1T-04

声質の類似性に基づく歌手マップの作成

冨塚　美歩†　　岩野　公司†
東京都市大学†

1. はじめに

現在，大量に存在するデジタル音楽からユーザが望む音楽を素早く見つけ出すための技術として，様々な楽曲検索が存在している．検索クエリとして，曲名や歌手名，歌詞，メロディ（鼻歌）などを利用する検索が実用化されているが，歌手の「声質」に着目し，お気に入りの歌手の声質に似た別の歌手の楽曲を検索する手法が提案されている[1]．従来研究[1]では，声質の類似度の計算手法を提案した上で，入力クエリ（楽曲）を歌っている歌手の声質に対し，類似度が上位の楽曲を検索結果として出力するシステムの実装を行っている．

本研究では，歌手の声質による楽曲検索を効率的に行うため，声質の類似度に基づいて歌手間の相互関係を 2 次元平面上に表現した「歌手マップ」を利用することを提唱する．このマップでは，声質が類似している歌手同士が近くに配置されるもので，お気に入りの歌手の周辺を調べることで，声質の似た歌手を素早く発見することができる．本研究では，この声質による歌手マップの作成方法を数種類提案し，主観評価実験によって作成されたマップを評価する．

2. 声質による歌手マップ生成手法の提案

2.1 提案手法の流れ

図 1 に歌手マップの作成の流れを示す．まず，対象となる各歌手の楽曲について，帯域通過フィルタに基づくボーカル抽出[2]を施し，音楽部分を抑制した歌声データを作成する．次に，その歌声データを 12 次元のメル周波数ケプストラム係数（MFCC）とその一次微分，対数パワーの一次微分成分の計 25 次元の音響特徴量に変換する．得られた特徴量を学習データとして，歌手ごとに，図 2 に示す 2 状態の隠れマルコフモデル（HMM）の学習を行う（歌手モデル）．モデルの第 1 状態（s_1）には，楽曲の開始，中間，終了時に存在する非歌唱区間が，第 2 状態（s_2）には歌唱区間が反映されると考えられるため，この

Vocalist map generation based on vocal timbre similarity
Miho Tomitsuka†, Koji Iwano†, †Tokyo City University

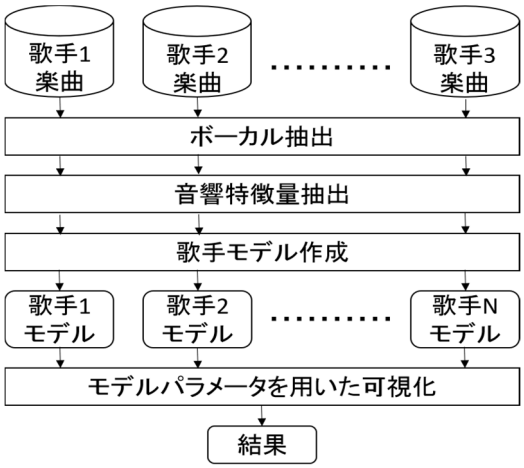

図 1　歌手マップの作成の流れ

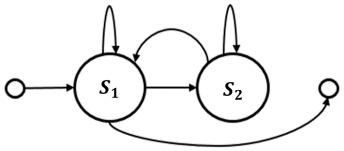

図 2　利用する歌手モデルのトポロジー

第 2 状態のパラメータを利用して，歌手の類似性の可視化を行う．可視化の方法としては，次節以降で説明する 3 つの方法を試みる．

2.2 平均ベクトルを利用した SOM による可視化

各歌手について，モデルの第 2 状態から最も混合重みの大きい正規分布の平均ベクトル（25 次元）を取り出して歌手の声質を表す特徴ベクトルとし，それに自己組織化マップ（SOM）[3]を適用することで，2 次元空間への可視化を行う．SOM では，ユークリッド距離が近い特徴同士が近くに写像されるようにニューラルネットワークが学習され，それにより可視化が実現される．

2.3 モデル間距離を利用した SOM による可視化

2.2 節の手法では，歌手モデルのパラメータのユークリッド距離に基づいて可視化が行われる．一方，話者照合における話者性の分析[4]では，話者間距離を話者モデル間のカルバック・ライブラー（KL）情報量に基づいて定義することの有効性が示されている．

そこで，SOM の入力特徴ベクトルに，この距離を明示的に導入する．歌手 a と歌手 b 間の距離 $D(a,b)$を式(1)のように定義し[4]，各歌手について，全ての歌手との間の距離を算出して，その数値を要素とする特徴ベクトルを構成し，SOM による可視化を行う．

$$D(a,b)=\frac{1}{2}\left\{\sum_{k}^{K} w_k^a \cdot \min_l SKL\left(N_k^a, N_l^b\right)+\sum_{k}^{K} w_k^b \cdot \min_l SKL\left(N_l^a, N_k^b\right)\right\} \quad (1)$$

ここで，N_k^x, w_k^x は歌手 x のモデルの第 2 状態の混合正規分布における k 番目の正規分布と，その混合重みを表している．また，SKL は対称化した KL 情報量である．

2.4 モデル間距離を利用した CMDS による可視化

古典的多次元尺度構成法（CMDS）[5]では，データ間の距離行列を直接入力として与え，可視化を行うことができる．そこで，歌手間の距離を 2.3 節の KL 情報量に基づく歌手モデル間の距離で定義し，全歌手間の距離情報（距離行列）を CMDS の入力とすることで可視化を行う．

3. 歌手マップの評価実験

3.1 実験条件

2 章で説明した，3 手法で歌手マップを作成し，10 名の被験者（大学生）による主観評価によって有効性の評価を行った．対象歌手は大学生に馴染みのある日本人の男性歌手 19 名とし，各歌手あたり 1〜5 曲の楽曲を利用した．また，歌手モデルの混合数は 16 とした．

被験者には，3 手法で作成したマップに加え，ランダムで歌手を配置したマップの 4 つを提示し，指定した歌手の近く（遠く）に位置している歌手を複数選んで実際のその歌手の楽曲を聴き，声質の類似性が適切に表現されているかを，評価してもらう．具体的には，全ての手法のペアについて，どちらの手法が優れているかを対比較してもらった．指定する歌手は被験者あたり 5 人で，被験者によって変更している．

3.2 実験結果

4 つの手法のプリファレンススコアを図 3 に示す．この結果，平均ベクトルを利用した SOM では十分な表現力が得られていないが，モデル間距離を利用した SOM はランダム配置に比べ，有意に高い評価となった．また，モデル間距離を利用した SOM と CMDS との間には有意差は見られなかった．したがって，KL 情報量に基づく歌手間距離が効果的であったと考えられる．図 4 に「モデル間距離に基づく SOM」で作成したマップを示す．歌手間の距離が大きいほど，声質が異なる傾向が，比較的良好に表現されている．

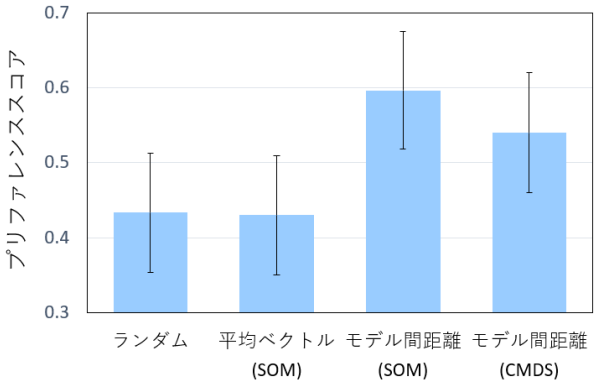

図 3　主観評価実験の結果

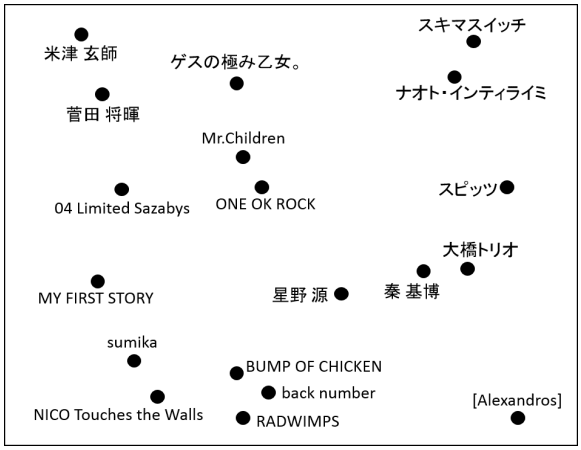

図 4　モデル間距離と SOM よる歌手マップ

4. まとめ

本研究では，楽曲検索のための声質の類似性に基づく「歌手マップ」の作成方法を提案し，その評価を行った．実験の結果，KL 情報量に基づくモデル間距離を利用して作成した歌手マップが優れた表現力を有していることが分かった．今後の課題として，ボーカル抽出の精度向上によるマップの高精度化や検索インタフェースとしての実装などが挙げられる．

参考文献

[1] 藤原他，情報処理学会研究報告，vol.2007, no.81, pp. 27-32, 2007.
[2] https://mahoroba.logical-arts.jp/archives/234
[3] T. Kohonen, Biol. Cybrtn., vol.43, iss.1, pp.59-69, 1982.
[4] 岩野他，信学技報，vol.117, no.189, pp.55-60, 2017.
[5] J. C. Gower, Biometrika, vol.53, pp. 325-328, 1966.