{o}_t)} w_{ig}\mathcal{N}_{ig}(V(\boldsymbol{o}_t))}
\end{aligned} \tag{21}$$

$$\begin{aligned}
\xi_t(i, j) &= P(q_t = i, q_{t+1} = j|\boldsymbol{O}, \lambda) \\
&= \frac{\alpha_t(i) a_{ij} b_j(\boldsymbol{o}_{t+1}) \beta_{t+1}(j)}{\sum_{h=1}^{N} \sum_{k=1}^{N} \alpha_t(h) a_{hk} b_k(\boldsymbol{o}_{t+1}) \beta_{t+1}(k)}
\end{aligned} \tag{22}$$

[1] Complete description of the proofs can be found in [6].

From the condition mentioned in Theorem 2, it can be shown that each $\mathbf{\Sigma}_{ig}$ is positive definite.

5. APPLICATION TO PITCH PATTERN MODELING

The MSD-HMM includes the discrete HMM and the continuous mixture HMM as special cases since the multi-space probability distribution includes the discrete distribution and the continuous distribution. If $n_g \equiv 0$, the MSD-HMM is the same as the discrete HMM. In the case where $S(\boldsymbol{o}_t)$ specifies one space, i.e., $|S(\boldsymbol{o}_t)| \equiv 1$, the MSD-HMM is exactly the same as the conventional discrete HMM. If $|S(\boldsymbol{o}_t)| \geq 1$, the MSD-HMM is the same as the discrete HMM based on the multi-labeling VQ [7]. If $n_g \equiv m > 0$ and $S(\boldsymbol{o}) \equiv \{1, 2, \ldots, G\}$, the MSD-HMM is the same as the continuous G-mixture HMM. These can also be confirmed by the fact that if $n_g \equiv 0$ and $|S(\boldsymbol{o}_t)| \equiv 1$, the reestimation formulas (16)-(18) are the same as those for discrete HMM of codebook size G, and if $n_g \equiv m$ and $S(\boldsymbol{o}_t) \equiv \{1, 2, \ldots, G\}$, the reestimation formulas (16)-(20) are the same as those for continuous HMM with m-dimensional G-mixture densities. Further, the MSD-HMM can model the sequence of observation vectors with variable dimension including zero-dimensional observations, i.e., discrete symbols.

While the observation of pitch has a continuous value in the voiced region, there exist no value for the unvoiced region. We can model this kind of observation sequence assuming that the observed pitch value occurs from one-dimensional spaces and the "unvoiced" symbol occurs from the zero-dimensional space defined in Section 2, that is, by setting $n_g = 1$ $(g = 1, 2, \ldots, G-1)$, $n_G = 0$ and

$$S(\boldsymbol{o}_t) = \begin{cases} \{1, 2, \ldots, G-1\}, & \text{(voiced)} \\ \{G\}, & \text{(unvoiced)} \end{cases}, \qquad (23)$$

the MSD-HMM can cope with pitch patterns including the unvoiced region without heuristic assumption. In this case, the observed pitch value is assumed to be drawn from a continuous $(G-1)$-mixture probability density function.

Experiments in [8] have shown that the likelihood is increased monotonically by calculating the reestimation formulas iteratively. From the trained MSD-HMMs, we can generate pitch patterns which approximate those of natural speech by using an algorithm [9] for speech parameter generation from HMMs with dynamic features. An example is shown in Fig. 3 without the explanation of experimental conditions [8] because of limitations of space.

6. CONCLUSION

A multi-space probability distribution HMM has been proposed and its reestimation formulas are derived. The MSD-HMM includes the discrete HMM and the continuous mixture HMM as special cases, and further can cope with the sequence of observation vectors with variable dimension including zero-dimensional observations, i.e., discrete symbols. As a result, MSD-HMMs can model pitch patterns without heuristic assumption.

In the near future, we will present a speech synthesis system in which sequences of speech spectra [10], pitch patterns [8] and state durations [11] are modeled by MSD-HMM in a unified framework. This system may synthesize speech with various voice characteristics by applying a speaker adaptation technique developed for speech recognition systems [12]. Pitch pattern modeling based on MSD-HMM may also be useful for enhancing the speech recognition performance.

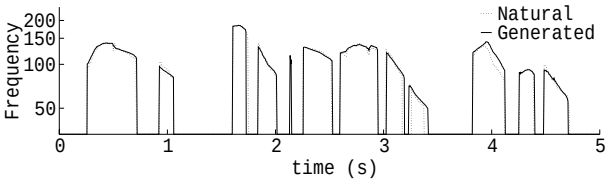

Figure 3: Pitch pattern generation based on MSD-HMM.

7. REFERENCES

[1] U. Jensen, R. K. Moore, P. Dalsgaard, and B Lindberg, "Modeling intonation contours at the phrase level using continuous density hidden Markov models," Computer Speech and Language, vol.8, no.3, pp.247–260, Aug. 1994.

[2] G. J. Freij and F. Fallside, "Lexical stress recognition using hidden Markov models," in *Proc. ICASSP*, 1988, pp.135–138.

[3] K. Ross and M. Ostendorf, "A dynamical system model for generating F_0 for synthesis," in *Proc. ESCA/IEEE Workshop on Speech Synthesis*, 1994, pp.131–134.

[4] L. A. Liporace, "Maximum Likelihood Estimation for Multivariate Observations of Markov Sources," IEEE Trans. Information Theory, vol.28, no.5, pp.729–734, 1982.

[5] B.-H. Juang, "Maximum-likelihood estimation for mixture multivariate stochastic observations of Markov chains," AT&T Technical Journal, vol.64, no.6, pp.1235–1249, 1985.

[6] N. Miyazaki, K. Tokuda, T. Masuko and T. Kobayashi, "An HMM based on multi-space probability distributions and its application to pitch pattern modeling," *IEICE Technical Report*, SP98-11, 1998 (in Japanese).

[7] M. Nishimura and K. Toshioka, "HMM-based speech recognition using multi-dimensional multi-labeling," in *Proc. ICASSP*, 1987, pp.1163–1166.

[8] N. Miyazaki, K. Tokuda, T. Masuko and T. Kobayashi, "A study on pitch pattern generation using HMMs based on multi-space probabilit y distributions," *IEICE Technical Report*, SP98-12, 1998 (in Japanese).

[9] K. Tokuda, T. Masuko, T. Yamada, T. Kobayashi and S. Imai, "An algorithm for speech parameter generation from continuous mixture HMMs with dynamic features," in *Proc. EUROSPEECH*, 1995, pp.757–760.

[10] T. Masuko, K. Tokuda, T. Kobayashi and S. Imai, "Speech synthesis from HMMs using dynamic features," in *Proc. ICASSP*, 1996, pp.389–392.

[11] T. Yoshimura, K. Tokuda, T. Masuko, T. Kobayashi and T. Kitamura, "Duration modeling for HMM-based speech synthesis," in *Proc. ICLSP*, 1998.

[12] M. Tamura, T. Masuko, K. Tokuda and T. Kobayashi, "Speaker adaptation for HMM-based speech synthesis system using MLLR," in *Proc. ESCA/COCOSDA Third International Workshop on Speech Synthesis*, 1998.