Vol.2019-MUS-122 No.12
Vol.2019-EC-51 No.12
2019/2/22

カバーソング同定法と音声指紋を組み合わせることによるマッシュアップを考慮したメドレー楽曲における楽曲断片検出法の提案

佐藤 僚太[1,a)]　竹川 佳成[2,b)]　平田 圭二[2,c)]

概要：本稿では，メドレー楽曲におけるマッシュアップを考慮した楽曲断片検出法について述べる．メドレー楽曲とは，複数楽曲の一部分の区間を接続することで作られる新たな形式の楽曲のことを指す．メドレー楽曲では原曲をテンポやキー，音の追加や削除などのアレンジやマッシュアップを行うことで，一曲であるかのように楽曲断片同士の接続を行っている．我々は，メドレー楽曲において何の曲がどこからどこまで登場しているか同定するため，Wang の音声指紋とカバーソング同定法（Cover Song Identification, CSI）である Serrà らの相互再帰定量化 (Cross Recurrence Quantification, CRQ) を組み合わせた楽曲断片検出法の提案する．メドレー楽曲とその構成楽曲から音声指紋を抽出し，2 曲の類似度行列である行列 CR とカバーソングを定量的に評価するための累積値行列 Q を作成し，開始地点を同定することで楽曲断片の検出っを試みる．実験結果から，やや低い精度ではあるものの，パラメータの検討によって本手法の楽曲断片検出の精度が向上することが示唆された．

A Proposal of Musical Segment Detection Method for Mashup Medley Using Cover Song Identification and Audio Fingerprinting

Ryota Sato[1,a)]　Yoshinari Takegawa[2,b)]　Keiji Hirata[2,c)]

1. はじめに

聴き手がメドレー楽曲を聴取することによる能動的音楽鑑賞 [4] がされている．メドレー楽曲とは，編曲された複数の楽曲断片から作られる新たな形式の楽曲を指す．メドレー楽曲では音楽的展開が考慮されているため，ひとつの楽曲の聴取であるかのように複数楽曲の聴取が行われる．ひとつの楽曲を聴取するようにメドレー楽曲を聴取することで，未知楽曲を知る機会を与え，未知楽曲を単体で聴取する衝動を促す．

[1] 公立はこだて未来大学大学院
Graduate School of Future University Hakodate
[2] 公立はこだて未来大学
Future University Hakodate
[a)] g2117024@fun.ac.jp
[b)] yoshi@fun.ac.jp
[c)] hirata@fun.ac.jp

本研究では，次のような特徴を持ったメドレー楽曲の音響情報を対象としている．

- メドレー楽曲内の楽曲断片は次のようなカバー曲的な編曲がされている
  - Digital Audio Workstation (DAW) 等によって編曲されている
  - キーやテンポが変更されている（図 2）
  - 主旋律に対して音の追加や削除などが行われている（図 3）
- 歌声が楽器音に置き換えられている
- 2 曲以上の楽曲が同時に進行する（マッシュアップ）区間がある（図 4）
- メドレー楽曲の曲長やメドレー楽曲内で使用する楽曲数において制限はないが，10 分以上で 50 曲以上のものが多い．

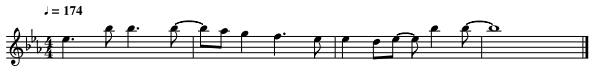


図 1 原曲に含まれる旋律の例

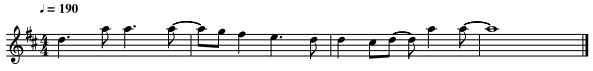


図 2 図 1 のキーを 2 つ下げ，テンポを早めた変更を加えた例

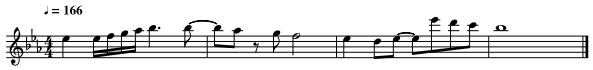


図 3 図 1 の主旋律に対して音の追加や削除を行った例

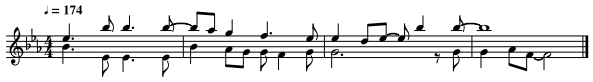


図 4 図 1 を別の楽曲とマッシュアップを行った例

メドレー楽曲の楽曲の変わり目は，音楽的展開が考慮されているため，楽曲が変わったことに気づかないような自然なものであることが多く，聴き手が楽曲断片を認識するのが困難である．また，マッシュアップされている区間においても，2 曲以上の楽曲が同時に発音しているため，楽曲断片を認識するのが難しいという問題がある．特に，楽曲断片が聴き手の未知楽曲である時の検出が困難であり，第三者の存在無しにはメドレー楽曲を構成している全ての楽曲を網羅することが難しい．メドレー楽曲の音響情報から自動で楽曲断片を検出し，メドレー楽曲のどの部分にどの楽曲が使われているかを知ることが可能なシステムが存在すれば，聴き手が未知である楽曲を第三者に頼ること無く知ることができる．メドレー楽曲内で使用されている楽曲を，第三者に頼ることなく未知楽曲も含めて全て認識できることで，聴き手の未知楽曲に対する能動的音楽鑑賞をより促すことができる．

本論文では，聴き手がメドレー楽曲を構成するの全ての楽曲を認識可能なシステムを構築するため，マッシュアップされているメドレー楽曲から楽曲断片を検出することを目指す．聴き手が楽曲断片を検出するプロセスでは，楽曲認識と，楽曲のどの部分がメドレー楽曲のどこからどこまで登場しているかという区間検出が必要となる．しかしメドレー楽曲では，ひとつの楽曲であるかのような音楽的展開をするために，楽曲断片同士を連結する際に前述のようなカバー曲的な編曲がされている．そのため，Foote の Self-Similarity Matrix[2] や Dannenberg ら [1] の楽曲構造分析では，多くの場合でメドレー楽曲中に繰り返し構造の境界が構成楽曲の構成とならないため，楽曲の変わり目を検出することが困難である．また，マッシュアップと同様に複数楽曲を一曲のように聴取する手段である DJ MIX を対象にした楽曲認識の研究がいくつかあるが [3][6][8]，メロディが同時刻に 2 つ以上存在して干渉し合うような楽曲を対象にした研究は少ない．我々が以前提案したメドレー楽曲を対象にした楽曲断片検出手法 [10] は，「メドレー楽曲中では複数の楽曲が同じ時刻において存在しない」という制約を設けたが，実世界で存在するメドレー楽曲を対象とするためには，マッシュアップを考慮する必要がある．そのため，マッシュアップがされているメドレー楽曲からの楽曲断片検出を実現するために，以下の要件を満たした手法を用いることが望ましい．

- 同時刻に 2 つ以上存在するメロディの成分を独立に扱うことが可能
- 原曲と編曲された楽曲のマッチングが可能
- メドレー楽曲の時刻上で楽曲断片が出現している開始時刻と終了時刻が同定可能

本論文ではこれらの要件を満たす楽曲断片検出システムの構築に向け，我々が以前提案した楽曲断片検出法を改良し，Serrà らの相互再帰定量化 (Cross Recurrence Quantification, CRQ) によるカバーソング同定法 (Cover Song Identification, CSI)[7] と，Wang の音声指紋を用いたアルゴリズム [9] を組み合わせた楽曲断片検出手法の提案を行う．CSI タスクの手法は原曲とそれを編曲した楽曲とのマッチングを目的としているため，編曲による影響を考慮した楽曲断片の検出を行うことができる．CRQ を用いた手法では，メロディ抽出の精度に依存することや，メロディ以外の要素がマッチングに影響する可能性がある．音声指紋は，メロディの成分の強い部分を独立に扱うことができるため，同じ時刻に 2 つ以上のメロディが存在する場合にそれぞれのメロディを独立に扱うことで，マッシュアップを考慮したマッチングが期待される．

2. 関連研究

本章では，提案システムで用いる CRQ 手法と音声指紋アルゴリズムの 2 つの先行研究について解説する．

2.1 音声指紋

本研究では，同時刻に 2 つ以上存在するメロディを独立に扱うため，音声指紋を特徴量としてシステムを構築している．音声指紋は，街中で流れている楽曲をスマートフォンなどに通して楽曲認識することを目的とした手法である．Wang はこれを実現するために，高速かつノイズや音質劣化に頑健な特徴量を用いることが必要だとしている [9]．音声指紋では，これを満たす特徴量としてスペクトログラム上でエネルギーの強いポイント（ピーク）を用いている．オーディオファイルから抽出したピークの 1 つを基点と

して他のピークとペアを形成し楽曲認識を行っている（図5）．このペア L_n は，ある 2 つのピーク P_1，P_2 の情報を用いて以下のように表される．

$$L_n : <t_1, f_1, f_2, \Delta t> \quad (1)$$

ここでの t_1，t_2 はそれぞれ P_1，P_2 の時刻を表し，Δt は t_1 と t_2 の差（$t_2 - t_1$）を表す．また，f_1，f_2 はそれぞれ P_1，P_2 の周波数を表す．

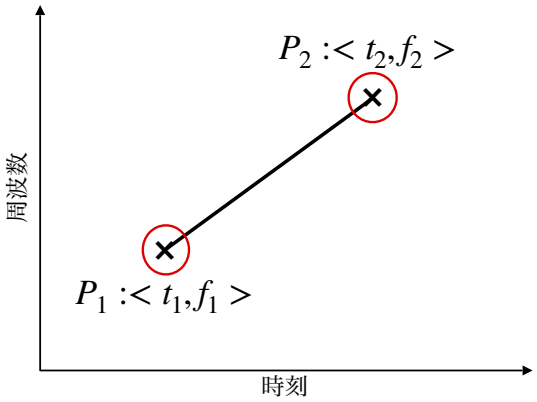


図 5　ランドマークの形成過程

2.2 Cross Recurrence Quantification for Cover Song Identification

本研究では，Serrà らの提案したカバーソング同定法 [7] を用いることで，原曲同士のマッチングに頑健な音声指紋をカバーソング同定に応用している．Serrà らの提案したカバーソング同定法 [7] は，Marwan らが提案した CRQ アルゴリズム [5] をカバーソング同定に応用したものである．Marwan らは CRQ アルゴリズムにおいて，2 つの信号の異なる時刻における類似度行列である行列 CR を定義した．Serrà らはカバーソング同定に行列 CR を応用するため，入力にクロマグラムを用いている．2 つの楽曲信号のクロマグラムから行列 CR を作成した後，動的計画法に基づいて累積値行列を作成し，カバーソング同定を行っている（図 6）．動的計画法に基づくことで，カバーソング同定のマッチングにおいて，テンポの違いを吸収することを期待している．また，この累積値行列は，行列内で最も高い値を持つ地点が類似区間の終了地点であるという特徴を持っており，この値を各楽曲間で比較することでカバーソング同定を行っている．

3. メドレー楽曲における楽曲断片検出手法

本節ではマッシュアップされているメドレー楽曲から楽曲断片を検出する提案手法ついて説明する．提案手法の構成を図 7 に示す．本手法ではまず，メドレー楽曲とそ

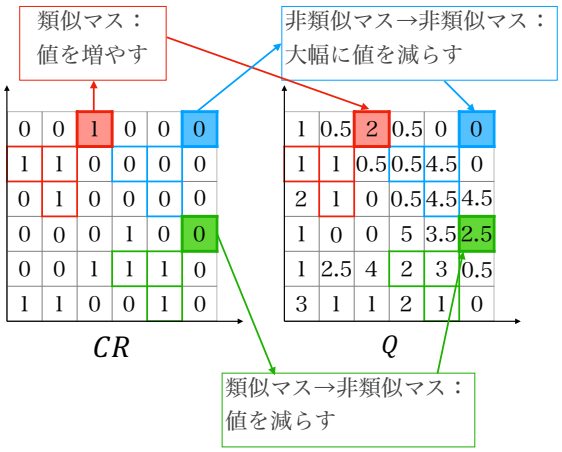


図 6　行列 CR と累積値行列 Q

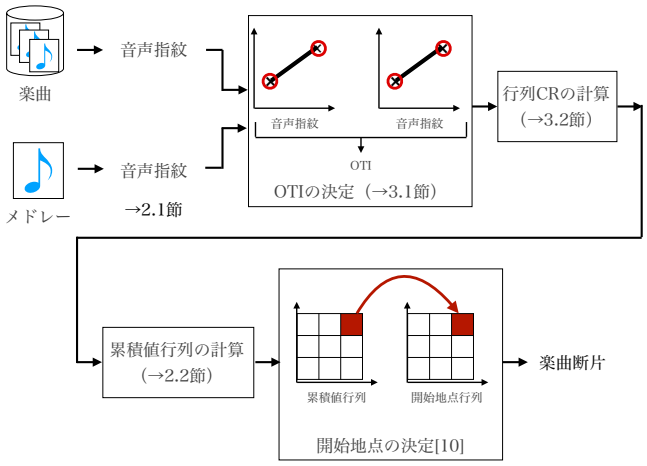


図 7　楽曲断片検出システム構成図

の構成楽曲から音声指紋を抽出し，Optimal Transposition Index(OTI) を計算する．次に，Serrà らの CSI 手法に基づき，行列 CR と累積値行列を作成し，累積値行列の最大値とそのインデックスを得る．そして，佐藤ら [10] の手法に基づき，累積値行列から得られた最大値とそのインデックスを用いて開始地点行列を作成し，同定した開始地点を用いて楽曲断片を検出する（図 8）．

3.1 音声指紋における OTI の計算

本項では，メドレー楽曲と構成楽曲の音声指紋から OTI を計算する手法について説明する．音声指紋によってある 2 つのピーク P_1，P_2 の情報を持つリストが得られる．

$$L_n : <t_1, f_1, f_2, \Delta t> \quad (2)$$

t_1，t_2 はそれぞれ P_1，P_2 の時刻を表し，Δt は t_1 と t_2 の差（$t_2 - t_1$）を表す．また，f_1，f_2 はそれぞれ P_1，P_2 の周波数を表す．このうち，時間の情報を削除した $L'_n : <f_1, f_2>$ を用いて OTI の計算を行う．

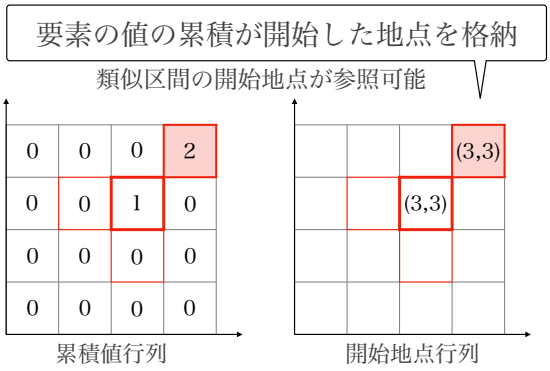


図 8　開始地点行列

$$OTI(L'_A, L'_B) = \underset{-12 \le id \le 12}{\arg\max} \{\sum_{i=0}^{n}\sum_{j=0}^{m} \Theta(L'_{A_n} \cdot \mathrm{circshift}(L'_{B_m}, id))\} \tag{3}$$

ここで，circshift(L', id) は，$L'_n : <f_1 + id, f_2 + id>$ のように，L' のインデックスを id 個ずらす処理を表す．また，$\Theta(\cdot)$ は，

$$\Theta(x) = \begin{cases} 1 & if\ x = 0 \\ 0 & otherwise \end{cases} \tag{4}$$

を表す．本手法では，インデックスをずらしながらペアの周波数が合致する要素を数え上げることで，キーの差分である OTI を計算している．

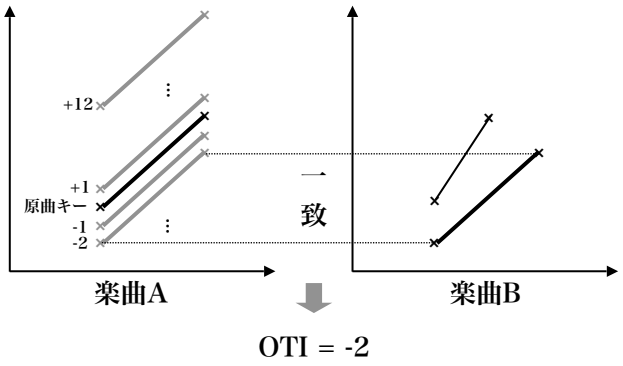


図 9　音声指紋における OTI の計算

最後に，計算した OTI を用いて以下のように L'_A のインデックスをずらす．

$$L'_{A_{Tr}} = \mathrm{circshift}(L'_A, OTI) \tag{5}$$

3.2　音声指紋による行列 CR の作成

本項では，メドレー楽曲と構成楽曲の音声指紋から行列 CR を計算する手法について説明する．3.1 項で計算したピークのリストを用いて，Serrà らの CSI 手法に基づき CRP 行列を作成する 10．Serrà らの CSI 手法では，ある楽曲信号 a における時刻 i のクロマベクトル x とある楽曲信号 b における時刻 j のクロマベクトル y の類似度を，式に基づいて計算し，行列 CR の (i, j) 成分における値としていた．

本手法では，以下の式に基づいて音声指紋を考慮した行列 CR を作成する．

$$\begin{cases} CR_{t_A, t_B} = 1 & if\ \Theta(L'_A \cdot \mathrm{circshift}(L'_B, id)) \\ 0 & otherwise \end{cases} \tag{6}$$

ここで，t_A，t_B はそれぞれ $L_n : <t_1, f_1, f_2, \Delta t>$ の t を表す．この式に基づいて行列を作成することで，ある楽曲信号 A における時刻 i のピーク x とある楽曲信号 B における時刻 j のピーク y の，それぞれ他のピークとのペアの形成の仕方が，周波数に関して同様のものであるかどうかという 2 値の CRP 行列を作成する．

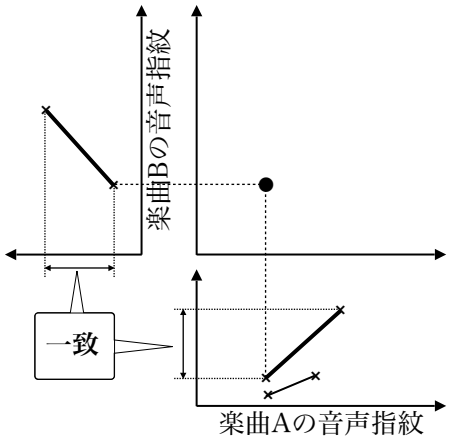


図 10　音声指紋を考慮した行列 CR の作成

4.　評価実験

マッシュアップのあるメドレー楽曲の楽曲断片の検出について，提案手法の精度評価を行った．

4.1　実験条件

検出された楽曲断片を正解データと比較し，F 値を用いて精度評価を行う（図 13）．構成楽曲数が 5〜10 曲かつ 2 分以下のマッシュアップを含むメドレー楽曲とその構成楽曲の，メロディと伴奏のみに限定したデータを用いた．各メドレー楽曲とその構成楽曲はそれぞれ既存のものを用いており，全ての時刻上で 2 曲以上の楽曲でマッシュアップ

されている．楽曲断片の開始時刻・終了時刻の正解データは，本実験の音源データを作成する際に手動で付与した．

音声指紋を抽出する際のパラメータは，デフォルトのものを用いた．CRP 行列を作成する際のパラメータは，Serrà らの検証に基づき，$k = 0.1$，行列 Q を作成する際のパラメータは，累積する値を 100，$\gamma_o = 5.0$, $\gamma_e = 0.5$ とした．

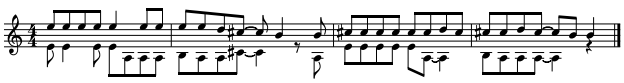


図 11 メドレー楽曲 a でマッシュアップがされている部分の例

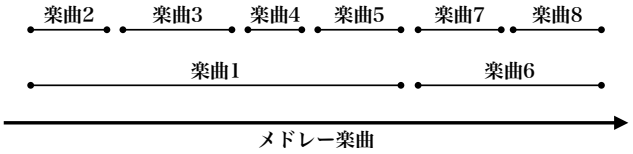
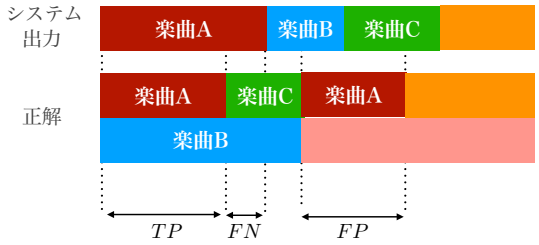


図 12 メドレー楽曲 a における構成楽曲の登場区間



$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F - measure = \frac{2Recall \cdot Precision}{Recall + Precision}$$

		正解	
		区間内	区間外
システム	区間内	TP	FP
	区間外	FN	

図 13 楽曲 A の F 値を計算する例

4.2 実験結果

実験による各メドレー楽曲に対する精度を表 1 に示す．—は構成楽曲が少ないため，ID に対応する楽曲が存在しないことを表す．

全てのメドレー楽曲における F 値の平均は 51.92%であった．F 値が最も高いメドレー楽曲は 54.91%であった（メドレー楽曲 d）．対して F 値が最も低いメドレー楽曲は 50.13%であった（メドレー楽曲は ID2）．

表 1 マッシュアップを含むメドレー楽曲における楽曲断片検出の F 値 (%)

		メドレー楽曲					平均
		a	b	c	d	e	
楽曲	1	70.06	69.89	56.41	67.93	63.02	
	2	21.02	53.85	70.38	18.87	69.24	
	3	31.69	60.84	62.45	53.98	57.15	
	4	72.38	17.33	51.08	68.97	11.16	
	5	57.25	67.95	59.42	48.31	43.68	
	6	67.92	64.08	13.67	66.61	61.80	
	7	68.66	19.41	—	59.86	58.75	
	8	12.07	—	—	53.98	49.91	
	9	—	—	—	60.53	—	
	10	—	—	—	50.03	—	
平均		50.13	50.48	52.24	54.91	51.84	51.92

4.3 結果に対する考察

マッシュアップされたメドレー楽曲に対する楽曲断片検出の結果は，全体的にやや低い精度となった．

全体的に精度を向上するために，音声指紋や CRQ のパラメータを検討する必要がある．本論文で作成される行列 CR に対応させる形で，本論文では CRQ の累積値行列を作成する際の加算値を 100 とした．また，その他のパラメータに関しては，音声指紋や CRQ のそれぞれにおいてデフォルトのパラメータを使用した．本論文の目的は，音声指紋や CRQ が想定している目的や入力と完全に合致していないため，パラメータの検討をする必要がある．

また，楽曲断片検出法が不適切である可能性が考えられる．佐藤ら [10] の手法では，マッシュアップを考慮しない楽曲断片検出を行っており，「メドレー楽曲中では複数の楽曲が同じ時刻において存在しない」という制約を設け，他の楽曲の結果と比較しながら楽曲の登場区間を決定していた．しかし本論文では，他の楽曲と比較することなく楽曲断片検出の結果を出力したため，精度が低くなってしまったと考えられる．

加えて，極端に精度が低くなってしまう楽曲（メドレー楽曲 a の楽曲 2 など）が存在する原因として，OTI の計算手法が適切でないことが考えられる．極端に精度が低くなってしまう楽曲のほぼ全てが，正しく OTI の検出ができていなかった．この問題は，音声指紋のプロット数などのパラメータ調整によって解消できる可能性があるが，本論文で提案した OTI の計算手法が適切でない可能性もあると考えられる．

5. おわりに

本論文では，マッシュアップによってメロディが同時刻に 2 つ以上存在するようなメドレー楽曲を対象に，音声指紋とカバーソング同定法である CRQ を組み合わせることによる楽曲断片検出法を提案した．メドレー楽曲とその構成楽曲から原曲同士のマッチングに頑健な音声指紋を抽出

することで，同時刻で 2 つ以上存在するメロディを独立して扱うことを期待した．音声指紋を原曲とその編曲された楽曲とのマッチングが可能なものにするため，キーとテンポの違いを吸収したマッチング手法を試みた．キーの違いを吸収するため，本手法で音声指紋における OTI の計算手法を提案した．テンポの違いを吸収するため，入力するデータを整形し，Serrà らのカバーソング同定法に基づいて行列 CR と累積値行列を作成した．作成した累積値行列から，佐藤らの手法に基づいて楽曲が類似している区間を検出することで楽曲断片検出を試みた．マッシュアップがされたメドレー楽曲から楽曲断片を検出する精度評価において，精度の低かった楽曲の考察から，パラメータの再検討や楽曲断片検出法を改良する必要性が示唆された．今後はまず，各パラメータの適切な値と本楽曲断片検出法の妥当性について検証する必要がある．また，精度評価においてメロディと伴奏のみに限定した音源を用いるなどの様々な制約を設けたため，実際の音源に対して頑健な手法となるよう改良を行う．

謝辞 本研究を通じて，ご指導を賜りました寺井あすか准教授（公立はこだて未来大学）に深く感謝いたします．本研究は JSPS 科研費（16H01744，16K12560）の助成を受けたものです．

参考文献

[1] Dannenberg, R. B. and Goto, M.:Music Structure Analysis from Acoustic Signals, In D. Havelock, Kuwano, S., Vorländer, M., editors, Handbook of Signal Processing in Acoustics, pp.477-482 (2011).
[2] Foote, J.: Visualizing Music and Audio using Self Similarity, In Proc. ACM International Conference on Multimedia, pp.77-80 (1999).
[3] Glazyrin, N.: Towards extraction of ground truth data from DJ Mixes, Music Information Retrieval Conference (ISMIR), (2017).
[4] Goto, M.: Active Music Listening Interfaces Based on Signal Processing, In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp..1441-1444 (2007).
[5] Marwan, N., Romano, M. C., Thiel, M. and Kurths, J.: Recurrence Plots for the Analysis of Complex Systems. Physics Reports, vol.438, No.5, pp237-329 (2007).
[6] Schwarz, D. and Fourer D.: Towards extraction of grand truth data from DJ Mixes, Music Information Retrieval Conference (ISMIR), (2015).
[7] Serrà, J., Serra, X. and Andrzejak, R. G.: Cross Recurrence quantification, New Journal of Physics, Vol.11, No.9, pp.093017 (2009).
[8] Sonnleitner, R., Arzt, A. and Widmer, G.: Landmark-based audio fingerprinting for DJ Mix monitoring, Music Information Retrieval Conference (ISMIR), (2016).
[9] Wang, A.: An Industrial Strength Audio Search Algorithm, In Proc. International Society for Music Information Retrieval Conference (ISMIR), pp.7-13 (2015).
[10] 佐藤僚太，竹川佳成，平田圭二：カバーソング同定法を応用した楽曲断片検出法の提案，（社）情報処理学会音楽情報科学 (SIGMUS) Vol.2018-MUS-118，No.17 (2018).