

単音の遷移に注目した単音認識処理

Note recognition using statistical information of musical note transitions

木下 智義 村岡 秀哉 田中 英彦

平成10年 2月 25日

1 はじめに

著者らは既に、音楽情景分析 (music scene analysis) の処理モデル OPTIMA (Organized Processing Toward Intelligent Music-scene Analysis) を提案し、その実験システムを実装している^{4,5)}。音楽情景分析とは、音楽音響信号を入力とし、各楽器の演奏情報 (単音や和音など) を記号表現として出力する情報処理を指す。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 では、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、確率をもった単音や和音などの仮説の組と、これらの仮説の組の間の条件つき確率を出力する。続いて、これらを用いて仮説ネットワークを構成する。仮説ネットワークにおいて、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 は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。それぞれの結果は緩い制約として統合されるため、処理結果の誤りを互いに補完しあうことができる。従って、処理モジュールを追加することによって単音認識の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、ある程度の傾向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われている⁶⁾。しかしながら、単音の遷移パターンを単音の認識そのものに応用する研究は少ない。

OPTIMA においても、仮説ネットワークに構造上の制約があったために²⁾、これまでは単音の遷移に関する情報を用いた処理は行われていなかった。著者らは OPTIMA について構造上の制約に関する検討を進めている²⁾が、それとは別に構造上の制約が解消された場合において、単

音の遷移に関する情報がどの程度単音の認識に有効であるかを明らかにすることも重要な課題である。そこで本論文では、楽曲データを解析し、その結果から単音の遷移に関する情報を定量的に抽出して、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 OPTIMA における音楽情景分析処理

著者らが既に提案している音楽情景分析の処理モデル OPTIMA の全体像を Fig. 1に示す⁴⁾。

OPTIMA は、入力として、複数楽器による演奏のモノラル音響信号をとる。現在は、簡単のために入力の演奏は実楽器の音を収録したサンプリャによって MIDI データを用いて手動の打ち込みによって行っている。ピッチバンドは用いられていない。また、対象とする楽曲は、曲を通じて一定の調性が保たれていることを仮定しており、特に本論文の範囲においてはあらかじめ調の長短が与えられているものとする。

OPTIMA では、各時点で得られた入力音響情報に基づいて、周波数成分 (frequency component)、単音 (musical note)、および和音 (chord) についての仮説を生成し、事後確率最大を評価基準として、全体として最も尤もらしい仮説の組を逐次求めていく枠組である。Fig. 1 に示すシステムの入力はモノラルの音楽音響信号であり、出力は、和音記号の列、楽器ごとに分類された単音記号の列、楽器ごとに分類された周波数成分の組、および拍位置を表す記号列である。これらの記号列は、音楽演奏に対する「知覚的な音」³⁾に相当する。なお、出力される周波数成分を元に、楽器ごとの音響信号波形を再合成することも可能である。

処理モデル OPTIMA は、前処理部 (preprocesses)、主処理部 (main processes)、知識源 (knowledge sources)、

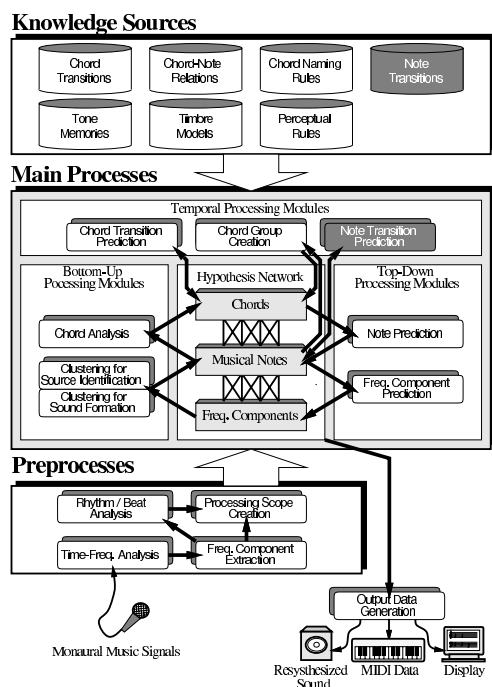


Fig. 1: 処理モデル OPTIMA の全体像

および出力データ生成部 (output data generation) の 4 つの部から成る⁴⁾。

主処理部は、音響事象の仮説を保持するためのベイジアンネットワーク (仮説ネットワーク; hypothesis network) を備えている。仮説ネットワークは、1. 周波数成分、2. 単音、3. 和音の 3 つの抽象度の階層をもっている。仮説ネットワークに対して、(a) 抽象度の低い階層の情報に基づいて抽象度の高い階層の情報を与えるボトムアップ処理モジュール (bottom-up processing modules) (b) 抽象度の高い階層の情報に基づいて抽象度の低い階層の情報を与えるトップダウン処理モジュール (top-down processing modules) (c) 時間の推移に関する情報を扱う処理モジュール (temporal processing modules) の 3 つの群に分けられる処理モジュールが情報を書き込む。

仮説ネットワークの特徴は、ネットワークに対し外部から情報が与えられるごとに、それまでに生成されている各仮説に対する確信度ベクトル (事後確率ベクトル) $BEL(n)$ が自動的に求まる点である (ここで n はあるノードにおける仮説の組を表す)^{4, 1)}。

(a) ボトムアップ処理モジュールとしては、周波数成分の情報を基に単音の情報を生成する処理 (単音仮説生成; sound formation および source identification)、単音の情報を基に和音の情報を生成する処理 (和音仮説生成;

chord recognition) の 2 つがある。

(b) トップダウン処理モジュールとしては、和音の情報を元に単音仮説の確からしさに関する情報を出力する処理 (和音構成音情報付与; note prediction)、単音の情報を元に周波数成分仮説の確からしさに関する情報を出力する処理 (単音構成周波数成分情報付与; frequency component prediction) の 2 つがある。

(c) また、時間方向の処理モジュールとしては、和音の推移に関する情報を出力する処理 (和音遷移情報付与; chord transition prediction)、時間的に連続する何個の処理単位が 1 つの和音を形成するかに関する情報を出力する処理 (和音グループ生成; chord group creation) の 2 つがある。

これまで OPTIMA においては仮説ネットワークは木構造でなければならず、そのため単音レベルで時間方向の情報を与えることはできなかったが、この問題に関して検討を進めた結果、単音遷移に関する情報の付与が可能となった²⁾。

そこで本論文では、単音の遷移に関する情報を出力する処理 (単音遷移情報付与; note transition prediction) を提案し、OPTIMA に時間方向の処理モジュールとして追加するものである。

本論文では以下、単音遷移情報付与に関する処理において用いられる楽曲に含まれる単音遷移に関する統計的信息について述べた上で、この情報を用いて遷移に関する情報を与える処理をどのように行うかを順に議論する。

3 楽曲における単音の遷移

本論文では、初めに単音列データを解析することにより、楽曲における単音の遷移のパターンに関する以下の情報を抽出した。

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。このデータには旋律・伴奏を含めた全パートが収録されている。また繰り返しのある曲に関しては、繰り返しの全てが含まれる。全単音数は 2,297,995、1 曲あたり平均 7,389 単音である。なお、1 つの曲に対して複数の演奏が含まれていることはない。

単音の出現頻度

楽曲において単音が出現している時間の統計をとった。

単音の遷移幅の出現頻度

時間的に連続する2つの単音について、その遷移における音高の差を遷移幅として抽出し、その幅ごとの出現頻度を調べた。

単音の遷移パターンの出現頻度

単音の音高を調性によって正規化した上で、遷移の前後の音階名の組を遷移パターンとして抽出し、パターンごとの出現頻度を調べた。ここでの調性は平行調を区別している。

以下、それぞれに関して詳しく述べる。

3.1 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。

単音の音高の出現頻度は、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はあまり意味がない。そこで、あらかじ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主音が0となるように、音高を表すMIDIノート番号をから定数を引いた上で12で割った余りをとることで行った。ここでの調性は平行調を区別している。

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 (Fig. 2)。ただし、ここで言う出現頻度は、単純な出現回数ではなく、その音の発音時間の和を表す。

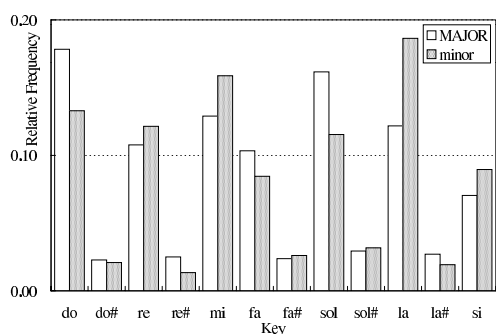


Fig. 2: 単音の出現頻度

結果を見ると、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比較することで、調性の長短が既知である入力の曲の調性の認識に応用できると考えられる。

3.2 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オクターブのずれが遷移幅12に相当する。

結果のグラフが Fig. 3である。

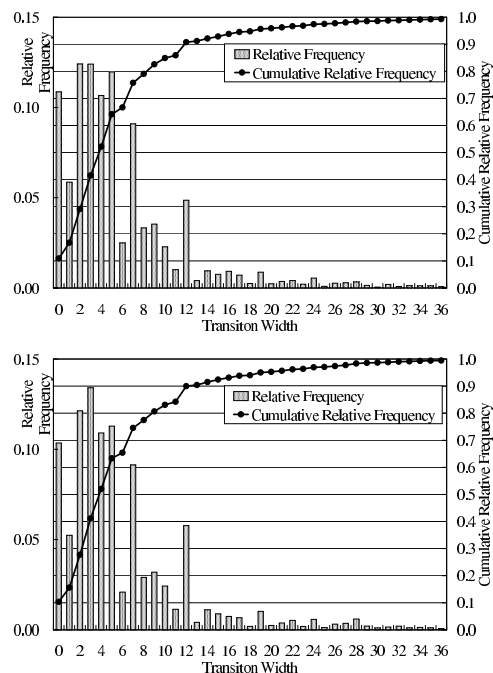


Fig. 3: 単音の遷移幅の出現頻度 (上: 長調、下: 短調)

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半音) 以内であることがわかる。

この結果を参照して、単音の遷移に対してその遷移幅に応じた確率値を与える。これにより、仮説生成の段階で音高の認識を誤ったことによる大きな遷移幅に対して小さな確率を与えることができ、音高の誤りの補完が可能になる。

3.3 単音の遷移パターン

単音列MIDIファイルから、単音遷移パターンの出現頻度を求めた。単音遷移パターンとは、音高の、調性ごとに正規化した値について、各遷移の前後の音高の組を指す。ここでは、単音遷移パターンの出現回数をカウントした (Table 1)。

単音出現頻度の大きいI、III、V、VI度といった音階の遷移が多くなっている。本論文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。

Table 1: 単音遷移パターンの出現回数上位 5 位の表

| 長調 (上行) | | 短調 (上行) | |
|---------|-------|---------|-------|
| 音階 | 出現数 | 音階 | 出現数 |
| I → I | 1.000 | I → I | 1.000 |
| V → V | 0.896 | I → III | 0.932 |
| VI → I | 0.786 | V → III | 0.680 |
| I → III | 0.720 | V → V | 0.663 |
| III → V | 0.667 | I → V | 0.600 |

「出現数」は、最上位の出現回数で正規化した値である。

4 単音遷移情報付与処理の概要

本論文では、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単音遷移情報を付与する OPTIMA の処理モジュールを新たに提案する。本節では、単音遷移情報付与モジュールの処理について述べる。

最初に処理の概念図を Fig. 4 に示す。

4.1 遷移幅確率を与える

あらかじめ MIDI データ群から統計的処理によって遷移幅確率データ (3.2項) 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く、処理を行う際の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。

4.2 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 (3.1項) 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、処理を行う曲の調を判定する。
3. 得られた調名に応じ、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 (3.3項) を参照することによって、遷移パターン確率を与える。

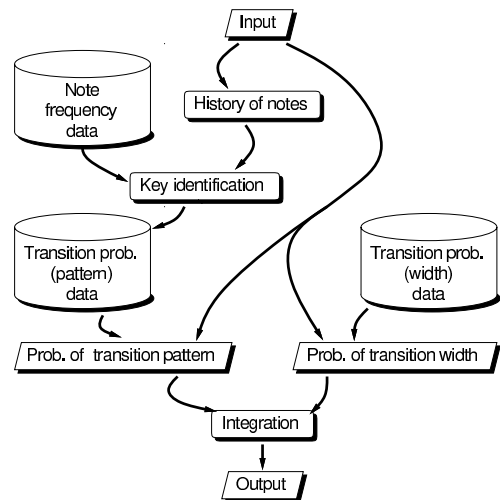


Fig. 4: 単音遷移情報付与モジュールの処理の概要

4.3 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、本論文では上記の確率値の積を与えることによって行っている。

4.4 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、各単音集合仮説が 1 つの単音のみからなるような場合を「単旋律 (single melody)」と呼ぶことにする。

現在、OPTIMA においては、一般に 1 つの単音集合仮説には複数の単音が格納されている。このような、1 つの単音集合仮説が複数の単音からなる場合を、単旋律と対比して「複旋律 (multiple melody)」と呼ぶ。本論文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単音集合仮説の組に対し、以下の手順で処理を行う (Fig. 5)。

1. 入力の単音仮説の組から、同一パートとみなせる単音の組を抽出する。

同一パートとみな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い、確率値の最も大きいものを最終的な出力結果とする。

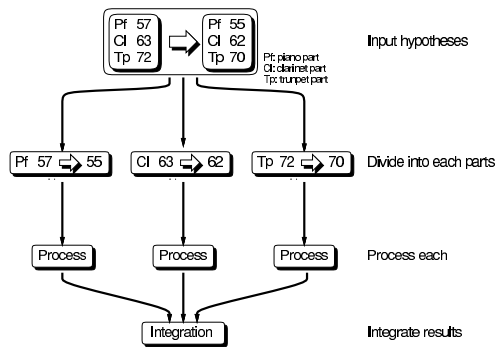


Fig. 5: 複旋律に対する処理

5 実装

本論文で提案した処理を、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⁴⁾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。

この処理においては、OPTIMA の仮説生成モジュールが出力した単音に関する確信度つき仮説と、本論文で実装するモジュールが出力する条件つき確率を統合し、仮説に付随する確信度を更新する。これによって仮説生成モジュールの出力に含まれる誤りを補完するものである。

6 評価

本論文で提案した手法に関し、評価実験を行った。

6.1 実験 1: 単体評価

OPTIMA 実験システムには組み込まず、あらかじめ用意した単音仮説データに対して処理を行い、その結果と用意した単音仮説データを統合することにより、処理による正答率の変化を調べた。

評価実験の流れは以下の通りである。

1. 正解の準備

「蛍の光」の室内楽アンサンブル⁵⁾の楽譜データから得られる単音列を正解として用意し、時間毎にまとめておく。これが正解仮説になる。それぞれの時間には、最大 3 つの単音が含まれる。

2. ノイズの重畳

各時間に対し、正解の単音集合の一部にノイズを加えた誤り仮説をいくつか用意する。本実験では、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、更に一様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った。

3. 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。本実験においては OPTIMA 実験システムにおける実際の出力結果を想定して、正規乱数を用いて確信度に揺らぎを与えた。

実際には、仮説 $H_i (i = 1 \dots 5, H_1$ が正解仮説) に対し、確信度 $P(H_i) = 1 / \exp(i + r)$ とした。ここで、 r は標準偏差 σ の正規分布に従う乱数値である。この σ は、後述する実験操作において、ノイズパラメータとして用いられている。OPTIMA において、確信度の値は正の値でなければならず、また複数の仮説が存在した場合には本質的には確信度の比のみが問題になる。そこで本論文では簡単に \exp を用いた。

また、この時点で正解仮説の確信度が最も大きな値となっている割合を調べ、これを処理前の正答率とする。

4. 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い、情報統

合を行った上で、正解仮説が最上位にきた割合を計算した。これを処理後の正答率とする。

単体の評価実験は以下の 5 種類行った。

6.1.1 実験 1.1: 処理の効果の評価

最初に、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、それをを用いて処理を行った。結果を Fig. 6 に示す。

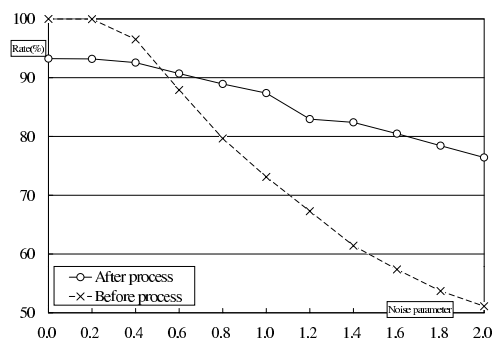


Fig. 6: 処理の効果の評価

グラフにおいて、横軸はノイズを加える際に用いたパラメータ (ノイズパラメータ) である。縦軸は、正解の仮説が最上位に現れた割合 (正答率) である。いずれの正答率も 95% 信頼区間は 0.5% 以内にとどまる。これらに関しては、以降の実験に関しても同様である。

処理前の正答率と処理後の正答率を比較すると、最大で約 25% の正答率の向上が見られる。加えるノイズが大きい程効果が高い傾向にあることがわかる。これによって、本論文で提案した処理が有効であることが示された。

一方、加えられたノイズが少ない場合に、処理後の正答率が処理前のものよりも低くなっている。これは、統計的に得られた遷移パターンに合致しなかった仮説が排除されたためと考えられる。

6.1.2 実験 1.2: 旋律数による差の評価

続いて、正解データのうちの伴奏パートのみの単音列を取り出し、それによって評価実験を進めた。結果を Fig. 7 に示す。

結果を見ると、ノイズパラメータの小さい箇所では単旋律・複旋律でほぼ同等の結果となっているが、ノイズパラメータが大きくなるにつれ、複旋律の方が正答が困難となっている。

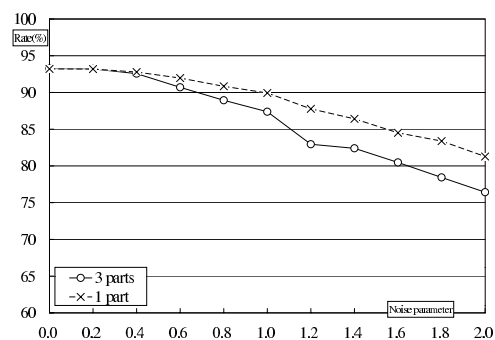


Fig. 7: 旋律数による差

複旋律のデータを対象とした処理では、単旋律のデータを対象とした場合と異なり、パート毎にデータを分割する処理を行う必要がある。そのため、パート毎への分割に失敗したために生じる誤りが多く含まれていると考えられる。

6.1.3 実験 1.3: 処理内容による差の評価

今回提案した処理においては、単音の遷移幅に関する情報と、単音の遷移パターンに関する情報を用いている。本実験ではこれら 2 つの情報のうち、一部のみを用いた場合の処理について調べる。

実験では、単音遷移情報付与モジュールにおいて上述の 2 つの情報のうち一方を用いずに処理を行った場合と、両方の処理を行った場合についてその正答率の差を調べた。結果を Fig. 8 に示す。

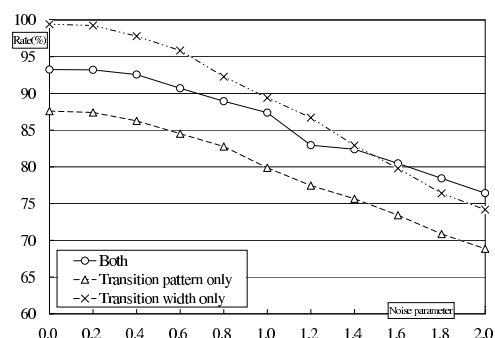


Fig. 8: 処理内容による差

実験結果から、特にノイズパラメータの大きな箇所では、単音遷移幅情報および単音遷移パターン情報の一方のみを用いた場合に比べ、両方の情報を用いた場合の方が 5~10% 程度正答率において優れている。これは、今

回提案した処理において、単音遷移幅情報と単音遷移パターン情報の両方が処理において有効に機能していることを示すものである。

一方、ノイズパラメータの小さい箇所では単音遷移幅情報のみを用いた場合に、両方の情報を用いた場合に優る正答率が得られている。従って、重量されているノイズが少ない場合には、単音遷移幅情報のみが有効に働いていると考えられる。

6.1.4 実験 1.4: ノイズの種類による差の評価

本実験では、入力として与えるデータに加えるノイズを 2 種類用意したが、ここではそれらの一方のみを与えた場合について実験を行った。

入力として、単音の音高を変更したもの、単音の音源名を変更したもの、および単音の音高と音源名を変えたもの。の 3 種類を用意した。ただし、音高と音源名を変えたものについては、音高と音源名が同時に変更されることはない。また、音高を変えたものと音源名を変えたものの比率は 1:1 である。

これらのデータにより処理を行い、その正答率の差を見た。結果を Fig. 9 に示す。

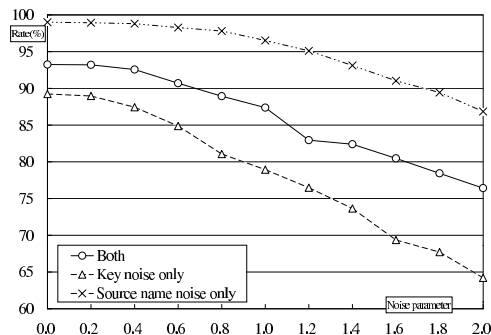


Fig. 9: ノイズの種類による差

結果を見ると、音高ノイズのみを加えたデータによる正答率が全ての場合において、3 つのうちで最低になっている。また、音源ノイズのみを加えた場合では、3 つのうちで最高になっている。これにより、本論文による処理は、音源ノイズの訂正に対してより有効に機能しているといえる。

6.2 実験 2: システム稼働評価

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 通り行い、その結果を比較した。

実験には、「蛍の光」の室内楽アンサンブル⁵⁾をサンプルを用いて演奏したデータを用いた。曲は 2 パートのもの 3 パートのものを用意した。

実験結果を Table 2 に示す。

Table 2: システム全体評価の結果

| 実験データ | 音源名 | モジュール追加前 | モジュール追加後 |
|-------|-----|----------|----------|
| 2fp | 含む | 66.5 | 67.5 |
| | 含まず | 88.5 | 89.3 |
| 3ctp | 含む | 49.6 | 49.7 |
| | 含まず | 79.7 | 79.9 |
| 3fcp | 含む | 51.5 | 51.6 |
| | 含まず | 82.8 | 83.0 |

ここで、実験データは、それぞれ以下のパートからなるデータである。

- 2fp フルート、ピアノ
- 3ctp クラリネット、トランペット、ピアノ
- 3fcp フルート、クラリネット、ピアノ

数字は認識率 (%)。認識率 R は以下の式で与えられる。

$$R = \frac{\alpha - \beta}{n} + \frac{1}{2}$$

- α 出力に含まれる正しい音符の数
- β 出力に含まれる誤った音符の数
- n 正解に含まれる総音符数

単音遷移情報付与モジュールを追加した場合と、追加しない場合を比較した場合、有意な差が見られない。その原因については以下のように考えられる。

本来、今回のように、条件つき確率を与える処理モジュールは、OPTIMA においては、他の処理モジュールとの出力結果と統合することで、誤って出力された仮説の訂正を行う効果が期待されるものである。

情報統合により、仮説に付随する確信度の更新が行われるものの、仮説の内容そのものの更新は OPTIMA の性質上行えない。そのため、仮説生成の段階で、仮説生成モジュールから正しい仮説が出力されなかった場合、その誤りを他の条件つき確率付与モジュールの出力による情報によって訂正することはできない。

今回の実験では、単音遷移情報付与モジュールを用いなかった場合における誤り箇所のうち、正解仮説が確信度が低いながらも出力されていた箇所、即ち単音遷移情報付与モジュールによって正解へ訂正される余地を残し

ていた箇所は極めて少なかった。

単体評価においては、単音認識率の大幅な改善が見られたため、システム稼働実験において認識精度の向上が少なかったのはこの原因によるものと考えられる。

また参考ながら、単音仮説生成モジュールの出力に、正解仮説を手動で含めた場合、15%程度の認識率の向上が見られた (Table 3)。このことから、本論文で提案した手法が、単音仮説生成の誤りの訂正に有効であることが明らかになった。

Table 3: 手動で正解仮説を補完した場合の結果

| 実験データ | 音源名 | モジュール追加前 | モジュール追加後 |
|-------|-----|----------|----------|
| 2fp | 含む | 72.5 | 88.0 |
| | 含まず | 91.2 | 93.3 |

数字は認識率 (%)。算出法は Table 2 参照。実験データは、フルート、ピアノのパートからなるデータである。正解仮説の初期確信度は、単音仮説生成モジュールが出力した仮説のうちで最も低い確信度を持つ仮説のものと同しくした。

7 おわりに

本論文では、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した。

評価実験において、今回検討した単音遷移情報を統合するモジュールをシステムから切り出して単体で評価したところ、最も効果が顕著な場合で約 25% 単音認識精度が向上した。

その一方で、システムに組み込んだ場合には、システム稼働評価の結果に見られたように、十分な効果が得られない可能性も存在する。手動で正解仮説を含めた場合に最大 15% 程度の精度向上が見られたことから、OPTIMA の他のモジュールの改善により精度向上を期待でき、この点を今後の課題として挙げるができる。

本論文の範囲においては調の長短が既知であることや曲を通じて調性が一定であることを仮定しており、これらの仮定により、対象となる楽曲がごく一部のものに限られている。長調、短調の判別、あるいは転調を検出するモジュールを OPTIMA へ追加して、処理の対象を広げることも検討したい。

また、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、本論文においては、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課題となる。

さらに、本論文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。

最後に、本論文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。また、本論文に関して議論の相手をして下さった NTT 基礎研の柏野邦夫氏に謝意を表します。

参考文献

- 1) Judea Pearl. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, Vol. 29, No. 3, pp. 241–288, 1986.
- 2) 木下智義, 村岡秀哉, 田中英彦. 音楽情景分析の処理モデル OPTIMA における 確率的情報統合機構の改善. 情報処理学会第 53 回全国大会, Vol. 2, No. 7T-3, pp. 369–370, September 1996.
- 3) 柏野邦夫. 計算機による聴覚的情景分析 – はじめの一步 –. 日本音響学会誌, Vol. 50, No. 12, pp. 1023–1028, December 1994.
- 4) 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.
- 5) 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.
- 6) 松田稔, 秋山好一. 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, Vol. 52, No. 4, pp. 253–260, 1996.