



## 自動採譜技術の展望\*

嵯峨山茂樹 (東京大学情報理工学系研究科)\*\*・

亀岡弘和 (NTT コミュニケーション科学基礎研究所)\*\*

43.75.Xz, Zz; 43.60.Uv

### 1. 自動採譜の目的

約 40 年前の今頃、筆者はある曲の採譜で悪戦苦闘していた。買って来たレコードを何十回も聴いて楽譜にして、それを楽曲分析して大学のゼミ科目のレポートにしようという魂胆が甘かった。それ以来、機械で楽に自動採譜できる日が早く来ないかな、と虫のいいことを思ったものである。

それが来つつある。

何の役に立つのか。もちろんレポート書きのためではない。録音しかしない音楽を楽譜化できれば演奏に役立つ。ジャズの即興演奏から「耳コピー」をして演奏したい人には福音だろう。もっと大きな用途も見えて来ている。音楽のデジタル化が発展し、音楽の供給も入手経路も増え、音楽検索や音楽推奨が大きな関心になっている。自分が好む未知の音楽を効率良く探し出せれば、ユーザ側にも供給側にも恩恵があるだろう。音楽がひとたび楽譜のような記号列に変換できれば、検索の可能性は高まるはずである。

また、音楽の能動的な観賞も今後の方向である。従来は音量調節やスペクトルのイコライザくらいしかなかった受動的な音楽鑑賞を、テンポを変え、音色を変え、リズムを入れ替え、演奏スタイルを入れ替え、更には編曲を変えるなど、音楽への能動的な関わりが容易にできるような音楽加工の技術があれば、一步踏みこんだ音楽鑑賞ができるだろう。これらには、音楽の基本的な記述形式である楽譜が重要な役割を果たせるだろう。

加うるに、音声と音楽は、人間が発し聴く膨大な音響情報の授受という点で双璧である。音楽の

自動採譜は、音声の自動認識と対をなす概念と言え、知能的な音響信号処理を象徴する目標と言えよう。音声認識が音声分析、符号化、合成、加工などの技術と連動して進歩したのと同様に、自動採譜も音楽情報処理全般に及ぼす恩恵が大きい。

### 2. 何が難しいか

図-1(a)~(d)をご覧いただきたい。ここには自動採譜の問題が端的に現れている。

まず第 1 は多重音解析である。図-1(a)の楽譜をピアノ演奏したオーディオ信号が入力されるとする。図-1(b)はそのスペクトログラムだが、倍音成分が複雑に重なりあっていて、これから何の音符がいつ発音されたか知るの容易ではない。しかし、人間はこれを耳で楽しんでいるわけで、専門家は音を聴いて図-1(c)のピアノロール表示のような MIDI データを(人手で)作成することができる。人間には、このような「時刻表」に従って難なくピアノ音色の音符が次々に聞こえているわけである。この専門家に代わる多重音解析の問題ははまだ難問である。一度混ざったものはなかなか分離できないのは、何でも同じである。

第 2 の問題はリズム認識である。仮に多重音解析により正確な MIDI データが得られたとしても、音楽的な演奏にはテンポの緩急、音符の長短が含まれることが多い。隣り合った音符間のオンセット時刻差 (IOI と呼ぶ) を単純に量子化して最も近い音符長で表すと図-1(d)のようになる。要するに、全音符が 16 分音符の 16 倍の長さで弾かれていない以上、分解能を 16 分音符に設定したらタイや符点だらけの、しかし実は演奏に極めて忠実な楽譜になる。目標とする図-1(a)の原楽譜にはほど遠い。

その他の問題は、派生音の # ♭ などの選択、拍子推定、調号推定、ピアノならば左右の手の楽譜への分割やペダリング、多種類楽器の合奏ならば音色

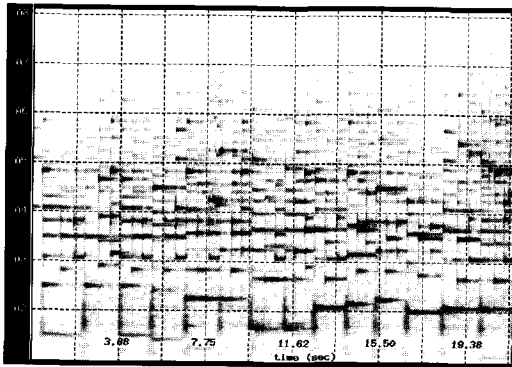
\* Automatic music transcription: A review.

\*\* Shigeki Sagayama (Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, 113-8656)

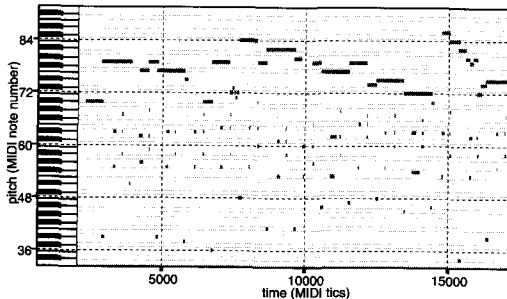
\*\*\* Hirokazu Kameoka (NTT Communication Science Laboratories, Atsugi, 243-0198)



(a) 楽譜例 (F. Chopin 作曲 Nocturne, op. 9, no. 2)



(b) 上記のピアノ演奏オーディオ信号のスペクトログラム



(c) 上記の「耳コピー」MIDI データのピアノロール表示



(d) 上記 MIDI データからの単純な楽譜化例

図-1 楽譜、演奏音楽信号、MIDI データ、従来の採譜の例

を手がかりにしたパートへの分割、各音符のアーティキュレーション指定の推定、クレシェンドやスラーやテヌートなどの奏法の指定、Andante などの速度や表情の指定など、原楽譜を復元するには様々な情報が必要である。これらの多くは、専門家にも一意に決定できないものも多い。音声認識に喩えると、原文がどこで改行されているか、漢字か仮名のどちらで書かれているか、ふりがなはあるか、などに相当する問題まで含んでいる。

これらをすべて扱った自動採譜システムはまだ存在しないし、本稿で扱うには問題が広すぎるので、書籍 [1, 2] や解説 [3] などを参照されたい。本

稿では主に上記の第 1 の問題について議論する。

### 3. 自動採譜と音の群化

我々人間は、多数の音が混じり合った音響信号から、個々の音を難なく聴き分けることができる。足し算が不可逆であるのと同じように、いったん重畳されてしまった波形から個々の波形を復元することは一般には困難である。にもかかわらず、混じり合っている個々の音の「輪郭」や「境界」を正確に把握できるのは、人間の聴覚の「アルゴリズム」がいかに優秀であることを示している。

この逆問題を、人間がどのようなアプローチにより解いているのかについては未解明な点が多い。両耳に入ってくる二つの波形の微妙な違いに基づいて知覚される、波源位置の情報は音を聴き分けるための手がかりの一つに違いないが、我々はモノラル録音された音響信号からですら個々の音を聴き分けられる能力を持っている。このことは、人間には空間の手がかり以外の手がかりに基づく何等かの音の群化<sup>1</sup>メカニズムが備わっていることを示唆し、この困難な逆問題を人間がどうにかして解いているという事実は、音を聴き分ける原理を追究することへの動機となっている。

更に、我々は音楽を聴くとき、その音響信号には様々な現象的な「揺らぎ」があるにもかかわらず、どのような楽器が、どの音高で、どのようなビート、リズムで奏でられているかを容易く理解できる。これは、低次の機能による信号処理モジュールだけでなく、知識に基づく高次の機能によるパターン処理モジュールを総動員した、人間の優れた音の認識メカニズムのなせる業である。計算機に、音楽の音響信号から自動的に楽譜化させることを自動採譜 [4, 5] という。これは、人間の低次機能と高次機能を統合した圧倒的な情報処理メカニズムに迫ろうとする大いなる挑戦である。

本項目では、特に音の群化の問題に焦点を当て、問題を整理しながらその計算論的アプローチに関する近年の取り組みについて紹介する。

<sup>1</sup>人間は、物体を見たときに、どこまでをひとまとまりなのかを捉え、物体の輪郭を把握することができるのと同じように、音を聞いたときにも、どこまでをひとまとまりなのかを捉え、音の「輪郭」を把握することができる。このように、音の「輪郭」を把握することを「音の群化」と言い、このように形成されたひとまとまりの音の「塊」のことを「音脈」と言う。

### 3.1 基本周波数推定（周波数方向の群化）

音の群化の問題と多重音の基本周波数推定の問題との間には、極めて密接な関係がある。このことを明快にするため、分かり易い例題として単一音のパワースペクトルから基本周波数を推定する問題について考えよう。もし信号が純音の場合、パワースペクトルのピーク周波数が基本周波数に対応する（図-2(a)）が、一般の周期信号には複数のピークがある（図-2(b)）。そして複数あるピークのうち最大のピークの周波数が必ずしも基本周波数に対応するとは限らない（図-2(c)）。また、基本周波数成分はいつも大きいとは限らないため、複数あるピーク周波数のうち最も低い周波数を基本周波数と見なすのは頑健なやり方ではない（図-2(d)）。以上より、基本周波数を推定するためには、スペクトルピークのような限られた情報だけで済ませようとするのではなく、対象とする音の信号波形やスペクトル構造の全体を手がかりにした頑健な方法が必要になる（単一音の基本周波数推定すら容易でないことは長い研究の歴史が物語っている [6]）。しかしながら、複数の信号が混合されて観測される音響信号には、どの成分がどの音に帰属するかという情報が欠落しているため、基本周波数を推定するための重要な手がかりが得られないのである。従って、音の群化の問題が解かれない限り、個々の基本周波数を推定することは容易ではないわけである。一方で、もし、個々の音の基本周波数が既知であれば（極めて特異な状況であるが）、各音に由来する成分の見当がつくため、音の群化の問題は大幅に解き易くなる。すなわち、音の群化の問題を解く手がかりになる基本周波数の情報が、音の群化が解かれない限り安定的に求められない、といういわゆる「鶏と卵」の状況に陥るのである。

この問題に対しては、音に関する先験的知識（調波性やスペクトル概形の仮定）を利用するのが主流な常套手段となる。例えば、観測音響信号をあらゆる基本周波数の音の重みつき混合としてモデル化してその重みを推定するアプローチ [7]、スペクトルクラスタモデルを用いて音の群化と基本周波数推定を反復的に行うスタイルのアプローチ [8, 9] や対数周波数領域で調波構造のシフト不変性を仮定して対数周波数スペクトルを調波構造パターンで逆畳み込みするアプローチ [10] などが試みられ

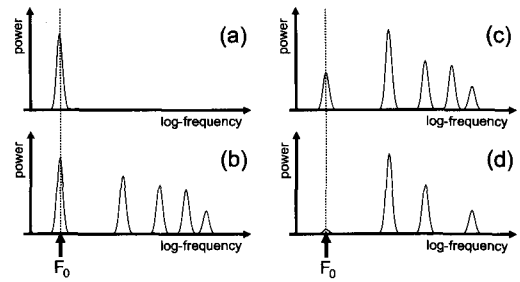


図-2 基本周波数推定の問題

ている。このほかにも、多重音から基本周波数を推定する手法は膨大にあるので、より詳しい動向については他の著書 [11] を参照されたい。

### 3.2 計算論的聴覚情景分析（時間方向の群化）

前節では、暗黙のうちに非常に短い時間区間における波形から個々の音に群化する問題について考えていた。我々人間でも、数十ミリ秒程度の混合信号から個々の音を聴き分けるのは必ずしも容易ではなく、容易に聴き分けるためにはある程度の信号の長さが必要になる。前節で考えていた問題は、周波数方向の群化と呼ぶものに相当し、人間はそれだけでなく音の時間的な連なりを形成する時間方向の群化も同時に行っているとされる。

近年、聴覚情景分析 [12] と呼ぶ心理学的アプローチの枠組みによって徐々に明らかになってきた人間の音の群化メカニズムに関する知見を積極的に利用して、音の群化問題の解決を図ろうとする試みが進められており、その枠組を総称して計算論的聴覚情景分析（Computational Auditory Scene Analysis: CASA）と呼ぶ。具体的には、知識を利用しない聴覚の低次の音の分離能力に関して、音響信号はスペクトログラムに似た要素に「分解」されること、同じ音源に由来する要素は「群化」されて音脈を形成すること、群化のされ易さ（分凝要件）は、(1) 調波性、(2) 調波成分の立ち上がりの共通性、(3) 調波成分の周波数及び振幅変化の共通性、(4) 成分の連続性、(5) 時間周波数の近接性、(6) 音源位置の共通性などに関係する、ことなどが心理実験を通して示されている。瞬時瞬時において調波の関係にある周波数成分を一つの音としてグルーピングすることを周波数方向の群化といい、それらを分凝要件 (2)～(5) に基づいて継時的にグルーピングすることを時間方向の群化と言う。これによって、例えば、二つの音声の基

本周波数軌跡がある時点で交差していたとしても、本来は分離不能なはずの交差の瞬間における個々の音声信号の各周波数成分がどのように重なっているかを前後の時刻から推論できるようになるわけである。CASAの目的は、このような人間の低次機能による音の群化メカニズムを模倣することであり、上記の分解と群化のプロセスを、分凝要件に関係する物理量を用いてアルゴリズムとして実現し、音脈の認識に有用な特徴量（基本周波数など）を抽出したり、目的音に相当する音脈の再構成を行うことである。

その具体的なアプローチとしては、周波数方向の群化に相当する処理により各離散時刻において個々の構成音の瞬時特徴成分（例えばスペクトルや基本周波数）を抽出したのちに、マルチエージェントシステム [7, 13] やベイジアンネットワーク [14] や隠れマルコフモデル [15] や Kalman フィルタ [16–19] などの手段を通して、時間的にどの成分が同じ一連の音に対応しているかを瞬時特徴成分の時間的滑らかさなどを評価尺度にして推定する方法が主流である。また、一方で、分凝要件 (1)~(5) から逸脱しない範囲の自由度を持った時変スペクトルを直接的にモデル化し、これを混合したもので観測時間周波数スペクトルにフィッティングする、周波数方向及び時間方向の群化を同時最適化問題として定式化されたアプローチ [9, 20–25] も考案されている。

### 3.3 スパース成分分析（記憶に基づく群化）

ところで我々は、ユニゾン（同一音高又はオクターブ違い）で弾かれたピアノとヴァイオリンの音を聴き分けることができる場合がある。一定の時間連続して一方の音の調波成分が完全に他方の音の調波成分と重なってしまうこの状況では、前後の時刻から調波成分の重なり具合を推論することが難しいため、これまで述べてきた群化メカニズムとは別の何等かのメカニズムが存在している可能性が示唆される。極めてわずかな基本周波数の違いによって二つの信号の間に干渉が生じており、それを手がかりにしている可能性もあるが、それよりもピアノやヴァイオリンがどのような音色であるかを漠然と記憶していて、それに基づいて個々の音脈を推論するような働きが関与していると予想される。

ピアノの音とヴァイオリンの音を過去にもっと

容易に聴き分け易い状況で聴き分けた経験があったとして、その経験から、それぞれの音響的特徴に関する「辞書」が作られているとすると、この「辞書」はユニゾンのような困難な状況においても高い精度で音を群化するための有用な手がかりになる。そして、このように音の群化が高い精度でなされるたびに、信頼性の高い学習データを得たことになり、「辞書」の再学習が可能になる。

スパースコーディングの考え方を基礎として、以上のような観点で音の群化の問題を捉えたアプローチが近年脚光を浴びている。具体的には、各時刻で観測される混合信号ないし混合スペクトルを、時刻によらず共通な基底セット（辞書）の重みつき和によってモデル化し、できるだけ重みをスパース<sup>2</sup>に、かつ復元誤差を小さくするように基底と重みを学習すると、一つ一つの基底が最大限の情報量を持った効率的な分解表現へと誘導される形となり、結果、各基底が観測中に頻発する信号あるいはスペクトルのパターンとなるはずだとする考え方である [26]。通常、基底と重み<sup>3</sup>は交互に更新されるため、丁度上述のとえと同様な反復学習が行われることになる。非負値行列分解 (Non-negative Matrix Factorization; NMF) は、基底と重みをいずれも非負制約のもとで学習する方法で、効率的な学習アルゴリズムが存在する点、非負制約以外の制約がなくとも副次的に重みがスパースになる基底の解が得られる点が特徴的である [27]。また、NMFを応用し、混合信号中の個々の音の間で共通しているスペクトル包絡や微細構造を自己組織的に発見する方法も試みられている [28]。

## 4. リズム認識とテンポ推定

第2の基本的な問題は各音符の音価決定である。従来は基本単位長の整数倍に音長を量子化する手法が一般的であったが、それでは図-1(d)のような楽譜になりかねない。これは演奏に忠実な楽譜であっても、ユーザが求める楽譜ではない。

量子化の処理を行うために演奏からテンポを推

<sup>2</sup>ほんの一部の基底の係数だけが大きな値を持ち、それ以外の係数は0であることを「重みがスパースである」と言う。

<sup>3</sup>重みの更新は、前段で更新された辞書信号あるいは辞書スペクトルを観測にフィッティングさせる操作に相当し、すなわち音の群化処理に外ならない。

定する研究は、主にテンポトラッキングと呼ばれる研究テーマとして研究されており、Cemgilら [29] は、テンポを隠れ変数としたカルマンフィルタを用いて、ポピュラ音楽のピアノ演奏のMIDI信号に対してテンポ推定が可能であることを示している。リズムとテンポを同時に推定する方法として、Raphael [30] は確率モデルを用いてリズムとテンポを同時に推定する手法を提案し、クラシック音楽のピアノ演奏のMIDI信号に適用可能であることを報告している。我々は、音声認識からの視点で、演奏される音長のゆらぎを確率的に扱うと同時に、ありふれたリズムや稀なリズムなどのリズムの出現し易さを学習した確率モデルを用いる、モデルベースのアプローチ [31-35] を取った。ここでの考え方は、大部分の音楽は典型的な「リズム単語」の連結でできていて、それにテンポや個別のゆらぎが掛けられて実際に演奏された個々の音長として観測される、というものである。そうすると確率モデルに基づく大語彙連続音声認識と同型の問題となり、HMM (Hidden Markov Model, 隠れマルコフモデル) によりリズム語彙をモデル化し、リズムパターン系列の統計的言語モデルを用いることで、意図されている音価が推定できる。

## 5. 今後の自動採譜への展望

音響信号を入力としてその内容を推定する技術の代表は音声認識であるが、自動採譜も連続音声音声認識と同型の問題と考えることができる。ただし、注目する情報が正反対あるいは双対関係にある。音声認識では音韻性を担うスペクトル形状に注目し、韻律情報を担う基本周波数や継続時間などの情報は捨てられることが多い。一方、楽器演奏音楽の自動採譜では、基本周波数は音高、時間パターンはリズムに関係して重要であるのに対し、頻繁な音色の変化は副次的で当面の対象ではない。両者が注目する情報は相補的である。

図-3に連続音声認識の概念図を示す。この図において音声信号入力を音楽信号入力に、音素モデルを音符モデルに、言語モデルを音楽モデルに置き換えれば、音声認識出力は自動採譜出力に置き換わるだろう。ここで特徴抽出部では、多重音解析による各音の基本周波数とオンセット時刻の情報の抽出を行う。音符モデルは、ピッチ変動(ビブラートなど)、音長変動、音色変動を表現する確

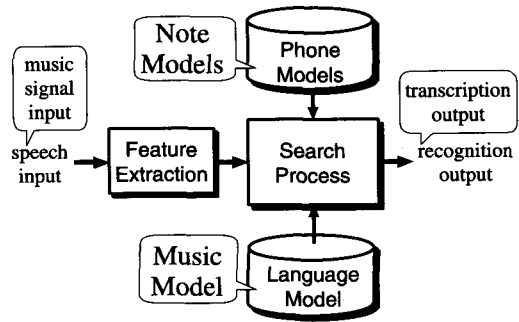


図-3 連続音声認識システムの構成を自動採譜へ転用

率モデルである。音楽モデルには、リズム語彙、和声構造、旋律、楽曲構造、曲スタイルなど、音楽に関する常識すべてが該当し、それらをいかに確率モデルとして表現するかが課題である。探索過程は、入力から抽出された特徴量を、音符モデルと音楽モデルの両方から最も尤もらしく説明できる(最尤の)原楽譜を効率良く探索する。

ここまで大きな野望はまだほとんど手がつけられていないのが現状である。しかし、部分問題はその方向を目指していると考えてよく、例えば和声解析にはHMMを使う [36] のが昨今の主流になって来ている。

音声認識が音声信号処理と言語処理を統合した問題設定の頂点であり続けるように、自動採譜は音楽信号処理と音楽情報処理の個別の問題を包含する問題の頂点であり、むしろ音声分野より応用の裾野は広いかもしれない。リズム認識の確率モデルは楽譜追跡や自動伴奏に、和声モデルは自動作編曲や検索に、旋律モデルは類似曲検索に、曲スタイルモデルは自動ジャンル分類などにそれぞれ応用できそうである。音声研究者に音楽研究を勧める所以である。

## 謝 辞

本稿執筆にあたって情報提供された武田晴登博士(現・ソニー株式会社勤務)に感謝する。

## 文 献

- [1] 長嶋洋一, 平賀 譲, 橋本周司, 平田圭二(編), コンピュータと音楽の世界—基礎からフロンティアまで—, bit 別冊 (共立出版, 東京, 1999).
- [2] A. Klapuri and M. Davy, Eds., *Signal Processing Methods for Music Transcription* (Springer-Verlag, New York, 2006).
- [3] 片寄晴弘, 井口征士, “知能採譜システム,” 人工知能学会会誌, 5, 59-66 (1990).
- [4] J.A. Moorer, “On the segmentation and analysis of continuous musical sound by digital computer,”

- Ph.D. Thesis, Stanford University (1975).
- [5] H. Katayose and S. Inokuchi, "The Kansei music system," *Comput. Music J.*, 13(4), pp. 72-77 (1989).
  - [6] W. Hess, *Pitch Determination of Speech Signals* (Springer-Verlag, Berlin, 1983).
  - [7] M. Goto, "A real-time music-scene-description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals," *Speech Commun. (ISCA J.)*, 43, 311-329 (2004).
  - [8] H. Kameoka, T. Nishimoto and S. Sagayama, "Separation of harmonic structures based on tied Gaussian mixture model and information criterion for concurrent sounds," *Proc. 2004 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP 2004)*, Vol. 4, pp. 297-300 (2004).
  - [9] H. Kameoka, "Statistical approach to multipitch analysis," Ph.D. Thesis, The University of Tokyo (2007).
  - [10] S. Saito, H. Kameoka, K. Takahashi, T. Nishimoto and S. Sagayama, "Specmurt analysis of polyphonic music signals," *IEEE Trans. Audio Speech Lang. Process.*, 16, 639-650 (2008).
  - [11] A. de Cheveigné, "Multiple F0 estimation," in *Computational Auditory Scene Analysis: Principles, Algorithms and Applications*, D.-L. Wang and G.J. Brown, Eds. (IEEE Press / Wiley, New York, 2006).
  - [12] A.S. Bregman, *Auditory Scene Analysis* (MIT Press, Cambridge, Mass., 1990).
  - [13] T. Nakatani, M. Goto and H.G. Okuno, "Localization by harmonic structure and its application to harmonic sound segregation," *Proc. 1996 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP '96)*, pp. 653-656 (1996).
  - [14] K. Kashino, K. Nakadai, T. Kinoshita and H. Tanaka, "Application of the Bayesian probability network to music scene analysis," in *Computational Auditory Scene Analysis*, D.F. Rosenthal and H.G. Okuno, Eds. (Lawrence Erlbaum Associates, Mahwah, N.J., 1998), pp. 115-137.
  - [15] M. Wu, D.L. Wang and G.J. Brown, "A multipitch tracking algorithm for noisy speech," *IEEE Trans. Speech Audio Process.*, 11, 229-241 (2003).
  - [16] 西 一樹, 安部素嗣, 安藤 繁, "聴覚情景分析のための多重ピッチ追跡と調波分離アルゴリズム," 計測自動制御学会論文集, 34, 483-490 (1998).
  - [17] 鶴木祐史, 赤木正人, "聴覚の情景分析に基づいた雑音下の調波複合音の一抽出法," 信学論, J82-A, 1497-1507 (1999).
  - [18] 安部素嗣, 安藤 繁, "共有 FM-AM の時間周波数統合に基づく聴覚情景分析 (I)—Lagrange 微分特徴量とその周波数統合—," 信学論, J83-D-II, 458-467 (2000).
  - [19] 安部素嗣, 安藤 繁, "共有 FM-AM の時間周波数統合に基づく聴覚情景分析 (II)—最適時間軸統合とストリーム音の再合成—," 信学論, J83-D-II, 468-477 (2000).
  - [20] H. Kameoka, T. Nishimoto and S. Sagayama, "A multipitch analyzer based on harmonic temporal structured clustering," *IEEE Trans. Audio Speech Lang. Process.*, 15, 982-994 (2007).
  - [21] 亀岡弘和, J. ルルー, 小野順貴, 嵯峨山茂樹, "調波時間構造化クラスタリングによる CASA へのアプローチ," 音響学会聴覚研究, 36(7), H-2006-103, pp. 575-580 (2006).
  - [22] H. Kameoka, T. Nishimoto and S. Sagayama, "Audio stream segregation of multi-pitch music signal based on time-space clustering using Gaussian kernel 2-dimensional model," *Proc. 2005 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP 2005)*, Vol. 3, pp. 5-8 (2005).
  - [23] J. Le Roux, H. Kameoka, N. Ono, A. de Cheveigné and S. Sagayama, "Single and multiple pitch contour estimation through parametric spectrogram modeling of speech in noisy environments," *IEEE Trans. Audio Speech Lang. Process.*, 15, 1135-1145 (2007).
  - [24] 宮本賢一, 亀岡弘和, 西本卓也, 小野順貴, 嵯峨山茂樹, "調波構造・時間包絡・音色の統合的クラスタリングによる楽音分析," 情報処理学会研究報告, 2007-MUS-71-25, pp. 155-160 (2007).
  - [25] 糸山克寿, 後藤真孝, 駒谷和範, 尾形哲也, 奥乃 博, "音色特徴量分布の利用による調波・非調波併用モデルのパラメータ推定," 情報処理学会研究報告, 2007-MUS-71-26, pp. 161-166 (2007).
  - [26] S.A. Abdallah and M.D. Plumbley, "Unsupervised analysis of polyphonic music using sparse coding," *IEEE Trans. Neural Networks*, 17, 179-196 (2006).
  - [27] P. Smaragdakis and J.C. Brown, "Non-negative matrix factorization for music transcription," *Proc. 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2003)*, pp. 177-180 (2003).
  - [28] 亀岡弘和, 柏野邦夫, "複合ソースフィルタモデルによる音響信号の三要素テンソル分解," 信学会 2008 年総合大会講演集, AS-5-5, pp. S-56-S-57 (2008).
  - [29] A. Cemgil, B. Kappen, P. Desain and H. Honing, "On tempo tracking: Tempogram representation and Kalman filtering" *J. New Music Res.*, 28, 259-273 (2001).
  - [30] C. Raphael, "Automated rhythm transcription," *ISMIR*, pp. 99-107 (2001).
  - [31] 齋藤直樹, 中井 満, 下平 博, 嵯峨山茂樹, "隠れマルコフモデルによる音楽演奏からの音符列の推定," 情報処理学会研究技術報告, 99-MUS-33, pp. 27-32 (1999).
  - [32] 大規知史, 齋藤直樹, 中井 満, 下平 博, 嵯峨山茂樹, "隠れマルコフモデルによる音楽リズムの認識," 情報処理学会論文誌, 43, 245-255 (2002).
  - [33] H. Takeda, N. Saito, T. Otsuki, M. Nakai, H. Shimodaira and S. Sagayama, "Hidden Markov model for automatic transcription of MIDI signals," *Proc. MMSP* (2002).
  - [34] 武田晴登, 西本卓也, 嵯峨山茂樹, "テンポ曲線と隠れマルコフモデルを用いた多声音楽 MIDI 演奏のリズムとテンポの同時推定," 情報処理学会論文誌, 48, 237-247 (2007).
  - [35] H. Takeda, T. Nishimoto and S. Sagayama, "Rhythm and tempo analysis toward automatic music transcription," *Proc. ICASSP 2008*, Vol. IV, pp. 1317-1320 (2007).
  - [36] 川上 隆, 中井 満, 下平 博, 嵯峨山茂樹, "隠れマルコフモデルを用いた旋律への自動和声付け," 情報処理学会研究技術報告, 99-MUS-34, pp. 59-66 (2000).