Vol.2017-MUS-115 No.49
2017/6/18

スペクトログラム間の類似度に基づくピアノ弾き語り用伴奏譜の自動生成

越井 琢巳[1,a)]　斎藤 博昭[1]

概要：本論文では，音響信号のみを入力とするピアノ弾き語り用伴奏譜の自動生成システムを提案する．本システムでは，作曲の知識がないユーザ向けに，原曲の再現度を重視したピアノ譜面を提示する．このシステムでは和音選択の際，原曲の再現度を高めるために，和音の正解率よりもスペクトログラム間の類似度を重視する．これはスペクトログラム形状が似ているほど曲の雰囲気が似ているという仮定に基づくものである．原曲のスペクトログラムと，候補となる伴奏譜のスペクトログラムとの類似度を，コサイン類似度を用いた手法によって比較し，適切な和音とリズムパターンを選択する．出力された楽譜をピアノ経験者に見てもらい，同時に曲を聴いてもらうことで評価実験を行った．その結果，和音にリズムパターンの情報を付加してからスペクトログラムを比較するという本手法が，原曲のリズム再現において有効であることが示された．

1. はじめに

自動作編曲という分野は，1959 年に作られた『イリアック組曲』[4] という作品から始まった [7]．それは，人間が行ってきた作曲という行為をコンピュータに代替させるという初の試みであった．この作品から 50 年以上経過した現在までに，コンピュータを用いた作編曲の研究がさまざまな手法で行われてきた．簡単な和音の進行を，ジャズ風の複雑な和音進行に編曲するパーピープン [5] や，歌詞の韻律に基づいたメロディーを自動生成し作曲を行う Orpheus[1] など，作編曲の方向性も多岐にわたっている．

本論文では，ピアノの弾き語り用の楽譜に焦点を当てる．ピアノの弾き語り譜は通常，ボーカルパートとピアノパートの右手と左手の三段組で構成される．JPOP を弾き語りする場合，ボーカルパートに関して言えば，原曲のボーカルパートをコピーするだけでよい．しかしながら，ピアノでその伴奏を弾くとなれば必ずしも原曲通りというわけにはいかない．ドラムやギターなどピアノ以外の楽器の部分を，どのようにしてピアノ向けにアレンジするかが重要となってくる．

今回提案する手法は，このピアノパートを自動で生成するというものである．ここで，原曲の雰囲気を保つ伴奏の生成を目指す為にスペクトログラムの概念を導入する．これは，音源に含まれる周波数の時間的変化を見ることができるグラフである．提案手法では，編曲スペクトログラムと原曲スペクトログラムとの類似度がより高いほど，原曲の雰囲気をより保っているという仮説を立て，ピアノ譜面を生成する．

[1] 慶應義塾大学大学院理工学研究科
Keio University Graduate School of Science and Technology
[a)] allegro@nak.ics.keio.ac.jp

2. 前提知識と関連研究

2.1 和音

和音とは複数のピッチクラス (同じ音名の音の集合) の音が混ざっている音を指す．和音中の最も低い音を根音と呼び，例えば根音が『ド』の和音はコードネームで “C” や “Cm” などのように表現され，根音は頭のアルファベットで表される．和音は含むピッチクラス数によって，図 1 のような五の和音，図 2 のような七の和音 (セブンス)，図 3 のような九の和音 (テンションコード) に分類される．五の和音は，根音の 3 度上の音と 5 度上の音を含む和音であり，それに 7 度上の音を加えるとセブンス，さらに 9 度上の音も加えるとテンションコードになる．それぞれ 3 つ，4 つ，5 つのピッチクラスを含み，ピッチクラス数が多くなるほど複雑な響きであるように感じられる．また上記 3 種の他に，3 度の音が上下にずれた sus2，sus4 の和音や，11 度上の音あるいは 13 度上の音を加えた和音なども存在する (図 4)．

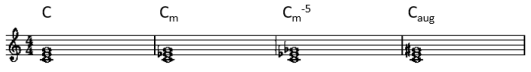

図 1　五の和音の例

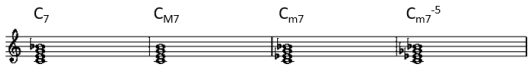

図 2　七の和音の例

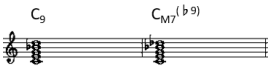

図 3　九の和音の例

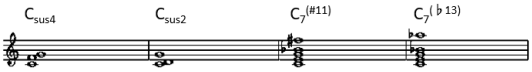

図 4　その他の和音の例

2.2　和音推定

音響信号または楽譜情報から和音の種類を判定する和音推定と呼ばれるタスクがある．一般に，和音中に含まれるピッチクラス数が多いほど推定が難しくなるとされる．また，楽譜情報を使う推定より，楽譜情報を使わずに音響信号のみを使う推定の方が難しいとされる．なぜなら，音響信号のみを使う推定では，その処理過程において複合音から単音への分解というタスクが発生するからだ．音響信号を使う推定手法には，楽譜を持たないユーザもシステムを利用できるというメリットがある．楽曲データから何らかの音楽的情報を得たいが，楽譜は無く，音源だけを持っているというようなユーザに対して，その需要を満たすことが可能となる．

音響信号から和音推定を行う研究では，クロマベクトルを用いた手法 [3] やスペクトルディップを用いた手法 [8]，隠れマルコフモデル (HMM) の学習と非負値行列因子分解 (Non-negative Matrix Factorization, NMF) を組み合わせた手法 [6] などが提案されている．それぞれの和音推定精度を表 1 に示す．

表 1　各手法での和音推定精度

手法	正解率	和音分類の種類
クロマベクトル [3]	69.9%	長三, 短三, 増三, 減三, 和音無し
スペクトルディップ [8]	83.3%	長三, 短三, 増三, 減三
HMM と NMF[6]	71.9%	長三, 短三, 和音無し

先に述べた通り，和音推定のタスクは和音中の音数が多くなるほど難しくなるが，表 1 においてその難しさは，和音分類の種類の多さによって示される．和音分類の種類が多いほど，難しいタスクを達成しているということになる．したがって，和音分類の種類が多く，かつ正解率の高いスペクトルディップ法が最も良い手法であるということが分かる．

2.3　自動作曲に関する関連研究

本節では，自動作曲の先行研究と本研究の立ち位置について述べる．

2.3.1　パーピープン [5]

パーピープンは単純なコード進行を入力するとジャズ風にアレンジされたコード進行を出力するシステムである．誰々風の演奏を生成したい，というリクエストに対し，まずその人の実際の演奏を集め，演奏をカデンツ*1に分解することで分析する．次に，実際に弾かれた音符や和音の発音タイミングの近いもの同士をグルーピングして和音列とし，各カデンツのボイシング*2を行う．

パーピープンは入力した和音進行に対して新たな和音進行を出力するものであるため，自動作曲というより自動編曲のシステムと言ったほうが良い．実際の演奏例を用いる事例ベースのシステムであり，演奏をカデンツに分解・再構成することにより，元の演奏の特徴を反映させることができるのがパーピープンの強みである．

一方で，入力が和音進行であり，また出力も和音進行である点から，システムを利用できるのは和音やコードについての知識を持ったユーザであり，その用途は作曲の支援と考えられる．

2.3.2　Orpheus[1]

Orpheus は，歌詞を入力すると，歌詞の韻律に基づいた旋律とその伴奏を生成するシステムである．旋律を横軸時間，縦軸音高の二次元平面状での遷移経路と見立て，動的計画法によって最尤経路を求める．伴奏はあらかじめ用意したものの中からランダムに選択されるか，ユーザが選択する．歌詞という非音楽的な入力を基に，一から作曲を行うことができる自動作曲のシステムである．

Orpheus は音楽的知識を持たないユーザでも，歌詞だけで気軽に作曲が行えるという強みがあるが，伴奏生成はランダムあるいは手動であり，あくまで旋律生成を主眼とした手法といえる．

2.3.3　本研究の立ち位置

本システムが目指すものは，原曲を基にピアノ伴奏を生成する自動編曲手法である．原曲の音響信号データのみを用いて，原曲の雰囲気に近いピアノ伴奏を生成できるシステム，すなわち Orpheus のように，作曲に関する知識を持たないユーザでも気軽にシステムを用いることができるシステムを目指す．出力和音には既存の和音推定で見られるような五の和音だけでなく，七の和音や sus2，sus4 を用い，パーピープンのように複雑な和音で原曲の再現を試みる．また，出力をそのまま弾き語り譜として即座に用いることができるシステムとする．この点において，いわゆる作曲支援システムとは異なる．

*1　論文中では “ケーデンス” と表現されている [5]
*2　ボイシングとは，和音に含まれるそれぞれの音をどのオクターブに置き，どの楽器に演奏させるのかを決めること．

3. ピアノ弾き語り譜の生成手法の提案

本節では，本論文が提案する伴奏譜の自動生成手法について述べる．

3.1 提案システムの俯瞰

提案システムの俯瞰を図 5 に示す．

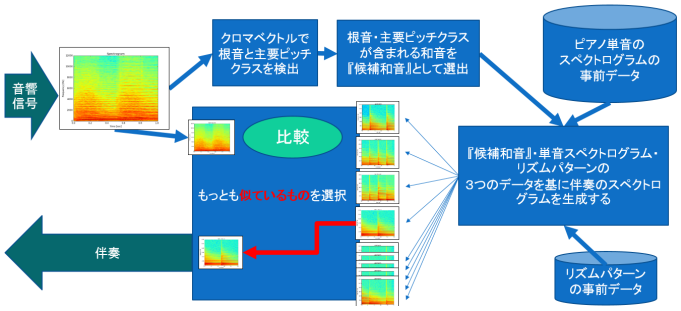

図 5 提案システムの俯瞰

提案システムは WAV の音響信号データを入力とする．提案システムではまず，音響信号から 2 拍分の長さのデータを取り出しスペクトログラムとクロマベクトルを作成する．このクロマベクトルを基に，複数の和音を候補に挙げる．事前に用意した複数のリズムパターンを候補の和音に当てはめ，ピアノ伴奏のスペクトログラムを作成する．候補の和音が n_c 個，リズムパターンが n_r 個とすると，$n_c \times n_r$ 個のスペクトログラムが作成されることになる．この $n_c \times n_r$ 個の候補スペクトログラムと，元の音響信号のスペクトログラムを比較し，最も似ている伴奏を選択する．この動作を 2 拍分の長さごとに行い，ピアノの伴奏を SMF 形式で作成していく．

後の節で，提案手法の各段階について詳しく述べる．

3.2 事前準備

本節では，提案手法を実現する上で必要な前準備について述べる．

3.2.1 システムに入力する情報

以下の項目をシステムに与えるものとする．

- 原曲の WAV ファイル
- 原曲の標本化周波数
- 原曲の BPM

3.2.2 ピアノ単音のスペクトログラムの作成

初めに，ピアノ単音 88 鍵分を録音する為に以下の手順で midi ファイル (dict.mid) を作成した．

(1) BPM を 120 とする．
(2) 88 鍵のピッチに対し下のピッチから順に，以下 3，4 の動作を行う．
(3) 1 拍 (0.5 秒) 分 velocity 値 100 でノートオン
(4) 1 拍分ノートオフ

次に，この dict.mid を電子ピアノの SMF 再生機能を用いて再生したものを録音し，dict.wav とした．こうすることで，dict.wav の s[sec] から $s+1$[sec] にかけて鳴っている音が，ピアノの鍵盤 88 鍵における s 番目 (ただし $s = 0 \sim 87$) の音となったため，アクセスが容易となった．

最後に，dict.wav ファイルを使って，WAV データから得られるピアノ単音のスペクトログラムを 88 鍵分すべて作成した．スペクトログラムの窓関数には式 (1) で表される ハニング窓を用い，チャンク幅を 0.04× サンプリング周波数，チャンクのスライド幅を 0.01× サンプリング周波数とした．

$$window(x) = 0.5 - 0.5\cos 2\pi x, \text{if } 0 \leq x \leq 1 \qquad (1)$$

3.2.3 リズムパターンの作成

リズムパターンは $r(0) \sim r(15)$ の計 16 パターン用意した (図 6)．

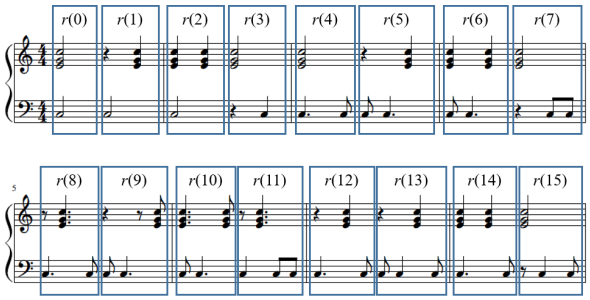

図 6 リズムパターン (和音が “C” の例)

図は，“C” の和音の例であるため，使用されている音が [C, E, G, C] となっている．和音 “C” のように和音中のピッチクラスが 3 つである時は，リズムパターンに適用されるピッチクラス列 $[p_0, p_1, p_2, p_3]$ の p_0 と p_3 には同じピッチクラスが格納される．別の例を挙げると，和音 “C_7” のように和音中のピッチクラスが 4 つである時は，$[C, E, G, A^{\#}]$ というように，すべてのピッチクラスが格納され，和音 “C_9” のように和音中のピッチクラスが 5 つ以上である時は，$[C, E, G, A^{\#}]$ と，5 つ目以降は格納しない．

図から分かるように，$r(0) \sim r(3)$ の 4 パターンは 4 分音符より細かく刻む音符は出現せず，残りの 12 パターンは 8 分音符より細かく刻む音符は出現しない．弾き語り譜として難易度の高すぎない楽譜にする為，8 分音符より細かく刻む音符は使用しなかった．このように，最も細かく刻む音符の種類を制限することで，難易度調整を図った，

3.3 候補和音の列挙

3.3.1 クロマベクトルの利用

本手法では，和音の候補列挙のためにクロマベクトルを用いた．

クロマベクトルがスペクトルディップ法よりも優れてい

る点は，音響信号中にどのピッチクラスの成分が多く含まれているかを測ることができる点である．本手法では，音響信号中に含まれるピッチクラスの情報を基に複数の候補和音を列挙する．複数列挙したのち，最後にスペクトログラムの比較によって正解を絞っていく．この為，候補の和音が少なすぎるとスペクトログラムの比較を行う意味が薄れてしまう．そこで，クロマベクトルによって得られた各ピッチクラスの振幅を利用する．これを使って，少なくともこの音は間違いなく含まれているという情報を取り出し，候補和音の列挙に用いる．

3.3.2 主要ピッチクラスの選定

まず，取り出した 2 拍分の原曲音響信号データからクロマベクトルを計算する．このうち振幅の大きい 4 つのピッチクラスを，主要ピッチクラスとして，候補和音列挙の為に用いる (図 7)．

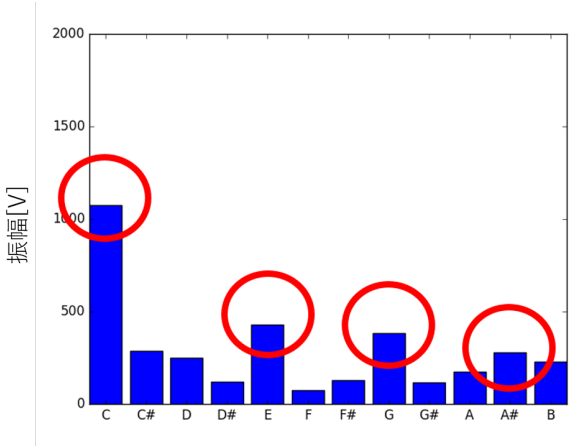

図 7 選択されたピッチクラスの例

3.3.3 候補和音の選択アルゴリズム

例えば，図 7 のように $\mathrm{C, E, G, A^{\#}}$ の 4 つのピッチクラスが選択されたとする．候補和音は，この 4 つのピッチクラスを根音に持つ和音 ($\mathrm{C_m, E_7, G_{dim}}$ など) の中から選択される．これらの和音のうち，$\mathrm{C, E, G, A^{\#}}$ のいずれか 3 つのピッチクラスを含むものを候補和音として絞り込む．今の場合，C と E と G を含む $\mathrm{C, C_7}$ の和音や，E と G と $\mathrm{A^{\#}}$ を含む $\mathrm{E_m^{-5}}$ の和音などが候補和音として選ばれる．

もし，この時に候補和音が一つも選ばれなかった場合は，$\mathrm{C, E, G, A^{\#}}$ のいずれか 2 つのピッチクラスを含むものでも候補和音として挙げてもよいという形で，含むピッチクラス数を減らした寛容な選出方法へと移行する．以下同様に，もし候補和音がない場合には，含むピッチクラスの数をさらに減らしていく．例えば図 8 のように，クロマベクトルによって選択されたピッチクラスが $\mathrm{C, C^{\#}, D, D^{\#}}$ であった場合は，含むピッチクラスの数を減らし寛容な選出方法へ移行する．

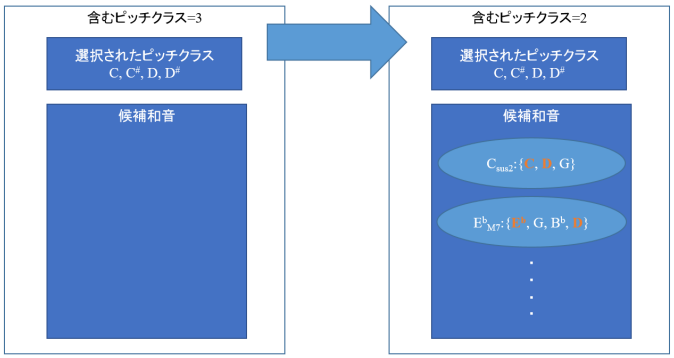

図 8 候補和音の選択

こうして選択された和音を候補和音とし，次のステップへと進む．

3.4 伴奏スペクトログラムの作成

混合音スペクトログラムの計算には，単音のスペクトログラムを用いる．第 n 鍵 ($n = 0 \sim 87$) の単音スペクトログラムを S_n，第 n 鍵のアクティベーションを x_n とすると，混合音スペクトログラム E は次のように表せる．

$$E = \sum_{n=0}^{87} S_n \times x_n \tag{2}$$

それぞれのリズムパターンのアクティベーションを基に E を求め，原曲のスペクトログラムとの比較に用いる．

3.5 スペクトログラムの比較

原曲スペクトログラムの計算には，前述と同様の窓関数，チャンク幅，チャンクのスライド幅を用いる．

3.5.1 比較範囲の検討

原曲スペクトログラムと，ピアノ単音から生成された混合音スペクトログラムの比較を行うにあたって，スペクトログラムの比較に用いる周波数の範囲を，計算時間，ピッチごとの周波数特徴，行列の疎密を考慮して実験的に求めた．本システムでは 100Hz〜1250Hz の範囲を比較に用いることとする．スペクトルは 25Hz ごとの値としたため縦軸 46 個のスペクトログラムとなる．横軸の個数は音源のテンポにより異なるので，全部で n チャンクあるとする．ただし前述の通り，チャンクの幅は 0.04× サンプリング周波数で表される．

以上のことから，本節は $46 \times n$ の 2 次元行列 S を処理する問題を扱うものといえる．

3.5.2 スペクトログラムの正規化

スペクトログラム S の比較を行う前に，式 (3) の関数 $\mathrm{normalize}(S)$ によって正規化し

$$\mathrm{normalize}(S) = \frac{1}{\left(\sqrt{\sum_{i=1}^{46}\sum_{j=1}^{n} s_{ij}^2}\right)} S \tag{3}$$

今後使用するスペクトログラムは原則として $\mathrm{normalize}(S)$ によって正規化されているものとする．

3.5.3 コサイン類似度による比較

スペクトログラムの比較にはコサイン類似度 $\cos\_\mathrm{sim}(\boldsymbol{vector}, \boldsymbol{vector_0})$ を用いる．あるチャンクでのスペクトル波形を，46 次元のベクトルとして見立て，このコサイン類似度の $\boldsymbol{vector}$ の部分に当てはめて考える．チャンクごとにコサイン類似度を計算し，その平均を求め，これをスペクトログラム間の類似度 $\mathrm{sim}(S, S_0)$ とする．$46 \times n$ の伴奏スペクトログラム S と原曲スペクトログラム O との間の類似度 $\mathrm{sim}(S, O)$ は以下の式 (4) で求められる．

$$\mathrm{sim} = \frac{\sum_{j=0}^{n} \cos\_\mathrm{sim}(\boldsymbol{s_j}, \boldsymbol{o_j})}{j} \tag{4}$$

スペクトログラム間の類似度 sim が最も高い伴奏譜を選択し，これを入力データに対して 2 拍ずつ行う．

3.5.4 コサイン類似度を用いることとなった経緯

当初，スペクトログラムの比較は画像処理の分野に関わると思われた．このため，ヒストグラムによる比較や，EMD(Earth Mover's Distance)[9] の利用も検討した．

しかし，画像比較では，ある画像とそれをわずかに平行移動させた画像では似ているとされるのに対し，スペクトログラムにおいてはわずかな周波数のずれは全く別の音を表すことになってしまう．このため，画像処理のアルゴリズムを用いるより，スペクトルでコサイン類似度を取る方が，各音高の特徴が得られてより良いという結論に至った．

4. 評価実験とその結果

完成したシステムに対し，被験者 2 人の協力のもと主観評価実験を行った．

4.1 実験

4.1.1 原曲として用いた楽曲

システムに入力する原曲には，RWC 研究用音楽データベース [2] のポピュラー音楽 1 番〜10 番を使用した．ポピュラー音楽データベースは，このデータベース用に制作された JPOP スタイルの楽曲を 100 曲収録したデータベースである．1 番から 10 番のサビの部分 32 拍 (4 分の 4 拍子でいうところの 8 小節) 分を取り出し，実験に用いた．

4.1.2 難易度の設定

8 分音符を使用するかしないかによって難易度を変化させて実験を行う．難易度は $r(0) \sim r(3)$ のみを使用した “易” と，$r(0) \sim r(15)$ の全てを使用した “難” の 2 種類で出力した．

4.1.3 使う音域の設定

使う音域によって伴奏に対する評価がどう変化するかを観察する．$\mathrm{F}^{\#}1$ キー以上の音を使用する “低” と，C2 キー以上の音を使用する “高” の 2 種類で出力した．

4.1.4 ボーカルパートの付加

出力されたピアノ伴奏に対し，ボーカルパートがあっているか観察する必要がある．そこで，データベースの SMF ファイルからボーカルパートのメロディーを取り出し，クラリネットの音色でピアノ伴奏の SMF ファイルに付加した．

4.1.5 比較用の出力

ポピュラー音楽データベースの正解和音ラベルを用いて，正しい和音で出力されたものも用意した．リズムパターンは 2 拍ごとに $r(0) \sim r(15)$ の中からランダムで選んだ．これを「参考」とし，比較用の目安として用いた．

4.1.6 被験者

本システムは，誰でも気軽に用いられるという方針のもと作成したものだが，ピアノ経験者ではないと出力された楽譜に対する評価が行えない．また，システムと同様に，ピアノへの編曲を行ったことがある者の方が，適切な評価を下せると考えられる．そこで，音のみ聴いて楽曲をコピーしピアノ演奏した経験のある者 (いわゆる “耳コピ” 経験者)2 人に協力してもらい主観評価実験を行った．

4.1.7 アンケート

次の 8 項目について 5 段階評価を行った．

A) 和音がボーカルパートに合っている
　そう思う 5 - 4 - 3 - 2 - 1 思わない
B) リズムがボーカルパートに合っている
　そう思う 5 - 4 - 3 - 2 - 1 思わない
C) 和音の響きが音楽的である
　そう思う 5 - 4 - 3 - 2 - 1 思わない
D) 楽曲としての面白みを感じる
　そう思う 5 - 4 - 3 - 2 - 1 思わない
E) 音の高さが適切か
　高すぎる 5 - 4 - 3 - 2 - 1 低すぎる
F) 難易度が適切か
　難しすぎる 5 - 4 - 3 - 2 - 1 簡単すぎる
G) 原曲らしさが再現できている伴奏である
　そう思う 5 - 4 - 3 - 2 - 1 思わない
H) ピアノ弾き語り譜として利用したいと感じる
　そう思う 5 - 4 - 3 - 2 - 1 思わない

4.1.8 実験手順

まず，原曲の WAV データを聴かせたのち，楽譜を見せた．次に，被験者に楽譜を見せながら，ROLAND 製の電子ピアノ HP603 の MIDI 演奏機能によって演奏された出力ファイルを聴かせた．その後，アンケートに記入させるという流れで実験を施行した．

出力ファイルとしては，難易度 “易・低”，“易・高”，“難・低”，“難・高”，“比較用” の 5 種類をランダムシャッフルしたものを用いた．この評価実験を 10 曲に対して行った．

図 9 は出力譜の例である．

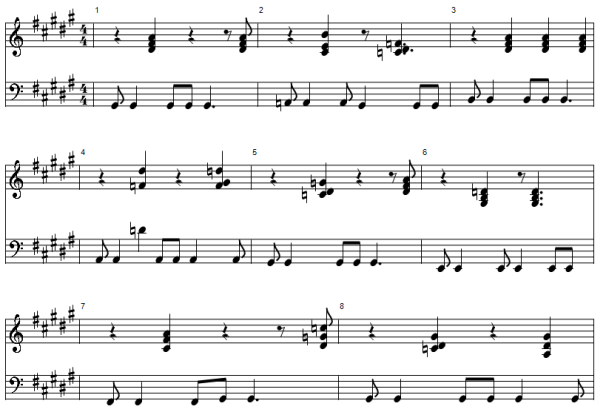

図 9　出力される楽譜の例 (1 曲目 “難・高”)

4.2　全 10 曲の結果の平均

結果には被験者二人の平均を用いた．グラフ (図 10) は全 10 曲の平均値である．

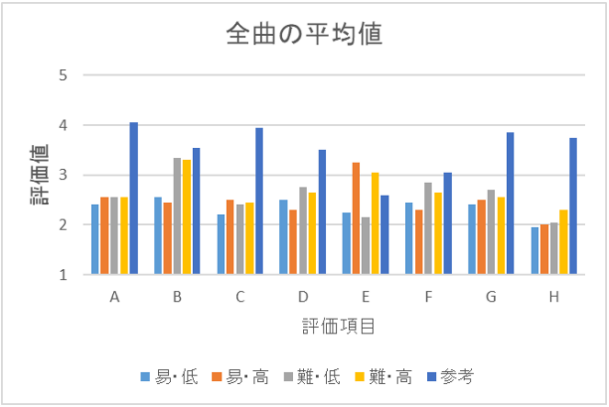

図 10　評価値平均

B の「リズムがボーカルパートに合っている」の評価値が高く，和音にリズムパターンの情報を付加してからスペクトログラムを比較するという本手法が，原曲のリズム再現において有効であることが示された．

5.　結論と今後の展望

本節では，本研究のまとめと今後の展望を述べる．

5.1　結論

本論文では，ピアノ弾き語りのための伴奏譜を自動生成する手法を提案した．入力に音響信号のみを用い完全な楽曲を出力する自動作編曲システムであることと，スペクトログラム間の類似度を基に最良の伴奏譜を選択する仕組みが，本手法最大の特徴である．

5.2　今後の展望

本システムで今後実装していく要素について述べる．

5.2.1　リズムパターンの学習

現状では，事前に用意した手入力のリズムパターン 16 種類を用いている．今後は，市販のピアノ弾き語り譜のリズムパターンを学習させて，テンポ・キーなどの情報も交えながら適切なリズムパターンを生成する機構を追加する．

5.2.2　スペクトログラム比較手法の強化と音源分離

現在，スペクトログラムの比較にはコサイン類似度を用いている．より強力な比較手法があれば，さらに原曲に近いピアノ伴奏の選出が行えるものと予想される．また，スペクトログラムの比較を行う前に，原曲からボーカルパート以外の部分の抽出を行うと，より原曲の伴奏部分に近いピアノ伴奏が生成できると考えられる．今後，音源分離の機構も備えたシステムにしていく必要がある．

参考文献

[1] 深山　覚，中妻　啓，米林裕一郎，酒向慎司，西本卓也，小野順貴，嵯峨山茂樹：Orpheus: 歌詞の韻律に基づいた自動作曲システム，情報処理学会研究報告， Vol. 2008, No. 78 (2008).
[2] 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, pp. 728–738 (2004).
[3] Harte, C. and Sandler, M.: Automatic chord recognition using quantised chroma and harmonic change segmentation, *Centre for Digital Music, Queen Mary University of London* (2009).
[4] Hiller, L. and Issacson, L.: Musical Composition with a High-Speed Digital Computer, 1958, *Journal of the Audio Engineering Society* (1958).
[5] 平田圭二，青柳龍也：パーピープン: 誰でもどこでもインタラクティブに使える知的ジャズ和音生成システム，情報処理学会研究報告音楽情報科学 (MUS)， Vol. 1999, No. 68, pp. 7–12 (1999).
[6] 丸尾智志：コード制約付き NMF を用いた音高推定に基づくコード認識，情報処理学会第 77 回全国大会， Vol. 5, p. 03 (2015).
[7] 松原正樹，深山　覚，奥村健太，寺村佳子，大村英史，橋田光代，北原鉄朗：創作過程の分類に基づく自動音楽生成研究のサーベイ，コンピュータソフトウェア， Vol. 30, No. 1, pp. 1_101–1_118 (2013).
[8] 宮田　聡，上野佑馬，蔵内雄貴，松原正樹，斎藤博昭：スペクトルディップテンプレートを用いた和音推定 (自然言語・音声・音楽, 一般論文)，情報科学技術フォーラム講演論文集 E-036，Vol. 8, No. 2, pp. 337–340 (2009).
[9] Rubner, Y., Tomasi, C. and Guibas, L. J.: The earth mover's distance as a metric for image retrieval, *International journal of computer vision*, Vol. 40, No. 2, pp. 99–121 (2000).