楽器固有の高調波構造モデルを用いたギター演奏に対する多重音高推定手法*

◎有元 慶太, 藤島 琢哉 (ヤマハ ST 開発センター), 後藤 真孝 (産総研)

1 はじめに

複数の楽器音が重畳する音楽音響信号から個々の音高 (基本周波数, F0) を推定する手法は, 古くから研究されてきており [1], 近年では Goto[2], Kameoka ら [3] の研究のように統計的推定の導入が盛んである. 例えば, 最大事後確率推定を用いた音高推定手法 PreFEst[2] は, 各単音の高調波構造をモデル化した確率分布 (音モデル) を用いて, それらをあらゆる F0 に並べた重み付き和のモデル (混合分布モデル) として入力の混合音を表現し, 各音モデルの重み (F0 の確率密度関数) と倍音比率をモデルパラメータとして推定することができる. 本研究では, 当初音楽 CD 等から単音のメロディ/ベースの F0 をリアルタイムに推定する手法として提案された PreFEst[2] を応用し, 従来ほとんど取り組まれていないギター独奏による最大同時 6 音までの F0 推定に特化するための改良を行った.

2 音高推定手法 PreFEst に対する改良

文献 [2] に述べられているように, PreFEst は, 様々な楽器の音色を音モデル (倍音比率) の事前分布 (これは初期値の設定にも用いられる) として与えることができる. しかし, 従来はガウス分布状に倍音比率の高域が減衰する事前分布等しか試されていなかった. また, 各 F0 の音モデルの音色 (倍音比率) と相対音量 (F0 の確率密度関数) は推定されるものの, 得られた倍音比率の妥当性を評価したり, 同時発音数を求めて個々の F0 を決定したりする方法は未検討であった.

そこで本研究では, 入力をギター独奏に限定し, 次の三つの改良を行った.

1. ギターの音色を事前分布として与える.
2. 推定された倍音比率がギターの音色と大きく違う場合には, その音が鳴っていないとみなす.
3. 人間に可能なギターのフレットの押さえ方 (フォーム) に関するトップダウンの知識を利用して, 同時発音数と各 F0 を決定する.

2.1 ギターの音色を反映した音モデル定義

PreFEst では, 同一 F0 に対して複数の音モデルを用意して推定できるが, 本研究ではギターの弦と音モデルを対応させ, 弦の数と同じ 6 種類の音モデルを定義した. その際に, 各弦の音域は重なりつつも範囲が違うため, 各音モデルの適用可能音域を限定する機能を追加した. 具体的には, Fig.1 に示すように, 各弦についてフレットに対応する半音単位で実楽器音をサンプリングして, 音モデルを設定した (以下, **ギター音モデル**と呼ぶ.). 特定の弦の特定のフレット位置でサンプリングして設定された音モデルは, その F0 での推定だけに用いられる.

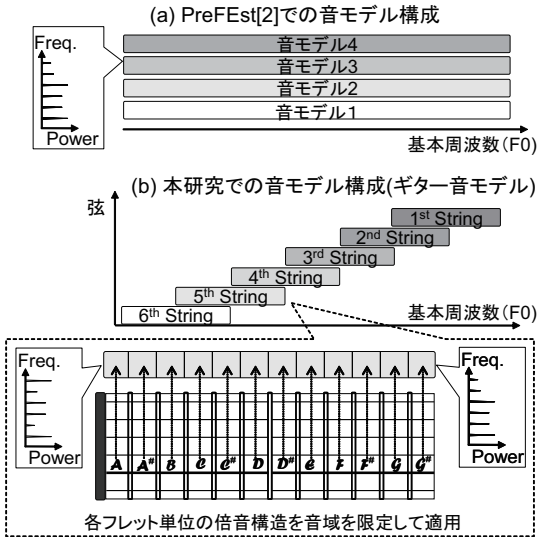

Fig. 1 ギターの楽器構造を反映した音モデル定義

2.2 倍音形状相違度による発音の有無の判断

ギター音モデルを用いて PreFEst で推定された倍音比率は, もしその F0 が入力中で鳴っているとしたら, その音はどんな倍音比率を持つのかを反映している. 実際に鳴っている F0 の場合には, ギター音モデルの倍音比率に近いことが多い (Fig.2 の右側のモデル推定値). 一方, 例えば半ピッチエラーのように, 実際に鳴っている音の 1 オクターブ下で推定されてしまった場合には, 偶数次倍音のみが強くなり, ギター音モデルとは大きく異なることが多い (Fig.2 の左側のモデル推定値). このように, 推定前後の倍音比率の変化を評価すれば, ある音が鳴っているかどうかの判断材料となる.

そこで, ギター音モデルと, 推定後の音モデル (モデル推定値) との間で Symmetric Kullback-Leibler (KL2) 情報量を求め, これを倍音形状相違度と定義した. 今回は第一歩として, 単にこの相違度が事前に定めた閾値よりも大きい場合は, 音モデルのその弦・フレットに対応する F0 の重み (F0 の確率密度関数) を 0 にして, その重みを初期値として再度 PreFEst で推定を行った. これにより, 再度の推定では, 鳴っていないと判断された音は使われなくなり, より適切な結果が得られることが期待できる.

2.3 ギターフォーム制約による音高候補の絞り込み

ギター独奏では, 各時刻において同時発音数が 0 音から 6 音までの間で変動するが, その各音の F0 を, F0

* "A Multiple F0 Estimation Method Using Specific Harmonic Structure Models for Guitar Performances" by Keita Arimoto, Takuya Fujishima (YAMAHA Corp. Center for Advanced Sound Technologies), and Masataka Goto (AIST)

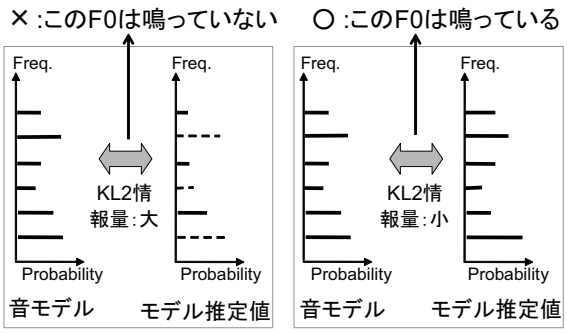

Fig. 2　ギター音モデルに基づく倍音形状相違度の評価

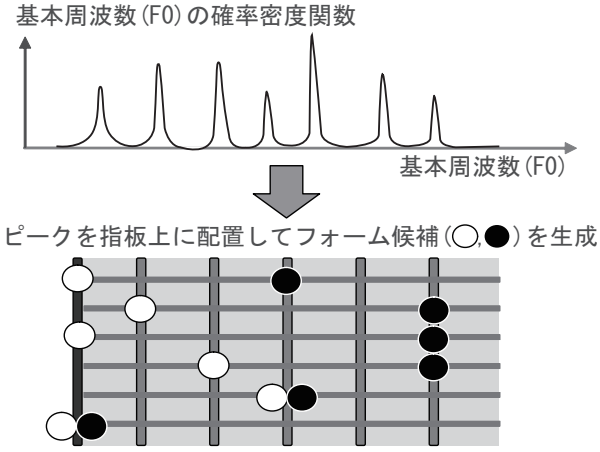

Fig. 3　ギターフォームの候補の生成

の確率密度関数から決定するには, その確率密度関数中のピーク（音高候補）を選べばよい. しかし, 推定結果には, 2.2 節の改良を加えても尚, 実際には発音していない偽りのピークが残ることがある. そこで, 人間に可能なフレットの押さえ方 (フォーム) に関するトップダウンの知識を用いて, 音高候補を絞り込んだ.

具体的には, まず F0 の確率密度関数を推定した後に, そのピークをギターの指板上に配置して, Fig.3 に 2 種類 (○, ●) 例示したように, 人間が手で押さえることが可能なフォームの候補を生成する. そして, 各候補の構成音の位置の確率密度関数の値の合計値を求め, 最も合計値が高い候補を採用する. 次に, そのフォームの全構成音が実際に鳴っていたかを検証するために, F0 の確率密度関数の事前分布 (通常は一様分布) として, 構成音の位置に等しい高さのピークを持つような分布を設定する. そして, 再度 PreFEst で推定して得られた F0 の確率密度関数に対して, その最大ピークに係数 (0.3) を乗じた値を閾値として, それより大きいピークを出力する F0 として決定した.

3　実験結果

ギター音モデルの設定に使用したエレキギターを, プロ演奏者 1 名が独奏した演奏音 (ドライソース) 計 307.4 秒を対象に評価した. これには同時発音数 0〜6 音の様々なフレーズの演奏が含まれている. これに対する F0 の正解データは, パワースペクトルを見ながら人手で作成した. その正解データの F0 と本手法で決定した F0 とをフレーム (10 ミリ秒) 単位で比較し, 一致 (C), 脱落誤り (FN), 挿入誤り (FP) の数を集計した. ただし, 正解データが無音 (発音数 0 音) のフレームは集計対象外とした. 正解率は, それらから $C/(C+(FP+FN)/2)$ によって計算される F-measure($\beta=1$) と定義した.

Table 1　各実験条件での正解率 (○:有効, ×:無効)

ギター音モデル	×	○	○	×	○
倍音形状相違度評価	×	×	○	×	○
ギターフォーム制約	×	×	×	○	○
正解率 (%)	81.3	82.9	83.2	85.1	90.0

Table 2　正解中の同時発音数ごとに集計した正解率

同時発音数	1	2	3	4	5	6
合計区間長 (秒)	84.2	53.7	19.3	27.5	42.1	18.6
正解率 (%)	85.2	90.9	84.3	92.2	91.4	91.3

個々の改良項目の効果を検証するために, 改良前の PreFEst[2] をベースラインとし, 2.1〜2.3 節の各改良項目について有効/無効を切替えて評価した. その結果を Table 1 に示す. 便宜上, ギターフォーム制約を無効にした場合は, 文献 [2] のマルチエージェントモデルで追跡されるピークから各時刻で最大 6 個を選び, それらに対して閾値処理した結果を評価した.

Table 1 から, 改良項目を有効にするほど正解率が向上することが分かる. 特に, フォーム制約は効果的であった. 全ての改良項目を有効にした提案手法の正解率は 90.0%に達し, ベースラインから 8.7%改善された. Table 2 に, そのときの正解中の同時発音数ごとに求めた正解率を示す. 各音数が正解中に出現する区間長の合計も付記した. この表から, 1 音と 3 音の正解率は他よりも低いことがわかった. 原因を調べたところ, 1 音の場合はオクターブ上の誤推定, 3 音の場合はパワーコード (根音+完全 5 度上+完全 8 度上) に対する誤推定が多かった.

4　まとめ

本研究では PreFEst[2] を改良し, ギター独奏音に特化した多重音高推定手法を実現した. エレキギターによる最大同時 6 音の独奏に対して評価した結果, F-measure による正解率は 81.3%から 90.0%へ改善し, 改良が有効であったことが確認された. 今後は, 他の楽器を対象に, 今回のような楽器に依存した音モデルや楽器演奏上の制約の導入を検討していく予定である.

参考文献

[1] A. Klapuri and M. Davy (Eds.), "Signal Processing Methods for Music Transcription," Springer-Verlag, 2006.

[2] M. Goto, "A Real-time Music Scene Description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals," Speech Communication (ISCA Journal), Vol.43, No.4, pp.311-329, 2004.

[3] H. Kameoka, T. Nishimoto, and S. Sagayama, "Harmonic-Temporal Structured Clustering via Deterministic Annealing EM Algorithm for Audio Feature Extraction," Proc. ISMIR 2005, 2005.