Vol.2013-MUS-99 No.20
2013/5/11

音声基本周波数の藤崎モデル指令列の統計的語彙モデル

石原達馬[1,a)]　吉里幸太[1,b)]　亀岡弘和[1,2,c)]　齋藤大輔[1,d)]　嵯峨山茂樹[1,3,e)]

概要：音声の基本周波数（F_0）軌跡は，話者性，感情，意図など豊富な非言語情報・パラ言語情報が含まれることが知られており，その分析は重要な課題である．我々は基本周波数軌跡の数理的なモデルの一つである，藤崎モデルのパラメータの生成過程を HMM によりモデル化することで，実測 F_0 軌跡から藤崎モデルのパラメータを推定する手法を開発してきた．本研究では，パラメータ推定精度の向上を目指して，藤崎モデルの指令列には典型的なパターン（テンプレート）が存在するという仮説に基づき，分析のための新しい HMM のトポロジーを提案する．定量評価実験により，モデルの持つテンプレート数に対する推定精度の変化を実験により確認した．

1. はじめに

音声には言語情報以外にも様々な情報が含まれており，日常的なコミュニケーションに利用される．我々はこれらの非言語的な情報を工学的に扱う枠組みを構築することを目標として，非言語情報の解析・合成のための情報処理と信号処理の研究を進めている．

音声の基本周波数 (F_0) 軌跡には，話者性，感情，意図などの非言語的な情報が豊富に含まれることが知られている．このため，F_0 軌跡のモデル化は，音声合成，話者認識，感情認識，対話システムなど，韻律情報が重要な役割を担う応用において極めて有効である．F_0 軌跡は，韻律句全体にわたってゆるやかに変化する成分 (フレーズ成分) と，アクセントに従って急峻に変化する成分 (アクセント成分) により構成される．これらはの成分は，ヒトの甲状軟骨の並進運動と回転運動にそれぞれ対応していると解釈できるが，この解釈に基づき対数 F_0 軌跡をこれらの成分の和で表した数学的なモデル (以後，藤崎モデル) が提案されている [1]．藤崎モデルは，フレーズ・アクセント指令の生起時刻，持続時間，各指令の大きさなどをパラメータとして有し，これらが適切に設定されたとき実測の軌跡を非常によく近似することが知られている．また，パラメータの言語学的対応の妥当性も広く確認されている．

[1] 東大院・情報理工
[2] NTT CS 研
[3] 国立情報学研究所
[a)] ishihara@hil.t.u-tokyo.ac.jp
[b)] yoshizato@hil.t.u-tokyo.ac.jp
[c)] kameoka@hil.t.u-tokyo.ac.jp
[d)] dsaito@hil.t.u-tokyo.ac.jp
[e)] sagayama@hil.t.u-tokyo.ac.jp

先述の藤崎モデルのパラメータは，韻律的特徴を効率よく表現できるため，実測の F_0 軌跡から藤崎モデルのパラメータを推定することは非常に重要な問題である [2]．しかしながら，この問題は元来不良設定問題であること，また藤崎モデルには言語学的な知見により守られるべき制約が存在することなどから，必ずしも容易ではなかった．これまで我々は，藤崎モデルをベースとした F_0 軌跡の確率的生成過程 [3], [4], [5] をモデル化し，藤崎モデルのパラメータ推定問題を EM アルゴリズムに基づく最尤推定問題に帰着させることに成功し，効果的なパラメータ推定アルゴリズムの開発を行ってきた．本手法の中心的なアイデアは，フレーズ・アクセント指令列の生成プロセスを隠れマルコフモデル (HMM) により表現した点にあり，HMM のトポロジーの設計や遷移確率の学習を通して，指令列に関する言語学的ないし先験的な知識を，パラメータ推定に効果的に組み込むことが可能である．一般に，適切な知識がモデルにうまく組み入れられれば，モデルパラメータの推定には有利である．例えば，HMM に基づく音声認識では，HMM のトポロジーや遷移確率を通して，言語的に自然または常識的でありつつ，観測特徴系列をなるべく良く説明するような音素系列を推定することが可能である．通常の発話においては，イントネーションは発話内容の言語的なアクセント構造に強く依存するため，いくつかの異なる内容の発話がしばしば共通のイントネーションをもつ．そこで，本稿では，指令列パターンのテンプレートの語彙仮説に基づく HMM のトポロジー設計とそれに基づく F_0 軌跡の生成モデルを定式化し，藤崎モデルパラメータ推定アルゴリズムを導出する．

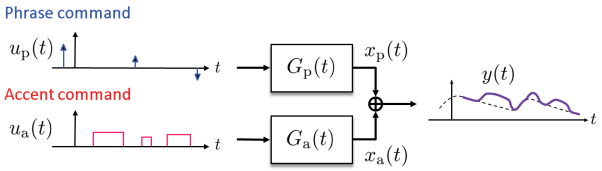

図 1 藤崎モデル [1].

2. 統計的 F_0 軌跡モデル

2.1 藤崎モデル [1]

藤崎モデル [1] では，対数 F_0 軌跡 $y(t)$ が以下のように 3 つの成分の和で表されると仮定する．

$$y(t) = x_p(t) + x_a(t) + x_b. \tag{1}$$

ここで，t は時間，$x_p(t)$ はフレーズ成分，$x_a(t)$ はアクセント成分，x_b はベースライン成分と呼ばれる，時間によらない定数である．さらにフレーズ成分，アクセント成分はそれぞれ，フレーズ指令，アクセント指令と呼ばれる信号の 2 次のフィルタの出力であると仮定される．

$$x_p(t) = G_p(t) * u_p(t) \tag{2}$$

$$G_p(t) = \begin{cases} \alpha^2 t e^{-\alpha t} & (t \geq 0) \\ 0 & (t < 0) \end{cases} \tag{3}$$

$$x_a(t) = G_a(t) * u_a(t) \tag{4}$$

$$G_a(t) = \begin{cases} \beta^2 t e^{-\beta t} & (t \geq 0) \\ 0 & (t < 0) \end{cases} \tag{5}$$

ここで $u_p(t)$ はフレーズ指令と呼ばれるデルタ列であり，$u_a(t)$ はアクセント指令と呼ばれる矩形パルス列である．これらのうち非ゼロの値をとるのは各時刻で高々 1 つである．α，β はそれぞれ 2 次フィルタの応答の速さを表す角周波数であり，個人や発話によらずおおよそ $\alpha = 3$ rad/s，$\beta = 20$ rad/s 程度の値をとることが知られている．

2.2 F_0 軌跡の確率的生成過程モデル

ここでは，これまで我々が開発してきた，藤崎モデルをベースにした F_0 軌跡の生成過程の確率モデル [3], [4], [5] について概説する．上述の藤崎モデルにおいて，フレーズ指令，アクセント指令はそれぞれデルタ列，矩形パルス列であり，さらにこれらは互いに重ならないという仮定が置かれる．我々はこれらの制約を満たすような指令列をうまく確率モデルの形として記述するために，フレーズ指令 $u_p[k]$，アクセント指令 $u_a[k]$ のペア $o[k] = (u_p[k], u_a[k])^\mathsf{T}$ を，HMM の出力として表現するモデルを考案した．各状態の出力分布を正規分布とした場合，出力系列 $\{o[k]\}_{k=1}^K$ は

$$o[k] \sim \mathcal{N}\left(o[k]; c_{s_k}, \Upsilon_{s_k}\right) \tag{6}$$

に従う．ここで s_k は時刻 k における状態を表す．すなわち，式 (6) は平均 $\mu[k] = (\mu_p[k], \mu_a[k])^\mathsf{T} = c_{s_k}$ と分散 $\Sigma[k] = \Upsilon_{s_k} = \mathrm{diag}(v_{p,k}, v_{a,k})$ が状態遷移の結果として時間とともに変化することを意味する．以上の HMM の構成は以下となる．

出力系列: $\{o[k]\}_{k=1}^K$
状態系列: $\{s_k\}_{k=1}^K$
出力確率分布: $P(o[k]|s_k) = \mathcal{N}(o[k]; c_{s_k}, \Upsilon_{s_k})$
平均値の系列: $\mu[k] = (\mu_p[k], \mu_a[k])^\mathsf{T} = c_{s_k}$
遷移確率: $\phi_{i',i} = \log P(s_k = i | s_{k-1} = i')$

上記の HMM から出力された指令関数 $u_p[k]$, $u_a[k]$ にそれぞれ異なるフィルタ $G_p[k]$ と $G_a[k]$ が畳み込まれたものがフレーズ成分とアクセント成分

$$x_p[k] = u_p[k] * G_p[k] \tag{7}$$

$$x_a[k] = u_a[k] * G_a[k] \tag{8}$$

となる．ただし，$*$ は離散時間 k に関する畳込みを表す．また，$G_p[k]$ と $G_a[k]$ はそれぞれ $G_p(t)$ と $G_a(t)$ を離散時間表現である．以上より，F_0 軌跡の離散時間表現 $x[k]$ は

$$x[k] = x_p[k] + x_a[k] + x_b \tag{9}$$

となる．x_b はベースライン成分を表す．

無声区間においては F_0 は観測されないことがあったり，観測されていたとしても信頼できない場合が多い．また，F_0 抽出において推定誤りが生じる場合もある．そこで観測 F_0 軌跡 $y[k]$ を，上述の F_0 軌跡モデル $x[k]$ とノイズ $x_n[k] \sim \mathcal{N}\left(0, v_n^2[k]\right)$ との和として表すことで，観測 F_0 系列の不確実性を分散 $v_n^2[k]$ の設定を通して組み込むことができる．よって，観測 F_0 系列 $y[k]$ は

$$y[k] = x[k] + x_n[k] \tag{10}$$

と表される．ここで，$x_n[k]$ を周辺化すると，$o = \{o[k]\}_{k=1}^K$ が与えられたもとでの $y = \{y[k]\}_{k=1}^K$ の条件つき確率密度関数 $P(y|o)$ は

$$P(y|o) = \prod_{k=1}^{K} \mathcal{N}(y[k]; x[k], v_n^2[k])$$
$$x[k] = G_p[k] * u_p[k] + G_a[k] * u_a[k] + u_b \tag{11}$$

となる．(6) より，状態系列 $s = \{s_k\}_{k=1}^K$ が与えられたもとでの $\{o[k]\}_{k=1}^K$ の条件つき確率密度関数 $P(o|s,\theta)$ は $P(o|s,\theta) = \prod_{k=1}^K \mathcal{N}(o[k]; c_{s_k}[k], \Upsilon_{s_k})$ で与えられる．ここで，θ は出力分布の平均と分散の系列を表す．状態系列 s の確率分布 $P(s)$ は HMM におけるマルコフ性の仮定より，遷移確率の積 $P(s) = \phi_{s_1} \prod_{k=2}^K \phi_{s_k, s_{k-1}}$ で与えられる．

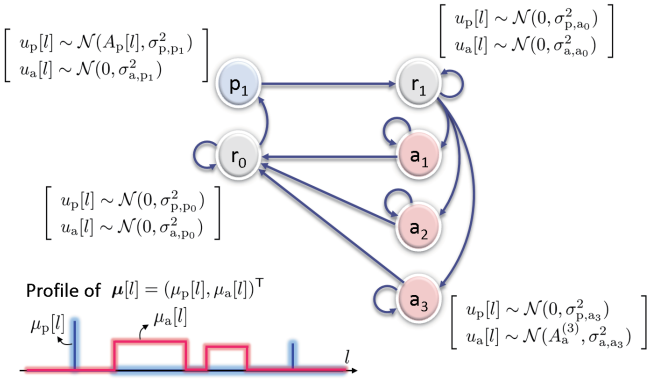

図 2 従来法におけるフレーズ・アクセント指令列の状態遷移モデル [3], [4], [5]. 状態 r_0 において $\mu_{\mathrm{p}}[k]$ と $\mu_{\mathrm{a}}[k]$ はゼロである．状態 p_1 において $\mu_{\mathrm{p}}[k]$ は非負値 $A_{\mathrm{p}}[k]$ をとることができ，$\mu_{\mathrm{a}}[k]$ はゼロである．状態 p_1 において自己遷移は禁止される．状態 r_1 において $\mu_{\mathrm{p}}[k]$ と $\mu_{\mathrm{a}}[k]$ はまたゼロのみに制限される．この状態は $\mu_{\mathrm{p}}[k]$ がパルス列になることを保証するものである．状態 r_0 は状態 $\mathrm{a}_1, \ldots, \mathrm{a}_N$ へのみ遷移することができ，これらの状態において $\mu_{\mathrm{a}}[k]$ はそれぞれ異なる値 $A^{(n)}_{\mathrm{a}}$ をとることができるが，$\mu_{\mathrm{p}}[k]$ はゼロに制限される．直接 a_n から $\mathrm{a}_{n'}$ へを通らずに r_1 遷移することは禁止される．これは $\mu_{\mathrm{a}}[k]$ が矩形パルス列であることを保証するためのものである．

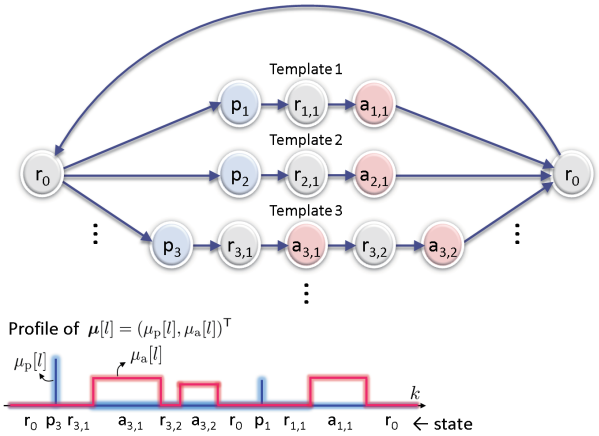

図 3 ピッチパターンテンプレートの語彙モデルに基づくフレーズ・アクセント指令列の状態遷移トポロジー

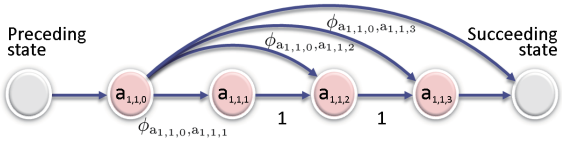

図 4 状態 $\mathrm{a}_{1,1}$ を 4 つの小状態 $\mathrm{a}_{1,1,0}$，$\mathrm{a}_{1,1,1}$，$\mathrm{a}_{1,1,2}$，$\mathrm{a}_{1,1,3}$ へ分割したもの．遷移確率 $\phi_{\mathrm{a}_{1,1,0},\mathrm{a}_{1,1,1}}$ は状態 $\mathrm{a}_{1,1}$ が 4 回持続する確率に対応する．

3. 提案モデル

3.1 フレーズ・アクセント指令の語彙モデル

以上のモデルにおいて重要なアイディアは，藤崎モデルの制約が HMM の状態遷移トポロジーで表現される点にあるが，これまでの状態遷移トポロジーのもとでは藤崎モデルにおける制約を満たす範囲のいかなる指令列も生成しえて，言語学的に必ずしも妥当でない指令列を生成することを許容していた．もし指令列のとりうる範囲を言語学的な先験知識に基づいて適切に制限できれば，提案モデルを用いて効果的に藤崎モデルパラメータの推定を行えるようになるはずである．以上より，言語的な先験的知識を HMM の状態遷移トポロジーの設計を通してモデルに組み込もうというのが本研究のアイデアの要点である．

通常の発話では，イントネーション型の種類は限られている．日本語の場合，ピッチアクセントは高いと低いの 2 値で表され，1 アクセント句に含まれるモーラ数には限りがあるためである．例えば，「あらゆる現実を」と「明日は輪講だ」のアクセントパターンは同一であるため，イントネーションはほとんど同一である．このことは，藤崎モデルの指令列ペアが，有限種類のテンプレートをつなぎ合わせて表現できる可能性があることを示唆する．そこでまず指令列のテンプレートに対応する有限種類の Left-to-Right HMM を考え，テンプレート間を遷移可能な HMM を考えることにより，上述のようなテンプレートベースの指令列の生成モデルを立てることができそうである．このような状態遷移トポロジーは例えば Fig. 3 のような HMM で表すことができる．この HMM をフレーズ・アクセント指令列の語彙モデルと呼ぶこととする．ここで，各テンプレートの時間伸縮をどれだけ許容するかを柔軟に扱えるようにする目的で，Fig. 4 に示すように各状態 (ただし p_1, p_2, ... を除く) を，同一な出力分布を有するよう拘束された小状態に分割することとした．これにより，各状態での停留時間を個別にパラメトライズすることが可能である．

以上の提案モデルに基づく藤崎モデルパラメータ（フレーズ・アクセント指令）推定法は，学習ステージと認識ステージの 2 つのステージにより構成される．学習ステージは，状態出力分布の平均と遷移確率を F_0 軌跡の学習データから推定するステージであり，学習データから指令列のテンプレートを学習することに相当する．認識ステージは，学習ステップで学習された状態遷移確率と状態出力分布を固定のもとで，最適な状態遷移系列（すなわち指令列推定値）を推定するステージである．どちらのステップも次節で説明する最適化アルゴリズムに基づくが，両ステージ間の唯一の違いは，HMM のパラメータ (遷移確率と出力分布) が固定されるかどうかという点のみである．

3.2 最適化アルゴリズム

この節では観測 F_0 系列 y が与えられたもとで，モデルパラメータ θ と o の事後確率 $P(o, \theta|y)$ の局所最適解を求める反復アルゴリズムを EM アルゴリズムと補助関数法に基づいて導出する．状態系列 s を隠れ変数とし，事後確率 $P(o, \theta|y)$ が $P(o, \theta, s|y) \propto P(y|o)P(o|s, \theta)P(s)$ を s につ

いて周辺化することで得られる点に注意すると，Q 関数 $Q(o,\theta,o',\theta')$ は

$$\begin{aligned}Q(o,\theta,o',\theta') &= \sum_s P(s|y,o',\theta')\log P(o,\theta,s|y)\\ &\stackrel{c}{=} \log P(y|o)\\ &\quad + \sum_s P(s|y,o',\theta')\log P(o|s,\theta)P(s), \quad (12)\\ \log P(y|o) &= \log\prod_{k=1}^{K}\mathcal{N}(y[k];x[k],v_{\mathrm{n}}^2[k]),\\ &= \sum_{k=1}^{K}\frac{-A[k]}{2v_{\mathrm{n}}^2[k]}, \quad (13)\\ A[k] &= \left(y[k]-\sum_{i\in\{\mathrm{p,a,b}\}}\sum_{l=1}^{K}G_i[k-l]u_i[l]\right)^2, \quad (14)\\ &\sum_s P(s|y,o',\theta')\log P(\theta,o|s)\\ &= \sum_s P(s|y,o',\theta')\sum_{i\in\{\mathrm{p,a}\}}\sum_{k=1}^{K}\frac{-B_i[k]}{2v_{i,s_k}^2}\\ &= \sum_{k=1}^{K}\sum_t P(s_k=t|y,o',\theta')\sum_{i\in\{\mathrm{p,a}\}}\frac{-B_i[k]}{2v_{i,t}^2} \quad (15)\\ B_i[k] &= \left(u_i[k]-\mu_{i,t}[k]\right)^2, \quad (16)\end{aligned}$$

と置ける．ここで，$\stackrel{c}{=}$ は定数項を除いて等しいことを表す．よって，$P(s|y,o',\theta')$ を Forward-Backward アルゴリズムにより計算するステップ，o と θ について $Q(o,\theta,o',\theta')$ を増加させるステップを繰り返すことで，$P(o,\theta|y)$ が局所最大となる解を得ることができる．o は藤崎モデルの指令関数のペアであるため，$Q(o,\theta,o',\theta')$ を増加させるステップにおいては，o の非負制約を考慮する必要がある．o の非負制約を満たしながら $Q(o,\theta,o',\theta')$ を増加させるような更新則は [6] と同様の考え方により導くことができる．[6] より，$Q(o,\theta,o',\theta')$ の下界が，Jensen の不等式

$$-\Big(\sum_{i\in\{\mathrm{p,a,b}\}}\sum_l G_i[k-l]u_i[l]\Big)^2 \geq -\sum_{i\in\{\mathrm{p,a,b}\}}\sum_l\frac{G_i^2[k-l]u_i^2[l]}{\lambda_{i,k,l}}, \quad (17)$$

を用いて設計することができる．ここで，$G_{\mathrm{b}}[k]=\delta[k]$ (クロネッカーのデルタ) である．また，$\lambda_{i,k,l}$ は，$0<\lambda_{i,k,l}<1$，$\sum_i\sum_l\lambda_{i,k,l}=1$ を満たす任意の変数である．以上をまとめると Q 関数の下界は，

$$\begin{aligned}Q(o,\theta,o',\theta') &\geq \sum_{k=1}^{K}\frac{-A'[k]}{2v_{\mathrm{n}}^2[k]}\\ &\quad + \sum_{k=1}^{K}\sum_t P(s_k=t|y,o',\theta')\sum_{i\in\{\mathrm{p,a}\}}\frac{-B_i[k]}{2v_{i,t}^2} \quad (18)\\ A'[k] &= y[k]^2\\ &\quad - 2y[k]\sum_{i\in\{\mathrm{p,a,b}\}}\sum_{l=1}^{K}G_i[k-l]u_i[l]\\ &\quad + \sum_{i\in\{\mathrm{p,a,b}\}}\sum_l\frac{G_i^2[k-l]u_i^2[l]}{\lambda_{i,k,l}}, \quad (19)\end{aligned}$$

と表される．この下界関数を $\lambda_{i,k,l}\geq 0$ に関して最大化するステップと o に関して最大化するステップを交互に繰り返せば $Q(o,\theta,o',\theta')$ を増加させることができる．いずれのステップの更新則も解析的に求めることができ，それぞれ

$$\lambda_{i,k,l}=\frac{G_i[k-l]u_i[l]}{\sum_{i\in\{\mathrm{p,a,b}\}}\sum_l G_i[k-l]u_i[l]}, \quad (20)$$

と

$$u_i[l]=\frac{\sum_t\frac{P(s_l=t|y,o',\theta')\mu_{i,t}[l]}{v_{i,t}^2}+\sum_{k=1}^{K}\frac{y[k]G_i[k-l]}{v_{\mathrm{n}}^2[k]}}{\sum_t\frac{P(s_l=t|y,o',\theta')}{v_{i,t}^2}+\sum_{k=1}^{K}\frac{G_i^2[k-l]}{v_{\mathrm{n}}^2[k]\lambda_{i,k,l}}}, \quad (21)$$

で表される．以上の反復が収束したあと，続けて θ を更新する．更新式は解析的に求めることができ，

$$\mu_{\mathrm{p}}[k]=\frac{\sum_{k=1}^{K}\sum_{t\in\mathrm{p}_n}\frac{P(s_k=t|y,o',\theta')u_{\mathrm{p}}[k]}{v_{\mathrm{p},t}^2}}{\sum_{k=1}^{K}\sum_{t\in\mathrm{p}_n}\frac{P(s_k=t|y,o',\theta')}{v_{\mathrm{p},t}^2}}, \quad (22)$$

$$\mu_{\mathrm{a}}^{(n)}=\frac{\sum_{k=1}^{K}\sum_{t\in\mathrm{a}_n}\frac{P(s_k=t|y,o',\theta')u_{\mathrm{a}}[k]}{v_{\mathrm{a},t}^2}}{\sum_{k=1}^{K}\sum_{t\in\mathrm{a}_n}\frac{P(s_k=t|y,o',\theta')}{v_{\mathrm{a},t}^2}} \quad (23)$$

と書ける．これらの更新値を改めて o' と θ' に代入して次の反復に進む．

以上の反復アルゴリズムが収束した後，Viterbi アルゴリズムにより求まる最適な s を指令列推定とする．

4. 評価実験

指令列の推定精度は語彙モデルの大きさ (イントネーションテンプレートの個数) に依存するため，異なる大きさの語彙モデルに対して指令列の推定精度の評価を行った．

学習ステージでは ATR 音素バランス文 [7] の男性話者 (MHT) によって読まれた最初の 50 文から [8] の手法を用いて抽出された F_0 軌跡を学習データとして提案モデルのパラメータが推定された．定数パラメータは以下のようにセットした．$t_0=8$ ms，$\alpha=3.0$ rad/s，$\beta=20.0$ rad/s，$v_{\mathrm{p}}^2[k]=3^2$，$v_{\mathrm{a}}^2[k]=0.03^2$，$v_{\mathrm{b}}^2=10^{-8}$，有声区間において $v_{\mathrm{n}}^2[k]=10^{15}$，無声区間において $v_{\mathrm{n}}^2[k]=0.1^2$ ．μ_{b} は全 $\log F_0$ の有声区間の値の最低値にセットされた．

認識ステージでは，学習ステージで用いたものと同一の

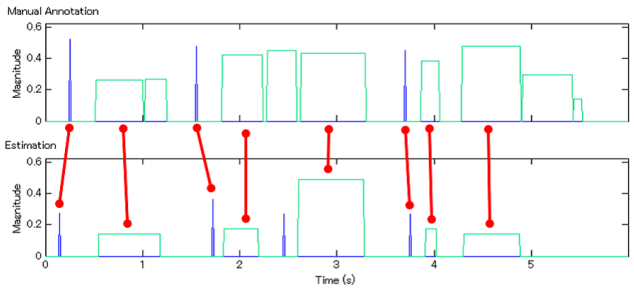

図 5 指令列のマッチングの例.

表 1 指令列の推定精度 (S=0.3s). 左，中央，右の列はそれぞれフレーズ指令とアクセント指令，フレーズ指令のみ，アクセント指令のみの推定精度を表す.「初期値」 の行は EM アルゴリズムの初期値（従って成澤らの手法によるもの）の 推定精度,「$T = N$」の行はそれぞれ提案アルゴリズムにおいてテンプレート数を N としたときの推定精度,「従来法」の行は我々の語彙モデルを組み込まない確率モデルによる手法 *[4], [5]* の推定精度である．$T = 5$ と $T = 10$ において全体の推定精度が改善しており，すべての場合においてフレーズ指令の推定精度が向上している.

	全指令	フレーズ指令	アクセント指令
初期値 [2]	65.9%	67.4%	65.1%
$T = 5$	**70.2**%	**70.3**%	70.0%
$T = 10$	**70.0**%	**71.0**%	69.5%
$T = 15$	67.3%	**72.5**%	64.7%
従来法 [4], [5]	69.3%	63.8%	72.1%

話者が読み上げた ATR 音素バランス文の最後の 53 文から，出力分布のパラメータを固定して指令列の推定を行い，結果に対して推定精度を評価した．テンプレート数は 5, 10, 15 個の場合についてそれぞれ評価された．Θ の初期値は [2] の手法を用いて行われた.

評価の枠組みは我々の従来の研究 [4], [5] と同じである．指令列の推定精度を評価するため，Fig. 5 にあるような動的計画法に基づく指令ごとのマッチングにより正解データとのマッチングを計算し，評価に利用した．フレーズ指令においては，位置の差が S 以下である指令をマッチしたと定義した．アクセント指令においては，指令の開始位置の差の大きさと指令の終了位置の差の大きさの平均が S 以下である 1 組の指令をマッチしたと定義した．指令の大きさの情報はマッチングに用いなかった．これは，指令の大きさはベースライン成分の値に影響を受けるが，その設定方法が提案手法と正解データとで異なるためである．N_{E} と N_{A} をそれぞれ提案法によって推定された藤崎モデル指令数と正解データの指令列数であるとし，N_{M} を推定指令列と正解指令列との間でマッチした指令数であるとする．N_{Esum}，N_{Asum} および N_{Msum} をそれぞれ N_{E}，N_{A}，N_{M} の全文に渡る総和であるとする．これらの値から，挿入エラー E_{I} を $(N_{\mathrm{Esum}} - N_{\mathrm{Msum}})/N_{\mathrm{Asum}}$ で定義し，脱落エラー E_{D} を $(N_{\mathrm{Asum}} - N_{\mathrm{Msum}})/N_{\mathrm{Asum}}$ で定義し，精度 A を $1 - E_{\mathrm{I}} - E_{\mathrm{D}}$ で定義した．Tab. 1 に，$S = 0.3$ s における評価結果を示す．提案手法はフレーズ指令の位置推定にアクセント指令の位置や大きさ，個数の情報を利用できるため，フレーズ指令の検出精度に大きな改善が見られたと考えられる．一方過学習によりアクセント指令の推定精度は語彙モデルの増大により減少する傾向にあることがわかった.

5. 結論

本稿では藤崎モデルパラメータ (フレーズ・アクセント指令列) を実際の F_0 軌跡から推定する問題に対して，F_0 軌跡の確率的生成モデルにピッチパターンテンプレートの語彙モデルを組み込むことで扱う手法を提案した．また確率モデルの補助関数法による最適化アルゴリズムを導出した．実験により，提案手法の異なるテンプレート数の語彙モデルに対しての推定精度を評価した．結果，指令列の解空間を制限することで指令列の検出精度がわずかに改善し，特にフレーズ指令については検出精度が向上することを実験により確かめた.

6. 謝辞

本研究の実験では，東京大学の広瀬啓吉教授が ATR 音声データベースの音声データに対し手動で付与した藤崎モデルのパラメータを正解データとして用いた．本データを提供して頂いた同氏に感謝する.

参考文献

[1] H. Fujisaki, *In Vocal Physiology: Voice Production, Mechanisms and Functions*, Raven Press, 1988.
[2] S. Narusawa, N. Minematsu, K. Hirose, and H. Fujisaki, "A method for automatic extraction of model parameters from fundamental frequency contours of speech," in *Proc. ICASSP*, 2002, pp. 509–512.
[3] H. Kameoka, J. L. Roux, and Y. Ohishi, "A statistical model of speech F_0 contours," in *Proc. SAPA*, 2010, pp. 43–48.
[4] K. Yoshizato, H. Kameoka, D. Saito, and S. Sagayama, "Statistical approach to fujisaki-model parameter estimation from speech signals and its quantitative evaluation," in *Proc. Speech Prosody 2012*, 2012, pp. 175–178.
[5] K. Yoshizato, H. Kameoka, D. Saito, and S. Sagayama, "Hidden Markov convolutive mixture model for pitch contour analysis of speech," in Proc. The 13th Annual Conference of the International Speech Communication Association (Interspeech 2012), Sep. 2012.
[6] H. Kameoka, T. Nakatani, and T. Yoshioka, "Robust speech dereverberation based on non-negativity and sparse nature of speech spectrograms," in *Proc. ICASSP*, 2009, pp. 45–48.
[7] A. Kurematsu, K. Takeda, Y. Sagisaka, S. Katagiri, H. Kuwabara, and K. Shikano, "ATR japanese speech database as a tool of speech recognition and synthesis," *Speech Communication*, vol. 27, pp. 187–207, 1999.
[8] H. Kameoka, "Statistical speech spectrum model incorporating all-pole vocal tract model and F_0 contour generating process model," in *Tech. Rep. IEICE*, 2010, in Japanese.