

楽器音の音響特徴の統計分布を用いた 音楽検索

津川博光[†] 小杉隼人^{††} 相川清明^{††}

楽器音の音響特徴の統計分布を用いた音楽検索を行うシステムについて述べる。従来のテキスト検索には、演奏内容に直接関係した検索を行いにくいことや、使用されている楽器などによって検索ができないといった問題点がある。そこで、本研究ではスペクトル時系列の特徴の統計的分布を用いた各種の楽曲検索法を比較した。その結果、ピアノを含む楽曲の検索において 0.92 程度の F 値を得ることができた。

Music Retrieval by the statistical distribution of Acoustical Features of Piano

Hiromitsu Tugawa[†], Hayato Kosugi^{††} and Kiyoaki
Aikawa^{††}

This paper proposes a method for music retrieval by the statistical distribution of acoustical features of musical instruments. Conventional text-based music retrieval uses only limited information on the contents of music. This report compares three statistical methods on retrieving music including piano. Experimental results indicated the maximum F-value was 0.92.

1. はじめに

近年のインターネットや mp3 プレーヤー、インターネットや携帯電話による音楽配信の普及などにより、個人が音楽を大量に所持し、管理することが可能になった。それに伴い、情報の肥大化により、管理が難しく、ユーザにとって本当に聞きたい楽曲を探すことが困難な状況が発生することがある。従来から 2 次元フィルタを用いて調波音と打楽器音を分離する[1]や多重音から短音と類似音色のクラスタリングを行う[2]のような楽音分析の方法が提案されている。また、感性表現を用いて音検索を行った[3]や、音声対話中の強調発話検出に音響特徴を用いた[4]などがある。また、パソコンや携帯端末などで歌検索 (midomi) [5]、波形による検索 (Shazam) [6]。などが実用化もされている。しかし、いずれもデータベースとの照合であり、それらは演奏内容に直接関係した検索を行いにくく、使用されている楽器などによって検索ができないなどの問題点がある。そこで、ユーザが楽曲の演奏内容をデータから検索できるシステムがあれば便利である。このような背景の中で、音楽波形データそのものを分析し、使用されている楽器などの音響特徴の分析を行い、音響特徴のうち音の立ち上がりと減衰の仕方に注目し、ピアノとシンバルの楽器音の検出を行った研究が楽器音の音響特徴を用いた音楽情報検索[7]である。本報告では、さらにピアノについての音響特徴の統計分布の形状による方法と高次モーメントによる方法と複数特徴による方法の比較を行った結果について述べる。

2. 検出方法の検討

2.1 予備検討

楽器音の音響特徴を用いた音楽情報検索[7]において楽器の時間周波数特徴をスペクトログラム上で分析した結果。ピアノとシンバルは音が鋭く立ち上がったあと、緩やかに減衰していることがわかった。ピアノを含む楽曲については 2000~5000Hz の範囲で、シンバルを含む楽曲については 12000~18000Hz の範囲で周波数を抜き出した時に正方向に大きく反応することがわかった。このような音の鋭い立ち上がりと緩やかな減衰を検出するために、まずスペクトログラムから特定の周波数帯を抜き出し、そのパワー時系列を時間微分する。その結果、正方向の振れ、即ち鋭い立ち上がりがあれば、プロットされる図は正の大きな値をとる。この値を閾値として、ピアノやシンバルを検出する。

先行研究ではこのように統計分布の形状を用いた方法を用いて実験を行ったが、ピアノ打鍵時のパワーの急激な立ち上がりを検出する方法として優れているかどうかの検討までは行われていなかった。また、正規分布からのずれを計測するために高次モーメントである歪度や尖度を用いる方法も考えられる。本報告ではこれらの方法を比較する。また、複数周波数を用いて検索を

[†]東京工科大学メディア学部

Tokyo University of Technology, School of Media Science

^{††}

行うことにより、一つの周波数上に他の楽器音が重畳した場合の誤検出を防げるのではないかと考えた。

2.2 二つの閾値を用いた方法

ピアノを含む楽曲で、4000Hz 付近を抜き出し、そのパワー時系列を時間微分した値のヒストグラムを表示してみると、正の方向に歪んだ分布になる。正方向へ大きく反応しているかどうかだけで判断すると、偶然に正方向へ反応した場合や、音の急激な減衰、即ち負方向への反応を含む場合も対象楽器を含んでいると判断するので、他の楽器を対象楽器として誤検出してしまう可能性が大きい。したがって、一定の区間内で、一定の頻度で正方向への反応を示し、負方向への反応は一定の頻度以下であった場合に限り対象楽器を含むという統計的状況から判断する。標準偏差 σ を求め、パワーの時間微分のすべての値を分析し、図 1 ように $\pm 2\sigma$ を超えた回数を数える。そして $+2\sigma$ を超えた数と -2σ を越えた回数の比率を求め、前者の占める割合の方が多ければ、対象楽器が含まれていると判断する。例えば $+2\sigma$ を超えた回数を n 、 -2σ を超えた回数を o 、閾値 x を 1 以上の値としたとき、

$$\frac{n}{o} \geq x \quad (1)$$

を満たせば、正方向へ反応しているとし、対象楽器が含まれていると検出する。楽器音の音響特徴を用いた音楽情報検索[5]では、この方法を用いてピアノの検索において F 値 0.9174、シンバルの検索において F 値 0.923 の値を得ることができた。

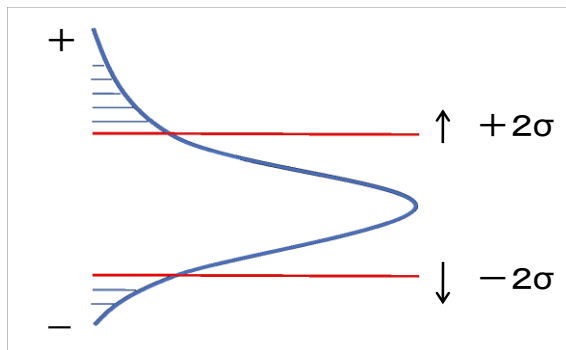


図 1 鋭い立ち上がりが多く含まれるパワー時系列微分の統計的分布

2.3 高次モーメントを用いた方法

パワーの時間微分のすべての値に対しての歪度の値を以下の式で求める。

$$Sk = \sum_{i=1}^n (X_i - \bar{X})^3 / (nV^{1.5}) \quad (2)$$

また、同様に尖度の値を以下の式で求める。

$$K\omega = \sum_{i=1}^n (X_i - \bar{X})^4 / (nV^2) - 3 \quad (3)$$

二つの閾値を用いた方法では $\pm 2\sigma$ の閾値を越えたデータのみを判断材料にしていたが、歪度を用いると楽曲の音響データ全体の正方向の偏りを判断材料として使えるので、さらに精度が上がるのではないかと予想した。

尖度においては、急激で大きすぎる立ち上がりや立ち下りではピアノを含んでいると判断しなくなるため、オルガンなどの誤検出を防ぎ、精度が上がるのではないかと予想した。

歪度や尖度の値が一定以上であるときにパワーの時間微分の統計的分布に偏りがあると判断し、対象楽器が含まれていると検出する。

2.4 複数周波数を用いた方法

二つから三つの周波数から検出した n と o の比率や、歪度や尖度を共変量として判別分析を行う。導き出される判別式は以下の通りである。

Mahalanobis 距離を計算するために必要な 2 次式を以下に示す。

$$SqDist = a * q1^2 + b * q1 * q2 + c * q2^2 \quad (4)$$

「F」(ピアノが含まれないと判定されるもの)の重心から各オブザベーションまでの Mahalanobis の距離をもとめる式を以下に示す。式(4)を式(5)に組み込んだものである。

$$SqDis[F] = SqDis[0] + d * q1 + e * q2 + f \quad (5)$$

「T」(ピアノが含まれると判定されるもの)の重心から各オブザベーションまでの Mahalanobis の距離を求める式を以下に示す。式(4)で係数を変えたものを式(6)に組み込んだものである。

$$SqDis[T] = SqDis + g * q1 + h * q2 + i \quad (6)$$

個々の対象から「F」グループの重心までのマハラノビスの距離 $SqDis[F]$ と「T」グループの重心までのマハラノビスの距離を $SqDis[T]$ とし、次のような判別関数を求める。

$$Z = SqDis[F] - SqDis[T] \quad (7)$$

式(7)を閾値とし、一定以上のときに対象楽器が含まれているかを判定する。

3. 評価実験の方法

3.1 使用楽曲

評価実験用の楽曲には RWC 研究用音楽データベース[8]を使用した。RWC 研究用音楽データベースは研究者が研究目的に利用する上で、著作権等による制約を受けずに、共通利用の自由、学術利用の自由が確保された音楽情報処理研究用データベースである。データベースの音源は製作者の意図したミックスバランスの「オリジナル版」を使用しており、サンプリング周波数 44.1kHz、16bit リニア量子化、ステレオ(2チャンネル)のデータである。データベースの音楽CDから wav ファイルの形式で取り出したものを使用する。RWC 研究用音楽データベースは、大きく5つのデータベースに分けることができる。楽器音データベースを除く4つのジャンル、全365曲の中から先行研究と全く同じ100曲を抜粋した。ピアノを含む楽曲の選定にあたっては、ピアノの楽器音が含まれている曲50曲と含まれていない楽曲を50曲とし、選曲が一つのジャンルに偏り過ぎないように考慮した。使用するにあたっては各楽曲を20秒に切り出したものを使用した。使用した楽曲は表1の通りである。

表1 使用した楽曲の内訳(ジャンル)

ジャンル名	曲数	データベース総曲数
クラシック	20 曲	50 曲
ジャズ	26 曲	50 曲
ポピュラー音楽	27 曲	100 曲
音楽ジャンル	27 曲	100 曲
合計	100 曲	200 曲

4. 実験結果

4.1 高次モーメントを用いた方法の実験結果

二つの閾値を用いた方法において一番性能が良かった 4000Hz を用いて実験を行った。各値は左から閾値を 0.1 刻みで変えていったときの値を示している。歪度では図2のように検索条件として歪度>0.5 のときに F 値 0.8991 で 89% の値を得た。尖度では図3のように尖度>4.5 のときに F 値 0.8224 で 81% の値を得ることができた。二つの閾値を用いた方法の方がよりも精度が高いことがわかった。

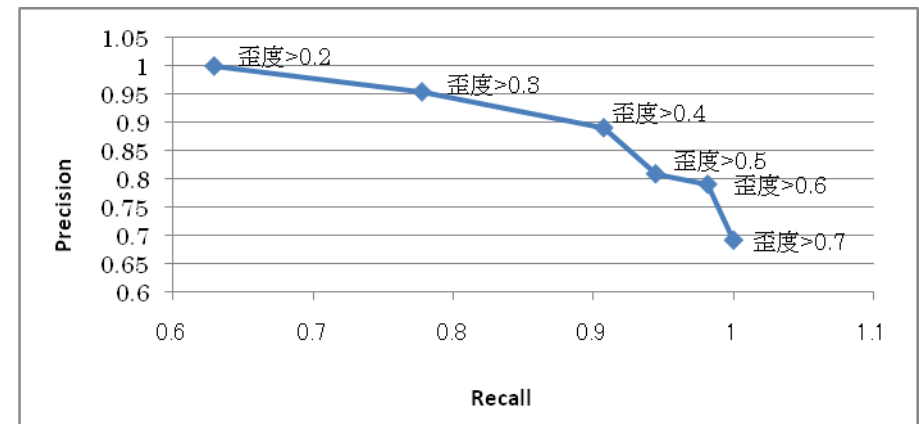


図2 パワー時系列微分の歪度を用いたピアノ検索における再現率と適合率

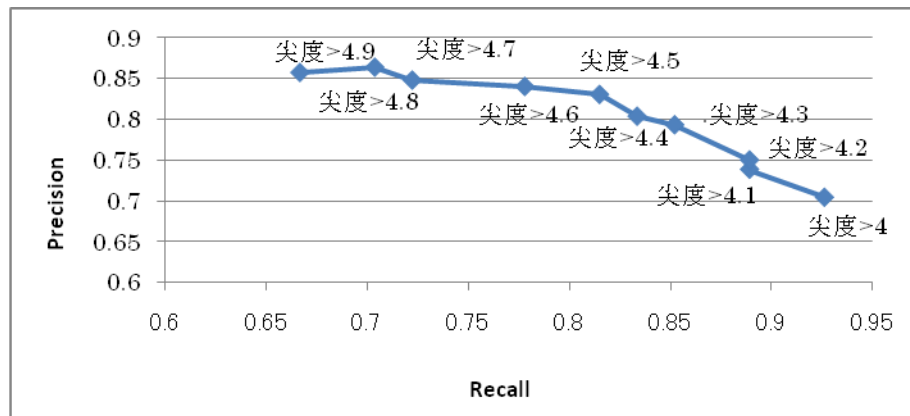


図3 パワー時系列微分の尖度を用いたピアノ検索における再現率と適合率

4.2 複数特徴による判別分析を用いた実験

ピアノの特徴を良く表している4000Hzを基本とし、他の楽器音が重畳した時を想定し、ある程度離れた8000Hz,12000Hz,16000Hz,20000Hzを組み合わせで実験を行った。その結果、4000Hzと12000Hzと4000Hzと16000Hzのときに、図4のように左から閾値を1刻みで変えて、検索条件として $z > -1$ のときにF値0.9174の値を得ることができた。結果としては二つの閾値を用いた方法と同じ値となった。また、三つの周波数を組み合わせても、同様に実験をおこなった。図5は最も高い値を出した4000Hzと8000Hzと12000Hzの組み合わせであり、左から閾値を1刻みで変えて、検索条件として $z > -1$ のときにF値0.9159の値となり、これも二つの閾値を用いた方法と同様程度の結果となった。図6のように歪度においても複数を組み合わせで実験をおこなったが、図6のように4000Hzと16000Hzでもっとも高いF値0.8991となり、パワー時系列微分の歪度をそのまま閾値として用いていた結果と同様な値となった。二つや三つの周波数を使っても結果に大きな変化は見られなかった。

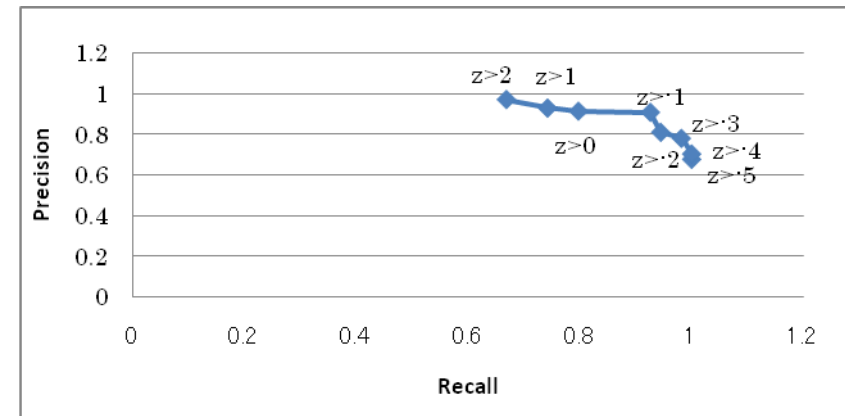


図4 4000Hzと12000Hzのパワー時系列微分の統計的分布を用いた判別分析によるピアノ検索における再現率と適合率

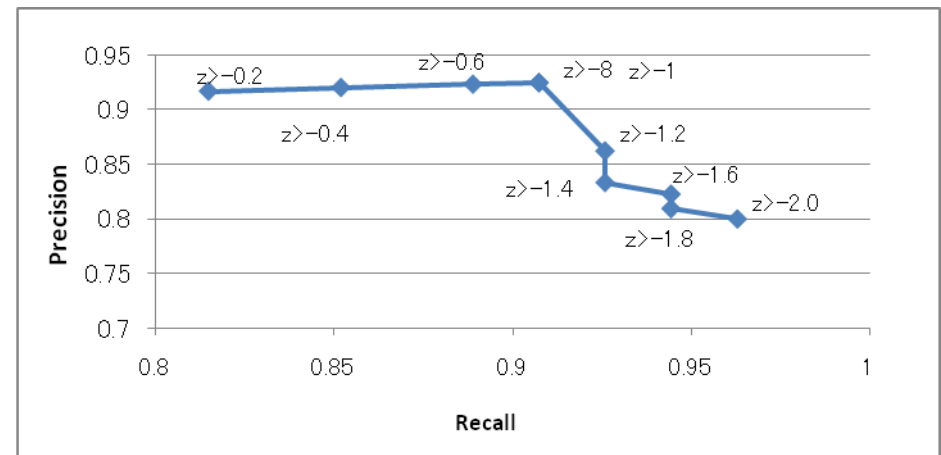


図5 4000Hzと8000Hzと12000Hzのパワー時系列微分の統計的分布を用いた判別分析によるピアノ検索における再現率と適合率

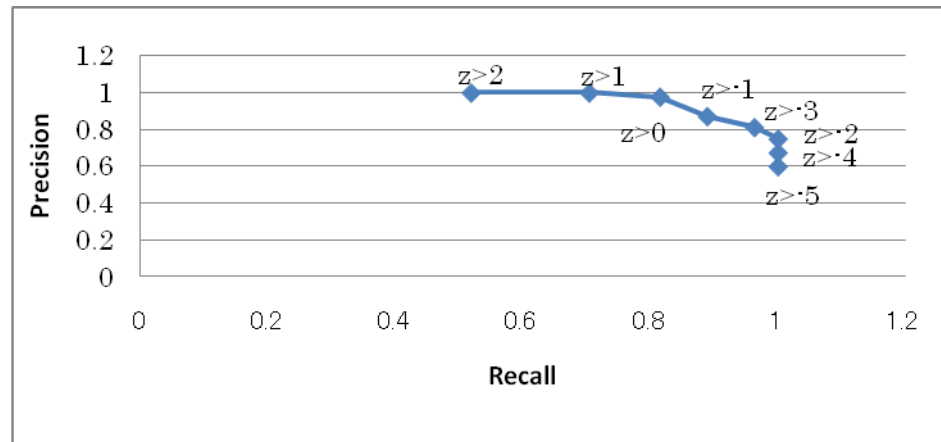


図6 4000Hzと16000Hzのパワー時系列微分の歪度を用いた判別分析によるピアノ検索における再現率と適合率

4.3 考察

各統計分布特徴の検出方法の結果を比較すると表4のような結果となった。表5のように二つの閾値を用いた方法と高次モーメントである歪度を用いた方法では、間違えた曲にばらつきがあった。二つの閾値を用いた方法と歪度を用いた方法に傾向や得意、不得意があると思われ、スペクトログラムで比較、分析を行った。二つの閾値を用いた方法ではピアノが含まれているものを含まれていないと間違えた曲が6曲で、ピアノが含まれていないものを含まれていると間違えた曲が3曲であったのに対して、歪度を用いた方法ではピアノが含まれているものを含まれていないと間違えた曲が8曲で、ピアノが含まれていないものを含まれていると間違えた曲が2曲であった。

表4 各統計分布特徴の検出方法の結果の比較

検索方法	最大F値
二つの閾値を用いた方法(4000Hz)	0.9174
歪度を用いた方法(4000Hz)	0.8991
尖度を用いた方法(4000Hz)	0.8224
複数周波数を用いた方法(4000Hzと12000Hz)	0.9174

表5 二つの閾値を用いた方法と歪度を用いた方法において間違えた曲の比較

二つの閾値を用いた方法で間違えた曲	歪度による方法で間違えた曲	両方の方法で間違えた曲
ジャズ_No.18	クラシック_No.33	クラシック_No.49
ジャズ_No.28	クラシック_No.37	ジャズ_No.25
ポピュラー音楽_No.86	ジャズ_No.23	ジャズ_No.32
音楽ジャンル_No.43	ポピュラー音楽_No.03	音楽ジャンル_No.01
	音楽ジャンル_No.74	音楽ジャンル_No.55

歪度を用いた方法において間違えた曲を考察すると、図7のようにピアノの特徴が見られるが、サステインペダルを多く踏んでいるために、強い立ち上がりを持たない。歪度を用いた方法の特徴である楽曲の音響データを全体で判定を行っているために、誤検出になったと考えられる。また、複数周波数を用いた方法においては、先行研究と複数周波数の方法とほぼ一緒であったが、1曲だけ間違えた曲が違う曲があった。今後、閾値と歪度の特徴を組み合わせや、周波数の組み合わせをさらに研究し、改善する必要がある。

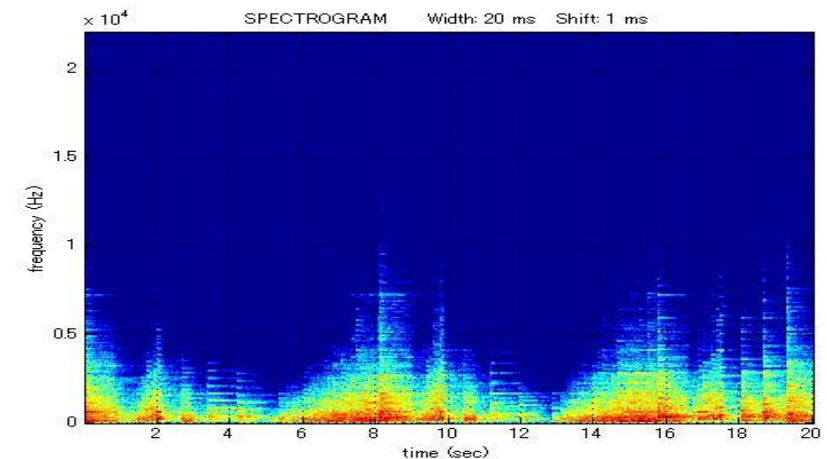


図7 パワー時系列微分の歪度を用いた方法においてピアノが含まれているのに含まれていないと誤判別した楽曲のクラシック_No.33のスペクトログラム

5. むすび

本研究では、ピアノが持つ、音が鋭く立ち上がり、緩やかに減衰するという音響特徴を、一つの判定方法や音響特徴でなく、複数の方法で検索する方法について検討した。そして、従来の音響特徴の音響特徴の統計分布において二つの閾値を用いた方法と高次モーメントを用いた方法と複数周波数を用いて検索する方法を比較し、その結果、F 値 0.9174 である二つの閾値を用いた方法の有効性を示した。さらに、二つ以上の音響特徴を用いて判別分析を行った。その結果、F 値 0.9154 が得られた。閾値と歪度においては誤検出した楽曲に違いがあるため、今後は判定式に用いる閾値や歪度や尖度や判定に用いる周波数や、閾値と歪度における検索の傾向などについてさらに細かく分析する必要がある。また、100 曲という曲数は決して十分な数とは言えず、一般的に使用するためにまだ改善点が多い。

謝辞 本研究では、評価実験のために RWC 研究用音楽データベースを利用しました。

参考文献

- [1] 宮本賢一、亀岡弘和、小野順貴、嵯峨茂樹, “スペクトログラムの滑らかさの異方性に基づいた調波音・打楽器音の分離”, 日本音響学会講演論文集, pp.903-904, (2008-03)
- [2] 宮本賢一、亀岡弘和、西本卓也、小野順貴、曾我山茂樹, “Source-Filter モデルを含めた調波構造・時間包絡・音色の統合的クラスタリング(HTTC)による楽音分析”, 日本音響学会講演論文集, pp.907-908, (2008-3)
- [3] 上野 智子、相川 清明, “音楽検索のための感性表現ベクトルと音響特徴量の関係の分析”, 情報処理学会研究報告, 2008-MUS-74, 2008-SLP-70, Vol. 2008, No. 12, pp.211-216, (2008-02).
- [4] 上野智子、相川清明, “音響特徴による音声対話中の強調発話検出”, 日本音響学会講演論文集, pp.183-184, (2007-03)
- [5] Midomi サイト, <http://www.midomi.co.jp> (2010,1 アクセス)
大竹俊介、菅村昇, “楽曲検索エンジン(midomi)の評価”, 工学院大学卒業論文, 2009-03
- [6] Shazam サイト, <http://www.shazam.com/> (2010,1 アクセス)
- [7] 小杉 隼人, “楽器音の音響特徴を用いた音楽情報検索”, 東京工科大学卒業論文, 2009-03
小杉 隼人、相川 清明 “音響特徴によるピアノ音とシンバル音を含む楽曲検索”, 日本音響学会講演論文集 ,3-8-5, (2009-03)
小杉 隼人、相川 清明, “ピアノとシンバルの音響特徴による音楽検索”, 情報処理学会研究報告, 2008-MUS-78, Vol. 2008, No. 127, pp.35-40, (2008-12)

[8] 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”, 情報処理学会論文誌, Vol.45, No.3, pp728-738, (2004)