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.
-

Speech Parameter Generation Algorithms for HMM

Presented by

Najeeb

June 26, 2014

SPEECH PARAMETER GENERATION BASED ON MAXIMUM LIKELIHOOD CRITERION

- For a given continuous mixture HMM λ, we derive an algorithm for determining speech parameter vector sequence

$$O = [o_1, o_2, \ldots, o_T]'$$

- Such that

$$P(O \mid \lambda) = \sum_{all\ Q} P(O, Q \mid \lambda)$$

is maximized where

$$Q = \{(q_1, i_1), (q_2, i_2), \ldots, (q_T, i_T)\}$$

SPEECH PARAMETER GENERATION BASED ON MAXIMUM LIKELIHOOD CRITERION

- We assume that the speech parameter vector o_t consists of the static feature vector c_t and dynamic feature vectors $\Delta c_t, \Delta^2 c_t$

$$c_t = [c_t(1), c_t(2), \ldots, c_t(M)]'$$

$$\Delta c_t = \sum_{\tau=-L_-^{(1)}}^{L_+^{(1)}} w^{(1)}(\tau) c_{t+\tau} \qquad (1)$$

$$\Delta^2 c_t = \sum_{\tau=-L_-^{(2)}}^{L_+^{(2)}} w^{(2)}(\tau) c_{t+\tau} \qquad (2)$$

Case 1: Maximizing P(O|Q, λ) with respect to O

- The logarithm of P(O|Q, λ) can be written as

$$\log P(O \mid Q, \lambda) = -\frac{1}{2} O^T U^{-1} O + O^T U^{-1} M + K$$

- Where

$$U^{-1} = diag[U_{q1,i1}^{-1}, U_{q2,i2}^{-1}, \ldots, U_{qT,iT}^{-1}]$$

$$M = [\mu_{q1,i1}^T, \mu_{q2,i2}^T, \ldots, \mu_{qT,iT}^T]$$

- P(O|Q, λ) is maximized when O = M without the conditions (1), (2)

$$O = M \Rightarrow Max\ P(O \mid Q, \lambda)$$

Case 1

- Conditions (1),(2) can be arranged in a matrix form

$$O = WC \qquad (3)$$

- Where

$$C = [c_1, c_2, \ldots, c_T]^T$$

$$W = [w_1, w_2, \ldots, w_T]^T \qquad w_t = [w_t^{(1)}, w_t^{(2)}, w_t^{(3)}]$$

$$w_t^{(n)} = [0_{M\times M}, \ldots, 0_{M\times M}, w^{(n)}(-L_-^{(n)}) I_{M\times M},$$
$$\ldots, w^{(n)}(0) I_{M\times M}, \ldots, w^{(n)}(L_+^{(n)}) I_{M\times M}, \ldots,$$
$$0_{M\times M}, \ldots, 0_{M\times M}]^T, \quad \text{n=0,1,2}$$

Case 1

- Under the condition (3), maximizing P(O|Q, λ) with respect to O is equivalent to that with respect to C

$$\frac{\partial \log P(WC \mid Q, \lambda)}{\partial C} = 0$$

$$\frac{\partial \left[-\frac{1}{2} [WC]^T U^{-1} [WC] + [WC]^T U^{-1} M + K \right]}{\partial C} = 0$$

$$W^T U^{-1} W C = W^T U^{-1} M^T \quad (4)$$

Case 1

- For direct solution of (4), we need $O(T^3M^3)$ operations
- To reduce computations Cholesky decomposition or the QR decomposition can be used
- A recursive algorithm can also be used for the computation of (4)

Recursive algorithm

$$b_j(\mathbf{o}_t) = \mathcal{N}(\mathbf{c}_t;\ \boldsymbol{\mu}_j, \mathbf{U}_j) \cdot \mathcal{N}(\Delta\mathbf{c}_t;\ \Delta\boldsymbol{\mu}_j, \Delta\mathbf{U}_j)$$

- maximizing P(O|Q, λ)

$$\mathbf{R}\mathbf{c} = \mathbf{r}$$

$$\mathbf{R} = \mathbf{U}^{-1} + \mathbf{W}'\Delta\mathbf{U}^{-1}\mathbf{W}$$

$$\mathbf{r} = \mathbf{U}^{-1}\boldsymbol{\mu} + \mathbf{W}'\Delta\mathbf{U}^{-1}\Delta\boldsymbol{\mu}$$

- Replace the above equation by

$$\hat{\mathbf{R}}\hat{\mathbf{c}} = \hat{\mathbf{r}}$$

$$\text{where} \quad \hat{\mathbf{R}} = \mathbf{R} + \mathbf{w}\mathbf{D}\mathbf{w}'$$

$$\hat{\mathbf{r}} = \mathbf{r} + \mathbf{w}\mathbf{d}$$

$$\mathbf{D} = \Delta\mathbf{U}_{\hat{q}_t}^{-1} - \Delta\mathbf{U}_{q_t}^{-1}$$

$$\mathbf{d} = \Delta\mathbf{U}_{\hat{q}_t}^{-1}\Delta\boldsymbol{\mu}_{\hat{q}_t} - \Delta\mathbf{U}_{q_t}^{-1}\Delta\boldsymbol{\mu}_{q_t}$$

$$\mathbf{w} = [\underset{1}{0}, \ldots, 0, \underset{t-L}{w(-L)}, \ldots, \underset{t}{w(0)}, \ldots, \underset{t+L}{w(L)}, 0, \ldots, \underset{T}{0}]'.$$

Recursive algorithm

To replace $\{\boldsymbol{\mu}_{q_t}, \mathbf{U}_{q_t}\}$ with $\{\boldsymbol{\mu}_{\hat{q}_t}, \mathbf{U}_{\hat{q}_t}\}$

$$\mathbf{D} = \mathbf{U}_{\hat{q}_t}^{-1} - \mathbf{U}_{q_t}^{-1}$$

$$\mathbf{d} = \mathbf{U}_{\hat{q}_t}^{-1}\boldsymbol{\mu}_{\hat{q}_t} - \mathbf{U}_{q_t}^{-1}\boldsymbol{\mu}_{q_t}$$

$$\mathbf{w} = [\underset{1}{0}, \ldots, 0, \underset{t}{1}, 0, \ldots, \underset{T}{0}]'.$$

Table 1: Summary of the proposed algorithm.

- Set $\mathbf{D}$, $\mathbf{d}$, $\mathbf{w}$ by (23)–(25) to replace $\{\Delta\boldsymbol{\mu}_{q_t}, \Delta\mathbf{U}_{q_t}\}$ with $\{\Delta\boldsymbol{\mu}_{\hat{q}_t}, \Delta\mathbf{U}_{\hat{q}_t}\}$.
- Set $\mathbf{D}$, $\mathbf{d}$, $\mathbf{w}$ by (26)–(28) to replace $\{\boldsymbol{\mu}_{q_t}, \mathbf{U}_{q_t}\}$ with $\{\boldsymbol{\mu}_{\hat{q}_t}, \mathbf{U}_{\hat{q}_t}\}$.

Substitue $\hat{\mathbf{c}}$ and $\hat{\mathbf{P}}$ obtained by the previous iteration to $\mathbf{c}$ and $\mathbf{P}$, respectively, and calculate

$$
\begin{aligned}
\boldsymbol{\pi} &= \mathbf{w}'\mathbf{P} \\
\boldsymbol{\kappa} &= \mathbf{D}^{-1} + \boldsymbol{\pi}\mathbf{w} \\
\mathbf{k} &= \mathbf{P}\mathbf{w}\boldsymbol{\kappa}^{-1} \\
\hat{\mathbf{P}} &= \mathbf{P} - \mathbf{k}\boldsymbol{\pi} \\
\hat{\mathbf{c}} &= \mathbf{c} + \mathbf{k}\left(\mathbf{D}^{-1}\mathbf{d} - \mathbf{w}'\mathbf{c}\right)
\end{aligned}
$$

Recursive algorithm

- The overall procedure for parameter generation from HMMs is summarized as follows
  - Solve the set of equations Rc=r for an initial state sequence, and obtain c and P
  - Replace the state qt of a frame t with q't according to a certain strategy, and obtain c' and P' by using the algorithm
  - if the value of log P[q', O'] is not smaller than that of log P[q, O] , discard the replacement
  - Repeat 2 and 3 until a certain condition is satisfied.

Case 2: Maximizing P(O,Q|λ) with respect to O and Q

- This problem can be solved by evaluating

$$\max_C P(O,Q \mid \lambda) = \max_C P(O \mid Q,\lambda) P(Q \mid \lambda) \quad \forall\ Q$$

- However this is impractical
- To control temporal structure of speech parameter sequence appropriately, HMMs should incorporate state duration densities

$$P(O,Q \mid \lambda) = P(O,i \mid q,\lambda) P(q \mid \lambda)$$

$$\log P(q \mid \lambda) = \sum_{n=1}^{N} \log p_{qn}(d_{qn})$$

Case 2: Maximizing $P(O,Q|\lambda)$ with respect to O and Q

- If we determine the state sequence q only by $P(q|\lambda)$ independently of O, maximizing $P(O,Q|\lambda) = P(O, i|q, \lambda)P(q|\lambda)$ with respect to O and Q is equivalent to maximizing $P(O, i|q, \lambda)$ with respect to O and I
- If we assume that state output probabilities are single-Gaussian, then the solution can be obtained as in case 1

Case 3: Maximizing P(O|λ) with respect to O

- An auxiliary function of current parameter vector sequence O and new parameter vector sequence O' is defined by

$$Q(O,O') = \sum_{all\ Q} P(O,Q\,|\,\lambda)\log P(O',Q\,|\,\lambda) \qquad (5)$$

- It can be shown that by substituting O' which maximizes Q(O,O') for O, the likelihood increases unless O is a critical point of the likelihood

Case 3

- (5) can be written as

$$Q(O,O') = P(O|\lambda)\left\{-\frac{1}{2}O'^T\overline{U^{-1}}O' + O'^T\overline{U^{-1}M} + \overline{K}\right\}$$

- Where

$$\overline{U^{-1}} = diag[\overline{U_1^{-1}}, \overline{U_2^{-1}}, \ldots, \overline{U_T^{-1}}]$$

$$\overline{U_t^{-1}} = \sum_{q,i} \gamma_t(q,i) U_{q,i}^{-1}$$

$$\overline{U^{-1}M} = [\overline{U_1^{-1}\mu_1}^T, \overline{U_2^{-1}\mu_2}^T, \ldots, \overline{U_T^{-1}\mu_T}^T]$$

$$\overline{U_t^{-1}\mu_t} = \sum_{q,i} \gamma_t(q,i) U_{q,i}^{-1}\mu_{q,i}$$

$$\gamma_t(q,i) = P\left(q_t = (q,i) \mid O, \lambda\right)$$

Case 3

- Under the condition

$$O' = WC'$$

- C' which maximizes Q(O,O') is given by the following set of equations

$$W^T \overline{U^{-1}} WC' = W^T \overline{U^{-1} M} \qquad (6)$$

- The above set of equations has the same form as case 1

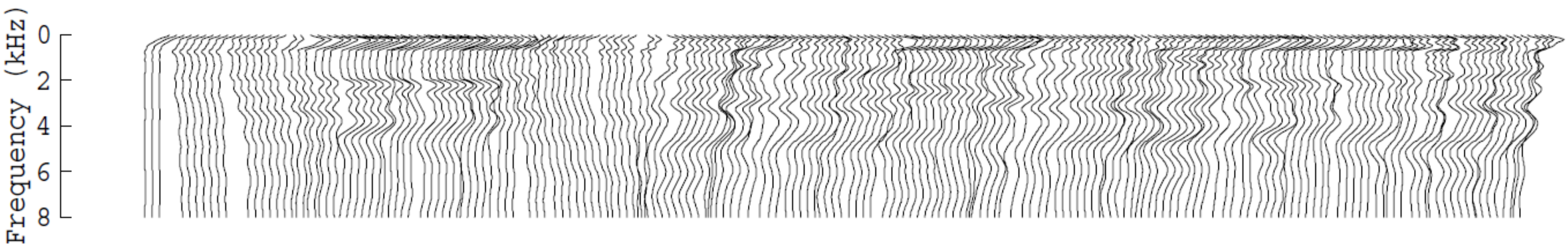
Case 3

- The whole procedure is summarized as follows:
  - Choose an initial parameter vector sequence C
  - Calculate $\Upsilon_t(q,i)$ with the forward-backward algorithm
  - Calculate U^{-1} and $U^{-1}M$ and solve (6)
  - Set C = C’ If a certain convergence condition is satisfied, stop; otherwise, go to Step 2

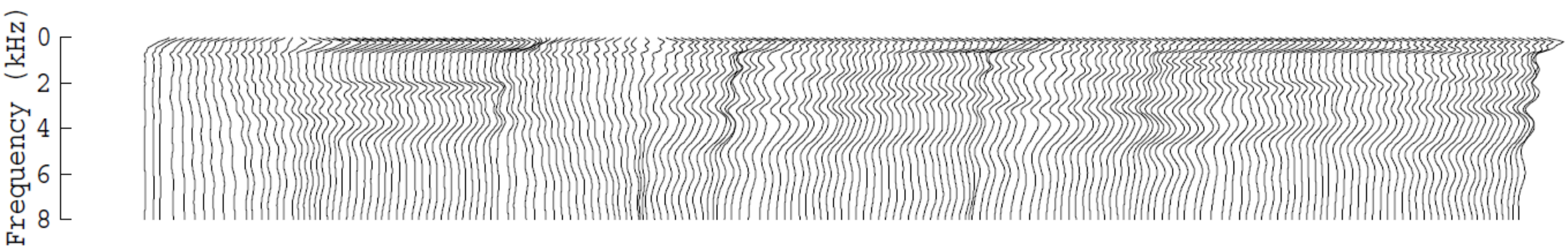
Effect of dynamic feature

- The parameter generation in the case 1, in which parameter sequence O maximizes $P(O|Q, \lambda)$ was considered
- State sequence Q was estimated from the result of Veterbi alignment of natural speech

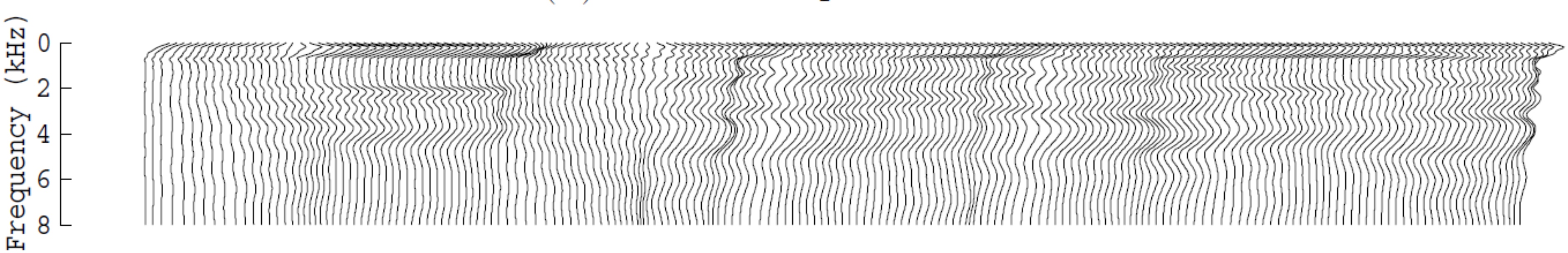
Effect of dynamic feature



(a) without dynamic features

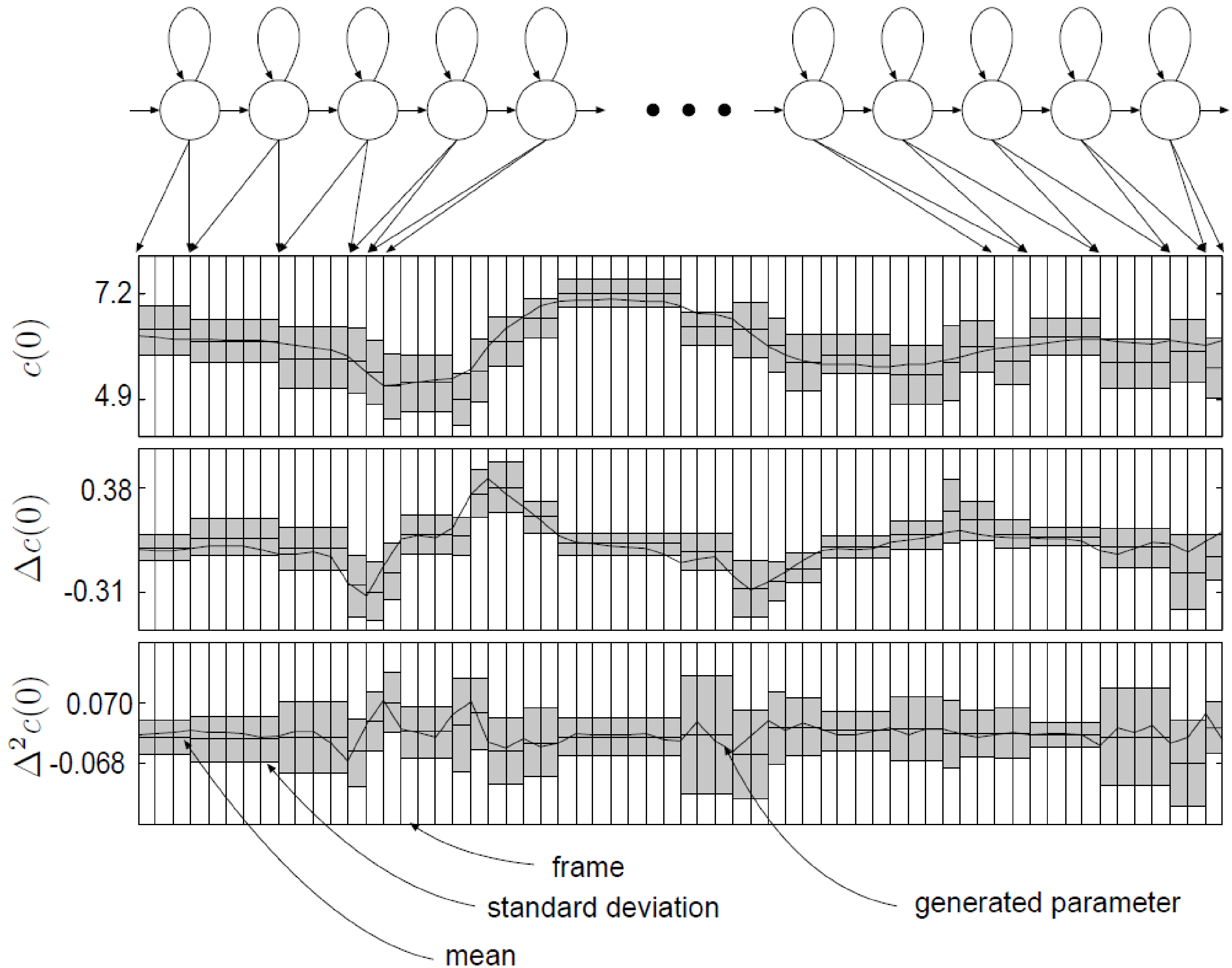


(b) with delta parameters



(c) with delta and delta-delta parameters

PDF and Generated Parameters



Parameter generation using multi-mixture HMM