

楽曲の音響信号に対する調性（長調 / 短調らしさ）の判定方法

山崎 道弘 岩崎 知弘

三菱電機（株） 情報技術総合研究所

1. はじめに

近年HDDを搭載したカーオーディオや携帯音楽プレーヤーが普及してきており、大量の音楽データから聴きたい音楽をすばやく検索するニーズが増えてきている。音楽をすばやく検索するための方法として、音楽を分類しユーザに提示することが考えられる。

調性（長調、短調）は楽曲の主観（明暗）に関連がある^[1]物理量であり、分類の尺度の一つとして適当であると考えられることから、調性（長調、短調）の判定方法について検討を行った。

2. 判定方式

長調・短調では使用される音高の分布が異なる。このため、楽曲で使用されている音高の分布を調べることで長調・短調を判定することが可能である。本稿では、以下の様な手順で楽曲の音響信号に対する長調らしさ、短調らしさの判定を行った。

2.1 特徴量の抽出

音高の分布を抽出するための特徴量として、音楽で使用されている音域のうち、主にメロディー部で使用されている音域と、主にベース部で使用されている音域それぞれに対し、12次元のクロマベクトル^[2]を作成した。このクロマベクトルは各次元の値が12平均律の異なる音名の正規化したパワーを表すものとし、以下のように求めた。

まず時刻 t の音響信号に対して短時間フーリエ変換(SFFT)を行い周波数軸が線形スケールのパワースペクトルを求めた後、周波数軸を対数スケールの周波数 f_{cent} に変換したパワースペクトル $\psi(f_{cent}, t)$ を求めた。次に各音名 c 、オクターブ位置 h に対する時刻 t の時のパワー $P(c, h, t)$ を以下の式で求めた。

$$P(c, h, t) = \int_{-\infty}^{\infty} W(c, h, f_{cent}) \psi(f_{cent}, t) df_{cent} \quad (1)$$

ここで、 $W(c, h, f_{cent})$ は音名 c 、オクターブ位置 h のパワーを通過させる窓長100centのハミング窓形状のバンドパスフィルターである。

次にノイズ成分を除去するため、 $P(c, h, t)$ の周波数軸方向に移動平均によりスムージングした値を引いて正規化したパワー $P_N(c, h, t)$ を式(2)により求めた。

$$P_N(c, h, t) = P(c, h, t) - \frac{(0.5P(c-1, h, t) + P(c, h, t) + 0.5P(c+1, h, t))}{2} \quad (2)$$

以上のようにして求めた正規化パワー $P_N(c, h, t)$ を式(3)により音名毎に全オクターブ・時刻に対して畳み込んだ値 $v(c)$ を求め、式(4)により $v(c)$ の絶対値の最大値で正規化を行なった値 $V(c)$ をクロマベクトル \vec{v} の各次元の値とした。

$$v(c) = \sum_{\forall t} \sum_{\forall h} P_N(c, h, t) \quad (3)$$

$$V(c) = \frac{v(c)}{\max_{\forall k} |v(k)|} \quad (4)$$

メロディー部、ベース部のクロマベクトルの算出に用いたパラメータを表1、表2に示す。

表1 メロディーの音程取得分析パラメータ

分析パラメータ	値
サンプリング周波数	22.05 kHz
FFT次元数	4096
フレーム周期	20 msec
フレーム長	20 msec
フレーム窓	ハミング窓
音程取得周波数範囲	220 Hz (A3) ~ 3322.44 Hz (G#7)

表2 ベースの音程取得分析パラメータ

分析パラメータ	値
サンプリング周波数	689.0625 Hz
FFT次元数	1024
フレーム周期	100 msec
フレーム長	1500 msec
フレーム窓	ハミング窓
音程取得周波数範囲	65.41 Hz (C2) ~ 246.94 Hz (B3)

"A judgment algorithm of a major scale / a minor scale for the audio signal of music"
Michihiro Yamazaki, Tomohiro Iwasaki. Information Technology R&D Center, Mitsubishi Electric Co.

2.2 中心音による正規化

中心音による正規化を行い、音名の正規化したパワーを表すクロマベクトル \vec{v} から、各次元が階名に対する正規化パワーを表すクロマベクトル \vec{v}_s を求める。これは、中心音が異なっても音楽を構成する音の階名が同じ場合、同じ判定結果が得られる様にするためである。ここで今回対象とした音階を表 3 に示す。

表 3 使用音階

種別	音階
7音音階	イオニアンスケール、エオリアンスケール、ハーモニックメイジャー、ハーモニックマイナー、他計8種
5音音階	琉球音階、陰旋法上行、陰旋法下行、他計9種
その他	6音2種類、4音2種類

なお、イオニアンスケールとエオリアンスケールの様に使用する音程の相対関係が同じになるものに関しては長音階に属するものを代表として一つ選択し使用した。中心音の推定は \vec{v} の各次元を順次中心音と仮定し、その時の使用音と表3の使用音とが一致した場合に仮定された次元の音を中心音とした。ここで使用音は $V(c)$ が閾値以上となる音として、閾値は最初0とし、表中一致する音階がない場合は閾値を変更して一致するものを検索する方式とした。

以上のようにして求めた中心音に対応する次元を c_i 、階名に対応する次元を s とすると、クロマベクトル \vec{v}_s の各要素値 $V_s(s)$ は式(5)により求まる。

$$V_s(s) = V((s + c_i) \bmod 12) \quad (5)$$

2.3 長調・短調評価値の算出

次にクロマベクトル $V_s(s)$ から、長調・短調判定用の評価値 x をクロマベクトル $V_s(s)$ の重み付き線形和として式(6)で求めた。

$$x = \sum_{\forall s} \alpha_m(s) V_{sm}(s) + \sum_{\forall s} \alpha_b(s) V_{sb}(s) \quad (6)$$

なお、式(6)において、 $V_{sm}(s)$ はメロディー部のクロマベクトル、 $V_{sb}(s)$ はベース部のクロマベクトルであり、 $\alpha_m(s)$ 、 $\alpha_b(s)$ は予め求めた重み係数である。

$\alpha_m(s)$ 、 $\alpha_b(s)$ は中心音・調性の判っている楽曲を学習用データとして、既知の中心音を用いて求めた x と、長調を1、短調を-1とした基準値とのピアソンの相関係数が最大になる様に求めた。なお、本稿では、学習用楽曲として、クラシック362曲を使用し、係数1となる音階を2個、係数-1となる音階を2個、他の音階に対する係数

を0とした。この結果ベース部、メロディー部共に係数1となる音階は*i*、*v*、係数-1となる音階は*iii*、*vi*であった。また、この時の相関係数は0.693であり、評価値 x の正負による調性判定精度は84.3%であった。

3. 評価実験

3.1 評価用データ

評価用データとして、予め調性の判っている楽曲(クラシック)81曲を用いた。なお評価用データは係数の学習に使用した楽曲とは異なる楽曲である。

3.2 評価結果と考察

評価用データに対して本方式を用いて実験を行った結果を表4に示す。また参考として中心音の推定を行わずに既知として正規化を行った場合の結果を合わせて示す。

表 4 評価結果

実験	正解率
中心音推定	79.0%
中心音既知	81.5%

本方式による長調・短調の正解率は中心音を推定した場合でも79.0%の精度を得ることが出来、中心音既知とした場合の81.5%とほぼ同等の正解率が得られた。

以上から、音響信号のパワースペクトルから音高のパワー分布を求めることにより、自動採譜等により旋律を求めることなく長調・短調の判定が可能であることが示せた。一方学習用データに対する正解率でも84.3%であり、今後正解率を上げていくためには重み係数の調整、転調への対応等が必要と考えられる。

4. まとめ

本稿では、音楽音響信号に対する長調・短調らしさの自動判定方法について検討を行った。その結果、パワースペクトルから音高毎のパワーの分布を求めることにより自動判定が可能である事を示した。

一方、中心音既知の場合と比べ正解精度が下がっていることから今後自動判定精度を上げるためには、使用音階の判定精度の向上が必要である。また今回の実験では楽曲全体に対する判定で転調への対応が出来ておらず、今後対応が必要である。

参考文献

- [1] 谷口高士「音は心の中で音楽になる」北大路書房
- [2] Shepard, R. N. Structural representation of musical pitch. In D. Deutsch(Ed.) The psychology of music. New York: Academic press 1992