

音楽音響信号から得られる音楽要素に基づく自動ピアノアレンジ

高森啓史^{†1} 深山覚^{†2} 後藤真孝^{†2} 森島繁生^{†3}

概要: 本研究では音楽音響信号からピアノ編曲譜面を自動で生成する。好みの楽曲をピアノで演奏できるように、ピアノ編曲譜面の自動生成を行うことには需要がある。これまで複数の演奏パートから構成される楽曲のピアノ編曲譜面を生成する研究はいくつか行われてきた。しかし、従来研究ではいずれも入力として原曲の総譜やバンドスコアが必要であった。そこで本研究では音楽音響信号のスペクトルフラックスから得られるリズムと能動的音楽鑑賞サービス Songle を通じて得られるメロディ、和音、楽曲構造情報をもとに、右手・左手のパートを書き分けた編曲譜面を生成する。右手パートの生成ではメロディにコード構成音を付加し、左手パートの生成では既存のピアノ譜面より構成した伴奏データベースから伴奏を自動で選択する。

キーワード: ピアノアレンジ, リズム, コード, 能動的音楽鑑賞サービス「ソングル」

Automatic Piano-Arrangement Based on Music Elements Obtained from Music Audio Signals

HIROFUMI TAKAMORI^{†1} SATORU FUKAYAMA^{†2}
MASATAKA GOTO^{†2} SHIGEO MORISHIMA^{†3}

Abstract: This research makes a piano-arranged staff from musical audio signals. There is a demand for automatic generation of a piano-arranged staff in order to play one's favorite piece of music. There are several studies that generate a piano-arranged staff from an original score composed of multiple instruments. However, in previous works, an original orchestral score or band score is necessary as an input. Hence, this study generates a piano-arranged staff from musical elements, namely rhythm, melody line, chords and a music structure. The rhythm is obtained from a spectral flux of music audio signals. The melody line, chords and a music structure are obtained from Songle: a web service for active music listening. The proposed method involves generating both right- and left-hand playing parts of a piano staff. With respect to the right part, optional notes from a chord are added to the melody. With respect to the left part, appropriate accompaniments are selected from a database composed of pop musical piano-arranged staves automatically.

Keywords: piano-arrangement, rhythm, chord, a web service for active music listening "Songle"

1. はじめに

近年、ピアノで演奏される音楽の種類は多岐にわたっている。中でもロック、ポップ、テクノポップといったポピュラー音楽や、歌声合成技術 VOCALOID を用いて作曲された曲（以下ボーカロイド曲）の popularity が高く、初心者からピアニストまで多くの人によってこれらのジャンルの音楽が演奏されている。実際にポピュラー音楽やボーカロイド曲をピアノで演奏するまでの過程は大きく分けて二つある。一つは市販されているピアノアレンジされた楽譜を購入し、その楽譜に従い演奏を行うという過程。もう一つは自らピアノアレンジを行い、演奏を行うという過程。基本的には完成されたピアノアレンジ楽譜に従い演奏を行う前者の過程が主であるが、必ずしも所望の楽曲のピアノアレンジ楽譜が入手できるとは限らない。一方、自らピアノアレンジを行うには多大な労力と専門知識を必要とする。従って、

好みの楽曲をピアノで演奏出来るよう自動的にアレンジできる技術には需要がある。

我々はピアノという楽器の特徴を考慮した質の高いピアノアレンジ譜面の自動生成を目標として、原曲から得られる四つの音楽要素（メロディ、リズム、コード、音数）に注目した手法を提案している[1]。従来の我々の手法では、原曲の楽譜より音楽要素の抽出を行い、抽出された音楽要素に基づいてピアノアレンジ譜面の生成を行っている。この研究の特徴は、事前に構築しておく既存ピアノ譜面中の伴奏のデータベースを活用しつつ、譜面から抽出された四つの音楽要素にのみ基づいてピアノアレンジ楽譜が生成されるということである。本稿では、この枠組みを拡張し、入力として原曲の総譜やバンドスコアを用いる代わりに原曲の音楽音響信号と能動的音楽鑑賞サービス Songle [2]の解析結果を用いたピアノアレンジ譜面の生成手法について述べる。

2. 関連研究

複数のパートから構成される楽曲のピアノソロアレンジに関する研究は旧来から多く発表されている。藤田ら[2]

^{†1} 早稲田大学

Waseda University.

^{†2} 産業技術総合研究所

National Institute of Advanced Industrial Science and Technology (AIST)

^{†3} 早稲田大学理工学術院総合研究所

Waseda Research Institute of Science and Technology

は原曲の総譜から旋律推定とベース推定を行い、楽譜を集約した。さらに、同時打鍵における最大度数や演奏不可能な連打について制限をかけ、音を減らす、またはオクターブ音を上下させるなどの処理を施し、習熟度に応じた難易度の譜面を出力した。Chiuら[4]は原曲スコアの各パートが持つ役割を分類し、その結果から重要なパートを選択し、ピアノアレンジ譜面を生成した。その際、原曲の再現性や最大同時打鍵数、片手での度数制限といった演奏可能性を考慮した。中村ら[5]は演奏可能性を考慮する際に運指モデルを導入し、時間的連続性に注目した制限を考慮することで、原曲の総譜からアレンジ譜面に用いる音の選択を行った。

これらの関連研究について共通していることは、いずれも入力が原曲の楽譜であり、出力結果が元の楽譜の音符(またはオクターブ上下させた音符)がそのまま用いられていることである。そのため、これらの手法を応用して音楽音響信号からピアノアレンジ譜面の出力を行う場合は、全てのパートの音高推定を行う必要があり、非常に困難である。一方、Grahamら[6]のSong2Quartetでは、音楽音響信号から弦楽四重奏によるカバーソングを生成している。その目的においては、全てのパートの音高推定を行う必要はなく、原曲のメロディ、リズム、コード、音数といった音楽要素が抽出出来ればよい。本手法は、それと類似した枠組みでありつつ、特にピアノアレンジを生成するための制約の考察に重点をおいている。

3. 良いピアノアレンジとは

本研究における「ピアノアレンジ」とは、ピアノ曲でない楽曲(複数楽器から編曲される楽曲や他の楽器で演奏するために作られた楽曲)を原曲の印象を保ちながらもピアノで演奏できるように変更を加えることを意味する。そこで「良いピアノアレンジ」とは以下の条件を満たすものと定める。

- ① メロディが最高音であること。
- ② 原曲のコードと一致していること。
- ③ 原曲のリズムが反映されていること。
- ④ 原曲の抑揚が考慮されていること。
- ⑤ 生成された楽曲が演奏可能であること。

これらの五つの条件については、それぞれ以下の理由がある。①:ポピュラー音楽におけるボーカルは、常にその楽曲におけるどの楽器のパートよりも高い音であるとは限らない。なぜなら「歌声」という他の楽器と差別化しやすいパートがメロディを担っていることで、その楽曲の旋律を強調することが出来る。しかし、ピアノ1台で演奏する場合、最高音の印象が強くなりやすく、内声を音色や強弱の

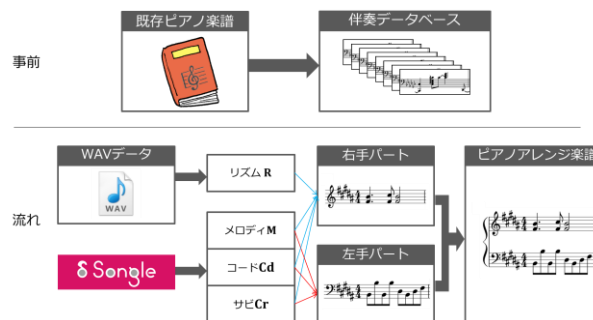


図 1 提案手法の概要図

Figure 1 An overview of the proposal method

みで強調することは難しい。従ってメロディは常に最高音であることは欠かせない条件である。②:アレンジ楽曲が原曲のコードと一致していない場合、不協和音が生じることが多く、聞いていて不快なアレンジ楽曲になってしまう。あえて原曲とは違うコードを割り当てリハーモナイズを行うアレンジ手法もあるが、本稿では原曲の印象を保ったアレンジを行いたいため、原曲のコードと一致していることを良いアレンジの条件として加える。③:リズムもまた楽曲を印象付ける重要な要素の一つである。主にリズムを変化させることでジャズ風やボサ・ノヴァ風といった表現が出来るほどリズムが楽曲に与える印象は大きいと言える[7]。従って、原曲の印象を保つためにリズムを変化させないようにする。④:ポピュラー音楽にはイントロ、Aメロ、Bメロ、サビといった構造がある。ピアノアレンジ版においても、サビでは音の厚みを増すといった操作を加えることで、より原曲の雰囲気を反映させることが出来ると考えられる。⑤:実際に人が弾くことを想定しているため、演奏可能なピアノアレンジ譜面である必要がある。具体的には、既存のピアノ楽譜が演奏可能であるという仮定の下、既存ピアノ楽譜から構築した伴奏データベースから伴奏を選択する。提案手法では、以上の五つの条件を満たすようピアノアレンジ譜面の生成を行う。

4. ピアノアレンジ譜面の生成手法

本手法の概要を図1 提案手法の概要図に示す。事前処理として伴奏データベースの構築を行う。次に能動的音楽鑑賞サービス Songle を通じてメロディ、コード、サビ情報の抽出を行い、音楽音響信号からリズムの抽出を行う。これらの音楽要素に基づき右手・左手のパートを書き分けたピアノアレンジ譜面を出力する。

4.1 伴奏データベースの構築

既存ピアノ楽譜の左手部分を基に伴奏データベースを構築する。伴奏データベースとは1小節ごとの発音位置と音高の情報が36行16列の行列(以下伴奏行列と呼ぶ)として表されたデータベースであり詳細は文献[1]の第3章を参

照されたい。また、伴奏データベースはDBと表しn番目の伴奏行列はDB_nと表す。

同時に、各伴奏行列からリズムDBR_nと音数DBAN_nを抽出する。リズムDBR_nは16次元ベクトルで表し、伴奏行列DB_nのj列目に1以上の数が存在する場合は、リズムDBR_nのj番目の要素に1を、なければ0を格納する。音数DBAN_nは伴奏行列DB_nに含まれる1以上の要素数を格納する。ただし、データベースに含まれる音数DBAN_nの最大値が1となるよう規格化する。

4.2 音楽要素の抽出

本手法では音楽要素としてメロディM、リズムR、コードCh、サビCrを使用する。メロディM、コードCh、サビCrについては能動的音楽鑑賞サービス Songle から抽出する。リズムRは入力楽曲の音楽音響信号から抽出する。以下、i小節目の各音楽要素をM_i, R_i, Ch_i, Cr_iと表す。

4.2.1 Songle を用いた音楽要素の取得

能動的音楽鑑賞サービス Songle とは、音楽をより深く理解しながら、能動的に音楽鑑賞が楽しめるサービスであり、110万曲以上のメロディ、コード、ビート、楽曲構造が解析されている。本手法では Songle の解析結果を用いて入力楽曲の音楽要素の抽出を行う。

Songle のメロディの解析結果から、メロディの各ノートに対する開始時間、音高、長さが取得できる。また、ビートの解析結果から、各ビートの打刻時間と小節番号が取得できるため、実時間から楽譜上の時間である音価とその位置を求めることができる。よって、メロディMには入力楽曲のメロディの各ノートの音高、音価、位置を格納する。格納方法は上記情報を小節ごとに88行16列の行列としてM_iに格納する。音高は列方向、位置は行方向、音価は格納される値で表現する。行数はピアノの鍵盤数を表しており、列数は16分音符の音価まで考慮するという分解能を表す。

Songle のコードの解析結果から、各コードに対するコード(根音、コードタイプ)、開始時間、長さが取得できる。これらの情報とビートの解析結果から各コードの含まれている小節番号、開始位置、長さが求められる。よって、コードChには入力楽曲の各コードの開始位置と長さを格納する。

Songle の楽曲構造の解析結果からサビ区間が取得できる。この情報とビートの解析結果から、サビである小節の小節番号がわかる。よって、サビCrには入力楽曲の各小節がサビであるかサビでないかの情報を{1,0}で格納する。

4.2.2 音楽音響信号からリズムの抽出

リズムR抽出の流れに入る前に、リズムはビートとは別物であることに注意されたい。ビートとは時間的に均等に配置されたアクセント、またはその繰り返し構造を表し、リズムは必ずしも時間的に均等な配置となっていないアクセントの集合を表す。当然、同じビート情報を持つ楽曲がリズムも一致するとは限らない。従って、本手法では楽曲

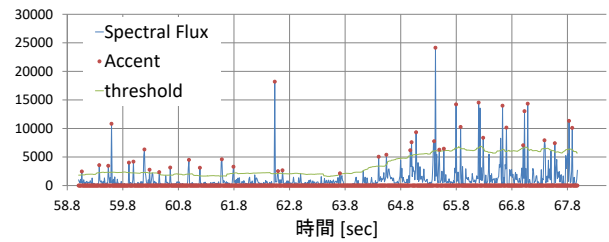


図2 スペクトルフラックスのピーク検出

Figure 2 The peak detection of spectral flux

のアクセント情報として、音楽音響信号から得られるリズムを使用する。

リズムR抽出の流れについて説明する。初めに入力楽曲にローパスフィルターをかけた波形のスペクトルフラックスF_tの計算を行う。スペクトルフラックスは各周波数におけるパワースペクトルの時間変化を表したものである。本手法ではパワースペクトルが増加する場合についてのみ考え、以下のように計算する。

$$F_t = \sum_{k=1}^K f_t[k] \quad (1)$$

$$f_t[k] = \begin{cases} (N_t[k] - N_{t-1}[k])^2 & N_t[k] > N_{t-1}[k] \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

N_t[k]とN_{t-1}[k]は、t-1番目のフレームにおけるスペクトルのk番目の周波数におけるパワースペクトルである。この処理を行うことで、入力楽曲のパワースペクトルが局所的に増加している部分、すなわちアクセントとなる箇所を探索することが出来る。次に、スペクトルフラックスが局所的にピークとなる箇所の検出を、2次多項式適合による平滑化微分を用いたピーク検出[8]と閾値処理によって行う。ピーク位置は、スペクトルフラックスF_tの平滑化微分

$$\sum_{\omega=-W_f}^{W_f} F_{t+\omega} \quad (3)$$

が正から負に変わる箇所でもまり、さらにスペクトルフラックスF_tが閾値を超えている場合、その箇所をアクセントと定める。しかし、1曲の中で比較的静かで落ち着いた箇所と激しく盛り上がる箇所が混在する楽曲においては一定の閾値を設けてしまうとアクセントの検出が困難となる。そのため、閾値φ_tを次のようにフレームごとに定めることでこの問題の解決を図った。

$$\phi_t = \alpha \times \frac{\sum_{\omega=-W_\phi}^{W_\phi} F_{t+\omega}}{2W_\phi + 1} \quad (4)$$

αは係数を表す。図2にアクセント検出の様子をグラフに示す。図2は、Songleに登録されている楽曲[9]の一部であり、サンプリング周波数は44100Hz、窓幅512サンプル

で計算したスペクトルフラックス F_t の外形と、 $W_f = 5$,
 $W_\phi = 70$, $\alpha = 2$ とした時のアクセントと閾値の変化を示す。

最後に、以上の過程より求めたアクセントをリズム \mathbf{R} に格納する。まず、Songle より各ビートの打刻時間を取得し、アクセントが発生した時間の小節番号と位置を求める。位置は1小節を8分割した際何番目の位置かで表す。i小節j番目の位置にあるアクセントは8次元ベクトルで表された \mathbf{R}_i のj番目の要素に1が格納される。アクセントがない要素はすべて0が格納される。

4.3 右手パートの生成

右手パートは抽出されたメロディ \mathbf{M} をそのまま割り当てる。ただし、サビ区間の小節における右手パートはメロディにコード構成音を付加する操作を行い、厚みと抑揚を与える。コード構成音は、アクセントのある箇所(リズム \mathbf{R}_i の要素が1となる箇所)についてのみ付加を行い、メロディよりも低い音高、かつ3度以上離れた音を付加する。以上の操作はメロディが最高音になるよう右手パートを生成し、入力楽曲のコードを保つよう、コード構成音から付加する音を選択しており、サビとサビ以外の部分で操作を変えているため、第3章における「良いピアノアレンジ」の①, ②, ③, ④の条件を満たしている。

4.4 左手パートの生成

左手パートは、4.1節で構築した伴奏データベース \mathbf{DB} から距離関数 Dist_i を最小にする伴奏データを選択する。距離関数 Dist_i はリズムと音数(厚み)に関して、入力楽曲と各伴奏データの類似度を表す関数である。選択された伴奏データはコード \mathbf{Cr}_i に基づきコードが与えられる。i小節目の左手パートを LeftPart_i としたとき、以下の式で与えられる。

$$\text{LeftPart}_i = \text{Hrm}(\arg \min \text{Dist}_i(\mathbf{DB}), \mathbf{Cr}_i) \quad (5)$$

$$\text{Dist}_i(\mathbf{DB}_n) = \omega_1 \times |\mathbf{DBAN}_n - \mathbf{Cr}_i| + \omega_2 \times \sum_{j=1}^8 |\mathbf{DBR}_n[2 \times j - 1] - R_i[j]| \quad (6)$$

コードを与える関数 Hrm は文献[1]の5.2節を参照されたい。 ω_1 と ω_2 はそれぞれの重みを表す。 $\mathbf{DBR}_n[2 \times j - 1]$ および $R_i[j]$ は4.1節および4.2.2項で記述したn番目のデータベースのリズム \mathbf{DBR}_n の $(2 \times j - 1)$ 番目の要素と入力楽曲のi小節目のリズム \mathbf{R}_i のj番目の要素を表す。本来、リズム \mathbf{R}_i も16次元ベクトルで表されるべきであるが、現段階では16分音符の音価での正確なアクセント位置を取得することが難しいため、8分音符、もしくはそれより長い音価について考える。以上の距離関数 Dist_i を最小にする伴奏を選択する操作は、「良いピアノアレンジ」の②, ③を満たす。

生成された左手部分の譜面は、既存のピアノ楽譜を基に構築されたデータベースを使用して生成されるため、既存楽譜の伴奏部分が演奏可能であれば基本的に生成された譜面も演奏可能であるといえる。従ってこの操作は「良いピ



図3 「PROLOGUE」のピアノアレンジ結果
 Figure 3 The piano arrangement of “PROLOGUE”

アノアレンジ」の条件⑤を満たすものである。

4.5 右手・左手パートの相互関係に基づく修正

最後に4.3節, 4.4節で生成した各譜面を合わせ一つの譜面にする。修正方法については文献[1]の5.3節に記述されている通りである。

5. 結果と考察

本手法を用いて Songle に登録されているポピュラー音楽及びボーカロイド曲についてピアノアレンジを行い、Midi データとして出力した。その結果を図3に示す。(a)はAメロ部分の一部を表し、(b)はサビ部分の一部を表す。Midi データは MuseScore 2.1[10]を使用して楽譜として表示した。使用楽曲として VocaListener(AIST)の「PROLOGUE」[11]を使用して楽譜を生成した際には、Aメロ部分では伴奏の音数が比較的少なく、サビ部分では多く配置されており、抑揚がつけられていた。また、おおむねアクセントがある位置に音がなるべく配置されているように生成されていた。右手部分ではサビにおいてメロディにコード構成音が付加されており、音の厚みが適切に増していた。すべてのメロディにコード構成音が付加されているわけではなく、アクセントが強い箇所のみ音を追加されていたため、より自然な音数の増し方がされていた。

一方、Songle よりメロディ情報を取得する際、音高はあっても、歌詞に合わせて音が途切れていない場合があった。Songle は解析誤りを誰でも訂正することができるが、Songle において誤り訂正後の楽曲であっても、連続する二つの音符が同じ高さの場合、そのように途切れずに一続きの音符になっている場合があった。本手法は Songle の解析結果を用いてメロディを生成するため、解析結果が出力さ

れるピアノアレンジ譜面の右手部分に直接関わってしまい、本来の歌詞では違う音になっているのに一続きの音符となって印象が変わるパートが出現してしまう問題が起きていた。この問題は、音楽音響情報を用いてメロディの途切れる箇所を検出をすることで解決が図れると考えられる。

6. まとめと今後の課題

本研究では原曲の音楽音響信号と Songle から音楽要素の抽出を行い、ピアノアレンジ譜面の生成を行った。特に、リズムの抽出においては Songle からは得られないアクセントが強い位置をスペクトルフラックスのピーク検出を行い、楽譜上のどの位置に対応するのかを求めた。抽出した音楽要素に基づき、右手パートはサビにのみコード構成音を付加し、左手パートは伴奏データベースから距離関数を最小にする原曲との類似度の高い伴奏を選び、コードを与えた。

今後は、伴奏データベースの増築を行う。また、難易度別のデータベースを構築することで、ユーザのレベルに合わせたピアノアレンジ譜面の出力を行う予定である。

謝辞 本研究の一部は JST ACCEL (JPMJAC1602)の支援を受けた。

参考文献

- [1] Hirofumi Takamori et al.. Automatic Arranging Musical Score for Piano using Important Musical Elements. Sound and Music Computing Conference. 2017, pp. 35-41.
- [2] 後藤真孝ら. Songle:音楽音響信号理解技術とユーザによる誤り訂正に基づく能動的音楽鑑賞サービス. 情報処理学会論文誌. 2013, vol. 54, no. 4, pp. 1363-1372.
- [3] 藤田顕次ら. 習熟度を考慮した複数楽譜からのピアノ譜生成手法の提案. 情報処理学会研究報告音楽情報科学(MUS). 2008, vol. 2008-MUS-77 No. 10, pp. 47-52.
- [4] Chiu Shih-Chuan et al.. Automatic system for the arrangement of piano reductions. IEEE International Symposium on Multimedia. 2009, pp. 459-464.
- [5] Nakamura Eita et al.. Automatic Piano Reduction from Ensemble Scores Based on Merged-Output Hidden Markov Model. ICMC. 2015, pp. 298-305.
- [6] Graham Percival et al., Song2Quartet: A System for Generating String Quartet Cover Songs from Polyphonic Audio of Popular Music, ISMIR, 2015, pp. 114-120.
- [7] 樋口拓志ら. 与えられたコード進行に基づくギター伴奏用ボサ・ノヴァ編曲システム. 情報処理学会研究報告音楽情報科学 (MUS). 2008, vol. 2008-MUS-078, pp. 47-52.
- [8] Savitzky, A. and Golay, M. J.. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry. 1964, vol. 36, no. 8, pp. 1627-1639.
- [9] “livetune feat. 初音ミク 『Tell Your World』 Music Video”. <https://www.youtube.com/watch?v=PqJNc9KVIZE>, (参照 2017-08-03).
- [10] “MuseScore”. <https://musescore.org/ja>, (参照 2017-08-03).
- [11] “【初音ミク】 PROLOGUE 【ばかりす+ぽかうお】 by VocaListener (AIST) staff.aist.go.jp”. <https://staff.aist.go.jp/t.nakano/music/VocaWatcher.Prologue.Mik.u.mp3>, (参照 2017-08-03).