

# 演奏動画を用いた作曲支援のためのバイモーダル型 TAB 譜採譜システム

## A Bimodal Music Dictation Method for Composition Support by using Guitar Performance Video

山上泰志

Yamagami Taishi

法政大学情報科学部デジタルメディア学科

E-mail: 08k1142@stu.hosei.ac.jp

### Abstract

This paper proposes a bimodal musical score dictation system for guitar performances. The system consists of two parts: recognition and transcription. In the recognition part, pitch is detected via a divided pick up, and fingering is estimated from color markers on fingers. In the transcription part, misrecognition of the previous part is corrected by rules based on attributes of guitar performance, and then the recognized score is output. Recognition accuracy was evaluated in a variety of conditions; performer, playing style and tempo were changed. As a result, the error rate of a whole of the system with and without rule were 50.4 percentage point and 75.1 percentage point in average, respectively. From this result, it can be concluded that correction based on the rules was effective.

### 1 まえがき

youtube やニコニコ動画など動画共有サイトを利用する人が増えている。中でも演奏動画やボカロイドの人気の高まり、演奏や作曲に興味を持つ人が増えている。

作曲を行う時、いきなり一曲を作り上げず、細かいメロディを思いつくたびに記録し、組み合わせてひとつの曲を作り上げる。メロディを記録する時、経験や知識があれば採譜や録音するのは容易だ。しかし初心者には採譜に時間がかかる上、楽譜の完成度も低い。また、録音した場合も、必要に応じて何度もメロディを聞き取ることで余計な手間が掛かってしまう。これらは初心者が作曲を行う妨げとなる。そこで本研究では、ギターの演奏動画を用いて自動採譜するシステムを提案する。

### 2 自動採譜

自動採譜とは、入力された演奏に対して音高やテンポを推定し楽譜を出力するシステムである。

楽譜には五線譜や一線譜などさまざまな種類がある。中でもギター用の楽譜として多く使用されているものに TAB 譜がある (図 1)。五線譜と違い、ギターの押弦すべき位置情報も記載されており直感的でわかりやすい。そこで本研究ではシステムの出力として TAB 譜を採用する。

TAB 譜を採譜するには、押弦位置の記述が必要となってくる。そのため、音高だけでなく押弦位置の推定も行わなくては



図 1. TAB 譜

はならない。従来研究では、ギター演奏音に対する自動採譜システムが考案されている [1][2]。[1] では潜在的調波配分法 (LHA) を用いて自動採譜を行う手法が用いられている。また、[2] では、[1] の問題点である、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に対して、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のルールを加えることで演奏に適した採譜を行う。さらに、手とギターの物理シミュレーションを使った手法 [9] もある。しかし、ギターには同じ音高が出せる位置が複数存在し、手の大きさや演奏の癖によって押弦位置に個人差があるため、これらの手法では対応できない。本研究では演奏動画を使うことで押弦した手の位置を認識し、ユーザの個人差を考慮した楽譜を出力する。

### 3 バイモーダルシステム

本研究は図 2 のように大きく分けると認識部、採譜部の 2 つの項目からなる。そのうち認識部では、入力したギター演奏動画に対してバイモーダル音声認識を行う。バイモーダル音声認識とは、音声データだけでなく、画像データからも特徴量を抽出する手法である。これにより、音声データだけより高精度に認識できる。しかし従来研究では、雑音下での認識率向上のために使われ、発話音声と唇の動きや口腔、歯の画像情報を入力し語句情報を出力するものが多く、楽器音に対して行った例は少ない。本研究では、耐雑音性ではなく、採譜に必要な特徴量を多く得るために行う。音声と画像は別々に処理を行い、それぞれの特徴量から推定を行う。

#### 3.1 認識部

認識部は、音声認識部と画像認識部に分かれる。音声認識部では、音声データから音階の推定を行い、画像認識部では、手やフレットの位置からの押弦位置の推定を行う。

##### 3.1.1 音声認識部

音声認識部では、演奏データから音階を推定する。本研究では、[5] で使用されている Roland 社のディバイデッドピック



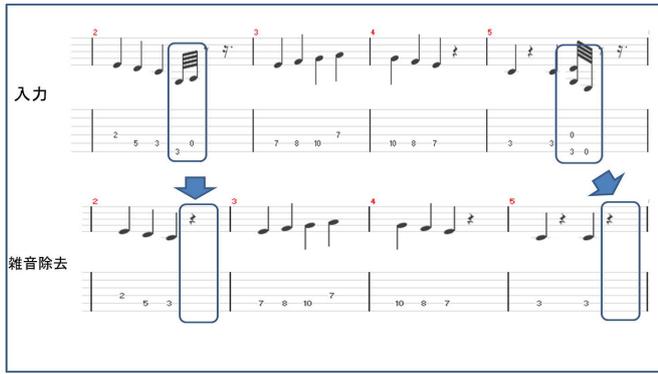


図 5. 雑音除去

的に、ギター演奏時に不必要な音がならないよう、発弦しない弦はミュートする。しかし、ミュートが不十分な場合、離弦時の音や前後の弦の音が鳴ってしまう。そこで本研究では、これらは弾いた音に比べて短い音であると仮定し、予めテンポと最小音符長を入力することで除去している。例えば、最小音符長が4分音符の曲を演奏し、36分音符の雑音があった。この場合、図4のように入力音にある36分音符の雑音が除去される。また、ギターの演奏において、和音は、各音が必ずしも同時に発音されているわけではない。各音をすばやく演奏することで和音として表現している。これを楽譜上に和音として表現したい。そこで、次の音のノートオンがノートオフより早く、ノートオンと次の音のノートオンの間隔がせまいものを和音と定義する。ここで、ノートオンとはMIDIの各音の発音する時刻、ノートオフとはMIDIの各音の切れる時刻を表す。これにより和音とアルペジオの区別が実現できる。

### 3.2.2 画像認識部での演奏ルール

演奏ルールを使い、画像認識部での認識結果に対して運指を補正する。画像認識部での認識で、指の角度や光からカラーマーカーが認識できない場合がある。これに対して、前後の指の位置から補正する。例えば、中指のマーカーが認識できておらず、前後の指の認識位置が1フレットの開きがある場合、中指の押弦位置は間のフレットとする。前後の指の認識位置が隣り合っている場合、前後のフレット内での指の位置関係から中指の位置を決定する(図5)。また、同フレット上でマーカーが重なっている場合、二つのマーカーの色が続いているとして下にあると考えられる認識面積の小さい指を運指とする。マーカーが重なっていない場合、弦方向の位置から認識面積の大きい指を運指とする(図6)。

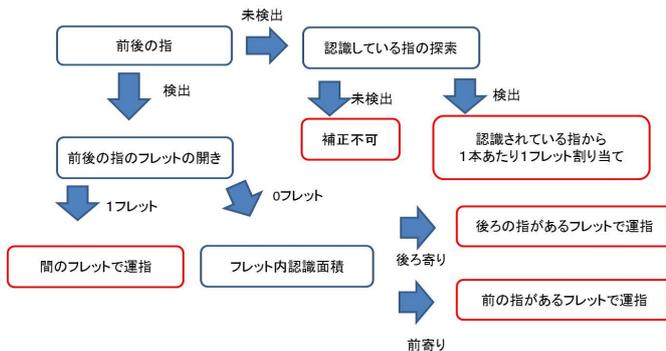


図 6. 未検出指検出チャート

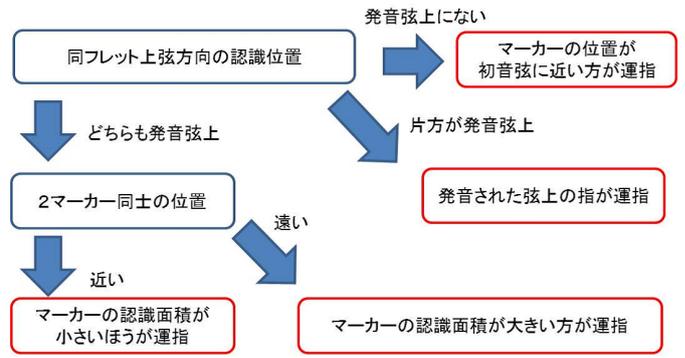


図 7. 同フレット上運指決定チャート

## 4 評価

実装したシステムの認識率を調べるため評価実験を行う。8小節の奏法の異なるフレーズを3種類用意し、ギター暦6年の被験者Aとギター暦4年の被験者B、ギター暦2年の被験者Cに弾いてもらう。それらに対して出力された楽譜の音高と押弦位置の正答率で評価する。楽譜の正答率が8割以上の場合、実用性があると判断する。それに加えて、演奏ルールの有無での正答率も比較する。

### 4.1 評価実験曲

図7は、評価に使った曲である。Aの曲はヘビーローテーションの冒頭サビの後の8小節で、単音のみで構成されている。Bの曲は雪どけの冒頭の8小節でアルペジオや和音、単音を含む。Cの曲はキミだけにはの冒頭の8小節で和音のみで構成されている。ロックバンドの曲でよく見るパワーコードを多用したフレーズである。

	曲名	作曲者
A	ヘビーローテーション	山崎耀
B	雪どけ	前田泰伸
C	キミだけには	花男・そら坊

図 8. 実験曲

### 4.2 評価実験結果

図8は曲ごとに演奏ルールの有無を比較した平均エラー率である。単音のみの場合が一番エラー率が低く、和音のみの場合がエラー率が高い。演奏ルールを使った場合の平均エラー率は50.4%、使わない場合の平均エラー率は75.1%だった。ギター用の演奏ルールを適用することで平均で21.5%誤認識を減らすことができた。

図8は、画像処理部の性能を評価するため曲Aに対して演奏データを使った場合とMIDI正解データを使った場合とのエラー率を比較したものである。図9は、曲Aに対しての音声データのエラー率である。

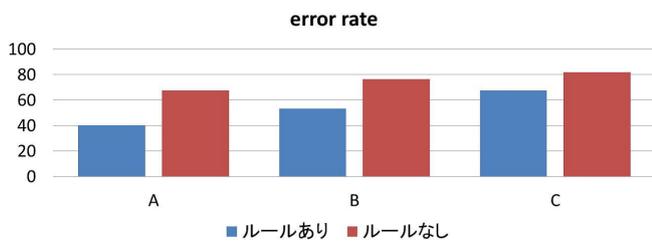


図9. 演奏ルールの有無での比較

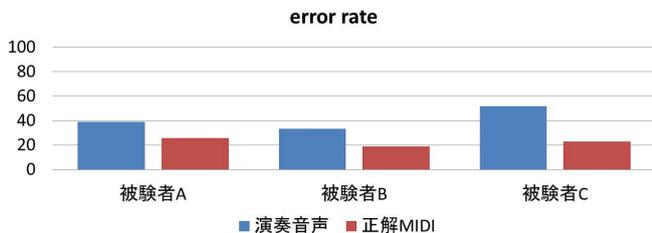


図10. 正解MIDIデータと演奏データでの比較

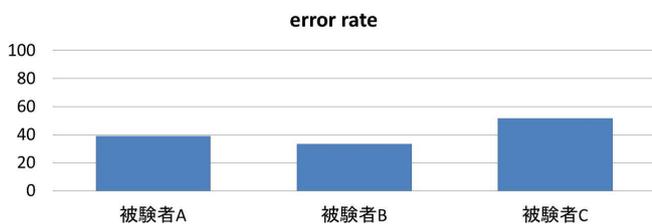


図11. 音声認識部エラー率

#### 4.2.1 誤認識パターン

■和音の誤認識 ギターの場合、低音弦と高音弦を弾き分けることで緩急を付けるため、与えた楽譜通りにいかない場合が多い。また、ミュートした弦からも音を拾ってしまう場合が多い。

■2音として認識 演奏音の途中で音が途切れて認識され2音として認識される場合がある。このように、正解データより多く音が検出され、楽譜としてのズレが生じる場合がある。

■指の誤認識 隣り合う指の位置が同フレット上で近かったためにマーカーの位置を誤認識し、そのせいで他の指に対してうまく演奏ルールが働かず運指が未検出になる。また、奏者によっては指を立てるためマーカーの認識がうまくできない。

#### 4.3 考察

本システムでは最小音符長を指定することで雑音の除去を行っているが、音間が狭くなるにつれて最小音符長に対して雑音が長くなってしまったため認識率が下がる。また、ノートオンやノートオフを比較する和音の認識も音の繋がりが曖昧になってしまう。これに対して調判別やコード進行の推定を導入することで、認識率を向上できる。

また、マーカーが検出できない場合がある。誤認識のパターンに示したように、演奏ルールで補正が行えていない場合がある。本システムではマーカーを利用しているため、指の角度や位置によっては認識できない。これに対し、手の形状を認識する従来研究で、シルエット画像を用いて手形状の認識を行う手法や、輪郭が不明瞭となる画像に対し、輪郭形状と位置と速度を特徴とするモデルと、位置と速度のみを特徴とするモデルからなる遷移ネットワークを学習することで形状追跡を可能とする手法を使うことで手の認識だけでなくコードの認識にも応用できる [3]。

図9より被験者Cの演奏データを使った場合とMIDI正解データを使った場合とのエラー率の差が大きい。そこで図10のように3人被験者に対して音声でのエラー率を比較したところ、音声でのエラー率が他の2人に比べて高かった。被験者Cの演奏データを確認したところ、ミュートを意識しておらず、不要な音が出ている。このような場合に雑音が処理できず音声認識部のエラー率が増えてしまう。このことから、システムを利用する際に演奏技術で影響がでる。

被験者3人に本システムやギターの演奏に関する聴取を行った。和音演奏時にすべての音を鳴らすように意識するかという問いに対して、特に意識しない、や緩急を意識するといった意見があった。カメラを固定した状態での演奏に対しては、弾きにくいという意見をもらった。

これらの意見から、和音の認識とギターの固定に対して、前後の音や上述した手形状の認識から補正することで認識率が向上できる。

## 5 おわりに

本研究では演奏動画を使うことでバイモーダルに認識を行いギター用の楽譜を出力するシステムを考案した。現段階でのシステムは音声認識部ではピックアップの性能に左右される。また、画像認識部ではカメラを固定して認識を行っている。今後はピックアップ無しの楽音の認識と、カメラの固定を行わずフレットを認識するシステムを目指す。

## 参考文献

- [1] 矢澤一樹ら他 “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”第75回情処学全大
- [2] K. Yoshii and M. Goto: “A nonparametric Bayesian multipitch an-alyzer based on infinite latent harmonic allocation”, IEEE Trans. ASLP, Vol.20, No.3, pp.717,730, 2012.
- [3] C. Duxbury, M. Sandler and M. Davies: “A hybrid approach to musi-cal note onset detection”, Proc. DAFx, pp.33-38, 2002.
- [4] 竹川佳成ら他: 鍵盤楽器のための実時間運指取得システムの構築, コンピュータソフトウェア, Vol.23, No.4,pp.51,59.2006.
- [5] 澤光映ら他”演奏ルールを用いたウッドベースのための実時間運指取得システムの設計と実装”コンピュータソフトウェア 27(1), 56-66, 2010-01-26
- [6] 田村哲嗣ら他”マルチモーダル音声認識における音声と画像の同期に関する調査” IEICE Tech.Rep. SP2008-70(2008-11)
- [7] 浜田康志ら他”手話認識のための複雑背景で高速に運動する手話形状推定” 信学誌 D vol.j90-D No.3 pp.617-627
- [8] 本川洋一 齊藤英雄”ギター演奏支援のための構造特徴の追跡を利用したマーカレスAR表示”日本VR学誌 13(2), 267-277, 2008-06-30bibitem
- [9] 藤原創太ら他”Fingerring Simulator:ギター単旋律の運指推定”2008 情処学研報