

# エレキベース演奏動画に対する譜面自動付加システム

関西学院大学 理工学部 人間システム工学科 片寄研究室 27019719 筒井勇多

## 1 はじめに

エレキベースの練習媒体は、かつては教則本を見ながら練習したりレッスンに通ったりする方法が主流であった。しかし、近年は動画投稿サイトに投稿された、演奏動画に譜面が付加されている譜面動画を見て練習する方法が主流である。しかし、譜面動画が投稿されている曲は有名曲がほとんどであり、演奏動画のみ投稿されている曲も多く、そのような曲の譜面が見たい人は多く存在する。また、投稿者の観点では、演奏動画に比べると譜面動画を投稿するには譜面の作成、演奏映像と譜面の同期の手順が増え、手間がかかる。そのため多くの曲は譜面動画が存在せず、演奏動画のみ投稿されている。

そこで、本研究では演奏映像に譜面を自動付加するシステムを開発する。このシステムにより視聴者は演奏動画の譜面を見ることができ、投稿者はより手軽に譜面動画を投稿することができるようになることを目標とする。なお、本研究での譜面はエレキベースの譜面として一般的に用いられている、押弦位置を示したタブ譜(図1)を用いる。これに伴い、エレキベースは楽器の特性上一つの音に対し押弦位置が複数存在することがあるため、譜面を生成するには押弦位置を推定する必要がある。

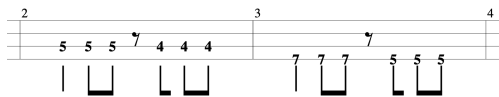


図 1: タブ譜

## 2 関連研究

関連研究に Paleari ら [1] の研究が挙げられる。彼らの研究では、アコースティックギターの単独演奏映像のオーディオ信号から音高を、画像情報から押弦位置を抽出し譜面を生成している。この研究では押弦せずに鳴らすことができる6種類のピッチ(開放弦)について、1音に対し複数の押弦パターンを提示してしまうエラーが発生している。また伴奏つきの入力波形に対応しておらず、本研究ではこれらの問題点を改善する。

Wang ら [2] の研究で開発した YOLOv7 [2] は画像中の物体の位置およびその物体が何であるかを推定する物体検出の機械学習モデルであり、実行速度が速いことが特徴である。

Défossez ら [3] の研究で開発した音源分離システム Demucs は、異種楽器の混合波形を楽器ごとに分離することに成功している。

## 3 研究概要

本研究では、エレキベースの演奏動画の入力に対し映像の下部に譜面を自動付加するシステムを構築する。システムの概要を図2に示す。入力動画の音声はベース単音のみでなく、異種楽器の混合音である場合も想定する。譜面は演奏動画が進むに従って自動で切り替わり、常に演奏中の譜面が表示されているものとする。本システムの対象者は、エレキベースの演奏者、または演奏動画投稿者とする。

### 3.1 採譜

本研究では、採譜は以下の二段階で構成される。

- (1) 演奏されている音高及び音価を推定する。
- (2) 指板上の手の位置を推定する。
- (3)(1) と (2) で得られた情報から譜面を生成する。

本研究では、(1) について、まず音源分離システムである Demucs [3] を用いることで入力波形からベースのみ抽出する。抽出したベースのみの波形に対し、音域を5弦24フレットのエレキベー

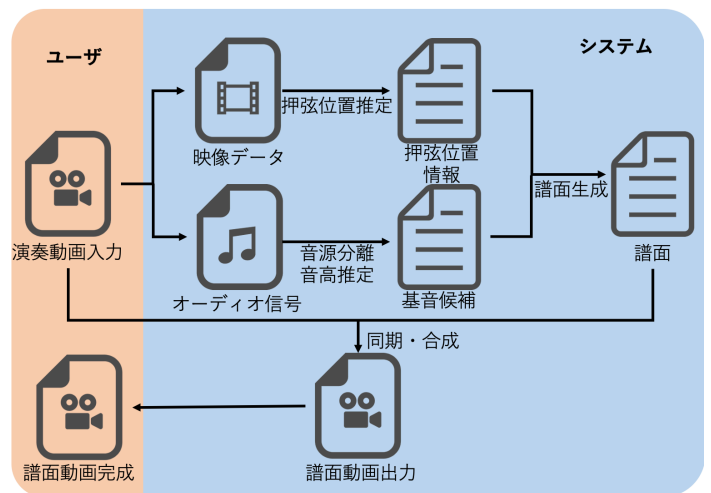


図 2: システム概要

スの音域45音(B0~G4)とした定Q変換を行う。この結果の各フレームに対して、45音それぞれの倍音構造を捉えるくし形フィルタを45個適用する。この結果より、当該フレームの音階を確率分布の形で得る。また、定Q変換の結果の各フレームごとの和をフレームのパワーとして扱い、これのピークにあたるフレームをオンセット、閾値未満であるフレームを無音として扱う。(2)について、本研究では、まず物体検出技術である YOLOv7 [2] を用いてベースの指板および左手を囲む矩形をそれぞれ推定する。推定結果の座標値に対し、指板領域の左上の点が(0,0)、右下の点が(1,1)となるような正規化を行い、この時の左手領域の中心座標値を指板上の左手の位置情報として扱う。(3)について、(1)で得た確率分布と(2)で得た左手の位置情報をもとに、押弦位置を最尤推定により求める。これの結果と(1)で得たオンセット及び無音の情報を元に譜面を生成する。

### 3.2 譜面の動画化

本システムでは、生成した譜面を映像の下部に動画と同期させて出力する。譜面と動画の同期は採譜中に音源から抽出した拍データをもとに行う。

### 3.3 評価実験

複数のエレキベース演奏者に対し実験を行う。それぞれの被験者による複数のフレーズパターンの演奏に対し本システムを適用し、手動での修正前の採譜結果と実際に演奏した譜面との音高、音価、撥音タイミングの採譜の精度により評価する。

## 4 現在の進捗と今後の予定

現時点で、3.1項(1),(2)のシステムが完成している。現在、(3)のシステム及びデータセットの調整を行っており、完了次第譜面化、評価実験を行う。

## 参考文献

- [1] M.Paleari, B.Huet, A.Schutz, and D.Slock. A multimodal approach to music transcription. *IEEE International Conference on Image Processing*, pp. 93–96, 2008.
- [2] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, 2022.
- [3] Alexandre Défossez, Nicolas Usunier, Léon Bottou, and Francis R. Bach. Demucs: Deep extractor for music sources with extra unlabeled data remixed. *CoRR*, Vol. abs/1909.01174, , 2019.