

音楽情景分析の処理モデル OPTIMA における和音の認識

柏野 邦夫^{†*} 木下 智義[†] 中臺 一博^{†**} 田中 英彦[†]

Chord Recognition Mechanisms in the OPTIMA Processing Architecture for Music Scene Analysis

Kunio KASHINO^{†*}, Tomoyoshi KINOSHITA[†], Kazuhiro NAKADAI^{†**},
and Hidehiko TANAKA[†]

あらまし 我々は、複数種類の楽器音を含む音楽演奏を対象とした音楽認識を、音楽情景分析の問題としてとらえ研究を行っている。ここで音楽情景分析とは、音楽演奏の音響信号から、単音や和音などの音楽演奏情報を記号表現として抽出することを指す。我々は先に、ベイジアンネットワークによる情報統合の機構を備えた音楽情景分析の処理モデル OPTIMA を提案した。本論文では、OPTIMA における処理のうち、特に和音の認識に絞って、情報統合機構の有効性を調べた。その結果、サンプル曲を用いた評価実験において、ボトムアップ処理のみによる和音認識を行った場合に比較して、和音を構成する単音に関する統計情報を統合した場合には 15.6%、また和音の時間的な遷移に関する統計情報を統合した場合には 18.7% の和音認識率の向上が見られたことから、提案する処理モデルにおいてこれらの情報を統合することの有効性が示された。

キーワード 聴覚的情景分析、和音認識、情報統合、ベイジアンネットワーク、自動採譜

1. ま え が き

音楽情景分析 (music scene analysis) 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。演奏された音楽に対し、単音や和音などの演奏情報を認識する技術は、将来的には、自動採譜システムや、音楽データベースの自動タグ付けシステムなどへの応用が考えられる。音楽の認識において、ただ単音を認識するだけではなく、和音なども含め音楽全体として演奏を認識することの必要性は、従来の自動採譜システムの研究においても指摘されている [1]。音楽情景分析は、認識すべき音楽全体を「情景」としてとらえ、音楽的事象を周波数成分・単音・和音などといった階層的な構造の中で記述しようとする試みである。周波数成分の集まりが単音をなし、単音の組合せが和音をなすという階層性は、画像を対象とする情景分析において、エッジの集まりや面が図形をなし、図形の組合せが物体をなすという階層性と類似のものである。

音楽情景分析においては、観測データに加え、対象のモデルや統計データなど、利用可能なさまざまな情報を統合して最終的な認識結果を得ることが重要である。そこで我々は、前論文 [2] において、ベイジアンネットワークに基づく情報統合の機構を備えた音楽情景分析の処理モデルを提案すると共に、単音の認識に対する情報統合機構の有効性を示した。本論文では、その情報統合機構について、和音の認識に対する有効性を示すことを目的とする。以下、まず 2. において、我々の提案する音楽情景分析の処理モデルの全体像を示し、和音認識処理の位置付けを示す。次に、3., 4., および 5. において、特に和音の認識に着目して、ボトムアップ処理、トップダウン処理、時間方向の処理の概要をそれぞれ説明する。6. において、和音の認識に関する評価実験を行い、和音に関する情報の統合を行った場合と行わなかった場合の実験結果とを比較することによって、提案する処理モデルにおける情報統合機構の有効性を示す。7. をむすびとする。

2. 処理モデル OPTIMA における和音認識の位置付け

我々が提案した音楽情景分析の処理モデル OPTIMA (Organized Processing toward Intelligent Music

[†] 東京大学工学部電気工学科, 東京都

Faculty of Engineering, University of Tokyo, Bunkyo-ku, Tokyo, 113 Japan

* 現在, NTT 基礎研究所

** 現在, NTT ソフトウェア本部

Scene Analysis) の全体像を図1に示す[2]。OPTIMAは、各時点で得られた情報に基づいて、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、全体として最ももらしい仮説の組を逐次求めていく枠組みである。図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もとに、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、(A) 前処理部 (pre-processes), (B) 主処理部 (main processes), (C) 知識源 (knowledge sources), および (D) 出力データ生成部 (output data generation) の四つの部からなる。

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共に、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、リズム

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。ここで周波数成分とは、サウンドスペクトログラム上で、時間的に連続した、周波数方向に見たときのパワーの極大点の集合をいう。また処理単位とは、立上り時刻が互いに近接した周波数成分の集合を指す。

主処理部は、音響事象の仮説を保持するためのペ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、および (3) 和音の三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もと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), 単音の情報をもとに和音の情報を生成する処理 (和音仮説生成; chord recognition) の二つがある。トップダウン処理モジュールとしては、和音の情報をもと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) と、単音の情報をもと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 (和音遷移情報付与; chord transition prediction) 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 (和音グループ生成; chord group creation) の二つがある。これらの処理モジュールのうち本論文で扱うものは、和音仮説生成、和音構成音情報付与、および和音遷移情報付与の三つである。

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 (和音遷移情報; chord transitions), 和音を構成する単音に関する統計データ (和音構成音情報; chord-note relations), 単音の集合に対しどのような和音名をつけるかをルールとしたもの

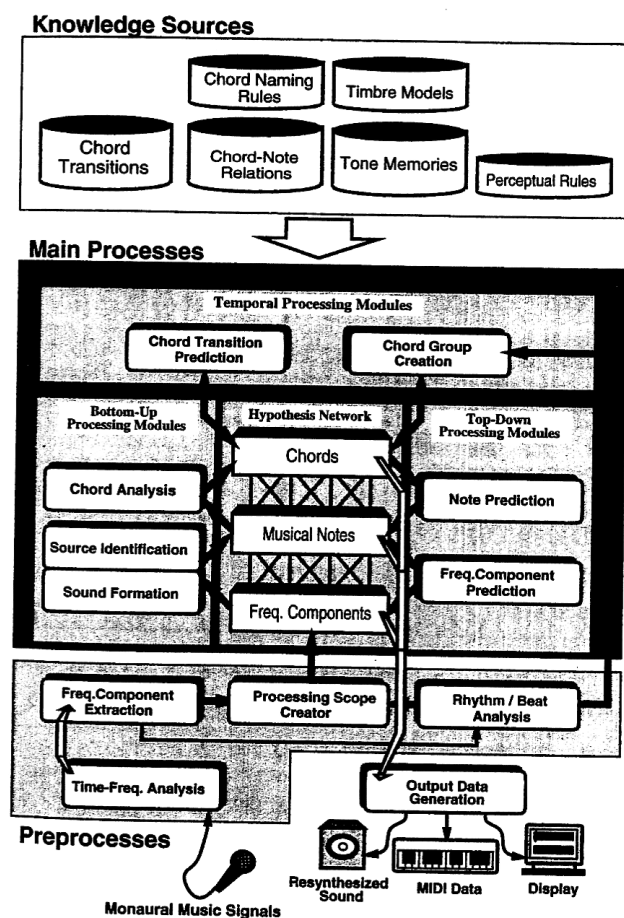


図1 処理モデル OPTIMA の全体像
Fig. 1 The OPTIMA processing architecture.

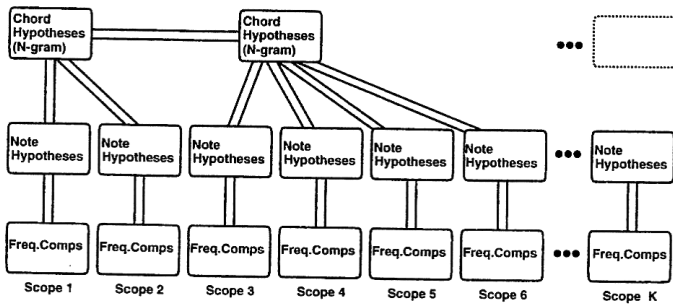


図2 構成される仮説ネットワークの例
Fig. 2 An example of the constructed hypothesis network.

(和音名ルール; chord naming rules), 単音を構成する周波数成分に関するデータ (tone memories), 音色の特徴空間 (timbre models), および単音形成のための知覚的ルール (perceptual rules) を備える。これらのうち本論文で扱うものは、和音仮説生成モジュールの参照する和音名ルール, 和音構成音情報付与モジュールの参照する和音構成音情報, および和音遷移情報付与モジュールの参照する和音遷移情報である。

出力データ生成部は、主処理部の仮説ネットワークにおいて最優秀となった仮説を、画面表示や MIDI (Musical Instrument Digital Interface) データなど目的に応じた形で出力するためのものである。

ここで、仮説ネットワークの和音レベルにおける時間的なつながりの生成について、図2に例を示して説明を加える。図2で、各ノードには仮説と確信度値が、また各リンクには、隣接するノードの仮説同士の間接する条件付確率がそれぞれ保持されている。OPTIMAの主処理部における処理は、前処理部から処理単位が入力されるごとに、逐次的に進行する。図2は、 K 番目の処理単位までの単音仮説が生成され、和音グループ生成モジュールが、3番目から6番目までの処理単位に対応する単音群が一つの和音に対応すると判定した時点を示したものである^(注1)。この時点で、和音仮説が確定され、先行の和音仮説との間に時間方向のリンクが張られる。このようにして、入力が増えらるごとに、仮説ネットワークは順次成長していく。

本論文で用いる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 N における仮説の組を表す)[2],[4]。これは、 $BEL(n)$ が、 α を正規化定数

として

$$BEL(n) = \alpha \lambda(n) \pi(n) \quad (1)$$

のように $\lambda(n)$ と $\pi(n)$ なる2要素の積で表されることと、その $\lambda(n)$ と $\pi(n)$ は、初期値が与えられれば、それぞれ各リンクの条件付確率(親のノードの仮説が与えられたときの子のノードの仮説の条件付確率を行列で表したもの) M を用いて順次隣接ノードに伝搬させられることによるものである。すなわち、ネットワークの一部に新たに加えられた情報の影響を、 $\lambda(n)$ と $\pi(n)$ の伝搬によって、既存の仮説の確信度に反映させていると見ることができる。

本論文では、対象観測前の事前知識に相当する $\pi(n)$ の初期値としては、等確率を仮定する。従って確率伝搬のために必要な情報は、

- (1) 各ノードにおける仮説に対する初期確信度 $\lambda(n)$
- (2) 隣接するノードの仮説同士の間接する条件付確率の行列 M

である。図1のシステムでは、前者をボトムアップ処理で与え、後者をトップダウン処理および時間方向の処理で与える。本論文では以下、和音認識に関する処理においてこれらの情報をどのように求めるかを順に議論する。

3. 和音仮説の生成

2. に述べたように、和音仮説の生成はボトムアップ処理によって行う。すなわち、各単音仮説に含まれる単音の音高から、知識源における和音名ルールを参照して和音仮説を生成する。

3.1 知識源

和音名の表現には、図3に示すような構造を用いた。音楽の分野で用いられている名前付け方法に基づいて、単音の集合と和音名との対応を与えた。例えば、和音のタイプ (Type) が m (マイナー) であれば、 m が無いときに比べて3度の音が半音低い、などといったルールを、図3に示すそれぞれの要素について与えた。

3.2 初期確信度の付与

和音仮説の生成においては、各単音仮説における単

(注1): 和音グループ生成モジュールは、処理単位ごとの和音仮説の生成結果を受け取り、これらを併合して一つの和音仮説として確定し出力する。併合すべき処理単位の判定は、前処理部で抽出する拍位置情報と、各処理単位の単音仮説に含まれる単音の立上り時刻および音高に基づいて行う。

音の欠落や誤りも考慮する必要がある。そこで、次のようなアルゴリズムによって初期確信度を与えた。ここで与える初期確信度は、名前付けの観点から見た和音名の確からしさを表すものである。

(1) 和音名の生成

まず、図3の構造をもつ和音名のうち、音楽的に無意味なもの（例えば B# など）を除いた組合せをすべて生成する。

(2) 和音名の評価

次に、得られた和音名を以下の方法で評価し、初期確信度を与える。

(a) 和音名から、前項の知識源を参照してその構成音（和音名構成音）を獲得する。但し、和音名構成音のオクターブの違いは考慮しないものとした。

(b) 和音名構成音と、入力として得られた単音仮説群とを比較して、一致する部分があれば点数を与え、一致しない部分があれば点数を減らす処理を行う。

(c) 得られた点数を、和音名構成音と単音仮説とが完全に一致した場合の得点で割って確信度とする。

(3) 和音名の選択

こうして得られた和音名のうち、確信度が上位にあるものを、あらかじめ設定した数だけ選択する。本論文では、予備実験の結果から、5位までの和音名を選択するものとした。

以上のようにして処理単位内における和音名仮説が得られるが、和音名仮説を「和音仮説」としてネットワークのノードに与える際には、和音の時間的な遷移について考慮する必要がある。これは、一つの和音ごとの遷移を考えると、OPTIMA の情報統合機構において仮定しているノードごとの仮説の独立性（あるノードの仮説の状態は、隣接するノードの仮説のみによって定まり、それより離れたノードにおける仮説の

状態には依存しないという仮定）が一般には成り立たないと考えられるためである。そこで、和音の遷移を N-gram で近似し（ある和音の出現確率は $N - 1$ 個前までの和音の状態にのみ依存すると近似し）、各ノードでは N-gram の仮説をもつことにする。このため、処理単位内における和音名仮説を、N-gram に展開した上で和音仮説とする。

4. 和音構成音情報の付与

本章では、トップダウン処理の一つである、和音に基づく単音情報の付与について述べる。

4.1 知識源

和音仮説に基づいて、出現する単音を予測し、ある和音の下での単音に対する条件付確率の付与を行う。処理に際しては、和音構成音情報を参照する。

和音構成音情報は、和音名つきで印刷された楽譜に対して、和音と単音に関する統計をとることによって構成することができる。本論文では、例として国内外の童謡や伝統的に歌われている曲を対象とし、2,365個の和音について分析を行った。これを表1に示す。この際、経過音と呼ばれる、和音の構成とは無関係に一時的に現れる単音については集計から除外した。

4.2 条件付確率の付与

この和音構成音情報から得られた確率値をもとに、条件付確率 P (単音仮説 | 和音仮説) を求める。単音仮説は一般に同時に発音している複数の単音を含んでいるが、和音構成音情報で与えられるのは1個の単音の存在確率であるから、 P (単音仮説 | 和音仮説) を

表1 和音構成音情報の例

印刷された楽譜における 2,365 和音の統計的分析から求めた条件付確率 P (単音 | 和音) の値を示す。

Table 1 Examples of chord-note relations obtained by statistical analysis of printed scores.

Example: A#m7+5 / C#

Base-name	Type	5,6,7,9 dig	others	Base note
A	#	m	7 +5	C #

Examples of possible variations:

A	(none)	(none)	(none)	(none)	(none)	(none)	(none)	(none)
B	#	m	M	6	+5	sus4	A	#
⋮	b	dim		7	-5		B	b
G		aug		9	+9		⋮	
					-9		G	

図3 本論文で用いた和音名の構造
Fig.3 Format of the chord names.

単音	和音の例		
	A	A7	Am
A	.983 ± .023	.852 ± .074	1.00 ± .000
A#	.000 ± .000	.023 ± .031	.030 ± .058
B	.150 ± .064	.364 ± .101	.091 ± .098
C	.008 ± .016	.023 ± .031	.848 ± .122
C#	.850 ± .064	.818 ± .081	.000 ± .000
D	.025 ± .028	.057 ± .048	.182 ± .132
D#	.067 ± .045	.023 ± .031	.030 ± .058
E	.842 ± .065	.545 ± .104	.909 ± .098
F	.017 ± .023	.023 ± .031	.000 ± .000
F#	.125 ± .059	.148 ± .074	.000 ± .000
G	.025 ± .028	.773 ± .088	.121 ± .111
G#	.075 ± .047	.045 ± .044	.000 ± .000

± は 95% 信頼区間の幅を示す。

与えるためには確率値の統合が必要である。そこで、Dempster の結合規則を用いた [6]。Dempster の結合規則は、一般に次式で表される。

$$m(A_k) = \frac{\sum_{A_i \cap A_j = A_k} m_1(A_i) m_2(A_j)}{1 - \sum_{A_i \cap A_j = \phi} m_1(A_i) m_2(A_j)} \quad (2)$$

これは、 A_i, A_j ($i, j = 1, 2, \dots$) なる焦点要素 (focal element) につき、独立した証拠に基づく二つの基本確率 m_1 および m_2 があつたとき、統合された確率値は $m(A_k)$ となることを主張するものである。ここでは、 m_i は単音仮説中の i 番目の単音によつてもたらされる単音仮説の確率を表すものとし、 A_1 はその単音仮説が正しいこと、 A_2 はその単音仮説が正しくないことに対応すると考える。具体的には次のような計算を行う。

(1) 最初に、和音仮説 C と、単音仮説 N に含まれる単音 n_i について、 $P(n_i|C)$ を前述の知識源から得る。

(2) 和音仮説 C の構成音数 l (G なら 3, $G7$ なら 4, \dots 等) から、その和音における平均の単音存在確率 p_a を求める。これは以下のようにして求める。

$$p_a = l/12 \quad (3)$$

ここで、12 は 1 オクターブ当りの単音の個数である (ここではオクターブの違いは考えない)。

(3) $P(n_i|C)$ と p_a を比較して、以下のように Dempster の基本確率を求める。

(a) $P(n_i|C) \geq p_a$ の場合、和音仮説 C に対して単音仮説が N であることの根拠となると考え、

$$m_i(E) = \frac{P(n_i|C) - p_a}{1 - p_a} \quad (4)$$

$$m_i(N) = 0 \quad (5)$$

$$m_i(U) = 1 - m_i(E) \quad (6)$$

とする。ここで E, N , および U は、それぞれ

$$E = \{A_1\}, \quad (7)$$

$$N = \{A_2\}, \quad (8)$$

$$U = \{A_1, A_2\} \quad (9)$$

である。すなわち、各々の値の意味するところは以下のとおりである。

$m_i(E)$: 単音仮説が N であるという基本確率。
 $m_i(N)$: 単音仮説が N ではないという基本確率。
 $m_i(U)$: 単音仮説が N であるかどうかかわからないという基本確率。

(b) $P(n_i|C) < p_a$ の場合、和音仮説 C に対して単音仮説が N ではないことの根拠となると考え、

$$m_i(E) = 0 \quad (10)$$

$$m_i(N) = \frac{p_a - P(n_i|C)}{p_a} \quad (11)$$

$$m_i(U) = 1 - m_i(N) \quad (12)$$

とする。

(4) 以上の基本確率を式 (2) に代入して、全体での Dempster の確率値を求める。単音仮説 N に含まれる単音数が 3 以上の場合には、式 (2) を繰り返し適用する。

(5) こうして得られた全体の $m(E)$, $m(N)$, および $m(U)$ に対して、上界確率

$$P^*(E) = m(E) + m(U) \quad (13)$$

と、下界確率

$$P_*(E) = m(E) \quad (14)$$

を求め、その中間点である

$$P = \frac{P^*(E) + P_*(E)}{2} \quad (15)$$

$$= m(E) + \frac{m(U)}{2} \quad (16)$$

を、条件付確率 $P(N|C)$ として出力する。

5. 和音遷移情報の付与

本章では、時間方向の処理のうち、和音遷移情報の付与について述べる。なお、時間方向の処理には、和音遷移情報付与の他に和音グループの同定があるが、和音グループの同定に関しては本論文では扱わないものとする。

5.1 知識源

和音遷移情報付与モジュールでは、知識源に蓄積された和音の遷移確率を参照して、和音仮説の生成および条件付確率の付与を行う。

本論文では $N = 3$ と仮定し、国内外の童謡や伝統的に歌われている曲など 206 曲を対象として、調性別に 3,570 の和音遷移について分析を行って遷移情報を蓄積した。蓄積した和音遷移情報の形式を図 4 に示し、

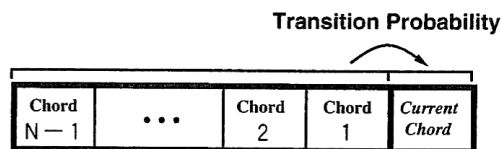


図4 和音 N-gram の形式
Fig. 4 Format of the chord N-gram.

表2 和音遷移情報の例

印刷された楽譜における 3,570 の和音遷移の統計的分析から求めた条件付確率 P (和音 | 和音群) の値を示す。

Table 2 Examples of chord transition information obtained by statistical analysis of printed scores.

和音遷移の例		確率値
I	V7 → I	0.075
V7	I → I	0.048
I	IV → I	0.038
V7	I → V7	0.036
I	V → I	0.022

内容の一部を表2に例示する。なお表2に示した確率値は、それぞれの遷移が出現した回数を全遷移数で割ったものである。

5.2 条件付確率の付与

条件付確率の付与においては、前章に述べた和音構成音情報の場合とは異なり、知識源に蓄積されている確率値をそのまま用いることができる。

6. 評価実験

処理モデル OPTIMA において、和音の認識に対する情報統合の有効性を調べることを目的として、評価実験を行った。まず、和音遷移情報を利用することの和音認識に対する効果を調べるため、典型的な和音遷移パターンを用いた実験を行う(実験1)。次に、和音遷移情報と和音構成音情報のそれぞれの効果を、サンプル曲を用いた実験によって調べる(実験2)。

6.1 実験1：和音遷移情報統合の効果

まず、正解の和音遷移を実際の楽譜における和音遷移をもとに設定した。設定した正解は、和音遷移情報の蓄積に用いた曲のうち、最も長い和音遷移をもつ曲の和音遷移を用いた。これは18個の和音からなる。次に、各和音の構成音からなる単音群データを作成し、これにノイズを重畳させたものを単音試料とした。ノイズの重畳は、単音をランダムに付加または削除することによって行った。これらの単音試料に対して和音認識処理を行い、和音遷移情報から得られる和音遷移確率の伝搬を行った場合と行わなかった場合の和音認

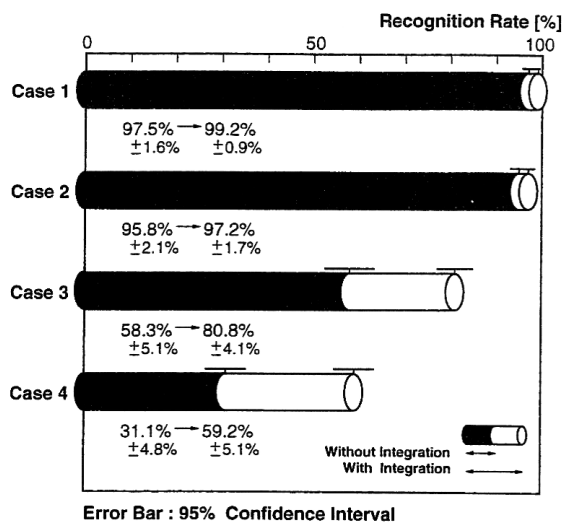


図5 和音遷移情報の効果
Fig. 5 Results of the experiment 1.

識率を調べた。ここで和音認識率は、出力した和音のうち正解として設定した和音と一致したものの割合である。

単音試料は、ノイズの重畳条件によって次の4通りのものを用いた。

条件1 和音18個のうちランダムに選択した1個について、単音一つのノイズを重畳(単音一つを付加または削除)した。

条件2 和音18個のうちランダムに選択した1個について、単音二つのノイズを重畳した。

条件3 和音18個のすべてについて、単音一つのノイズを重畳した。

条件4 和音18個のすべてについて、単音二つのノイズを重畳した。

本実験の結果を図5に示す。これに見られるように、和音遷移情報の統合後における和音認識の誤りは、統合前に比べ2/3以下になった。このことから、和音遷移情報の統合は、和音認識率の改善に対して顕著な効果があることが示された。

但し、この実験の対象とした曲は、統計情報の抽出に用いた206曲のうちの1曲であり、もともとシステムが蓄積している和音遷移の統計情報との相性が良いと言える。例えば、ジャンルが違う音楽の和音進行を処理の対象として用いた場合に、常にこのような効果が得られるとは限らない。

6.2 実験2：サンプル曲に対する情報統合の効果

本実験では、より実際に近い場合として、サンプル曲を演奏した音響信号試料を対象としてシステムを動作させ、和音構成音情報および和音遷移情報の統合の

有無と和音認識率との関係を調べる。

サンプル曲は「蛍の光」とした。これは、和音構成音情報および和音遷移情報の蓄積には用いていない曲である。この曲を、プロの作曲家に依頼して3パートのアンサンブルに編曲し、各パートを自然楽器音を蓄積したサンプラ（あらかじめ楽器音を記録蓄積しておき、これを再生することによって発音する方式の音源装置）によって演奏して音響信号試料とした。音色は、フルート、クラリネット、およびピアノとした。

この音響信号試料に対して実験システムを動作させ、和音構成音情報および和音遷移情報に基づく情報統合を行った場合と行わなかった場合について、和音認識率を比較した。本実験では、単音認識率および音高認識率についても併せて測定した。ここで、和音認識率は、出力された和音名のうち正解と判定されるものの割合である。正解は、編曲者が与えた和音名とした。また、和音認識では必ずしも正解が一意に定まらないので、編曲者の見解に基づき、次に例示するような許容範囲を設けた。

(1) 箇所によっては、和音名のうち6, 7, 9の有無については問わない（システムの出力した和音名に6, 7または9が余計についていても、また足りなくても、いずれも正解とみなす）ものとした。

(2) 箇所によっては、CmとEbの両者を許容した。実験に用いた楽譜および和音の正解例を付録の図6に示す。

一方、単音認識率 R は、以下のように定義した。

$$R = \frac{1}{2}(\alpha - \beta) + \frac{1}{2}. \quad (17)$$

但し、

$$\alpha = \frac{a}{n}, \quad \beta = \frac{b}{n} \quad (18)$$

であり、 n は入力（正解）に含まれる総音符数、 a は出力に含まれる音符のうち正しく認識された音符の数、 b は出力に含まれる音符のうち、正しく認識された音符でない音符の数である。音高認識率とは、単音認識率において、音源同定の誤りを問わない（音源同定の誤りは正解とみなす）場合の値である。なお、式(17)における $1/2$ の乗算と加算はスケール調整のためのものである。すなわち、システムが入力に含まれる音符と同数の音符を出力した場合、この正規化によって、すべてが誤っていれば $R = 0\%$ 、すべてが正しければ $R = 100\%$ となる。

表3 和音に関する情報統合の有無と和音/単音認識率との関係

Table 3 Chord/Note recognition rates with/without integration of chord transition information and chord-note relation information.

	行う (○) / 行わない (●)			
和音遷移情報の統合	●	●	○	○
和音構成音情報の統合	●	○	●	○
和音認識率 (%)	59.4	75.0	78.1	78.1
単音認識率 (%)	59.5	59.5	59.5	59.5
音高認識率 (%)	87.6	87.6	87.6	87.6
単音が正解の場合の和音認識率 (%)	59.4	78.1	81.3	81.3

なお本実験では、和音グループの作成は人手により行った。すなわち、和音グループの区切りとなる時刻をあらかじめ和音グループ生成モジュールに与えることにより、弱起の部分を除き、以後1小節に二つずつの和音グループを生成させた。

本実験の結果を表3に示す。和音構成音情報を統合した場合、これを統合しない場合に比べて15.6%、また和音遷移情報を統合した場合、これを統合しない場合に比べて18.7%の和音認識率の向上が得られた。また正しい単音を与えて和音認識を行ったところ、和音構成音情報を統合した場合、これを統合しない場合に比べて18.7%、また和音遷移情報を統合した場合、これを統合しない場合に比べて21.9%の和音認識率の向上が得られた。これらのことから、実験システムにおいて和音構成音情報および和音遷移情報を統合することの有効性が確かめられた。なお、単音認識を含めて動作させた場合と正しい単音を与えた場合との和音認識率の差は、主に単音認識結果における単音の欠落に起因すると考えられる。

また、認識された和音の時系列を演奏させて試聴したところ、和音遷移情報の統合を行わない場合における演奏では極めて不自然な演奏になるのに対し、和音遷移情報の統合を行った場合では、もとの楽譜に対しては誤りが含まれているものの、音楽としては比較的自然に聴くことのできる演奏となった。このことは、和音遷移情報の統合によって、音楽の進行としての性質にそわない誤りの多くが排除されたためであると考えられる。

しかし本実験では、和音遷移情報を統合した場合に生じた和音認識誤りの中で、更に和音構成音情報を統合することによって修正できた例は見られなかった。

これは、和音遷移情報を蓄積した際に、極めて簡単に典型的な和音進行で編曲された楽譜を多く用いたために、和音構成音情報よりも和音遷移情報の方が強力に認識結果に影響したためであると考えられる。

また、本実験においては、和音構成音情報や和音遷移情報の統合による単音認識率の向上は見られなかった。本実験における単音認識の誤りの大部分は、音源同定の誤り、音高同定のオクターブの誤り、または単音の欠落であった。前2者は、本実験システムの処理では和音とは直接関係しないので、正しく和音を認識することによる修正が見込めない種類のものである。一方、単音の欠落については和音認識による修正が期待できるが、生成された単音仮説に正解が含まれていなかった場合には修正され得ない。このため、和音認識率の向上にもかかわらず単音認識率の向上が見られなかったものと考えられる。

7. むすび

本論文では、先に我々が提案した音楽音響信号を対象とする音楽情景分析の処理モデル OPTIMA において、和音の認識に最も関連の深い処理に絞って、情報統合の有効性を調べた。その結果、サンプル曲を用いた評価実験において、ボトムアップ処理のみによる和音認識を行った場合に比較して、和音構成音情報を統合した場合には 15.6%、また和音遷移情報を統合した場合には 18.7% の和音認識率の向上が見られたことから、これらの情報を統合することの和音認識に対する有効性が示された。

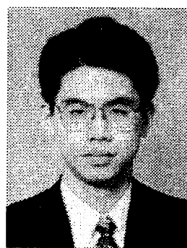
しかし、本論文の実験では、和音遷移情報や和音構成音情報の統合によっても、単音認識率は向上しなかった。これは単音認識に関するモジュールにおいて、音源同定の誤りなど、和音による修正が効かない種類の誤りが多く見られることなどによるものである。

和音遷移情報を統合した場合には、和音遷移情報の蓄積に用いた曲の多くが極めて簡単に典型的な和音付けをされていたために、典型的な和音進行パターンに強力に引きずられ、実際の構成音からかい離れた和音仮説の確信度が高くなった結果、誤認識となった例が多く見られた。従って、今後、より高精度の和音認識を行うためには、幅広く和音遷移の統計データを蓄積することが必要である。また、フレーズのまとまりや類似パターンの繰返しなど、曲の構成に関する情報を抽出し利用することができれば、和音認識精度は更に向上すると考えられる。これらの点について、今後研

究を進める予定である。

文 献

- [1] 片寄晴弘, 井口征士, “知的採譜システム,” 人工知能学会誌, vol.5, no.1, pp.59-66, 1990.
- [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,” 信学論 (D-II), vol.J79-D-II, no.11, pp.1751-1761, Nov. 1996.
- [3] 柏野邦夫, “計算機による聴覚の情景分析 —はじめの一步,” 日本音響学会誌, vol.50, no.12, pp.1023-1028, 1994.
- [4] J. Pearl, “Fusion, propagation, and structuring in belief networks,” Artificial Intelligence, vol.29, no.3, pp.241-288, 1986.
- [5] C. Roads, “Research in music and artificial intelligence,” ACM Computing Surveys, vol.17, no.2, pp.163-190, 1985.
- [6] 石塚 満, “Dempster & Shafer の確率理論,” 信学誌, vol.66, no.9, pp.900-903, 1983.



柏野 邦夫 (正員)

平2東大・工・電子卒。平7同大学院電気工学専攻博士課程了。工博。同年 NTT に入社、基礎研究所情報科学研究部勤務、現在に至る。聴覚的情景分析の研究に従事。音響的情報を対象とする信号処理および知識処理に興味をもつ。平6情報処理学会奨励賞受賞。情処学会、人工知能学会、日本音響学会、IEEE 各会員。



木下 智義 (学生員)

平7東大・工・電子情報卒。現在同大学院情報工学専攻修士課程在学中。聴覚的情景分析の研究に従事。情処学会会員。

付 録
Auld Lang Syne

Scotch Traditional

The musical score is presented in three systems, each with three staves (treble, middle, and bass clefs). Chord annotations are placed above the first staff of each system. The first system includes chords: B♭, B♭, Cm, F9, Gm, B♭, Cm, E♭. The second system includes: B♭, B♭7, Cm, Cm, B♭, F9, B♭, E♭. The third system includes: B♭, B♭7, Cm, Cm, Dm, B♭7, E♭, Cm. The score concludes with a double bar line.

1小節あたり2個ずつの和音を認識対象とした。

図6 実験2で用いた楽譜と和音認識の正解例
Fig.6 The score used in experiment 2.

(平成7年10月20日受付, 8年5月17日再受付)



中臺 一博 (正員)

平5東大・工・電気卒。平7同大学院情報工学専攻修士課程了。同年NTTに入社、ソフトウェア本部勤務、現在に至る。在学中、聴覚的情景分析の研究に従事。情処学会、人工知能学会、日本音響学会各会員。



田中 英彦 (正員)

昭40東大・工・電子卒。昭45同大学院博士課程了。工博。同年東大・工・講師、昭46同大助教授、昭62同大教授、現在に至る。この間昭53~54ニューヨーク市立大客員教授。計算機アーキテクチャ、並列推論マシン、帰納推論、オブジェクト指向計算システム、分散処理、CAD等の研究に従事。著書「非ノイマンコンピュータ」、「情報通信システム」、共著書「計算機アーキテクチャ」、「VLSI コンピュータ I, II」、「ソフトウェア指向アーキテクチャ」。情処学会、人工知能学会、日本ソフトウェア科学会、IEEE、ACM各会員。