

音楽情景分析の処理モデル OPTIMA における

5Z-6

単音・和音仮説生成の改善

村岡 秀哉 木下 智義 田中 英彦

東京大学工学部

1 はじめに

著者らは音楽情景分析の処理モデル OPTIMA を実装している [1, 2, 3]。

OPTIMA では、複数の独立したモジュールに確率を持った仮説の組を出力させ、これを確率伝搬によって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。

2 音楽情景分析の処理モデル OPTIMA

OPTIMA 実験システムでは、抽象度の低い順に

- 周波数成分集合
- 単音集合 (周波数成分集合と 1 対 1 対応)
- 和音 (コード)

の3つの階層をもち、各々のレベルで仮説を生成し、互いに確率伝搬することで各仮説の確信度を評価・再計算している。

現在の処理形態は、音楽音響信号から時間軸及び周波数軸に音響エネルギーを展開したサウンドスペクトログラムを生成し、更に一連の音響エネルギーの連なりを周波数成分として抽出し、立ち上がり時間で分割して周波数成分集合仮説としている。

単音集合・和音仮説については、低レベル仮説のクラスタリングにより生成される (図1)。

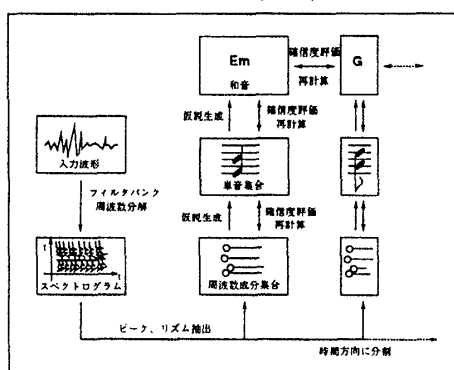


図 1: OPTIMA 仮説ネットワーク

3 下位仮説に依存しない仮説生成

前節のように OPTIMA 実験システムは構成されているが、認識率は十分なレベルとは言いがたい。

この原因の一つとして、上位レベルの仮説を生成する際に、下位レベルの仮説に含まれる誤りが残ってしまうことが挙げられる。実験システムでは、たとえば単音仮説の誤りが、和音仮説の誤りを引き起こしている。

そこで図2のようにサウンドスペクトログラムから直接上位仮説を生成することでシステムの認識率向上を図る。

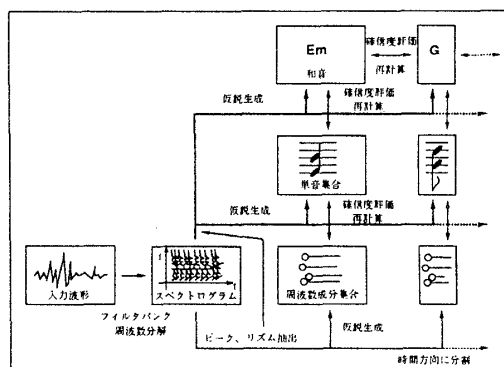


図 2: OPTIMA 仮説ネットワーク (2)

本稿ではこのうちの和音仮説における具体的な実装、評価実験結果及び考察について述べる。

4 単音仮説を用いない和音仮説生成

まず最初に、サウンドスペクトログラムを音の立ち上がり時間によって切断する。音の立ち上がり時間の抽出については本稿の範囲外であるため、従来の処理 [2] で生成されるリズム情報を利用してスペクトログラムを切断する。各々の時間領域における音響パワーの時間平均をとると、図3のようなパワー曲線が得られる。(このパワーのピークは当然ながら単音の周波数成分を表している)。

ここでそれぞれの音階 (C~Bb) に相当する周波数についてパワーを調べ、基本確率とする。

このままでは高次周波数による雑音が入るので、統計的に見込まれる雑音の大きさだけパワー値をカットする。例えば、5度の音 (基音がCならばG) に3倍音、6倍

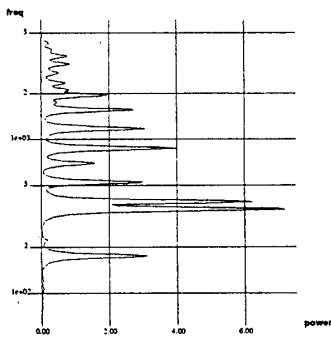


図 3: ある時間のパワー曲線

音、1 2 倍音...が入ってくるのでその分パワーをカットする。これにより、存在しない音階が高確率になることが防がれる。

また、単純にパワー値を確率とするのでは、音量の大きな音階が高確率となってしまうので、必要以上に大きなパワー値はカットする。そうすることで音量の小さな楽器の音階が埋もれる事なく確率値を得る。音階の確率が求まったら、全ての和音についてその構成音(Cm ならば C、Eb、G)の確率を元に和音確率を算出する。

### 5 評価実験・結果

自然楽器音を収録したサンプラーを用いてサンプル曲(「蛍の光」の室内楽アンサンブル)を演奏したモノラルの音響信号を入力として、和音の認識率を現在の実験システムと比較した。

音色は2音の場合はピアノ(p)とクラリネット(c)、3音の場合は更にトランペット(t)を加えたものを用いた。結果を図4、5に示す。ここで、「和音構成音情報を利用」とは和音-単音集合仮説間の確信度評価/再計算を行った場合で、「和音遷移情報を利用」とは同様に和音仮説間の時間方向の確信度評価/再計算を行った場合である。

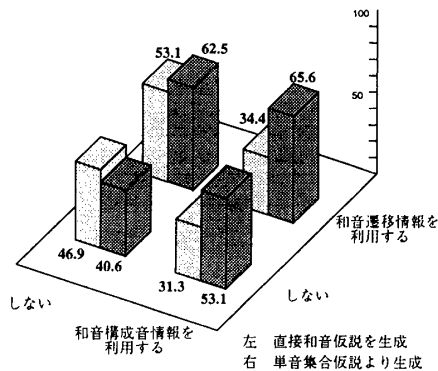


図 4: 音源 2 色の場合の和音認識率

### 6 考察

和音構成音情報、和音遷移情報とともに利用しなかった場合において、今回実装した方が良い認識率が得られた。これは、周波数成分及び単音集合仮説を中途に經由

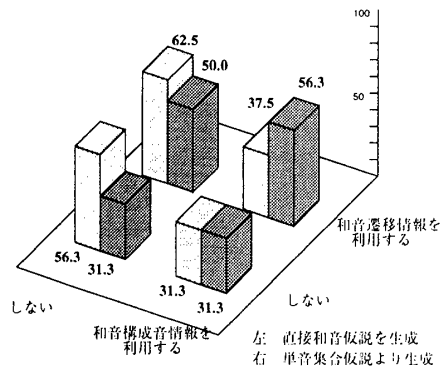


図 5: 音源 3 色の場合の和音認識率

して生成される現在の実験システムよりも直接生成した方が単独のモジュールとしては性能がよい事を示すものである。

また、音数を増やした場合、今までのシステムでは認識率が下降するのに対し、今回実装したものでは逆に認識率が上昇する傾向にある。これは、従来システムでは周波数成分の増加により下位レベル仮説の認識率が悪化し、そのために和音仮説の認識率も低下したと考えられる。逆に、今回のようにサウンドスペクトログラムから和音仮説を直接生成する場合は下位仮説の影響を受けず、音数が増える事によって和音の構成音の不足(例えば2音では Cm なのに Eb と G のみが鳴っている場合など)が解消されたといえる。

和音構成音情報を利用した場合に今回実装した手法で認識率が低下した原因であるが、現在の実験システムではもともと単音集合仮説から和音仮説を生成していたため、両仮説間の類似度(傾向)が高く、ゆえに両仮説間で競合による確信度低下は起きにくかったのに対し、直接生成の場合は競合により確信度の低下が生じたものと思われる。また、この結果は和音-単音集合仮説間の確信度評価モジュールの性能が不十分であることを示している。

### 7 おわりに

本稿では和音仮説を直接生成する手法の実装及び評価実験を行い、その有効性を示した。今後は単音仮説を直接生成する手法の評価についても考察したい。

### 参考文献

- [1] 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 の実装”, 情処 '95 春全大, 6D-2 (1995).
- [2] 木下 智義, 柏野 邦夫, 中臺 一博, 田中 英彦: “音楽情景分析の処理モデル OPTIMA における音楽シーン情報の抽出と利用”, 情処 '95 春全大, 6D-3 (1995).
- [3] 中臺 一博, 柏野 邦夫, 木下 智義, 田中 英彦: “音楽情景分析の処理モデル OPTIMA における統計的単音仮説生成処理”, 情処 '95 春全大, 6D-4 (1995).