

# 既存楽曲を利用した旋律生成システムの試作

## A Prototype System for Melody Synthesis using Existing Music

出口 幸子  
Sachiko Deguchi

大塚 健一  
Kenichi Otsuka

近畿大学工学部  
School of Engineering, Kinki University

This paper describes the prototype system that extracts phrases from existing music and that synthesizes new music using those phrases. This prototype system builds phrase database for each composer and synthesizes the music by choosing the phrases alternately from two databases. The system makes the feature vector for each phrase, and then chooses the nearest one to generate the next phrase. To make the feature vector, the system calculates the average of pitches in a phrase, the average of durations in a phrase, the difference between max pitch and min pitch in a phrase, etc. The melodies generated by this synthesis method are more natural than a random sequence of phrases, however, it is difficult to evaluate the effects of the feature vector. This system can synthesize the melodies by reusing huge music resources.

### 1. はじめに

本研究は計算機上でジャムセッションのモデルを検討することを目的に開始した。ジャムセッションの研究では複数の演奏者が互いに関連しながら即興演奏を生成する方式が研究課題であるが[1][2], 問題が複雑でモデル化が容易でない。そこで, 本研究では, 問題を単純化し, 既存楽曲のフレーズを用いて, フレーズ間の適合度を判定して旋律を合成するモデルを検討する。モデルの概要, 楽譜情報, 特徴ベクトルと適合度, およびモデルの実現と実行結果について述べる。

### 2. 旋律合成のモデルと楽譜情報

#### 2.1 旋律合成のモデル

ジャムセッションを単純化したモデルとして, 2名の音楽家が交互にフレーズを演奏するモデルを検討する。本モデルでは, 図1に示すように, それぞれの音楽家が自分の楽曲で使用しているフレーズの中から, 相手が演奏したフレーズに適したものを選択して演奏することにより旋律が合成される。

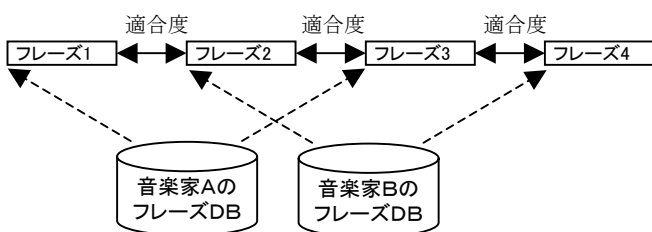


図1: 旋律合成のモデル

#### 2.2 楽譜情報

本研究では, 楽譜情報を入力したデータを使用する。入力形式には Humdrum の kern 形式[3]を使用してテキストファイルを図2のように作成する。ここで, 音高は, 中央のドからシを c, d, e, f, g, a, b, と表し, 1オクターブ下を C, D, E, ..., 1オクターブ上を cc, dd, ee, ..., 等と表している。音価は, 4分音符を 4, 2分音符

を 2, 付点4分音符を 4., 等としている。例えば, 4分音符の中央ドは 4c となる。このように記号で表された音高と音価を図2に示すように数値に変換する。ここで, 音高は MIDI と同じ表記を用い, 中央ドを 60 として, 音程を音高の差で表し, 半音の音程を 1 としている。音価は4分音符を 96, 2分音符を 192, 等と変換する。さらに, 複数の調の曲が混在しているので, 1つの調に移調する必要がある。本研究ではハ長調に移調する。

今回使用した楽曲は, 平井堅, 桜井和寿(Mr. Children), およびつんくの3名の音楽家の作曲した8曲であり, 声のパートを対象として作成した楽譜データベース[4]を使用している。

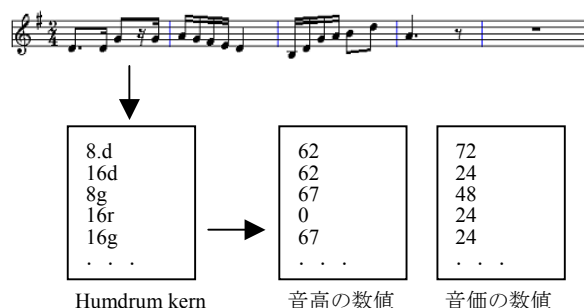


図2: 楽譜情報

#### 2.3 フレーズの切り出し

本研究では楽譜データベースからフレーズを切り出して旋律の合成に用いる。切り出しには以下の方法が考えられる。

- ・ 固定長: n 小節(n=1~4)の固定長のデータを 1 フレーズとする。n=2, 3, 4については, 以下の2通りの方法がある。
  - ・ n 小節毎に順に切り出す方法。
  - ・ n 小節を切り出し, 次に 1 小節ずらして n 小節を切り出し, 重複して順次切り出して行く方法。
- ・ 可変長: 休符で切り出し, 長さが 1~4 小節のものをフレーズとする。なお, 1 小節より短い場合は前後の短い方のフレーズと合わせ, 4 小節より長い場合は最長の音価の後に分割する。今回の試作では, 2 小節毎に順に切り出す方式を実現している。切り出したフレーズを音楽家毎にまとめて持ち, 旋律合成に用いる。なお, 曲に繰返しがある場合, 楽譜データベースにはフレーズが重複しているため, フレーズの切り出しの際に, 同一フレーズを除去する。

### 3. 特徴ベクトルと適合度

各フレーズに対して、特徴ベクトルを作成して、ユークリッド距離を計算することにより、フレーズ間の適合度を判定し、最も適合度の高いフレーズを順次選択して生成する。一度生成したフレーズは除外して次のフレーズを選択する。

#### 3.1 特徴ベクトル

フレーズにおいて、次の特徴量を検討する。

a: 音高の平均, b: 音価の平均, c: 音高の最高値と最低値の差, d: 音高の最高値, e: 音高の最低値, f: 特定の音価(4分音符, 8分音符)の数。

また、これらの特徴量の組み合わせから成る特徴ベクトルとして、次のものを検討対象とする。

(a, b), (a, b, c), (a, b, d, e), (a, b, f)。

なお、各特徴量に重み係数を掛けることも検討する。

#### 3.2 適合度の判定方式

2つのフレーズの適合度を判定する方式を以下に記す。

方式 1: フレーズAの特徴ベクトルと、他のフレーズ $B_j$  ( $j=1,2,3, \dots$ )の特徴ベクトルとのユークリッド距離 $L$ を比較し、最小となるフレーズ $B_k$ をフレーズAとの適合度が高いとする。

方式 2: フレーズAとフレーズ $B_j$  ( $j=1,2,3, \dots$ )の特徴ベクトル間のユークリッド距離を $L$ 、フレーズAの最後の音高とフレーズ $B_j$ の最初の音高の差を $D$ とし、 $L^2+D^2$ を最小とするフレーズ $B_k$ をフレーズAとの適合度が高いとする。つまり、特徴量に加えて、フレーズ間の音のつながりを考慮する方式である。

### 4. モデルの実現と実行結果

#### 4.1 モデルの実現

本研究では、Java を用いてモデルを実現している。音の生成には MIDI 音源を使用している。合成した旋律を MIDI ファイルに書き出してそれを再生するのではなく、次に演奏するフレーズを決定した直後に旋律を生成する。これは、今後インタラクティブな旋律生成を目指しているためである。また、MIDI シーケンスを作成して再生するのではなく、本システムで保持しているフレーズの情報(音高と音価の列)に基づき、ノートオンを送信して音を生成し、スレッドをスリープさせることによりリズムを実現する。

#### 4.2 旋律生成の実行

前述した特徴量 a~f を用いて、以下の特徴ベクトルおよび適合度の判定方式により旋律を生成する。ここで、 $L$  はユークリッド距離、 $D$  はフレーズ間の音高の差を表す。

- ①特徴ベクトル(a, b),  $L$  最小
- ②特徴ベクトル(a, b, c),  $L$  最小
- ③特徴ベクトル(a, b, d, e),  $L$  最小
- ④特徴ベクトル(a, b),  $L^2+D^2$  最小
- ⑤ランダム
- ⑥特徴ベクトル(a, b, c),  $L$  最小・重み付
- ⑦特徴ベクトル(a, b, f),  $L$  最小

以降、平井堅を H、桜井和寿を M、つんくを T と表す。音楽家の組み合わせ(演奏順を含む)は、HM, HT, MH, および MT について実行する。HM は上記①~⑦の全方式、HT, MH, および MT は上記①②④⑦の方式を実行する。最初に生成するフレーズは、先に演奏する音楽家の一曲を選び、その A メロの第一フレーズ、およびサビの第一フレーズとする。

### 4.3 実行結果

それぞれに異なる旋律が生成されることを確認した。しかし、例えば、特徴ベクトル(a, b)に対し、(a, b, c)は音高の幅を特徴量に加えているが、その効果は認められるものの評価は難しい。また、(a, b, d, e)は音高の幅に代え、音高の最高値と最低値を独立に特徴量としているが、その効果の評価はやはり難しい。特徴ベクトル(a, b, c)の各特徴量に異なる重みを掛けた場合(⑥)、特徴ベクトル(a, b)に加えてフレーズ間の音のつながりを考慮した場合(④)、あるいは特徴ベクトル(a, b, f)として特定の音価(8分音符)の数を特徴量に加えた場合(⑦)も効果の評価は難しい。一方、ランダムにフレーズを接続する方式は、実行の度に様々な旋律が生成されるが、違和感を与える旋律となることもある。以上から、特徴量の優劣は付け難いが、ランダムに接続する場合の違和感を除去する効果はあると考えられる。

音楽家の組み合わせについては、HT の組に違和感がある場合がある。これは、T のリズムが H に対して単調であること[4]によると考えられる。

また、実行結果の一部について、工学部の学生 14 名を対象として主観評価実験を行った。生成された 2 種類の旋律を続けて聴き、どちらがよいかを回答する方法を取った。本実験結果も、上述の結果と同様の傾向となった。

### 5. おわりに

複数作曲家の複数楽曲の楽譜情報を用いて、フレーズを切り出し、それらを接続して旋律を生成するシステムを試作した。各フレーズに対して特徴ベクトルを作成し、ユークリッド距離の近いものを順次選択する方式、および距離に加えてフレーズ間の音のつながりを考慮する方式を検討した。特徴量としては、音高の平均、音価の平均、音高の最大値と最小値の差、等を検討した。特徴ベクトルを使用することで違和感のない旋律を生成できることが示されたが、特徴量の優劣を付け難いことも分かった。また、音楽家の組み合わせには、適した組と適さない組があることも示された。

課題としては、他の特徴量の検討、マハラノビス距離の検討、他のフレーズ切り出し方法(2 小節毎以外)の実現、等が挙げられる。合成された旋律を定量的に評価する方式についても検討したい。また、現在、楽譜から作成した楽譜データベースを利用しているが、将来的に SMF (Standard MIDI File) を使用できるようにする必要があると考えている。

本研究は、ジャムセッションモデルの基礎研究として行ったが、ここで検討した方式により既存楽曲を用いた旋律合成が可能である。特に、著作権の切れた膨大な音楽資源を再利用できると考える。また、本研究の結果を基に、今後、ジャムセッションのモデルの検討をする予定である。

謝辞: 本研究の機会を頂いた近畿大学工学部電子情報工学科の先生方に感謝致します。

#### 参考文献

- [1] 浜中雅俊, 後藤真孝, 麻生英樹, 大津展之: Guitarist Simulator: 演奏者の振舞いを統計的に学習するジャムセッションシステム, 情報処理学会論文誌, Vol.45, No.3, pp.698-709, 2004.
- [2] 金森森, 片寄晴弘, 新美康永, 平井宏, 井口征士: ジャズセッションシステムのための音楽認識処理の一実現法, 情報処理学会論文誌, Vol.36, No.1, pp.139-152, 1995.
- [3] D. Huron: The Humdrum Toolkit Reference Manual, CCARH, Stanford University, 1994.
- [4] 三家本祥平, 井手綾香, 出口幸子: 楽譜データベースを用いた日本ポピュラー音楽の旋律分析, 情処学報, MUS-65, 2006.