

# Instrumeteor : ギター演奏動画の制作支援システム

新 由衣\* 加藤 淳† 後藤 真孝†

**概要.** 楽器演奏者は自身の演奏を視聴してもらうために SNS や動画共有サイトで演奏動画を投稿している。本稿では、ギターに焦点を当て、ギター演奏動画における定量調査を行い、既存の演奏動画に用いられる複数の演出上の工夫を特定した。続いて、調査で判明した典型的な制作工程を進むにあたり従来の制作支援システムが抱える問題について議論する。そしてこの問題を解決するため、本稿では、ギター演奏動画の制作支援システム Instrumeteor を提案する。本システムでは、従来の制作過程にあった定型的な手作業の労力を削減し、簡単に魅力的な演奏動画を制作することを可能にした。

## 1 はじめに

多くの楽器演奏者が自身の演奏を視聴してもらうために SNS や動画共有サイトで演奏動画を投稿している。例えばニコニコ動画<sup>1</sup>上で、演奏動画に付与される事が多い「演奏してみた」動画のタグは、2020年11月現在25万件[1]を超えている。しかし、従来の動画制作支援システムは、演奏動画という特定の場面を想定していない。

そこで本稿では、演奏動画制作に特化した制作支援システムを提案する。また、演奏動画は楽器の種類によって着目点が異なり制作方法も異なるため、本稿はギターの演奏動画のみを扱う。

はじめに、ギター演奏動画における定量調査を行い、既存の演奏動画に見られる典型的な演出上の工夫を特定した。そして調査によって判明した定型的な作業を自動で行う、ギター演奏動画の制作に特化したシステム Instrumeteor を提案する。Instrumeteor は、楽曲の音楽構造を自動的に分析し、撮影した複数の動画トラックを楽曲の音楽的要素に合わせて単一のタイムラインに揃える。また、調査で判明した典型的なトランジション加工や歌詞のアニメーションを自動で適切な位置に生成する。これにより、演奏動画制作者の負担を削減し、より簡単に演奏動画を制作する支援を実現する。

## 2 動画制作に関する既存研究

動画制作に関する研究は、これまで数多く提案されている。Chiら[4]は、ウェブページの情報から動画を制作するシステム、Caoら[3]は、漫画のページの情報から動画を制作するシステムを実現した。そ

の他に、会話動画[8][5]、インタビュー動画[2]、解説動画[10]に焦点を当てた制作支援システムが実現されている。本研究では、楽器演奏動画の制作に焦点を当てる。

音楽動画 (Music Video; MV) の制作に関する既存研究も多数存在する。DanceReProducer[9]は Web 上の MV を再利用して新たなダンス動画を自動生成する N 次創作支援システムである。既存のダンス動画の映像部分を、新たな別の音楽に合うように切り貼りして映像を自動生成し、かつインタラクティブに編集することが可能である。土田ら[11]は、多視点ダンス映像のためのインタラクティブ編集システムを提案している。多視点カメラで撮影した複数の動画を音楽的要素に合わせて切り替え、簡単に魅力的な動画を制作できる自動編集機能を実現した。しかし、これらはダンス動画という特定の場面における限定的な情報を活用しており、演奏動画における情報は想定されていない。

Songle<sup>2</sup>[6]は、楽曲構造や階層的なビート構造等の音楽情景記述を音楽理解技術を用いて自動推定するサービスである。中嶋ら[14]は、Songle を利用し、音楽の構造情報を反映した MV を簡単に制作できる Web システムを提案している。また、MV の自動要約を行う研究[12]や演奏動画の盛り上がり検出に関する研究[13]が存在する。TextAlive<sup>3</sup>[7]は、歌声に合わせて歌詞のアニメーションが自動で表示される動画制作支援のシステムである。本システムでは、TextAlive を利用し、楽曲構造の取得、歌詞の表示を行う。

## 3 ギター演奏動画の制作について

本節では、対象とするギター演奏動画の一般的な演出上の工夫についての調査結果を報告し、既存の動画制作支援ツールが抱える問題について議論する

Copyright is held by the author(s). This paper is non-refereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

\* 筑波大学

† 産業技術総合研究所

<sup>1</sup> <https://www.nicovideo.jp>

<sup>2</sup> <https://songle.jp>

<sup>3</sup> <https://textalive.jp>

とともに、演奏動画制作支援ツールのデザイン指針を明らかにする。

なお、本節の分析にあたっては、第一著者が持つ演奏動画の投稿経験、および、Web上で収集した演奏動画の定量調査を参考にした。

### 3.1 ギター演奏動画の特徴

まず、一般的な演奏動画の特徴を特定、分析するために、YouTube<sup>4</sup>上で「弾いてみた ギター」を検索キーワードとして、2018年1月1日～2020年11月10日の期間に投稿されたギターの演奏動画を再生回数順にソートした上位100件を収集した。その際に、より多くの投稿者の傾向を分析するため、投稿者の重複が生じないように100件を収集した。

収集した演奏動画には、大別して8つの演出上の工夫が見られた(図1)。

- (A) 歌詞表示の加工：楽曲の歌声のタイミングと同期し、小節ごとに歌詞の文字アニメーションを表示していた。
- (B) トランジション加工：曲調の変わり目に用いられていた。また、収集した演奏動画には複数の種類のトランジションが見られた。種類の詳細については後述する。
- (C) 視点切り替えの加工：多視点撮影された演奏動画を楽曲の音楽的要素に合わせて切り替えていた。楽曲のサビでは、演奏全体の雰囲気映す傾向があり、ソロでは、運指に着目して映す傾向があった。そして、楽曲のリズムに合わせて演奏する体の姿勢を見せる箇所も存在した。
- (D) 複数の演奏動画を同時に表示する加工：部分的に、異なるパートの演奏を同時に表示していた。
- (E) 曲調に合わせて彩度を変化させる加工：曲調が静かになる箇所では、彩度が低くモノクロになる傾向があった。
- (F) スタンプや揺れのエフェクト加工：楽曲のビートやリズム等の音楽的要素に合わせたエフェクトを表示していた。
- (G) MVを表示する加工：演奏動画内に原曲のプロモーション用MVが表示されていた。
- (H) 楽譜を表示する加工：演奏動画内に楽曲の進行にあった楽譜が表示されていた。

また、収集した演奏動画には複数の種類のトランジションが存在した(図2)。トランジションは、サビの開始時刻等の曲調が変わる際に用いられている。今回収集した演奏動画には、(a)クロスディゾルブが

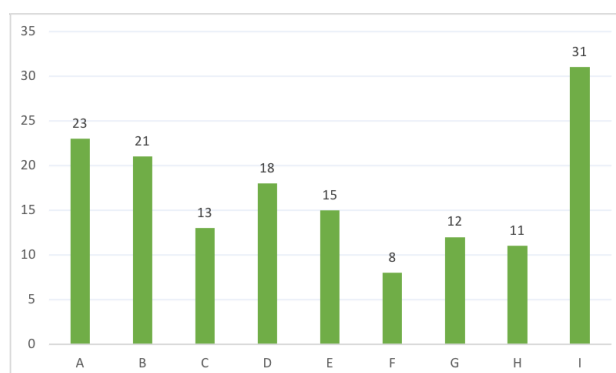


図1. A:歌詞表示, B:トランジション加工, C:視点切り替えの加工, D:複数の演奏動画表示の加工, E:彩度の加工, F:スタンプや揺れ等のエフェクト加工, G:MV表示, H:楽譜表示, I:加工無し

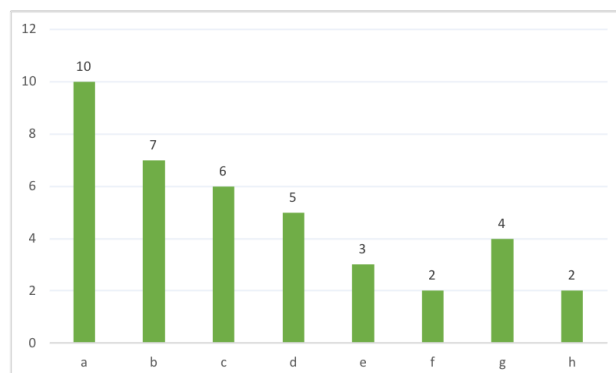


図2. a:クロスディゾルブ, b:ホワイトアウト, c:ブラックアウト, d:ワイプ, e:揺れ, f:クロスズーム, g:画像の表示, h:歌詞のみの表示

10件, (b) ホワイトアウトが7件, (c) ブラックアウトが6件, (d) ワイプが5件, (e) 揺れが3件, (f) クロスズームが2件であった。また, (g) 画像を一時表示するトランジションが4件, (h) 歌詞のみを一時表示するトランジションが2件である。

### 3.2 ギター演奏動画制作における問題

以下では、本稿で分析した演奏動画の演出上の工夫の項目(A)-(H)における問題についてそれぞれ説明する。(A)歌詞表示の加工は、常に楽曲の歌声と合わせる必要があり、細かいタイミング合わせとアニメーションのパラメタ調整の作業が必要となる。(B)トランジション加工は、サビの開始時刻等の曲調の変わり目、つまり楽曲の構造情報に合わせる必要があり、ユーザは複数の曲調の変わり目を把握し適切なタイミングに細かく調整する必要がある。

次に多視点で撮影した演奏動画の(C)視点切り替えの加工, (D)複数の演奏動画表示の加工は、楽曲

<sup>4</sup> <https://www.youtube.com>

の構造・中身をユーザが理解し、撮影した複数の動画トラックを単一のタイムラインに揃える作業を要する。(E) 彩度の加工に関しては、ユーザが曲調の変化・中身を認識して、曲調により適した彩度を考慮し、調整する作業が求められる。そして(F) スタンプや揺れ等のエフェクト加工は、楽曲のビートやリズム等の音楽的要素に合わせる必要があり、細かい調整の作業が必要となる。(G)MV 表示の加工は、ユーザが楽曲のMVをダウンロードする手間がかかり、(H) 楽譜表示の加工は、曲の進行に合わせて楽譜の一部を載せる必要があり、ユーザは楽曲の小節ごとに楽譜と照らし合わせて編集する作業が必要である。

これら(A)-(H)の演出上の工夫に要する作業は、従来の動画制作支援ツールにおいて手作業で行われているため、ユーザの負担が大きく編集に時間がかかる。また、従来の制作支援ツールは演奏動画という特定の目的に特化したものではないため制作に関する自由度が高く、特に多数の種類が存在するトランジション等の加工において、演奏動画により適したエフェクトを探す工程を要する。

本稿では、上記で説明した問題を解決する取り組みの第一段階として、(A) 歌詞表示、(B) トランジション加工、(C) 視点切り替えの加工、(D) 複数の演奏動画表示の加工、(F) スタンプや揺れ等のエフェクト加工、(G)MV 表示の加工に焦点を当て、これらの制作過程における問題を解決するシステム Instrumenteor を提案する。

## 4 Instrumenteor

### 4.1 ユーザインタラクション

ギター演奏者は、演奏風景全体が入る視点、運指に着目した視点、楽器の特定の箇所を映す複数視点で撮影を行い、同時にギター演奏の録音も行う。これらの動画、録音素材、楽曲の音源を本システムへの入力とする。

(A) 歌詞表示の加工は、従来は歌詞の表示タイミングの細かい調整とアニメーションのパラメタ調整の作業が必要であったが、本システムでは TextAlive の歌詞情報を利用し、歌声にあった歌詞のアニメーションを自動で適切な位置に表示する。(B) トランジション加工は、曲調の変わり目を認識して手動で調整する作業が求められる。種類も多数あるため、演奏動画に適したトランジションを探す必要があった。本システムでは、TextAlive から得られるサビの区間情報(各サビ区間の開始時刻と終了時刻の情報)を利用し、前述した調査で判明した典型的なトランジションが適切な位置に自動で生成される。また、トランジションの種類は修正可能であり、演奏動画に適したユーザ好みのトランジションに調整できる。

(C) 視点の切り替え加工、(D) 複数の演奏動画の表

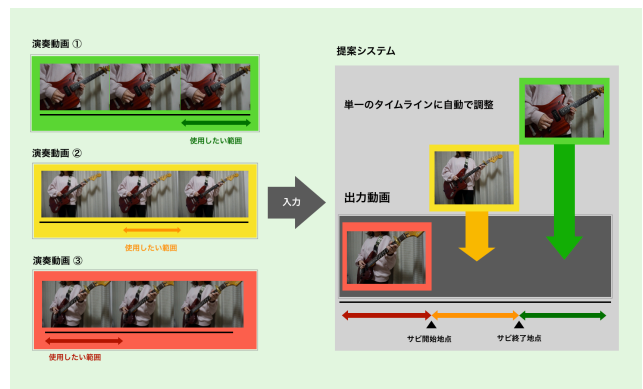


図 3. 多視点撮影された複数の動画を自動で単一のタイムラインに揃える。



図 4. 提案システムを用いて出力された演奏動画。上はサビ区間前、下はサビ区間中の動画の様子である。

示の加工は、複数の動画トラックを単一のタイムラインに揃える細かな手作業が必要であった。本システムでは、前述したサビの区間情報を利用する。ユーザはサビより前の区間に表示したい動画、サビの区間で表示したい動画、サビより後の区間で表示したい動画を選択する。その後選択した動画トラックが単一のタイムライン上の楽曲構造に合った適切なタイミングで自動的に挿入される(図3, 図4)。また、複数の演奏動画を同時に表示させるように指定すると、同様に自動で単一のタイムラインに揃えられて実現される。(F) スタンプや揺れ等のエフェクト加工は、ユーザが楽曲の音楽的要素に合わせて、手動で調整する必要があった。本システムでは、TextAlive から得られるビート情報を利用することで楽曲の音

楽的要素に適したエフェクトが生成される。(G)MV表示は、YouTube上の動画を直接埋め込むことで実現される。

以上の操作により、ユーザは楽曲に適した、より魅力的な演奏動画の制作を簡易に行うことが可能となる。

## 4.2 実装

本システムは、TextAliveの開発者向けAPI (Application Programming Interface)<sup>5</sup>を利用しており、ウェブ上で動作可能なアプリケーションである。また、動画のトランジションやエフェクト加工の実装には、p5.jsを使用している。

TextAliveで、YouTube等のMVを指定すると、その楽曲と歌詞の自動同期(時間的対応付け)とサビ区間検出の処理が行われる(実際にはTextAliveと連携している前述のSongle上で処理される)。本システムでは、サビ区間の開始時刻、終了時刻の情報を使用し、多視点撮影された演奏動画の切り替えを行う。その際、前述した調査によって判明した典型的なトランジションをランダムに使用するように実装している。また、TextAliveから歌詞情報を取得し歌声にあった歌詞のアニメーション表示を行っている。

## 5 まとめと今後の課題

本稿では、ギター演奏動画における定量調査を行い、従来の演奏動画に見られる一般的な演出上の工夫を特定した。そして、定型的な作業の労力を削減して簡単に魅力的な動画を制作できる、ギター演奏動画の制作支援システムInstrumeteorを提案した。

今後、本システムを用いて、既存の演奏動画制作支援ツールとの比較を行う被験者実験を予定している。また、オプティカルフローを利用して運指の動きが激しい動画、演奏全体の風景を映した動画を判定し、ユーザが選択する必要なく楽曲の区間に適した演奏動画を自動で挿入してくれる機能を検討している。最後に、本稿ではギターの演奏動画に焦点を当てた。将来、他の楽器の演奏動画の調査も行い、各々の楽器に特化した編集モードを実装する予定である。

## 参考文献

- [1] 人気の「演奏してみた」動画 252,355本 - ニコニコ動画. <https://www.nicovideo.jp/tag/演奏してみた>, (2020/11/14 確認).
- [2] F. Berthouzoz, W. Li, and M. Agrawala. Tools for Placing Cuts and Transitions in Interview Video. *ACM Trans. Graph.*, 31(4), July 2012.
- [3] Y. Cao, X. Pang, A. B. Chan, and R. W. H. Lau. Dynamic Manga: Animating Still Manga via Camera Movement. *IEEE Transactions on Multimedia*, 19(1):160–172, 2017.
- [4] P. Chi, Z. Sun, K. Panovich, and I. Essa. Automatic Video Creation From a Web Page. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20, p. 279–292, 2020.
- [5] O. Fried, A. Tewari, M. Zollhöfer, A. Finkelstein, E. Shechtman, D. B. Goldman, K. Genova, Z. Jin, C. Theobalt, and M. Agrawala. Text-Based Editing of Talking-Head Video. *ACM Trans. Graph.*, 38(4), July 2019.
- [6] M. Goto, K. Yoshii, H. Fujihara, M. Mauch, and T. Nakano. Songle: A Web Service for Active Music Listening Improved by User Contributions. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pp. 311–316, 2011.
- [7] J. Kato, T. Nakano, and M. Goto. TextAlive: Integrated Design Environment for Kinetic Typography. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 3403–3412, 2015.
- [8] M. Leake, A. Davis, A. Truong, and M. Agrawala. Computational Video Editing for Dialogue-Driven Scenes. *ACM Trans. Graph.*, 36(4), July 2017.
- [9] T. Nakano, S. Murofushi, M. Goto, and S. Morishima. DanceReProducer: An automatic mashup music video generation system by reusing dance video clips on the web. In *Proceedings of the 8th Sound and Music Computing Conference (SMC 2011)*, pp. 183–189, 2011.
- [10] A. Truong, F. Berthouzoz, W. Li, and M. Agrawala. QuickCut: An Interactive Tool for Editing Narrated Video. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, p. 497–507, 2016.
- [11] S. Tsuchida, S. Fukayama, and M. Goto. An Automatic System for Editing Dance Videos Recorded by Multiple Cameras. In *Proceedings of the 14th International Conference on Advances in Computer Entertainment Technology (ACE 2017)*, pp. 1–18, 2017.
- [12] C. Xu, X. Shao, N. C. Maddage, and M. S. Kankanhalli. Automatic Music Video Summarization Based on Audio-Visual-Text Analysis and Alignment. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '05, p. 361–368, 2005.
- [13] 小山 健一, 石先 広海, 帆足 啓一郎, 小野 智弘, 甲藤 二郎. E-041 演奏動画の盛り上がり検出に用いる特徴量の検討. 情報科学技術フォーラム講演論文集, pp. 305–306, 2011.
- [14] 中嶋 誠, 五十嵐 健夫. 音楽の構造情報を反映したミュージックビデオを簡単に制作・共有できるWebサービス. 日本ソフトウェア科学会第21回インタラクティブシステムとソフトウェアに関するワークショップ(WISS 2013) 論文集(デモ・ポスター発表), 2013.

<sup>5</sup> <https://developer.textalive.jp>