Vol.2015-MUS-108 No.14
2015/9/2

楽曲のビート類似度及び潜在トピックの類似度に基づくDJプレイの自動化

平井 辰典[1,a)]　土井 啓成[2]　森島 繁生[3,4]

概要：本稿では，楽曲同士を滑らかにミックスして切り替えるDJプレイの自動化手法を検討する．DJプレイにおいて最も重視されるものの一つにビートであり，ビートを一致させることで曲を繋げる研究は提案されてきた．一方で人間のDJはビート等の機械的な事象だけでなく，楽曲の意味等の高等な事象についても考慮して曲を繋げることが多い．本研究では，楽曲のビートの繋がりを維持するためのビート類似度の考慮と，さらに楽曲の音響的な意味にまで踏み込んだ潜在トピックの類似度を考慮する．それによって，人間が行うDJプレイに近い楽曲同士のミックスの実現を目指す．本研究では楽曲の意味を考慮するために，音響信号に対するトピック分析を行う．そこで，各トピックに明確な意味を持たせるために音響信号の言語化手法を導入し，従来の自然言語処理におけるトピック分析に近い形で楽曲の音響信号の潜在的な意味を分析する．これによって，ビートの類似度だけでなく潜在的トピックの類似度までも考慮した楽曲同士のミックスとそれに基づくDJプレイの自動化を提案する．また，主観評価実験によって本手法による潜在トピックの考慮の有効性を示す．

1. はじめに

音楽は街中に流れるBGMやパーティ等のBGM，作業中，移動中のBGM等，様々な形で音楽を享受しているが，それらの場合，複数の曲を連続して聞くことが多い．そういった場面において，音楽を楽しむためには選曲をすることが重要な要因である．その一方で選曲をするには手間がかかり，一曲ごとに作業が中断されることとなるため，事前に作ったプレイリストの再生やランダム再生に頼ることが多い．例えば運転中にドライバーが次の曲を選曲することは危険を伴う行為である．そのため，音楽を自動で選曲して再生する手法は重要な研究対象である．

音楽を選曲，再生することを専門とする職業にDJがある．DJは，クラブやパーティにおいて曲を選曲してエンドレスにミックス[*1]することで場を盛り上げる．DJプレイにおいて最も重要なことは音楽を止めないことであり，あるDJから別のDJに切り替わるときでさえ曲を止めることはない．また，DJプレイはプレイリストの曲を一曲ずつ流す場合とは違い曲の切り替わりを滑らかに，なるべく雰囲気や盛り上がりを保つようにする．もしも選曲のプロであるDJがドライブ中や運動時などの個人的な音楽鑑賞においても選曲，ミックスしてくれれば音楽鑑賞体験は向上することが期待できる．本研究ではDJが実際に行うような曲のミックスを自動化する手法について検討する．

DJプレイにおいて曲を繋げる際に最も重視される点はビートであり，曲をミックスする際にテンポを近づけることで，曲の切り替えをなるべく自然に保つことが多い．切り替えの際に二曲のビートを一致させ，徐々にフェードさせることで二曲がまるで一曲の曲であるかのように聞かせる．このようなミックスの手法はロングミックスと呼ばれ，DJプレイにおける最も基本的な曲の切り替え方の一つである．曲のビートの考慮はDJプレイにおいて重要かつ基本的な要素であり，テンポ及びビートを一致させるように楽曲のミックスを行う手法はすでに提案されている [1]．

一方で，ビートは表層的な要素であり，DJが楽曲を繋げる際にはビートだけではなく楽曲の持つ意味も考慮してミックスを行う．本研究では，いかに楽曲の背後に潜む意味を汲み取るかを検討し，ビートだけでなく意味も近いような箇所で楽曲をミックスする手法の実現を目指す．楽曲の意味は非常に複雑な概念であり，歌詞から読み取れるもの，音響的に汲み取れるもの，コンテクストが関わる複雑なものなどがある．本稿では，楽曲の意味を音響的に汲み

1 早稲田大学
Waseda University, Shinjuku, Tokyo 169–8555, Japan
2 株式会社ドワンゴ
3 早稲田大学理工学術院総合研究所
4 JST CREST
a) tatsunori_hirai@asagi.waseda.jp

*1 DJプレイにおいて「ミックス」とは，曲と曲とを滑らかに繋いで切り替えることを指す．

© 2015 Information Processing Society of Japan

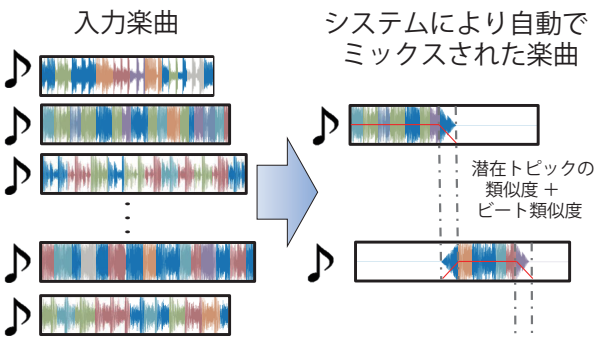

図 1　本システムによる楽曲のミックスの概念図
Fig. 1　The conceptual image of our system mixing songs.

取ってモデル化する手法を提案する．具体的には，音響信号を言語化し，音響信号に対するトピック分析を行う．トピック分析によって推定された楽曲のトピックの混合比が音楽の潜在的な意味に相当する．

本手法では DJ プレイの自動化を実現するために，ビートの類似度及び潜在トピックの類似度の両方を考慮し，二つの類似度が高い箇所同士で曲をミックスする（**図 1**）．ユーザが入力した楽曲集合に対して，類似度の高い箇所で次々と曲を切り替えていくことで自動 DJ プレイを実現する．

2.　関連研究

2.1　選曲と合成

Ishizaki らは曲同士のテンポを伸縮して一致させることによって曲を切り替える DJ システムを提案した [1]．このシステムでは，主観評価実験に基づいて定義されたテンポの伸縮によって生じるユーザの不快感に関する尺度を定義して楽曲のミックスを行っている．この手法では，テンポやビートの一致のみに焦点を当てておりそれ以外の要素については考慮していない．莪山は和音進行の連続性を用いて，楽曲を連続再生する手法を提案した [2]．この手法では，曲をミックスするのではなく，本来繋がるはずだった再生中の楽曲の続きに近い和音進行を持つ楽曲に瞬間的に切り替えることで違和感を抑えて曲を連続再生する．このような曲の切り替え手法は，DJ プレイにおいてカットインと呼ばれており，ロングミックスと並び DJ プレイにおける基本的なテクニックである．堀内らは，莪山による連続再生手法を応用した音楽再生システム Song Surfing を提案しており [3]，DJ プレイという観点ではないが，音楽を連続的に楽しむことを実現している．

DJ プレイに近い要素として，音楽プレイリストを生成する研究が提案されている（[4], [5], [6], [7], [8]）．Platto らはシード楽曲を基に，ガウス過程回帰を使ったプレイリスト生成システム AutoDJ を提案した [4]．AutoDJ のプロジェクトチームは，楽曲同士の類似度を推定してプレイリストを生成する手法も提案している [7]．Goto らは音楽との多様な出会い方を可能にする音楽検索・再生インタフェース Musicream を提案した [9]．このインタフェースにおいて，楽曲をくっつけてプレイリストを作るメタプレイリスト機能が提案されている．これらのプレイリストに関連したアプローチは，曲の連なりとしてのプレイリストに特化しており，曲同士のミックス等の DJ プレイにおいて重要なポイントについては言及されていない．

楽曲を混ぜ合わせる別のアプローチとして，マッシュアップが挙げられ，その生成技術についても研究がなされている [10], [11], [12]．マッシュアップ楽曲とは，複数の楽曲を使って作り出された一つの楽曲である．Tokui らはインタラクティブに楽曲をマッシュアップするシステム Massh を提案している [11]．Davies らによる AutoMashUpper では独自に定義した尺度に基づいてマッシュアップ楽曲の生成を実現している [12]．マッシュアップは DJ のテクニックとして行われることもあるが，本稿で注目する DJ プレイは楽曲を次から次へと繋ぐミックスである．

DJ プレイにおけるミックスに注目した研究はプレイリスト生成に関する研究に比較するとあまり行われていない．一方で，曲同士のミックスは DJ プレイという観点においては非常に重要であると共に，プレイリスト生成手法と組み合わさることでさらに強力な音楽鑑賞体験が実現できるものである．そこで，我々は楽曲同士のミックスに注目して自動 DJ プレイを実現するような手法を，テンポやビートよりもさらに踏み込んで実現する．DJ プレイにおいて，テンポやビートと同様に重要なものが曲の持つ意味であり，本研究では楽曲の意味の考慮を目指す．

2.2　トピックモデル

自然言語処理分野において，文章の背後に潜むトピックを推定するトピックモデルという潜在的意味解析手法が盛んに研究されている．トピックモデルは，文章において出現する単語はその文章のトピックに影響されることを利用し，文章そのものを分析することでその背後に潜むトピックについても推定する分析手法である．もしも潜在するトピックが同じであれば，文章同士は高次な意味レベルにおいて似ていると言うことができる．Sasaki らは歌詞に対してトピック分析を行うことで，楽曲の潜在トピックの分析を行った [13]．

トピックモデルは単語だけではなく，画像や音響特徴量の値に対しても応用することが可能である．その際，特徴量は Bag-of-features 表現 [14] によってベクトル量子化される．Nakano らはトピックモデルを歌声特徴に適用して歌声の分析を行った [15]．Nakano らはトピック混合比の KL-divergence に基づく類似度を定義し，似ている歌声の歌手が似たような潜在トピックを持つことを示した．しかし，このような特徴量の値に対して適用されたトピックモデルは，モデル化されたそれぞれのトピックがどのような

© 2015 Information Processing Society of Japan

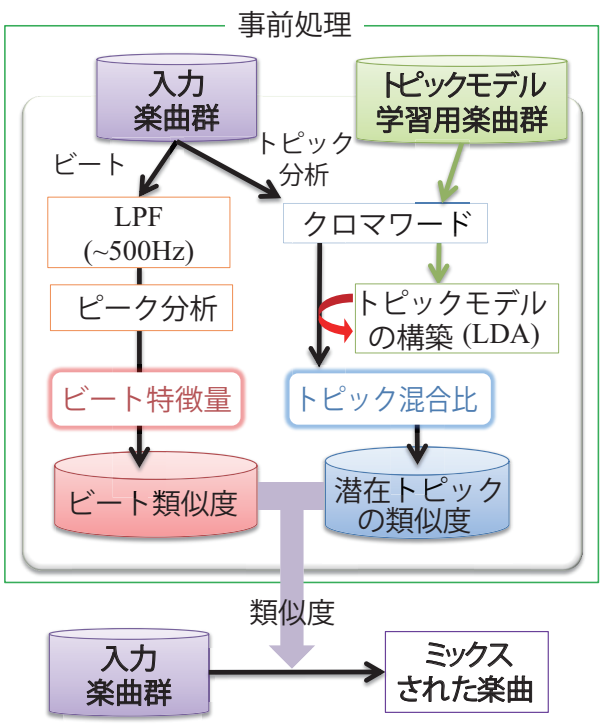

図 2　システムのフロー
Fig. 2　Overview of the system.

意味を持つのか，陽に理解することができない．それは，各トピックが単語の混合によって構成されるのではなく特徴量ベクトルの混合によって表現されるためである．

Hu らは楽曲の音名を単語と見立てて，トピック分析により楽曲の調を推定する手法を提案している [16]．この結果は，音名を用いたトピック分析が楽曲の背後に潜む潜在的な意味（この場合は調）を推定することに効果的であることを示唆している．Hu らはさらにこの手法を音響信号に応用し，音名ではなくクロマベクトルから楽曲の調を推定する手法を提案している [17]．このアプローチは，クロマベクトルを用いたトピック分析による楽曲の潜在的な意味推定の可能性を示している．

3.　システム構成

本システムのフローを図 **2** に示す．本システムは事前処理として楽曲群のビートと潜在的トピックの類似度を算出する．また，潜在的トピックの分析のために事前にトピックモデルを構築も行う．

ビート類似度を算出するにあたって，入力楽曲群のビートに関する情報を抽出する．ここでビートとは，楽曲のテンポとは違い，楽曲内でも常に変化するものである．クラブ音楽やダンス音楽におけるビートは，主に低音のリズム音（バスドラムやベース音等）が支配的な役割を担っている．DJ プレイにおいて重要なビートの情報を抽出するにあたって，本研究ではスネアドラムの音が含まれる程度の周波数までの低音部のみを処理の対象とする．そこで，入力楽曲にローパスフィルタをかけることによって 500Hz 以下の低音成分のみを抜き出す．抜き出した低音部音響信号にはビートに関連する音響信号が多く含まれる．さらに，抜き出した低音部の振幅に対してピーク分析を行うことで，入力楽曲群のビート位置に関する情報を抽出する．

楽曲のトピックモデルは，潜在的ディリクレ配分法（LDA: Latent Dirichlet Allocation）と呼ばれるモデリング手法によって構築する．本研究では，トピックモデルの学習用の楽曲として，様々な音楽ジャンルを包括しているデータベースを用いて変分ベイズ法によって学習する．トピック分析を行うにあたり，本研究ではクロマベクトルを元に音響信号を言語化したクロマワード（Chroma Word）を提案し，楽曲をクロマワードからなる文章として表現する．楽曲を文章表現することにより，自然言語処理的アプローチによる潜在トピック推定が可能となる．また，クロマワードはクロマベクトルにおける音名情報に基づく音響信号の単語化であるため，クロマワードの集合によって表現されるトピックからは音名が陽に推定可能である．クロマワードを使って学習されたトピックモデルを基に，未知の入力楽曲のトピック混合比が推定可能となる．我々の目的は楽曲全体のトピック分析をすることではなく，曲のミックスのための小区間におけるトピック混合比を推定することにある．そこで本稿では，5 秒をトピック分析のためのひとまとまり（言語における文章）とし，楽曲の 5 秒毎のトピック混合比を推定する．

最後に，ビート類似度及び潜在トピック類似度に基づいて楽曲のミックスを行う．事前に算出済みの類似度情報を用いて，任意の再生中の楽曲の小区間に対して最も類似度が高い楽曲の区間を求め，該当箇所同士でクロスフェードをすることによって楽曲同士をミックスする．このようにして，楽曲同士がビートと潜在トピックの両面において類似している箇所において楽曲がミックスしていくことによって，曲を次から次へと切り替えていく DJ プレイを自動で行うことができる．

本提案手法の要素であるビート類似度については 4 章，潜在トピックの類似度については 5 章，楽曲のミックスについては 6 章でそれぞれ詳細に述べる．

4.　ビート類似度

ビートは DJ のミックスにおいて最も基本的かつ重要な要素の一つである．我々は，ビートの特徴をシンプルに保つために，その他の音がビートに与える影響は少ないと仮定して，主にバスドラムからベース音，スネアドラムの音までが抽出できる程度の周波数領域でビート分析を行う．そのような低音部を抽出するために，入力楽曲の音響信号をローパスフィルタに通して 500Hz 以上の信号をカットする．500Hz は一般的なスネアドラムの低音域の周波数が包括される程度の周波数である．抽出した低音部にはビー

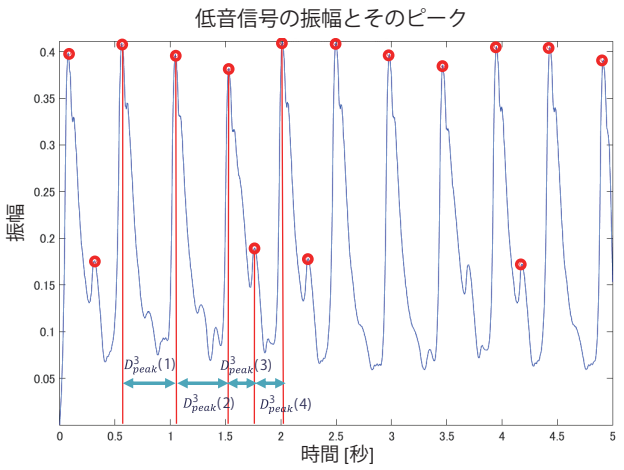

図 3 低音部の振幅波形からの楽曲のビート特徴量の抽出
Fig. 3 Extracting beat features from the amplitude of low frequency audio signal.

トに関連する音響信号が多く含まれており，この信号の振幅に対してピーク分析によって一定以上の極大値を見つけることでビート情報を抽出する．ローパスフィルタリング及びピーク分析は Lartillot らによる音楽信号解析ツールである mirtoolbox1.6.1 を用いて行った [18]．これにより，バスドラム等の音のアタック音に起因するビートの情報を抽出することができる．

楽曲のビートに関する情報を記述する特徴量として楽曲のビート特徴量を定義する．**図 3** は低音部の振幅波形から抽出したピーク情報を示す．ビート特徴量は，ユーザが決めた注目ピーク数に基づいて，各ピークに対して抽出される．注目ピーク数が N だった場合，i 番目のピークにおけるビート特徴量は，i 番目のピークを基準として N 個先のピークまでのピーク間隔 $\{D^i_{peak}(1), ..., D^i_{peak}(N)\}^T$ の N 次元ベクトルによって表現される．図 3 には一例として注目ピーク数 $N = 4$ の場合の 3 番目のピークにおけるビート特徴量を示している．ビート特徴量は楽曲毎に（総ピーク数 $-N$）サンプル抽出される．

楽曲の小区間同士のビート類似度はビート特徴量を用いて求められる．小区間 i と小区間 j のビート類似度 $S_{beat}(i, j)$ は以下の式によって表される．

$$S_{beat}(i, j) = \frac{1}{\sum_{k=1}^{N} ||D^i_{peak}(k) - D^j_{peak}(k)|| + 1} \quad (1)$$

ここで，N が大きいほどビートという観点において良いマッチが図れるが，N を大きくするほどミックスの候補となる区間は少なくなる．また，各小区間の長さは基本的には一致しないが，ビート類似度の値が大きい小区間同士の長さは極めて近くなる．$S_{beat}(i, j) = 1$ の場合すべてのピーク間隔が一致して，該当区間同士で曲をミックスした場合，ビートの観点からは自然に曲を切り替えることが可能となる．

5. 潜在トピックの類似度

本研究では，楽曲の潜在トピックの類似度に注目することで，人間の DJ が曲をミックスすることにより近い楽曲のミックスの実現を目指す．楽曲には，歌詞やジャンル，曲の構造，楽器構成，音色等，様々な要因があり，楽曲の潜在的な意味を捉えるためには，それら様々な要因を検討する必要がある．本稿では，楽曲の小区間において，ある音名を生成させる要因となった潜在的な意味を考慮すべく，楽曲の音響信号を用いた潜在的意味解析を行う．

5.1 クロマワードの抽出

本稿では，トピックモデルの構築のための特徴量として，クロマベクトルを基に音響信号を単語化したクロマワードを提案し，自然言語処理的な枠組みの中で楽曲の潜在的意味解析を行う．

クロマベクトルの各ビンには音名が割り振られており，その各音名を単語における文字として優勢な音名から順に単語を構築することでクロマワードの抽出ができる．この音名文字は A から G#まで，12 文字あることになる．ここで，#は一文字には含まない．

クロマベクトルの値には打楽器音等に起因するノイズが混ざることが多いため，クロマベクトル全体の上位 70% までに含まれる音名文字を利用するとともに，クロマワードの長さの上限は四文字までとする．このようにして楽曲の音響信号を単語化する．

クロマベクトルを Bag-of-features 表現によって記述するのではなく，クロマベクトルの単語表現としてのクロマワードを導入することによって，音名に関する意味合いが陽に理解可能な形でトピックを学習する．これは，トピックモデルの妥当性を議論する上で，有用である．

図 4 に実際の楽曲のクロマベクトルから抽出したクロマワードの例を示す．図 4 には，クロマベクトル 0.5 秒ごとに三語ずつ等間隔にクロマワードを抽出した例を示している．クロマワードにおける最も左の文字が最も支配的な音名を表し，右の文字ほど影響が小さい文字である．

楽曲の音響信号 1 フレームに対して一つのクロマワードが割り当てられ，複数フレーム分のクロマワードを集めることで文章と見なすことが可能となる．ここで，音声はサンプリング周波数 16,000Hz のモノラルデータとし，クロマベクトルを抽出する際のフレーム長は 200ms で，シフト幅は 10ms とした．つまり，一秒あたり 100 語のクロマワードが抽出される．

5.2 潜在的ディリクレ配分法

音響信号が単語として表現されるため，自然言語処理と同等の手法でトピックモデルを構築することができ

© 2015 Information Processing Society of Japan

表 1　各トピックを構成する上位 5 語のクロマワード
Table 1　Top-5 ChromaWords allocated to each topics.

Topic22	Topic90	Topic7	Topic98	Topic78	Topic52	Topic9	Topic79	Topic43	Topic80	Topic85	Topic44
CBC#A	silent	AG	CFG	AA#BC	ED	AFB	silent	AEA#	DAA#C	DGG#	EB
CC#BA	GDCA	AA#G	CFGD	ABA#C	E	AA#BF	GA	AEA#B	DACA#	CGB	BE
CC#AB	GCDA	AGA#	CGFA	BAA#C	F	silent	CF	silent	ADA#C	AEA#	BEG
CBAA#	GDAC	ADGA#	CGFD	AA#CB	DE	AA#GB	CFAG	AEA#G	DACC#	CFA	EBG
CABA#	DA	ADA#G	CGF	ABCA#	EF	AA#FB	FC	AEA#C	AA#BC	CGE	EA

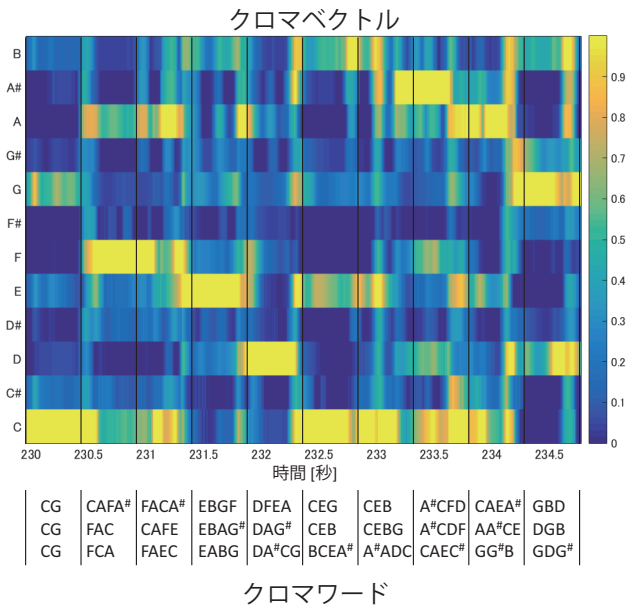

図 4　クロマベクトルからのクロマワードの抽出
Fig. 4　Extraction of ChromaWords from chroma vector.

る．本研究では，Sasaki らが楽曲の歌詞に対して適用した LDA と同様の変分ベイズ LDA をクロマワードに対して適用することでトピックモデルを構築した [13]．LDA のトピックの数は実験的に 100 とした．クロマワードのボキャブラリー数は，完全な無音状態を含めて 13345 ($= {}_{12}P_4 + {}_{12}P_3 + {}_{12}P_2 + {}_{12}P_1 + 1$) である．

トピックモデルの学習には，DJ プレイに使用する楽曲とは別の学習用楽曲群を用いる．具体的には，様々な音楽ジャンルを包括している学習用の楽曲データには前述のように RWC 研究用音楽データベース（音楽ジャンル）の 100 曲を用いた [19]．RWC 音楽研究用音楽データベースを用いて学習された 100 のトピックのうち，混合比の大きい上位 12 トピックにおける上位五単語を**表 1** に示す．左に示されているトピックほど学習データにおける混合比の和が大きかったトピックである．この結果から，各トピックがどのような音名が支配的なトピックであるかがクロマワードから推測できる．例えば表 1 のトピック 44 は E と B の音が支配的なトピックであり，さらに G の音が少し含まれるようなトピックである．このように学習された 100 のトピックの重みを求めることで，音名の連なりのみではない微細なニュアンスまでも表現可能なモデルとなる．我々はこのトピックモデルをある音名の音を生成させる要因となった潜在的な意味であると考え，トピック混合比が近い楽曲区間同士は潜在的な意味も近いものとする．ただし，前述のように楽曲の意味には様々な要因があり，本手法がそれらも含めた楽曲の意味を包括するわけではなく，音名の生成に関する潜在トピックの組み合わせによる潜在的な意味の推定を可能とするものである．また，各トピックが直接調やコードに対応しているわけではないが，100 トピックの混合により間接的に調やコードの表現もなされるものであると考えており，その検証についても今後行っていきたい．

5.3　未知の入力楽曲のトピック混合比の推定

本研究の目的は潜在トピックの類似度に基づいて楽曲の潜在的な意味を考慮した楽曲のミックスを実現することにある．そのため，未知の入力楽曲の任意の区間におけるトピックの混合比を推定するために予測分布を求める．楽曲の一定以上の長さの区間を一つの文章とする都合上，入力楽曲のトピック混合比は一定区間毎に推定する．本稿では，楽曲を実験的に 5 秒間の小区間に分けてそれぞれのトピック混合比を推定することとした．実際に入力した未知の楽曲の 5 秒間の小区間に関して推定したトピック混合比を**図 5** に示す．この混合比において最も大きな比率を持つトピック 3 は C と E が特に支配的なトピックであり，次に大きな比率を持つトピック 10 は F や E や G が支配的なトピックである．また，この楽曲の該当区間のコード進行は C，G，Am，F であり実際のコード進行とトピックを構成するクロマワードとの間に関係性が見られる．

5.4　潜在トピックの類似度

ここまでに，5 秒の小区間毎に入力楽曲の潜在トピックの混合比を求めることができた．これにより，小区間 i と小区間 j の潜在トピックの類似度 S_{topic} が求められる．潜在トピックの類似度はビート類似度と同様の以下の式によって表される．

$$S_{topic}(i,j) = \frac{1}{\sum_{k=1}^{K} ||f_i(k) - f_j(k)|| + 1} \qquad (2)$$

ここで，K はトピック数であり，本稿の実装では 100 に設定されている．

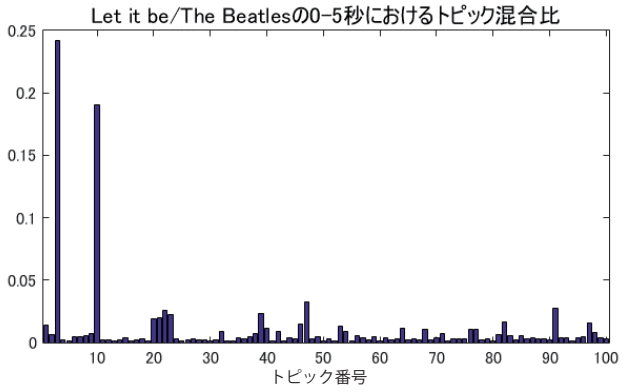

図 5　実際の楽曲の 5 秒間の区間に対して推定したトピック混合比
Fig. 5　A result of topic analysis applied to the 5-seconds fragment of a song.

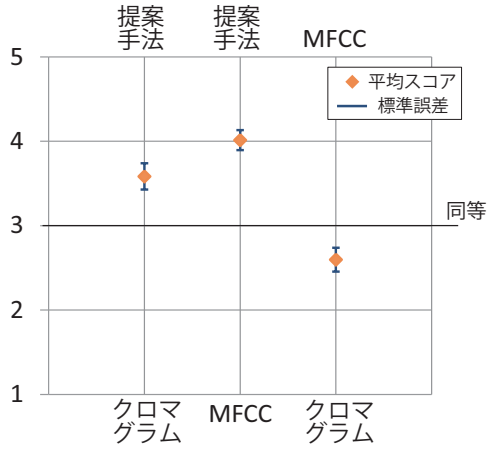

図 6　主観評価実験の結果
Fig. 6　The result of subjective evaluation experiment.

5.5　評価実験

本トピック分析手法を用いた場合と，音響特徴量を直接用いてトピック分析をした場合の楽曲の類似度を主観評価実験により検証した．本手法では，13345 語彙のクロマワードを用いて，100 のトピックを LDA で学習したものである．それに対して，比較手法にはクロマベクトル及び MFCC を用いてそれぞれトピックモデルを学習したものを用いた．比較手法の学習は，中野らによる音楽の確率的生成モデルの構築方法を参考にした [20]．具体的には，学習用の楽曲群から抽出した音響特徴量を k-means クラスタリングによって $k = 64$ 個にクラスタリングし，そのセントロイドを代表点としてベクトル量子化を行う．これは単語の語彙を 64 語として，特徴量ベクトルをいずれかの単語に割り当てることに相当する．さらに，データ数を削減するためにフレームの 15% をランダムに選択し，変分ベイズ LDA によって学習する．ここで特徴量には，クロマベクトル（12 次元）を用いたものと，MFCC（12 次元），Δ MFCC（12 次元），Δ パワー（1 次元）の計 25 次元を用いたものとをそれぞれ学習した．

まず，各手法でそれぞれ RWC 研究用音楽データベースを用いてトピックモデルを学習する．それに対して，未知の入力楽曲 50 曲の 5 秒間隔のトピック混合比をそれぞれ推定する．50 曲の楽曲には主に J-pop やロック，ダンス音楽が含まれている．50 曲を 5 秒毎の小区間に分割した結果，2192 の小区間が得られた．各小区間と 2192 区間のうち同一楽曲ではない楽曲の小区間との潜在トピックの類似度を式（2）によって算出した．3 手法それぞれで類似度が高かった上位 30 組の小区間の中から，無音を含む区間や曲の重複を避けて手作業でそれぞれ 3 組を選出し，5 段階評価によって比較した．

被験者は 20 代の男女 8 名で，いずれも DJ 経験者ではない．被験者は手法 A によって類似していると判定された 5 秒間の楽曲素片の組を聴き，次に手法 B によって判定された楽曲素片の組を聴いてもらい，「どちらの組が似ていたか」について，「1：A の組の方が似ていた」，「2：どちらかというと A の組の方が似ていた」，「3：どちらも同等であった」，「4：どちらかというと B の組の方が似ていた」，「5：B の組の方が似ていた」の 5 段階で評価してもらった．ビート等の影響を防ぐために，曲の組はミックスとして提示するのではなく，各小区間をそれぞれ聴いてもらう形式で提示した．全 9 組のペアの各組に対して，比較手法の 6 組との間の比較を行った．楽曲組の提示順はランダムにし，半分の被験者には手法 A から先に，残りの半分の被験者には手法 B から先に聴いてもらった．

本評価実験の結果を図 **6** に示す．各プロットは比較対象である二つの手法との三組対三組の組み合わせ 9 個のスコアに関する平均を被験者 8 名分でさらに平均を取ったものである．ここで，エラーバーは標準誤差を表している．スコアが 3 の場合比較した両手法が同等であることを表しており，提案手法はクロマグラム及び MFCC を用いたトピック分析の結果よりもより似ている小区間の組み合わせを見つけることができているといえる．

また，被験者にとったアンケートにおいて，ビートやテンポが似ているかどうかが判断基準になったと回答したケースがあった．これは，類似度におけるビートの影響を示唆していると共に，本実験ではビートの影響を取り除き切れていなかったことを示している．ビートの類似度を考慮した評価実験は本稿では行っていないため，今後，ビートの類似度と潜在トピックの類似度の両方を考慮した評価実験を行う必要がある．それによってどちらの要素を主に考慮すべきであるかが検証できると考えている．

6.　楽曲のミックス

潜在トピックの類似度 S_{topic} は 5 秒毎に算出されている一方で，ビート類似度はピークの数に応じた不規則な小区間毎に算出されている．そこで，両者を混合するためにビートの小区間毎に潜在トピックの類似度を求めなおす．

具体的には，5 秒毎の区間のトピック混合比をその区間に含まれる各ビート特徴の小区間に割り当てる．ビート特徴が複数のトピック区間に割り当てられている場合，最も多くのトピックの区間が含まれている混合比を割り当てる．10 秒以上のビート小区間の場合には中心に近いトピック混合比を割り当てる．これにより，ビート類似度と潜在トピックの類似度の両方を同じ間隔で算出することができる．今回はトピック混合比の推定に必要な計算時間の都合で 5 秒毎にトピック混合比を求めたが，今後はビートの小区間においてより正確なトピック混合比の推定ができるように処理を改善する予定である．

ビート類似度と潜在トピックの類似度を混合した楽曲小区間同士の類似度は以下の式で表される．

$$S(i,j) = w \times S_{beat}(i,j) + (1-w) \times S_{topic}(i,j) \quad (3)$$

ここで，w はビートとトピックを重視する際の重みを表すパラメータであり，現在の実装では 0.5 に設定している．

入力楽曲群が与えられ，最初の楽曲が決まると，システムはその楽曲の後半以降の小区間から，最も類似度が高い楽曲の小区間を検索してクロスフェードによってミックスする．そのようにして曲が切り替わると，また次の楽曲に対して最も類似度が高い楽曲の小区間を検索してミックスすることで，エンドレスに楽曲を繋ぐことができ，これが自動 DJ プレイとなる．ここで，楽曲の終わりに近い箇所から始まったり，始まりに近い箇所で切り替わってしまったりすることを避けるために，楽曲の後半からは始まらない，楽曲の前半では曲が変わらないといった制約を与える．これによって曲が数秒で切り替わってしまったりすることを防ぐ．また，一度再生された楽曲は再生しないといった制約を与えることで特定の楽曲が繰り返し再生されてしまうということも防げる．

本手法では，まだ楽曲の構造を考慮していないため，サビの直前で別の曲に切り替わってしまったり，歌の途中で曲が切り替わってしまったりすることなどがある．今後，考慮すべき点を増やし，より効果的なミックスの実現を目指したい．

7. まとめ

本稿では，楽曲同士を滑らかにミックスして切り替える DJ プレイの自動化を目的として，楽曲のビート類似度及び潜在トピックの類似度の両方を考慮した楽曲のミックス手法を提案した．本研究の特徴は，クロマワードによる楽曲の音声信号の言語化と，それに付随してトピックモデルにおける楽曲の各トピックの意味を陽に推定できるようにした点にある．また自動 DJ プレイを目的として，ビートよりもさらに一歩踏み込んだ楽曲の潜在的意味の考慮も本研究の特徴である．さらに，主観評価実験を通じて，本研究で提案した潜在トピックの類似度の考慮が，従来のベクトル量子化された特徴量を用いた潜在トピックの類似度に比べて有効であったことを示した．

今後，ユーザインタフェースを実装することで，全自動の DJ プレイだけでなくインタラクティブに楽曲をミックスすることも目指したい．それにより，選曲というコンテクストや場の雰囲気に依存する要素が大きい作業を人が行い，最も良いタイミングで楽曲をミックスするという難しい作業をコンピュータが担当するといった人とコンピュータとの協調作業が可能となる．また，コンピュータによる支援の度合を変えていくことで，DJ プレイの上達の支援なども可能となると考えている．

本稿では，元の楽曲に変更を加えずに最も類似度の高い箇所でミックスするというアプローチを取ったが，楽曲の伸縮やピッチシフトを行って，そのままではうまくミックスし難い楽曲をうまくミックスできる可能性がある．実際に DJ はテンポの変更やイコライザによる音に調整も頻繁に行ってプレイする．今後は楽曲に変更を加えることも視野に入れた自動 DJ プレイの実現も目指したい．

謝辞　本研究における変分ベイズ LDA の実装にあたって佐々木将人氏の助言を受けた．本研究の一部は，日本学術振興会特別研究員奨励費及び，JST CREST「OngaCREST プロジェクト」の支援を受けて実施された．

参考文献

[1] Ishizaki, H., Hoashi, K. and Takishima, Y.: Full-Automatic DJ Mixing System with Optimal Tempo Adjustment based on Measurement Function of User Discomfort., *Proc. of ISMIR*, pp. 135–140 (2009).

[2] 莪山真一：和音進行の連続性を用いた楽音信号の連続再生手法，情報科学技術フォーラム一般講演論文集，Vol. 4, No. 2, pp. 277–278 (2005).

[3] 堀内直明，薗田俊行，田中浩司，田中淳一，長沢秀哉，莪山真一：Song Surfing: 類似フレーズで音楽ライブラリを散策する音楽再生システム，*PIONEER R&D*, Vol. 17, No. 2, pp. 78–84 (2007).

[4] Platt, J. C., Burges, C. J., Swenson, S., Weare, C. and Zheng, A.: Learning a Gaussian Process Prior for Automatically Generating Music Playlists, *Proc. of NIPS*, pp. 1425–1432 (2001).

[5] Aucouturier, J.-J. and Pachet, F.: Scaling up music playlist generation, *Proc. of ICME*, Vol. 1, pp. 105–108 vol.1 (2002).

[6] Pampalk, E., Pohle, T. and Widmer, G.: Dynamic Playlist Generation Based on Skipping Behavior., *Proc. of ISMIR*, Vol. 5, pp. 634–637 (2005).

[7] Ragno, R., Burges, C. J. and Herley, C.: Inferring Similarity Between Music Objects with Application to Playlist Generation, *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*, pp. 73–80 (2005).

[8] 中野倫靖，濱崎雅弘，後藤真孝：PlaylistPlayer: 再生終了時刻と再生方針が指定できる音楽プレイリスト再生インタフェース，*Proc. of WISS2014*, pp. 13–18 (2014).

[9] Goto, M. and Goto, T.: Musicream: Integrated Music-Listening Interface for Active, Flexible, and Unexpected Encounters with Musical Pieces, *Information and Media*

Technologies, Vol. 5, No. 1, pp. 139–152 (2010).

[10] 宮島　靖：Music Mosaic Generator: 高精度時系列メタデータを利用した音楽リミックスシステム，*Proc. of WISS2007*, pp. 13–18 (2007).

[11] Tokui, N.: Massh!: a web-based collective music mashup system, *Proc. of DIMEA*, ACM, pp. 526–527 (2008).

[12] Davies, M. E., Hamel, P., Yoshii, K. and Goto, M.: AutoMashUpper: Automatic Creation of Multi-Song Music Mashups, *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, Vol. 22, No. 12, pp. 1726–1737 (2014).

[13] Sasaki, S., Yoshii, K., Nakano, T., Goto, M. and Morishima, S.: LyricsRadar: A Lyrics Retrieval System Based on Latent Topics of Lyrics, *Proc. of ISMIR* (2014).

[14] Sivic, J. and Zisserman, A.: Video Google: a text retrieval approach to object matching in videos, *Proc. of ICCV*, pp. 1470–1477 vol.2 (2003).

[15] Nakano, T., Yoshii, K. and Goto, M.: Vocal timbre analysis using latent Dirichlet allocation and cross-gender vocal timbre similarity, *Proc. of ICASSP*, pp. 5202–5206 (2014).

[16] Hu, D. J. and Saul, L. K.: A Probabilistic Topic Model for Unsupervised Learning of Musical Key-Profiles., *Proc. of ISMIR*, pp. 441–446 (2009).

[17] Hu, D. J. and Saul, L. K.: A probabilistic topic model for music analysis, *Proc. of NIPS*, Vol. 9 (2009).

[18] Lartillot, O., Toiviainen, P. and Eerola, T.: A Matlab Toolbox for Music Information Retrieval, *Data Analysis, Machine Learning and Applications*, pp. 261–268 (2008).

[19] Goto, M.: Development of the RWC music database, *Proc. of ICA*, Vol. 1, pp. 553–556 (2004).

[20] 中野倫靖，吉井和佳，後藤真孝：確率的生成モデルに基づく音楽の類似度とありがち度の推定に関する検討，情処音楽情報科学研報，Vol. 2014, No. 2, pp. 1–7 (2014).

© 2015 Information Processing Society of Japan