音楽情景分析の処理モデル OPTIMA の実装

6D－2

柏野 邦夫　中臺 一博　木下 智義　田中 英彦
東京大学 工学部

1　はじめに

われわれは、聴覚的情景分析を「知覚的な音」の分離抽出 (知覚的音源分離) と構造化の問題と捉え[1]、モノラルの楽器演奏の音響信号を題材として、音楽情景分析 (音楽音響信号を対象とする聴覚的情景分析) の処理モデルについて検討を行っている。ここで、知覚的音源分離とは、人間がひとつのものとして知覚または認識するような音響エネルギーのまとまり (これを知覚的な音と呼ぶ) を一つのものとして記号化することを指す。

われわれは既に、ベイズの定理に基礎を置く定量的かつ階層的な情報統合のメカニズムを備えた音楽情景分析の処理モデル OPTIMA (Organized Processing toward Intelligent Music Scene Analysis) を提案した[2]。この処理モデルに基づき、音楽情景分析の実験システムを実装し検討を行ったので、本稿でその概要を報告する。

2　処理モデルの全体像

音楽情景分析の処理モデル OPTIMA の全体像を 図1 に示す。処理モデルは、複数種類の楽器音を含むモノラルの音楽音響信号を入力とし、楽器種類ごとの演奏情報を抽出して、和音記号列、単音記号列 (MIDI データ, 画面表示) および分離・再合成した各楽器ごとの音響信号の形で出力する実験システムとして実装されている。

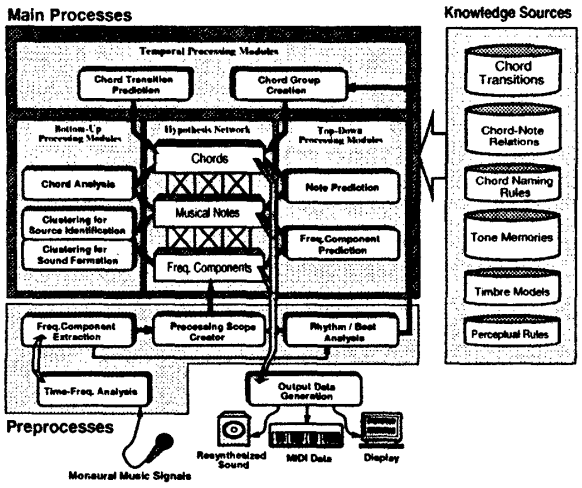


図 1: 処理モデル (OPTIMA) の全体像

OPTIMA は、(A) 前処理部、(B) 主処理部、(C) 知識源、および (D) 後処理部の四つの部から成る。

前処理部は、入力音響信号を時間と周波数に関するエネルギー表現に変換するとともに、このエネルギー表現上における特徴を周波数成分として抽出し、拍位置情報によりこれを整形して、主処理部に対する入力となる処理単位 (processing scope) を形成する部分である。ここで処理単位とは、近接した時刻に開始端点を持つ周波数成分の集合である。

主処理部は、抽象度の低い順に (1) 周波数成分、(2) 単音、および (3) 和音の三つの抽象度の階層を持つ。ここで、それぞれの階層は、時間に対応する次元を持っている。主処理部は、これら三つの階層に対応する仮説ネットワークを備えており、それぞれの階層のある処理単位において、一般に複数の仮説を保持する。この仮説ネットワークに対して、(a) 抽象度の低い階層から抽象度の高い階層への情報表現の変換を行うボトムアップ処理モジュール、(b) 抽象度の高い階層から抽象度の低い階層への情報表現の変換を行うトップダウン処理モジュール、(c) 時間の推移に関する情報を扱う処理モジュール、の三つの群に分けられる処理モジュールが情報を書き込む。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元に単音の情報を生成する処理、単音の情報を元に和音の情報を生成する処理の二つがある。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正当性に関する情報を出力する処理と、単音の情報を元に周波数成分仮説の正当性に関する情報を出力する処理の二つがある。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と、時間的に連続する何個の処理単位が一つの和音を形成するかに関する情報を出力する処理の二つがある。

仮説ネットワークによる階層的な情報統合の方法として、Pearl のベイジアンネットワークを用いる[3]。これによれば、木または単結合グラフで表される事象系において、確率として与えられる情報を 2 種類のリンクを用いてネットワーク全体に矛盾なく伝搬させることができる。実装した実験システムでは、事象系を抽象度の階層と時間的区分とに対応させ、図2 のようなネットワーク構造をとっている。

主処理部における各処理モジュールは、それぞれ必要に応じて知識源を参照する。知識源としては、和音遷移に関する統計データ、和音構成音に関する統計データ、和音認識ルール、単音を構成する周波数成分に関するデータ、音色の特徴空間、および単音形成のための知覚的ルールを備える。

後処理部は、主処理部の仮説ネットワークにおいて最尤となった仮説を、目的に応じた形で出力するためのものである。後処理としては、(a) 画面などに楽譜形式で表示するための処理、(b) システムによって認識された単音や和音を、新たに電子楽器で演奏させることによって聴覚的に確認するための MIDI (Musical Instrument

---

Implementation of OPTIMA : Organized Processing toward Intelligent Music Scene Analysis
Kunio Kashino, Kazuhiro Nakadai, Tomoyoshi Kihoshita and Hidehiko Tanaka
University of Tokyo, Department of Electrical Engineering
7-3-1 Hongo, Bunkyo-Ku, Tokyo, 113, Japan.

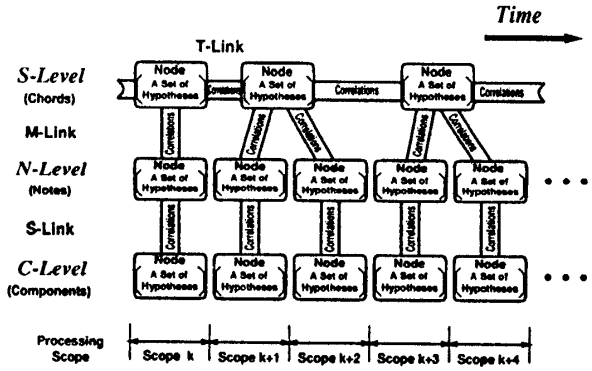


図 2: 実験システムにおける仮説ネットワークの構造

Digital Interface) データ生成処理、および (c) システムによって認識された単音や和音を、周波数成分データから直接再合成するための処理の三種の処理を支持している。

3 処理モデルの実装

前処理部 (図 1 の preprocesses) と主処理部 (main processes) においては、各モジュールが独立かつ非同期的に動作し、TCP/IP ソケットにより通信を行って処理を進める。このうち主処理部の実装の概要を図 3 に示す。実装には C 言語を用いており、システム全体でのソースコードの総量は 7 万行余である。

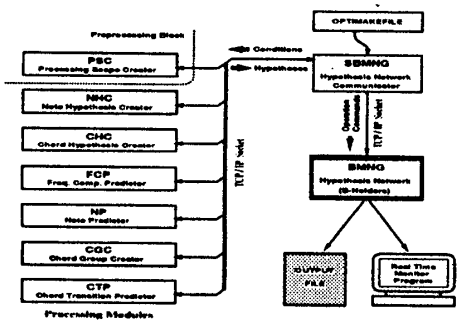


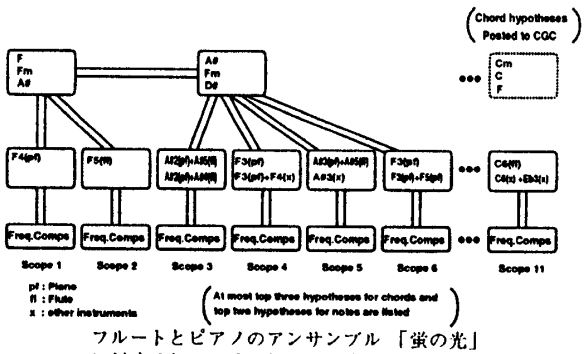
図 3: 主処理部の実装

4 実験システムの動作

図 4 に、実験システムの動作の例を示す。図 4 は、サンプル曲 (フルートとピアノによる「蛍の光」のアンサンブル演奏) の音響信号に対し、11 番目の処理単位まで処理が進んだ時点における仮説ネットワークの状態を示している。図 4 に示すように、仮説ネットワークは、曲の進行につれて徐々に成長する。

5 処理モデルの特徴

聴覚的情景分析に関連する従来の研究ではボトムアップ処理に重点を置いたものが多く[4] 、複数種類の情報の統合に関する検討には乏しかった。情報統合を考慮した研究の例として、黒板モデルに基づく類似の試み [5] を、本稿に述べた処理モデルによるアプローチをと比較すると、まず、本稿の処理モデルでは処理の制御が極めて容易である点が特徴である。黒板モデルに基づくシス



フルートとピアノのアンサンブル「蛍の光」に対する処理の途中経過を示す

図 4: 実験システムの動作例

テムがルールなどの形で処理の制御のための知識を必要とするのに対し、本稿の処理モデルは、システム全体が独立かつ非同期的に動くモジュールの集合として実装されている。各モジュールは局所的な起動条件のみによって起動されるため、グローバルな制御用の知識を必要とせず、実装および各部のメンテナンスが極めて容易であった。それに加え、本稿のモデルでは、得られた処理の結果に対して、各処理モジュールが出力する情報に基づく最尤推定という意味での定量的裏付けが与えられている点も特徴である。

6 おわりに

本稿では、仮説ネットワークによる情報の統合を基盤とする音響エネルギーの群化と構造化の機構を提案し、その具体的応用例として実装された、楽器演奏を対象とする音楽情景分析システムの全体像を示した。実装された実験システムにおいては、情報統合の顕著な効果を示す一連の評価実験結果を得ており、その一部を以下 2 稿 [6, 7] において報告する。

今後は、旋律など、和音以外の時間的構造の利用や、確率だけでは捉えきれない音楽的構造・意味的情報の利用についても検討を進める予定である。

参考文献

[1] 柏野 邦夫: "計算機による聴覚の情景分析", 日本音響学会誌, **50**, 12, pp.1023-1028 (1994).

[2] 柏野 邦夫, 中臺 一博, 田中 英彦: "音楽音響信号から単音記号列を生成するシステム OPTIMA の全体像", 情処研報 **94**, 71, pp.57-64 (1994).

[3] Pearl J. : "Fusion, Propagation, and Structuring in Belief Networks", *Artificial Intelligence*, **29**, 3, pp.241-288 (1986).

[4] Brown G. J.: *"Computational Auditory Scene Analysis: A Representational Approach"*, Ph.D. Thesis, Department of Computer Science, University of Sheffield (1992).

[5] Nawab S. H. and Lesser V.: "Integrated Processing and Understanding Signals", in Oppenheim A. V. and Nawab S. H. (eds.) : *"Symbolic and Knowledge-Based Signal Processing"*, Prentice Hall, pp.251-285 (1992).

[6] 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).

[7] 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).