

音楽情景分析における単音遷移に関する 統計的情報を利用した単音認識処理

木下 智義 村岡 秀哉 田中 英彦

東京大学大学院 工学系研究科

{kino,hideya,tanaka}@mtl.t.u-tokyo.ac.jp

筆者らは、知覚的な音の階層的な内部像を構築するために、これまでに音楽情景分析の処理モデル OPTIMA を提案し、その実験システムを実装した。しかし、そのシステムは認識率の面で実用的とは言えず、その改善が課題となっている。OPTIMA は、複数の処理モジュールからの出力を統合して外界の最尤推定像を得る枠組であるため、処理モジュールの追加によって認識精度の向上を図ることができる。本稿ではまず単音の遷移に注目し、単音の遷移に関する統計的分析を行って知識源を構成した。次に、単音の遷移に関する確率値を与える処理を提案した上で、OPTIMA に追加した。評価実験の結果、単音の認識誤りの訂正に成功し、処理の有効性を示すことができた。

音楽情景分析, 単音遷移, 聴覚的情景分析, 確率的情報統合

Note recognition using the statistical information about note transition

Tomoyoshi Kinoshita Hideya Muraoka Hidehiko Tanaka

University of Tokyo

{kino,hideya,tanaka}@mtl.t.u-tokyo.ac.jp

Previously, we have proposed a processing model OPTIMA for music scene analysis, and implemented its experimental system. However, its recognition accuracy was not practical. The OPTIMA processing architecture is the framework where multiple sources of information are integrated. Due to this feature, we can improve its accuracy by adding new processing modules.

The objective of this paper is improvement of recognition accuracy using statistics of note transition. In this paper, first we analyze the transition of notes and construct a knowledge source for OPTIMA. We then propose a novel process which gives probabilistic information about note ptransition. Experimental results show the effectiveness of this process.

Music scene analysis, Note transition, Auditory scene analysis, Probabilistic information integration

1 はじめに

筆者らは既に、知覚的な音の内部像を構築するために、音楽情景分析の処理モデル OPTIMA を提案し、その実験システムを実装している [1, 2]¹。しかしながら、現在の OPTIMA は、認識率において実用的なレベルには達しておらず、その改善が課題となっている。

OPTIMA は、複数の独立した処理モジュールを用意し、それぞれが入力データに対し処理を行って、確率をもった仮説の組と、これらの仮説の組みの間の条件つき確率を出力する。これらを用いて仮説ネットワークを構成する。仮説の組はノードとして、条件つき確率はリンクとして表される。その後、確率伝搬によって確率情報を統合することにより、外界の音響的事象に関する最尤推定像を求める。この様に、OPTIMA を入力を多角的に解析し、その結果を統合するという特徴を持つ。そのため、処理モジュールを追加することによって認識精度の向上を図ることができる。

一方、一般に、楽曲における単音の遷移には、ある程度の法則が存在すると言われている。実際、単音の遷移パターンに関する研究がなされており、遷移パターンから調性を判別する試みなども行なわれている [3]。

そこで本稿では、楽曲データを解析し、その結果から定量的に単音の遷移に関する情報を抽出して、それを OPTIMA による音楽情景分析処理へ応用することにより、処理精度の向上を図る。

2 楽曲における単音の遷移

初めに、楽曲における単音の遷移のパターンに関する情報を抽出するために、単音列データを解析した。

解析に用いたデータは、日本のポピュラー音楽を主とする 311 曲の MIDI データである。全単音数は 26,282,277、1 曲あたり 84,509 単音である。

本稿で抽出した情報は以下の通りである。

- 単音の出現頻度
- 単音の遷移幅の出現頻度
- 単音の遷移パターンの出現頻度

以下、それぞれに関して詳しく述べる。

¹OPTIMA に関するソフトウェアはフリーソフトとして公開されている。連絡先は以下の通り。

E-mail: optima@mt1.t.u-tokyo.ac.jp

URL: <http://www.mt1.t.u-tokyo.ac.jp/~optima>

2.1 単音の出現頻度

各曲について、単音の出現頻度を音高ごとに調べた。

単音の音高は、その曲の調性に大きく依存するため、全体でそのまま統計をとることにあまり意味がない。そこで、予め調性によって正規化した上で統計をとった。

正規化は、それぞれの調性に関してドの音が 0 となるように、定数を引いた上で音高を 12 で割った。

この正規化した音高値を用いて単音の出現頻度を調べた(図 1)。ただし、ここで言う出現頻度は、単純な出現回数ではなく、その音の継続時間の和を表す。

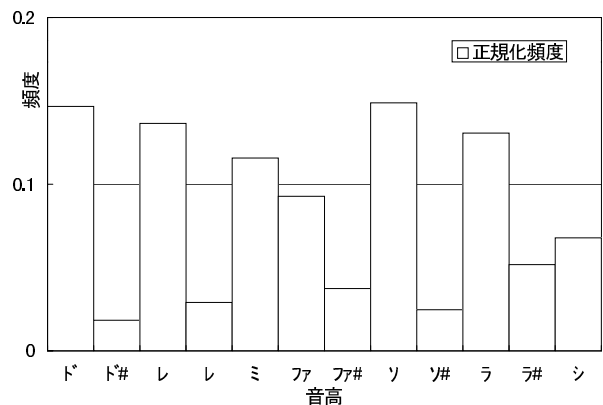


図 1: 単音の出現頻度

出現頻度を見ると、その大きい音高と小さい音高との差が非常に大きい。

そこで、入力された単音列の単音出現頻度をと比較することで、入力の曲の調性の認識にも応用できると考えられる。

2.2 単音の遷移幅

次に、パートごとに単音の遷移幅の出現頻度を調べてみた。

ここでいう単音の遷移幅とは、連続する単音の音高の差を、半音を 1 として数えたもので、1 オクターブずれると遷移幅は 12 である。

結果のグラフが図 2 である。

解析結果を見ると、9 割の遷移は、遷移幅が 1 オクターブ(つまり、12 単音)以内であることがわかる。

これによって、遷移幅の大きすぎる遷移に関しては、その確率が低いと判断できる。

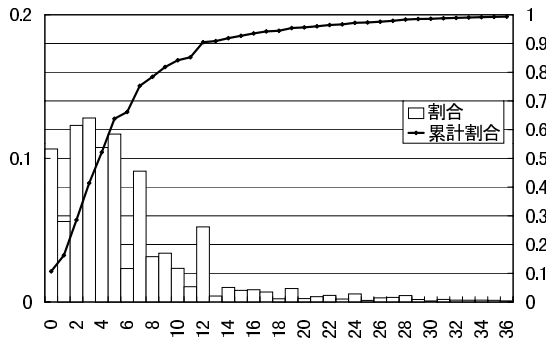


図 2: 単音の遷移幅の出現頻度

2.3 単音の遷移パターン

単音列 MIDI ファイルから、単音遷移の頻度を求めた。ここで、単音遷移は前項のように、調性ごとに正規化した値で、各遷移パターンの出現回数をカウントした(表 1)。

単音番号	音階	出現数(正規化値)
0→0	ド→ド	1.000
0→9	ド→ラ	0.946
0→7	ド→ソ	0.844
4→0	ミ→ド	0.837
9→9	ラ→ラ	0.794

表 1: 単音遷移パターンの出現回数上位 5 位の表

当然ながら、出現頻度の大きいド・ミ・ソ・ラといった音階の遷移が多くなっている。本稿ではこのデータを参照して、その遷移に関する条件つき確率を与えるものとする。

3 単音遷移情報付与処理の概要

本稿では、前項の結果を踏まえて、解析結果を知識源データとして用いる処理モジュールを新たに提案する。

提案・実装する単音遷移情報付与モジュールは、以下のような流れで処理を行なう。

最初に処理の概念図を図 3 に示す。

3.1 遷移幅確率を与える

予め MIDI データ群から統計的処理によって遷移幅確率データを用意しておく。

これを、遷移幅とその出現頻度の組として格納しておき、処理における知識源として用いる。

入力された単音の遷移幅(音高値の差)に応じ、上記確率データを参照して遷移幅確率値を得る。

3.2 遷移パターン確率を与える

1. 入力された単音仮説の組は、最初に履歴記憶に格納される。また、それとは別に、予め MIDI データ群から統計的処理によって調性別単音出現確率データを用意しておく。
2. 履歴データに含まれる単音の音高別の出現頻度と、調性別単音出現確率データと照らし合わせることで、処理を行なう曲の調性を判断する。
3. 得られた調性名に応じ、遷移パターン確率データを用意する。
4. 入力の遷移パターンに応じて、用意した確率データを参照することによって、遷移パターン確率を与える。

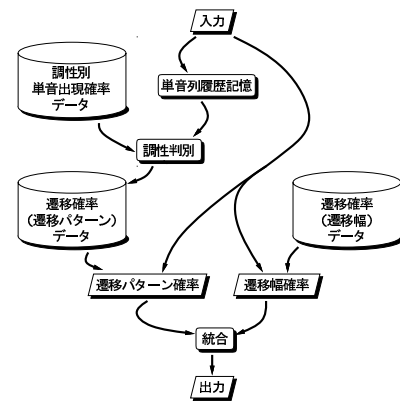


図 3: 単音遷移情報付与モジュールの処理の概要

3.3 結果の統合

最後に、得られた上記 2 つの確率値を統合することで全体の遷移確率を得る。統合は、本稿では上記の確率値の積を与えることによって行なっている。

3.4 複旋律への対応

前項で示した処理は、入力として与えられた単音仮説が、それぞれ1つの単音のみからなる場合の処理に相当する。このように、各単音仮説が1つの単音のみからなるような場合を「単旋律」と呼ぶことにする。

現在、OPTIMAにおいては、一般に単音仮説として1つの仮説には複数の単音が格納されている。単旋律に対し、このような1つの単音仮説に複数の単音からなる場合を「複旋律」と呼ぶ。本稿で提案する処理においてはこのような複旋律のデータに対しても処理を行なう。

複旋律を対象とした処理においては、入力として与えられた2つの単音仮説に対し、以下の手順で処理を行なう。

1. 入力の2つの単音仮説から、同一パートと見做せる単音の組を抽出する。
同一パートと見做すための条件は以下の通りである。
 - 同じ音源名である。
 - 2つの単音に関し、時間的な重なりがない。
2. このようにして得られたパート毎に、前項の処理を行ない、得られた処理結果を統合する。ここでは確率値の積を計算するものとする。
3. パートの組合せが複数考えられる場合には、それぞれについて同様の処理を行ない、確率値の最も大きいものを最終的な出力結果とする。

4 実装

本稿で提案した処理は、OPTIMA 実験システムの処理モジュールとして実装した。このモジュールは、仮説ネットワークモジュールからの処理要求に応じ、時間的に連続する単音レベルの2つの仮説ノードが保持する仮説の内容を入力として得、それらの間の遷移が起こる確率を出力する。出力された結果は、仮説ネットワークモジュールが受け取り、それを仮説ノード間のリンクとして実現する。

出力されたデータは、OPTIMA の確率的情報統合機構により、他の処理モジュールからの出力結果と統合される。

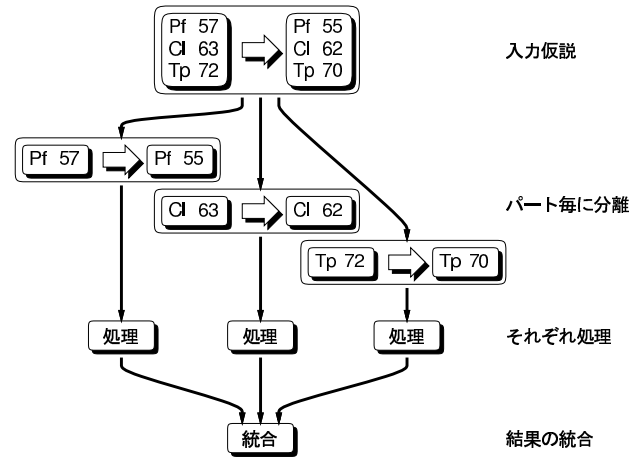


図 4: 複旋律に対する処理

5 評価

本稿で提案した手法に関し、評価実験を行なった。

5.1 評価実験の方法

OPTIMA 実験システムに単音遷移情報付与モジュールを組み込み、処理を行なった。実験は、単音遷移情報付与モジュールを組み込んだ場合と、組み込まなかった場合の2通り行ない、その結果を比較した。

評価実験の流れは以下の通りである。

正解の準備

サンプル楽譜(「蛍の光」室内楽アンサンブルの、3パートからなるもの)から得られる単音列を正解として用意し、時間(処理単位)毎にまとめておく。これが正解仮説になる。

ノイズの重畳

各時間(処理単位)に対し、正解の単音集合の一部を変えた誤り仮説をいくつか用意する。ここでは、1つの単音の音高もしくは音源名を変更したものを用意した。

ノイズの重畳は、パートのうちからランダムに1つを選び、更に乱数を用いて音源名を変えるか、もしくは上下1オクターブの範囲内で音高を変えることを行なった。

入力データの準備

用意した正解仮説と誤り仮説を混ぜて、確信度がある程度ランダムに、正解仮説のそれが高くなりやすいように与えておく。本稿において、確信度は、OPTIMA システムにおける実際の出力結果を想定して、最初に正解仮説に確信度 1 を与え、以下誤りの仮説には 0.8, 0.6, … のような確信度を与えた。更に乱数を用いてそれに変動を加えた。

このようにすると、正解仮説が最上位仮説となる箇所が多くなるが、誤り仮説が最上位になる箇所もある程度見られるようになる。

処理実行および評価

以上のように用意された仮説に対し、今回作成した単音遷移情報付与モジュールの処理を行ない、情報統合を行なった上で、正解仮説が最上位に来た割合を計算した。

単体の評価実験は 3 種類行なった。その 3 種類は以下に述べる通りである。

1. 単旋律評価

最初に、正解データのうちから伴奏パートの単音列のみを取り出し、それをを用いて処理を行なった。この実験では、パートの追跡処理の必要がないので、単音遷移確率の付与に相当する部分に対する評価となる。

2. 複旋律評価

続いて、正解データのうちの全パートについて単音列を取り出し、それによって評価実験を進めた。この実験では単旋律で行なった評価に加えて、パート追跡処理も必要になるため、この処理が全体の処理に及ぼす影響も結果に反映されるものと予想できる。

3. 処理別評価

最後に、単音遷移情報付与モジュールの処理において、単音遷移幅情報および単音遷移パターン情報のいずれかのみを用いて処理を行なった。これにより、各処理の特性を明らかにする。

5.2 評価実験の結果および考察

5.2.1 単旋律評価

図 5 に結果を示す。

グラフにおいて、横軸はノイズ重畳におけるパラメータである。また、いずれも 95% 信頼区間は 1% 以内にとどまる。横軸は、初期確信度へのノイズ重畳の際のパラメータで、これが大きい程重畳させるノイズが多くなる。縦軸は、正解の仮説が最上位に現れた割合(正答率)である。

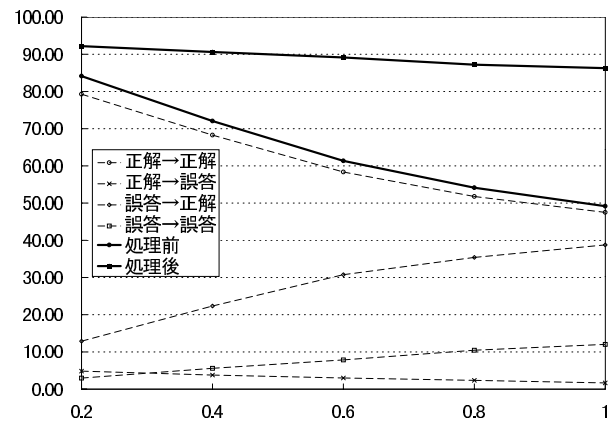


図 5: 単体評価：単旋律の結果

「誤答→正解」とは、処理前に誤りであったものが処理を加えたために正解へ訂正された箇所の割合で、その他も同様の意味である。

グラフを見ると、特に重畳しているノイズが多い場合に正答率について大幅な改善(最大で 40% 程度)が得られていることが分かる。このことは、正解仮説が出力されているもののその確信度が低いという状況において、今回提案した単音遷移情報付与モジュールを追加することによって、認識結果の改善が行なえることを示している。従って、本稿において提案した処理が有効なものであると結論できる。

また、「正解→誤答」あるいは「誤答→正解」のみを見ても、ノイズが多い方が望ましい結果となっていることが分かる。

5.2.2 複旋律評価

図 6 に結果を示す。

グラフに記載された数値の意味は単旋律評価のものと同じである。

単旋律に対する実験結果と較べて認識率の上昇は少ないものの、それでも最大で 30% 程度の認識率の上昇を示している。

この実験により、本稿において提案した処理が、複旋律のデータに対しても有効なものであると結論できる。

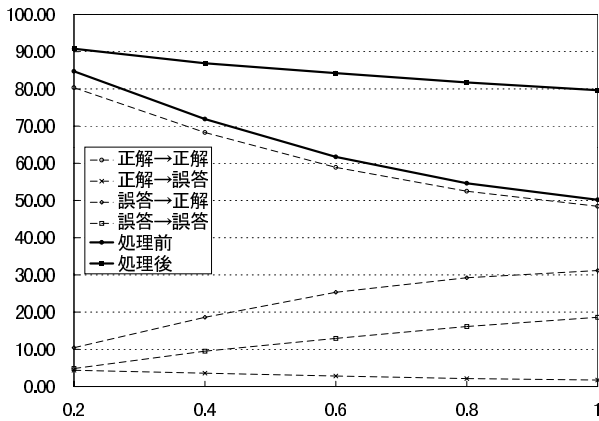


図 6: 単体評価：複旋律の結果

単線率評価の結果に比べて正答率の上昇が鈍った原因としては、ノイズの重畳により音源名が変更された結果、パートの追跡を誤ったことが考えられる。

5.2.3 処理別評価

処理内容のうち、単音遷移パターン情報のみを用いた場合、単音遷移幅情報を用いた場合の 2 つの結果、及び通常 of 処理を行なったもの、さらに処理前の正答率を同じグラフに表したものを図 7 に示す。

グラフに記載された数値の意味は単旋律評価のものと同じである。

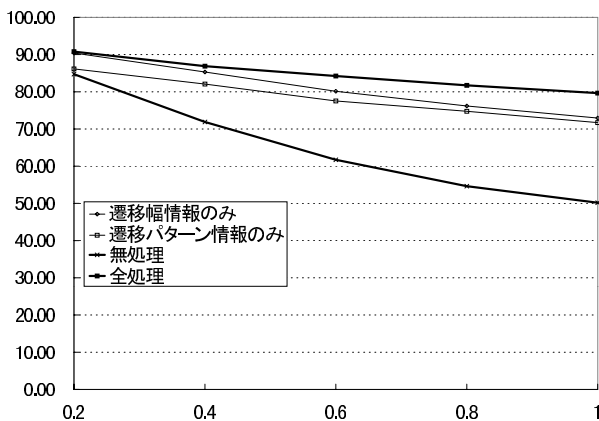


図 7: 単体評価：処理内容による差

実験結果を見ると、両方の処理を施した場合に比べ、遷移パターン情報のみを用いた処理は、正答率が 10% 程度下回っている。このことは、今回実験に用いたようなデータについて、遷移パターン情報のみでは改善できない箇所が約 10% あることを示して

いる。

さらに、ノイズパラメータが低い部分において、両方の処理を用いた場合と、遷移幅情報のみを用いた場合の処理の正答率がほぼ等しい値となっている。これにより、低ノイズ時には遷移幅情報のみによる処理でも十分な効果が得られることが結論づけられる。

逆に、ノイズが多い場合には遷移パターン情報を用いた処理によってさらなる効果が得られることが分かる。

6 おわりに

本稿では、楽曲データを解析し、その結果から定量的に単音の遷移に関する情報を抽出し、それを OPTIMA による音楽情景分析処理へ応用することの有効性を示すことができた。

一方、本来は単音の遷移は時間的に連続したもののみならず、その前後のある程度の時間における遷移に依存するものと考えられる。しかしながら、本稿においては、OPTIMA が隣接しないノード間の独立性を仮定しているためなどの理由により、このような幅を持たせた遷移確率の付与ができなかった。今後の研究においては、このような問題の解決もその課題となるだろう。

また、本稿のような局所的な遷移に注目するだけでなく、メロディーなどの大域的な情報を利用することでより精度の高い処理が期待できる。

謝辞

本稿で用いた単音列データを提供して下さった、神戸コンサルティング社の縄谷正明氏に感謝します。

参考文献

- [1] 柏野, 中臺, 木下, 田中. 音楽情景分析の処理モデル OPTIMA における単音の認識. 電子情報通信学会論文誌, J79-DII(11):1751-1761, 11 1996.
- [2] 柏野, 木下, 中臺, 田中. 音楽情景分析の処理モデル OPTIMA における和音の認識. 電子情報通信学会論文誌, J79-DII(11):1762-1770, 11 1996.
- [3] 松田, 秋山. 単旋律の進行パターンに基づく調性判別と主音推定. 日本音響学会誌, 52(4):253-260, 1996.