

# 音楽情景分析における予測に基づく単音の認識

3M-5

木下 智義\* 村岡 秀哉 田中 英彦

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

東京大学大学院 工学系研究科†

## 1 はじめに

これまでに我々は、人間が音楽を聞いてそこに含まれる音符列や和音名などの情報を抽出する機能を音楽情景分析と呼び、その計算機上での実現に向けて検討を重ねてきた。既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構築している [2, 3]。

OPTIMA では、複数楽器による演奏のモノラル音響信号から音符列、和音名、リズムなどを抽出・認識して出力する。しかしながら、その認識精度は実用上十分とは言えず、その改善が課題となっている。

この問題に関して、我々は既に単音の認識における時間的情報の利用の一面として、単音遷移の統計的情報の利用に関し検討を行った [1]。しかしながら、単音の局所的な遷移に関する処理のみでは十分ではなく、より大局的な処理を行うことが課題の 1 つとなっていた。

本稿では、単音の認識におけるより大局的な時間的情報の利用として、単音の履歴を利用した単音の予測を行う。

## 2 単音の予測

一般に、人が音楽を聞く時に単音を予測し得る手がかりとして、以下に示すものがあるものと考えられる。

1. 繰り返しによる、同一メロディーの出現
2. コード進行に従う音符の遷移
3. 盛り上がりなど、曲に含まれる情感

番号が小さい方が記号処理的で、大きいほど知識処理的になる。

計算機上でこれらの処理を実現することを考える。3. は情感の抽出が可能となったとしても、情感と単音とが直接関連を持たないものであるため、そこから単音の予測を行うことは困難である。一方 1. および 2. は、情感等の感性的な情報を抽出する必要がなく、また繰り返し部分の抽出やコードの予測が可能となれば比較的容易に単音の予測を行える。本研究ではこのうち 1. の繰り返しに注目した処理について検討する。

## 3 繰り返しに注目した予測処理

本稿では、前項で述べた予測機構のうち、曲の繰り返しに注目した方法を用いた単音の予測を行う。

本手法では、以下のような手順により単音の予測を行う。

1. 最初に予測を行う時刻を与え、そこから時間的に前の方向へ一定数の音符を抽出する。これを予測テンプレートと呼ぶことにする。(図 1(a))
2. 続いて、それ以前の全ての時刻に対して、予測テンプレートと実際の音符データの類似度を計算する。(図 1(b))
3. 最後に、得られた類似度の最も大きかった音符から予測される音符を得る。具体的には、予測テンプレートとの類似度が最も大きかった箇所の次に位置する音符を予測結果として得る。(図 1(c))



図 1: 予測テンプレートによる単音の予測

ここで言う類似度は、テンプレートから予測される時間的位置にある音符を対象として、以下の点について計算し、その重みをつけた和を計算用いた。

- それぞれの音符の音源名の一致
- 音高の遷移パターンの一致
- それぞれの音符の開始時刻と継続時間の一致
- 予測位置との時間的距離

計算を行う音符として、各時刻に関して複数の候補を抽出する。これにより、同時に複数の音符が存在するような複数パートの曲に対しても処理を行うことができる。

\* 日本学術振興会特別研究員

† "Music note recognition based on prediction of notes"  
Tomoyoshi Kinoshita, Hideya Muraoka and Hidehiko Tanaka

University of Tokyo, Graduate School of Engineering,  
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

### 3.1 評価実験

今回提案した手法に関し、実際の楽譜データを対象に実験を行った。

実験は、楽譜（「蛍の光」の室内楽アンサンブル編曲版）の各音符について、それ以前の音符列を元に予測を行い、実際の音符と一致したものを正解として再現率と適合率を計算した。実験結果を表1に示す。

#### 3.1.1 予測精度の評価

表1: 予測機構単体の評価

パート	音符数	再現率	適合率
旋律	59	28.8%	81.0%
伴奏	122	2.5%	50.0%
両方	181	11.1%	74.1%

再現率の低さは、この手法が曲の繰り返しに注目しているために、繰り返してない箇所での予測ができないことに起因する。実際、旋律パートで繰り返されている箇所は音符数で24であり、これを認識対象とすれば、再現率87.5%となる。

#### 3.1.2 繰り返しが多い場合の評価

繰り返しの多い曲を想定し、同じ曲を2度続けた場合の結果を表2に示す。

表2: 2度繰り返す場合の結果

パート	音符数	再現率	適合率
旋律	118	45.8%	85.7%
伴奏	244	48.0%	95.9%
両方	362	47.8%	92.5%

この場合、繰り返しとなっている音符は全部で205であり、再現率は84.4%となる。これにより、今回提案した手法が、繰り返しを含む曲での単音予測に関して十分な精度を持つことが分かる。

#### 3.1.3 音符が欠落している場合の評価

実際に音楽の認識処理を行っている場合には、一部の音符を認識できないことが多い。そこでそのような場合を想定し、予め音符列から一定の割合の音符を取り除いたものを対象として同様の実験を行った。その結果を表3に示す。ここでは10%に相当する音符をランダムに取り除いて実験した。

表3: 音符の欠落がある場合の結果

パート	音符数	再現率	適合率
旋律	51	15.7%	72.7%
伴奏	108	2.8%	30.0%
両方	173	6.1%	62.5%

音符の欠落により、再現率・適合率がともに低下して

いるのがわかる。これは、予測テンプレートとの類似度の計算において、音符が欠落している箇所での類似度が低くなり、予測され得る単音が予測結果として出力されなかったことが原因として挙げられる。

#### 3.2 実際の処理結果に対する予測の評価

最後に、これまでのOPTIMAによる単音認識結果を用いた評価を行った。これまでのOPTIMAの認識結果では、単音認識での誤りが多く、本稿で示す手法によってその一部を補完することを目的とする。その結果を3に示す。

表4: 音符の欠落がある場合の結果

パート	音符数	再現率	適合率
両方	171	9.9%	77.3%

出力されたもののうち、正しく出力されたいものは、全て伴奏パートのものであった。これは、OPTIMAでは高い音である旋律パートの単音を認識できない場合が多く、そのため予測が行えなかったことが原因であると考えられる。

### 4 おわりに

本稿では、音符列の繰り返しに注目することで出現する単音を予測し、単音の認識処理に応用する手法について検討した。実験の結果、良好な精度で単音の予測が可能であることが明らかとなった。一方で、音符の欠落など不完全な履歴を元に予測する場合には、予測精度が大幅に悪化するという問題があり、今後の課題として挙げられる。

今回は、実験対象を1つの曲にのみ限ったが、今後はより多くの曲を対象とした実験により予測手法の特性を解析することが課題の1つとなる。また、本稿で提案した手法では、計算量が音符数 $n$ に対して $O(n^2)$ となるため、効率的な枝刈り方法などの検討も行う必要がある。

### 謝辞

本研究は、文部省科学研究費補助金（課題番号09-07628）による研究成果の一部である。

### 参考文献

- [1] 木下智義, 村岡秀哉, 田中英彦. 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. 電子情報通信学会技術研究報告, 音声研究会, SP96-116, Vol. 96, No. 540, pp. 15-20, March 1997.
- [2] 柏野邦夫, 中臺一博, 木下智義, 田中英彦. 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, Vol. J79-DII, No. 11, pp. 1751-1761, 11 1996.
- [3] 柏野邦夫, 木下智義, 中臺一博, 田中英彦. 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, Vol. J79-DII, No. 11, pp. 1762-1770, 11 1996.