

太鼓の達人の譜面難易度デザインにもとづく 楽曲中のキータイミングの推定に向けて

奥村 綾¹ 辻野 雄大^{2,a)} 山西 良典^{1,b)}

概要: 映像やコンサートでは、楽曲中の局所的な盛り上がりであるキータイミングに合わせて、エフェクトや会場の照明、花火などの演出が用いられている。本稿では、キータイミングを推定するために、音楽ゲームの一つである「太鼓の達人」の譜面を利用し、音響特徴量を入力として、ある時刻に対して全ての難易度で共通して存在する指示符のタイミングを推定する機械学習モデルを構築した。音楽ゲームの指示符のタイミングを、楽曲のキータイミングを推定するための擬似的な学習データとして利用するアイデアの有用性を検討した。

1. はじめに

Spotify や Apple Music などのサブスクリプションの音楽配信アプリでは、膨大な楽曲がデジタル化されており、多くの楽曲を容易に聴くことが可能である。音楽の楽しみ方としては、楽曲そのものを楽しむのみならず、楽曲を用いたミュージックビデオ (MV) のような映像作品や音楽ライブやコンサート、ダンスなども存在する。これらの音楽が付随したエンタテインメントにおいては、楽曲中の曲調や局所的な盛り上がりに合わせて、エフェクトや演出が考えられて作られている。MV であれば、演者 (キャスト) に近づくズームインや、グリッチエフェクトといった画面がまるで壊れたかのようなノイズを出すエフェクトなどの映像エフェクトがある。音楽ライブやコンサートでは、照明や花火、紙吹雪、水、炎、シャボン玉、レーザーなど数多くの演出が挙げられる。例えば、Perfume のライブ^{*1}では、プロジェクションマッピングやダイナミック VR などの演出が用いられている。他にも、ダンスでは決めタイミングで決めポーズをしたり、ダイナミックな動きしたり、フォーメーションを組んだりする。これらを総合すると、音楽にひもづいた映像などの視覚表現においては、視聴者にとって飽きが来ないよう、楽曲中のハイライトに相当するような局所的に盛り上がるタイミングに合わせて演出を付与するなどの工夫がなされていると言える。このような局所的な盛り上がりや、本稿では「キータイミング」

と呼称する。楽曲中のキータイミングを推定し提示することは、映像制作者やコンサートの演出家、ダンスの振付師を支援することにつながるだろう。

これまでも音楽情報処理の分野では、楽曲に存在する音響特徴を抽出し、楽曲構造を推定することで盛り上がり箇所を可視化する技術が提案されてきている。後藤ら [1] は Songle といった音楽音響信号理解技術を用いた音楽鑑賞サービスを提供した。このサービスは楽曲構造、階層的なビート構造、メロディライン、コードの 4 種類の音楽情景記述を自動推定して可視化するものになる。楽曲構造推定を通して、楽曲中のサビの自動検出が可能になる。サビに変わる瞬間は、映像にとって印象を強く残すことができると大野ら [2] と佐藤ら [3] の研究で述べられている。しかし、これらの技術ではサビのようなパート以外の局所的なキータイミングを推定することは難しいと言える。

本研究では、楽曲中に局所的に偏在するキータイミングを推定するため、「太鼓の達人」^{*2}の譜面を利用する。太鼓の達人は和太鼓 (打楽器) をモチーフとした音楽ゲームの一つであり、リズムに合わせて指定された指示符に沿って、和太鼓型のコントローラを叩くゲームである。指示符には、叩き方が異なる複数の種類が存在する。ある楽曲に対して用意された指示符の配置パターンを「譜面」と呼ぶ。太鼓の達人では、一つの楽曲に対して、指示符の配置タイミングや順序が異なる、複数の難易度の譜面が存在する。和太鼓というモチーフの特性上、太鼓の達人は音楽ゲームの中でも特にリズムを刻むことを重視して譜面が作成されていると考えられる。この点において、キータイミングと

¹ 関西大学総合情報学部

² 立命館大学情報理工学部

a) yudait@fc.ritsumeit.ac.jp

b) ryama@kansai-u.ac.jp

*1 <https://youtu.be/G0UcCrji0u0?t=148>, 2:29-10:15

*2 太鼓の達人シリーズ公式ポータルサイト, <https://taiko-ch.net/index.php>, 最終閲覧日 2022 年 7 月 26 日

表 1 対象とした映像エフェクトと意味

名称	動作
カット編集	必要な場面だけを抜き出して繋ぎ合わせる
ズームイン	画を拡大する
ズームアウト	画を縮小する
ディゾルブ	徐々に重ね合わせて次のカットへと移行する
VFX	実写映像と現実にはあり得ないものと組み合わせる

いう楽曲中の特定の瞬間を推定する本研究の目標達成に向け、太鼓の達人の譜面が有用と考えた。本稿では、ある時刻において、複数の難易度に共通して指示符が存在するタイミングは重要であるという仮定のもと、このような時刻をキータイミングとみなす。2節でこの仮定の検証を行う。

本研究は、音楽ゲームを対象とした計算科学研究として捉えることができる。代表的な課題として音楽ゲーム譜面の自動生成を試みたものが存在する。Chrisら [4] は Dance Dance Revolution (DDR)*³ という音楽ゲームを対象とし、音響特徴量を入力して譜面を自動で生成する手法を提案した。譜面の制作を「指示符をどのタイミングに置くのか」「どの種類の指示符を置くのか」という2つのタスクに分解し、それぞれ深層学習モデルを構築することで、自動生成を達成した。辻野ら [5] は DDR の譜面を操作を要求される頻度、リズム構成の複雑さなどの特徴に基づいて分類することで、「面白さ」の特性を客観的に表現し、特徴的な譜面を生成する手法を提案している。これらの研究は、音響特徴量から音楽ゲームの譜面生成を目標としている。一方、本研究は楽曲内に存在するキータイミングを推定するために音楽ゲームの譜面を利用しており、音楽ゲーム譜面の生成に限らない、音楽にひもづくコンテンツ全般の制作支援を目標としている。また、香川ら [6] は beatmaniaIIDX *⁴ という音楽ゲームの譜面自動生成に向け、楽曲中の重要音を抽出する手法を提案している。メロディやハーモニー、リズムなどの繰り返し現れるフレーズが重要音として捉えている。楽曲中から重要音を抽出するという点で、本稿の研究課題は香川らと通じる部分もあるが、香川らは単一の音の抽出を目標としており、本研究では楽曲中の盛り上がるタイミングを検出することを目標としている、という点において課題の性質が大きく異なる。

1.1 映像制作におけるキータイミング利用可能性の検討

音楽映像や MV に対するキータイミングの利用可能性を検討する。検討対象の楽曲は、太鼓の達人に収録されている楽曲のうち、公式の MV が存在する楽曲の中から、無作為に選定した。その結果、“CandyPop”^{*5} が検討対象となった。本分析においては、映像のエフェクトを「視聴者

*³ DanceDanceRevolution A20 PLUS - e-amusement, <https://p.eagate.573.jp/game/ddr/ddra20/p/>, 最終閲覧日 2022 年 7 月 26 日

*⁴ beatmaniaIIDX29 CastHour, <https://p.eagate.573.jp/game/2dx/29/>, 最終閲覧日 2022 年 6 月 28 日

*⁵ https://www.youtube.com/watch?v=wQ_POfToaVY

表 2 データセットの情報

楽曲数	90(2.6 時間)
難易度数	3
譜面数	270(7.2 時間)
音符数/秒	4.582

表 3 譜面データの数字記法と指示符の種類の対応

数字	名称	動作
0	音符なし	なし
1	ドンッ	面を叩く
2	カッ	ふちを叩く
3	ドンッ (大)	面を強く叩く* ⁶
4	カッ (大)	ふちを強く叩く
5	連打の始まり	区間中太鼓を連打する
6	連打の終わり	

が見飽きないように抑揚をつける映像演出のこと」と定義した。映像のエフェクトにはさまざまな種類が存在するが、今回の検討においては、表 1 に列挙する後付けのエフェクトを分析対象とした。

太鼓の達人における“CandyPop”の「かんたん」「ふつう」「むずかしい」3種類の譜面を分析し、全てに共通して指示符が存在した時刻(以下「全難易度共通操作タイミング」と呼ぶ)を集めたデータと、“CandyPop”の公式 MV から見られた映像エフェクト・時刻を記録したデータを使用し、分析を実施した。全難易度共通操作タイミングの定刻ちょうどに映像エフェクトが存在した数と、全難易度共通操作タイミングの前後1秒を含めた時刻において映像エフェクトが存在した数を計測した。

“CandyPop”に存在する45個の全難易度共通操作タイミングのうち、定刻に映像エフェクトが存在したものは4個、前後1秒以内に映像エフェクトが存在したものは37個であった。定刻ちょうどにエフェクトが少なかった要因として、“CandyPop”のMVの内容が、パフォーマンスメインだったことが挙げられる。アーティストのパフォーマンスを魅力的に見せるために、キータイミングで被写体がしっかり視認できるよう、エフェクトはキータイミングの前後に付与されていたと考えられる。例えば、サビ開始のダンスシーンにおいて、キータイミングの定刻はサビ開始のタイミングと一致していた。しかし、サビ開始のタイミングで、被写体がダンスを始める瞬間を映し出すため、カット編集はサビ開始の0.5秒前に行われていた。このような映像エフェクトが他にも多く見受けられた。

1.2 データセット

本研究では、太鼓の達人の譜面データを著者が書き起こ

*⁶ 叩いた強さを検知するセンサが搭載されている場合はセンサの値で判定されるが、センサが存在しない場合は「両手で叩く」という操作に代替される。

したものをデータセットとして用いる。データセットの情報表 2 に示す。太鼓の達人では、1 楽曲に対して難易度が異なる複数の譜面が収録されている。本研究では、1 楽曲に対し「かんたん」「ふつう」「むずかしい」の 3 種類の譜面を対象とする。譜面データは、太鼓の達人の譜面を集積し、画像の形でまとめている Web サイト^{*7}を参照し、著者らがテキストデータの形式で書き起こした。譜面テキストデータには、楽曲名、対応する音楽ファイル名、音楽ファイルの開始点に相当する時刻、テンポを示す BPM 値の 4 種類のメタデータに加えて、指示符の時系列を数字によって表現した文字列によって構成されている。表 3 に、指示符が要求する動作の種類を、データ内で対応する数字とともに示す。連打の指示符は、楽曲の収録時期などによって風船やくす玉など特殊な形で表現されることもあるが、一般的には黄色の横棒で示される。参照している Web サイトでは、指示符の表現形式によらず、区間の始点と終点の情報が示されている。本研究では全ての連打指示符が、最も一般的な黄色の横棒で表現されるものとした。

2. キータイミングに見られる太鼓の達人譜面の特徴についての仮説と検証

本研究で対象とする「かんたん」「ふつう」「むずかしい」の 3 つの譜面は、この順でより複雑な動作を指示する難解な譜面になっている。以下で、太鼓の達人の難易度デザインに関する仮説を立て、キータイミングと考えられるものを見出す。

2.1 仮説

同一楽曲に対して用意された 3 種類の譜面間の関係について、以下の仮説を立案した。

仮説 1 曲中に出てくる指示符の数が多ほど複雑で難解になる。逆を言えば、難易度が低い譜面では指示符の数が少ない。

仮説 2 かんたんな譜面では人間が知覚しやすいタイミングに指示符が配置され、難易度が上がるにつれ、裏拍やシンコペーションなど知覚しづらいタイミングに指示符が配置されうる。

仮説 3 仮説 1 および 2 が真ならば、複数の譜面で共通して指示符が存在するタイミングは、指示符の数を少なくしても残すべきタイミングであり、複雑なリズムで指示符を配置する際にも核となるタイミングである、重要なキータイミングであると言える。

2.2 検証

表 4 に、難易度ごとの 1 秒あたりの指示符頻度をそれぞれ示す。譜面の指示符頻度は「かんたん」「ふつう」「むず

表 4 難易度別情報

難易度	かんたん	ふつう	むずかしい
指示符の総数	8,891	13,010	22,220
指示符の総数/秒	0.95	1.39	2.37

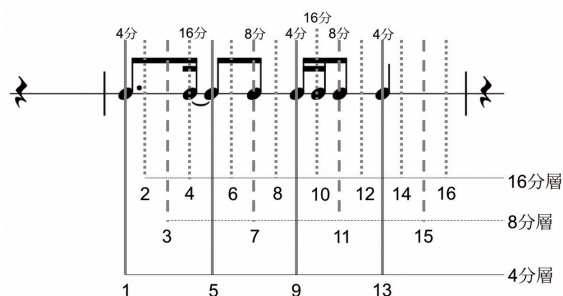


図 1 拍階層・拍位置の概念。4 分層に属する音符は、8 分層と 16 分層にも属するが、最下層である 4 分層に存在するものとする。各層の下に記した 1~16 の数字は、小節内での拍位置を示す。(文献 [5] の図 2 を引用)

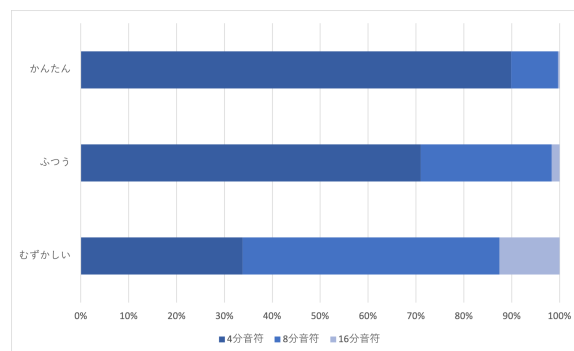


図 2 指示符が存在する拍階層の比

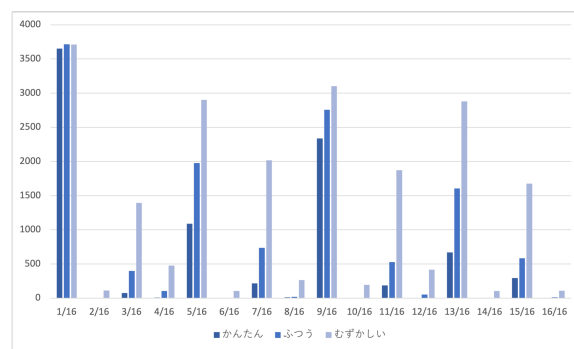


図 3 指示符が存在する拍位置の分布 (難易度別)

かしい」の順で増えている。したがって、仮説 1 の妥当性が示された。

仮説 2 の検証に向け、タイミングの知覚のしやすさを、辻野らにならって拍階層表現によって表現する (図 1)。図 1 中で、音符の上に書かれた「4 分」や「8 分」が各音符の拍階層表現である。実際の楽器の楽譜における「4 分音符」や「8 分音符」などの定義とは異なるが、音楽ゲームにおいてはこのように表現されることが多い。太鼓の達人では連打でない限り、太鼓を叩く時間は瞬間であるため、「音

^{*7} 太鼓の達人 新筐体の収録曲, <https://www.wikihouse.com/taiko/index.php?FrontPage>, 最終閲覧日 2022 年 1 月 18 日

長」の概念が存在しないことに起因すると考えられる。

図 2 に、拍階層ごとに指示符が存在する構成比を難易度別に示す。16 分音符単位で分析しており、3 連符や 6 連符は存在しないものとしている。図 2 から、かんたんの譜面では 4 分音符が 90%使われている一方で、むずかしいの譜面では 4 分音符は約 30%で、8 分音符や 16 分音符の割合が多いことが分かる。図 3 には、難易度ごとに小節内における各拍位置に存在する指示符の統計を示した。用意した太鼓の達人の譜面データはすべて 4 分の 4 拍子を基準として作成している。「かんたん」の譜面は、4 分層にあたる 1,5,9,13 に位置する指示符が配置されやすい。一方で「むずかしい」の譜面は、4 分層だけでなく 8 分層や 16 分層にも分布している。加えて、「かんたん」で指示符が存在する 8,891 箇所の時刻のうち、「むずかしい」では指示符が存在しない時刻が 699 箇所あることが判明した。このことから、「かんたん」で指示符が存在する知覚しやすいと言えるタイミングの中には、リズムの複雑化に伴って「むずかしい」では指示符が存在しなくなった事例が生じていると言える。これらから、**仮説 2** については、妥当性があると考えた。

仮説 1 と **仮説 2** に対する検証結果より、**仮説 3** の妥当性が示された。すなわち、複数の譜面で共通して指示符が存在するタイミングは、少数の指示符で構成される譜面において配置されるタイミングであり、かつリズムの複雑化に伴って失われない、重要なキータイミングであると言える。全ての譜面で共通して指示符が存在する時刻のことを、本稿では全難易度共通操作タイミングと呼ぶ。全難易度共通操作タイミングはキータイミングであると考えられる。

3. 音響特徴量からのキータイミングの推定

キータイミングであると考えられる全難易度共通操作タイミングを、深層学習を利用して音響特徴量から推定する。精度の高い学習モデルが得られれば、太鼓の達人の譜面が存在していない楽曲に対しても、音響特徴量を入力することでキータイミングの推定が可能になると考えられる。

3.1 全難易度共通タイミングを用いたキータイミングデータの生成

キータイミング推定に向けて、1.2 節のデータセットを対象として難易度間で共通する操作タイミングを抽出したデータを作成した。全難易度共通操作タイミングのデータには、同一楽曲に対する「かんたん」「ふつう」「むずかしい」の 3 譜面全てに共通して指示符が存在する時刻の一覧が記録されている。ここで、指示符の種類的一致は問わないこととした。

3.2 学習手法

学習における入力に WAV 形式の「楽曲データ」を、出

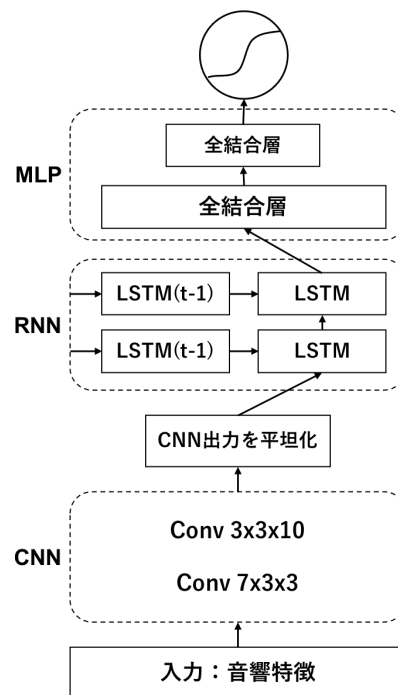


図 4 学習モデルの構造 (文献 [4] の Figure.5 を参考に作成)

力として「全難易度共通操作タイミング」を、それぞれ使用する。データセット内の楽曲 90 曲を、「学習」72 曲、「検証」9 曲、「テスト」9 曲にランダムに振り分けた。「学習」「検証」で全難易度共通操作タイミングを推定可能なモデルを学習させる。「テスト」では、学習済みモデルを使用し、音響特徴量から全難易度共通操作タイミングを推定する。推定した全難易度共通操作タイミングと、オリジナルの譜面から読み取れる全難易度共通操作タイミングを比較し、モデルの精度を評価する。モデルの構成、および入力となる楽曲データの表現は、Chris ら [4] のものを参考とした。

楽曲の長さおよび曲の構成は、太鼓の達人に収録されているものと同じになるように編集した。楽曲をサンプリングレート 44.1kHz で読み込み、短期間フーリエ変換 (FFT) を行うことでスペクトログラムを得る。ここで FFT の窓サイズは 23ms, 48ms, 92ms の 3 種類、ステップサイズは 10ms とした。窓サイズが短いほどピッチや音色などの低レベルの特徴を、窓サイズが長いほどメロディやリズムなどの高レベルの特徴を表すことができる。得られた 3 種類のスペクトルそれぞれに、80 個の周波数帯域のメルフィルタバンクを適用し、メルスペクトログラムを得た。楽曲中のある時刻フレーム (幅 10ms) について、対応する時刻を中心とする 15 フレームのメルスペクトログラムを、モデルに入力する音響特徴量とした。すなわち、1 フレームに対するモデルへの入力は、 $15 \times 80 \times 3$ のベクトルで表現される。

譜面データからは、正解データとなる全難易度共通操作タイミングを抽出する。「かんたん」「ふつう」「むずかしい」の 3 種類の譜面データを比較し、全ての譜面に共通し

表 5 テストデータに対する全難易度共通操作タイミング推定についての評価結果

楽曲名	全難易度共通タイミング数		適合率	再現率	F -score
	正解	推定結果			
あなたとトゥラッタッタ♪	47	162	0.216	0.745	0.335
私以外私じゃないの	114	241	0.021	0.044	0.028
残響	88	174	0.345	0.682	0.458
crossing field	74	199	0.020	0.054	0.029
ピースサイン	90	264	0.235	0.689	0.350
シュガーソングとビターステップ	108	328	0.101	0.306	0.151
ゴッドソング	161	255	0.294	0.466	0.361
紅蓮華	58	179	0.168	0.517	0.253
流星	89	183	0.022	0.045	0.029
平均	92.111	220.556	0.158	0.394	0.222

て指示符が存在する時刻を、共通操作タイミングとして抽出する。抽出した時刻を10ms単位にクオンタイズし、対応する楽曲の先頭から、共通操作タイミングであれば1、そうでなければ0を記した正解データ系列を作成する。

上記の入力および正解データを用いて、Chrisら[4]と同じ構造*8のC-LSTMモデルを学習させる。図4に示すように、学習モデルは2つのCNN層、2つのLSTM層、2つの全結合層、そしてsigmoid関数からなり、最終的に楽曲中の各時刻フレームに対して0以上1以下の数値を出力する。モデルの詳細は文献[4]を参照されたい。「学習」と「検証」に振り分けられた楽曲を用いて、モデルを学習させた。

学習済みモデルを用いることで、任意の楽曲に対して全難易度共通操作タイミング、すなわちキータイミングを推定する。推定対象の楽曲から音響特徴量を取得し、学習済みモデルに入力することで、楽曲中の各時刻フレームに対して0以上1以下の出力値が得られる。この出力値は、対応する時刻が全難易度共通操作タイミングである確率を示す。出力値を時刻順に並べた出力系列に対して、極大値検出を行い、検出された極大値が存在する時刻をキータイミング推定結果とした。この時、辻野らの研究[7]を参考に、「検証」データに対する推定性能(キータイミング単位のマイクロ F -score)が最も良くなるように閾値を定め、その閾値を超える極大値のみを検出する。

3.3 キータイミングの推定結果

「テスト」に振り分けられた楽曲に対して、学習済みモデルが推定したキータイミングと、元の譜面から読み取れる全難易度共通操作タイミングを比較し、評価を行った。評価指標には F -scoreを用いた。楽曲ごとの適合率と再現率から得られる F -scoreを合計し、平均したものを F -score^cとする。キータイミング単位で算出したマイクロ

F -scoreは、 F -score^mとする。評価結果として、 F -score^cは0.222、 F -score^mは0.219となった。表5に、楽曲ごとの結果を、全難易度共通操作タイミングの数と共に示した。

3.4 考察

全ての楽曲について、正解の全難易度共通タイミング数と比較して、学習モデルから推定したキータイミングの個数が多い結果となった。このため、再現率において一定の性能を示す一方で、適合率が非常に低く、これに伴って F -scoreも低くなるという結果となった。現状の提案モデルは、キータイミングの候補を提示するという利用方法であれば有用性があるが、完璧にキータイミングを的中させることは難しいと言える。完全な自動推定の実現のためには、より少数の候補を選び出すような仕組みを実装する必要がある、今後の課題である。

学習モデル構造の参考とした先行研究[4]で報告されている F -scoreは、おおよそ0.7、低くとも0.389であり、これと比較すると本研究の提案手法は性能が低いと言える。音楽ゲームという同一ジャンルのゲームであるにも関わらず、同一のモデル構造でうまく学習できなかった背景として、先行研究が対象としたダンスゲームと本研究が対象とした太鼓の達人の間で入力となる楽曲の特性や推定対象の配置傾向に違いがあった可能性が考えられる。入力楽曲の特性として、ダンスゲームではバスドラムが等間隔で鳴らされるような、いわゆる「4つ打ち」のダンスミュージックが多く収録されている。一方で、太鼓の達人は流行のポップス・アニメソングなどをジャンル問わず収録する傾向があるため、推定対象が存在する時刻について音響的な共通点がダンスゲームほど明確に見られず、学習がうまく進まなかった可能性が考えられる。また、指示符の配置時刻そのものを推定対象としていた先行研究に対して、本研究は「指示符配置時刻の中でも特に複数難易度で共通するもの」という狭い対象の推定を目標としており、タスクとしての難易度が上昇しているとも考えられる。本研究で推定対象

*8 ただし出力の特性上、先行研究で用いられていた難易度 one-hot ベクトルは実装していない

とした全難易度共通タイミングの頻度は、最も低い曲で約 0.45 回/秒、最も高い曲で約 1.84 回/秒であり、曲あたりの平均は約 0.89 回/秒、分散は約 0.07(回/秒)²であった。ダンスゲームの譜面生成を目指した先行研究が用いていた Fraxtil データセットにおいて、指示符の頻度の平均が最も 0.89 回/秒に近い難易度は、Beginner 難易度 (曲当たり平均約 0.67 回/秒) である。Beginner のような指示符の頻度が少なく、難易度が低い譜面については、指示符配置タイミングの推定が難しいことが文献 [4] 中でも述べられていた。太鼓の達人の譜面を利用したキータイミング推定性能を向上させるためには、対象データの特性を捉えて適切な学習モデルを構築することが重要であると考えられ、これについても今後の課題としたい。

4. キータイミングの利用可能性についての検討

4.1 音楽関連メディアコンテンツにおけるキータイミングの利用可能性

1.1 節で検討した通り、キータイミングの推定結果は映像作品制作支援に利用可能であると考えられる。1.1 節では、無作為に選出した一つの楽曲と映像のみを検証した。本稿で取り扱っている全 90 曲は、全て公式に MV が存在する楽曲でもある。音楽のジャンルも映像のジャンルも様々であるので、90 曲全全ての楽曲と MV を調べることで、多種多様な結果が得られるだろう。加えて、より詳細な分析を行うために、対象とする映像エフェクトの種類についても検討を進めたい。

楽曲と映像の関係性を示した研究は複数存在する。平井ら [8] の研究では、任意の楽曲と映像を同期させるために、既存の音楽動画コンテンツを再利用し、時間軸上で新しく音楽映像を自動生成するシステムを提案している。また、人間から見て楽曲と映像が合っていると感じる同期手法について主観的評価実験を行っている。主観的評価実験では、楽曲中のテンポやリズムに対して映像中のアクセントとなるシーンが対応している場合、同期していると見受けられた。Yagya ら [9] は、音楽映像にて表現される感情に焦点を当てた研究をした。視覚と音響環境を同時にどのように知覚し、どのように意味付けし、どのように時間的な判断を行うかを、深層学習を使って解明しようとした。音楽の構造的特徴 (テンポ・モード・メロディ・リズム・歌詞・ラウドネス) と映像の視覚的感情 (顔の表情・体の動き・目の動き) が MV に共存しているため、複雑であると述べた。Laure ら [10] の研究は、国際的に著名な音楽映像の専門家にインタビューを行い、公式ミュージックビデオを対象に音楽映像の分析を行った。インタビューの中には、サビや節といった音楽的な構造、MV の演出に考慮されていることやリズム、ビート、ダウンビートなどの音声イベントを考慮した編集を行うと意見があった。また、見

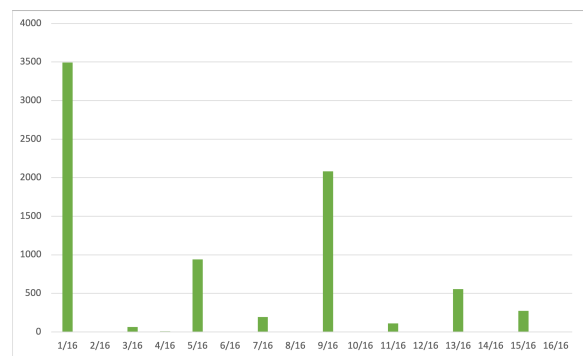


図 5 全難易度共通操作タイミングの小節内拍位置分布

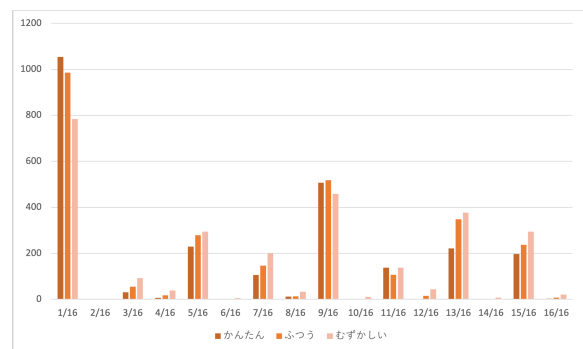


図 6 特殊動作指示符が存在する小節内拍位置の分布 (難易度別)

る人を驚かせたり、音楽の感情的なクライマックスを説明したりといったダイナミクスを生み出すために、映像中に意外性のあるものを取り入れることがある。これらの研究から、楽曲中のキータイミングに映像制作者が考えるシーンを入れれば、映像シーンをより良く際立たせることが期待できるだろう。本稿で提案したキータイミング推定手法は、映像制作者のスキルアップを考慮した映像制作支援の手法の提案にもつながると考えられる。

本稿では、映像制作に限らず音楽にひもづくコンテンツ全般の制作支援を目標とするため、楽曲からのキータイミング推定に注力した。音楽にひもづくコンテンツの一例として、コンサートの演出が挙げられる。神野ら [11] は、楽曲の歌詞、長短調とテンポを用いることで、コンサート中の照明の色を定めていた。楽曲の特徴を 1 曲通して見ているため、時間の変化に対しての音響特徴量の変化を見ていない。

4.2 キータイミングを用いた特殊動作タイミングの推定

仮説を立てた上で検証を行った 2.2 節より、全難易度共通操作タイミングはキータイミングであると言える。その一方で、太鼓の達人における単一の譜面の中でも、キータイミングに関連して存在すると考えられるものに特殊な動作を要求する指示符が挙げられる。表 3 の下半分の部分である「ドンッ (大)」「カッ (大)」「連打」の 3 つの動作は、通常の「ドンッ」「カッ」と比べて、要求される動作が大きくなる。これら 3 つの動作を「特殊動作」と名称する。

寺澤ら [12] によると、楽曲中の盛り上がりに合わせて体性感覚に負担がかかるとき、脳はダイナミックな情動の変化だと認知する。特にアンサンブルでは、演奏者同士で情動変化の共有が起こりやすくなっている。太鼓の達人は楽器を使ったゲームであり、楽曲に合わせて演奏すると捉えると、アンサンブルに似た状況下でのプレイになる。したがって、大きな動きを要求する特殊動作のタイミングは、楽曲中のキータイミングと関わりがあると考えた。

図 5 に、各拍位置に存在する全難易度共通操作タイミングの統計を示し、図 6 に各拍位置に存在する特殊動作指示符の統計を難易度ごとに示した。「かんたん」の特殊動作指示符配置タイミングの拍位置分布と、全難易度共通操作タイミング拍位置分布の間の相関係数は約 0.983 となり、強い正の相関を示した。同様に、「ふつう」特殊動作タイミングと全難易度共通操作タイミング間の相関係数は約 0.971、「むずかしい」特殊動作タイミングと全難易度共通操作タイミング間の相関係数は約 0.929 となり、いずれも強い正の相関を示した。これらの分析結果から、特殊動作指示符の配置タイミングの分布は全難易度共通操作タイミング、すなわちキータイミングの分布と傾向が類似していると判断できる。これらは、キータイミングを入力として特殊動作のタイミングを推定するモデルが構築できる可能性を示唆しており、キータイミングは太鼓の達人ほか音楽ゲームの譜面制作の支援に利用できる可能性があると考えられる。

5. おわりに

本稿では、音楽におけるキータイミングの利用可能性について、映像エフェクトのタイミングとの関連を分析した。キータイミングの前後 1 秒にキータイミング上に多く映像エフェクトの存在していることを確認し、音楽が付随するエンタテインメントコンテンツにおけるキータイミングの利用の可能性を述べた。そして、音楽ゲームのひとつである太鼓の達人に存在する譜面データを応用することでの楽曲中のキータイミングの推定を試みた。提案手法では、太鼓の達人の難易度デザインの関連性の分析から、「かんたん」「ふつう」「むずかしい」の全ての難易度において共通して指示符が存在する全難易度共通操作タイミングをキータイミングと定義した。実験の結果、音響特徴量を用いたキータイミングの推定では、本来の譜面にはないタイミングまでも推定していた。キータイミングを高い適合率で推定することはできなかったものの、利用者に対して候補を提示する形であれば利用可能と考えられる。精度の高い自動推定の実現に向けては、より少数の適切な候補を選び出すような仕組みを実装する必要がある。

提案手法では、全ての難易度で共通するタイミングのみをキータイミングとみなしたが、全ての難易度で共通していなくとも、二つの難易度に存在するタイミングも含めてキータイミングと捉えることや、そもそもいずれかの難易

度で指示符が存在するだけでキータイミングと捉えることも可能である。1.1 節で例に挙げた映像エフェクトなど、楽曲中の局所的な盛り上がりが必要となるものと、太鼓の達人の譜面を比較することで、太鼓の達人の譜面内でキータイミングに相当するものが何なのかを確認可能であると考える。

謝辞

本研究の実施にあたり、一部、立命館大学情報理工学部西原陽子教授、立命館大学大学院情報理工学研究科間城大輔氏からの支援を受けた。記して謝意を表す。

参考文献

- [1] 後藤真孝, 吉井和佳, 藤原弘将, Matthias, M., 中野倫靖: Songle: 音楽音響信号理解技術とユーザによる誤り訂正に基づく能動的音楽鑑賞サービス, 情報処理学会論文誌, Vol. 54, No. 4, pp. 1363-1372 (2013).
- [2] 大野直紀, 土屋駿貴, 中村聡史, 山本岳洋: 独立した音楽と映像に対する印象評価と音楽動画の印象の関係性に関する研究, 情報処理学会論文誌, Vol. 59, No. 3, pp. 929-940 (2018).
- [3] 佐藤晴紀, 平井辰典, 中野倫靖, 後藤真孝, 森島繁生: 映像の盛り上がり箇所に音楽のサビを同期させる BGM 付加支援手法, 情報処理学会研究報告. [音楽情報科学], Vol. 2015, No. 10, pp. 1-6 (2015).
- [4] Donahue, C., Lipton, Z. C. and McAuley, J.: Dance Dance Convolution, *Proc. of ICML 2017*, pp. 1039-1048 (2017).
- [5] 辻野雄大, 山西良典, 西原陽子, 福本淳一: 時系列深層学習に基づく難易度間関係モデルを用いたダンスゲーム譜面難易度の自動調整, 情報処理学会論文誌, Vol. 59, No. 11, pp. 1953-1964 (2018).
- [6] 香川俊宗, 手塚宏史, 稲葉真理: 音楽の重要な構成要素の抽出の提案-音楽ゲーム用譜面自動生成のために, エンタテインメントコンピューティングシンポジウム 2015 論文集, Vol. 2015, pp. 326-333 (2015).
- [7] Tsujino, Y. and Yamanishi, R.: Dance Dance Gradation: A Generation of Fine-Tuned Dance Charts, *Proc. of ICEC 2018*, pp. 175-187 (2018).
- [8] 平井辰典, 大矢隼士, 森島繁生: 既存音楽動画の再利用による音楽に合った動画の自動生成システム, 情報処理学会論文誌, Vol. 54, No. 4, pp. 1254-1262 (2013).
- [9] Pandeya, Y. R. and Lee, J.: Deep learning-based late fusion of multimodal information for emotion classification of music video, *Multimedia Tools and Applications*, Vol. 80, No. 2, pp. 2887-2905 (2021).
- [10] Pr  tet, L., Richard, G. and Peeters, G.: Is there a "language of music-video clips" ? A qualitative and quantitative study, *Proc. of ISMIR 2021*, pp. 539-546 (2021).
- [11] 神野満里奈, 福原義久: 歌詞や曲調の印象に応じた舞台照明の自動調光・調色システムの実現, 第 83 回全国大会講演論文集, Vol. 2021, No. 1, pp. 743-744 (2021).
- [12] 寺澤洋子, 星-芝玲子, 柴山拓郎, 大村英史, 古川 聖, 牧野昭二, 岡ノ谷一夫: 身体機能の統合による音楽情動コミュニケーションモデル, 認知科学, Vol. 20, No. 1, pp. 112-129 (2013).