Vol.2015-MUS-109 No.6
2015/11/7

楽器音辞書を用いた楽曲の非負値疎分解による楽譜の生成

野村　亮[†1,a)]　栗田　多喜夫[†1,b)]　相原　玲二[†1,c)]

概要：楽曲に使用されている楽器音の計算機による解析は，音楽情報処理において重要な研究課題である．この論文では楽器音辞書を用いた，音響信号の自己相関関数を特徴量とし，楽曲の非負値疎分解による楽譜の生成アルゴリズムを提案する．楽曲は正の値により重みづけをされた複数の楽音の足し合わせにより構成されると考えることが出来る．足し合わせる楽音の種類は膨大な数が存在するが，楽曲のある瞬間に同時に使用される楽音はごく少数であり，各楽音の重みの多くは 0 である．すなわち，疎なベクトルとして楽音の重みは表現可能である．多くの楽音はあらかじめサンプルを入手する事が可能である．サンプルから事前に楽音のモデルを学習し，それらのモデルを集めたものを辞書として使用する．この時，楽曲中の楽音の解析は，楽曲の非負値疎分解により実現できる．本手法の有効性を二種類の楽器で作成した楽曲の楽音を推定する実験により確認した．

キーワード：non-negative sparse coding，事前学習，音源同定，主成分分析，自己相関関数

The Generation of Music Score by Non-negative　Sparse Decomposition Musicalsignals Using Pre-trained Dictionary

Abstract: The automatic analysis of identification of sound source in musical signal is important task for music information analysis. This paper proposes the generation of music score algorithms by non-negative sparse decomposition musicalsignals using pre-trained dictionary. We can consider music are composed as cojmposition of several musical insturment sound signals summation. There are enourmous musical insturment sounds. Only a few musical instrument sounds appear at the same time. Thus, the coefficient vector of musical insuturment sound signals must be positive and sparse. We can gather many samples of musical insturment sounds easily and make dictionary from these samples beforehand. Then, we estimate the coefficents of instruments sounds in musical signals using non-negative sparse coding. We confirm that the proposed generative method is applicable.

Keywords: non-negative sparse coding, pretraining, sound source identification, principle compornent analysis, auto correlation function

1. 序論

楽曲に使用されている楽器音の計算機による解析は，音楽情報処理において重要な研究課題である．計算機の性能向上，インターネット通信技術の発展と普及により，音楽の専門家ではない一般ユーザーでも，手軽に作曲や作品の発信，大量の楽曲にアクセスする事が可能になっている．現在，アクセスする事が出来る楽曲は大量に存在し，今後も止まることなく楽曲が増え続けていく．ユーザーが効率よく莫大な楽曲の中から好みに合った楽曲を見つけ出すニーズが高まっている．楽曲のメロディーやハーモニーといった音楽情景情報を楽曲の推薦や検索に活用する事が可能であれば，ユーザーが効率よく音楽コンテンツを楽しめるようになる．この論文では楽器音辞書を用いた，音響信号の自己相関関数を特徴量とし，楽曲の非負値疎分解による楽譜の生成アルゴリズムを提案する．

楽曲は正の値により重みづけをされた複数の楽音の足し合わせにより構成されると考えることが出来る．足し合わせる楽音の種類は膨大な数が存在するが，楽曲のある瞬間

[1] 情報処理学会
IPSJ, Chiyoda, Tokyo 101–0062, Japan
[†1] 現在，広島大学
Presently with Hiroshima University
[a)] m140951@hiroshima-u.ac.jp
[b)] tkurita@hiroshima-u.ac.jp
[c)] ray@hiroshima-u.ac.jp

ⓒ 2015 Information Processing Society of Japan

に同時に使用される楽音はごく少数であり，各楽音の重みの多くは 0 である．すなわち，疎なベクトルとして楽音の重みは表現可能であり，音楽情報処理においてもスパースコーディングが注目を集めている．S. A. Abdallah et al. はデータ駆動型処理とスパースコーディングを用いて音程の同定を行っている [1]．また，得られる結果がスパースになりやすい性質を持つ Non-negative Matrix Factorization(NMF) も音楽情報処理において注目を集めている．NMF は全ての要素が正である行列から，基底行列とアクティベーション行列の二つの非負値行列を推定するアルゴリズムである．NMF は D.D.Lee at el. によって画像処理にも用いられ [2]，文書解析 [3] や，音楽情報処理にも広く応用されている [4] [5] [6] [7]．NMF は拡張性も高く，音楽情報処理において，板倉斎藤距離基準 NMF[8]，複素 NMF[9]，無限半正定値テンソル分解 [10]，NMF のスパース性をより強めたスパース NMF[11] 等が提案されている．

多くの楽音はあらかじめサンプルを入手する事が可能であるため，基底行列とその重み行列であるアクティベーション行列を同時に推定する事は必ずしも必要ではない．そこで，我々はサンプルから事前に楽音のモデルを学習する事は規定と重みを同時に推定する事よりも比較的容易であり，学習した楽音モデルを集めたものを楽音辞書として使用する事を提案する．この時，楽曲中の楽音の解析は，楽曲の非負値疎分解により実現できる．本手法の有効性を二種類の楽器で作成した楽曲の楽音を推定する実験により確認した．

2. 楽音辞書学習

NMF で得られる基底は，必ずしも楽器音の物理的な特徴を表現しているとは言えない．そこで我々の手法では，単独で鳴らされている単音の楽器音信号から特徴量を抽出し，楽器音モデルを作成する．楽曲で使用されている楽器音の多くは，事前に入手し集める事が容易である．集めた楽音から特徴量を抽出し，機械学習を行うことにより，楽音を良く表現したモデルの作成を行い，楽曲の非負値疎分解を行う際の楽音辞書として使用する．

我々は楽器音信号から特徴量を抽出する手法として自己相関関数 (ACF) を用いた．解析する楽曲音響信号は，重みづけされた複数の楽器音の線形和として表せるので，音響信号を変換した先でも加法性が満たされる手法が望ましい．自己相関関数には，互いに独立している信号による混合信号の自己相関関数は，各信号の自己相関関数の線形和として表すことが出来るという性質があり，この問題で使用するのに適している．また，自己相関関数とパワースペクトル密度関数は互いにフーリエ変換対であるという関係がある．すなわち，信号の特徴量として，自己相関関数を使用する事と，パワースペクトル密度関数を使用する事とは，本質的に同一である．以上より，我々は信号から直接求める事が可能である自己相関関数を特徴量として使用する事とした．

信号から得られる自己相関関数は高次元なベクトルとして得られる．高次元なベクトルをより扱いやすい低次元なベクトルにするため，主成分分析 (PCA) により自己相関関数の次元削減を行う．主成分分析は，サンプルが持つ説明変数間の相関を無くし，低次元の説明変数によって元のサンプルを表現しなおす手法であり，多くの分野で利用されている．得られた自己相関関数に主成分分析を行うことで，情報損失を抑えつつ次元を圧縮する事が可能である．

本項では，単独で演奏される楽器音から特徴量を抽出し，次元圧縮を行う手法について解説する．

2.1 自己相関関数

自己相関関数は，自信号との相互相関関数として定義される．時間方向にずれた自分自身との類似性を計算して得られる関数であり，基本的な信号処理手法である．時間 $t=0$ から $t=T-1$ までの音響信号を $\{x(t)|t=0,....,T-1\}$ とする．$\{x(t)|t=0,....,T-1\}$ における $x(t)$ の平均を μ_x とし，$x(t)$ の分散を σ_x とすると，信号 $\{x(t)|t=0,....,T-1\}$ の自己相関関数 $c(\tau)$ は，

$$c(\tau)=\frac{E\big[\big(x(t+\tau)-\mu_x\big)\big(x(t)-\mu_x\big)\big]}{{\sigma_x}^2} \tag{1}$$

で得られる．二つの信号 $x_1(t)$，$x_2(t)$ をそれぞれ w_1，w_2 に重みづけして混合した信号を $z(t)=w_1x_1(t)+w_2x_1(t)$ とする．$x_1(t)$，$x_2(t)$ が互いに独立であると仮定すると，それらの混合信号 $z(t)$ の自己相関関数 $c_z(\tau)$ は，

$$c_z(\tau)=\frac{E\big[w_1^2c_1(\tau)+w_2^2c_2(\tau)\big]}{\sigma_x^2} \tag{2}$$

となる．$c_1(\tau)$， $c_2(\tau)$ はそれぞれ， $x_1(t)$ ， $x_2(t)$ の自己相関関数である．信号自身の分散で正規化された自己相関関数は， $[-1,1]$ の範囲の値を取る．本研究では後述する 使用するデータが全て正であることを前提に提案された Non-negative Sparse Coding（NNSC) のアルゴリズムを使用するために，非負の値を取らないように自己相関関数に以下の様な処理を行い，全ての値が正の値を取るように正規化を行った．

$$y(\tau)=\frac{c(\tau)+1}{2} \tag{3}$$

この処理により正規化された自己相関関数 $y(\tau)$ は $[0,1]$ の値に収まり，NNSC アルゴリズムを適用できる様になる．こうして得られたベクトル $\mathbf{y}=\big(y(0),....,y(L)\big)^T$ を本項では non-negative normalized auto correlation vector(nnACV) と呼び，特徴ベクトルとして扱う．

2.2 辞書学習

単独で演奏される楽器音の録音サンプルを集める事は難

しいことではない．学習の必要がある楽器の音を単独に録音することにより簡単に集めることが可能である．我々は録音された楽器音から抽出した nnACV から，非負値疎分解を行う為の楽音辞書を学習する．使用する学習手法として，サンプル間平均と主成分分析を用いて楽音辞書を作成した．

サンプル群の平均は各サンプルとの誤差の総和を最小にする，機械学習の基本的な手法である．N 個の異なる楽器音があり，それぞれから M 個の自己相関関数が得られるとする．録音された音から得られる nnACV を $\{\mathbf{y}_i^{(j)}|i=1,....,M;j=1,....,N\}$ とする．i はサンプル数であり，j は楽音のインデックスである．

この時，得られた nnACV のサンプル間平均ベクトルは，

$$\bar{\mathbf{y}}^{(j)} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{y}_i^{(j)} \tag{4}$$

となり，j 番目の楽音を代表する特徴ベクトルと呼ぶ．楽音全ての nnACV のサンプル間平均ベクトルを集めたものは，

$$Y = \left(\bar{\mathbf{y}}^{(1)},\ldots,\bar{\mathbf{y}}^{(N)}\right) \tag{5}$$

となり，これを平均楽音辞書とする．得られた平均楽音辞書に主成分分析を適用し，平均楽音辞書の次元削減を行う．主成分分析は，サンプルの説明変数間の相関を無くす様な直交変換基底 V を学習する．直交変換基底 V は集めたサンプルの説明変数に関する共分散行列の固有ベクトルを求めることによって得られる．

平均ベクトル楽音辞書 Y の共分散行列 Σ_Y とすると，直行変換基底 V は，

$$\Sigma_Y = V\Lambda V^T \tag{6}$$

から求めることが出来る．Λ は 固有値 $\{\lambda_k|k=1,...,N\}$ を対角要素に持つ固有値行列である．固有値が大きい順に対応する固有ベクトルを l 個まで取り出した直交変換行列 V_l を Y に作用させる事により次元削減を行う．

$$D = V_l^T Y \tag{7}$$

学習された楽音辞書 D は負の値も含まれる．得られた直行変換行列 V_l を，同様に解析する楽曲信号から得られた特徴ベクトルにも適応し，楽音辞書 D との最小値を比較し，両者の要素すべてから最小値を引くことにより，正に正規化する事が出来，後に説明する非負値疎分解を適用する事が可能となる．

3. 非負値疎分解による楽譜の生成

この項では楽曲音響信号に含まれる楽音を推定する手法について説明する．楽曲は，複数の楽音が正の値の重みで足し合わされて構成されていると考える事ができる．楽曲に現れる可能性のある楽音は膨大な数になるが，楽曲のある瞬間に同時に使用される楽音は極小数である．即ち，楽音の重みベクトルはほとんどの要素が 0 の疎（スパース）表現となる．事前に学習した楽音辞書の線形結合で楽曲の近似を行えば，正で疎な重みベクトルを推定する事が可能である．近似する手法として，最小二乗法に L1 正則化を適用させたものを使用する．線形結合を行う楽音辞書は正領域への正規化を行うので，重みベクトルの推定に Non-negative Sparse Coding (NNSC)　の基底学習ステップを除いだアルゴリズムを利用した．Non-negativeSparse Coding (NNSC) とは，Non-negative Matrix Factorization (NMF) で得られる結果のスパース性を高めるように拡張したアルゴリズムである．NMF は全ての値が正である行列を，二つの正値をもつ行列の積に近似するアルゴリズムである．得られる二つの行列の内一つは，基底行列と呼び，もう一つはアクティベーション行列と呼ばれる．NMF には非負の値が含まれないため，基底の足し合わせで元データの近似を行う．その性質上，引き算が存在せず，得られるアクティベーションはスパースになりやすい．P.O.Hoyer はこの性質をより強くするために二つの行列を計算する際に L1 正則化を行い，より強いスパース性を NMF に与える拡張を施した [11].

3.1 非負値疎分解

楽音辞書内の j 番目の基底楽音モデルを $\{\mathbf{d}^{(j)}|j=1,....,N\}$ とする．音響音楽信号の特徴ベクトル $\mathbf{x}$ を基底楽音モデルで近似すると考えると，以下の様な式になる．

$$\mathbf{x} \approx \sum_{j=1}^{N} w_j\mathbf{d}^{(j)} = D\mathbf{w} \tag{8}$$

$\mathbf{w} = (w_1,....,w_N)^T$ は非負のベクトルであり，対応する基底楽音モデルの存在確立を表す．楽音辞書を $\mathbf{w}$ で重み付けをしたものと，$\mathbf{x}$ との差を二乗したものを $E(\mathbf{w})$ とすると，

$$E(\mathbf{w}) = \|\mathbf{x} - D\mathbf{w}\| \tag{9}$$

となる．$E(\mathbf{w})$ 最小にするような $\mathbf{w}$ を求めればよい．これに L1 正則化を適用した式は，

$$Q(\mathbf{w}) = E(\mathbf{w}) + \lambda\|\mathbf{w}\| \tag{10}$$

である．この時，以下の様な制約条件を課すと，

$$\min_{\mathbf{w}} Q(\mathbf{w}) \quad \text{with constraints} \quad \mathbf{x},\mathbf{w},D \geq 0 \tag{11}$$

という問題に帰着する．この条件下での最小二乗は結果的に，P.O.Hoyer が提案した Non-negative Sparse Coding (NNSC) の基底行列を固定した場合の推定と同一になる．

そこで，本手法では NNSC の基底行列を固定し，アクティベーション行列だけを計算するアルゴリズムを使用し

Data: $\mathbf{x}$ and D
Result: $\mathbf{w}$
Initialize $\mathbf{w}^0$ to a random positive vector. Set $h = 0$.;
repeat
  $\mathbf{w}^{h+1} = \mathbf{w}^h .* (D^T\mathbf{x}./(D^T D\mathbf{w}^h + \lambda))$;
  Increment h;
until $\mathbf{w}^h$ *converges*;

Algorithm 1: Algorithm to estimate the non-negative sparse coefficients $\mathbf{w}$.

た．NNSC では，基底行列もしくは，アクティベーション行列，またはその両方に L1 正則化を行い，NMF のスパース性をさらに高めたアルゴリズムである．基底行列を事前に学習を行った楽音辞書とすると，基底を学習する必要がなくなる．NNSC は NMF と同様に基底行列とアクティベーション行列を誤差が収束するまで更新を繰り返す．更新式は以下である．

$$\mathbf{w}^{h+1} = \mathbf{w}^h .* (D^T\mathbf{x}./(D^T D\mathbf{w}^h + \lambda)), \tag{12}$$

h は更新回数である．$.*$ と $./$ という演算記号は，要素毎に演算を行う事を示す．このアルゴリズムを 1 に記す．

4. 実験

事前学習を行った楽音辞書での非負値疎分解による楽譜の生成が可能かどうかを検証した．辞書学習に使用するサンプル，楽譜を生成したい音源ともに KOMPLETE10 に同封されている KONTAKT5 で制作した．

4.1 学習用データセット

我々は Native Instrument から発売されている KONTAKT5 を音源として，学習用データと実験用音源を作成した．学習に適した形で作成した MIDI 信号を KONTAKT5 で出力したものを，WAVE 形式で録音した．我々はアルトサックスとコントラバスから各々 3 つの音の強さで鳴らされる 12 種の音階を使用した．教師信号を $\{\mathbf{s}_l^{(j)}|j = 1, ..., 24|l = 1, ..., 3\}$ として定義た．j は音階を，l は音の強さを表すタグである．アルトサックスの音域は，$C4$ から $B4$ でり，コントラバスの音域は， $G1$ から $F\flat 2$ である．各楽器 12 音であり，合計で 24 音である．録音のサンプリングレートは 44100 [Hz] である．アルトサックス，コントラバスともに音量の異なる 3 タイプの音を用意した．各音の音源は実時間で 4 秒とした．各音の最初と最後は 0.5 秒の無音区間である．サンプリングレートを Fs とすると，信号 $\mathbf{s}_l^{(j)}$ の実時間の長さを Ts s とすると，データ長は $U = Fs \times Ts$ である．**表 1** は学習用データセットのパラメーターである．

辞書を作成する為に，信号を長さ T の矩形窓で切り出した．切り出した信号 $\{\mathbf{x}_i^{(j)} = (x_i^{(j)}(0), ..., x_i^{(j)}(T-1))^T|i = 1, ..., M; j = 1, ..., 24\}$ をとする．各信号毎に得られるサンプル数は，$M = 3 \times \left(\frac{U-T-1}{K} - 1\right)$ である．$\mathbf{x}_i^{(j)}$ から，特徴ベクトル $\mathbf{y}_i^{(j)}$ を得る．特徴ベクトル $\mathbf{y}_i^{(j)}$ から平均楽音辞書 Y を作成し，主成分分析を行い，第 l 主成分まで使用した直交変換行列 V_l によってベクトル $\mathbf{y}_i^{(j)}$ を写像したものを楽音モデル $\{\mathbf{d}^(j)|j = 1, ..., 24\}$ とする．それらを集めて，楽音辞書 D を作成した．平均楽音辞書 Y を主成分分析をして得られる直行変換 V を **図 1** に記す．l の値は， $l = 23$ とし，寄与率は 99.9 %である．

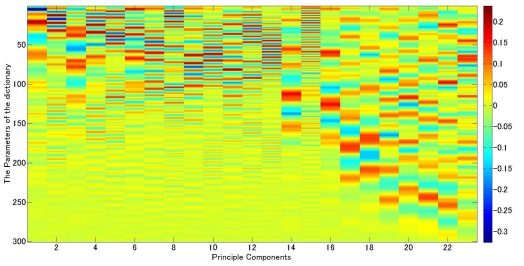


図 1 直交変換行列 V_{23} .
Fig. 1 Orthogonal transformation matrix V_{23}.

	Alto-Saxophone	ContraBass
MIDI sound source	KONTAKT5	
MIDI sequencer	Domino	
Recording Soft	Audacity	
Sampling-rate	44100[Hz]	
pitch range	C4 - B4	G2 - G♭3
Bit	16[bits]	
Velocity	30, 60, 120	

表 1 トレーニングデータのパラメーター
Table 1 The basic information of the training data.

4.2 楽曲データセット

使用する楽曲データセットとして，自作した楽曲信号 $\mathbf{s_t}$ として使用した．楽曲に使用した楽器種とそれらの使用音域は，学習用データセットと同じである．テンポは 120 bpm である．音の最小単位は 4 分音符を 3 つに分割した，3 連符とした．コントラバスは 2 分音符のみを使用した．2 は楽曲の音源パラメーターである．**図 2** に MIDI シーケンサを記す．

楽曲信号 $\mathbf{s_t}$ から特徴ベクトル $\{\mathbf{y}_i|i = 1, ..., l\}$ を計算する．自己相関関数のラグ，サンプリングレート，矩形窓長，窓シフト長はトレーニングデータを得る時の値を使用する．

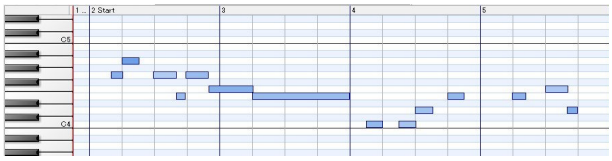
4.3 次元削減

得られた楽音モデル同士の相互相関が 0 に近いほど，推定する重みベクトル $\mathbf{w}_i$ の値の精度が増す．サンプリングレート Fs 下における楽音辞書 $D = \{\mathbf{d}^{(j)}|j = 1, \ldots, 24\}$ の相互相関行列を，

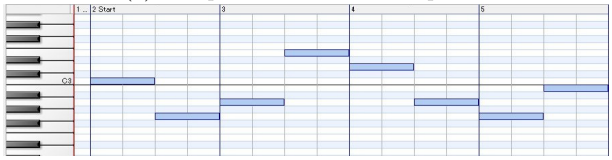
	Alto-Saxophone	ContraBass
MIDI sound source	KONTAKT5	
MIDI sequencer	Domino	
Recording Soft	Audacity	
Sampling-rate	44100[Hz]	
pitch range	C4 - B4	G2 - G♭3
Bit	16[bits]	
Velocity	61-97	50
Beat Per Minutes	120[bpm]	
Rhythm	2 beats	
The number of bars	4 bars	

表 2　楽曲データのパラメーター

Table 2　The basic information of music data.



(a) The piano roll of alto-saxophone



(b) The piano roll of contrabass

図 2　MIDI シークエンス

Fig. 2　MIDI sequences.

	平均楽音辞書 Y	楽音辞書 D
Alto-sax	0.6717	0.5578
Contrabass	0.2363	0.2859
Mixed	0.2901	0.2723

表 3　相互相関の合計と対角要素の和の比.

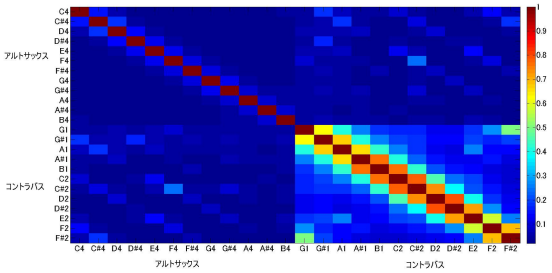
Table 3　The ratio of the sum of diagonal elements to the total sum of all cross-correlations.

$$\rho_{kl} = \frac{\mathbf{d}^{(k)^T}\mathbf{d}^{(l)}}{||\mathbf{d}^{(k)}||||\mathbf{d}^{(l)}||} \quad (k,l = 1,\ldots,24). \tag{13}$$

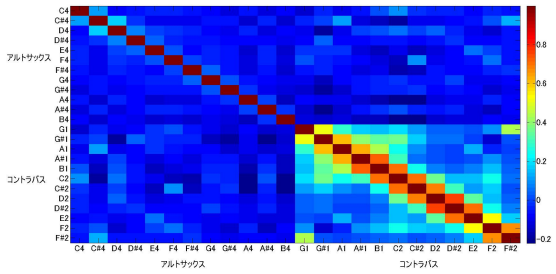
として計算する．得られた相互相関行列の全要素の和と，対角成分の和の比率を以下の式で求める．

$$Idp = \frac{\sum_i^N |\rho_{ii}|}{\sum_{k=1}^N \sum_{l=1}^N |\rho_{kl}|} \leq 1. \tag{14}$$

Idp は楽音辞書内のモデル同士の相関度合いを表し，1 に近いほど，楽音モデル同士は無相関であることになる．平均楽音辞書 Y の Idp と Y に寄与率 99.99 % 楽音辞書 D の Idp の値を**表 3** に記す．この表から，平均楽音辞書 Y に直交変換行列 $\{V_l|l = 23\}$ を適用して作成した楽音辞書 D を比較すると，アルトサックスの Idp と全体の Idp は低下している．しかし，コントラバスの Idp は向上が見られた．計算機による非負値疎分解による楽譜の生成に使用する楽音辞書として，平均楽音辞書 Y を直交変換した D を使用する．平均楽音辞書 Y と楽音辞書 D の相関行列を



(a) 平均楽音辞書 Y の相関行列



(b) 楽音辞書 D の相関行列

図 3　辞書行列の相関行列を可視化したもの．上が平均楽音辞書 Y であり，下が Y に直交変換行列 V_{23} を適用したものである．

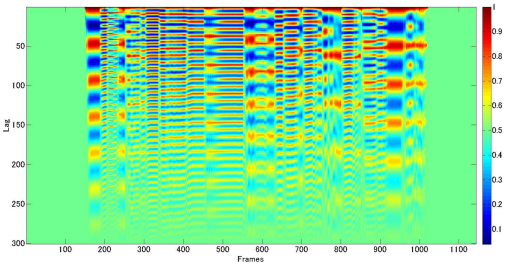


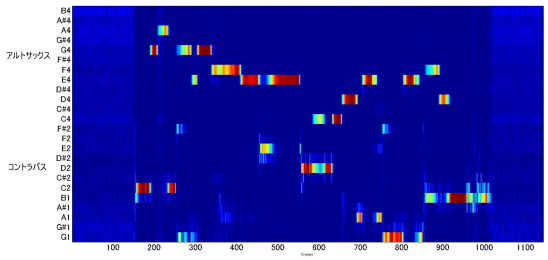
図 4　直交基底行列を疑似カラーでプロットしたもの．

Fig. 4　This figure visualize the auto correlations of Test data with psuedo color.

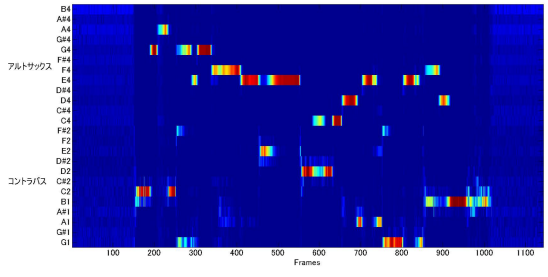
疑似カラーでプロットしたものを**図 3** に記す．

4.4　非負値疎分解による楽譜生成

作成したテスト用音源を用い，事前学習を行った辞書による非負値疎分解による楽譜生成アルゴリズムの動作性を確認した．我々は信号を $T = 300$ の矩形窓で切り出したものから音の特徴を抽出した．その結果，テストデータからは 1143 のサンプルを取り出せた．テスト用音源から抽出されたデータを**図 4** に示す．平均楽音辞書 Y から主成分を抽出し，300 から 23 へと次元削減を行った．この際の寄与率は 99.99 % であった．正則化パラメーター λ は幾つかの値を試しながら実験的に定めた．平均楽音辞書 Y を用いたものには，$\lambda = 50$ とした．楽音辞書 D を用いたものには，$\lambda = 0.7$ とした．アルゴリズムの最大試行回数は 1000 回に設定し，最大試行回数に到達する前に収束したと判断されたら更新を止めた．**図 5** に非負値疎分解の結果を示した．縦軸は楽音を示している．図の上半分がアルトサックスのパートを示し，下半分がコントラバス



(a) 平均楽音辞書 Y による非負値疎分解の結果.



(b) 楽音辞書 D による非負値疎分解の結果.

図 5 作成したデータから生成された非負値疎分解の結果を疑似カラーでプロットしたもの．(a) が平均楽音辞書を用いたもので，(b) が楽音辞書を用いたもの．

Fig. 5 Decomposition results of the test music. These figures visualize the coefficients of the dictionary components with psuedo color. (a) is music sound dictionary of nnACV vectors. (b) is music sound dictionary of nnACV vectors with PCA.

のパートを示している．アルトサックスとコントラバスの音階は下から順にそれぞれについて，$C4, C\#4, ..., B4$ と，$G1, G\#1, ..., F\#2$ となっている．推定された重みの強さは，疑似カラーで表現した．濃い赤い色程，重みの値が強く，深い青に近いほど，重みは 0 に近くなっている．図 5 は MIDI シーケンスと同様の音の配置になるようになっている．

MIDI シーケンスである図 2 と解析結果である図 5 を比べると，MIDI 信号として出力された楽音に対して重みづけが行われ，出力されていない楽音に対しては 0 の値に近い深い青色を示している事がわかる．特にアルトサックスのパートにおいては図 2 に沿った信号が強く出力されている．この結果から，提案手法による楽譜の生成は可能であることが示せた．平均楽音辞書 Y と楽音辞書 D の結果を比較すると，次元削減による楽音規定の重み推定性能の目立った低下も見られなかった．

5. 結論

本論文では，楽器音辞書を用いた楽曲の非負値疎分解による楽譜の生成について述べた．自己相関関数を用いて音響信号から抽出した特徴量を使用した楽音辞書に主成分分析を利用して次元削減を行った．しかし，次元削減を行う以前と，行った後の結果を比較する定量的判断材料が無い為，楽音辞書の性能表以下を客観的に示すことが出来ていない事は問題である．それゆえ，L1 正則化を行う際の正則化パラメーター λ を実験的に定めてしまい，定めた値が今回扱った問題に対して最適な値であるとは言えない．トレーニングデータと，解析する楽曲データの両方の音源を，同一のものとしている為，音色の異なる同一の楽器種に対しても適用できるかも定かではない．

今後は，得られた楽譜の評価方法を構築し，比較できるようにし，正則化パラメーターの最適値を推定する課題が一つ．モデル同士の無相関化を目指す事が一つ．楽器種が同一で音色の異なる音に対しても対応できるような汎化性能の高い楽音辞書の作成．以上三点の研究を行っていく．

参考文献

[1] Abdallah, S. A. and Plumbley, M. D.: Unsupervised analysis of polyphonic music by sparse coding, *IEEE Trans. on Neural Networks*, Vol. 17, No. 1, pp. 179–196 (2006).
[2] Lee, D. D. and Seung, H. S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol. 401, No. 6755, pp. 788–791 (1999).
[3] W.Xu, X.Liu and Y.Gong: Document clustering based on nonnegative matrix factorization, *Proc. ACM SIGIR*, pp. 267–273 (2003).
[4] Kawamoto, T., Hotta, K., Mishima, T., Fujiki, J., Tanaka, M. and Kurita, T.: Estimation of single tones from chord sounds using non-negative matrix factorization, *Neural Network World*, Vol. 10, No. 3, pp. 429–436 (2000).
[5] Virtanen, T.: Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria, *IEEE Transaction on Audio, Speech, and Language Processing*, Vol. 15, No. 3, pp. 1066–1074 (2007).
[6] Nakano, M., Roux, J. L., Kameoka, H., Kitano, Y., Ono, N. and Sagayama, S.: Nonnegative matrix factorization with markov-chained bases for modeling time-varying patterns in music spectrograms, *Latent Variable Analysis and Signal Separation*, pp. 149–156 (2010).
[7] Smaragdis, P. and Brown, J.: Non-negative matrix factorization for polyphonic music transcription, *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2003*, pp. 177–180 (2003).
[8] Cedric F 卒 evotte, Bertin, N., Durrieu, J.-L. : Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis, *Neural Computation*, Vol. 21, pp. 793–830 (2009).
[9] Kameoka, H., Nishimoto, T. and Sagayama, S.: Complex NMF: A new sparse representation for acoustic signals, *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*, pp. 45–48 (2009).
[10] Yoshii, K., Tomioka, R., Mochihashi, D. and Goto, M.: Infinite positive semidefinite tensor factorization for source separation of mixture signals, *International conference on machine learning 2013* (2013).
[11] Hoyer, P. O.: Non-negative sparse coding, *Neural Networks for Signal Processing XII (Proc. IEEE Workshop on Neural Networks for Signal Processing)*, pp. 557–565 (2002).

ⓒ 2015 Information Processing Society of Japan