

押弦制約付きギター演奏自動採譜システム

矢澤 一樹[†] 阪上 大地[‡] 糸山 克寿[‡] 尾形 哲也[‡] 奥乃 博[‡]

[†] 京都大学 総合人間学部 認知情報学系 [‡] 京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

音楽情景分析・演奏支援・音楽学習などを目的として、音楽 CD などの楽器演奏をコンピュータによって五線譜、タブ譜、ピアノロールなどの楽譜に変換する自動採譜が取り組まれている。特にギターは演奏人口が多いため、ギター用タブ譜の高精度な生成は重要な課題である。ギターの和音のような多重基本周波数の高性能な推定法の一つとして潜在的調波配分法 (latent harmonic allocation; LHA)[1] がある。

LHA の問題点は、楽器の構造や人間の身体的制約上演奏できない音の組み合わせを出力しうる (例えば 6 弦のギター演奏に対して 7 音以上の同時発音を出力することである。このような推定結果を直接楽譜に変換しても、演奏支援に用いることはできない。これは LHA が特定の楽器の使用を前提としていないためであり、演奏楽器を限定し楽器固有の特徴 (発音可能な音域や同時発音可能な音の組み合わせなど) を制約として用いると、この問題は解決できると考えられる。

そこで本稿では演奏楽器をギター 1 本に限定し、LHA に後処理的に押弦制約を加えることで演奏不可能な音の組み合わせを排除する手法を報告する。本手法では押弦可能なギターフォームをあらかじめ列挙し、各時間フレームごとの最適フォームを推定することで、そのフォームで発音不可能な音を抑圧する。押弦可能フォームはフレット幅と押弦箇所数によって選別し、ギター演奏で頻繁に用いられる演奏法セーハについても考慮する。評価実験では、押弦制約を用いることで閾値に対する頑健性が得られ、システムの応用性が向上したことを示す。

2. ギターフォーム列挙法

あるギターフォームが押弦可能かどうかは、主に押弦箇所数 (何本の弦を同時に押さえるか) とフレット幅 (指をどのくらい広げるか) という 2 つの条件で判断する。またギター演奏においては基本的には 1 本の指で 1 本の弦を押さえるが、人差し指で複数の弦を同時に押さえるセーハ (図 1) と呼ばれる演奏法も頻繁に用いられるため、それを考慮してパターンを列挙する。

2.1 基本パターンの列挙

まずはじめに、手首の位置を考えず指の相対的な位置関係だけを考えた基本パターンを生成する。後でセーハを考えるため、人差し指を使用せずに押弦可能な基本パターン (BP1) と人差し指の使用を許した基本パターン (BP2) を、それぞれ次の条件で別々に列挙する。

BP1 押弦箇所数 3 以下、フレット幅 3 以下。

BP2 押弦箇所数 4 以下、フレット幅 4 以下。

また上記の条件を満たさないが比較的よく出現するパターンとして、図 2 の例外基本パターンを BP2 に加える。

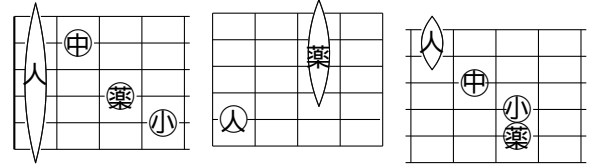


図 1: セーハの例 図 2: 例外基本パターン 図 3: P1 の例

2.2 全パターンの列挙

次に BP1 と BP2 をギターの指板上で水平方向に移動することで、手首の位置を考慮した押弦可能フォーム P1, P2 を作成する。標準的なギターは 21 のフレットをもつため、P1 の移動で 19 通り、P2 の移動で 18 通りのフォームが生成できる。さらに P1 についてはセーハを用いたギターフォームを含み、BP1 の左隣のフレットの複数の弦 (最高 6 本) を余った人差し指を使って同時に押さえるフォームを追加した。

P1 BP1 を水平方向に移動し (19 通り)、左隣の複数の弦 (0-6 本) をセーハした全パターン (図 3)。

P2 BP2 を水平方向に移動 (18 通り) した全パターン。

最後に、P1 と P2 を合わせ重複したフォームを削除した全てのパターンを、押弦可能フォーム F_1, \dots, F_P とする。本手法で列挙された押弦可能フォームの総数 P は 89479 となった。各フォームで発音可能な音の組み合わせを K_p とする。ここでは通常の 6 弦ギターを考えているので、各組み合わせ K_p には 6 つの音が含まれる (ただし重複を許す)。

3. 押弦制約法

本節では LHA の概要とギターの押弦制約を用いた多重基本周波数推定法について述べる。

3.1 潜在的調波配分法

LHA でははじめに、観測された音響信号に対してウェーブレット解析を行い、得られた振幅スペクトルの系列に対して基本周波数推定を行う。時間フレーム数を D とし、 D フレーム合わせた全ての観測変数を $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_D\}$ とする。ここで $\mathbf{X}_d = \{x_{d1}, \dots, x_{dN_d}\}$ は各フレームにおける観測周波数の系列であり、例えばフレーム d の振幅スペクトルにおける周波数 f のパワーが a であれば、フレーム d で周波数 f は a 回観測されたとみなす。 N_d はフレーム d での観測周波数の総数である。

次に上記の振幅スペクトルを、それぞれが倍音数 M の調波構造をもつ K 個の基底を混合した、以下のネスト型混合ガウス分布で定式化する。

$$\mathcal{M}_d(x) = \sum_{k=1}^K \pi_{dk} \left\{ \sum_{m=1}^M \tau_{km} \mathcal{N}(x | \mu_k + o_m, \lambda_k^{-1}) \right\} \quad (1)$$

ここで μ_k, λ_k は基底 k のガウス分布の平均および精度であり、 o_m は M 個のガウス分布を倍音関係に配置する

Music Transcription of Guitar Sound using Fingering Position Restriction: Kazuki Yazawa, Daichi Sakaue, Katsutoshi Itoyama, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

表 1: 基本周波数推定結果の F 値．最適値は各楽曲で F 値がもっとも大きくなるような閾値パラメータの値を表す．

閾値変数	LHA(従来法)					押弦制約付き LHA(提案法)				
	最適値(閾値)	0.20	0.10	0.05	0.01	最適値(閾値)	0.20	0.10	0.05	0.01
RM-J006	0.63(0.17)	0.63	0.57	0.40	0.13	0.63(0.17)	0.63	0.60	0.51	0.41
RM-J007	0.83(0.08)	0.54	0.82	0.77	0.37	0.81(0.09)	0.55	0.80	0.80	0.72
RM-J009	0.85(0.15)	0.82	0.76	0.59	0.25	0.86(0.15)	0.83	0.82	0.74	0.60
RM-J010	0.82(0.16)	0.79	0.73	0.61	0.28	0.82(0.16)	0.81	0.78	0.73	0.66
平均	0.78	0.70	0.72	0.59	0.26	0.78	0.70	0.75	0.69	0.60

オフセット値である．さらに τ_{km} は基底 k における倍音 m の相対強度， π_{dk} はフレーム d での基底 k の相対強度である．生成モデルの詳細およびパラメータの更新式は，紙面の都合上省略する．各パラメータは変分 EM アルゴリズムで推定する．

パラメータの推定後，各フレーム d における基底 k の有効観測数 $N_{dk} = \pi_{dk} N_d$ がある閾値以上となるような k を求め，その音高の音を実際に演奏された音と判断する．すなわち t を閾値パラメータとし， $N_{dk} \geq t \max_{dk} N_{dk}$ を満たす全ての k に対応する音高をフレーム d での推定結果とする．

3.2 押弦制約法

LHA の問題は，楽器の構造や身体的制約を考慮しないため，ギター演奏として不自然な音の組み合わせが生成されることである．本手法では LHA による推定結果を後处理的に修正し，時間フレーム毎に最適ギターフォームを推定してそのフォームで発音不可能な音を排除する．事前に列挙した押弦可能なフォーム F_1, \dots, F_p ごとにフレーム d でのフォーム F_p の有効観測数 $N_{dp} = \sum_{k \in \mathbf{K}_p} N_{dk}$ を計算する．最後に，各フレーム d でこの値が最大となるような p に対応する F_p をそのフレームでの最適フォームとし，このフォームで発音できない音高 k については N_{dk} の値を 0 とする．

$$p_d = \operatorname{argmax}_p N_{dp} \quad (2)$$

$$\tilde{N}_{dk} = \begin{cases} N_{dk} & (\text{if } k \in \mathbf{K}_{p_d}) \\ 0 & (\text{otherwise}) \end{cases} \quad (3)$$

こうして得られた \tilde{N}_{dk} を元の N_{dk} の代わりに用いて最終的な音高を出力することで，最適フォームで発音可能な音のみに推定結果を制限することができる．

4. 評価実験

本手法の性能を評価するため，LHA に押弦制約を加えた本システムと制約を加えていない従来のシステムの両方を使って，ギター演奏音から時間フレーム毎の F0 を推定する実験を行い，結果を比較した．

4.1 実験条件

実験データには，RWC 研究用音楽データベース [2] に含まれるジャズ楽曲のうち，ギター 1 本で演奏された 4 曲の冒頭 24 秒を用いた．楽曲は全て MIDI 音源 (YAMAHA MOTIF-XS) を使用して音響信号を作成し，ウェーブレット変換を行った．正解データ (各時間フレームでの F0 の正解値) は MIDI ファイルから作成し，F0 推定の評価尺度として時間フレームレベルでの F 値を用いた．

モデル中の各パラメータは変分 EM アルゴリズムを

100 回繰り返すことで推定し，事前分布は全て無情報事前分布とした．基底数 K および倍音数 M はそれぞれ 73, 6 とし，閾値パラメータ t については各楽曲に対する最適値，0.20, 0.10, 0.05, 0.01 の 5 条件で実験を行った．また従来法と性能を比較するため，上記と同様の条件で押弦制約を用いない通常の LHA による推定も行った．

4.2 実験結果

表 1 に実験結果を示す．最適な閾値を用いたときの F 値は従来法と提案法で等しいが，提案法では閾値を変化させたときに F 値が低下しにくいことから，提案法の閾値に対する頑健性が示されている．このことは，従来の LHA で閾値を下げると推定されてしまう余分な (実際には演奏されていない) 音が，押弦制約によってある程度抑圧できていることに対応している．

5. おわりに

本稿では，多重基本周波数推定法 LHA にギターの押弦制約を加える方法について述べた．本システムでは押弦制約による閾値頑健性のため，ユーザーが閾値を演奏環境や楽曲に応じて調整しなくても安定した認識精度が得られる．またもう 1 つの利点として，押弦制約において用いた各時間フレームでの最適フォーム F_{p_d} を利用することで，人間が演奏するのに適したタブ譜を出力できる．さらに本手法では演奏楽器をギターに限定していたが，押弦パターンの列挙法を変えることで他の弦楽器の採譜にも応用することが可能である．

列挙した押弦可能フォームは，一般的なギターコード表 [3] に載っているギターフォームを全て含んでおり，通常用いられるフォームの大部分をカバーしていると考えられるが，親指を使った特殊なフォームなどが含まれていない，押弦が難しくほとんど用いられないフォームを含んでいる，といったフォームの過不足の問題が残されている．そのため，本手法における押弦可能フォームの列挙法にはまだ改善の余地があると考えられる．また，時間フレーム間の指の動きを考えた運指制約を加える，音楽的なコードの推移条件を用いるなどの改善に取り組む予定である．なお，本研究は科研費 (S)，GCOE の支援を受けた．

参考文献

- [1] 吉井 他: 多重基本周波数解析のための無限潜在的調波配分法, 情報研報, 2010-MUS-86, 2010.
- [2] 後藤他: RWC 研究用音楽データベース, 情報処論, Vol45, No.3 (2004), 728-738.
- [3] 夏林 他: 「ギターコードダイアグラム (コード表) 初心者のためのアコースティックギターサイト」 <http://www9.ocn.ne.jp/knatsu/chords/chords.html> (2011 年 12 月 17 日)