

## 主観評価に基づく楽曲間類似度算出モデル

平賀 悠介<sup>†1</sup> 大石 康智<sup>†2</sup> 武田 一哉<sup>†1</sup>

本研究では、複数の音響的特徴の類似度から楽曲の主観的な類似度を表現する手法を提案する。我々は楽曲間類似度の“ground-truth”を構築するため、現在 RWC 研究用音楽データベースの楽曲 202 トラックの全組み合わせに対し、楽曲の類似性を主観評価したデータを大量に収集している。収集する主観評価データの性質を予め調査するため、16 名の被験者に対し予備的にデータを収集し、その性質について調査した。またそのデータから楽曲の組に対し主観的な類似度を表すスコアを算出し、音響的特徴の類似度と重回帰分析により関連付けた。その結果、音響的特徴の類似度の重み付け和から主観評価に基づく楽曲間類似度を推定できることを示した。

### Method for Calculating the Subjective-based Music Similarity Measure

YUSUKE HIRAGA,<sup>†1</sup> YASUNORI OHISHI<sup>†2</sup>  
and KAZUYA TAKEDA<sup>†1</sup>

In this paper, we propose a method for representing a subjective music similarity by several acoustic similarities. For establishing “ground-truth” similarity, we are collecting subjective similarity data for pairs of 202 tracks of RWC music database. A preliminary experiment is conducted for subjective-similarity data collected from 16 subjects. The subjective similarity is associated with the acoustic similarities using linear regression model. The result indicates that weighted sum of several acoustic similarity measures well predict the subjective similarity.

<sup>†1</sup> 名古屋大学大学院情報科学研究科

Graduate School of Information Science, Nagoya University

<sup>†2</sup> NTT コミュニケーション科学基礎研究所

NTT Communication Science Laboratories, NTT Corporation

### 1. はじめに

近年、インターネットを通じた大規模楽曲データベースのアクセスや、大容量メディアによる個人の大量の楽曲ファイルの所有が可能になった。そのような大量の楽曲から即座に好みの曲を検索したり<sup>1)2)</sup>、ユーザの嗜好に応じた曲を推薦するための研究やアプリケーション開発が進められている<sup>3)4)</sup>。そのようなシステムを研究する上で、楽曲の類似度の定義の仕方についても様々な議論がこれまでになされている。音響的特徴の類似度を楽曲の類似度として用いる手法は多くの音楽情報検索システムで用いられており、代表的な特徴量としてメル周波数ケプストラム係数 (Mel-Frequency Cepstrum Coefficient: MFCC)、スペクトル形状に基づく特徴量などが挙げられる。また、人間の音楽知覚に基づく特徴量を用いる手法もいくつか提案されている<sup>5)6)7)</sup>。

一方、これらの客観的な評価尺度と楽曲の類似性に対する主観評価や個人の嗜好を関連付けることがこれまで試みられている。Hoashi ら<sup>3)</sup> 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ている。Vignoli と Pauws<sup>4)</sup> 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<sup>8)</sup> 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。また彼らは、個人の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が選べるような仕組みにしている。

また、楽曲間類似度の“ground-truth”を構築する試みもなされている。国際音楽情報検索会議 ISMIR の音楽情報検索システムのコンテスト (MIREX) のホストである IMIRSEL は、MIREX のタスク “Audio Music Similarity and Retrieval” (類似楽曲検索)、“Symbolic Melodic Similarity” (シンボリック形式からのメロディ類似度判定)、“Audio Mood Classification” (ムード分類)において、“Evalutron 6000”<sup>9)</sup>を用い楽曲検索システムの主観的な性能評価を行うことで楽曲間類似度の主観評価データを収集している<sup>10)</sup>。この実験は、システムが “query” の楽曲 1 曲とそれに対する “candidates” の楽曲数曲を提示し、被験者がそれぞれの “candidate” が “query” に似ているかどうかを 3 段階評価 (“NOT Similar”, “Somewhat Similar”, “VERY Similar”) と 100 段階評価 (0 ~ 10 点, 0.1 点刻み) で評価



図 1 主観評価データ入力用インタフェース

を与えるというものである。また楽曲データは IMIRSEL が管理しているもので、“Audio Music Similarity and Retrieval” の場合 7000 曲 10 ジャンルの楽曲データセットを利用している。

本研究では、音楽情報検索研究用の楽曲データベースである RWC 研究用音楽データベース<sup>11)12)</sup> の楽曲 202 トラックの全ての組み合わせに対する、楽曲類似性の主観評価データを収集し、楽曲間類似度の“ground-truth”を構築することを一つの目的とする。さらに収集した主観評価データに基づき、楽曲の類似性に対する人の主観的な判断を推定できるような楽曲類似尺度を構築することを目的とする。

本稿ではまず、楽曲間類似度主観評価実験の概要と先に行った予備実験に対する評価を行う。次に、楽曲の主観的な類似度は音響的な類似度の重み付け和により表されると仮定し、その重み係数を収集した主観評価データに基づき線形回帰モデルにより推定する実験を行う。

## 2. 主観評価データ収集実験

我々は楽曲間類似度の“ground-truth”を与えるデータベースを構築するための主観評価実験を行っている。web ブラウザ上で動作する主観評価データ収集システム(図 1)を構築し、web を通じて実験を行っている。楽曲を再生し、主観評価データを入力するためのインタフェースは Flash 形式のアプリケーションで、FlashPlayer がインストールされている PC であれば使用環境を問わず実験を行うことができる。

### 2.1 使用楽曲

被験者に提示する楽曲は、RWC 研究用音楽データベースの「ポピュラー音楽」<sup>11)</sup>の楽曲 100 トラックと「音楽ジャンル」<sup>12)</sup>の楽曲 102 トラックの計 202 トラックを用い、同一曲同士を除く全ての楽曲ペアについて実験を行う。楽曲の再生区間は先行研究<sup>10)</sup>と同様に、トラックの中間の前後 15 秒ずつを切り出した区間を用いる。

### 2.2 実験手順

ログインした被験者に対し、システムはランダムに「参照曲」と「評価曲」の 2 曲を提示する(以後、参照曲と評価曲のペアを「楽曲ペア」と呼ぶ)。被験者は各楽曲ペアについて評価曲が参照曲に似ているか似ていないかと、どのような観点で評価したか(以後「評価基準」と呼ぶ)をシステムに入力する。ただし被験者には、楽曲が似ている、似ていないとはどういうことを示すような基準は提示していない。一方で評価基準は複数回答を可能として、該当する項目を全て選択するように指示している。例えば被験者は提示された楽曲ペアのメロディと雰囲気が異なるため似ていないと感じたのであれば、「似ていない」を選択し、「メロディ」と「雰囲気」にチェックを入れる。評価基準として提示した項目を表 1 に示す。なお、楽曲は繰り返して再生することができるようにし、実験の休憩、再評価、再ログインは被験者の判断で自由に行えるようにした。評価データは評価された時間も含んでおり、再評価を与えた際は前回の評価データとは別々にデータは保管されている。また、

表 1 評価基準の項目

メロディ	歌やその曲で主となる楽器のメロディライン
テンポ	曲の速さ
雰囲気	曲の印象：繊細、力強い、など
ジャンル	音楽のジャンル
声質	ボーカルの声質
楽器構成	使用されている楽器
その他	上記以外の要素

表 2 収集実験の条件

楽曲データベース	RWC 研究用音楽データベース： ポピュラー音楽 100 トラック 音楽ジャンル 102 トラック
総楽曲ペア数	40602 ペア
一人当たり評価楽曲ペア数	200 曲/人
楽曲再生区間	中間の前後 15 秒ずつ
ファイル形式	MP3, 192kbps, CBR, joint-stereo

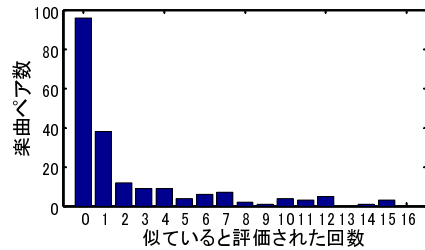


図 2 似ていると評価される回数の分布

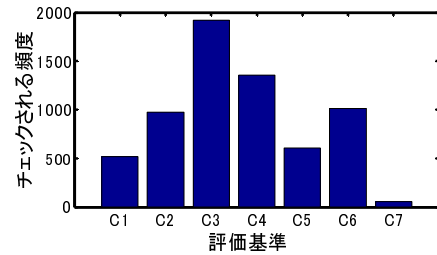


図 3 評価基準の出現頻度 (C1:メロディ, C2:テンポ, C3:雰囲気, C4:ジャンル, C5:声質, C6:楽器構成, C7:その他)

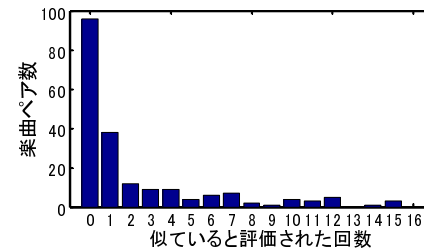


図 4 似ていると評価する回数の分布

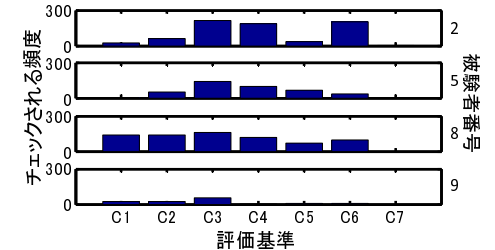


図 5 被験者別に見た評価基準の出現頻度 (C1:メロディ, C2:テンポ, C3:雰囲気, C4:ジャンル, C5:声質, C6:楽器構成, C7:その他)

上記のデータの他に被験者の年齢, 性別, 職業, 日常的な楽器演奏経験の有無についてアンケートを行っている. 被験者は年齢, 性別に偏りなく募集している. 現在 75 名が参加しており, 楽曲ペア 1567 ペアに対して評価を収集した.

### 2.3 予備実験

収集するデータベースの傾向を調査するため, 学生 16 名に対し予備実験を行った. 予備実験は, 被験者全員に同一の楽曲ペア (200 ペア) を提示したこと以外は実際の実験と同様の手順で行った.

図 2 は楽曲ペア 200 ペアが, 似ていると評価された回数の分布を表す. 図によると, 誰にも似ていると評価されなかったペアが最も多く, 似ていると評価された回数が増加するに従って指数関数的に楽曲ペア数は減少している. なお, 被験者の半数以上に似ていると評価されたペアは, 全体の 9.5% であった. 図 3 は, 各評価基準が選択された頻度を示す. 最も選ばれた回数が多かったのは「雰囲気」で, 次いで「ジャンル」「テンポ」であった. 逆に, 最も少なかったのは「その他」を除けば「メロディ」であった.

図 4 に被験者が似ていると評価した回数の分布と, 図 5 に各被験者の選んだ評価基準のヒストグラムの例をそれぞれ示す. 似ていると感じる頻度や基準は, 被験者により大きく異なっていることが確認できる.

## 3. 主観評価と音響的類似度の関連付け

本節では収集した楽曲類似性の主観評価データと楽曲の音響的類似度の関連付け手法について述べる.

### 3.1 主観評価スコア

主観評価スコア  $y$  は, 次のように定義する.

$$y = [y_1, y_2, \dots, y_N]^T \quad (1)$$

$$y_i = \frac{\text{楽曲ペア } i \text{ を似ていると評価した被験者数}}{\text{楽曲ペア } i \text{ を評価した被験者数}} \quad (2)$$

$y_i$  は楽曲ペア  $i$  の主観評価スコアである. つまり, 5 人中 3 人が似ていると判定した楽曲ペア  $i$  の主観評価スコア  $y_i$  は,  $y_i = 3/5 = 0.6$  である.

### 3.2 音響的類似度

本研究では楽曲の音響的特徴を表す音響ベクトルを算出し, その類似度を音響的類似度として用いる. 以下では音響特徴量の抽出と音響ベクトルの作成および類似度の算出について述べる.

#### 3.2.1 音響特徴ベクトルの算出

楽曲を短時間フーリエ変換 (STFT) して求めた短時間パワースペクトルから以下の音響特徴量を算出し, 音響特徴ベクトルを作成する. STFT の条件を表 3 に示す. 音響特徴量

サンプリング周波数	44.1kHz
フレーム長	50ms
フレームシフト	25ms
窓関数	ハニング窓

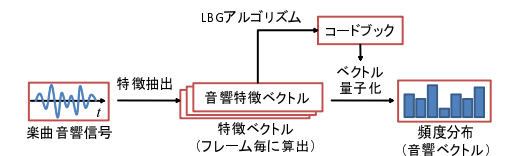


図 6 音響ベクトルの作成

は、音量に関する特徴量であるインテンシティ<sup>13)</sup>、音色を表す特徴量<sup>13)</sup>、MFCC、和音を表す特徴量であるクロマベクトルの4種類を用いる。インテンシティ、音色を表す特徴量、MFCCについては動的変動成分として1次の回帰係数

$$\Delta \mathbf{x}(t) = \frac{\sum_{k=-L}^L k \cdot \mathbf{x}(t-k)}{\sum_{k=-L}^L k^2} \quad (3)$$

を求め、特徴ベクトルの成分として加えた。本実験では  $L = 2$  とした。なお、インテンシティについてはSTFTを行う前に楽曲信号をサンプリング周波数16kHzにダウンサンプリングした。

#### インテンシティ

インテンシティは音量に関する特徴量で、フレーム全体の音量  $I(t)$  と、フレーム全体に対するサブバンドの音量の比  $D_j(t)$  により表される。

$$I(t) = \sum_{f=0}^{f_s/2} X(t, f) \quad (4)$$

$$D_j(t) = \frac{1}{I(t)} \sum_{f=L_j}^{H_j} X(t, f) \quad (5)$$

ここで、 $f_s$  はサンプリング周波数、 $X(t, f)$  はパワースペクトル、 $t$  はフレーム番号、 $f$  は周波数を表す。また  $j$  はサブバンド番号を表わし、サブバンドの下限の周波数を  $L_j$ 、上限の周波数を  $H_j$  とする。サブバンドへのパワースペクトルの分割はオクターブスケールフィルタバンク

$$\left[0, \frac{f_0}{2^n}\right), \left[\frac{f_0}{2^n}, \frac{f_0}{2^{n-1}}\right), \dots, \left[\frac{f_0}{2^2}, \frac{f_0}{2^1}\right] \quad (6)$$

を用いた。ここで  $n$  はサブバンドの数で、本実験では  $n = 7$  とした。したがって、インテンシティは動的変動成分も含め16次元のベクトルとなる。

#### 音色

音色を表わす特徴量はスペクトルの形状とコントラストを表す特徴量で構成される。スペクトル形状を表す特徴量は

##### (1) スペクトル重心

$$f_c(t) = \frac{\sum_{f=0}^{f_s/2} X(t, f) \cdot f}{\sum_{f=0}^{f_s/2} X(t, f)} \quad (7)$$

##### (2) バンド幅

$$B(t) = \sqrt{\frac{\sum_{f=0}^{f_s/2} (f - f_c)^2 X^2(t, f)}{\sum_{f=0}^{f_s/2} X^2(t, f)}} \quad (8)$$

##### (3) スペクトルロールオフ

スペクトルロールオフ  $f_r$  は、スペクトル分布の全帯域の95%を占める周波数で、

$$\sum_{f=0}^{f_r} X(t, f) = 0.95 \times \sum_{f=0}^{f_s/2} X(t, f) \quad (9)$$

として定義される。

##### (4) スペクトルフラックス

$$F = \sqrt{\sum_{f=0}^{f_s/2} (X(t, f) - X(t-1, f))^2} \quad (10)$$

の4種類を用いる。また、スペクトルのコントラストを表す特徴量は、 $j$  番目のサブバンドのパワースペクトル  $X_{j,1}, X_{j,2}, \dots, X_{j,N}$  を降順に並び変えた系列  $X'_{j,1}, X'_{j,2}, \dots, X'_{j,N}$  を用い、以下の3種類で表される。

##### (1) サブバンドピーク

$$P_j = \log \frac{1}{\alpha N} \sum_{k=1}^{\alpha N} X'_{j,k} \quad (11)$$

##### (2) サブバンドバレー

$$V_j = \log \frac{1}{\alpha N} \sum_{k=1}^{\alpha N} X'_{j,N-k+1} \quad (12)$$

##### (3) コントラスト

$$C_j = P_j - V_j \quad (13)$$

ここで、 $\alpha$  はピークとバレーの値を安定して抽出するためのパラメータで、 $\alpha = 0.2$  と

した。サブバンドの数はインテンシティの算出時と同様 7 としたので、コントラストを表す特徴量は  $3 \times 7 = 21$  次元となり、音色の特徴量の次元数は  $(4 + 21) \times 2 = 50$  次元である。

MFCC

MFCC は音声認識で広く用いられている特徴量で、類似楽曲検索では音色を表現する特徴量として用いられているが、本研究ではスペクトル形状に基づく音色の特徴量とは異なる特徴量として扱った。本実験では 16 次の MFCC を用い、動的成分を含めた 32 次元のベクトルとした。

クロマベクトル

和音を表す特徴量としてクロマベクトル (HPCP<sup>14</sup>) を用いた。クロマベクトルは 12 次元のベクトルで、各次元は A から G<sup>#</sup> までの各音階に対応するパワースペクトルの和で構成される。抽出には MIRtoolbox1.1<sup>15</sup> を用いた。

本研究では音響特徴ベクトルを各音響特徴について、フレームごとに作成する。従って 4 種類の音響特徴ベクトル系列が各楽曲から算出され、音響ベクトルも 4 種類算出される。本稿では、特徴量全てを 1 つの音響特徴ベクトルにまとめる従来法<sup>16</sup> との比較実験を行う。従来法では、音響ベクトルは 1 楽曲につきただ 1 つ算出される。なお、音響特徴ベクトルは、楽曲データセット全体から算出した平均と分散を用い正規化した。

3.2.2 音響ベクトルの作成と類似度算出

次に、3.2.1 節で求めた音響特徴ベクトルをベクトル量子化し、音響ベクトルを作成する。ベクトル量子化のためのコードブックは、全楽曲から求めた音響特徴ベクトルの集合から LBG アルゴリズムを用い作成した。コードブックサイズは 2048 とした。こうして求めた音響ベクトルのコサイン類似度を楽曲間の音響的類似度として用いる。

3.3 関連付け手法

3.1 節で求めた主観評価スコア  $y$  と 3.2 節で求めた音響的類似度を線形回帰モデルで関連付ける。音響特徴  $m$  についての楽曲ペア  $i$  の音響的類似度を  $d_{mi}$  とすると、主観評価に基づく新しい類似度  $\hat{y}$  は

$$\hat{y} = \Phi(\Phi^T \Phi)^{-1} \Phi^T y \tag{14}$$

$$\Phi = \begin{pmatrix} 1 & d_{11} & \dots & d_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & d_{M1} & \dots & d_{MN} \end{pmatrix} \tag{15}$$

で表される。 $M, N$  はそれぞれ音響特徴量数と総楽曲ペア数で、提案法では  $M = 4$ 、従来法では  $M = 1$  となる。

4. 関連付け性能の評価実験

音響特徴量の類似度と主観評価による楽曲の主観的類似度との関連付け性能を検証した。

4.1 実験条件

主観評価スコアの算出に用いたデータは、2.3 節で述べた予備実験により収集された 16 名による主観評価データで、楽曲ペア  $N = 200$  ペアに対する評価である。楽曲 200 ペアで出現した楽曲は 178 曲である。偏回帰係数の値を比較するため、音響的類似度は平均と分散を用いて正規化する。評価は leave-one-out 法で行った。

実験結果の評価には平均二乗誤差 (Root Mean Square Error: RMSE) を用いた。主観評価スコア  $y$  に対する算出した類似度  $\hat{y}$  の RMSE は

$$RMSE = \sqrt{\frac{1}{N} \sum_i (y_i - \hat{y}_i)^2} \tag{16}$$

で表される。従属変数である主観評価スコア  $y$  の標準偏差よりも RMSE の値が低ければ、提案したモデルは従属変数の分散を説明できている、すなわちモデルが有効であるといえる。

4.2 実験結果と考察

RMSE の値を表 4 に示す。主観評価スコアの標準偏差よりも平均二乗誤差が小さいことから、モデルの有効性が確認された。また、従来法に比べ提案法の RMSE は小さくなった。偏回帰係数の値を表 5 に示す。音色の特徴量に強く重み付けされており、楽曲の類似度に対する音色の特徴量、即ちスペクトル形状に基づく特徴量の重要性が確認された。

ところで線形回帰モデルは、各説明変数は互いに独立であることを仮定する。しかし、表 7 で示した各類似度の相関係数  $r$  を調査したところ、各説明変数には相関があることが確認され、特にインテンシティと音色で  $r = 0.7937$  と高い相関が認められた。また、本稿では主観的類似度を式 (2) として定義し、主観的類似度が音響的類似度の重み付け和で表されると仮定したが、主観的類似度の定義や、音響的類似度との関連付け手法の双方には検討の余地があると考えられる。

5. まとめと今後の展開

楽曲類似性の主観評価データ収集のためのシステムを構築し、収集実験を行った。事前に

予備実験を行い、収集するデータの性質を分析し、多様なデータが収集できることを確認した。また、予備実験で収集した主観評価データと楽曲の音響的類似度を線形回帰モデルにより関連付ける手法を検討した。類似度算出実験の結果、手法の有効性を確認した。

今後は主観評価データ収集実験を継続しデータベースを大規模化することを目指す。また、主観評価と音響的類似度の関連付けについてさらなる検討を行う。さらに楽曲検索システムへの適用と、主観評価実験による性能評価を行う。

### 参 考 文 献

- 1) Rauber, A., Pampalk, E., and Merkl, D., "Using Psycho-Acoustic Models and Self-Organizing Maps to Create a Hierarchical Structure of Music by Sound Similarity," *Proc. ISMIR*, pp.71-80 (2002).
- 2) Goto, M. and Goto, T., "Musicream: New Music Playback Interface for Streaming, Sticking, and Recalling Musical Pieces," *Proc. ISMIR* (2005).
- 3) Hoashi, K., Matsumoto, K. and Inoue, N., "Personalization of User Profiles for

表 4 主観評価スコアの標準偏差と RMSE 値

主観評価スコアの標準偏差	従来法の推定誤差	提案法の推定誤差
0.2287	0.1925	<b>0.1886</b>

表 5 偏回帰係数の値

バイアス項	インテンシティ	音色	MFCC	クロマベクトル
0.1838	-0.0158	0.1044	0.0264	0.0282

表 6 主観評価スコアの区間ごとに見た推定誤差

主観評価スコア $y_i$	$0 \leq y_i < \frac{1}{3}$	$\frac{1}{3} \leq y_i < \frac{2}{3}$	$\frac{2}{3} \leq y_i \leq 1$
推定誤差	0.1316	0.2978	0.4436

表 7 音響的類似度の相関行列

	インテンシティ	音色	MFCC	クロマベクトル
インテンシティ	1.0000			
音色	<b>0.7937</b>	1.0000		
MFCC	0.4590	0.4161	1.0000	
クロマベクトル	0.2901	0.3826	0.0386	1.0000

Content-based Music Retrieval on Relevance Feedback," *Proceedings of ACM Multimedia 2003*, pp.110-119 (2003).

- 4) Vignoli, F. and Pauws, S., "A Music Retrieval System Based on User-Driven Similarity and its Evaluation," *Proc. ISMIR* (2005).
- 5) McKinney, M.F. and Breebaart, J., "Features for Audio and Music Classification," *Proc. ISMIR* (2003).
- 6) Pampalk, E., Flexer, A. and Widmer, G., "Improvements of Audio-Based Music Similarity and Genre Classification," *Proc. ISMIR* (2005).
- 7) Lidy, T. and Rauber, A., "Evaluation of Feature Extractors and Psycho-Acoustic Transformations for Music Genre Classification," *Proc. ISMIR*, pp.34-41 (2005).
- 8) Lampropoulos, A.S., Sotiropoulos, D.N. and Tsihrantzis, G.A., "Individualization of Music Similarity Perception via Feature Subset Selection," *IEEE, International Conference on Systems, Man and Cybernetics 2004* (2004).
- 9) Gruz, A.A., Downie, J.S., Jones, M.C. and Lee, J.H., "Evalutron 6000: Collecting Music Relevance Judgments," *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries* (2007).
- 10) Jones, M.C., Downie, J.S., and Ehmann, A.F., "Human similarity judgements: Implications for the design of formal evaluations," *Proc. ISMIR2007* (2007).
- 11) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R., "RWC Music Database: Popular, Classical, and Jazz Music Databases," *ISMIR*, pp.287-288 (2002).
- 12) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R., "RWC Music Database: Music Genre Database and Musical Instrument Sound Database," *ISMIR*, pp.229-230 (2003).
- 13) 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 (2006).
- 14) Gómez, E., "Tonal description of music audio signals," Ph.D. thesis, UPF, Barcelona, Spain (2006).
- 15) Lartillot, O. and Toiviainen, P., "MIR in Matlab (II): A toolbox for musical feature extraction from audio," *Proc. ISMIR* (2007).
- 16) Takahashi, R., Ohishi, Y., Kitaoka, N. and Takeda, K., "Building and Combining Document and Music Spaces for Music Query-By-Webpage System," *INTER-SPEECH 2008*, pp.2020-2023 (2008).