Disclaimer

- The material provided in this document is not my original work and is a summary of some one else’s work(s).
- A simple Google search of the title of the document will direct you to the original source of the material.
- I do not guarantee the accuracy, completeness, timeliness, validity, non-omission, merchantability or fitness of the contents of this document for any particular purpose.
- Downloaded from najeebkhan.github.io

Multi-Space Probability Distribution HMM

Presented by

Najeeb

July 7th 2014

Introduction

- We cannot apply both the conventional discrete and continuous HMMs, to observation sequences which consist of continuous values and discrete symbols
- The proposed HMM includes discrete HMM and continuous HMM as special cases and furthermore can model sequences which consists of observation vectors with variable dimensionality and discrete symbols

Introduction

- The fundamental frequency pattern of speech cannot be modeled by discrete and continuous HMMs since F0 values are not defined in the unvoiced region
- The observation sequence of an F0 pattern is composed of one dimensional continuous values and discrete symbols which represents unvoiced

Introduction

- This paper describes a new kind of HMM in which the state output probabilities are defined by multi-space probability distributions
- Each space in the multi-space probability distribution has its weight and a continuous pdf whose dimensionality depends on the space
- An observation vector consists of an n-dimensional continuous vector and a set of space indices which specify n-dimensional spaces

Multi-Space Probability Distribution

- We consider a sample space Ω, which consists of G spaces

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

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

- For simplicity

$$\mathcal{N}_g(\boldsymbol{x}) \equiv 1 \text{ for } n_g = 0.$$

Multi-Space Probability Distribution

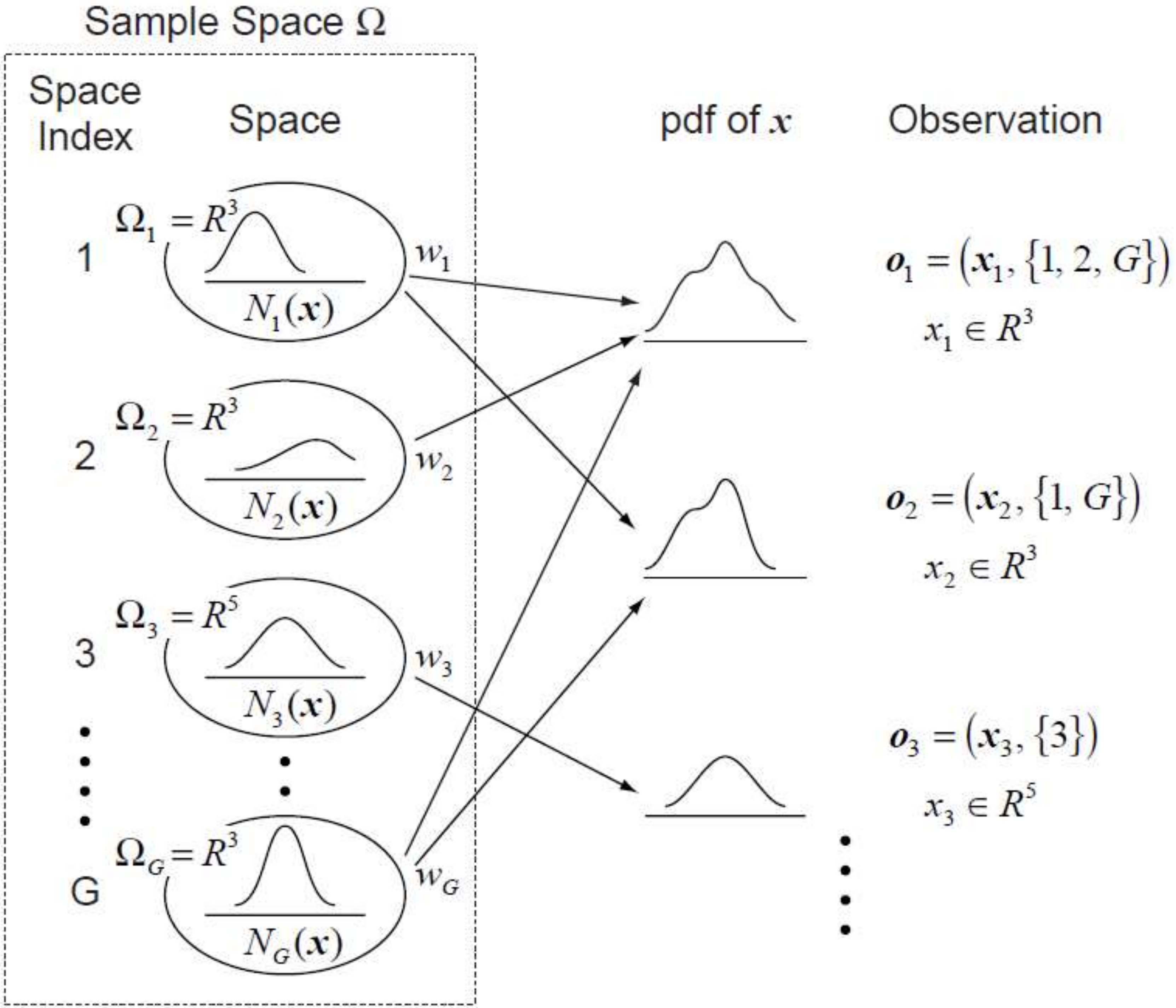
- Each event is represented by a random variable o which consists of a continuous random variable x ∈ R^n and a set of space indices X, that is

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

- where all spaces specified by X are n-dimensional

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

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


Sample Space Ω
Space Index
Space
pdf of $\boldsymbol{x}$
Observation
1
$\Omega_1 = R^3$
$N_1(\boldsymbol{x})$
w_1
2
$\Omega_2 = R^3$
$N_2(\boldsymbol{x})$
w_2
3
$\Omega_3 = R^5$
$N_3(\boldsymbol{x})$
w_3
G
$\Omega_G = R^3$
$N_G(\boldsymbol{x})$
w_G
$\boldsymbol{o}_1 = (\boldsymbol{x}_1, \{1, 2, G\})$
$x_1 \in R^3$
$\boldsymbol{o}_2 = (\boldsymbol{x}_2, \{1, G\})$
$x_2 \in R^3$
$\boldsymbol{o}_3 = (\boldsymbol{x}_3, \{3\})$
$x_3 \in R^5$

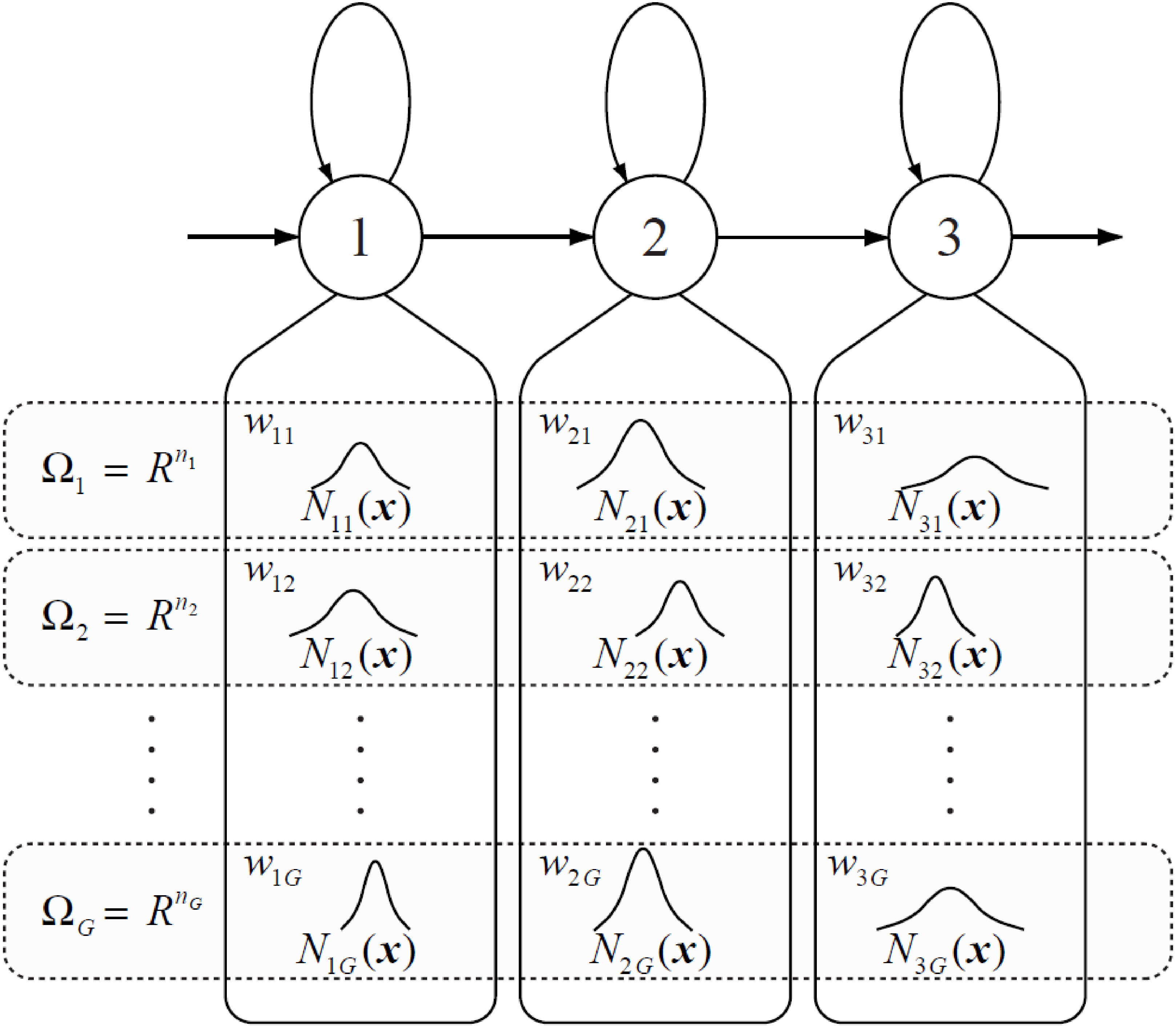
Multi-Space Probability Distribution

- MSD is the same as the discrete distribution and the continuous distribution when $n_g \equiv 0$ and $n_g \equiv m > 0$, respectively
- If S(o) ≡ {1, 2, . . . , G}, the continuous distribution is represented by a G-mixture probability density function
- Thus multi-space probability distribution is more general than either discrete or continuous distributions

Multi-Space Distribution HMM

- The output probability in each state of MSD-HMM is given by the multi-space probability distribution

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


1
2
3
$\Omega_1 = R^{n_1}$
$\Omega_2 = R^{n_2}$
$\Omega_G = R^{n_G}$
w_{11}
$N_{11}(\boldsymbol{x})$
w_{21}
$N_{21}(\boldsymbol{x})$
w_{31}
$N_{31}(\boldsymbol{x})$
w_{12}
$N_{12}(\boldsymbol{x})$
w_{22}
$N_{22}(\boldsymbol{x})$
w_{32}
$N_{32}(\boldsymbol{x})$
w_{1G}
$N_{1G}(\boldsymbol{x})$
w_{2G}
$N_{2G}(\boldsymbol{x})$
w_{3G}
$N_{3G}(\boldsymbol{x})$

Multi-Space Distribution HMM

- Observation probability of O = {o_1, o_2, … , o_T } is written as

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

- where q = {q_1, q_2, … , q_T} is a possible state sequence, l = {l_1, l_2, … , l_T} ∈ {S(o_1)×S(o_2)×…×S(o_T)} is a sequence of space indices which is possible for the observation sequence O

Multi-Space Distribution HMM

- The forward and backward variables

$$\begin{aligned}\alpha_t(i) &= P(\boldsymbol{o}_1, \boldsymbol{o}_2, \ldots, \boldsymbol{o}_t, q_t = i|\lambda) \\ \beta_t(i) &= P(\boldsymbol{o}_{t+1}, \boldsymbol{o}_{t+2}, \ldots, \boldsymbol{o}_T|q_t = i, \lambda)\end{aligned}$$

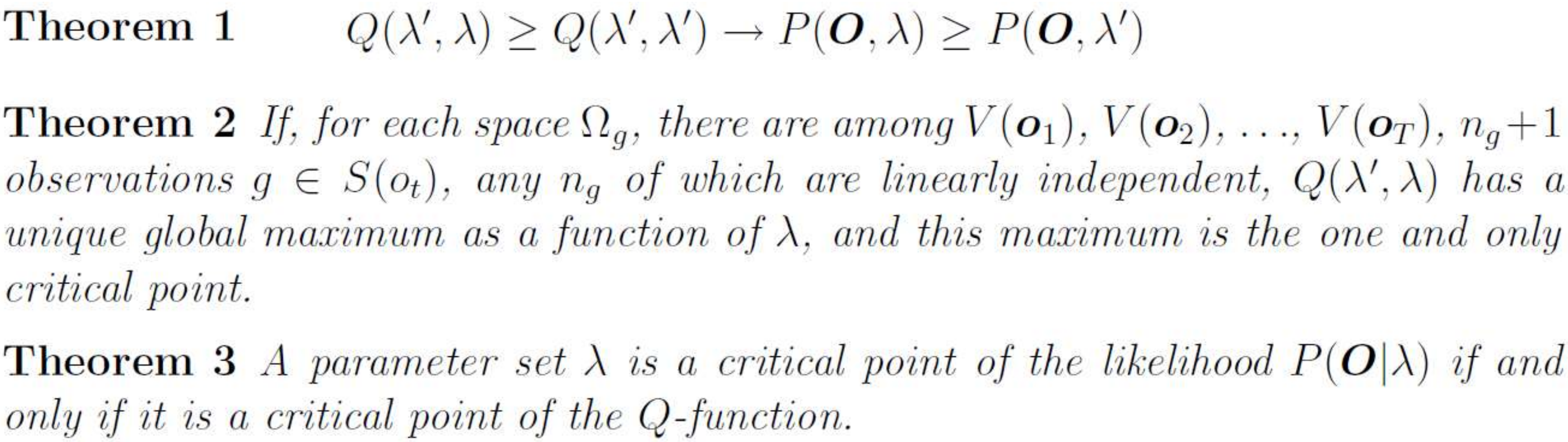
- can be calculated with the forward-backward inductive procedure in a manner similar to the conventional HMMs

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

Re-estimation algorithm for MSD-HMM training

- An auxiliary function Q(λ’, λ) of current parameters λ’ and new parameter λ is defined as follows

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



Theorem 1 $Q(\lambda', \lambda) \geq Q(\lambda', \lambda') \rightarrow P(\boldsymbol{O}, \lambda) \geq P(\boldsymbol{O}, \lambda')$

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.*

Re-estimation algorithm for MSD-HMM training

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

Re-estimation algorithm for MSD-HMM training

- For given observation sequence O and model λ’, we derive parameters of λ which maximize Q(λ’, λ)
- log P(O, q, l|λ) can be written as

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

Re-estimation algorithm for MSD-HMM training

- The Q function can be written as

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

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

Re-estimation algorithm for MSD-HMM training

- The parameter set λ = (π,A,B) which maximizes Q, subject to the stochastic constraints can be derived as

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

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

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

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

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

$$n_g > 0$$

Re-estimation algorithm for MSD-HMM training

- 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

$$\begin{aligned}
\gamma_t(i,h) &= P(q_t = i, l_t = h|\boldsymbol{O}, \lambda) \\
&= \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)} \cdot \frac{w_{ih}\mathcal{N}_{ih}(V(\boldsymbol{o}_t))}{\sum_{g \in S(\boldsymbol{o}_t)} w_{ig}\mathcal{N}_{ig}(V(\boldsymbol{o}_t))} \\
\xi_t(i,j) &= P(q_t = i, q_{t+1} = j|\boldsymbol{O}, \lambda) \\
&= \frac{\alpha_t(i)a_{ij}b_j(\boldsymbol{o}_{t+1})\beta_{t+1}(j)}{\sum_{h=1}^{N}\sum_{k=1}^{N} \alpha_t(h)a_{hk}b_k(\boldsymbol{o}_{t+1})\beta_{t+1}(k)}
\end{aligned}$$

Application to F0 pattern modeling

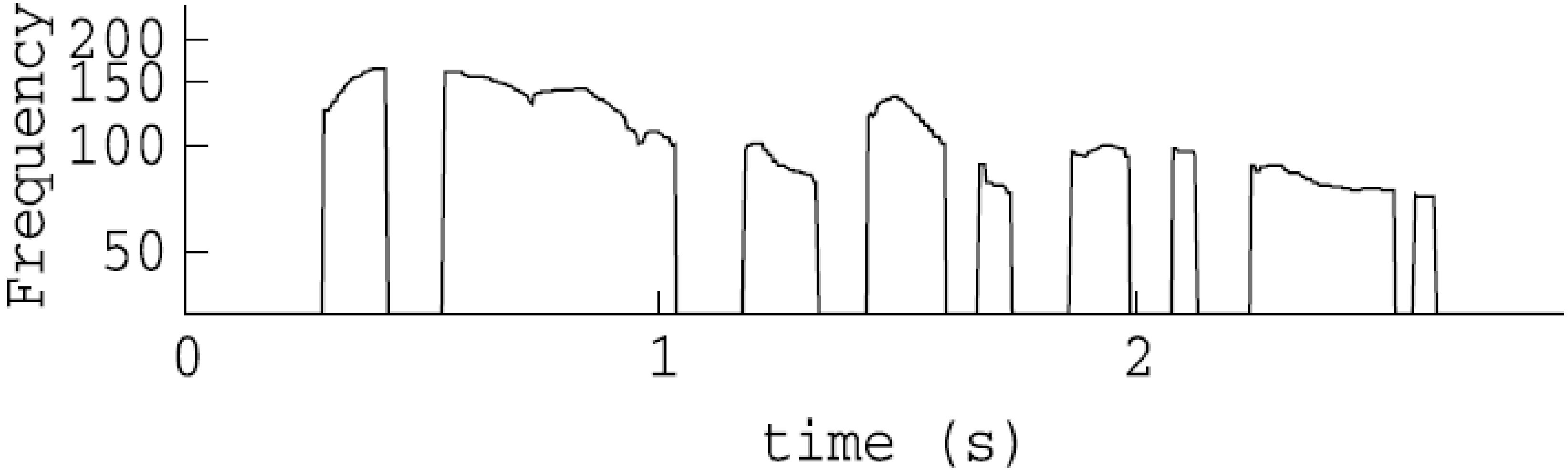
- If $n_g \equiv 0$, the MSD-HMM is the same as the discrete HMM
- If $n_g \equiv m > 0$ and $S(o) \equiv \{1, 2, . . . , G\}$, the MSD-HMM is the same as the continuous G-mixture HMM
- We can model F0, assuming that observed F0 value occurs from one-dimensional spaces and the “unvoiced” symbol occurs from the zero-dimensional space

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

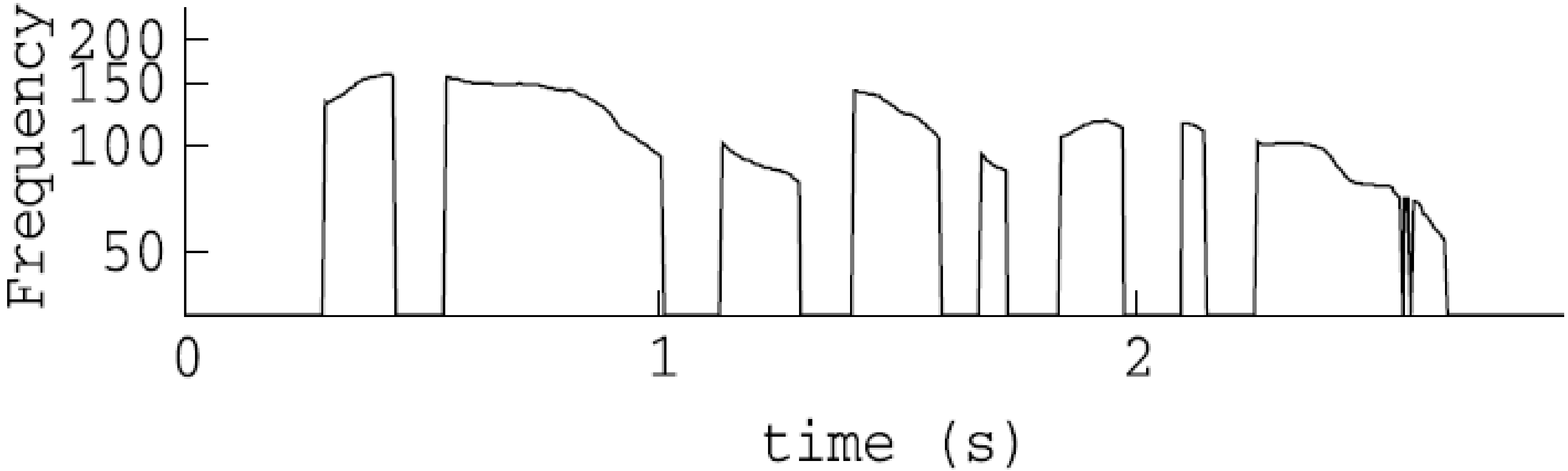
Application to F0 pattern modeling

- From the trained MSD-HMMs we can generate F0 patterns using an algorithm (case 1) for speech parameter generation from HMMs with dynamic features

Application to F0 pattern modeling



(a) natural F0



SIMULTANEOUS MODELING OF SPECTRUM, PITCH AND DURATION IN HMM BASED SPEECH SYNTHESIS

INTRODUCTION

- A speech synthesis system is constructed in which spectrum, pitch and state duration are modeled simultaneously in a unified framework of HMM
- In the system, pitch and state duration are modeled by MSD HMMs and multi-dimensional Gaussian distributions, respectively
- The feature vector of HMMs used in the system consists of two streams, i.e., the one for spectral parameter vector and the other for pitch parameter vector, and each phoneme HMM has its state duration densities

INTRODUCTION

- The distributions for spectral parameter, pitch parameter and state duration are clustered independently by using a decision tree based context clustering technique

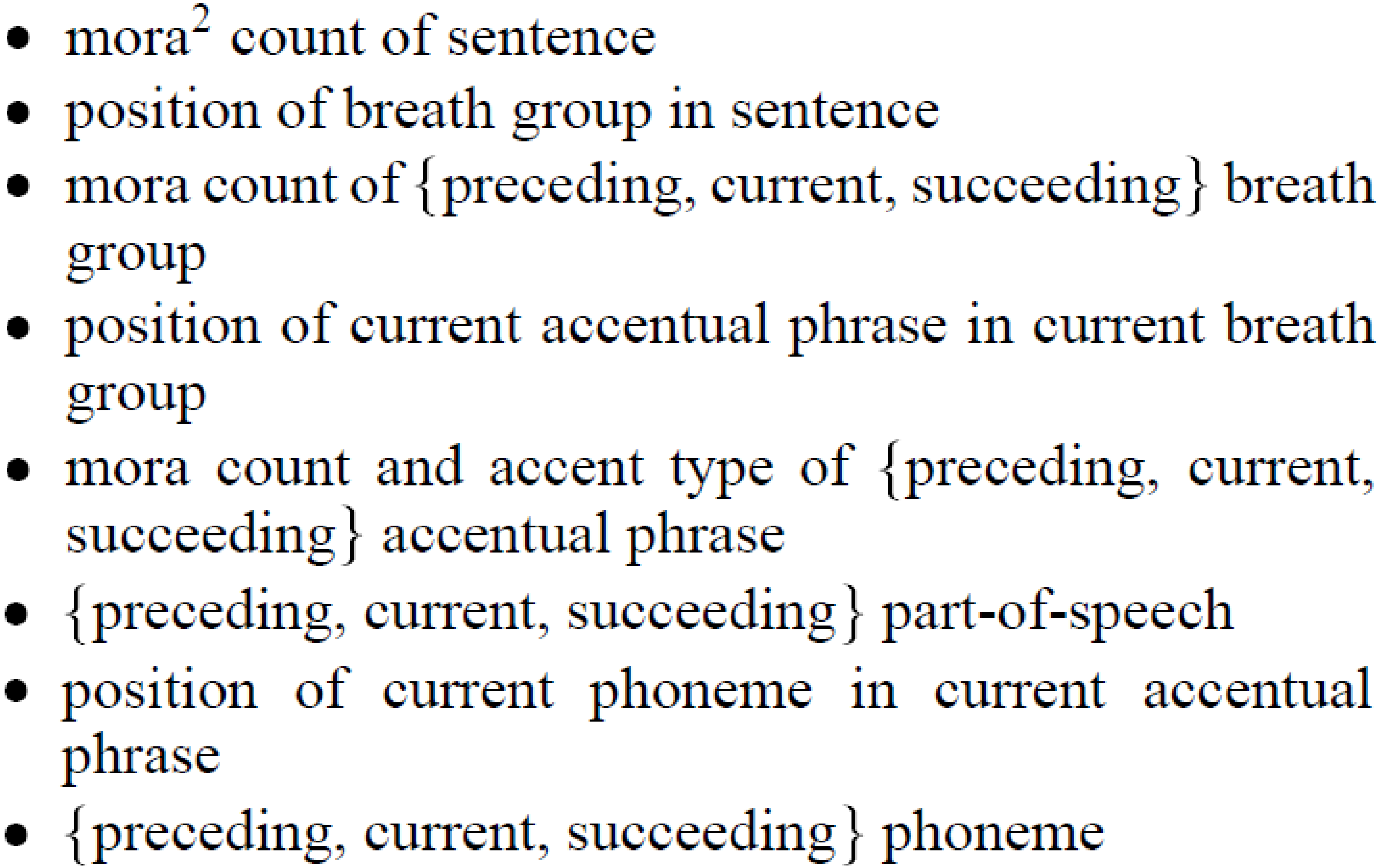
SIMULTANEOUS MODELING

- Spectrum and Pitch Model
  - If spectrum models and pitch models are embedded trained separately, speech segmentations may be discrepant between them
  - To avoid this problem, context dependent HMMs are trained with feature vector which consists of spectrum, pitch and their dynamic features

SIMULTANEOUS MODELING

- State Duration Model
  - State duration densities are modeled by single Gaussian distributions
  - Dimension of state duration densities is equal to the number of state of HMM
  - Since state durations are modeled by continuous distributions, our approach has the following advantages
    - The speaking rate of synthetic speech can be varied easily
    - There is no need for label boundaries
  - State duration densities are estimated by using state occupancy probabilities which are obtained in the last iteration of embedded re-estimation

CONTEXT DEPENDENT MODEL



- mora[2] count of sentence
- position of breath group in sentence
- mora count of {preceding, current, succeeding} breath group
- position of current accentual phrase in current breath group
- mora count and accent type of {preceding, current, succeeding} accentual phrase
- {preceding, current, succeeding} part-of-speech
- position of current phoneme in current accentual phrase
- {preceding, current, succeeding} phoneme

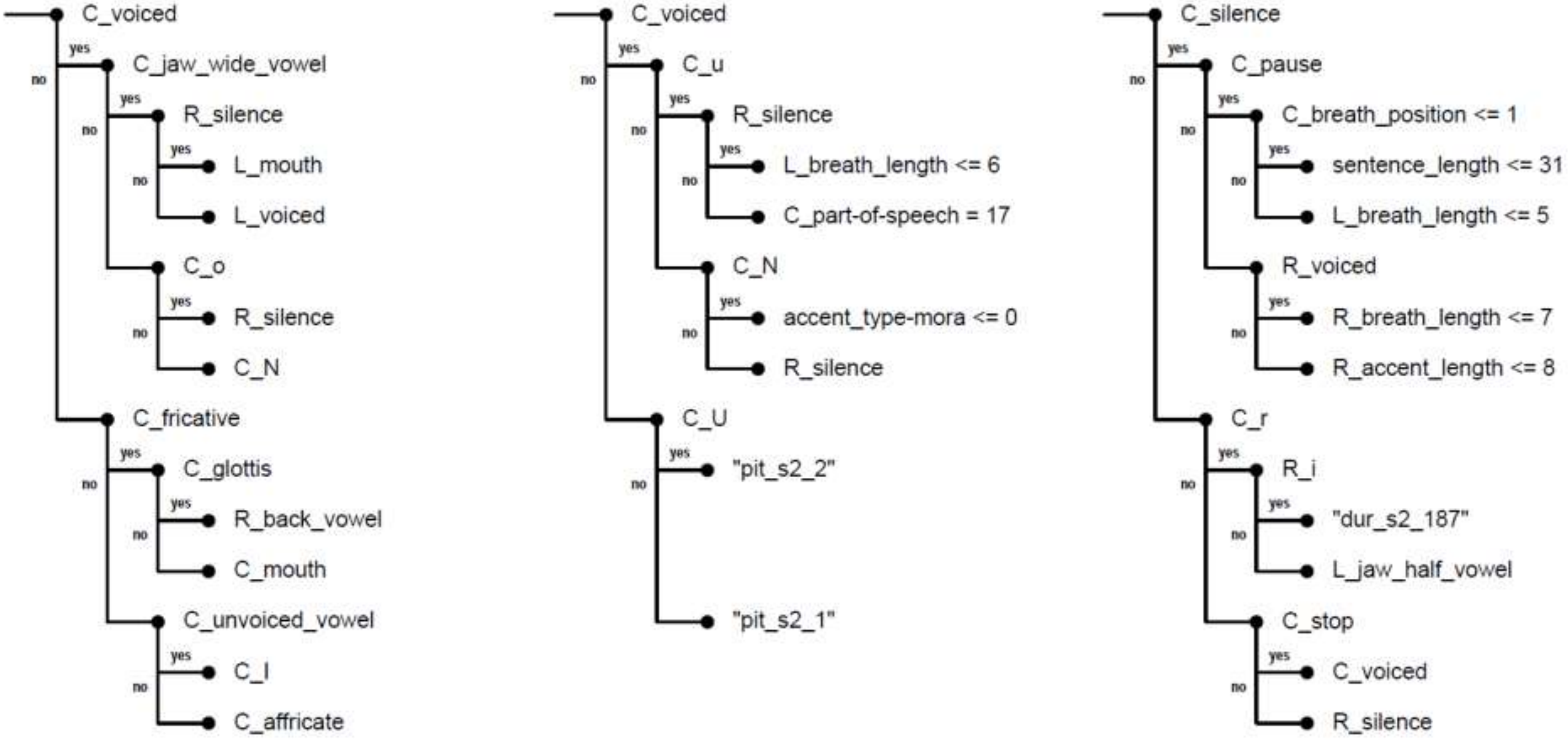
CONTEXT DEPENDENT MODEL

- As contextual factors increase, their combinations also increase exponentially
  - Model parameters with sufficient accuracy cannot be estimated with limited training data
  - It is impossible to prepare speech database which includes all combinations of contextual factors
- To overcome this problem, we apply a decision tree based context clustering technique to distributions for spectrum, pitch and state duration

CONTEXT DEPENDENT MODEL

- The decision tree based context clustering algorithm have been extended for MSD HMMs
- Since each of spectrum, pitch and duration have its own influential contextual factors, the distributions for spectral parameter and pitch parameter and the state duration are clustered independently

CONTEXT DEPENDENT MODEL



(a) Tree for Spectrum Model
(1st state)

(b) Tree for Pitch Model
(1st state)

(c) Tree for State Duration Model

CONTEXT DEPENDENT MODEL

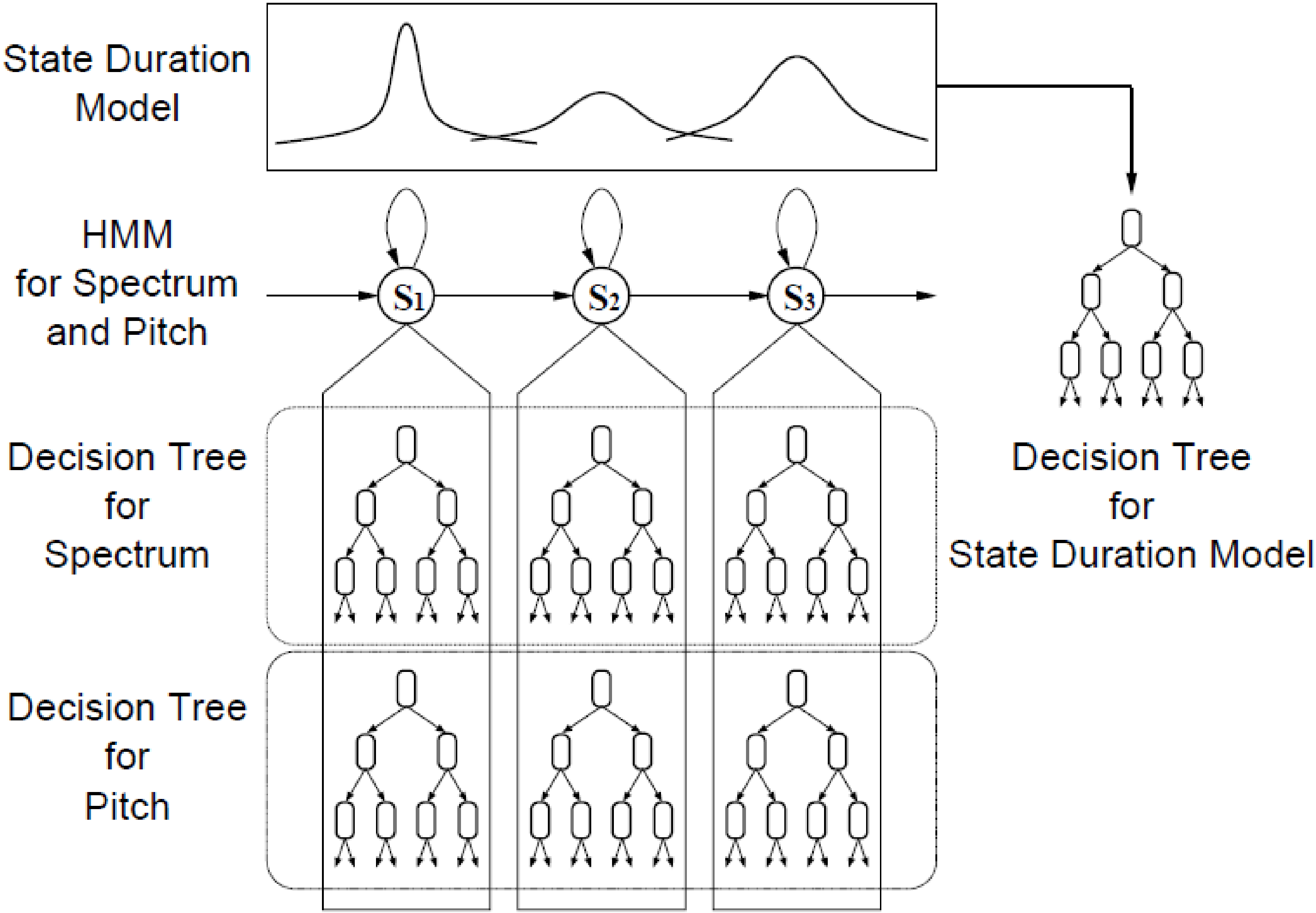

State Duration Model
HMM for Spectrum and Pitch
S_1
S_2
S_3
Decision Tree for Spectrum
Decision Tree for Pitch
Decision Tree for State Duration Model

Figure 2. Decision trees.

Synthesis

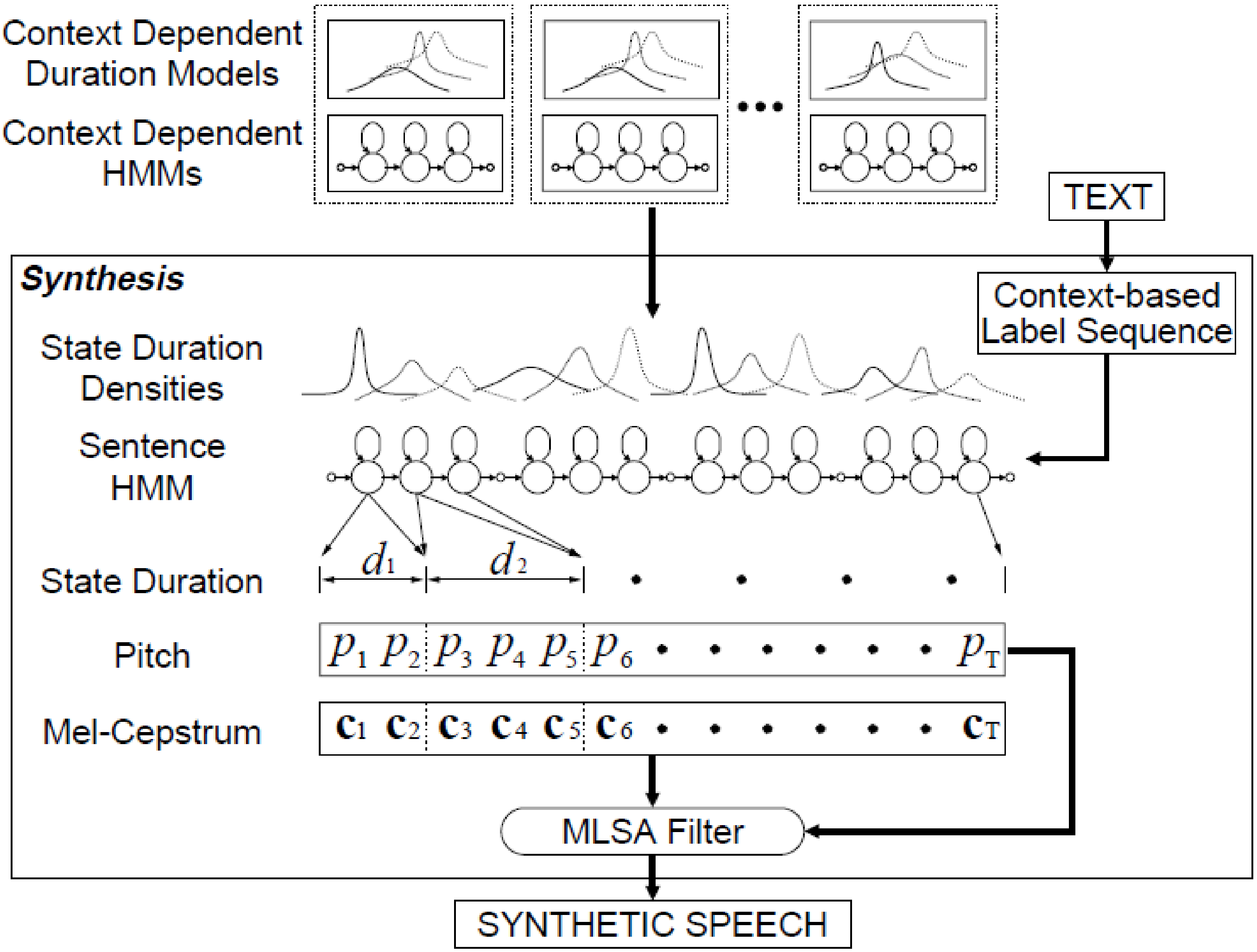

Context Dependent Duration Models
Context Dependent HMMs
TEXT
Synthesis
Context-based Label Sequence
State Duration Densities
Sentence HMM
State Duration
d_1
d_2
Pitch
p_1 p_2 p_3 p_4 p_5 p_6 • • • • • • p_T
Mel-Cepstrum
$\mathbf{c}_1$ $\mathbf{c}_2$ $\mathbf{c}_3$ $\mathbf{c}_4$ $\mathbf{c}_5$ $\mathbf{c}_6$ • • • • • • $\mathbf{c}_T$
MLSA Filter
SYNTHETIC SPEECH

Figure 3. Synthesis part of the system.