

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*

山上 泰志 (法政大学情報科学部), 伊藤 克亘 (法政大学 情報科学部)

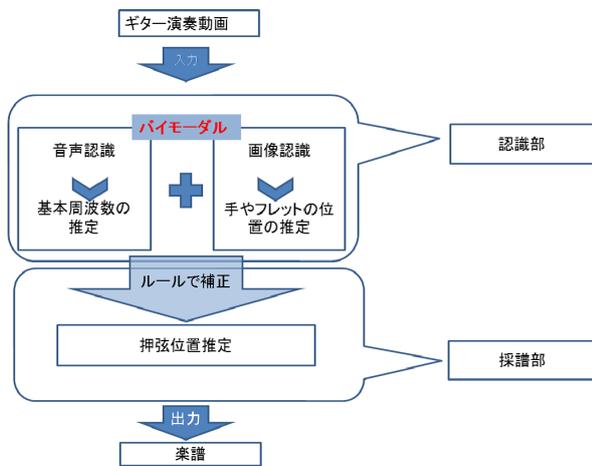
1 まえがき

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になり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。

ギター用の楽譜としてTAB譜が多用されている。これは、五線譜と違い、ギターの押弦すべき位置情報も記載され、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。

2 バイモーダルシステム

本研究は下図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。



2.1 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。

音声認識 音声認識部では、演奏データから音階を推定する。録音には、Roland社のディバイデッドピックアップを使用してMIDIデータに変換し、音高を得る。このピックアップは各弦ごとに処理を行い、MIDIデータに変換してくれるため音声認識部での誤認識を減らすことができる。ピックアップから得たMIDIデータ

から周波数に変換し各弦ごとのフレット番号を得ることによって押弦する位置を推定できる。

画像認識 画像認識部では、手やフレットの位置から押弦位置と押弦した指を推定する。手法はカラーマーカーを利用し手の位置を認識するカラーマッチングを行う。

本研究ではカメラとギターを固定して撮影する。カメラに対してギターのフレットが正面を向くようにする。カラーマーカーの認識では、カラーマーカーのサンプル値を取得し、範囲を指定することで認識を行う。これをフレームごとに繰り返し、各指がフレットに対してどの位置にあるかをフレームごとに取得する。フレット間の比率を用いて各フレットの長さを計算する。各フレット内のマーカーの認識面積を比較し、認識面積が広いフレットをその指の運指位置とする。さらに、弦方向の認識面積でもフレットの場合と同様に比較し運指位置を決定する。

画像と音声の統合 統合する際には、フレームレートの違いや同期のずれといったバイモーダル音声認識に特有の問題が認識に悪影響を及ぼす。そこで本研究ではカメラに録音された音声データとピックアップから録音された音声データに対して相互相関を用いて二つのデータの類似度から演奏開始時間のずれの補正する。本研究ではピックアップでの録音は発音と同時に開始するためMIDIデータの先頭に無音部分はない。よって動画側のデータをMIDIの発音タイミングにあわせるため動画の演奏開始までの区間を省いて同期する。

2.2 採譜部

採譜部では、演奏ルールを用いて音声、画像それぞれに対して押弦位置の補正を行ってから採譜を行う。

ギターの構造や人間の手の構造から認識部で得た情報の補正が行える。演奏ルールを用い、動画から運指を推定できる[1][2]。本研究では、ギター演奏を考慮し新たに本システム用のルールの検討を行う。同じ弦楽器であるウッドベースのルールの手の構造に関するルールを本研究で応用する。以下で演奏ルールについて説明する。

音声認識部での演奏ルール 音声認識部での補正では、和音とアルペジオの区別と雑音の除去を行う。まず、ピックアップから得た各弦ごとのMIDI情報を周波数に変換し、押弦したフレットに変換する。ピックアップは弦の移動時の小さな音も拾ってしまうため補正が必要である。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する。また、ギターの演奏において、演奏音が楽譜に対応していない場合がある。例えば、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。

* A Bimodal Music Dictation Method for Composition Support by using Guitar Performance Video.: Taishi Yamagami (Hosei Univ.) et al.

これを楽譜上に和音として表現するために、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。

画像認識部での演奏ルール 画像認識部での補正では、認識結果に対して運指の補正を行う。画像認識部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を行う。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。また、マーカーが重なっている場合、同フレット内に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。

3 評価

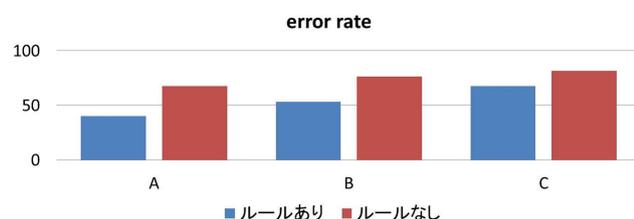
実装したシステムの認識率を調べるため評価実験を行う。8小節の奏法の異なるフレーズを3種類用意し、ギター暦の異なる3人に弾いてもらう。それらに対して出力された楽譜の音高と押弦位置の正答率で評価する。楽譜の正答率が8割以上の場合、実用性があると判断する。それに加えて、演奏ルールの有無での正答率の比較も行う。

3.1 評価実験曲

Aの曲は単音のみで構成されている。Bの曲はアルペジオや和音、単音を含む。Cの曲は和音のみで構成されている。

3.2 評価実験結果

下図は、評価実験を行い、各曲ごとに演奏ルールの有無を比較した認識結果である。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。



3.2.1 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 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いう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、認識率を向上させられる。

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]。

被験者3人に対して、和音演奏時にすべての音を鳴らすように意識するか聴取を行ったところ、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。和音の認識では、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を向上させられる。

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、画像認識部ではカメラを固定して行っている。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] 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.
- [2] 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
- [3] 浜田康志ら: 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