楽曲メディアコンテンツのビート特徴に合致したスライドショー動画メディアコンテンツ生成方式

山口 凜華† 岡田 龍太郎† 峰松 彩子† 中西 崇文†

†武蔵野大学データサイエンス学部 〒135-8181 東京都江東区有明 3-3-3

E-mail: † s2122069@stu.musashino-u.ac.jp,
{ryotaro.okada, ayako.minematu , takafumi.nakanishi }@ds.musashino-u.ac.jp,

あらまし 本研究では，楽曲メディアコンテンツのビート特徴に合致したスライドショー動画メディアコンテンツ生成方式について示す．本方式は，楽曲メディアコンテンツと複数の画像メディアコンテンツ群を入力とし，楽曲と映像が連動したスライドショー動画メディアコンテンツを自動生成する．本方式では，楽曲メディアコンテンツのリズムを特徴づけているドラムパートを抽出した上で，その音量の大きい箇所を時系列情報として取り出したものをビート特徴として定義し，その特徴に合わせて画像メディアコンテンツが切り替わるようなスライドショー動画メディアコンテンツを生成する．本方式を実現することで，ユーザは楽曲と映像が連動した動画メディアコンテンツを容易に作成することが可能となる．

キーワード 音声・音楽，マルチメディア，スライドショー動画メディアコンテンツ

1. はじめに

近年，YouTube などの動画共有サービスや Instagram などの画像の共有を目的としたソーシャルネットワーキングサービスにおいて，ショート動画と呼ばれる 90 秒以内の短い動画が若い年代を中心に流行している．その中でも，スライドショー動画メディアコンテンツと呼ばれる，音楽に合わせて画像を切り替える形式の動画が人気を博している．Instagram においては，スライドショー動画メディアコンテンツを手軽に作成することができる機能を内包するリールと呼ばれる機能が実装されている．

これらの流れによって，多くのユーザが自身の作成したメディアコンテンツを発信できるようになると，他のユーザよりも更に質の高い動画を作成したいと考えるユーザが現れる．そのようなニーズに合わせて動画編集ソフトもかなり身近な存在になりつつある．

スライドショー動画メディアコンテンツを作る際に，一般的に手軽に作成することができるタイプのものは，画像は一定の決められた時間間隔で切り替わるものとなっている．それに対して，動画クリエイターが作成した質の高いスライドショー動画メディアコンテンツでは，画像の切り替えのタイミングが一定ではなく，BGM のテンポやリズムに合わせて切り替わるため，視聴者は音楽と画像がマッチした感覚を味わうことができる．しかし，このような動画は動画クリエイターが手動で画像の切り替えタイミングを指定していることが多く，作成のためには多くの時間や労力が必要となる．

以上のことから，任意の音楽に連動して画像が切り替わるスライドショー動画メディアコンテンツの生成を支援するシステムを作ることができれば，コンテンツの作成にかかる時間や労力を低減させることが可能になると考える．

本稿では，楽曲メディアコンテンツのビート特徴に合致したスライドショー動画メディアコンテンツ生成方式について示す．本方式は以下の手順で構成される．まず，1 つの楽曲メディアコンテンツと複数の画像メディアコンテンツを入力する．入力された楽曲メディアコンテンツは音声分離機能と切り替えタイミング抽出機能によってリズムタイミングデータを抽出する．そのリズムタイミングデータは，入力された楽曲メディアコンテンツと複数の画像メディアコンテンツとともに，タイミングマッチ機能によって楽曲付き動画メディアコンテンツとして生成，出力する．本方式では，楽曲メディアコンテンツのリズムを特徴づけるものとしてドラムパートに着目する．ドラムパートの音量が大きい箇所を時系列情報として取り出したものをビート特徴として定義し，画像メディアコンテンツが切り替わるタイミングとして利用する．

本方式により，楽曲メディアコンテンツや画像メディアコンテンツを効率的に選択，統合を実現することで，ユーザに新たなメディアコンテンツを提供することが可能となる．

本稿は，次のように構成される．2 節では，関連研究について示す．3 節では，本方式である楽曲メディアコンテンツのビート特徴に合致したスライドショー動画メディアコンテンツ生成方式について示す．4 節では，本方式を実現する実験システムを構築し，実験

結果を示す．5 節では，本稿をまとめる．

2. 関連研究

本節では，2.1 節で楽曲と映像が連動した動画コンテンツ生成に関する研究について述べる．2.2 節ではスライドショーの自動生成に関する研究について述べる．2.3 節では，ドラムパートによる音楽のセグメンテーションに関する研究について述べる．2.4 節では，本研究の位置付けについて述べる．

2.1. 楽曲と映像が連動した動画コンテンツ生成に関する研究

平井[1]らは，音のエネルギーを表す RMS の変化に，映像の明滅や動きなどのアクセントを対応させるように，既存の音楽動画コンテンツを切り貼りする音楽動画自動生成システムを提案している．

F.Jonathan ら[2]は，セマンティックビデオコンテンツの選択と自動ビデオ構成による音楽スポーツビデオ生成システムを提案している．

2.2. スライドショーの自動生成に関する研究

舟澤ら[2]はユーザが指定した楽曲に対し，その歌詞情報を基に検索した Web 画像を用いて，スライドショーを自動で生成するシステムを提案している．

X.Songhua ら[3]は，個人の写真と音楽の歌詞との関連性を推測し，音楽に最適な個人写真を選択し，音楽スライドショーを生成するシステムを提案している．

2.3. ドラムパートによる音楽のセグメンテーションに関する研究

H.Tan ら[3]は，ドラムループパターンが音楽の顕著なリズム構造をどのように特徴付けるかを説明し，それの抽出アプローチについて示している．これは，抽出されたドラムループパターンが音楽のセグメンテーションに有用性があることを示唆している．

2.4. 本研究の位置付け

本方式では，選択した楽曲メディアコンテンツのドラムビートに合わせて画像メディアコンテンツ群を切り替えることで，スライドショー動画メディアコンテンツを生成するシステムを実現する．これにより，楽曲の音楽的な特徴に合わせたスライドショー動画を生成することが可能になる．

3. 楽曲メディアコンテンツのビート特徴に合致したスライドショー動画メディアコンテンツ生成方式

本節では，提案方式である楽曲メディアコンテンツのビート特徴に合致したスライドショー動画メディアコンテンツ生成方式について述べる．

本方式は，1 つの楽曲メディアコンテンツと複数のメディアコンテンツを入力し，楽曲メディアコンテンツのビート特徴に合致したスライドショー動画メディアコンテンツを出力する．

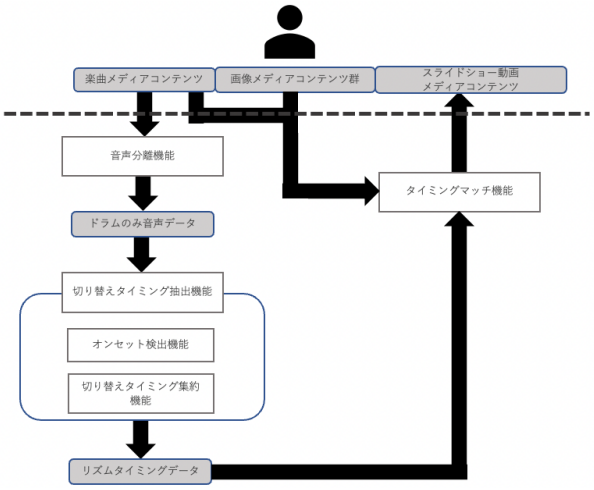

図 1　提案方式の全体像

3 章の構成について述べる．3.1 節では，提案方式の概要について述べる．3.2 節では，音声分離機能について述べる． 3.3 節では，切り替えタイミング抽出機能について述べる．3.4 節では，タイミングマッチ機能について述べる．

3.1. 提案方式の概要

本節では提案研究の概要について述べる．提案方式の全体像を図 1 に示す．本方式は，音声分離機能，切り替えタイミング抽出機能，タイミングマッチ機能から構成される．切り替えタイミング抽出機能はオンセット検出機能と切り替えタイミング集約機能からなる．

本方式の入力は楽曲メディアコンテンツと画像メディアコンテンツ群であり，出力は楽曲付きスライドショー動画メディアコンテンツである．まず，入力された楽曲メディアコンテンツに対して音声分離機能を適用しドラムの音声データを抽出する．次にドラムの音声データに対して切り替えタイミング抽出機能を適用し，リズムタイミングデータを抽出する．最後に，タイミングマッチ機能によって，このリズムタイミングデータに合わせて入力した画像メディアコンテンツ群を配置し楽曲と合成することによって，スライドショー動画メディアコンテンツを生成しユーザに提供する．

3.2. 音声分離機能

音声分離機能は，入力された楽曲メディアコンテンツをピアノ，ボーカル，ベース，ドラムのパートに分離し，ドラムパートのみを抽出する機能である．本機能の入力は楽曲メディアコンテンツであり，出力はドラムパートの音声データである．

この機能では，事前学習済みモデルを用いて音源を分離するツールである Spleeter[6]を楽曲メディアコン

テンツに適用させることで楽曲を 4 つのパートに分離し，ドラムパートの音声データのみを出力する．

本機能によって，画像メディアコンテンツの切り替えタイミングを判断するための音声データの出力が可能となる．

3.3. 切り替えタイミング抽出機能

切り替えタイミング抽出機能とは，ドラムの音声データからドラム音が強く鳴っている箇所を抽出する機能である．

本機能の入力はドラムパートの音声データであり，出力は時系列形式のリズムタイミングデータである．

本機能は，オンセット検出機能と切り替えタイミング集約機能からなる．

この機能によって，楽曲メディアコンテンツのビート特徴に合致したタイミングでの画像メディアコンテンツの切り替えの実現が可能となる．

3.3.1. オンセット検出機能

オンセット検出機能は，音声データからドラムの音が強く鳴っている箇所を抽出する機能である．オンセットとは楽器の音が鳴り始める箇所のことであり，そのオンセットを抽出する方式をドラムの音声データに適用することによって，打楽器であるドラムの強く鳴っている箇所を抽出する．

本機能の入力はドラムパートの音声データであり，出力は時系列形式のリズムタイミングデータである．

まずドラムパートの音声データにおいて，Onset Envelope と呼ばれる直前の音との音量の差分を算出する．その後，Onset Envelope の値が設定された閾値を超えた時点を Onset として検出する．

3.3.2. 切り替えタイミング集約機能

切り替えタイミング集約機能とは，ドラムパートの音声データにおいてオンセットが連続している箇所のタイミングを集約させる機能である．

本機能の入力はオンセット検出機能の出力のリズムタイミングデータであり，出力は集約されたリズムタイミングデータである．

本機能では，リズムタイミングデータにおいて隣り合う時点の時間間隔が 0.2 秒以下の場合，後ろのタイミングの除去を行う．

これにより，画像メディアコンテンツの切り替えが早すぎることによる画像の認識不可を防止することが可能となる．

3.4. タイミングマッチ機能

タイミングマッチ機能とは，リズムタイミングデータを基準に楽曲メディアコンテンツと画像メディアコンテンツ群を合成する機能である．

本機能の入力は，切り替えタイミング抽出機能の出力であるリズムタイミングデータと，本方式の入力である楽曲メディアコンテンツと画像メディアコンテンツ群であり，出力はスライドショー動画メディアコンテンツである．

以上の一連の機能によって，楽曲メディアコンテンツのビート特徴に合致したスライドショー動画メディアコンテンツ生成方式が実現する．

4. 実験

本節では本方式の評価実験について述べる．

実験ではプロトタイプシステムに対して，ドラムのテンポが早い楽曲と遅い楽曲を入力し，切り替えタイミング抽出機能の出力結果を確認する．これにより，ビート特徴の反映の検証を行う．

4.1. 実験環境

本方式のプロトタイプシステムを実装し，実験をおこなった．入力する楽曲として,「Baby I Need You」(作詞:Mr.Black 作曲:Ann sung hyun，Joosiq)のサビの 7 秒間と,「ccMixter」(作詞: Creative Commons)のサビの 11 秒間の 2 曲を使用した．「Baby I Need You」はテンポの速い楽曲であり,「ccMixter」はテンポの遅い曲である．

4.2. 実験:ビート特徴の反映の検証

本研究では，ビート特徴の反映の検証を目的としてテンポの早い楽曲と遅い楽曲を入力し，切り替えタイミング抽出機能における Onset Envelope とオンセットの様子を確認する．

4.3. 実験結果

本節では，実験結果とそれに対する考察を述べる．

「Baby I Need You」に対してオンセット検出機能まで適用した段階の Onset Envelope とオンセットを図 2 に示す．切り替えタイミング集約機能まで適用した段階の Onset Envelope とオンセットを図 3 に示す. 図 2, 図 3 の横軸は時間を表していて, 縦軸は Onset Envelope を表している．Onset Envelope は 0-1 に正規化した値を用いている．実線は Onset Envelope を示し，点線はオンセットの時点を示している．

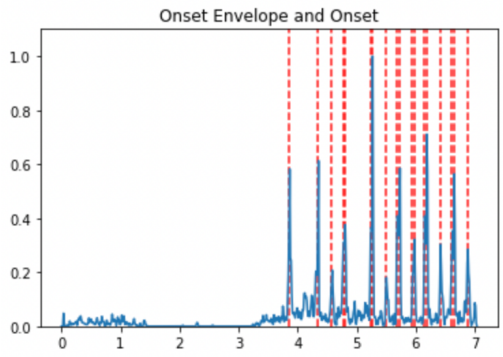

図 2 「Baby I Need You」のオンセット検出

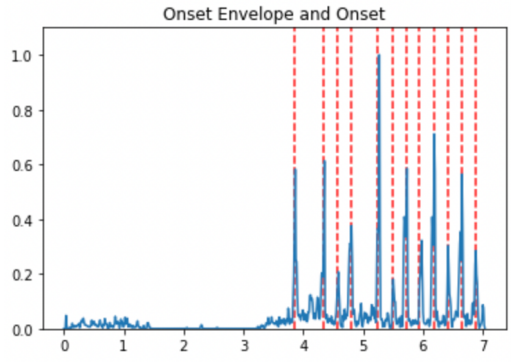

図 3 「Baby I Need You」の
切り替えタイミング集約

図 2 と図 3 を見比べると，図 3 では図 2 からオンセットが減少していることが分かる．オンセットの数は，図 2 のときは 18、図 3 のときは 12 であった。近くに存在するオンセットが切り替えタイミング集約機能によって集約され，極端に短いタイミングでの画像の切り替えが行われなくなったことが分かる．

「ccMixter」に対してオンセット検出機能まで適用した段階の Onset Envelope とオンセットを図 4 に示す．切り替えタイミング集約機能まで適用した段階の Onset Envelope とオンセットを図 4 に示す．軸は図 2，図 3 と同様である。「Baby I Need You」の場合と同様に，切り替えタイミング集約機能が有効に機能していることが分かる．オンセットの数は，図 4 のときは 26，図 5 のときは 18 であった．

続いて，「Baby I Need You」と「ccMixter」の切り替えタイミング抽出機能の出力の比較を行う．ただし，「Baby I Need You」の開始 4 秒まではドラムパートがないため正常な比較を行うため，4-7 秒の範囲に限定した結果を示し，考察を行う．

4-7 秒の範囲の切り替えタイミングの数を確認すると，「Baby I Need You」は 11 回，「ccMixter」は 7 回存在していることがわかる．

本実験では，画像の切り替えタイミングが楽曲のビート特徴と合致しているかを検証した．切り替えタイミング抽出機能の出力から，テンポの速い「Baby I Need You」は画像の切り替えタイミングが 11 回，テンポの遅い「ccMixter」は画像が 7 回存在することがわかった．このことから，テンポの早い楽曲ほど画像の切り替えタイミングが多くなるといえる．

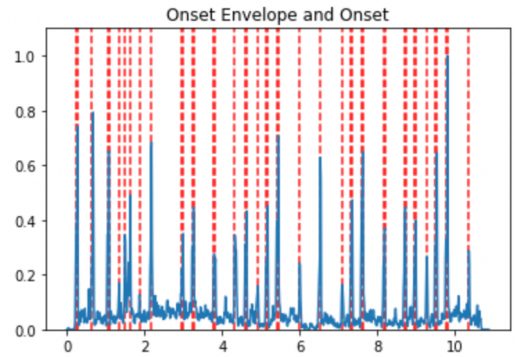

図 4 「ccMixter」のオンセット検出

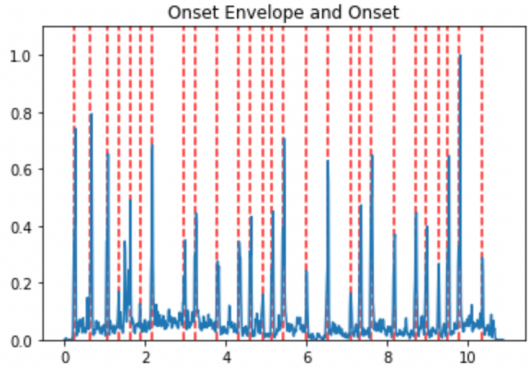

図 5 「ccMixter」
の切り替えタイミング集約

5. 終わりに

本稿では，楽曲メディアコンテンツのビート特徴に合致したスライドショー動画メディアコンテンツ生成方式について述べた．本方式は，楽曲メディアコンテンツと複数の画像メディアコンテンツ群を入力とし，楽曲と映像が連動したスライドショー動画メディアコンテンツを自動生成するものである．本方式では，楽曲メディアコンテンツのリズムを特徴づけているドラムパートを抽出した上で，その音量の大きい箇所を時系列情報として取り出したものをビート特徴として定義し，その特徴に合わせて画像メディアコンテンツが切り替わるようなスライドショー動画メディアコンテンツを生成することが可能となる．

本方式は，ユーザの保持している画像メディアコンテンツを使って容易にスライドショー動画メディアコンテンツを生成されるようになり，ユーザに新たなメディアコンテンツを提供することが可能になった．

また,本稿では，本方式を実現する実験システムを構築し，実験を行うことで有効性の検証を行った．今後の課題としては，ユーザが選択した画像メディアコンテンツ群の中からも，その楽曲の雰囲気に一致する画像メディアコンテンツを選択できるようにすることが挙げられる．

参　考　文　献

[1] 平井辰典，大矢隼士，森島繁生，“音楽と映像が同期した音楽動画の自動生成システム”， IPSJ SIG Technical Report，Vol.2013-MUS-99，No.26，2013/05/11．

[2] Wang, Jinjun, Changsheng Xu, Chng Eng Siong, Ling-yu Duan, Kong-Wah Wan and Qi Tian. “Automatic generation of personalized music sports video.” MULTIMEDIA，2005.

[3] 舟澤慎太郎，石先広海，帆足啓一郎，滝嶋康弘，甲藤二郎，“歌詞情報を利用した Web 画像・楽曲連動 スライドショー自動生成 スライドショー自動生成システム”, IPSJ SIG Technical Report，Vol.2010-MUS-84，No.13，2010/02/16.

[4] Xu, Songhua, Tao Jin and F. Lau. “Automatic Generation of Music Slide Show Using Personal Photos.” 2008 Tenth IEEE International Symposium on Multimedia，pp．214-219，2008.

[5] Tan, Hui Li, Yongwei Zhu, Susanto Rahardja and Lekha Chaisorn. “Rhythm analysis for personal and social music applications using drum loop patterns.” 2009 IEEE International Conference on Multimedia and Expo，pp．1672-1675，2009.

[6] Hennequin, R., Khlif, A., Voituret, F., & Moussallam, M. Spleeter: a fast and efficient music source separation tool with pre-trained models. Journal of Open Source Software, 2020.