

ギター演奏からの押弦パターン・発音時刻・フォーム変化時間制約を用いたタブ譜自動生成システム

矢澤 一樹 阪上 大地 柳楽 浩平 糸山 克寿 奥乃 博
 京都大学 大学院情報学専攻 知能情報学専攻

1. はじめに

本研究の目的は、ギター演奏初心者の演奏支援や楽曲の二次利用のための、音響信号を入力としたタブ譜自動生成システムの実現である。LHA [1] などの多重基本周波数推定法の結果から自動生成されたタブ譜は、以下の3つの問題のため演奏に適さないものであった。1. 演奏不可能な音高の組み合わせ、2. 発音時刻以外での押弦フォームの変化、3. 過剰に頻繁な押弦フォームの変化。そこで我々は、既存の推定手法 LHA に新たに 1. 押弦可能フォーム、2. フォーム変化時刻、3. 同一フォーム継続時間に関する制約を加えることで、演奏に適したタブ譜を生成することに成功した。

2. 従来法とその問題

潜在的調波配分法 (LHA) [1] は観測スペクトルから複数音の音高を推定するための機械学習アルゴリズムである。出力として、時刻 t での音高 k の相対的なパワー N_{ik} が得られる。従来の音高推定では、 N_{ik} を閾値判定することで音の有無を判断する。すなわち、閾値パラメータを α とし、 $N_{ik} \geq \alpha \max_{ik} N_{ik}$ をみたす全ての音高 k を時刻 t で鳴っている音高とみなす。

この出力結果をタブ譜に変換しようとした場合、我々は次の3つの問題に直面する。

1. 7音以上の同時発音など、ギター一本では演奏不可能な音の組み合わせが推定される。
2. 使用押弦フォームが発音時刻以外で変化する。
3. フォーム変化に要する時間を考慮していないため、使用押弦フォームが短時間で頻繁に変化しすぎる。

3. タブ譜生成手法

提案法では、LHA の推定結果を元に以下の3つの制約下で最適な押弦フォーム系列を一つ決定し、各フォームで演奏不可能な音高を排除することで、上記の問題を解決する。演奏可能な音の組み合わせを列挙する際には、6本の弦と20のフレットを持つノーマルチューニング (EADGBE) のギターを仮定する。

3.1 3つの制約

3.1.1 押弦パターン制約

まず、ギターで演奏可能な押弦フォームの一覧を列挙する [2]。フォームの押弦可能性は、指を広げる幅と使用する指の数で判断する。一般的に、押弦フォームはオープンコードとバレーコードの2種類に分けられるため、指幅と使用指数の条件を元にこれら2種の押弦可能フォームをそれぞれ列挙した (図1)。ただしバレーコードについて、今回は人差し指で6弦を押さえる形の押弦フォームのみを考慮し、それ以外の種類のバレーコード (薬指で複数弦を押さえるフォームや5弦以下の弦を同時に押さえるフォームなど) は実際の演奏ではほとんど用いられないため無視した。

列挙されたフォームの総数は 38,119 となった。以降では、 p をフォーム番号、発音可能な音高の組み合わせを

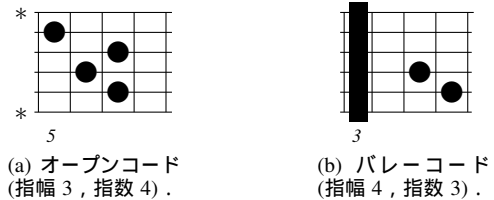


図1: 押弦可能フォームの例。星印は開放弦、長方形は人差し指、数字5と3はフレット番号をそれぞれ表わす。

\mathbf{K}_p ($p = 1, \dots, 38,119$) と表わす。ギターは通常6弦であるため、 \mathbf{K}_p には最大6つの音高が含まれる。

推論の際、各時刻で用いられる押弦フォームを、上述の押弦可能フォームの中から一つ選ぶこととする。そのため、時刻 t での各フォーム p の出現度を、 $N_{tp} = \sum_{k \in \mathbf{K}_p} N_{tk}$ と定義する。最適な押弦フォーム系列は、以降で述べる残り2つの制約下で、 $\sum_t N_{tp_i}$ が最大となるように選ばれる。ここで、 p_t は時刻 t で用いられるフォームを表わす。

3.1.2 発音時刻制約

演奏中の押弦フォームの変更は通常、その時点のフォームで演奏不可能な音高を演奏するために行われる。そこで我々は、入力音に対し発音時刻検出を行い、その時刻でのみ使用押弦フォームを変更できるという制約を加える。

発音時刻は spectral flux ($SF_t = \sum_f \max(0, X_{tf} - X_{t-1,f})$) を用いて検出する [3]。ここで、 X_{tf} は時刻 t での周波数 f のスペクトルパワーを表わす。 SF_t がある特定の閾値 $\beta \max_t SF_t$ よりも大きくなる時刻 t を発音時刻とみなす。

3.1.3 フォーム変化時間制約

人間が指を動かせる速度には限界があるため、フォーム変化が頻繁に起こりすぎるようなタブ譜は不適切である。そこで我々は、フォーム変化後最低 D 時間の間は同一のフォームを用いなければいけないという制約を加える。本稿では、最低継続時間 D は全ての押弦可能フォームについて同一の値であると仮定する。

3.2 制約つき音高推定

各時刻 t での最適押弦フォーム \hat{p}_t を、上記3つの制約下で推定する (図2)。本手順は重み付き有向グラフ上での最長経路問題として、動的計画法を用いて解くことができる (図3)。 v_{tp} を時刻 t 、フォーム p に対応する頂点とし、 e_{tupq} を頂点 v_{tp} から v_{uq} への有向辺とする。 e_{tupq} は以下のA, Bのどちらかを満たす場合のみ存在する。A) $p = q$ かつ $u = t + 1$, B) $p \neq q$ かつ $u = t + D$ であり t が発音時刻である場合。条件Aは、使用フォームが時刻 t で変化しない場合である。このとき、有向辺 $e_{t,t+1,p,p}$ の重みを N_{tp} とする。一方条件Bは、使用フォームが、時刻 t でフォーム p からフォーム q へ変化する場合である。このとき、有向辺 $e_{t,t+D,p,q}$ の重みを $\sum_{t'=t}^{t+D} N_{t'q}$ と定義する。本グラフ上での最長経路を求めることで、最適押弦フォーム系列 $\hat{p}_1, \dots, \hat{p}_T$ が得られる。

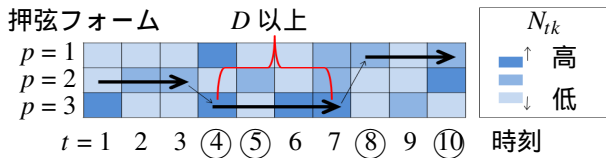


図 2: 最適押弦フォーム系列推定のイメージ. 矢印は推定されたフォーム遷移を, 丸で囲まれた数字は発音時刻をそれぞれ表わす. 本図は $D = 3$ の場合である.

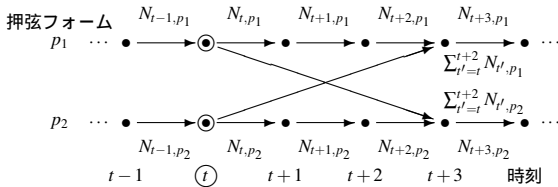


図 3: 最適フォーム推定に用いるグラフ図. 矢印は遷移可能なフォーム系列を表わす. 本図は $D = 3$, 押弦可能フォーム数 2 で, 時刻 t のみが発音時刻の例である.

最適フォーム系列の推定後, 各フォーム \hat{p}_t で発音不可能な全ての音高を排除する. すなわち, 時刻 t での音高 k の修正後のパワーを \tilde{N}_{tk} とすると, $k \in \mathbf{K}_p$ のとき $\tilde{N}_{tk} = N_{tk}$, そうでないときは $\tilde{N}_{tk} = 0$ とする. 元の N_{tk} の代わりにこの \tilde{N}_{tk} の値を用いて閾値判定を行うことで, 演奏不可能な音高を排除することができる.

3.3 タブ譜生成

提案手法では, 推定された各最適押弦フォーム \hat{p}_t および制約後の各音高のパワー \tilde{N}_{tk} を用いて, 各時刻で演奏された弦およびフレットが推定できる. さらに, この情報と拍検出 [5] を組み合わせることで, 演奏可能なタブ譜を生成することができる. 出力タブ譜の例を図 4 に示す. 本タブ譜を出力するにあたって, 楽曲のテンポやビートパターンが正確に推定されていることを仮定した.

4. 評価実験

本手法と従来法の性能を定量的に比較するため, 各時刻での音高を推定する実験を行った.

4.1 実験条件

実験データには, RWC 音楽データベース [4] のジャズ楽曲から 9 つのギターパートおよびポピュラー楽曲から 70 パートを抽出し, それらの冒頭 60 秒を用いた. 音響信号は MIDI シンセサイザー (YAMAHA MOTIF-XS) を用いて録音し, 20ms の時間分解能でウェーブレット変換した. 正解データは対応する MIDI データから作成した.

我々は発音時刻制約のパラメータ β の値を変えて, 提案手法による実験を複数回行った. 音高推定の閾値パラメータ α は, 各条件および各実験データで最適値となるように定めた. フォームの最低継続時間 D は, 200 ms とした. 比較のため, 同一の実験データを用いて従来法である LHA の評価も行った. 評価基準には F 値を用いた.

4.2 実験結果

実験結果を表 1 に示す. 両ジャンルにおいて, 提案法の性能が従来法を上回っていることが確認された. さらに, 平均発音数が多くなるほど, 提案法の従来法に対する優位性が増すことが分かった. これは, 提案法が演奏可能な音の組み合わせを適切に推定できているためである.

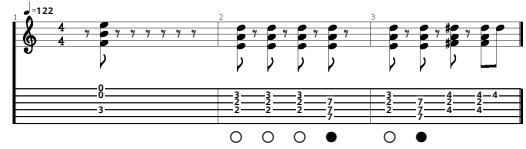


図 4: 出力タブ譜の例. 白丸と黒丸の箇所では, 同じ音高の組み合わせに対して異なるフォームが推定されている.

表 1: 音高推定の F 値. LHA は従来法 (制約なし) を表わし, 太字は平均発音数ごとの最大性能を表わす.

Genre	平均発音数	LHA	$\beta = 0.2$	0.3	0.4	0.5
Jazz	(2, 3)	0.701	0.721	0.719	0.713	0.691
	(3, 4)	0.583	0.645	0.652	0.652	0.648
	(4, 5)	0.515	0.648	0.643	0.664	0.640
	(2, 3)	0.615	0.649	0.651	0.644	0.638
Popular	(3, 4)	0.604	0.669	0.672	0.674	0.667
	(4, 5)	0.712	0.808	0.813	0.820	0.823
	(5, 6)	0.736	0.837	0.840	0.837	0.821

5. 考察

列挙された押弦可能フォームは, ギター演奏で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. 実際, 一般的なギターコード表 [6] に載っているフォームの約 98.7% (1186 / 1202) を列挙できている.

また, ある時刻での同時発音数が 5 音以下の場合, これら全ての音高を発音可能な押弦フォームは複数存在しうる (図 4). 本手法ではこれらのフォームに対し優劣を定めてはいない. しかし実際の演奏では, 演奏者は前後のフォームを考慮し, フォーム変化 (運指) の容易なものを選択することが適切である. このため, 今後以下の 2 点を検討する必要がある: (1) 各フォームの押弦難易度, (2) あるフォームから別のフォームへの変化の難易度.

6. おわりに

我々は既存の音高推定法 LHA に制約を加えることで, 人間の演奏に適したタブ譜を生成可能にし, さらに音高推定精度も向上させた. 今後の課題として, 指幅と押弦数以外のより厳密な判定基準の検討や, 他の押弦可能フォーム (親指を使ったフォームや人差し指以外を用いたバレーコードなど) の追加, 過剰フォーム (実際には演奏し難かったり, ほとんど用いられないフォーム) の削減などを行いたい. また押弦可能フォームの列挙法を変えることで, ギター以外の楽器の採譜にも応用していきたい.

なお, 本研究の一部は科研費 (S) No. 24220006 の支援を受けた.

参考文献

- [1] K. Yoshii and M. Goto: "A nonparametric Bayesian multipitch analyzer based on infinite latent harmonic allocation", *IEEE Trans. on ASLP*, Vol.20, No.3, pp.717-730, 2012.
- [2] 矢澤一樹, 阪上大地, 糸山克寿, 尾形哲也, 奥乃博: "押弦制約付きギター演奏自動採譜システム", 情報処理学会第 74 回全国大会, 4S-6, Mar. 6-8, 2012.
- [3] C. Duxbury, M. Sandler and M. Davies: "A hybrid approach to musical note onset detection", *Proc. DAFX*, pp.33-38, 2002.
- [4] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一: "RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース", 情報処理学会稿誌, Vol.45, No.3, pp.728-738, 2004.
- [5] A. Maezawa, H. G. Okuno, T. Ogata and M. Goto: "Polyphonic audio-to-score alignment based on Bayesian Latent Harmonic Allocation Hidden Markov Model" *Proc. ICASSP*, pp.185-188, 2011
- [6] "GuitarGearHeads - Gear Up Your Sound - Tiny Content", <http://www.guitargearheads.com/modules/tinycontent/index.php?id=5>, 2012.