

単一楽曲の切り貼りによる 動画の盛り上がり同期した BGM 自動付加手法

佐藤 晴紀[†] 平井 辰典^{**} 中野 倫靖^{***} 後藤 真孝^{***} 森島 繁生^{****}

[†]早稲田大学/JST CREST ^{**}早稲田大学 ^{***}産業技術総合研究所 ^{****}早稲田大学理工学術院総合研究所/JST CREST

1. はじめに

近年、動画共有サービスの利用者の増加に伴い、動画を共有する文化が広がっており、個人が動画制作に携わる機会が増加している。動画制作における重要な過程に BGM の付加があり、視聴者の動画に対する印象を強くさせる効果がある[1]。また、動画をより印象的に見せるために、編集者が「動画の決定的なシーンに楽曲のサビを合わせる」などのこだわりを持って BGM を付加することがある。こだわりを実現するために、楽曲の切り貼りによる編集(以降、再編と呼ぶ)がたびたび行われているが、再編した BGM に違和感が残らないようにしつつ、長さを調節するためには反復的な編集作業が必要であり、これらを同時に考慮した再編には多くの労力が必要となる。この問題を解決するため、これまでも動画に BGM を付加する手法の研究が行われてきた[2], [3]。しかし、既存研究の多くは BGM を自動で付加することを重視しており、こだわりを反映させて動画の特定のタイミングに楽曲の特定の箇所を付加するための手法は提案されていない。

本稿では、ユーザが動画と楽曲の指定箇所を入力として与えると、動画と楽曲の始端と終端を合わせつつ、動画と楽曲の指定箇所を同期させた BGM を動画へ自動的に付加する手法を提案する。それにより、楽曲の編集を経験したことのないユーザでもこだわりを反映した BGM の付加を可能とする。

2. BGM 付加手法

本稿における提案手法では、動画と楽曲の始端と終端を合わせつつ、動画と楽曲の指定箇所が同期した BGM の付加を実現させる。一般に動画と楽曲の長さ及び、指定箇所は時間的に異なるため、長さ及び指定箇所の同期は困難であるが、本手法では小節単位での楽曲の再編により実現する。

本手法の流れを図 1 に示す。まず、小節ごとに音響特徴量を抽出する(図 1, A)。次に、動画と楽曲の始端、動画と楽曲の終端を合わせ、動画と楽曲の指定箇所を入力する(図 1, B)。ここで、動画と楽曲の指定箇所の対は複数でもよい。そして、楽曲の第一小節の終わり、指定箇所の始まり、指定箇所の終わり、楽曲の最終小節の始まりの間を小節単位での再編

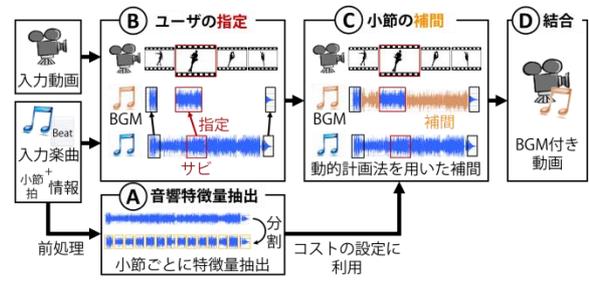


図 1. 提案手法の流れ

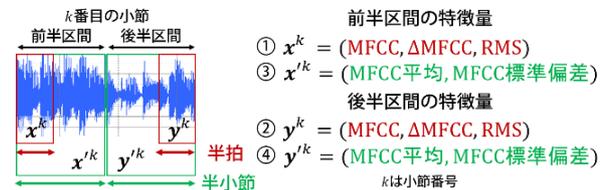


図 2. 音響特徴量の抽出

によって補間する(図 1, C)。最後に、動画の長さ、BGM の長さを一致させながら結合させることで、動画と楽曲のユーザの指定箇所が同期した BGM の動画への付加を行う(図 1, D)。

2.1 音響特徴量の抽出

小節単位で楽曲を再編する際、小節の繋ぎ目の違和感を軽減させるために、類似した小節同士を接続させる。そこで、楽曲を小節ごとに分割したのち、小節を前半/後半の二区間に分割し、それぞれの区間ごとに音響特徴量を抽出する(図 2)。楽曲の小節構造情報は既存のアノテーション情報を用いた。

小節の始まり/終わりから半拍分の MFCC(低 12 次元), Δ MFCC(低 12 次元), RMS(1 次元)を抽出する(図 2, ①②)。RMS とは音のエネルギーを表す特徴量であり RMS を E_{RMS} とした時、標本数 n と i 番目の波形の標本値 z_i を用いて式(1)のように表せる。

$$E_{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n z_i^2} \quad (1)$$

次に、前半/後半区間の MFCC の平均と標準偏差を抽出する(各低 12 次元, 図 2, ③④)。ここで、サンプリング周波数は 44.1kHz, 分析窓幅・シフト幅は 10ms とした。

2.2 小節の補間

小節の補間は図 3 に示すような経路選択により行う。まず、楽曲の第一小節と指定箇所と最終小節の間を補間する小節の候補として全小節を配置する。次に各小節間の繋がりやすさを式(2)に示すコストに基づいて決定する。

“An automatic BGM to video addition method synchronizing attention points of a video clips via rearrangement of a musical piece”

[†] Haruki SATO, Waseda University/JST CREST

^{**} Tatsunori HIRAI, Waseda University

^{***} Tomoyasu NAKANO, Masataka GOTO, National Institute of Advanced Industrial Science and Technology (AIST)

^{****} Shigeo MORISHIMA, Waseda Research Institute for Science and Engineering/JST CREST

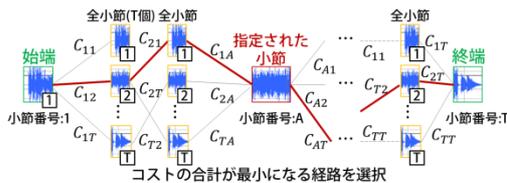


図3. 小節の補間方法

$$C_{jk} = \alpha \sqrt{\sum_{i=1}^{D_1} (y_i^k - x_i^j)^2 + \sum_{i=1}^{D_2} (y_i^k - x_i^j)^2} \quad (2)$$

$$\alpha = \begin{cases} 0 & (k = j + 1) \\ 1.0 & (\text{その他}) \end{cases} \quad (3)$$

ここで、 x, x', y, y' は 2.1 章で述べた小節の前半/後半区間の特徴量ベクトル、 j, k は小節番号、 D_1, D_2 は特徴量の次元数である。また、 α は元の楽曲における後続の小節の接続されやすさを制御する重み係数であり、元の楽曲と同じ接続である場合には 0 とする。最後に、楽曲の第一小節の終わりから指定箇所を経由し最終小節の始まりにかけて、動的計画法を用いてコストの合計が最小となる経路を決定し、その経路上の小節を接続することでBGMを作成する。

2.3 動画と楽曲の指定箇所の同期

本手法では、小節単位での再編を行うため、再編されたBGMの長さは小節の長さの自然数倍となる。従って、動画の指定箇所が小節の自然数倍でない場合、再編されたBGMにおいて動画と楽曲の指定箇所が最大で1/2小節分ずれてしまう。

そこで、楽曲の第一小節から指定箇所、指定箇所から最終小節の区間に相当する映像部を、ずれに応じて均等にフレームを複製・削除することで伸縮させる。これにより、動画と楽曲の指定箇所が同期したBGM付き動画の作成が可能となる。

3. 楽曲の再編結果の評価

本手法が適用可能な楽曲の特徴を調査するため、RWC 研究用音楽データベース[4]の音楽ジャンルデータベースの10楽曲に本手法を用いてBGMを作成し、自然さを順位法により評価した。被験者間の評価に相関があれば、順位中央値が上位である楽曲は自然であると言える。楽曲は、音楽ジャンルデータベースの大分類の中から1曲ずつ選曲し、小節構造情報はAIST Annotation for RWC Music Database (Beat structure)を用いた[5]。20代の男性4人、女性1人を被験者とし、BGMの長さは評価の簡易化のため冒頭30秒とした。被験者はBGMを聴取し、順位付けを行った(BGMには小節の繋ぎ目が複数箇所あるようにした)。2つの順位の間には、Spearmanの順位相関係数を用いた[6]。同順位のない順序ベクトル a, b の順位相関係数 ρ は式(4)の様に表示される。

$$\rho = 1 - \frac{6}{N^3 - N} \sum_{i=1}^N (a_i - b_i)^2 \quad (4)$$

ここで、 N は楽曲数、 a_i, b_i は順序ベクトル a, b における i 番目の要素の順位を表す。 N が10の場合は $\rho \geq 0.7333$ 、 $\rho \geq 0.5636$ ならばそれぞれ1%・5%水準で有意な相関がある。

表1. 順位相関係数 ρ の相関がある組の数と統計量

1%水準	5%水準	平均	標準偏差	最大	最小
3	8	0.6448	0.1278	0.8424	0.4545

表2. BGMに用いた楽曲と順位中央値

曲番号・ジャンル	中央値	曲番号・ジャンル	中央値
No.1・ポップス	4	No.53・クラシック	9
No.8・ロック	4	No.55・行進曲	9
No.21・ダンス	8	No.72・ワールド	2
No.29・ジャズ	3	No.87・声楽	7
No.38・ラテン	3	No.91・邦楽	7

被験者の組み(10組=5C2)ごとに順位相関係数を計算し、1%・5%水準で相関がある組の数と統計量を求めた結果及び、順位中央値を表1、表2に示す。表1より平均が5%水準の値を超えているため順位に相関があることが分かる。中央値が小さい楽曲はNo.72、No.29、No.38であり、同じメロディの繰り返しが多いため小節の繋ぎ目が自然になりやすいと考えられる。一方、中央値が大きい楽曲はNo.53、No.55であり、メロディやテンポが変化しやすいため、小節の繋ぎ目で違和感が大きくなりやすかった。

これらの問題については、音楽構造やテンポを考慮したコスト関数の設計により、対応可能となると考えている。また、歌声の有無を考慮していないため、小節の繋ぎ目で歌声が出現・消失する可能性がある(No.1及びNo.8の楽曲は歌声があるが、小節の繋ぎ目が歌声区間に現れなかったため中央値が小さくなったと考えられる)。以上より、本手法は同じメロディが多いインストゥルメンタルの楽曲であれば、自然な再編が可能であると考えられる。

また、本手法の性能として、指定箇所を増やすと小節の探索の自由度が減少してしまい、小節の繋ぎ目の違和感が大きくなりやすいという傾向があった。

4. まとめと今後の課題

本稿では、動画と楽曲の指定箇所が同期したBGMを動画へ自動付加する手法を提案した。

今後、テンポの変動や歌声がある楽曲に本手法を適用可能にするようなコスト関数の設計を行いたい。また、一度再編されたBGMに対して、ユーザが気に入らなかつた部分を他の候補から選択し置換することで容易に編集できるシステムとすることで、よりユーザ好みなBGMの付加を実現したい。

謝辞 本研究の一部はJST CRESTの支援を受けた。また、RWC研究用音楽データベースを使用した。

参考文献

- 1) 岩宮真一郎, “オーディオ・ビジュアル・メディアを通しての情報伝達における視覚と聴覚の相互作用に及ぼす音と映像の調和の影響,” 音響学会誌 Vol.48(9), pp.31-39, 1992.
- 2) Feng et al., “Auto-generation of professional background music for home-made videos,” Proceedings of the 2nd International Conference on Internet Multimedia Computing and Service, pp.15-18, 2010.
- 3) 小野ら, “音楽のムード分類結果を利用したホームビデオへの自動BGM付与・同期手法,” 情報科学技術フォーラム講演論文集 9.2(2010):295-296.
- 4) 後藤ら, “RWC研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース,” 情報処理学会論文誌, Vol.45, No.3, pp.728-738, 2004.
- 5) Masataka Goto, “AIST Annotation for the RWC Music Database,” Proceedings of the 7th International Conference on Music Information Retrieval, pp.359-360, 2006.
- 6) Kendall et al., “Rank Correlation methods,” 5th edition, p.260, Oxford University Press(1990)