

楽曲間の類似判定における許容度の推定

川瀬 将太^{1,a)} 宮島 千代美¹ 北岡 教英¹ 武田 一哉¹

概要: 楽曲検索に関して、楽曲の音響情報を用いて楽曲間の主観的類似度を推定する手法について検討する。本研究では、楽曲間の主観的類似度は楽曲間の音響的類似度と聴取者の個人性により決定されると考える。本研究はこのうち聴取者の個人性に焦点を当て、聴取者間にどのような差異があるかを明らかにし、主観的類似度推定のモデルに組み込むことを目的としている。聴取者の個人性に関する先行研究の結果より、楽曲が音響的にどの程度似ていたら似ていると感じるかに大きな個人差が存在することが示唆された。本稿ではこの「音響的にどの程度似ていたら似ていると感じるか」を聴取者の「許容度」と呼び、許容度を含んだ主観的類似判定のモデルを提案する。実験では、楽曲間類似度の主観評価データを用いて聴取者の許容度を推定すると共に、実用の場面においてこの許容度を少数の類似性評価の結果を用いて推定することが可能であるかを確認する。

1. はじめに

記憶媒体の大容量化や高性能な圧縮技術の開発などを背景に、インターネットを通じた大規模楽曲データベースへのアクセスや、個人の大量の楽曲ファイルの利用が可能になった。反面、大量の楽曲に埋もれて所望の楽曲を発見することができないということが問題となっている。本研究ではこの問題を解決するための技術として楽曲間主観的類似度の推定手法について研究している。楽曲間の主観的類似度を推定し、類似した楽曲を検索することができれば、所望の楽曲を発見することが容易になるからである。

主観的類似度の推定は、人間の音楽知覚において重要な音響特徴量を抽出し、人間が行うのと同じやり方でそれらの特徴量を比較することができれば実現できる。しかし、人間の音楽の知覚については明らかでない部分が多く、主観的な類似度を音響的特徴に基づき正確に推定することは非常に困難である。また、どのような楽曲同士を似ていると感じるかは人により異なっていると考えられる。従って、本研究では楽曲間の主観的類似度が、楽曲間の音響的類似度及び、聴取者の個人性という2つの要因から成ると考える。楽曲間類似度に寄与する音響的特徴とそれを処理する聴取者の個人性を明らかにすることにより、楽曲間主観的類似度を客観的情報から推定する手法を明らかにすることを目指す。

楽曲間主観的類似度を音響的特徴から推定するための

研究は音楽情報処理の分野で広く行われてきた。例えば Pampalk[1] は楽曲からメル周波数ケプストラム係数 (Mel-Frequency Cepstrum Coefficient: MFCC) を短時間フレームごとに算出し、その分布を正規分布もしくは混合正規分布で表現し、その分布間距離を KL 情報量で算出することで音色の類似度とした。音響特徴量として代表的なものには、音色を表すものとして上で述べた MFCC の他にスペクトル重心やゼロ交差数などスペクトル形状を表す特徴量が、和音を表すものにクロマベクトル、リズムを表すものに Fluctuation Pattern[2] や Rhythm Histogram[3] などが挙げられる。

楽曲検索における個人性に関して、聴取者の嗜好や類似評価の個人性を客観的な尺度に反映させる方法についての研究が行われている。Hoashi ら [4] は、ユーザに好みのジャンルや楽曲を選んでもらい、その情報からユーザの嗜好を表すベクトルを作成することでユーザの好みに合わせ楽曲を推薦する手法を提案した。Vignoli と Pauws[5] は、楽曲の類似度を音色、ジャンル、テンポ、発売年、雰囲気 の5つの特徴量の重み付け和で表現し、この重みパラメータをユーザが手動で設定することで個人の嗜好をシステムに反映することを試みた。Lampropoulos ら [6] は、音響特徴量を入力としたニューラルネットワークにより類似楽曲を検索するとともに、出力された楽曲群に対してユーザが順位と類似度を与えて再度学習させることでニューラルネットワークを最適化するシステムを提案している。またこの研究では、音楽知覚に影響を与える特徴量は個人により異なるという仮説を立て、特徴量セットからいくつかのサブセットを構築し用いることで最適な特徴量セットをユーザ

¹ 名古屋大学
Nagoya University, Nagoya, Aichi, 464-8603, Japan
^{a)} shota.kawabuchi@g.sp.m.is.nagoya-u.ac.jp

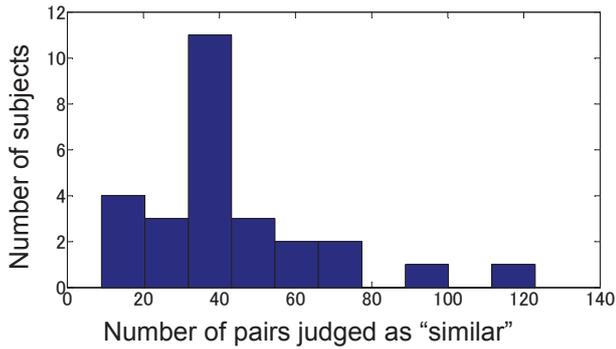


図 1 被験者が似ていると評価した回数のヒストグラム。各被験者が 200 回の評価試行のうち何回「似ている」と判定したかを示している (平均 43.4 回)。平均付近の被験者が最も多いが、似ていると判定した回数が平均から大きく離れている被験者もいる (最低: 9 回, 最大: 123 回)。

Fig. 1 Histogram of number of pairs judged as "similar". This indicates how many pairs were judged as similar in 200 trials (average 43.4 times). Most of the subjects are concentrated around the average, however, outlier subjects also exist (least 9 times, most 123 times).

が選べるような仕組みにしている。Kawabuchi ら [7] は、楽曲ペアの類似性を 2 段階で評価する実験を行い、収集されたデータを用いて楽曲間の距離尺度を個人毎に学習することで被験者の個性を表現した。

文献 [7] では、収集されたデータを分析した結果、楽曲ペアを似ていると評価する頻度が被験者毎に大きく異なることが示されていた (図 1)。このことから、楽曲が音響的にどの程度似ていたら似ていると感じるかには大きな個人差が存在すると考えられる。しかし、文献内で用いられていた楽曲間類似度のモデルにはこの情報が用いられていなかった。本稿ではこの「音響的にどの程度似ていたら似ていると感じるか」を聴取者の「許容度」と呼び、許容度を含んだ類似判定のモデルを提案する。

本稿の構成は以下の通りである。2 章では、許容度を含んだ楽曲類似判定のモデルを提案する。3 章では、楽曲間類似性の主観評価データに対し類似判定モデルを適用することで、被験者の許容度を推定する。4 章では、本稿のまとめと今後の課題について述べる。

2. 類似判定モデル

類似判定が楽曲ペアの類似度と聴取者の許容度より決定されると仮定する。類似度が高い楽曲ペアほど似ていると評価される可能性が高く、許容度が大きい聴取者ほど似ていると評価する可能性が高い。また、類似度が判定境界に近い楽曲ペアについては類似評価に揺らぎがある (「似ている」と判定されたり、「似ていない」と判定されたりする) と考えられる。以上のことを考慮に入れ、ある聴取者 i が楽曲ペア j を「似ている」と評価する確率を、ロジス

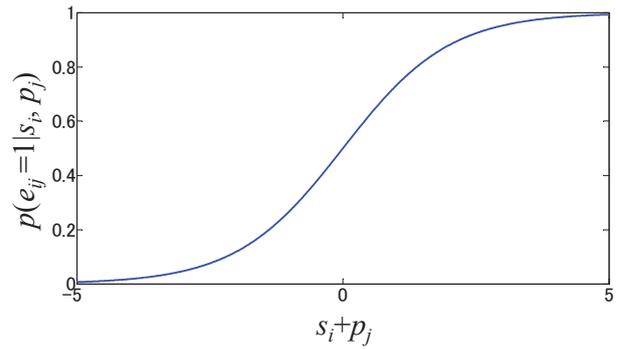


図 2 $s_i + p_j$ の値を変化させたときの $p(e_{ij}|s_i, p_j)$ の値の変化。

Fig. 2 Change of the value of $p(e_{ij}|s_i, p_j)$ as changing the value of $s_i + p_j$.

ティック関数を用いて以下のようにモデル化する。

$$p(e_{ij} = 1 | s_i, p_j) = \frac{1}{1 + e^{-(s_i + p_j)}} \quad (1)$$

ここで、 e_{ij} は聴取者 i の楽曲ペア j に対する類似評価結果であり、 $e_{ij} = 1$ ならば聴取者 i が楽曲ペア j を「似ている」と評価したことを、 $e_{ij} = 0$ ならば聴取者 i が楽曲ペア j を「似ていない」と評価したことを表す。 s_i, p_j はそれぞれ聴取者 i の許容度と楽曲ペア j の類似度に相当するパラメータである。 s_i が大きいほど被験者 i が「似ている」と評価する可能性が高く、 p_j が大きいほど楽曲ペア j が似ていると評価される可能性が高い (図 2)。

2.1 パラメータの推定

M 人の聴取者が N 対の楽曲ペアに対して類似性評価を行ったデータがあるとする。このデータに対し前節で述べた類似判定モデルを適用すると、モデルの尤度は以下のように表せる。

$$\begin{aligned} p(\mathbf{E} | \mathbf{s}, \mathbf{p}) &= \prod_{i=1}^M \prod_{j=1}^N p(e_{ij} = 1 | s_i, p_j)^{e_{ij}} p(e_{ij} = 0 | s_i, p_j)^{1 - e_{ij}} \\ &= \prod_{i=1}^M \prod_{j=1}^N p(e_{ij} = 1 | s_i, p_j)^{e_{ij}} \{1 - p(e_{ij} = 1 | s_i, p_j)\}^{1 - e_{ij}} \end{aligned} \quad (2)$$

ただし、 $\mathbf{E} = (e_{11}, \dots, e_{ij}, \dots, e_{MN})$, $\mathbf{s} = (s_1, \dots, s_M)$, $\mathbf{p} = (p_1, \dots, p_N)$ である。この尤度ができるだけ高くなるようパラメータ \mathbf{s}, \mathbf{p} を求めたい。

尤度が高くなるパラメータ \mathbf{s}, \mathbf{p} を求めるために、モデルの対数尤度

$$\begin{aligned}
 & \log p(\mathbf{E}|\mathbf{s}, \mathbf{p}) \\
 &= \sum_{i=1}^M \sum_{j=1}^N [e_{ij} \log p(e_{ij} = 1|s_i, p_j) + \\
 & \quad (1 - e_{ij}) \log p(e_{ij} = 0|s_i, p_j)] \\
 &= - \sum_{i=1}^M \sum_{j=1}^N [(1 - e_{ij})(s_i + p_j) + \log \{1 + e^{-(s_i + p_j)}\}]
 \end{aligned} \tag{3}$$

を $s_i (i = 1, \dots, M), p_j (j = 1, \dots, N)$ でそれぞれ偏微分する。

$$\frac{\partial}{\partial s_i} \log p(\mathbf{E}|\mathbf{s}, \mathbf{p}) = \sum_{j=1}^N \left[\frac{e^{-(s_m + p_j)}}{1 + e^{-(s_m + p_j)}} - (1 - e_{ij}) \right] \tag{4}$$

$$\frac{\partial}{\partial p_j} \log p(\mathbf{E}|\mathbf{s}, \mathbf{p}) = \sum_{i=1}^M \left[\frac{e^{-(s_i + p_n)}}{1 + e^{-(s_i + p_n)}} - (1 - e_{ij}) \right] \tag{5}$$

式 (4), (5) を $(i = 1, \dots, M), (j = 1, \dots, N)$ についてそれぞれ 0 とおく。尤度が極大値をとるパラメータ \mathbf{s}, \mathbf{p} が、式 (4), (5) を 0 とおいた連立方程式の解として得られる。

2.2 パラメータの事前分布の導入

式 (4), (5) は、似ていないと評価される回数の期待値から、実際に似ていないと評価された回数を引いた形になっている。従って、たとえば、ある楽曲ペア j が全ての聴取者に「似ていない」と評価されたとすると、式 (5) が 0 になる時、 $p_j = -\infty$ である。このような極端なモデルパラメータが推定されるのを防ぐために、パラメータ \mathbf{s}, \mathbf{p} が何らかの事前分布 $p(\mathbf{s}), p(\mathbf{p})$ に従うと仮定し、データ \mathbf{E} が与えられた下でのパラメータ \mathbf{s}, \mathbf{p} の事後確率 $p(\mathbf{s}, \mathbf{p}|\mathbf{E})$ を最大化するパラメータを求めることを考える。本稿では、許容度 $s_i (i = 1, \dots, M)$ と類似度 $p_j (j = 1, \dots, N)$ の事前分布がそれぞれ互いに独立な正規分布 $\mathcal{N}(s_i|\mu_s, \sigma_s^2), \mathcal{N}(p_j|\mu_p, \sigma_p^2)$ に従うと仮定する。従って、事後確率 $p(\mathbf{s}, \mathbf{p}|\mathbf{E})$ を最大化するためには

$$\begin{aligned}
 & p(\mathbf{s})p(\mathbf{p})p(\mathbf{E}|\mathbf{s}, \mathbf{p}) \\
 &= \prod_{i=1}^M \mathcal{N}(s_i|\mu_s, \sigma_s^2) \prod_{j=1}^N \mathcal{N}(p_j|\mu_p, \sigma_p^2) \\
 & \quad \prod_{i=1}^M \prod_{j=1}^N p(e_{ij} = 1|s_i, p_j)^{e_{ij}} p(e_{ij} = 0|s_i, p_j)^{1-e_{ij}}
 \end{aligned} \tag{6}$$

を最大化すればよい。式 (6) を最大化するパラメータ \mathbf{s}, \mathbf{p} を求めるために、目的関数

$$\begin{aligned}
 L &= \log \{p(\mathbf{s})p(\mathbf{p})p(\mathbf{E}|\mathbf{s}, \mathbf{p})\} \\
 &= - \sum_{i=1}^M \left\{ \frac{(s_i - \mu_s)^2}{2\sigma_s^2} + \log \sqrt{2\pi}\sigma_s \right\} \\
 & \quad - \sum_{j=1}^N \left\{ \frac{(p_j - \mu_p)^2}{2\sigma_p^2} + \log \sqrt{2\pi}\sigma_p \right\} \\
 & \quad - \sum_{i=1}^M \sum_{j=1}^N [(1 - e_{ij})(s_i + p_j) + \log \{1 + e^{-(s_i + p_j)}\}]
 \end{aligned} \tag{7}$$

とおく。 L を $s_i (i = 1, \dots, M), p_j (j = 1, \dots, N)$ でそれぞれ偏微分すると

$$\frac{\partial}{\partial s_i} L = \sum_{j=1}^N \left[\frac{e^{-(s_m + p_j)}}{1 + e^{-(s_m + p_j)}} - (1 - e_{ij}) \right] - \frac{1}{\sigma_s^2} (s_i - \mu_s) \tag{8}$$

$$\frac{\partial}{\partial p_j} L = \sum_{i=1}^M \left[\frac{e^{-(s_i + p_n)}}{1 + e^{-(s_i + p_n)}} - (1 - e_{ij}) \right] - \frac{1}{\sigma_p^2} (p_j - \mu_p) \tag{9}$$

最適なパラメータ \mathbf{s}, \mathbf{p} は、式 (8), (9) を 0 とおいた連立方程式の解として得られる。

2.3 パラメータ推定アルゴリズム

目的関数 L が極大値をとるとき $\mathbf{s} = (s_1, \dots, s_M), \mathbf{p} = (p_1, \dots, p_N)$ を求めるため、以下の式によりパラメータを反復的に更新する。

$$s_i \leftarrow s_i + \eta \frac{\partial}{\partial s_i} L \quad (i = 1, 2, \dots, M) \tag{10}$$

$$p_j \leftarrow p_j + \eta \frac{\partial}{\partial p_j} L \quad (j = 1, 2, \dots, N) \tag{11}$$

ただし、 η は学習係数である。この手続を目的関数 L が収束するまで繰り返すことで \mathbf{s}, \mathbf{p} の推定値が得られる。

3. 主観評価データに対する類似判定モデルの適用

楽曲間類似性の主観評価データに対し類似判定モデルを適用することで聴取者の許容度、及び、楽曲ペアの類似度を推定する実験を行う。また、実用の場面においては少数の類似性評価から聴取者の許容度を推定できる必要があるため、少数の類似性評価から許容度を推定できるかを調べる実験も行う。

3.1 実験に使用するデータ

ポピュラー楽曲 80 曲から選出した 200 ペアに対して、被験者がペアの類似性を 2 段階で評価したデータ [7] を実

験に使用する。実験の手続きは以下の通りである。被験者はまず 80 曲の中から選択された 2 つの楽曲 (以降、楽曲ペア) を聴き、その類似度を「似ている」「似ていない」かの 2 段階で評価した。その後、その楽曲について楽曲の構成する要素 (メロディ、テンポ・リズム、声質、楽器構成) の中から似ていると考えたものを複数回答可で選択した。各被験者は一人当たり 200 ペアを評価した。各々の被験者が評価した 200 ペアは全て同じ楽曲ペアであった。実験に参加した被験者は男性 13 名、女性 14 名の 27 名であった。

本稿では、このデータのうち類似度の判定のデータのみを用いる (構成要素についての類似評価データは用いていない)。

3.2 類似判定モデルの適用

文献 [7] で収集された 27 名 × 200 ペアの主観評価データを \mathbf{E} として、各被験者の許容度と楽曲ペアの類似度を求める実験を行った。反復アルゴリズムのための $s_i (i = 1, \dots, M), p_j (j = 1, \dots, N)$ の初期値としては、期待値 0、分散 1 の正規乱数を用いた。 \mathbf{s}, \mathbf{p} の事前分布 $p(\mathbf{s}), p(\mathbf{p})$ のパラメータは $\mu_s = 0, \sigma_s^2 = 1, \mu_p = 0, \sigma_p^2 = 1$ とした。更新学習は、目的関数 L が十分に収束するまで反復した。

推定された許容度と類似度を図 3 に示す。図中の直線は類似判定モデルにおける判別境界 (「似ている」と評価する確率が 50% となる面) の位置を示している。全点数に対する、正しく識別された点の比率 ($s + p > 0$ となる「似ている」点の数と $s + p \leq 0$ となる「似ていない」点の数の比率) は 0.842 であった。図より、「似ていない」点の識別が高い精度 (識別率 95.8%) でできているのに対し、「似ている」点の識別はあまりできていない (識別率 42.5%)。これは、「似ていない」点の数が「似ている」点の数よりも多いことによると考えられる (5400 (= 27 × 200) の評価のうち、「似ている」評価数が 1171, 「似ていない」評価が 4229 数であった)。

3.3 少数データからの許容度の推定

類似判定モデルを用いて類似判定を予測するためには、楽曲間の類似度と聴取者の許容度が分かっている必要がある。実用の場面において、楽曲間の類似度は楽曲の特徴を用いることで推定することができると考えられるが、聴取者の許容度については聴取者の評価を用いなければ推定することができない。しかし、実用において聴取者に大量の楽曲ペアについて類似度を評価をさせるというのは現実的ではない。従って、より少数のペアについての評価データから正確な許容度の推定ができることが望ましい。前節では、主観評価が与えられた 27 名 × 200 ペア分の評価データに対し類似判定モデルを適用することで許容度と類似度の推定を行った。以下では、200 ペアから少数のペアを選択し、選択した少数のペアに対する類似性評価データから

聴取者の許容度を推定する実験を行う。

許容度を推定したい被験者 i の許容度を s_i 、200 ペアから選択された少数ペアの類似度を $\mathbf{p} = (p_1, \dots, p_n)$ とする。データ \mathbf{E} が与えられた下での事後確率 $p(s_i, \mathbf{p} | \mathbf{E})$ を最大にする s_i と \mathbf{p} は

$$\begin{aligned} & \log \{p(s_i)p(\mathbf{p})p(\mathbf{E}|s_i, \mathbf{p})\} \\ &= - \left\{ \frac{(s_i - \mu_s)^2}{2\sigma_s^2} + \log \sqrt{2\pi}\sigma_s \right\} \\ & \quad - \sum_{j=1}^n \left\{ \frac{(p_j - \mu_p)^2}{2\sigma_p^2} + \log \sqrt{2\pi}\sigma_p \right\} \\ & \quad - \sum_{j=1}^n [(1 - e_{ij})(s_i + p_j) + \log \{1 + e^{-(s_i + p_j)}\}] \end{aligned} \quad (12)$$

を最大にする s_i と \mathbf{p} を求めることで得られる。目的関数 $L = \log \{p(s_i)p(\mathbf{p})p(\mathbf{E}|s_i, \mathbf{p})\}$ とおき、 L を s_i について偏微分すると

$$\begin{aligned} \frac{\partial}{\partial s_i} L &= \sum_{j=1}^n \left[\frac{e^{-(s_i + p_j)}}{1 + e^{-(s_i + p_j)}} - (1 - e_{ij}) \right] \\ & \quad - \frac{1}{\sigma_s^2} (s_i - \mu_s) \end{aligned} \quad (13)$$

であるから、

$$s_i \leftarrow s_i + \eta \frac{\partial}{\partial s_i} L \quad (14)$$

のように反復アルゴリズムを用いて聴取者 i の許容度 s_i を推定できる。

3.4 許容度の推定実験

主観評価データから数ペアを選択し、選択したペアの評価データのみを用いて許容度を推定する実験を行った。実験の手続きは以下のとおりである。まず、200 ペアから推定に用いるペアをランダムに選択した。選択したペアを用いて式 (14) により反復的に被験者の許容度を推定した。許容度を推定した後、選択されなかったペアを用いて識別率および推定精度 (3.2 節で推定された許容度との差の絶対値) を算出した。推定に使用するペア数として 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 を試した。ランダムに選択されたペアによる結果の偏りを避けるため、全被験者・全条件に対し許容度の推定を同じ手順で 100 回行い、100 試行 × 全被験者の平均により識別率と推定精度を評価した。

許容度の推定に必要な楽曲間類似度 \mathbf{p} としては、

- 3.2 節で推定された楽曲間類似度
 - 音響特徴ベクトル間のユークリッド距離 d_j を入力にとり p_j の値を予測する線形回帰モデルの出力 \hat{p}_j (図 4)
- の 2 種類を試した。音響特徴ベクトルの算出法は文献 [7]

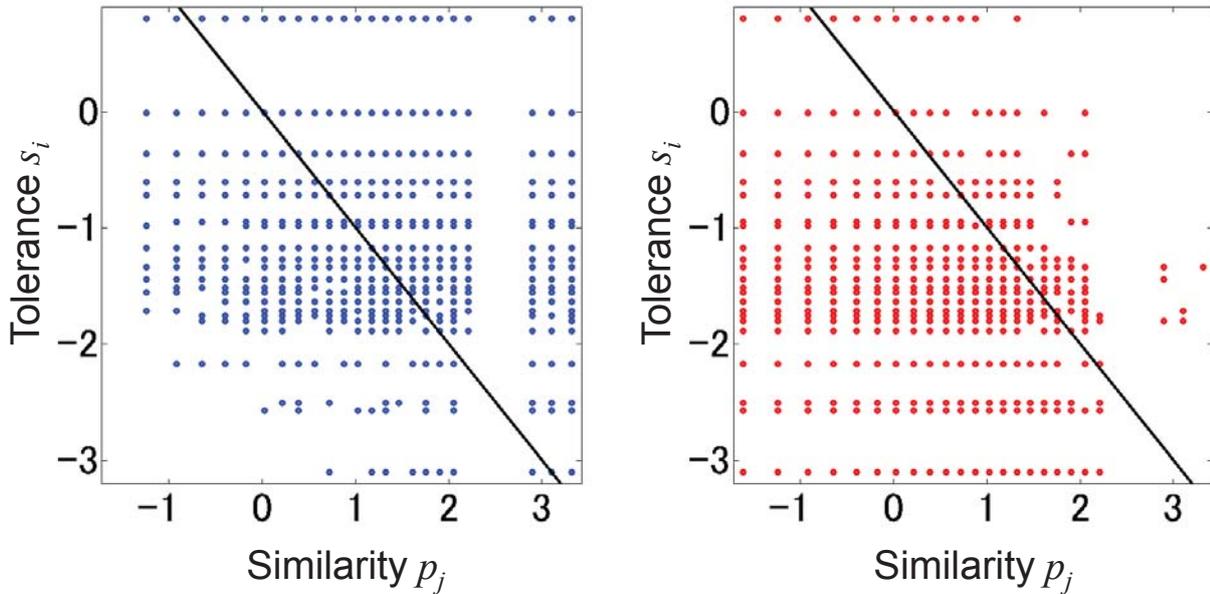


図 3 許容度 × 類似度の平面上における「似ている」点 (左図) と「似ていない」点 (右図) の散布図。縦軸は被験者の許容度, 横軸は楽曲ペアの類似度である。図中の線は $s_i + p_j = 0$ となる識別面であり, この線よりも右に「似ている」点, 左に「似ていない」点があることが望ましい。「似ている」点の識別率は 42.5% 「似ていない」点の識別率は 95.8% であった。

Fig. 3 Scatter plots of "similar" points (left figure) and "dissimilar" points (right figure) on the tolerance-similarity plane. Vertical axis is subject's tolerance, and horizontal axis is pair's similarity. Lines on the figure means discrimination surface $s_i + p_j = 0$, and it is desirable that "similar" points are on the right side of this surface and "dissimilar" points are on the left side. Discrimination rate of "similar" points was 42.5%, and of "dissimilar" points was 95.8%.

を参考にした。具体的には, メル周波数ケプストラム係数 (1-13 次), インテンシティ [8], スペクトルセントロイド, スペクトルフラックス, スペクトルロールオフ, 高周波数エネルギー [9] を短時間特徴量として抽出し, ベクトル量子化することで楽曲毎にクラスタのヒストグラムを得る。得られたヒストグラムの各ビンの対数をとったものを楽曲を表す音響特徴ベクトルとして用いた。

実験の結果を結果を図 5, 6 に示す。図 5 は, 許容度の推定に用いるペア数を変化させた時の平均識別率の推移を表している。青の実線が 200 ペアを用いて推定された楽曲間類似度を用いた場合, 赤の点線が音響特徴により推定した類似度を用いた場合である。図中の黒の水平線は 3.2 節での実験の結果得られた識別率であり, 本手法における識別率の上限を示すと考えられる。図より, 音響特徴より推定した類似度を用いた場合に, 識別性能が若干低下すること, どちらの類似度においても最大限に高い識別率を得るために必要なペア数は 20~30 ペア程度であるということが分かる。図 6 は, 許容度の推定に用いるペア数を変化させた時の推定精度の推移を表している。図 5 と同様, 青の実線が 200 ペアを用いて推定された楽曲間類似度を用いた場合, 赤の点線が音響特徴により推定した類似度を用いた場合である。図より, 音響特徴を用いた場合は若干精度が下

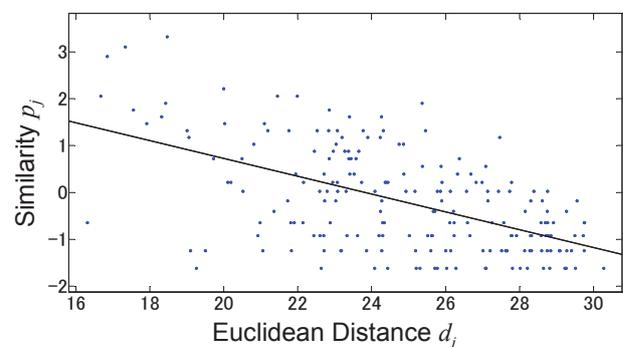


図 4 特徴ベクトル間のユークリッド距離 d_j と類似度 p_j との関係。 d_j と p_j の線形相関は $\rho = -0.553$ 。図中の黒線は回帰直線 $\hat{p}_j = -0.190d_j + 4.543$ 。

Fig. 4 Relation of Euclidean distance between feature vectors d_j and similarity p_j . Linear correlation between d_j and p_j is $\rho = -0.553$. Black line on the figure is regression line $\hat{p}_j = -0.190d_j + 4.543$.

がることや, 推定に用いるペア数を増やしていくと 20~30 ペア程度までは推定精度が急激に向上するが, それ以降は緩やかであることなどが分かる。

4. まとめ

本稿では, 楽曲ペアの類似判定における「許容度」を含

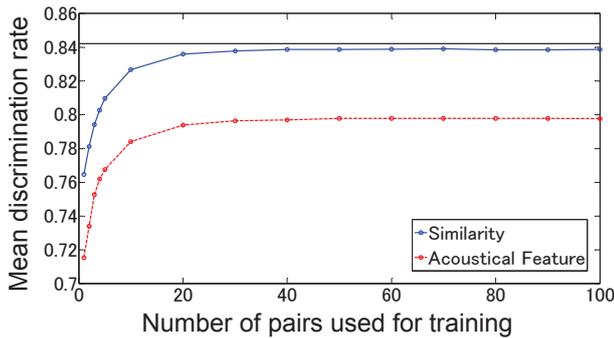


図 5 許容度の推定に用いたペア数を変化させたときの平均識別率の推移。横軸は推定に用いたペア数、縦軸は平均識別率である。青の実線 (Similarity) が 3.2 節で推定された楽曲間類似度を用いた場合、赤の点線 (Acoustical Feature) が音響特徴により推定した類似度を用いた場合の結果である。図中の水平線は 3.2 節での実験において得られた識別率を示し、この手法における識別率の上限と考えることができる。

Fig. 5 Plot of the mean discrimination rate according to the number of pairs used for training. Horizontal axis is the number of pairs used for training, and vertical axis is the mean discrimination rate. Blue solid line (Similarity) indicates the result obtained by using similarity estimated by using 200 pairs. Red dotted line (Acoustical Feature) indicates the result obtained by using similarity estimated from acoustical features. A horizontal line in this figure indicates the discrimination ratio obtained in the experiment in section 3.2. It can be interpreted as upper bound of performance of this method.

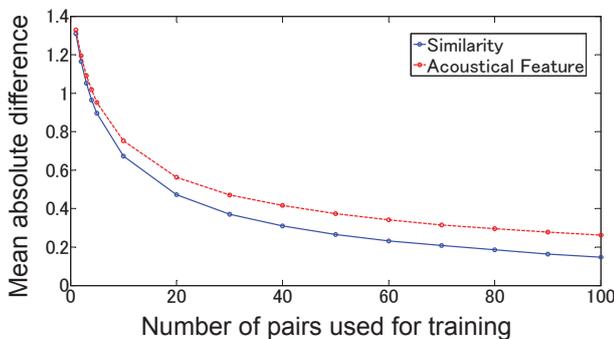


図 6 許容度の推定に用いたペア数を変化させたときの推定精度の推移。横軸は推定に用いたペア数、縦軸は推定した許容度と 3.2 節で推定された許容度との差の絶対値の全被験者・全試行にわたる平均である。

Fig. 6 Plot of the mean absolute difference with the various number of pairs used for training. Horizontal axis is the number of pairs used for training. Vertical axis is mean absolute difference between estimated tolerance and tolerance estimated in section 3.2 over all subjects and all trials.

む類似判定モデルを提案した。類似判定モデルは楽曲ペアの類似判定が聴取者の許容度と楽曲ペアの類似度により決定されるというモデルである。このモデルの定式化とパラメータを推定アルゴリズムについて説明した。また、実際

の楽曲間類似評価のデータを用いてモデルのパラメータを推定する実験を行った。実験の結果、「似ている」と評価される場合と「似ていない」と評価される場合を 84.2% の精度で識別可能なモデルが学習された。また、実用において少数の評価データを用いて許容度を推定することができるかどうかを調べる実験を行った。実験の結果より、最大限に高い識別率を得るためには、今回用いたデータにおいては、概ね 20~30 ペアに対する評価データがあれば良いことが分かった。しかし、どの程度高い識別率・推定精度が得られれば十分であるかについては検討を行っていない。これについては今後の課題である。また、類似判定モデルの識別性能は、音響特徴を用いた楽曲間類似度の推定手法に大きく依存すると考えられる。本稿では音響特徴ベクトル間のユークリッド距離を入力にとる単純な線形回帰モデルにより楽曲間類似度を推定したが、これを改善することで類似判定モデルの性能をより高めることができると考えられる。

謝辞 本研究の一部は科学技術振興機構 CREST, 及び科研費 No.23650088 によるものである。

参考文献

- [1] Pampalk, E.: Computational models of music similarity and their application in music information retrieval, PhD Thesis, Vienna University of Technology (2006).
- [2] Pampalk, E., Rauber, A. and Merkle, D.: Content-based organization and visualization of music archives, *ACM Multimedia*, pp. 1-6 (2002).
- [3] Lidy, T. and Rauber, A.: Evaluation of feature extractions and psycho-acoustics transformations for music genre classification, *the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pp. 34-41 (2005).
- [4] Hoashi, K., Matsumoto, K. and Inoue, N.: Personalization of user profiles for content-based music retrieval on relevance feedback, *ACM Multimedia*, pp. 110-119 (2003).
- [5] Vignoli, F. and Pauws, S.: A music retrieval system based on user-driven similarity and its evaluation, *the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pp. 272-279 (2005).
- [6] Lampropoulos, A., Sotiropoulos, D. and Tsihrantzis, G.: Individualization of music similarity perception via feature subset selection, *IEEE International Conference on Systems, Man and Cybernetics*, Vol. 1, pp. 552-556 (2004).
- [7] Kawabuchi, S., Miyajima, C., Kitaoka, N. and Takeda, K.: Subjective similarity of music: Data collection for individuality analysis, *Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1-5 (2012).
- [8] Lu, L., Liu, D. and Zhang, H. J.: Automatic mood detection and tracking of music audio signals, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 1, pp. 5-18 (2006).
- [9] Juslin, P. N.: Cue utilization in communication of emotion in music performance: relating performance to perception, *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 26, No. 6, pp. 1707-1813 (2000).