Vol.2013-AVM-83 No.6
2013/12/5

ガウス過程回帰モデルを用いた多重音信号における楽器同定

鍵本哲宏†1　　笠井裕之†1

多重音信号解析において，各音がどの音でどの楽器に属しているかを認識することは，一般的に困難である．Specmurt 法では，同じ楽器カテゴリに属する音高は同じスペクトルパワー比（調波構造）を持つと仮定し，固定の調波構造を用いて解析を行なっている．しかしながら，一般的に調波構造は一定ではない．そこで本稿では，楽器音を事前に学習しガウス過程回帰でモデル化することで，可変的に調波構造を作り，スペクトル基底から最適な組み合わせを見つけることで，認識したい音高と属するカテゴリを特定する．提案方式と調波構造一定の場合を仮定した方式との認識精度の比較を行い，提案方式の有効性を示す．

キーワード　ガウス過程回帰，Specmurt，非負値行列因子分解

1. はじめに

近年，音楽の情報処理技術の飛躍的な向上により，音楽の創造や編集が容易にできるようになってきた．中でも自動採譜システムや音楽検索等を対象にした様々な音楽アプロケーションに応用が可能であることから，音楽音響信号から個々の基本周波数（音階）を推定する楽音解析の研究が特に注目を浴びている．しかし依然として，訓練を受けた人の力を借りなければ楽音の採譜は難しい．そこで本研究では，ある音楽を構成する楽器やピッチを自動的に認識することで，楽器ごとの楽譜を作成することを目標としている．一方，音楽データは時間に対する音の振幅の情報しか与えられない．しかし、人間の聴覚は多重音であってもそれぞれの音を認識することができる．多重音の解析は単音の解析に比べて困難であり，これまでも数多くの手法が試されてきた．中でもモノラル信号における音源分離は難しい問題とされている．こうしたモノラル信号における音源分離のうち，音声認識技術としてケプストラム法[1]やメル周波数ケプストラム(MFCC)[2]といった技術が提案された．これらの研究は単一楽器あるいは音声のピッチを推定するための技術である．さらに，単音だけでなく和音でもピッチを推定するために Specmurt 法が提案された[3]. Specmurt 法は多重音解析手法の一つであり，共通調波構造と呼ばれるモデルを自ら反復的に生成することで擬似的な楽器情報を得て，それを元に基本周波数を求めるというものであった．しかしこの手法は同じ楽器のもつ全ての音が同じ調波成分を持つと仮定したもので，実際は 1 音ごとに調波成分は異なる．よって多重音を単音ごとに分離するためには，調波成分を可変的に扱えるモデルを考える必要がある．更に最近では非負値行列因子分解（Non-negative Matrix Factorization）やスパースコーディングといった音楽の疎性を利用した解析も行なわれている[4]-[6],[9]. 我々は，1 楽器カテゴリが同じ確率スペクトル包絡を持つものと仮定し，調波成分のパワー比を可変的に扱うことが可能な，ガウス過程回帰モデルを用いたスペクトル基底に基づく多重音解析方式を提案してきた[7]．そこでは Specmurt 法と提案手法による音の認識率の比較によって，必要とする音のテンプレート数を削減した上での認識精度の向上を示し，単数の楽器による和音解析をおこなった．本稿では更に複数の楽器が含まれている場合の楽器の同定を行い評価した．

†1 電気通信大学大学院　　情報システム学研究科
The University of Electro-Communications
Graduate school of Information systems

2. 従来研究

2.1 Specmurt 法

多重音を解析する手法の一つとして Specmurt 法が提案されている．この手法によって，単一楽器によって演奏された和音を含む楽曲の基本周波数を求めることができる．

周波数成分間のパワー比が基本周波数に関わらず共通である調波構造を想定し，多重音スペクトルが基本周波数分布と共通調波構造の畳み込みで表すことができるものと仮定する．共通調波構造が既知であれば，多重音スペクトルとの逆畳み込みをすることで基本周波数分布を求めることができる．Specmurt 法の具体的な適用方法を以下に示す．

共通調波構造を$h(x)$，基本周波数を$u(x)$とする時，多重音スペクトル$v(x)$は式(1)で表される．

$$v(x) = h(x) * u(x) \tag{1}$$

ここで，共通調波構造とは全ての音の周波数精文館のパワー比が基本周波数に関わらず普遍的に一定である調波構造のことであり，基本周波数分布とは基本周波数がどの値にどれだけの成分を持つかを表したものである．

共通調波構造$h(x)$が既知であるなら，基本周波数分布$u(x)$は(x)と$h(x)$の逆畳み込みで求められ，式(2)で表される．

$$u(x) = h(x)^{-1} * v(x) \tag{2}$$

ただし，線形周波数スケールでは基本周波数が$\Delta\omega$変化するとn次の高調波周波数は$n\Delta\omega$変化してしまい，式(1)は成立しない．そこで，これら各関数を対数周波数スケールで扱うことにする．対数周波数スケールでは基本周波数がΔx

変化すると全ての高調波周波数もΔx変化し，式(1)が成立する．以上のように，対数周波数スケールで逆畳み込みをすることにより，基本周波数分布を求める方法は Specmurt 法と呼ばれる[3] .

3. 提案方式の概要

本研究では楽器カテゴリごとに分散を含めたスペクトル包絡を学習し，それを基にランダムにスペクトルを発生させ，確率的なアプローチにより音楽信号を解析する．具体的には，事前に楽音を楽器カテゴリごとに学習しておき，その学習データのテンプレートを用いて多重音楽解析を行なう．

3.1 非負値行列因子分解

シングルチャネルにおける混合楽器解析を実現するための手法として，非負値行列因子分解（Non-negative matrix factorization; NMF）を用いた手法が提案されている[4]-[6]．音響信号の振幅スペクトログラムを一つの行列とみなして NMF を実行するで，この行列を音源固有の情報（スペクトル）を表す基底行列と，基底ごとの時間的なゲイン変動を表すアクティビティ行列の積に分解することができる．得られた二つの行列から，音階情報，時間情報（発音，音価，強度）を求めることができ，これにより楽音イベントが推定可能である．

3.2 確率スペクトル包絡

本研究では「一つの楽器カテゴリには，一つの確立的なスペクトル包絡を一つもつ」と仮定する．確率スペクトル包絡は平均曲線と分散曲線によって表される確率的なスペクトル包絡である．包絡線の導入によって楽器音の調波構造は音高によって可変的に扱うことができる．この時，ある音高に対する調波構造は一定ではなく，分散値によって変化する．

3.3 提案方式の流れ

本研究では楽器カテゴリごとに分散を含めたスペクトル包絡を学習し，それを基にランダムにスペクトルを発生させ，確率的なアプローチにより音楽信号を解析する．本研究手法の処理手順を図 1 に示す．ここでは，確率スペクトル包絡を求める学習ステップ（4 節）と楽音解析を行なう解析ステップ（5 節）に分かれる．

学習ステップでは，予めいくつかの学習データを用意する．学習データは楽器カテゴリごとに用意され，それぞれの音階が順番に鳴らされた音響信号を用いる．次に学習データの振幅スペクトログラムに対して教師無しの NMF を実行する．音源数を既知として教師無しの NMF を実行し場合，理想的な基底行列とアクティビティ行列に極めて近い行列に分解することができる．ここで理想的な基底行列とはそれぞれのイベントのスペクトルを列要素とする行列を意味する．教師無し NMF の実行から求められた基底行列からスペクトルのピーク値を取り出し，ガウス過程回帰モデルに入力することで確率スペクトル包絡を楽器ごとに学習する．

一方，解析ステップでは，教師あり NMF を実行して適応度の高い基底の組み合わせを探す．

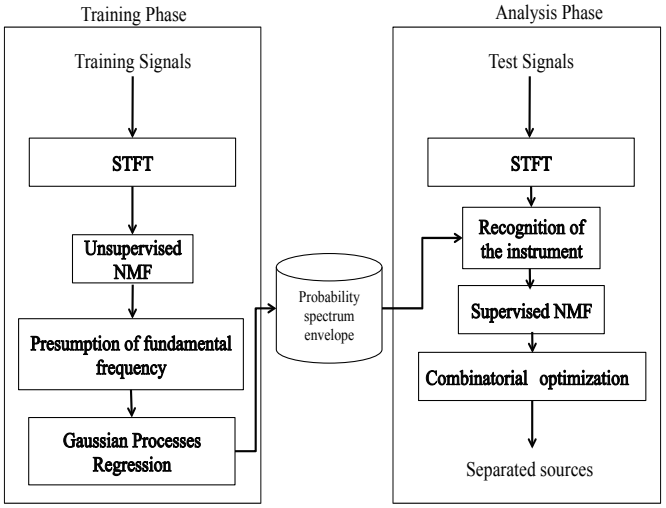

図 1　提案手法の処理手順
Figure 1　Configuration of style set.

4. 学習ステップ

4.1 教師無し NMF による基底スペクトルの抽出

確率スペクトルの学習は，楽器のカテゴリ毎に行なわれる．あるカテゴリに属する信号を，サンプリング周波数f_dで短時間フーリエ変換する．得られた振幅スペクトログラム$\mathbf{V}(\in \mathbb{R})$に対して，教師無しの NMF の実行は式(3)(4)で表される．

$$\mathbf{V} \approx \mathbf{WH} \tag{3}$$

$$\forall i,j,k, \mathbf{W}_{ij} \geq 0, \mathbf{H}_{jk} \geq 0 \tag{4}$$

以上のように$\mathbf{V}$を２つの非負行列の積で表すことができる．ここで，$\mathbf{W}(\in \mathbb{R})$は基底行列，$\mathbf{H}(\in \mathbb{R})$はアクティビティ行列である．

NMF では，二乗誤差基準により各行列要素毎更新を行なう．すなわち$D_{EUC}(\mathbf{V},\ \mathbf{WH}) = (\mathbf{V}-\mathbf{WH})^2$を最小化するような$\mathbf{W}$と$\mathbf{H}$を求める．各行列の更新式は式(5)(6)で表される．

$$\mathbf{W}_{ij} \leftarrow \mathbf{W}_{ij} \frac{(\mathbf{VH}^T)_{ij}}{(\mathbf{WHH}^T)_{ij}} \tag{5}$$

$$\mathbf{H}_{jk} \leftarrow \mathbf{H}_{jk} \frac{(\mathbf{W}^T\mathbf{H})_{jk}}{(\mathbf{W}^T\mathbf{WH})_{jk}} \tag{6}$$

4.2 ガウス過程回帰モデルによる確率スペクトル包絡

本稿では各楽器が共通のスペクトルの包絡線を有するものと仮定している．スペクトルのピーク点を入力として線形

予測をし，スペクトルの包絡を作成する．このスペクトル包絡を確率スペクトル包絡(Probability Spectrum Envelope;PSE)と呼ぶこととする．PSE の作成の仕方を説明する．

まず前節で得られたスペクトルの行列$\mathbf{W}$から，全てのピーク点を抽出する．$\mathbf{W} = [w_1(f)\, w_2(f) \cdots w_R(f)]$として，$r(= 1, \cdots, R)$番目の音源のスペクトル$w_r(f)$の基本周波数$f_r$を求める．倍音のインデックスを$h(= 1, \cdots H_r)$とすると，$w_r(f)$の$h$倍音目のピーク点は$(f_{hr}, y_{hr})$と表すことができる．ただし，$H_r$は$w_r(f)$のピーク数$f_{hr} = h \cdot f_r$，$y_{hr} = w_r(h \cdot f_r)$である．

ここで，$N = \sum_r H_r$個のピーク集合$(\boldsymbol{f}, \boldsymbol{h}) = \{(f_{hr}, y_{hr})\}_{h,r}$を入力としてガウス過程回帰モデルに入力する．尚，ガウス過程を回帰問題に適用するには観測される目標変数の値に含まれるノイズを考える必要がある[8]．ここでt_nは式(7)で表される．

$$t_n = y_n + \epsilon_n \tag{7}$$

ここで$y_n = y(f_n)$であり，またϵ_nはn番目の観測値に加えられるノイズで，各観測値に対して独立に決定される．ここではノイズもガウス分布に従うものとし，$p(t_n|y_n) = \mathcal{N}(t_n|y_n, \beta^{-1})$とする．ここで，$\beta$はノイズの精度を表す超パラメータである．ノイズはデータ点に対して独立に決まるため，$\mathbf{y} = (y_1, \cdots y_N)^T$が与えられた上での目標値$\mathbf{t} = (t_1, \cdots t_N)^T$の同時分布は$p(\mathbf{t}|\mathbf{y}) = \mathcal{N}(\mathbf{t}|\mathbf{y}, \beta^{-1}\mathbf{I_N})$の等方的なガウス分布に従う．ここで$\mathbf{I_N}$は$N \times N$の単位行列とする．ガウス過程の定義より，周辺分布$p(\mathbf{y})$は平均が 0 で共分散がグラム行列$\mathbf{K}$で与えられるガウス分布となる．

入力値$f_1, \cdots f_N$で条件付けられたときの$\mathbf{t}$の周辺分布$p(\mathbf{t})$は式(8)で表される．

$$p(\mathbf{t}) = \int p(\mathbf{t}|\mathbf{y})p(\mathbf{y})\, d\mathbf{y} = \mathcal{N}(\mathbf{t}|\mathbf{0}, \mathbf{C}) \tag{8}$$

ここで$\mathcal{N}(\mu, \sigma^2)$は平均$\mu$，分散$\sigma^2$の正規分布を表す．また，共分散行列$\mathbf{C}$の要素は式(9)で表される．

$$C(f_n, f_m) = k(f_n, f_m) + \beta^{-1}\boldsymbol{\delta_{nm}} \tag{9}$$

を持つ．$k(f_n, f_m)$はカーネル関数で，ガウス過程回帰に用いるカーネル関数は，式(10)で表される．

$$k(f_n, f_m) = \theta_0 \exp\left\{-\frac{\theta_1}{2}\|f_n - f_m\|^2\right\} + \theta_2 + \theta_3 f_n^T f_m \tag{10}$$

超パラメータ$\theta_0, \ldots, \theta_3$の値の学習は尤度関数$p(\mathbf{t}|\boldsymbol{\theta})$の評価に基づいており，勾配法によって求めることができる．

回帰において訓練データの集合として入力が与えられた時に，新しい入力に対する目標変数の値を$\boldsymbol{f}_1, \ldots, \boldsymbol{f}_N$と，対応する$\mathbf{t_N} = (t_1, \cdots t_N)^T$が与えられているときに，新しい入力ベクトル$\boldsymbol{f}_{N+1}$に対する目標変数$\mathbf{t_{N+1}}$を予測したいものとする．

ここで，条件付き分布$p(t_{N+1}|\mathbf{t})$を求めることを考える．$t_1, \cdots t_{N+1}$の同時分布は式(9)より式(11)で表される．.

$$p(\mathbf{t}_{N+1}) = \mathcal{N}(\mathbf{t}_{N+1}|\mathbf{0}, \mathbf{C}_{N+1}) \tag{11}$$

$\mathbf{C}_{N+1}$は，$(N+1)\times(N+1)$の共分散行列である．この同時分布はガウス分布であるため，条件付きガウス分布が得られる．これを行なうような共分散行列の分割は式(12)で表される．

$$\mathbf{C}_{N+1} = \begin{pmatrix} \mathbf{C}_N & \mathbf{k} \\ \mathbf{k}^T & c \end{pmatrix} \tag{12}$$

$\mathbf{C}_N$は$n, m = 1, \ldots, N$に対する要素が式(9)であるような$N \times N$の共分散行列である．$\mathbf{k}$は，要素$k(f_n, f_{N+1})(n = 1, \ldots, N)$を持つベクトルである．またスカラー$c = k(f_{N+1}, f_{N+1}) + \beta^{-1}$とする．条件付き確率分布$p(t_{N+1}|\mathbf{t})$は平均曲線$\mu_f$と分散曲線$\sigma_f$を持つようなガウス分布になる．平均曲線$\mu_f$と分散曲線$\sigma_f$は式(13)(14)で表される．

$$\mu_f = \mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{t} \tag{13}$$

$$\sigma^2 = c - \mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{t} \tag{14}$$

確率スペクトル包絡$E(f; \mu_f, {\sigma^2}_f)$をデータベースに保存する．

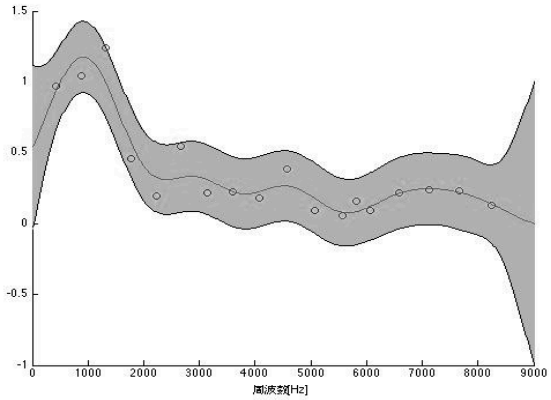

図 2　確率スペクトル包絡の例．丸は入力点，真ん中の線は平均曲線，その上下の曲線は平均曲線±分散曲線を表す．
Figure 2　The example of probability spectrum envelope. A circle is an input point, the line of middle is an average curve, and the curve of the upper and lower sides expresses an average curvilinear ± dispersion curve.

5. 解析ステップ

5.1. 確率スペクトル包絡に基づくスペクトルのランダム生成

確率スペクトル包絡$E(f, y; \mu_f, {\sigma^2}_f)$に基づくスペクトル包絡$e(f)$はランダムに生成される（図 3）．$e(f)$は式(15)で表される．

$$e(f) \approx \mathcal{N}\left(\mu_f, {\sigma^2}_f\right) \tag{15}$$

このスペクトル包絡にそった基本周波数νのスペクトル$p(f)$は式(16)で表される．

$$p(f) \approx \max\left(e(f), 0\right) \cdot \Psi\left(f; \nu\right) \tag{16}$$

最大値をとっているのはスペクトルが非負値をとらない制約によるものである．$\Psi(f;\nu)$は基本周波数νの櫛形フィルタ（図３）であり，式(17)で表される．

$$\Psi(f;\nu) = \sum_l \exp\left\{-\frac{(f-\nu\cdot l)}{2\lambda_0^2}\right\} \tag{17}$$

ここでlはコンポーネントを表すインデックス，λ_0は各コンポーネントの尖度を決定するハイパーパラメータであり，実験的に定めることができる．

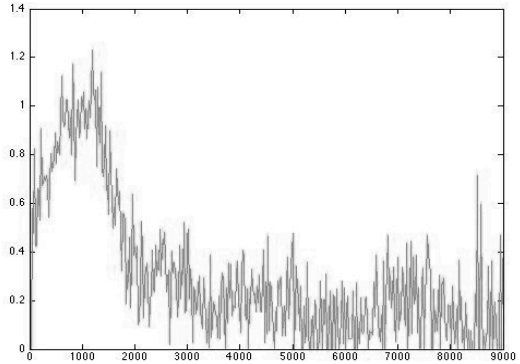

図 3　ガウス分布に従う確率スペクトル包絡の例
Figure 3　The example of Gaussian distributed the probability spectral envelope.

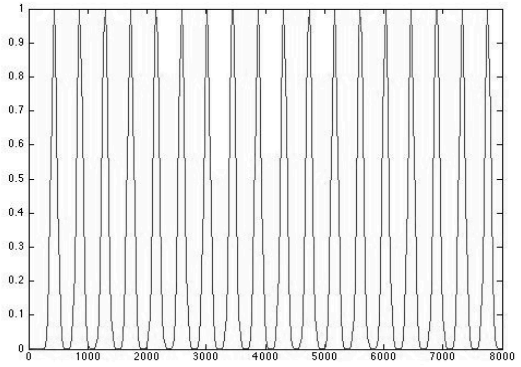

図 4　櫛形フィルタ例
Figure 4　The example of comb filter

5.2. スペクトル包絡による楽器の同定と分解

多重音解析をする際，学習した確率スペクトル包絡から生成される基底スペクトルの内，観測データのスペクトルに最も近づくような基底スペクトルの組み合わせを求める．この処理は楽音全ての総当たり的なアプローチによって行なわれる．多重音に含まれている楽器カテゴリが分かれば総当たりの行程数を削減する事ができる．

5.2.1. 楽器同定による楽音認識

判別操作の流れを説明する．まず観測データスペクトルの内，最も低い周波数を基本周波数とするスペクトルピーク点集合$(\boldsymbol{f}, \boldsymbol{h}) = \{(f_{hr}, y_{hr})\}_{h,r}$を選択する．次に選択されたピーク点セットベクトルが学習されている分散曲線内に属しているかを判定する．属しているピーク点の多い最も多い楽器カテゴリを決定する．複数の楽器カテゴリに属した場合，学習平均曲線と基底のピーク点ベクトルと比較し，ノルムが最小となる楽器カテゴリを選択する．楽器カテゴリが決定したら観測スペクトルから選択されたピーク点分を引く．次に最も低い周波数ピーク点を基本周波数とするスペクトルピーク点を抽出し，同様の操作を繰り返し行なう．選択されたスペクトルの組み合わせを基底スペクトルの集合$\tilde{\mathbf{W}}$に記憶する．

5.2.2. 教師あり NMF

学習ステップで作成された基底スペクトルの集合$\tilde{\mathbf{W}}$を教師データとして振幅スペクトログラム$\mathbf{V}$に NMF を適用し，アクティビティ行列$\tilde{\mathbf{H}}$を逆算する．$\tilde{\mathbf{H}}$は式(20)で表される．

$$\tilde{\mathbf{H}} = (\tilde{\mathbf{W}}^{\boldsymbol{T}}\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}^{\boldsymbol{T}}\mathbf{V} \tag{20}$$

$\tilde{\mathbf{W}}$が正しい場合，$\tilde{\mathbf{H}}$は非負値の要素を持つ．逆に正しくない場合，$\tilde{\mathbf{H}}$の要素は負値が多くなる．そして距離$D_{EUC}(\mathbf{V},\ \tilde{\mathbf{W}}\tilde{\mathbf{H}})$が小さいものを選択することで，解析データに含まれている音とアクティビティが分かる．従来研究では音の数が既知として，$\tilde{\mathbf{H}}$の負値成分が最も少なくなる基

底の組み合わせから，音の識別を行なった．しかし本稿では音源数や楽器数を先に判定する事でその過程を飛ばす事ができる．ただし，$\tilde{\mathbf{H}}$の負値成分数がしきい値を超えた場合は$\tilde{\mathbf{W}}$を総当たりのアプローチで求めた．総当たりの場合，N個の基底の最大の組み合わせを考えると楽器数をLとして最大で$\Sigma_{L\,N}C_i$となる．一方楽器の同定を行なう事で全ての基底探査を行なう必要がなくなる．

6. 評価実験

調波構造を可変的に扱うことで認識率が向上することを示すために，単一楽器による和音の解析を実行し，性能を評価する．まず 3 和音の多重音解析を実行し，和音の組み合わせによる認識率の変化と，各種法ごとの認識率の精度を比較する．次に，和音の数を変化させたときの認識率の変化を示し，Specmurt 法と従来手法，そして提案手法の性能を評価する．最後に複数の楽器が含まれた楽音に対して楽器の同定を行ない評価する．

6.1. 単一楽器の 3 和音の多重音解

midi のピアノ音源を用いて確率スペクトル包絡から教師データとしてのスペクトルを作成した．今回は C1 から B6 までの 12 音階 6 オクターブの音を学習させた．テストデータは Apple 社の GarageBand を用いて作成した．また使用音源は標準装備されていたソフトウェア音源を使用した．なお多重音楽データは GarageBand を使用して作成している．音源データの再生時間はすべて 3 秒である．

実験では，GarageBand で作成した Piano の和音に対し，従来研究の Specmurt 法と調波構造が既知である場合，そして本研究手法の 3 方式の認識率を比較した．具体的には，Specmurt 法では 1/f 特性に則った固定された調波構造を用いる[1]．既知である場合とは事前に単音のスペクトルを確認しておき，フレームごとに調波構造を記録しておく．そしてパワー比を正規化してそろえ，最も出現頻度の高いスペクトルを既知の調波構造とした．そして 5.2 と同じ手順で適応度の最も高い基底の組み合わせを選択している．

まずは 3 和音（D3F3A3, A3F3A4, C#3D3D#3, A2A3A4 ）に対する解析を行い，認識率を比較した．
4 種類の和音を解析対象にした理由は，基本周波数と倍音が重なる多重音スペクトルに対しての各手法の有効性を比較するためである．各和音の特徴として，D3F3A3 は 1 オクターブ内の楽音で構成される和音で，倍音と基本周波数が被りにくい．A3F3A4 は 1 オクターブ上の楽音を含む和音で，倍音と基本周波数が重なる．C#3D3D#3 は 1 オクターブ以内だが半音ずつずれており，スペクトルのピークの間隔が狭い．A2A3A4 はオクターブ違いの音高のみで構成される和音で，倍音と基本周波数の重なりが多い．

図 5 に，3 和音に対する各手法の認識率の結果を示す．図 4 において，縦軸は認識率(%)を表している．“初期 Specmurt 法”とは従来研究[1]で提案されている手法である．調波構造がある固定された構造を持つものと仮定され，Specmurt 法に適用している．“調波構造既知”は各音のある時間の調波構造が既知である時の認識率である．

この結果より従来研究の Specmurt 法よりも認識率が高いことが分かった．音の調波構造が既知であるものと比較して同等の性能を示した．これはそれぞれの音色ごとに調波構造を可変的に扱うことで実際の調波構造に近い概形を表現することができるためと考えられる．さらにピッチが離れた和音とピッチの近い和音の場合を比較すると，前者の向上が大きいことが分かる．これはピッチの変化が大きい場合，調波構造の差異が大きくなるため，Specmurt 法では認識率が下がっていると考えられる．一方で，ピッチの離れた音であっても表現できる提案手法は，和音を構成するピッチの高さに依存しないため，Specmurt 法との相対的な差異が現れやすいと考えられる．以上より，確率スペクトル包絡はテンプレートを 1 つだけ用意すれば良いことから，確率スペクトル包絡は教師データとして有効であると言える．

6.2. 和音数の変化に伴う認識率の評価

単一楽器で構成される音楽データの和音の数を変化させたときの認識率の挙動を調べた．2 和音, 3 和音, 4 和音は，C1 から B6 までの 12 音階 6 オクターブの音のうちからランダムに選択して作成した．実験に用いたデータは同様に Apple 社の GarageBand の内蔵ソフト音源を使用した．音色はピアノである．

各実験の認識率はフレーム単位での全音符数を分母としたときの，正しく基本周波数が推定されていた割合を表す．

図 6 は単音，2 和音，3 和音，4 和音の単楽器多重音解析に対する認識率を示している．Specmurt 法と従来手法，提案方式を比較している．単音の場合，2 音和音の場合はどちらも高い認識率を示しているが，3 和音以上の時急激に認識率が下がっている．Specmurt 法の場合，和音数が増えると和音の組み合わせバリエーションが増えるだけでなく，音の高低差が大きく出る場合があり，その場合に認識率が大きく下がっていると考えられる．一方で，PSE を用いた手法の場合，高低差による調波構造の変化に対して比較的に頑強であると言える．更に従来 PSE と提案方式を比較した場合，ほぼ同様の性能を示している事から提案方式手法でも基底スペクトルの集合$\tilde{\mathbf{W}}$を表現できている事が分かる．

6.3. 楽器同定評価

単音，２和音，３和音，４和音の音楽データを用意し，それぞれの音源に含まれる楽器の同定を行なった．用いた音源はピアノ, ピアノとギター, ピアノとギターとフルート，ピアノとギターとフルートとサックスの和音である．各楽器は常に 1 音以上鳴っており音源の長さは 3 秒である．フ

レーム数はピアノ，ギター，フルート，サックスそれぞれから学習した確率スペクトル包絡と，観測データのスペクトルから作成した確率スペクトル包絡を比較する．学習した確率スペクトル包絡を組み合わせ，観測データのスペクトルの概形に近づける．

図 7 は音源中で検出された楽器カテゴリの回数と実際に同定が成功した回数を示している．

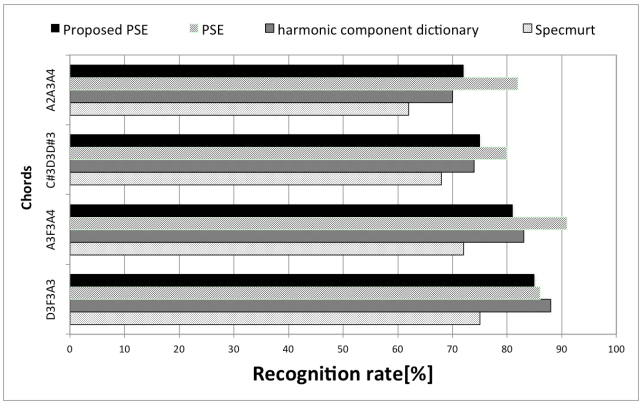

図 5　各手法による 3 和音の認識率の関係

Figure.5　The relation of the recognition rate of three chords by each technique

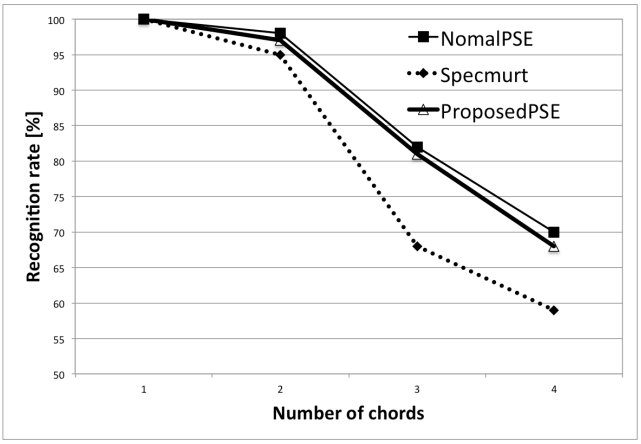

図 6　和音数を変化させたときの認識率の変化

Figure.6　Change of a recognition rate when changing the number of chords

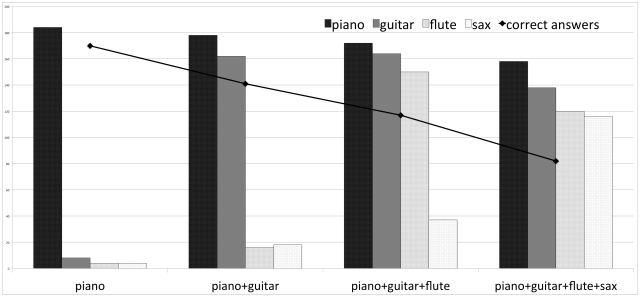

図 7　多重楽器音の検出数

Figure.7　The number of detection of multiplex musical instrument sound

7. まとめ

同一の楽器の音高の調波構造は一定ではない．そこで確率スペクトル包絡を同一楽器で共通のモノとして扱うことで，調波構造の揺らぎを表現することができ，また一つの確率スペクトル包絡を用いれば全ての音高を表現できる．これは従来の多重音解析手法で膨大なテンプレート必要とする問題を解決すると考えられる．そしてランダムにスペクトルを生成し，最適なものを見つけていくことができる．学習された確率スペクトル包絡を今後の課題として，より多くの教師データを用いて確率スペクトル包絡を導出し，また教師あり NMF によって最適な基底の組み合わせを探索するプログラムを実装する予定である．また楽器の同定については本来観測データに含まれていない楽器が誤検知されてしまっていた．これは確率スペクトル包絡そのものが楽器特徴を十分表現できていないことが原因と考えられる．今後は確率スペクトル包絡以外にも同定判別要素を追加することで，検知精度の向上を目指したい．

謝辞　笠井准教授をはじめとする作成にご協力頂いた方々に，謹んで感謝の意を表する．

参考文献

1)　B. P. Bogert,M. J. R. Healy, and J. W. Tukey, “The quefrency alanysis of time series for echoes:cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking” in Proceedings of the Symposium on Time Series Analysis (M.Rosenblatt,Ed) ”Chapter15, 209-243.New York:Wiley, (1963).

2)　Kiyoharu Aizawa, Yuichi Nakamura, and Shin’ichi Satoh, “HMM-based audio keyword generation,” Advances in Multimedia Information Processing - PCM 2004: 5th Pacific Rim Conference on Multimedia. Springer, (2004).

3)　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, pp.61-66, Dec. (2003).

4)　P. Smaragdis, and J.C. Brown, “Non-negative matrix factorization for polyphonic music transcription,” In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp.177–180, 2003.

5)　T. Virtanen, “Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria,” IEEE Transactions on Audio, Speech, and Language Processing, vol.15, no.3, pp.1066-1074, (2007).

6)　O. Dikmen, and A.T. Cemgil, “Unsupervised single-channel source separation using Bayesian nmf,” IEEE Workshop on Applications of Signal Processing to Audio and Acoustics WASPAA’09, pp.93-96, 2009.

7)　鍵本哲宏，笠井裕之，“ガウス過程回帰モデルを用いたスペクトル基底に基づく多重音解析方式の提案,” 研究報告マルチメディア通信と分散処理（DPS）, ,no.2,pp.1–6,(2013).

8)　C.M. Bishop：Pattern Recognition and Machine Learning, Springer,（2006).

9)　Boris Mailhe□, Daniele Barchiesi and Mark D. Plumbley,” INK-SVD: LEARNING INCOHERENT DICTIONARIES FOR SPARSE REPRESENTATIONS”, ICASSP,.pp3573-pp.3576,(2012)