

音声スペクトルの動的特徴を反映したHMMとその有効性

坪香, 英一
松下電器産業株式会社中央研究所

高木, 英行
松下電器産業株式会社中央研究所

<https://hdl.handle.net/2324/4481552>

出版情報：日本音響学会講演論文集，pp.33-34，1989-03. 日本音響学会
バージョン：
権利関係：



1-6-17 音声スペクトルの動的特徴を 反映したHMMとその有効性*

○坪香英一

高木英行

(松下電器中研)

1. まえがき

筆者らは、音声信号の時間構造を反映するために、パラメータが時変であるHMMを提案し、そのパラメータの推定方法を示した¹⁾²⁾。ここでは、過度部に特徴を持つ二三の時系列信号について、本HMMと従来のHMMとの比較を行う。

2. パラメータ時変型HMM

図1は従来の典型的なHMMである。 a_{ij} は状態 q_i から状態 q_j への遷移確率、 $b_i(\mathbf{x})$ は状態 q_i でベクトル $\mathbf{x}$ が観測される確率密度である。このモデルの問題点は、 a_{ij} および $b_i(\mathbf{x})$ を規定するパラメータが状態 q_i の連長に関わらず一定で、各状態内部の時間構造が反映されないことである。図2は本提案によるHMMで、この欠点を補うべくパラメータが状態 q_i の連長 k に応じて変化するものである。

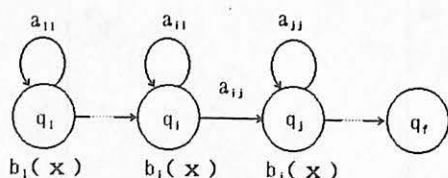


図1 従来のHMM

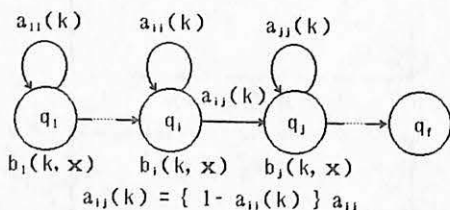


図2 パラメータ時変型HMM

3. パラメータの推定

観測ベクトル系列を $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$ 、状態 q_i の連長を τ 、その開始フレームを $t-\tau$ とし、状態 q_i が k 連続したときの観測ベクトル $\mathbf{x}_{t-\tau+1+k}$ の発生確率密度を正規分布

$$b_{ik}(\mathbf{x}_{t-\tau+1+k}) = N(\mathbf{x}_{t-\tau+1+k}, \boldsymbol{\mu}_i(\tau, k), \boldsymbol{\Sigma}_i)$$

とする。ただし、平均値 $\boldsymbol{\mu}_i(\tau, k)$ は、状態 q_i 毎に定まるベクトル $\boldsymbol{\mu}_{i0}$ 、 $\mathbf{u}_i$ に関し、

$$\boldsymbol{\mu}_i(\tau, k) = \boldsymbol{\mu}_{i0} + \left[k - \frac{\tau+1}{2} \right] \mathbf{u}_i$$

与えられ、 k に関して直線的に変化するとす

る。状態 q_i が τ 連続する確率密度 $d_i(\tau)$ は、Poisson分布や Γ 分布の場合については既に提案されており³⁾⁴⁾、本モデルにも勿論導入可能であるが、ここでは、スペクトルの時間構造のみ考慮した場合の効果調べるために、従来のHMMと同様に、 $a_{ii}(k) = \gamma_i$ (一定)として、次式で与える。

$$d_i(\tau) = \gamma_i^{\tau-1} (1 - \gamma_i).$$

以上の条件で、文献2)の方法でパラメータを推定する。 T をフレーム数、 λ をモデルとする時、

$$p(i, j, t, \tau) = P(\mathbf{x}_1, \dots, \mathbf{x}_T, s_{t-\tau} = q_i, \dots, s_{t-1} = q_i, s_t = q_j | \lambda)$$

とすれば、次式が得られる。

a) 遷移確率 a_{ij} の推定

$$\bar{a}_{ij} = \frac{\sum_{t=2}^{T+1} \sum_{\tau=1}^{t-1} p(i, j, t, \tau)}{\sum_{t=2}^{T+1} \sum_{\tau=1}^{t-1} \sum_{j=1, j \neq i} p(i, j, t, \tau)}$$

b) 遷移確率 γ_i の推定

$$\bar{\gamma}_i = \frac{\sum_{t=2}^{T+1} \sum_{\tau=1}^{t-1} p(i, i, t, \tau) (\tau - 1)}{\sum_{t=2}^{T+1} \sum_{\tau=1}^{t-1} \sum_{j=1, j \neq i} p(i, j, t, \tau) \tau}$$

c) 平均ベクトル $\boldsymbol{\mu}_{i0}$ の推定

$$\bar{\boldsymbol{\mu}}_{i0} = \frac{\sum_{t=2}^{T+1} \sum_{\tau=1}^{t-1} p(i, i, t, \tau) \sum_{k=1}^{\tau} \mathbf{x}_{t-\tau+1+k}}{\sum_{t=2}^{T+1} \sum_{\tau=1}^{t-1} \sum_{j=1, j \neq i} p(i, j, t, \tau) \tau}$$

d) 平均値の方向ベクトル $\mathbf{u}_i$ の推定

$$\bar{\mathbf{u}}_i = \frac{\sum_{t=2}^{T+1} \sum_{\tau=1}^{t-1} p(i, i, t, \tau) \sum_{k=1}^{\tau} \mathbf{f}(t, \tau, k)}{\sum_{t=2}^{T+1} \sum_{\tau=1}^{t-1} \sum_{j=1, j \neq i} p(i, j, t, \tau) \frac{(\tau-1)\tau(\tau+1)}{12}}$$

ただし

$$\mathbf{f}(t, \tau, k) = \sum_{k=1}^{\tau} \left\{ k - \frac{\tau+1}{2} \right\} \mathbf{x}_{t-\tau+1+k}$$

e) 分散共分散行列 $\boldsymbol{\Sigma}_i$ の推定

$$\bar{\boldsymbol{\Sigma}}_i = \frac{\sum_{t=2}^{T+1} \sum_{\tau=1}^{t-1} p(i, i, t, \tau) \sum_{k=1}^{\tau} \mathbf{S}(i, t, \tau, k)}{\sum_{t=2}^{T+1} \sum_{\tau=1}^{t-1} \sum_{j=1, j \neq i} p(i, j, t, \tau) \tau}$$

*"Time-Variant Parameter Hidden Markov Model", by E.TSUBOKA and H.TAKAGI, Matsushita Electric Industrial Co., Ltd.

ただし

$$S(i, t, \tau, k) = \sum_{k=1}^T \left[\mathbf{x}_{t-\tau-1+k} - \boldsymbol{\mu}_i(\tau, k) \right] \left[\mathbf{x}_{t-\tau-1+k} - \boldsymbol{\mu}_i(\tau, k) \right]^T$$

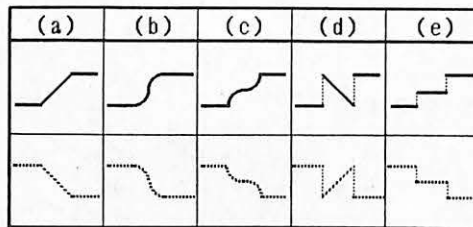
4. 実験方法および結果

図3に実験に用いたパターンを示す。各々は2次元のベクトル系列で30フレームからなる。実線は各フレームにおける平均値で、これに一樣乱数を重畳し、(a)~(e)のカテゴリ毎に50のパターンを準備した。モデルは図2において、 $I=f=4$ および $j=i+1(a_{ij}=\delta_{i+1,j})$ とした。これら50パターンで各カテゴリに対応するモデルの作成を行い、同じパターンを入力して識別実験を行った。

図4は confusion matrices である。図5はカテゴリ間の分離度を示している。入力パターンwの属すべきカテゴリに対するHMMを λ^i とする時、

$$l_1(w) = \log P(w | \lambda^i), \quad l_2(w) = \max_{j(j \neq i)} \{\log P(w | \lambda^j)\}$$

として、点 $(l_1(w), l_2(w))$ を前述の全パターンについてプロットしたものである。図5の、斜線下部の三角領域の点は正しく、上部の点は誤って識



—— 第1要素, - - - - 第2要素

図3 実験に用いたデータ

出力 \ 入力	従来型					H ² パラメータ時変型				
	(a)	(b)	(c)	(d)	(e)	(a)	(b)	(c)	(d)	(e)
(a)	13			8		50				
(b)		37	15				50			
(c)			13	23				50		
(d)	37			12	42				50	
(e)					49					50

(1) S/N ≈ 37[dB]

出力 \ 入力	従来型					H ² パラメータ時変型				
	(a)	(b)	(c)	(d)	(e)	(a)	(b)	(c)	(d)	(e)
(a)	49	49	49		49	48		16		
(b)							42	30		4
(c)			1			2	2	4		
(d)	1	1		50					50	
(e)					1		6			46

(2) S/N ≈ 11[dB]

図4 コンフュージョンマトリクス

別された場合に対応し、右下の方向に斜線から離れる程、カテゴリ間の分離度がよいと言える。

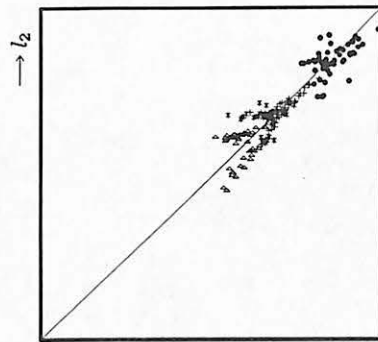
以上の実験結果から、本提案モデルは、ここで用いたような過渡部に特徴のあるパターンに対して、従来モデルに比べ、識別率、分離度の点で優れていると言える。

5. むすび

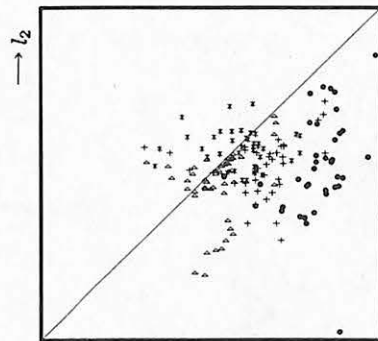
音声スペクトルの動的特徴を反映した新しいHMMとしてパラメータ時変型HMMを提案し、その可能性を確認した。今後は種々の実際の音声データについて、その振舞いを明らかにして行きたい。

参考文献

- 1) 坪香: “音声スペクトルの動的特徴を考慮したHMM”, 信学技報, SP87-97(1987).
- 2) 坪香: “時系列信号の動的特徴を反映したHMMの定式化と線形予測係数をパラメータとする場合への応用”, 信学技報, SP88-107(1988).
- 3) Russell, M.J. and Moore, R.K.: “Explicit modeling of state occupancy in hidden Markov models for automatic speech recognition”, Proceedings of ICASSP-85, pp.5-8, March 1985.
- 4) Levinson, S.E.: “Continuously variable duration hidden Markov models for automatic speech recognition”, COMPUTER SPEECH & LANGUAGE, vol.1, no.1, pp.29-45, March 1986.



(1)従来型HMM → l_1



→ l_1
(2)パラメータ時変型HMM

● : S/N ≈ 37[dB] + : S/N ≈ 23[dB]
* : S/N ≈ 17[dB] △ : S/N ≈ 11[dB]

図枠外の点は $l_1 \gg l_2$ のパターンに対応

図5 第1候補の尤度対第2候補の尤度