

Convolutional HMMに基づく歌声の基本周波数制御モデルの提案とそのパラメータ最尤推定*

大石 康智 (名大), 亀岡 弘和, 柏野 邦夫 (NTT CS 研), 武田 一哉 (名大)

1 はじめに

歌声の基本周波数 (F0) 軌跡は, 階段状の旋律概形に様々な動的変動成分が複雑に重ね合わさった状態で観測され, この変動成分は, 歌声の個人性知覚に影響を与えるという知見が報告されている [1]. つまり, 旋律に相当する譜面情報に加え, 譜面には載っていない歌唱者の個性や感情のような非譜面情報が, F0 軌跡に含まれていると考えられる. 本研究では, 文献 [2] の F0 制御モデルを参考にして, F0 軌跡から旋律概形を表す階段状の入力信号と動的変動因子を表すインパルス応答を同時推定 (分離) することを考える. ただし, 譜面情報無し状態で, 観測される F0 軌跡だけから, それぞれの信号を推定する逆問題は, そもそも不良設定問題であるため, 従来のモデルだけではこの問題を解くことができない. そこで, 我々は階段状の拘束をもつ特殊な入力信号を隠れマルコフモデル (HMM) でモデル化し, システムの伝達関数を全極モデルで表現することによる制約条件を設けた上で, Viterbi 学習と線形予測分析 (LPC) 的な解法の反復により, モデルパラメータを効率的に推定するアルゴリズムを提案する. その定式化と実装を行い, F0 軌跡から旋律概形と動的変動成分をともに推定できること, さらに推定されたモデルパラメータによって, F0 軌跡を生成可能であることを確認する.

歌声の F0 軌跡から旋律概形と動的変動成分を分離する技術は様々な応用に応用できる. 例えば, 歌声素片の連結や HMM を利用した歌声合成手法 [3, 4] によって, 表情豊かな歌声合成音が実現されているが, 知覚に影響を与える動的変動成分を精緻にモデル化する提案手法によって, さらに人間らしい合成音に近づけると考えられる. その他, 歌唱スタイルの転写や歌唱力評価などへの応用が考えられる.

2 歌声の F0 制御モデルの提案

提案する歌声の F0 制御モデルは, Fig. 1 に示すような 3 つのアイデアに基づくモデルである.

アイデア 1 旋律概形を表す階段状の入力信号は, HMM によってモデル化される.

アイデア 2 F0 制御モデルは, 全極モデルで表される伝達関数 (式 (1)) によってモデル化される.

アイデア 3 各モデルパラメータは, 最尤推定に基づく以下の 2 つのステップの反復推定によって最適化される.

ステップ 1 状態系列と HMM のパラメータは Viterbi 学習によって推定される.

ステップ 2 インパルス応答 a_0, \dots, a_M は, LPC 的な解法によって推定される.

以下の節で各アイデアの詳細を述べる.

2.1 動的変動因子を表すインパルス応答のモデル化

F0 制御モデルの動的特性を表すインパルス応答は, 以下の伝達関数によってモデル化される.

$$\frac{Y(z)}{U(z)} = \frac{1}{a_0 + \sum_{j=1}^M a_j z^{-j}} \quad (1)$$

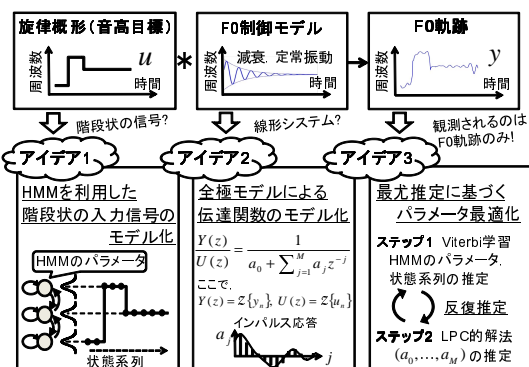


Fig. 1 提案する歌声の F0 制御モデルの概略図

ここで, 階段状の入力信号 u_n と観測される F0 軌跡 y_n の z 変換を $U(z)$, $Y(z)$, 伝達関数の次数を M と表す. 音声分析の用途においては, 零平均の Gauss 性雑音の入力信号を仮定し, フィルタと入力スケールの任意性を除く目的で通常 $a_0 = 1$ と置くが, これに対し, ここで我々が考えたい F0 制御モデルには, 動的変動成分のスケールを入力音高信号に依らず自由に調節できる機構が備わっていることが望ましい. そこで, a_0 をパラメータ化することで, ビブラート等の動的変動成分のスケールを入力信号とは独立に調節できるようにする. すなわち, 音声分析の用途では通常定数として扱われる a_0 を, 自由パラメータとして扱うことがここでは本質的に重要な意味をもつ. 式 (1) を逆 z 変換すると,

$$a_0 y_n + \sum_{j=1}^M a_j y_{n-j} = u_n \quad (2)$$

となり, 入出力関係を表す差分方程式が導かれる. いま, F0 軌跡と入力信号の関係はこの差分方程式に従うと想定するが, 実際にはこれにさらに微細変動成分 [2] が加わる. この微細変動成分を平均 0, 分散 σ^2 の正規分布に従う Gauss 性白色雑音 ε とし, $\hat{u}_n = \sum_{j=0}^M a_j y_{n-j}$ と u_n との間に,

$$\hat{u}_n - u_n = \varepsilon, \quad (\varepsilon \sim \mathcal{N}(0, \sigma^2)) \quad (3)$$

を仮定する.

2.2 階段状の入力信号のモデル化

階段状の拘束をもつ入力信号 u_n は, Fig. 2 に示す状態集合 $S = \{S_1, \dots, S_I\}$ からなる HMM によってモデル化される. 1 回の状態遷移により実数が 1 個生成されるモデルであり, 観測される F0 軌跡から状態遷移 (どの音高間でどのタイミングで遷移したか) を一意に決定できない. つまり,

$$u_n = m_{q_n}, \quad (m_{S_i} \in \mathbb{R}, q_n \in \{S_1, \dots, S_I\}) \quad (4)$$

と表し, ここで m_{S_i} は状態 S_i における出力確率分布 (式 (3) で仮定する正規分布に相当する.) の平均を表す. 一様な Markov 連鎖を想定し, 状態 S_j から S_i への遷移確率は $\mathbb{P}(S_i|S_j)$ と表す. q_n は状態集合 S の要素 (HMM の状態番号) の中のいずれかの値をとる.

* Maximum Likelihood Parameter Estimation of F0 Control Model for Singing Voices based on Convolutional HMM. by Yasunori OHISHI, Hirokazu KAMEOKA, Kunio KASHINO, Kazuya TAKEDA

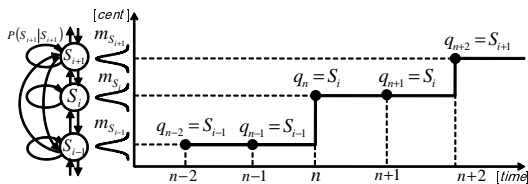


Fig. 2 階段状の入力信号のモデル化

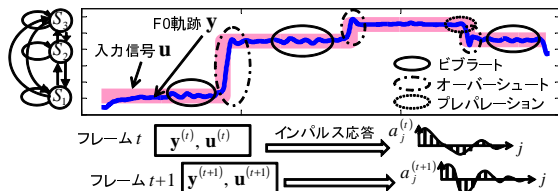


Fig. 3 時変な F0 制御モデルへの拡張

2.3 各モデルパラメータの解釈

インパルス応答 a_0, \dots, a_M は、歌声の動的変動因子を表すパラメータである。音高が安定するときの振動や、音高目標に到るまでの連続的な音高遷移が表現される。状態系列 q_1, \dots, q_N は、音長を決定するパラメータである。これは、必ずしも譜面に記される音符の音長に対応するわけではなく、歌唱者の意図やスタイルに基づいて生成される運動指令の長さを表現するものと想定する。最後に、HMMのパラメータ m_{S_1}, \dots, m_{S_I} は、歌唱者が意図する旋律の音高（音高目標値）に対応するパラメータである。これは、必ずしも譜面に記される音符の音高（客観的に定まっている音高値）には対応しない。

2.4 時変な F0 制御モデルへの拡張

Fig. 3 より、旋律の変化とともに、ビブラートやオーバーシュートのような動的変動成分が次々と変化することがわかる。これは、図 1 の F0 制御モデルのインパルス応答が、時々刻々と変化するためであると考えられる。そこで、長さ N の F0 軌跡 $\mathbf{y} = \{y_1, \dots, y_N\}$ と入力信号 $\mathbf{u} = \{m_{q_1}, \dots, m_{q_N}\}$ をフレームに分割し、フレームごとにインパルス応答を推定する。長さ L のフレーム t における F0 軌跡と入力信号をそれぞれ $\mathbf{y}^{(t)} = \{y_1^{(t)}, \dots, y_L^{(t)}\}$, $\mathbf{u}^{(t)} = \{m_{q_1^{(t)}}, \dots, m_{q_L^{(t)}}\}$, インパルス応答を $\theta^{(t)} = \{a_0^{(t)}, \dots, a_M^{(t)}\}$ と表す。ここで入力信号を生成する HMM のパラメータ m_{S_1}, \dots, m_{S_I} はすべてのフレーム（フレームの総数は T ）にわたって共通するものとする。この状態系列と HMM のパラメータを合わせて、 $\omega = \{q_1, \dots, q_N, m_{S_1}, \dots, m_{S_I}\}$ と表すと、推定しなければならないパラメータの集合は、 $\Theta = \{\theta^{(1)}, \dots, \theta^{(T)}, \omega\}$ となる。

3 F0 制御モデルのパラメータ最尤推定

F0 軌跡の集合 $\{\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(T)}\}$ が与えられたとき、パラメータ集合 Θ を最尤推定する方法を提案する。 $\{\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(T)}\}$ の対数尤度は、式 (3) より、

$$\begin{aligned} \log \mathbb{P}(\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(T)} | \Theta) &= \sum_{t=1}^T \log \mathbb{P}(\mathbf{y}^{(t)} | \theta^{(t)}, \omega) \\ &= \sum_{t=1}^T \left\{ - (L - M) \log(\sqrt{2\pi}\sigma^{(t)} a_0^{(t)}) \right. \\ &\quad \left. - \frac{1}{2\sigma^{(t)2}} \sum_{l=M+1}^L \left(\sum_{j=0}^M a_j^{(t)} y_{l-j}^{(t)} - m_{q_l^{(t)}} \right)^2 \right\} \end{aligned} \quad (5)$$

となる。ここで各フレームの F0 軌跡は独立に観測されるものと想定する。また、 $\sigma^{(t)2}$ はフレーム t において仮定する式 (3) の正規分布の分散値に相当する。パラメータ集合 Θ の事後確率は、 $\mathbb{P}(\Theta | \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(T)}) \propto \mathbb{P}(\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(T)} | \Theta) \mathbb{P}(\Theta)$ なので、 $U(\Theta) \triangleq \mathbb{P}(\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(T)} | \Theta) \mathbb{P}(\Theta)$ とおくと、

$$\begin{aligned} \log U(\Theta) &= \sum_{t=1}^T \log \mathbb{P}(\mathbf{y}^{(t)} | \theta^{(t)}, \omega) + \sum_{t=1}^T \log \mathbb{P}(\theta^{(t)}) \\ &\quad + \log \mathbb{P}(m_{S_1}, \dots, m_{S_I}) + \log \mathbb{P}(q_1, \dots, q_N) \end{aligned} \quad (6)$$

と記述できる。ここで事前確率 $\mathbb{P}(\theta^{(t)})$ と $\mathbb{P}(m_{S_1}, \dots, m_{S_I})$ は一様分布を想定し、 $\mathbb{P}(q_1, \dots, q_N)$ は、一様な Markov 連鎖を想定するため、

$$\log \mathbb{P}(q_1, \dots, q_N) = \log \mathbb{P}(q_1) \mathbb{P}(q_2 | q_1) \dots \mathbb{P}(q_N | q_{N-1}) \quad (7)$$

となる。遷移確率 $\mathbb{P}(S_i | S_j)$ は事前に決定する定数である。簡単のため、 $P_{S_i, S_j} \triangleq \log \mathbb{P}(S_i | S_j)$ とおく。以上より、式 (5) と式 (7) を式 (6) に代入し、定数項を除いた以下の式

$$\begin{aligned} J &\equiv - (L - M) \sum_{t=1}^T \log \sigma^{(t)} a_0^{(t)} \\ &\quad - \sum_{t=1}^T \frac{1}{2\sigma^{(t)2}} \sum_{l=M+1}^L \left(\sum_{j=0}^M a_j^{(t)} y_{l-j}^{(t)} - m_{q_l^{(t)}} \right)^2 + \sum_{n=1}^N P_{q_n, q_{n-1}} \end{aligned} \quad (8)$$

がパラメータの集合 Θ に関して最大化したい目的関数である。しかし、式 (8) を最大化する Θ は解析的に求めることができない。そこで、 Θ の各要素に関して、他の要素を固定したもとで式 (8) を最大化するステップを式 (8) の値が収束するまで繰り返す (Fig. 4)。

3.1 インパルス応答 $\theta^{(t)}$ の推定

パラメータ集合 ω を固定して、式 (8) を最大にする $\theta^{(t)}$ を求める。 J を $a_{j'}^{(t)}$ ($j' = 0, \dots, M$) に関して偏微分して 0 とおくと、

$j' = 0$ のとき

$$a_0^{(t)} \sum_{l=M+1}^L \left(\sum_{j=0}^M a_j^{(t)} y_{l-j}^{(t)} - m_{q_l^{(t)}} \right) y_l^{(t)} + \sigma^{(t)2} (L - M) = 0 \quad (9)$$

$j' = 1, \dots, M$ のとき

$$\sum_{l=M+1}^L \left(\sum_{j=0}^M a_j^{(t)} y_{l-j}^{(t)} - m_{q_l^{(t)}} \right) y_{l-j'}^{(t)} = 0 \quad (10)$$

となる。 J を $\sigma^{(t)}$ に関して偏微分して 0 とおくと、

$$\sigma^{(t)2} = \frac{1}{L - M} \sum_{l=M+1}^L \left(\sum_{j=0}^M a_j^{(t)} y_{l-j}^{(t)} - m_{q_l^{(t)}} \right)^2 \quad (11)$$

となる。式 (5) の正規化項に動的変動成分のスケール $a_0^{(t)}$ を含むため、 j' が 0 とそれ以外で、偏微分の式が異なる。したがって、以上の連立方程式を解析

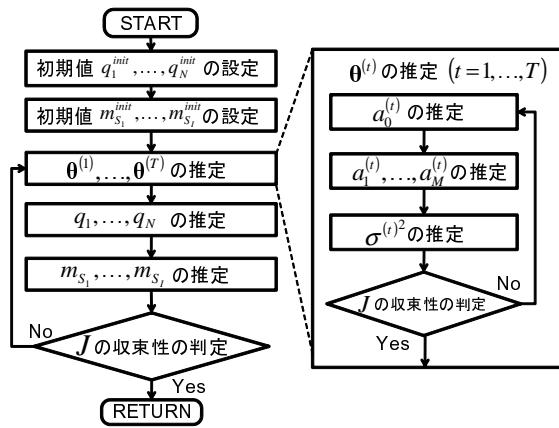


Fig. 4 反復法によるモデルパラメータの推定手順

的に解くことはできない．そこで、まず、式 (9) の $a_1^{(t)}, \dots, a_M^{(t)}, \sigma^{(t)}$ を固定して、 $a_0^{(t)}$ に関する 2 次方程式を解析的に解く．次に、式 (10) の $a_0^{(t)}$ を固定して、 $a_1^{(t)}, \dots, a_M^{(t)}$ に関する連立方程式を解析的に解く．最後に式 (11) の $\sigma^{(t)^2}$ を計算する．この計算過程を目的関数 J が収束するまで順番に繰り返して、 $a_0^{(t)}, \dots, a_M^{(t)}, \sigma^{(t)}$ を更新させる．

3.2 状態系列 q_1, \dots, q_N の推定

インパルス応答の集合 $\{\theta^{(1)}, \dots, \theta^{(T)}\}$ と HMM のパラメータ m_{S_1}, \dots, m_{S_I} を固定して、状態系列 q_1, \dots, q_N に関して、式 (8) の最大化を考える．この問題は、Viterbi アルゴリズム (動的計画法) により効率的に解ける．最初から時刻 k に状態 S_i に至るまでの最適な状態系列は漸化式を利用して、

$$\delta_k(S_i) = \max_{S_j} \left[\delta_{k-1}(S_j) - \sum_{(t,l) \in C_k} \frac{1}{2\sigma^{(t)^2}} (\hat{u}_l^{(t)} - m_{S_i})^2 + P_{S_i, S_j} \right] \quad (12)$$

$$(C_k = \{(t, l) | y_k \in \mathbf{y}^{(t)}, 1 \leq t \leq T, 1 \leq l \leq L\})$$

と記述できる．ここで、 C_k は、1 から T までのフレームの中で、 y_k に相当する $y_l^{(t)}$ のフレーム番号 t とインデクス l の組み合わせからなる集合である．式 (12) を $k = N$ まで計算すれば、最適な経路が求められる．

3.3 HMM のパラメータの推定

インパルス応答の集合 $\{\theta^{(1)}, \dots, \theta^{(T)}\}$ と状態系列 q_1, \dots, q_N を固定して、式 (8) が最大になるように HMM のパラメータ m_{S_1}, \dots, m_{S_I} を推定する． J を m_{S_i} に関して偏微分して、0 とおくと、

$$m_{S_i} = \frac{1}{|D_i|} \sum_{n \in D_i} \frac{1}{|C_n|} \sum_{(t,l) \in C_n} \hat{u}_l^{(t)} \quad (13)$$

を得る．ここで、集合 $D_i = \{n | q_n = S_i\}$ とし、 $|D_i|$ をその要素数とする．最終的に、3.2 節と本節で推定される状態系列 q_1, \dots, q_N と HMM のパラメータ m_{S_1}, \dots, m_{S_I} から、階段状の入力信号 $u_n = m_{q_n}$, ($n = 1, 2, \dots, N$) が決定される．

以上のように階段状の入力信号を HMM によってモデル化し、そのパラメータ学習方法が入力信号と F0 軌跡に基づく逆畳み込みと Viterbi 学習から成ることから、提案モデルを畳み込み HMM (Convolutional HMM) に基づく歌声の F0 制御モデルと呼ぶ．

3.4 初期値の設定

局所解に収束することを防ぐために、2 つの初期値設定を行う (Fig. 4)．まず、観測系列 \mathbf{y} に 3.2 節で述べた Viterbi アルゴリズムを適用して初期状態系列 $q_1^{\text{init}}, \dots, q_N^{\text{init}}$ を決定する．次に、3.3 節と同様に、初期状態系列から初期の HMM のパラメータ $m_{S_1}^{\text{init}}, \dots, m_{S_I}^{\text{init}}$ を決定する．

3.5 F0 軌跡の生成

フレーム t において推定されたインパルス応答 $\theta^{(t)}$ と状態系列 $q_1^{(t)}, \dots, q_L^{(t)}$, HMM のパラメータ m_{S_1}, \dots, m_{S_I} から、F0 軌跡を生成する．式 (2) より、フレーム t において生成される F0 軌跡 $\hat{y}^{(t)}$ は、

$$\hat{y}_l^{(t)} = \frac{1}{a_0^{(t)}} (u_l^{(t)} - \sum_{j=1}^M a_j^{(t)} \hat{y}_{l-j}^{(t)}), \quad (l \geq M + 1) \quad (14)$$

となる．ただし、 $\hat{y}_l^{(t)} = y_l^{(t)}$, ($l \leq M$) とする．

4 評価実験

畳み込み HMM に基づく F0 制御モデルのパラメータ学習アルゴリズムによって、各フレームにおいて推定される階段状の入力信号 $\mathbf{u}^{(t)}$ とインパルス応答 $\theta^{(t)}$, それらから生成される F0 軌跡 $\hat{\mathbf{y}}^{(t)}$ の性能を評価する．

4.1 歌声データベース

クラシックの音楽家、プロのポップス歌手、音楽的な訓練を受けていない素人 (それぞれ男女 1 名ずつの計 6 名) の歌声からなるデータベースを構築した．歌唱者は、ヘッドフォンで歌唱曲の旋律 (ガイドトーン) を聴きながら伴奏なしで歌唱した．歌唱曲は、「きらきら星」、「喜びの歌 (Beethoven の交響曲第 9 番第 4 楽章の歌の部分) を岩佐東一郎氏によって作詞されたもの」である．日本語歌詞による歌唱、ハミングによる歌唱の 2 パターンを収録した．

4.2 実験条件

F0 は、de Cheveigné らの提案した YIN[5] を利用して 10ms ごとに推定される．なお、 Hz で表される周波数 y_{Hz} を、 $cent$ で表される対数スケールの周波数 y_{cent} に変換する．200ms 以内の F0 が推定されない区間は、無声音の区間とみなし、F0 を線形補間する．また、ガイドトーンからも F0 を推定し、提案手法により推定される階段状の入力信号の評価のために利用する．

推定された F0 軌跡は、500ms ずつシフトさせて得られる長さ N の素片に分割される．本実験では $N = 2s$ とする．この素片が 2.4 節で述べた \mathbf{y} に相当し、素片ごとに提案手法を適用する．素片に F0 が推定されない区間 (休符区間) が含まれる場合は、今回は分析を行わない．各素片におけるフレームは 100ms ずつシフトさせる．また、各素片の HMM の状態数 I は相平面を利用して決定する．相平面に描かれる F0 軌跡のアトラクタの分布を混合ガウス分布 (GMM) によって学習し、その極大値の数を HMM の状態数 I とする [6]．各状態の初期状態確率は $1/I$ 、自己遷移確率は 0.9、状態間の遷移確率は $0.1/(I-1)$ とした．各フレームにおいて仮定する正規分布の分散の初期値は 2500 とした．

4.3 フレーム長と伝達関数の次数の評価

ある素片の推定結果を Fig. 5 に示す．フレーム長 L は 250ms、伝達関数の次数 M は 3 とした．音高遷移の部分を除いて、入力信号 \mathbf{u} は対応する区間のガイドトーンの F0 軌跡 \mathbf{g} と近いものが推定される．し

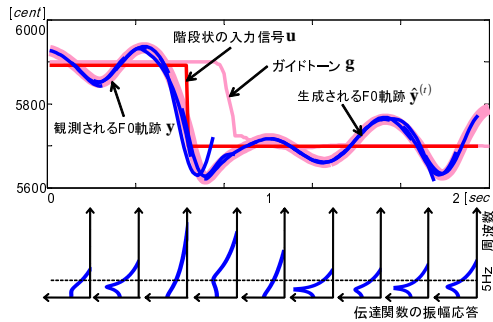


Fig. 5 推定結果の例

しかし、F0 軌跡 $y^{(t)}$ と生成される F0 軌跡 $\hat{y}^{(t)}$ との間に誤差が生じる箇所がいくつか存在する。素片によっては $\hat{y}^{(t)}$ が発散してしまうフレームも存在した。

以上を踏まえて、 L と M を変化させたときの各フレームの $u^{(t)}$ と $\hat{y}^{(t)}$ の推定性能を評価した。フレームごとに以下の平均二乗誤差 (RMS)

$$\text{RMS}_{u^{(t)}} = \sqrt{\frac{1}{L-M} \sum_{l=M+1}^L (u_l^{(t)} - g_l^{(t)})^2} \quad (15)$$

$$\text{RMS}_{\hat{y}^{(t)}} = \sqrt{\frac{1}{L-M} \sum_{l=M+1}^L (\hat{y}_l^{(t)} - y_l^{(t)})^2}$$

を計算し、分析を行ったすべてのフレームのうち正しく推定された割合を正解率として以下に定義する。

$$\text{正解率}_{u^{(t)}} = \frac{\text{RMS}_{u^{(t)}} < \xi \text{ となるフレーム数}}{\text{すべての素片の総フレーム数}} \times 100$$

$$\text{正解率}_{\hat{y}^{(t)}} = \frac{\text{RMS}_{\hat{y}^{(t)}} < 10 \text{ となるフレーム数}}{\text{すべての素片の総フレーム数}} \times 100 \quad (16)$$

ここで、 ξ は各素片の y と g の RMS とする。ガイドトーンを聴きながら歌唱しても、歌唱者によってはガイドトーンからずれて歌唱するため、 ξ 以内で $u^{(t)}$ が推定されれば正解とした。 $\hat{y}^{(t)}$ に関しては、 $\text{RMS}_{\hat{y}^{(t)}}$ が 10cent 以内で生成されれば正解とした。

推定性能を Fig. 6 に示す。 L を短くするにつれて、 $\hat{y}^{(t)}$ の正解率が向上した。つまり、時変な F0 制御モデルとして、素片をフレームに分割し、フレームごとに伝達関数を推定することの有効性を確認できた。また、 M の増加によって伝達関数の自由度が増すため、正解率が向上した。しかし、総フレーム数の 3 割は、伝達関数が適切に推定されないため、F0 軌跡を生成させることが困難であった。また、 L を短く、 M を大きくしすぎると、素片内で扱うモデルパラメータ数が増えてしまう問題もある。したがって、情報量基準などを利用して L や M を変化させながら推定を行うなどの提案手法の改善が必要とされる。

4.4 伝達関数の振幅応答に基づく歌唱者の個性に関する考察

Fig. 5 の下段に式 (1) から計算される伝達関数の振幅応答を示す。音高が安定するフレームでは共振周波数を持ち、そのいくつかはビブラート (4~7Hz 程度の周期的変動 [1]) に対応すると考えられる。そこで、式 (16) で正解とされたフレームの共振周波数の分布を Fig. 7 に示す。歌唱者ごとに分布の形状が異なり、特に女性の声楽家は 2.5Hz あたりに分布が集中した。この歌手は音高が安定するときに顕著に振動させる歌い方をすることを確認した。一方、素人の歌唱者ほど平坦な分布となり、推定された伝達関数の振幅応答から個性や歌唱力の違いを確認することができた。

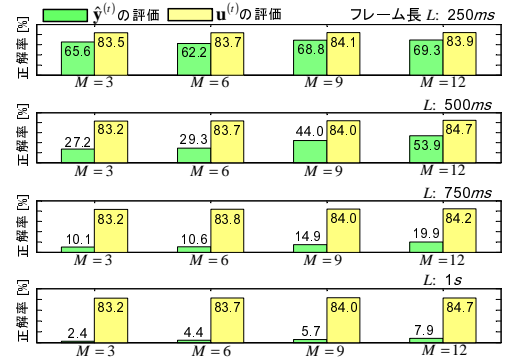


Fig. 6 フレーム長と伝達関数の次数に関する評価

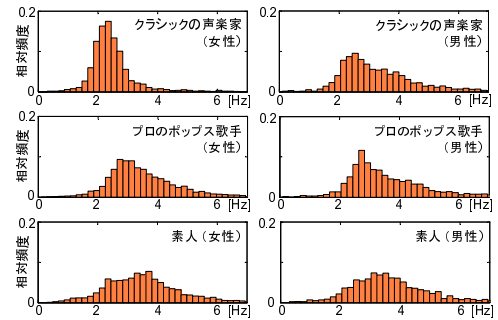


Fig. 7 歌唱者ごとの共振周波数の分布

5 まとめと今後の展開

歌声の F0 軌跡から、階段状の入力信号と動的変動因子を表すインパルス応答を同時推定するための F0 制御モデルとパラメータ学習アルゴリズムを提案した。実験結果より、分析を行った総フレーム数の 69.3% は正確にそれぞれの信号を推定できることを確認した。また、推定される伝達関数の振幅応答に、歌唱者の個性や歌い方の違いが表れることを確認した。

性能を向上させるためには、フレーム長や伝達関数の次数を可変にしてパラメータを推定することが必要とされる。また、F0 軌跡の符号化、歌唱者の歌い方を反映した歌声合成、歌唱力評価など本研究の応用先について検討する必要がある。楽器音や生体信号などの時系列信号への適用や、提案モデルを多変量化し、MFCC ベクトルの時系列などを動的モデリングして音声分析に利用することも今後の展開である。

謝辞 本研究は日本学術振興会特別研究員 (DC2) 科研費の補助を受けた。また、これまで本研究に対し、有益なご助言を頂いた後藤真孝氏 (産総研)、伊藤克亘氏 (法政大) に感謝致します。また、F0 制御モデルに関して有益なご意見を頂いた南泰浩氏 (NTTCS 研) に感謝致します。

参考文献

- 齋藤 毅ほか：歌声の個人性知覚に寄与する音響特徴の検討，音講論集，2-Q-26，pp. 601-602 (2007)。
- 齋藤 毅ほか：SingBySpeaking: 歌声知覚に重要な音響特徴を制御して話声を歌声に変換するシステム，情処研報音楽情報科学，Vol. 2008，No. 74，pp. 25-32 (2008)。
- Bonada, J. et al.: Synthesis of the Singing Voice by Performance Sampling and Spectral Models, *IEEE Signal Processing Magazine*, Vol. 24, No. 2, pp. 67-79 (2007)。
- 酒向慎司ほか：声質と歌唱スタイルを自動学習可能な歌声合成システム，情処研報音楽情報科学，Vol. 2008，No. 74，pp. 51-58 (2008)。
- de Cheveigné, A. et al.: YIN, a fundamental frequency estimator for speech and music, *Journal of the Acoustical Society of America*, Vol. 111, No. 4, pp. 1917-1930 (2002)。
- 大石康智ほか：相平面を利用した歌声の F0 軌跡の新しい表現方法，電子情報通信学会総合大会，pp. S-51-S-52 (2008)。