

音楽的特徴に基づくメドレー曲の自動生成手法

A Method of Musical Medley Generation based on Musical Features

岡村 亮一*
Ryoichi Okamura

大野 将樹*
Masaki Oono

沼尾 雅之*
Masayuki Numao

1 はじめに

近年, MP3 などのオーディオデータにおける圧縮技術の発展や, iPod などの携帯音楽機器の普及に伴い, 個人が容易に大量の音楽データを所有することが可能となっている. ユーザはいつでも大量の音楽を聴取できるようになったが, 個人のライブラリであっても楽曲は数千曲にも上るため, 楽曲名やアーティスト名などのメタデータで楽曲を検索する従来のインタフェースでは, すべての楽曲を把握するのは非常に困難である. 複数の楽曲の聴取を支援する機能としては, ランダム再生がある. ランダム再生は, ライブラリにあるすべての楽曲に対しランダムに連続再生を行う再生方法であるが, ライブラリの曲数が増加するに従い楽曲のジャンル, 曲調の幅が広がるため, 鑑賞中の楽曲の雰囲気がかかりと変わりユーザが違和感を覚えることがある. そこで, 複数の楽曲の聴取を支援するサービスが求められている.

音楽, 音声を認識する技術は日々進歩しているが, それを活かしたインタフェースが少ないため, エンドユーザはその恩恵を十分に享受できていない. そのため, 日常的に使用できる音楽インタフェースの研究が重要である. 後藤は, 能動的音楽鑑賞インタフェース [1] という音楽音響信号理解研究が, 音楽の聴き方をどのように豊かにできるかを明らかにする取り組みを行っている. これは, 例えば, 楽曲のサビだけを聴いたり, ドラムパートの編集をするインタフェースのことで, 近年, この研究に注目が集まっている. 本研究は, 能動的音楽鑑賞インタフェースの一つの手法として "複数の楽曲からメドレーを自動生成する手法" を提案する.

2 関連研究

堀内らは, 再生中の楽曲と特徴が類似する楽曲を複数提示し, ユーザがその候補を選択すると違和感を覚えないういタイミングで再生を切り替える音楽再生インタフェース Song Surfing [2] を提案した. このシステムでは, 楽曲を 1 秒ごとのフレームに分割し, その波形データの特徴量を得て類似度を比較する. この類似度の高いフレーム同士をクロスフェードさせ, 異なる楽曲を滑らかに繋ぎ合わせている.

佐藤らは, 複数の音楽からメドレーを自動生成するシステム [3] を提案した. このシステムは, メドレーの時間軸に沿った盛り上がりを入力することで, ユーザ好みのメドレーを生成することを可能にしている. まず, ボーカルの音声の有無によって楽曲を分割する. 次に, 分割された楽曲の特徴量を抽出する. 最後に, 特徴量を基に

表 1: AIST Annotation のパート構造とその意味

パート構造	意味
"intro"	楽曲の前奏を表す.
"verse A"	楽曲のサビ (主題) 以外のメロディを表す. 2 種類以上存在する場合は順番に "verse A", "verse B", "verse C", ... となる.
"chorus A"	楽曲のサビを表す. 2 種類以上存在する場合は順番に "chorus A", "chorus B", "chorus C", ... となる.
"bridge A"	楽曲の間奏を表す.
"ending"	楽曲の後奏を表す.
"pre-chorus"	サビの直前にある特徴のある特殊なメロディを表す.
"post-chorus"	サビの直後にある特徴のある特殊なメロディを表す.
"nothing"	無音部分を表す.

入力した盛り上がり沿って分割された楽曲の並び順を求め, 連続再生することでメドレーとしている.

Song Surfing は, ランダム再生を発展させた形で複数の楽曲を延々と連続再生する手法なので, 当然 "1 曲としては出力できない". また, "楽曲中の不自然な位置で楽曲が切り替わる" という問題がある. 文献 [3] のシステムは, "拍の位置が合わない" という問題がある. そこで, それら問題点を解決した手法を提案する.

3 提案手法

3.1 パートの定義

一般に楽曲における 16 小節程度の中区間の繰り返し構造をパートと呼ぶ. すなわち, イントロ, A メロ, B メロ, サビ, エンディングといったものを指す. 本研究では, 表 1 に示す AIST Annotation [4] のパート構造をパートと呼ぶことにする.

3.2 メドレーの定義

本研究では, 次の 3 つの要件を満たすものをメドレーと定義する.

1. すべての楽曲から歌詞の存在するパートを 1 パート

*電気通信大学大学院情報理工学研究科情報・通信工学専攻

ずつ用いる。

2. イントロで始まりエンディングで終わる。
3. 楽曲の途中にイントロとエンディングが入らない。

1 は、ある数曲をメドレーにしようとするとき、一般的に対象の楽曲すべてから歌詞の存在するパートを用いるためである。2 は、B メロから急に再生を開始したり、サビが終了したら急に再生を終えるのは一般的ではないためである。3 は、ある 1 つのメドレー内で楽曲が終了したと思ったら再び楽曲が開始することは一般的ではないためである。

次に、メドレー内の楽曲 (パート) 間の繋がりとして以下の 2 つの繋がりを考える。

1. 局所的接続
2. 大局的接続

局所的接続とは、メドレーにおける隣り合う 2 パート間の繋がりである。大局的接続とは、メドレー全体を通してのパートの展開のことを言う。パートの展開とは、時間軸に沿った盛り上がりの変化である。例えば、ポップスでは、A メロ B メロ サビとなるにつれて盛り上がり、サビ A メロとなると落ち着く。

メドレーの自動生成におけるナイーブな手法としてランダムにパートを選択しメドレーを生成する手法が考えられる。前述のメドレーの定義より、一口にメドレーといってもイントロ B メロ B メロ B メロ エンディングというものもメドレーに含まれる。ナイーブな手法では、このようなメドレーが生成されることは十分に有り得る。そこで、人が違和感を覚えないメドレーを生成する手法を考える必要がある。そのために、滑らかな局所的接続、滑らかな大局的接続という 2 つの接続方法を定義する。本研究が目指す自然なメドレーとは、滑らかな局所的接続と滑らかな大局的接続を満たすメドレーである。

3.3 滑らかな局所的接続

音響的に滑らかに繋がる局所的接続である。ある楽曲から異なる楽曲に再生を切り替える際に人が違和感を感じる要因として、音響的な要素、例えば、テンポや音量の差が大きく異なることが挙げられる。滑らかな局所的接続を満たすための指標として以下の 5 つの指標を考える。

- 和音
- テンポ
- 音量
- 楽曲の盛り上がり度合い
- 楽曲の複雑さ

楽曲の盛り上がり度合いは、どれだけ高音が含まれているかを表す指標である。一般的に楽曲が盛り上がっているほど高音が多く含まれる。楽曲の複雑さは、どれだけ多くの音階が同時に鳴っているかを表す指標である。

3.4 滑らかな大局的接続

曲らしいパートの展開になっており、主題を 1 つ以上含むような大局的接続である。メドレー全体を通して聴いたときに違和感を感じる要因として、主題 (ポップスでは一般的にサビのことを言う) が含まれていないことが挙げられる。また、急激に盛り上がり、落ち着いたりすることも違和感の要因の 1 つである。滑らかな大局的接続を満たすためにはパートの遷移に制約をつけることでそれを解決する。実際のポップス曲 100 曲を調査しパートの遷移図を作成した。パートの遷移図を図 1 に示す。

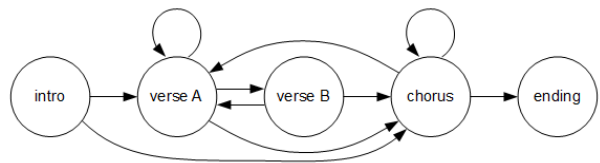


図 1: パートの遷移図

4 提案システム

提案システムの流れを図 2 に示す。提案システムは、入力部、特徴抽出部、接続部、出力部からなる。以下、詳細を述べる。

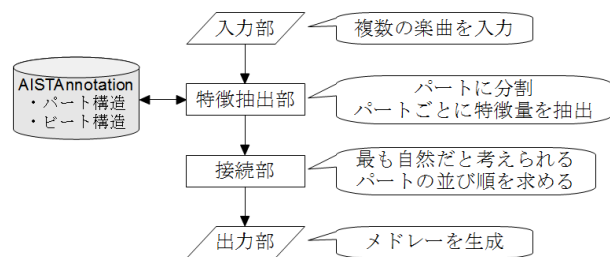


図 2: 提案システム

4.1 入力部

提案システムへの入力、任意の数の楽曲である。各楽曲のフォーマットは wav とする。

4.2 特徴抽出部

入力楽曲をパートに分割し、パートごとに特徴量を抽出する。特徴量は、クロマベクトル、BPM (Beats Per Minute)、聴感上の音量、周波数重心、音程パワー分散を用いる。特徴抽出区間は各パートの開始から 1 秒間である。一般に和音進行が 1 秒内で変化し続けることは極めてまれであることから、少なくとも 1 秒間は同じ和音進行が続くという仮定に基づいて特徴抽出を行う。

4.2.1 パート構造の抽出

AIST Annotation に基づき各入力楽曲のパート構造を得る。ある連続したパート同士を連結した方が良い場合がある。連結するとは、"開始時間 t_{s_A} 、終了時間 t_{e_A} 、

パート名 A ”と ”開始時間 t_{s_B} , 終了時間 t_{e_B} , パート名 B ”の 2 つのパートが存在するとして, パート A の直後にパート B が来るとき ($t_{e_A} = t_{s_B}$), 2 つのパートを ”開始時間 t_{s_A} , 終了時間 t_{e_B} , パート名 A ”という 1 つのパートとみなすことを言う. ただし, パート B が ending の場合は, ”開始時間 t_{s_A} , 終了時間 t_{e_B} , ending ”とみなす.

まず, intro と intro の直後のパート, ending と ending の直前のパートを連結する. これは, AIST Annotation の intro, ending に歌詞が含まれないためである.

次に, 隣り合う chorus 同士を連結する. AIST Annotation では, 例えば, 1 番のサビに ”chorus A ”, または ”chorus B ”が複数存在する場合があるためである. 一般的なポピュラー音楽のサビに見られる同じ繰り返し構造を 2 回以上用いたメロディの場合に, このようなアノテーションになる. また, 同様に隣り合う bridge も連結する.

最後に, 楽曲に bridge が存在したら bridge の前後のパートと合わせて 3 つのパートを図 3 のように 4 つのパートとして扱う. 一般的なメドレー, 特にカラオケでのメドレーがそうであるように, 楽曲をメドレーにする際にはどの楽曲も歌詞の存在するパートを用いる. よって, 提案システムでは, bridge 単独のパートを存在しないようにし, bridge は前後のパートと連結する. ただし, bridge を含まない方が違和感がなくなる場合も当然存在し, 逆に含んだ方が良い場合も考えられるので, 図 3 に示すようなパート構造にした.

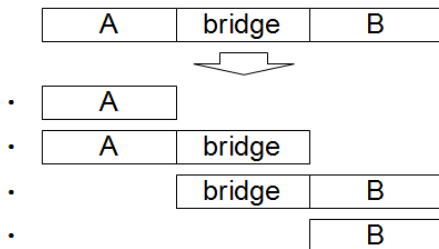


図 3: bridge の処理

4.2.2 ビート構造の抽出

AIST Annotation に基づき入力楽曲のビート構造を得る. ビート構造とは, 楽曲における拍の位置の構造である. 得られたビート構造は, メドレー生成の際に小節における最初の拍の位置を合わせる目的の他, BPM の計算に用いられる.

4.2.3 特徴抽出

入力の各楽曲に対して, 3.3 節の 5 つの指標と対応させパートごとに次の 5 つの特徴量を抽出する.

- (1) クロマベクトル
- (2) BPM
- (3) 聴感上の音量
- (4) 周波数重心

(5) 音程パワー分散

特徴抽出区間は, 各パートの開始から 1 秒間である.

楽曲 X のパート X_i における音響信号の開始から 1 秒間のデータを $w = (w_1, w_2, \dots, w_N)$ とする. ただし, 楽曲のサンプリング周波数を F_s とすると, N は $1 \times F_s$ である. BPM を除き, w を用いて各特徴量を計算する. w の例を図 4 に示す.

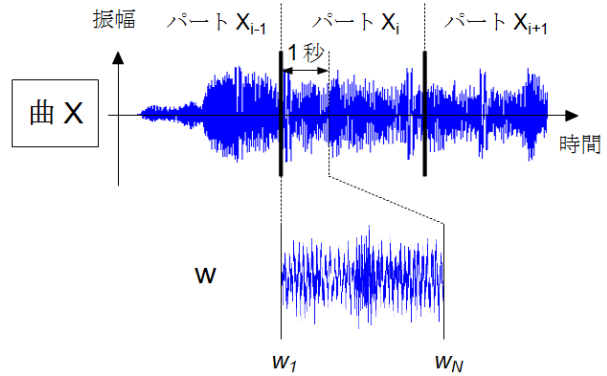


図 4: 特徴抽出区間

- (1) のクロマベクトルは, 平均律音階に相当する周波数パワーを 1 オクターブ (12 音階) に射影・加算した 12 次元のベクトルである. クロマベクトルは, Shepard の研究 [5] で用いられたクロマを周波数軸としてパワーの分布を表現した特徴量である. これは和音を表す.
- (2) の BPM は, 60 秒間の拍数である. これはテンポを表す. 現在のビートトラッキングは, BPM が本来の 2 倍, または 1/2 倍の数値になることが多い. しかし, テンポは特徴量の中でも非常に重要な要素であるため 2 倍 (1/2 倍) の数値になることは致命的である. 本研究では, BPM の計算においてのみ AIST Annotation のビート構造情報を用いて, より正確な数値になるように計算した.
- (3) の聴感上の音量 f_{vol} は, 振幅をデシベルに変換した値の平均値である. これは音量を表す.
- (4) の周波数重心 f_{fg} は, 周波数パワーの周波数重み付き平均値である. これは楽曲の盛り上がり度合いを表す.
- (5) の音程パワー分散 f_{cv} は, クロマベクトルの分散である. これは楽曲の複雑さを表す.

4.3 接続部

得られた特徴量から最も自然だと考えられるパートの並び順を求める. まず, パート X_i からパート Y_i に連続再生したときに音響的に滑らかに繋がるような距離 $D(X_i, Y_j)$ の計算を行う (局所的接続). 次に, その $D(X_i, Y_j)$ を基にメドレーとして最も自然だと考えられるパートの並び順を求める (大局的接続). 最も自然だと考えられるパートの並び順は, 動的計画法によって求める.

4.3.1 距離の定義

特徴ベクトル $f = (\text{BPM}, f_{vol}, f_{fg}, f_{cv})$ を用いて, 楽曲 X のパート i と楽曲 Y のパート j の距離 $D(X_i, Y_j)$

を式 (1) のように定義する．ただし，和音が滑らかな局所的接続を満たすために最も優先されるべきで，和音が等しければある程度距離が離れていても滑らかな局所的接続を満たすという仮定の基，クロマベクトルは特徴ベクトルに含めない．

$$D(X_i, Y_j) = \|f_{X_{i+1}} - f_{Y_j}\| \quad (1)$$

式 (1) の 1 つ目の特徴ベクトルが $i + 1$ になっていることに注意する．つまり， $D(X_i, Y_j)$ は図 5 のように X_{i+1} の特徴ベクトルと Y_j の特徴ベクトルのユークリッド距離をとる．これは， X_{i+1} の出だしと Y_j の出だしを比較

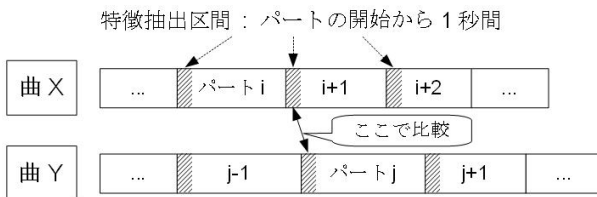


図 5: 特徴ベクトルの比較位置

することによって， X_i から X_{i+1} への繋がりを考慮し， $D(X_i, Y_j)$ が 0 に近いほど X_i から Y_j にクロスフェードしたときに音響的に滑らかに繋がりが，滑らかな局所的接続を満たすだろうというねらいがある．また，各特徴量は， $D(X_i, Y_j)$ の範囲が 0 から 1 になるようにあらかじめ正規化しておく．

4.3.2 パートの並び順を求めるアルゴリズム

ノードをパート，エッジを曲の繋がりを，エッジの重みを距離 D とすると，これはグラフ問題に帰着できる．まず，任意の ending を部分グラフに加える (図 6(a))．次に，図 1 の遷移を満たし，直前に部分グラフに加えられたパートと距離 D が最小のパートを部分グラフに加える (図 6(b))．これを繰り返すことで最も自然だと考えられるパートの並び順を求める．

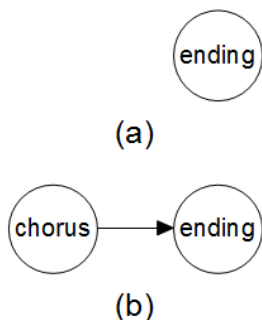


図 6: アルゴリズムの手順

距離 D が最小のパートを順々に加えていくと，後半に加えられるパートの距離 D は，どうしても最初の方に加えられた距離 D よりも大きくなってしまふ．そこで，始点を ending とした．これは，メドレーを通して聴いたときに終盤のパートの方が印象に残りやすいので，終盤のパートから違和感のないパートを選択していくこ

とによりメドレー全体を通しての印象を良くするねらいがある．パートの並び順を求めるアルゴリズムを下記に示す．

1. x, s : パート $S_{candidate}$, S_1 : パートの集合
2. M : リスト t_{chroma} , t_{dist} : 定数
3. 初期条件: $x \leftarrow$ 任意の ending, $M \leftarrow \phi$
- 4.
5. $insert(M, x)$
6. loop
7. $S_{candidate} \leftarrow x$ に接続可能なパートの候補の集合
8. if($S_{candidate} = \phi$) break
9. $S_1 \leftarrow \phi$
10. for each($s \in S_{candidate}$)
11. if(x と s のクロマベクトルのコサイン類似度 $\geq t_{chroma}$ and $D(x, s) \leq t_{dist}$)
12. $S_1 \leftarrow S_1 \cup s$
13. end
14. end
15. if($S_1 \neq \phi$) $S_{candidate} \leftarrow S_1$
16. $x \leftarrow S_{candidate}$ のうち x との距離 D が最小のパート
17. $insert(M, x)$
18. end
19. M を出力する

5 行目，17 行目の $insert(M, x)$ は，リスト M の先頭にパート x を追加する関数である．

全ての入力楽曲の ending についてこのアルゴリズムを適用する．そして，全ての入力楽曲の M に対して隣同士のノードにおける距離 D の総和が最小の M を最も自然だと考えられるパートの並び順とした．7 行目の x に接続可能なパートの候補の集合は， M にない楽曲のうち図 1 の遷移を満たすパートの集合である．本研究では， $t_{chroma} = 0.8, t_{dist} = 0.2$ とした．以下に例を示す．

入力楽曲を "楽曲 1", "楽曲 2", "楽曲 3", "楽曲 4" の 4 曲とし，各入力楽曲のパート構造を "intro", "chorus", "ending" とする．まず，ステップ 3 の任意の ending として "楽曲 1 の ending" を選択する．ステップ 5 で M に "楽曲 1 の ending" を加える．ステップ 7 で図 1 を参照し，"楽曲 1 の ending" へ遷移できる "楽曲 2 の chorus", "楽曲 3 の chorus", "楽曲 4 の chorus" を $S_{candidate}$ の要素とする．メドレーの定義 "すべての楽曲から歌詞の存在するパートを 1 パートずつ用いる．", "楽曲の途中にイントロとエンディングが入らない．" より，"楽曲 1 の intro", "楽曲 1 の chorus", "楽曲 2 の intro", "楽曲 3 の intro", "楽曲 4 の intro" は候補には含まれない．ステップ 10 で $S_{candidate}$ の各パートと "楽曲 1 の ending" のコサイン類似度，距離 D を比較し，条件を満たしていたらそのパートを S_1 に加える．条件を満たさなかったとする．ステップ 16 で $S_{candidate}$ のうち "楽曲 1 の ending" との距離 D が最小のパート ("楽曲 3 の chorus" だとする) を x に代入する．ステップ

17で"楽曲3のchorus", "楽曲1のending"を M の要素とする。ステップ7で"楽曲2のchorus", "楽曲4のchorus"を $S_{candidate}$ の要素とする。ステップ10で条件を満たさなかったとする。ステップ16で $S_{candidate}$ のうち"楽曲3のchorus"との距離 D が最小のパート("楽曲2のchorus"だとする)を x に代入する。ステップ17で"楽曲2のchorus", "楽曲3のchorus", "楽曲1のending"を M の要素とする。ステップ7でメドレーの定義"イントロで始まりエンディングで終わる。"より,"楽曲4のintro"を $S_{candidate}$ の要素とする。ステップ10で条件を満たさなかったとする。ステップ16で $S_{candidate}$ のうち"楽曲2のchorus"との距離 D が最小のパート"楽曲4のintro"を x に代入する。ステップ17で"楽曲4のintro", "楽曲2のchorus", "楽曲3のchorus", "楽曲1のending"を M の要素とする。 M をendingを"楽曲1のending"としたときのパートの並び順とする。同様に,"楽曲2のending", "楽曲3のending", "楽曲4のending"についても適用し,各 M のうち隣同士のノードにおける距離 D の総和が最小の M を最も自然だと考えられるパートの並び順とする。

4.4 出力部

接続部で求めたパートの並び順でパートをクロスフェードして出力する。クロスフェードするには,距離の定義から元の楽曲の展開を保つために特徴抽出区間の部分を一致させるようにクロスフェードしなければならない。また,小節における最初のビートの位置も一致させる。よって,図7のようになる。

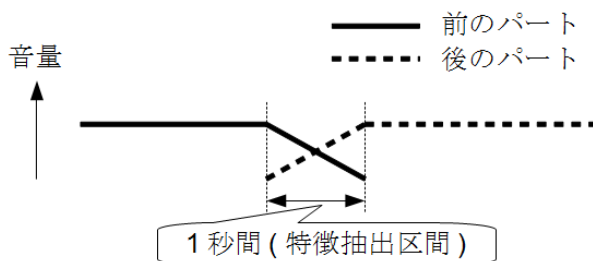


図 7: クロスフェード位置

5 評価実験

5.1 実験環境

実験の目的は,提案システムが滑らかな局所的接続,滑らかな大局的接続を満たしているか,そして,提案システムがナイーブな手法と比較して改善されているかを評価することである。実験はMATLABでMIRtoolbox[6]を用いてプログラミングを行った。実験データとしてはRWC研究用音楽データベース[7]のポピュラー音楽データベースの楽曲を用いた。これらの楽曲にはパート構造やビート構造のアノテーションがついており,提案システムのパート構造の抽出とビート構造の抽出はこのAIST Annotationに基づいて行った。RWC研究用音楽データベースの楽曲は,サンプリング周波数44.1kHz,

量子化ビット数16ビットである。また,ステレオであるため,計算のしやすさからモノラルに変更した。

5.2 滑らかな局所的接続の評価実験

5.2.1 目的

本実験の目的は,提案システムが滑らかな局所的接続を満たしているかを評価することである。

5.2.2 方法

違和感なくパートを連続再生させるための特徴抽出区間としては,他にも図8のように,パートの終了とパートの開始を比較して類似していたら滑らかに繋がる,という考え方もある。この距離計算方法を距離計算方法Aと呼ぶことにする。距離計算方法Aと比較して提案システムが滑らかな局所的接続を満たしているかを評価する。

まず,被験者に滑らかな局所的接続の例を聴いてもらう。次に,特定のパート X から異なるパートへの連続再生を考えると,提案システムにより計算されたパート X に近いパート Y と,距離計算方法Aにより計算されたパート X に近いパート Z を求める。パート X からパート Y への連続再生と,パート X からパート Z への連続再生をそれぞれ聴かせ,どちらが滑らかな局所的接続を満たしているかを回答させる。この際,あらかじめ元の楽曲のパート X からパート $X+1$ への連続再生も聴かせておく。被験者は26名である。実験データはRWC研究用音楽データベースのポピュラー音楽データベースの楽曲番号1から100を用いた。パート X は,ランダムに5パート選択した。

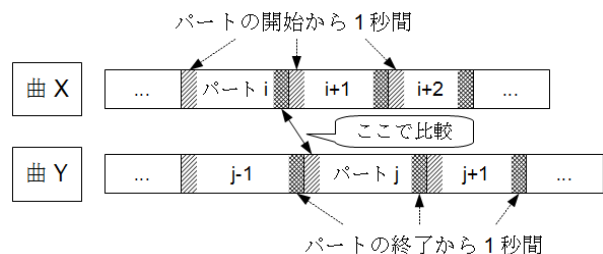


図 8: 距離計算方法 A

実験の結果,全被験者の評価に対し,帰無仮説を"被験者が提案システムによる連続再生を選択する確率は50%"として有意水準5%の二項検定を行うと有意差が認められた。つまり,被験者は無作為に回答していたわけではなく,意図的に提案システムと回答していたことになる。よって,提案システムは滑らかな局所的接続を満たしていることが示された。

5.3 滑らかな大局的接続の評価実験

5.3.1 目的

本実験の目的は,提案システムが滑らかな大局的接続を満たしているか,また,滑らかな大局的接続を満たすメドレーが自然であるかを評価することである。

5.3.2 方法

提案システムに図 1 で示したパートの遷移の制約をつけたメドレーとつけないメドレーの 2 種類を聴かせ、滑らかな大局的接続について説明したうえで滑らかな大局的接続を満たしている方と自然なメドレーだと思おう方を回答させる。ただし、制約をつけないメドレーも滑らかな局所的接続は満たすものとする。被験者は 6 名である。実験データは RWC 研究用音楽データベースのポピュラー音楽データベースの楽曲番号 1 から 50 を用いた¹。同一の入力楽曲 (10 曲) に対し、パートの遷移の制約をつけたメドレーとつけないメドレーのうち "滑らかな大局的接続を満たしていると思おう方" と "自然なメドレーだと思おう方" を回答させる設問を 5 つ用意した。

実験の結果、"滑らかな大局的接続を満たしていると思おう方" に関し、帰無仮説を "被験者が制約をつけたメドレーを選択する確率は 50%" として有意水準 5% の二項検定を行うと有意差が認められた。つまり、被験者は無作為に回答していたわけではなく、意図的に制約をつけたメドレーを回答していたことになる。よって、人は滑らかな大局的接続を認識でき、提案システムは滑らかな大局的接続を満たしていることが示された。

同様に、"自然なメドレーだと思おう方" に関しても有意水準 5% の二項検定を行ったが有意差が認められなかった。これは、滑らかな局所的接続と滑らかな大局的接続に対する重みの個人差によるものと考えられる。提案システムが滑らかな大局的接続を満たしていると回答しているにも関わらず、自然なメドレーとしては制約をつけないメドレーの方が自然だと回答するような逆転するケースの数は、被験者によって大きな開きがあった。

5.4 提案システムの評価実験

5.4.1 目的

本実験の目的は、提案システムがナイーブな手法と比較して改善されているかを評価することである。

5.4.2 方法

メドレーの自動生成におけるナイーブな手法としてランダムにパートを選択しメドレーを生成する手法が考えられる。このナイーブな手法と提案システムによるメドレーを比較し、自然なメドレーだと思おう方を回答させる。被験者は 6 名である。実験データは RWC 研究用音楽データベースのポピュラー音楽データベースの楽曲番号 1 から 50 を用いた¹。同一の入力楽曲 (10 曲) に対し、ナイーブな手法、提案システムによるメドレーのうち "自然なメドレーだと思おう方" と "特に違和感を感じた箇所" を回答させる設問を 5 つ用意した。

実験の結果、被験者が提案システムによるメドレーを選択した数は、設問 1 が 5 人、設問 2 が 5 人、設問 3 が 2 人、設問 4 が 3 人、設問 5 が 4 人となった。帰無仮説を "被験者が提案システムによるメドレーを選択する確率は 50%" として有意水準 5% の二項検定を行ったが有

意差は認められなかった。原因を分析するために、まず特に違和感を感じた箇所についてを分析する。

メドレー内の楽曲のパートと次の楽曲のパートの平均距離と、被験者が特に違和感を感じた箇所についての平均距離を計算した。1 箇所の平均距離よりも被験者が特に違和感を感じた箇所についての平均距離の方が大きくなっているため、被験者はメドレーの中でも特に距離 D の離れている箇所に違和感を感じていることがわかる。

また、メドレー内の楽曲のパートと次の楽曲のパートにおける特徴量の差の絶対値の平均と、被験者が特に違和感を感じた箇所についての平均を計算すると、特徴量の中でも特に前後の楽曲の BPM の差に敏感であることが示された。よって、各特徴量に重みを付けるべきで、特に BPM に高い重みを付けるべきであることが示された。

次に、パートの展開に関して分析する。A メロ B メロ、B メロ サビ、A メロ サビは一般的に盛り上がる展開で、サビ A メロは落ち着く展開である。被験者が特に違和感を感じた箇所の中で盛り上がる展開と落ち着く展開それぞれについて、前後のパートにおける BPM の差の平均を計算すると、盛り上がる展開の箇所では BPM が下がったときに、落ち着く展開の箇所では BPM が上がったときに違和感を感じていることを示された。よって、盛り上がる展開では BPM が上がるように、落ち着く展開では BPM が下がるようにメドレーを生成することで、提案システムを改善できると考えられる。

被験者からは、"ボーカルの性別が前後で変わると違和感を感じた。" という意見が多かった。全被験者の特に違和感を感じた箇所計 64 箇所 (異なる被験者が同じメドレーの同じ箇所を指定した場合の重複を許す) のうち、32 箇所は前後でボーカルの性別が異なっていた。また、"ラップが入ると違和感を感じた。" という意見もあった。

提案システムによるメドレーのうち、被験者が違和感を感じた箇所の合計が一番多かったメドレーは、設問 4 のメドレーであった。これは、間奏が多く含まれていたからだと考えられる。

最後に、以上の考察を踏まえてナイーブな手法と比較して改善されなかった原因を分析する。被験者の半数以上がナイーブな手法によるメドレーの方が自然だと回答した設問 3, 4 のメドレーが提案システムがナイーブな手法と比較して改善されなかった原因であると考え、この 2 つのメドレーについて分析する。設問 3 のメドレーで全被験者の特に違和感を感じた箇所は 12 箇所 (重複を許す) であったが、そのうち 6 箇所 (重複を許す) がラップを含む局所的接続、5 箇所 (重複を許す) がボーカルの性別が前後で異なる局所的接続であった。設問 4 のメドレーは、前述の通り間奏が非常に多かった。よって、提案システムの改善点としては、間奏の数を考慮しパートの遷移の制約に組み込む、ボーカルの性別を抽出する、ラップのパートを除外する、という改善点が挙げられる。また、特徴量を重み付けする、パートの展開による前後の楽曲における BPM の差の考慮することで、さらなる改善が見込める。

¹メドレーの入力楽曲を 10 としているため、実験のコスト考えてメドレーが 5 つとなるようにデータベースの楽曲は 50 曲とした。

6 おわりに

6.1 まとめ

本研究は、複数の楽曲から人が聴いて違和感のないメドレーを自動生成する手法を提案した。提案手法をシステムとして実装し、提案手法の評価実験を行った。

実験の結果、ナイーブな手法(ランダムにパートを選択しメドレーを生成する手法)によるメドレーと比較して提案システムによるメドレーの方が違和感がないと被験者が回答した数は過半数を超えたものの、有意差は認められなかった。また、いくつかの改善の余地があることが示唆された。

6.2 今後の課題

今後の課題として、次の 6 つの改善点が挙げられる。

- (1) パートの遷移の制約に対するオプション
- (2) 特徴量の重み付け
- (3) パートの展開による前後の楽曲における BPM の差の考慮
- (4) 間奏の考慮
- (5) ボーカルの性別の考慮
- (6) ラップを含むパートの除外

(1) 5.3 節の実験から、人は滑らかな局所的接続と滑らかな大局的接続に対する重みに個人差があることが示された。接続部で滑らかな大局的接続を満たすためのパートの遷移の制約を入れるか入れないかのオプションを付けることで提案システムの改善が見込める。

(2) 滑らかな局所的接続を満たすための特徴量に重み付けをする必要がある。実験から、人は BPM の差に大きく左右されることが示されたので、BPM の重みを特に大きくする必要がある。

(3) パートの展開によって前後の楽曲における BPM の差を考慮する必要がある。盛り上がる展開の場合には前の楽曲よりも BPM が上がるように、落ち着く展開の場合には前の楽曲よりも BPM が下がるようにすることで提案システムを改善できる。

(4) 間奏が非常に多い場合に違和感を感じることを示された。滑らかな大局的接続を満たすためのパートの遷移の制約に間奏を組み込み、さらに間奏の数を考慮することで提案システムを改善できる。

(5) 5.4 節の実験から、前後の楽曲でボーカルの性別が異なると違和感を感じることを示された。できるだけ同性のボーカルが続くようにする必要がある。ボーカルの性別を抽出するためには、メタデータであるアーティスト名などを用いれば良い。アーティスト名は、ウェブから容易に取得できるので現実的である。

(6) 提案システムによるメドレーよりもナイーブな手法によるメドレーの方が自然であると多くの被験者が回答したメドレーがあった。このメドレーはラップを含んでおり、そこに違和感を感じるという意見があった。しかし、パートにラップを含んでいるかどうかを得ることは困難である。ラップは、同じ楽曲の他のパートの歌詞に

比べてモーラ数が多くなる傾向があるので、歌詞に着目しモーラ数を計算することでラップを含むパートを除外できると考えられる。

謝辞 本研究の実験において "RWC 研究用音楽データベース: ポピュラー音楽" (RWC-MDB-P-2001) を使用した。

参考文献

- [1] 後藤真孝. 音楽音響信号理解に基づく能動的音楽鑑賞インタフェース (インタフェース). 情報処理学会研究報告. [音楽情報科学], Vol. 2007, No. 37, pp. 59–66, 2007-05-10.
- [2] 堀内直明, 園田俊行, 田中浩司, 田中淳一, 長沢秀哉, 裁山真一. 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, 2007.
- [3] 佐藤学, 鈴木健史, 新美怜志, 濱川礼. 4r-7 複数の音楽からメドレーを自動生成するシステム (音楽演奏・生成, 学生セッション, 人工知能と認知科学). 全国大会講演論文集, Vol. 71, No. 2, 2009.
- [4] M. Goto. AIST annotation for the RWC music database. In *Proc. ISMIR*, pp. 359–260. Citeseer, 2006.
- [5] R.N. Shepard. Circularity in judgments of relative pitch. *Psychol. Rev.*, Vol. 36, pp. 172–180, 1929.
- [6] MIRtoolbox. <https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox>.
- [7] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一. 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 (<特集> 音楽情報科学). 情報処理学会論文誌, Vol. 45, No. 3, pp. 728–738, 2004-03-15.