

Specmurt 分析と Chroma Vector を用いた HMM による音楽音響信号の調認識

齊藤 翔一郎[†] 武田 晴 登[†]
西 本 卓 也[†] 嵯峨山 茂樹[†]

本稿では、音楽音響信号の調を認識する新しい手法として、Specmurt 分析によって得られた chroma vector 時系列から確率モデルを用いて調を推定する手法を提案する。提案手法は、音響信号からの音高特徴量抽出と、確率モデルによる調推定の 2 つの処理からなる。音高情報を表す特徴量である chroma vector を、多重音楽信号から基本周波数を強調する手法である specmurt 分析を施して得られる多重音のピッチ情報から得る。また、chroma vector は和声状態に対して確率的に変動し、和声状態のマルコフモデルでモデル化する。この 2 つのモデルは、調を HMM (Hidden Markov model, 隠れマルコフモデル) とする確率モデルとして統合される。調推定は、音響信号から得られた chroma vector の時系列に対して事後確率を最大にする HMM の探索として行われる。評価実験において、RWC 音楽データベースを用いた調の推定について 93.8%、また MIREX 評価用データに対して 86.5% の正解率を得た。

Key Detection of Music Audio Signals via HMM using Chroma Vector through Specmurt Analysis

SHOICHIRO SAITO,[†] HARUTO TAKEDA,[†] TAKUYA NISHIMOTO,[†]
and SHIGEKI SAGAYAMA[†]

This paper describes a method for key detection from polyphonic audio signals based on probabilistic modeling using HMMs (Hidden Markov Models) with chroma vectors extracted through specmurt analysis. The proposed method consists of two processes. First, a chroma vector sequence is extracted to obtain robust multi-pitch information from audio signals. In extracting a chroma vector, specmurt analysis is applied to emphasize fundamental frequencies in power spectrum structured by multi-pitch harmonic pattern. Next, we model fluctuation of chroma vectors by probabilistic distribution characterized by chord states in a key, and also time sequence of chord in a key with Markov model. These two probabilistic models are integrated in the framework of HMM. For a given chroma vector sequence, we find the key that maximizes *a posteriori* probability. This MAP estimation is performed by an efficient search algorithm known as time synchronous Viterbi search. Experimental evaluation shows 93.8% key detection accuracy for 64 pieces in RWC music databases, and 86.5% for 96 pieces in the MIREX data.

1. はじめに

本稿では、specmurt 分析と隠れマルコフモデルを用いて音楽音響信号を対象とした調の推定手法を提案する。

音楽信号から調 (tonality) を推定する技術は、音楽情報検索、自動採譜、自動編曲、自動伴奏などの多くの場面で重要な役割を果たす。自動採譜においては、調号推定、転調の検出、情報検索においては、調の正規化

後の類似度計算、調性をキーとする音楽検索、自動編曲や伴奏においては、旋律の調性推定とその調に基づく編曲や和声づけ、和声機能の解析などの応用が考えられる。

調は、西洋音楽の音階に基づいて決められるものである。音響信号からの調推定は音響信号から得られる音高情報をもとに行うのが自然だと考えられる。実際、人間が聞いている音楽の調を明確に認知するには、演奏されている音楽の旋律や和声などの音高情報を認識し、その音高情報をもとに音楽的知識を用いて調を推測すると考えられる。現在提案されている音響信号からの手法では、まず音響信号から chromagram など

[†] 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology,
The University of Tokyo.

の特徴量を抽出し、それを prob tone profile method (K-S's method) を用いて調を決定する Pauws¹⁾ や Gómez ら²⁾ の研究がある。また、音高信号から調を推定する方法として、特徴量として和声の意味から音高情報をモデル化した spiral array を用いる手法が提案されている³⁾。

今回、我々はこれまでに行ってきた音高情報取得の方法と和声のモデル化を組み合わせる音響信号から調を推定する手法を提案する。我々はこれまでに、音楽音響信号から音高情報を得る手法として、調波構造の精密なモデル化に基づく基本周波数推定法であるハーモニッククラスタリング⁶⁾、及び、高調波成分を抑圧し音高らしさを推定する specmurt 分析^{5),7)} を提案してきた。また、楽曲に現れる和声のモデルとして、与えられた旋律に対して自動的に適切な和声を付与する目的で、隠れマルコフモデルでモデルを用いて旋律と和声を含む音高情報をモデル化し、調推定も行った⁸⁾。今回は、音高情報を specmurt 分析により得た特徴量をもとに HMM を用いて調を認識する手法を検討する。

2. Specmurt 分析による chroma vector

2.1 Specmurt 分析による音高情報取得の意義

我々は、和声の構成音の音高を取得する為に、音響信号の短時間周波数解析から、演奏されている音の厳密な基本周波数と音源数を求めるのではなく、和声の推定の手がかりとして用いるのに十分なであろう複数の音の音高情報らしさを得る為に、specmurt 分析を行う。一般に、短時間周波数解析で得られる音響信号のパワースペクトルは、各楽音の倍音に対応する周波数のパワーから本来の和声の構成音と異なる成分が現れるので、和声を構成する音高情報を取得するには、この倍音成分を除去しなければならない。Specmurt 分析は、簡潔な計算で効率的に倍音成分を抑圧する手法であり、この目的に適している。

2.2 Specmurt 分析の概要

Specmurt 分析は調波構造を持つ楽音に対し、周波数領域において倍音成分を抑圧し基本周波数を強調する分析法である。調波構造の倍音パワー比は基本周波数に依らず共通（「共通調波構造モデル」と呼ぶ）と仮定すれば、調波構造を持つ音の基本周波数がシフトした場合、線形周波数領域では倍音の位置関係は線形伸縮となる。一方、対数周波数領域で考えれば、倍音の位置関係も一定で調波構造全体がシフトする。従って、パワースペクトルの加法性を仮定すれば、多重音スペクトルは対数周波数領域における基本周波数の分布と調波構造のモデルとの畳み込みで表せる。図 1 に概念図を示す。

このとき、基本周波数分布を $u(x)$ 、共通調波構造を

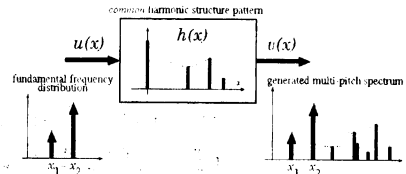


図 1 基本周波数分布と共通調波構造の畳み込みによる多重音スペクトルの生成モデル

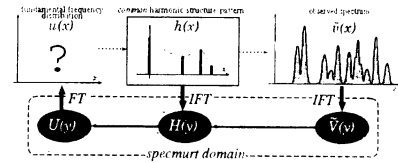


図 2 specmurt 分析の概要

$h(x)$ (x は対数周波数) とすると、多重音スペクトル $v(x)$ は

$$v(x) = u(x) * h(x) \quad (1)$$

と表せる。 $h(x)$ 、 $v(x)$ の逆フーリエ変換をそれぞれ $H(y)$ 、 $V(y)$ と定義すると、フーリエ変換の畳み込み定理から、基本周波数分布 $u(x)$ のフーリエ逆変換 $U(y)$ は、

$$U(y) = \frac{V(y)}{H(y)} \quad (2)$$

と表される (y は対数周波数の逆フーリエ変換領域)。求めたい基本周波数分布 $u(x)$ は $U(y)$ をフーリエ変換することによって得られる。このようにして基本周波数分布を求める手法を specmurt 分析という。概要を図 2 に示す。

今回、我々はこの specmurt 分析を効果的に行うために、以下の 2 つの処理を行う。まず、短時間周波数解析を行うための窓の影響が対数周波数領域で一定の周波数成分の広がりとなるように wavelet 変換を用いる。また、各短時間フレームで最適な共通調波構造を反復推定⁷⁾ により求める。

なお、ここで specmurt 分析と Mellin 変換との関連について述べる。式 (2) における y は前述のように対数周波数の逆フーリエ変換領域であるから、

$$\begin{aligned} G(y) &= \int_0^{\infty} e^{jy \log \omega} F(\omega) d \log \omega \\ &= \int_0^{\infty} e^{jy \log \omega} F(\omega) \frac{d\omega}{\omega} \\ &= \int_0^{\infty} \omega^{s-1} F(\omega) d\omega, \quad \text{但し } s = jy \quad (3) \end{aligned}$$

と変形できるので、 $s = jy$ (純虚数) の場合の Mellin 変換に一致する。しかし、(実数) Mellin 変換が比例伸

縮不変な特徴量の抽出に使われるのに対し、specmurt分析は伸縮係数情報を保存すること。対数周波数は音楽の半音階によく一致すること、また、ここでは対数周波数領域でのフーリエ変換の畳み込み定理が重要な役割を担うことなど、Mellin 変換としての理解は必ずしも有用でない。

2.3 Specmurt から chroma vector への変換

音階名で表される音高情報 (例えば「C3」) は、オクターブ (「3」) と半音階名 (「C」) を指定することによって一意に定めることができる。このうち、オクターブ違いの同じ音階の成分を全て重ね合わせて1オクターブ内の半音階の12音の成分に縮約したものが、chroma vector⁹⁾ である。楽曲の旋律や和声は、全体をオクターブ単位で上下に平行移動しても調性は変化しないことから、chroma vector は、オクターブ方向の分布の情報を取り除くため、調推定に必要な音高情報を圧縮していると言える。このため我々は、音響信号から和声に関する音高情報を表す特徴量として、chroma vector を使用する。^{*}

Specmurt 分析により得た音高情報から chroma vector を求める。Specmurt 分析後の周波数分布を各音階の区間に分け、各区間での強度の最大値をその音階の強度として採用し、その後オクターブ違いの音階の強度を足し合わせ、各要素の和を正規化した。即ち、chroma vector $\mathbf{x} = (x^{(1)}, \dots, x^{(12)})$ を

$$x^{(d)} = \sum_m \max_{R(m,d)} u(\log \omega) \quad (4)$$

とした。ここで d は半音階名、 m はオクターブの高さを表し、 $R(m, d)$ は音階が d 、オクターブが m であるような音であると見なす対数周波数 $\log \omega$ の範囲を表す。

2.4 観測 chroma vector の確率的変動

前節で求めた chroma vector は厳密には真の音高位置のみに成分を持つことは保証されず、実際の specmurt 分析による抑圧しきれていない成分や、周波数分布上での小さな成分の影響や、実際の楽曲の旋律や音強の変化により、同一の調の同一の和声であっても、chroma vector は大きく変動する。我々は、各短時間周波数解析によって得られる chroma vector は、確率な変動を伴って観測されると仮定する。

今回は、正規化された chroma vector は多次元正規分布に従うと仮定する。即ち、調 K で和声 c であるとき、chroma vector が \mathbf{x} として観測される確率は、次

の確率密度分布で与えられるとモデル化する。

$$P(\mathbf{x}|c, K) = \frac{1}{\sqrt{(2\pi)^{12}|\Sigma_{c,K}|}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \mu_{c,K})^T \Sigma_{c,K}^{-1}(\mathbf{x} - \mu_{c,K}) \right\} \quad (5)$$

となる。この分布のパラメータである平均 μ と分散 Σ は学習により適切な値を求めることができる。

更に、各時刻の chroma vector が 独立にその時刻の調と和声によって定められる確率分布に従うと仮定すると、ある調 K で chroma vector の時系列 $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ が観測される確率は、 \mathbf{x}_t に対応する和声 c_t であるとして、

$$P(\mathbf{X}|\mathbf{C}, K) = \prod_{t=1}^T P(\mathbf{x}_t|c_t, K) \quad (6)$$

と表せる。ここで、 c_t は時間 t での和声を表す。

3. 調と和声のモデル化

3.1 和声状態のマルコフモデル

我々は chroma vector を特徴付ける楽曲中の和声に注目し、その時間変化を和声の時系列と見なし、その局所的な特徴を確率モデルで数理的に扱う。楽曲の調は旋律と和声によって定められるが、和声は曲の一部分を扱う場合でも調の特徴をよく反映し、短時間周波数解析においても十分に情報を得ることができるので、和声によって調認識を行うことは十分妥当だと考えられる。

今回、我々はある調は複数の和声状態でモデル化し、音響信号から得られた chroma vector はモデルの中のいずれかの和声状態に対応するものと仮定する。ここで用いる和声状態は、I, V などのように和声学に基づいて明示的に定められるものではない。一般に和声学の定める和声は多種であり、調認識の為に厳密な和声を推定する必要はないと考えられる。さらに、あらゆる音響に対応するように和声学において考えられる和声を全て和声状態として定義すると、状態数が多くなり、データスパースネスに陥る可能性もある。今回は、学習時に初期状態として I, II, IV, V, VI の5和音を与えて得られる、5状態として最適化された和声状態のモデルを用いる。

一般に和声の時間変化は、局所的には調性感を損なわないようカデンツを形成しながら変化し、一方大域的には楽曲形式などに依存した曲の構成によっても特徴づけられる。今回、我々は曲全部からではなくその一部から演奏部分の調性を推定することを目的としているため、局所的な性質によってのみモデル化することが好ましい。そこで、和声状態の時間変化の無記憶

^{*} 反面、展開形やバスの保続音など和声上の音高配置、楽器特様の音域などの絶対音階に依存する手がかりは失われてしまう。

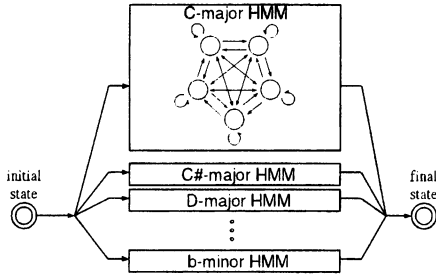


図3 調推定における調のネットワークモデル。観測された chroma vector の時系列は和声状態を隠れ状態とする HMM ネットワーク上の遷移に対応する。

性と、全ての和声状態間の遷移が許されると仮定し、和声状態をエルゴード的なマルコフモデルでモデル化する。

ある調 K の楽曲で、時刻と共に和声 $C = \{c_1, c_2, \dots, c_T\}$ と進行する確率は

$$P(C|K) = P(c_1, c_2, \dots, c_T|K) \\ = P(c_1|K) \prod_{t=2}^T P(c_t|c_{t-1}, K) \quad (7)$$

と書ける。

3.2 調のモデル

次に、調のモデルについて述べる。図3のように、マルコフモデルが長調と短調それぞれ12個ずつ24個あり、入力である chroma vector 時系列はどれかひとつの調を通過して終端へと到達するようなモデルを考える。楽曲が何調であるかという問題は、入力から何調のマルコフモデルを通して終端に達したかという問題として扱える。ここで調 K の用いられやすさを表す事前確率を $P(K)$ として考えることが出来るが、我々は調は基本的に音高情報から十分に推定できると考えるので、今回は調出現に対する事前知識を用いないことにする。ただし、楽器の種類に応じて用いられ易い調性などが存在することから、演奏に対する事前情報があればこの確率は有効に用いることが可能である場合も存在する。

3.3 調・和声からなる HMM ネットワーク

Chroma vector の時系列 X を音楽音響信号から観測するのに対し、和声進行や調は未知である。我々のモデル化では、各調の仮説において、観測信号に対して隠れている Markov 性を持つ和声進行のモデル化しているのので、これは HMM と見なすことができる。

調が K であり和声進行が C のように進行し、chroma vector の時系列データが X である確率は、式(6)・(7)を用いると

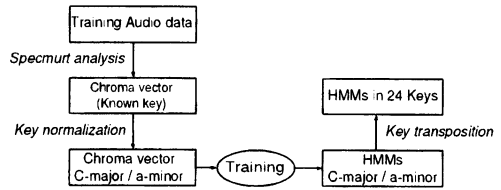


図4 HMM の学習の模式図

$$P(X|C, K) \cdot P(C|K) \cdot P(K) \\ = \prod_{t=1}^T \left\{ P(x_t|c_t, K) \cdot P(c_t|c_{t-1}, K) \right\} P(K) \quad (8)$$

となる。ただし $P(c_1|c_0, K) = P(c_1|K)$ としている。楽曲の調を前節で述べたモデルによって HMM で推定することを考えると、式(8)の $P(x_t|c_t, K)$, $P(c_t|c_{t-1}, K)$ はそれぞれ出力確率、遷移確率に相当する。

4. HMM による調認識

4.1 MAP 推定による調認識

我々は、観測された chroma vector の時系列 X に対して、確率的に最も尤もらしい調 K を推定する。これは、

$$\hat{K} = \operatorname{argmax}_K P(K|X) \quad (9)$$

と表される。ここで、Bayes の定理を用い、また観測からは隠れている和声にも注目すると、この事後確率は次のように書き直される。

$$\hat{K} = \operatorname{argmax}_K P(X|K)P(K) \\ \simeq \operatorname{argmax}_{K,C} P(X|K, C) \cdot P(C|K) \cdot P(K) \quad (10)$$

この式の argmax の中身は式(8)と等しくなり、この式の値を最大化する \hat{K} を求めることになる。即ち、調推定の事後確率最大化問題は、式(8)を最大化する問題に帰着される。

音声認識の場合と同様に、これは HMM のネットワーク上での探索問題であり、時間同期 Viterbi 探索を行うことで最適性を保証しながら効率的な探索が行われ、最大尤度とその状態遷移列(和音進行)が求められる。さらに、このような HMM を図3のように、長調、短調各12個用意し、それぞれについて尤度を計算し比較することで、楽曲の調を最大尤度を出力する HMM の調であるとして推定できる。

4.2 HMM の学習

HMM の状態遷移確率と出力確率は学習データを用いて学習できる。しかし、24種のすべての調を平等に扱うためには、調に依存した偏りが全く無い学習用データを得ることが必要になる。そこで、長調と単調

のそれぞれについて、これらをまず単一の調に移調し、単一の調について学習する。ここでは長調の曲はハ長調、短調の曲はイ短調に揃えている。これは音響信号から得られた chroma vector の要素を回転することで実現できる。例えば二長調の学習データがあった場合、その chroma vector の 3 番目の要素を 1 番目に、2 番目の要素を 12 番目に、などすることで、その曲がハ長調であるとしたときの学習データが得られることになる。

学習用データを探索すると、ハ長調もしくはイ短調において初期パラメータに対してもっとも尤度の大きい状態遷移列が求まる。状態遷移確率 $P(c_t|c_{t-1}, K)$ (K =ハ長調 or イ短調) は、状態 c_{t-1} から c_t へ遷移した頻度を、状態 c_{t-1} をとった頻度で割った値によって更新する。一方出力確率は、状態 $c \in \{I, II, IV, V, VI\}$ をとった時刻の x_t について平均と分散を計算することで、 μ_{cK} と σ_{cK}^2 が更新される。

こうして学習されたハ長調とイ短調のパラメータを、chroma vector の空間上で回転して戻してやることで、24 個の HMM のパラメータが更新される。ここで、長/短調の各 12 の HMM の状態遷移確率や出力確率は調に依存しないと仮定し、各パラメータは長/短調で値を共有させた。この更新されたパラメータによって、調が未知のデータに対して探索を行い、調の推定を行う。図 4 に HMM の学習の模式図を示す。

なお、今回の評価実験では共分散行列は対角行列と仮定し、式 (5) によって出力確率を計算している。

5. 評価実験

5.1 実験条件

提案手法を実装し、RWC 研究用音楽データベース¹⁰⁾、MIREX2005¹¹⁾ で提供されているデータを対象に評価実験を行った。MIREX とは音楽情報検索に関連する技術の国際コンテストで、音響信号や MIDI 信号からの調推定、ジャンル推定、旋律類似度などの項目が設定されている。また、アルゴリズムの学習に用いるための Training Data Set が公開されている。評価実験全体に共通の実験条件を表 1 に示す。楽曲は 16kHz サンプリングのモノラル音声で、フレームシフトは 16ms、周波数帯域 30-7609Hz で分析する。specmurt 分析の共通調波構造の初期値としては、調波成分を 10 倍音まで考えた周波数の 1.5 乗で減衰する調波モデルを仮定する。

学習、認識に用いた音楽音響信号は、MIREX2005 の “Audio and Symbolic Key Finding” のトピックで公開されている Training Data Set 96 曲 (30 秒程度) と、RWC 研究用音楽データベースのクラシック

音楽*から抜粋した曲のうち、転調していないと思われる 20 秒前後の切出し曲 64 曲である。学習は open data では 4 分割の交差検証法を用いた。また、学習の反復回数は 5 回である。HMM は 5 状態で、各状態の chroma vector 出力確率の平均ベクトルの初期値は I, II, IV, V, VI であるとし、和声の構成音とそれ以外の音階が 5:1 となるように配分したうえで正規化したベクトルを用いている。また、chroma vector 出力確率の共分散行列を対角共分散とし、学習を行うときの初期値を $\sigma^2 = 0.1, 0.05, 0.02$ の 3 種類で行った。

表 1 実験条件

周波数解析	サンプリング周波数	16 kHz
	フレームシフト	16 ms
	ウェーブレット関数	Gabor 関数
	周波数分解能	16 cent
	周波数帯域	30-7609 Hz
HMM	状態数	5
学習	データ分割	4
	反復数	5

5.2 実験結果

表 2 は MIREX の Training Data Set、表 3 は RWC のデータを用いた認識結果である。太字が提案手法の認識正解率であり、分散の初期値によって 3 つの結果を表示している。比較として、specmurt 分析を用いずに周波数分布をそのまま用いた場合 (“wavelet”) と、HMM の状態の時間的遷移を用いずに、曲全体を単一の chroma vector にして認識した場合 (“histogram”) をあわせて掲載してある。

表に示すように、MIREX で 85% 程度、RWC で 90% 以上の正解率を得ることができた。完全五度、平行調、同主調以外の誤りは全体の 1-2% に過ぎなかった。MIREX の結果では、specmurt 分析を実行した上で HMM を用いる提案手法が最も認識率が高いという結果が出ている。RWC では全体の認識率はより高いものの、手法によってどちらが優れているという有意な差は出ず、むしろ specmurt 分析を施さない方が認識率がよい傾向もある。これは RWC の方が複数の楽器を用いた楽曲が多いことから、specmurt 分析の精度が低下しているということが原因のひとつとして考えられる。

なお、MIREX2005 では、調認識の得点として、完全に一致した場合を 1 点、完全五度の場合 0.5 点、平行調の場合 0.3 点、同主調の場合 0.2 点という計算法を用いている。この方式に従うと、MIREX の Training Data Set に対しては提案手法 (open data) の最高点

* 使用した楽曲番号は 1-3, 6, 8, 10, 11, 19, 20, 22, 24-27, 29-32, 34, 36, 38, 41-43, 47-50 である。

表 2 調認識の正解率 (MIREX2005)

	open data				closed data			
	$\sigma^2 = 0.1$	$\sigma^2 = 0.05$	$\sigma^2 = 0.02$	histogram	$\sigma^2 = 0.1$	$\sigma^2 = 0.05$	$\sigma^2 = 0.02$	histogram
specmurt	86.5%	83.3%	85.4%	76.0%	91.7%	90.6%	90.6%	76.0%
wavelet	84.4%	80.2%	83.3%	71.9%	85.4%	89.6%	89.6%	71.9%

表 3 調認識の正解率 (RWC)

	open data				closed data			
	$\sigma^2 = 0.1$	$\sigma^2 = 0.05$	$\sigma^2 = 0.02$	histogram	$\sigma^2 = 0.1$	$\sigma^2 = 0.05$	$\sigma^2 = 0.02$	histogram
specmurt	92.2%	93.8%	93.8%	89.1%	96.9%	100%	96.9%	95.3%
wavelet	95.3%	95.3%	93.8%	89.1%	98.4%	96.9%	96.9%	100%

は 87.6 点 (96 点満点) であった。

6. まとめと展望

本稿では specmurt 分析と HMM を用いた調認識手法を述べた。和声を調認識の手がかりにするために短時間周波数解析で specmurt 分析を用いて chroma vector として音高特徴量を抽出し、chroma vector の時系列を調の背後にある和声状態から確率的に生起するとして、観測と調の関係を HMM でモデル化した。調認識は事後確率最大化問題と定式化され、HMM の最尤経路探索により調が求められた。提案モデルの性能評価実験より、調認識率は 83.3%–93.8% という結果を得た。

本報告では調認識可能であることを示したが、今後は特徴量やモデルを拡充させて、転調の検出、調認識の精度向上を目指したい。転調の検出は、提案手法の HMM で HMM 間の遷移を学習することにより転調が扱える可能性がある。また、特徴量の確率分布や、chroma vector 以外の音高特徴量の使用も検討したい。

謝辞 本研究では、東京大学工学部計数工学科 4 年生諸岡孟氏が RWC 研究用音楽データベースから作成した評価・学習用のデータセットを、また同研究室の亀岡弘和氏提供の wavelet 変換プログラムを使用しました。

参考文献

- 1) Pauws, S., "Musical key Extraction from Audio," 5th International Conference on Music Information Retrieval (ISMIR), pp. 96–99 (2004).
- 2) Gómez, E., Herrera, P., "Estimating The Tonality Of Polyphonic Audio Files: Cognitive Versus Machine Learning Modeling Strategies," 5th International Conference on Music Information Retrieval (ISMIR), pp. 92–95 (2004).
- 3) Chew, E., "The Spiral Array: An Algorithm for Determining Key Boundaries," In Proceedings of the 2nd International Conference

on Music and Artificial Intelligence (ICMAI), Vol.2445, pp. 18–31 (2002).

- 4) Richard A. Altes, "The Fourier-Mellin transform and mammalian hearing," Journal of Acoustical Society of America, 63(1), pp. 174–183 (1978).
- 5) 高橋佳吾, 西本卓也, 嵯峨山茂樹, "対数周波数逆畳み込みによる多重音の基本周波数解析," 情報処理学会研究報告, 2003-MUS-53-13, pp. 61–66 (2003).
- 6) 亀岡弘和, 西本卓也, 嵯峨山茂樹, "ハーモニック・クラスタリングによる多重音信号音高抽出における音源数とオクターブ位置推定," 情報処理学会研究報告, 2003-MUS-51, pp. 29–34 (2003).
- 7) 亀岡弘和, 齊藤翔一郎, 西本卓也, 嵯峨山茂樹, "Specmurt における最適共通調波構造パターンの反復推定による多声音楽信号の可視化と MIDI 変換," 情報処理学会研究報告, 2004-MUS-56-7, pp. 41–48 (2004).
- 8) 川上隆, "HMM を用いた旋律への自動和声付けの研究," 修士論文, 北陸先端科学技術大学院大学情報科学研究科 (2000).
- 9) M. A. Bartsch and G. H. Wakefield, "To Catch A Chorus: Using Chroma-Based Representations for Audio Thumbnailing," Proc. WAS-PAA'01, pp. 15–18 (2001).
- 10) 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: クラシック音楽データベースとジャズ音楽データベース," 情報処理学会研究報告, 2002-MUS-44-5, pp. 25–32, 2002.
- 11) http://www.music-ir.org/mirexwiki/index.php/MIREX_2005