

5 和声解析・リズム認識・自動伴奏・運指決定



嵯峨山 茂樹

東京大学大学院情報理工学系研究科

まえがき

タイトルに並ぶキーワードは、一見無関係に見えるかもしれない。しかし、確率モデル（特に隠れマルコフモデル、HMM）の観点から、これらには共通したアプローチが有効であることを、筆者らの研究を通じてこの解説で紹介したい。話題は、和声解析、和音推定、調性推定、転調検出、旋律への和声づけ、リズム認識（音価列推定）、拍子推定、拍節推定、テンポ推定、小節分割、楽譜追跡、自動伴奏、自動運指決定などである。音声認識分野で発展してきた確率統計的な定式化や手法は、音楽情報処理でも大いに役立つ。

和音認識

■ 和音と和声進行

和音 (chord) や和声進行 (chord progression) あるいは和声 (harmony) は、クラシックでもポピュラーでも、調性のある音楽においては音楽の構成原理の最重要要素の1つである。もちろん、これは西洋音楽を源流に持つ音楽を前提にしているわけで、日本伝統音楽を始めとする民族音楽では和声の概念がないものもあるが、現在の音楽市場の大部分が調性のある和声音楽であることは間違いない。

人が音楽を聴くときには、旋律や歌唱以外の音には注意を払っていないつもりでも、実は俗に「響き」と呼ばれる一種の印象として和声の効果は十分感じ取っている。音楽に慣れた人なら、聴きながらコード進行を書き下すこともできるし、再現して演奏することもできる。

では、機械に和声を聴き取らせることはできるだろうか？ それが和音検出問題 (chord detection) である。

直接の応用としては、音楽信号あるいは楽譜からギターなどのコードを自動作成してくれれば便利であるのはもちろんだが、それとともに音楽情報検索 (Music

Information Retrieval, MIR) や音楽分類等、音楽の内容に基づいた解析に対しても和音は有用と考えられている。具体的には、曲間の類似度、曲のスタイルやジャンル、異なる編曲の同一音楽の識別問題 (cover song identification) など、音楽の検索や著作権行使などへの利用が考えられる。

人間でも、鳴っているすべての音を認識 (多重音解析) するのは難しくても、和声進行は感じ取れることが多い。人間が採譜をするときでも、和声的な曲は和音構成音を多く含むから、和声進行が分かれば全構成音を推定する上でも役立つ。また和声進行を把握することで非和声音にも注目できる。つまり、和声の把握は自動採譜の1段階としても重要である。

■ 和声理論

和声に関しては「和声学」という分野があり、作曲の原理・音楽の文法とも言える最重要な音楽理論であり、音楽大学の主要な科目の1つになっていて、既存の曲の和声解析や、与えられた旋律への適正な和声づけなどを学習する。クラシック音楽の理解に不可欠な古典的な和声学と、ジャズ和声から始まったポピュラー和声（コード理論）の両方があるが、本質は大きくは変わらない。一方、複旋律音楽 (ポリフォニー) の理論として対位法という理論があり、現代の目から見るとその方が複雑で高度に見えるが、面白いことに、西洋音楽史では、そちらが先に確立し、そのうちに複数の旋律を同時に重ねたときの響きの効果に次第に関心が移り、16～18世紀ころに徐々に和声学が確立されたとされている。

和声学は、基本的には長調や短調などの調性を基礎として、その音階音のどれかを根音として選び、その上に三度間隔で音を重ねて、三和音 (triad) を作る。第三度の音が根音から長三度ならば長三和音、短三度ならば短三和音と呼ばれる。つまり、根音の動きと、その上に積み重ねる三度音と五度音による響きと、それらの間の文法

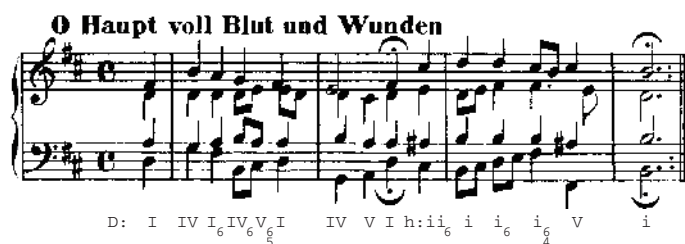


図-1 和声構造が明瞭なコラール編曲例
(J.S. Bach 「マタイ受難曲」から)



図-2 非和声音が多い例—赤丸は倚音，青丸は経過音(G. Verdi 「椿姫」第一幕から)

のような仕組みを覚えれば音楽が作れるという経験則を理論化した，西洋文化らしい合理的な理論である。いわば，現代の世界に広がっている音楽という共通言語の文法であり，それを知っても知らなくても，プロもアマチュアも，万人が大いにその恩恵を受けていることになる。

和声学は，和音の転回形，七の和音や九の和音，変化和音や借用和音，各部の声部進行規則と禁則と続く。図-1は，比較的単純な和声構造の例である。和声記号はいろいろな流儀があるが，ここでは古典的な記法の1つを用いて示した。自動和声解析では，このような記号が出力できればよいということになる。しかし，転調の判断，和音の複数の解釈などの問題もある。

さらに厄介なのは非和声音である。和音はド・ミ・ソを想定していても，旋律はド，ミ，ソの音だけ使っているはおもしろくないから，和声内音だけでなく非和声音も使う。和声学では，音楽的に良好な非和声音のルールを作っていて，経過音，刺繍音(補助音)，繋留音，先取音，倚音などの分類と，使用法の規則がある。一口で言えば，和音の外にある音だが，和声に吸い込まれるような使い方をすれば，旋律と和音の両方を活かした旋律が作れるというわけである。図-2はその例の1つだが，ヴェルディ特有の倚音の多い旋律が，明瞭な和声進行の中で魅力的な旋律を作っている。

和声学は音楽の文法と言えるものの，文法自体が単純でなく，さらに実態は文法理論が追いつかないほど多様である点は，自然言語の文法と同じである。文法は極論すれば目安でしかない。

■ HMM による和声と調のモデル

コンピュータで和声解析を行う問題を考えよう。まずは，楽譜あるいは演奏の MIDI 信号が与えられたとして，

その和声解析ができるか。

和声進行はどんな和音の次にどんな和音がきてもよいわけではなく，つまり和音の並びは決まてはいないが，ランダムでもなく，確率的に偏りがある。これは曲のスタイルや作曲家の個性によっても異なるだろう。一方，旋律と和音を実現するためには音符が置かれる。たとえばドミソの和音であれば，旋律にも伴奏にもドかミかソの音が使われる確率が高いと予想されるが，非和声音も出現する。

ある和音の次にある和音が続くのは確率の問題と考えて，和声進行は(正しくは n 重)マルコフ過程に従うと見なすことにしよう。和声進行が源泉となつて，和音に関連した旋律や音符を生み出す。これは隠れマルコフモデル(Hidden Markov Model; HMM)と呼ばれるものになっている。HMMは隠れ状態間で確率的に遷移(マルコフ遷移)し，各隠れ状態から観測量が確率的に出現するモデルで，80年代から音声認識分野で大きく発展したが，いまや非定常時系列の強力なモデルとして多くの分野で広く使われている。1999年に川上らの研究¹⁾で和声モデルへ導入された。

HMMによる和声モデルは，和音を状態として，各時刻に状態遷移するかとどまる(自己遷移)かして，そのたびに音符を生み出すような，確率的な作曲のモデルということになる。もちろんこれだけで常に良い曲が作れるわけではないが，与えられた音楽をこのモデルから生成されたものと見なせば，逆にそれらの音符を生み出した和声進行を推定することは，確率的な逆問題と考えられる。図-3にこの概念を示す。判定の原理は Bayes の定理に基づけばよい。原因が S で観測が X であれば，

$$\hat{S} = \operatorname{argmax}_S P(S|X) = \operatorname{argmax}_S P(X|S)P(S)$$

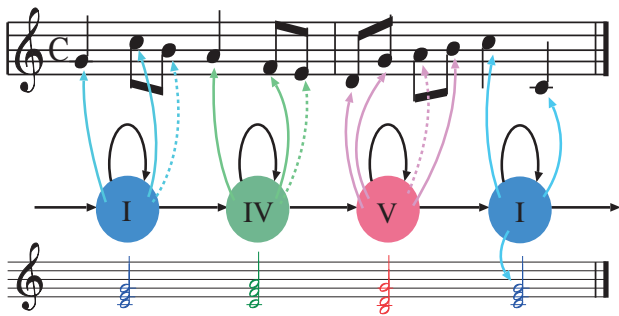


図-3 HMMによる和声から旋律の生成のモデル。和声進行は状態遷移で表され、旋律は状態からの確率的出力としてモデル化される。

により、事後確率最大の原因推定ができる。和声解析では、 S が和声、 X が音符である。この解は、Viterbi アルゴリズムと呼ばれる効率の良い計算法で求められる。

以上を行うには、HMMのパラメータをあらかじめ学習する必要がある。和声ラベルがつけられた楽曲データから、Baum-Welch アルゴリズムのような学習アルゴリズムを用いて、和声と音符との確率的な関係を統計的に求められる。

■ HMMによる旋律の調性推定

これを調性の推定に拡張しよう。同一の和音は複数の調で出現するが、その確率は大きく異なる。また和声進行も調によって異なる。そもそも調性は音階の概念というよりむしろ和声学の概念であり、どのような調として和声進行がうまく説明できるかが調性であり、和声解析より先に調を決めることは原理的にできない。そこで、相互依存の関係にある調と和音を同時に推定するため、図-4のような和声モデルを用いる。平行長短調を同一調と扱い12種類の調ごとに、和音がエルゴディックに遷移するモデルを作成する。なお、ここではさらに1つの調の中で許される和声進行が和声進行の常套句の集合で確率的に表されるというモデルも含まれている。

■ 音響入力との和音認識

以上は楽譜あるいはMIDI信号のような離散記号が入力である場合(symbolic)であった。一方、音楽音響信号を入力とする場合(audio)も、基本的には同様に考えればよい。音楽入力信号を短時間フレームごとにスペクトル分析し、それを、フレームごとに状態遷移(自己遷移を含む)するHMMからの確率的な出力と見なす。これで、音響入力から和声進行を推定する確率的逆問題として定式化できる。しかし、音楽音響信号のスペクトルは倍音が多重に重なり合っており、和音の構成音を知ることは容易でない。その例を図-5に示す。赤線で旋律中のいくつかの音符と対応するスペクトル中の周波数と時刻を

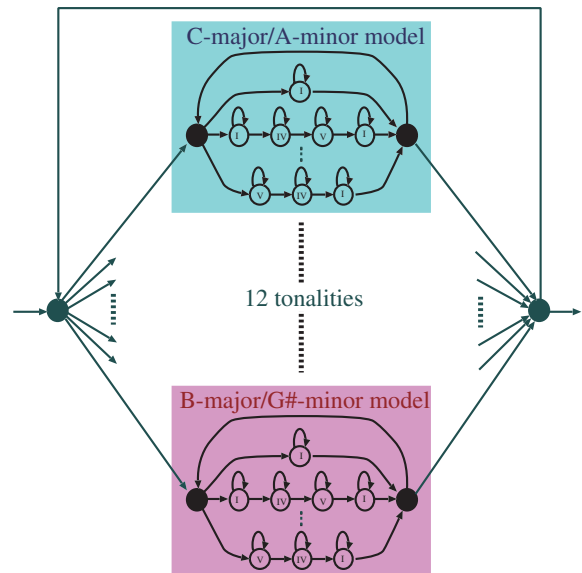


図-4 HMMによる調性と和声のモデル。調性内、調性間ともに確率的に状態遷移する。

示してあるが、スペクトルから音符を推定することは容易でないことが理解されるだろう(多重音解析に関しては本特集の亀岡氏の解説を参照されたい)。

和音は、さまざまなオクターブにわたって演奏されたり、いくつかの転回形や開離形、密集形などさまざまな音高配置で演奏される。このような和音の音高配置によらない特徴量として、クロマ特徴量がよく使われる。これは、パワースペクトルを半音名ごとに複数オクターブ間で足し合わせることで得られ、それぞれの半音ピッチに対応するパワーを表している。2003年にShehらは、これを特徴量とするHMMにより、音響入力の和音認識の研究を行った。

音楽音響信号を扱う上での問題の1つに、音楽は一般的に打楽器音やアタック音のようにピッチを持たない非調波的な音を含むことが多いという点がある。和音は音楽の調波的な要素であり、非調波的な音は邪魔であるので、調波音・打楽器音分離手法(HPSS, 本特集の小野氏の解説を参照)により得られた調波音から求めたクロマベクトルを特徴量に用いるのは効果的と考えられる。図-6にあるポピュラー曲の原信号スペクトル、HPSSによる打楽器抑制スペクトル、それぞれから抽出したクロマ特徴量の時系列(クロマグラム)を示す。内山ら²⁾はこの手法により2008年の音響入力と和音認識コンテスト(MIREX2008)で最高位を得た。

今後の方向としては、和音境界の精度を向上させるために、和音境界は拍の位置に置かれて打楽器音が重畳する確率が高いという仮説のもとに、打楽器音を和音境界の手がかりとして利用することなどが考えられる。

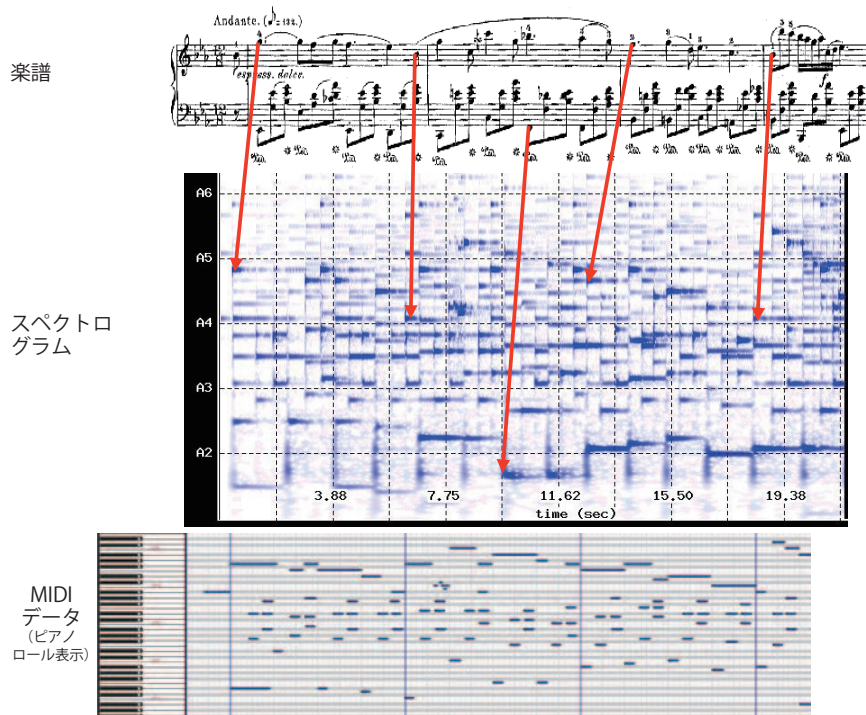


図-5 楽譜とスペクトルと手作業で作成された MIDI データの対比。倍音が複雑に重なり合って、音符が何であるかを知るの容易でない。特に、赤線で示した音符とスペクトルの対応に注目。(F. Chopin 作曲 ノクターン Op.9, No.2 冒頭から)

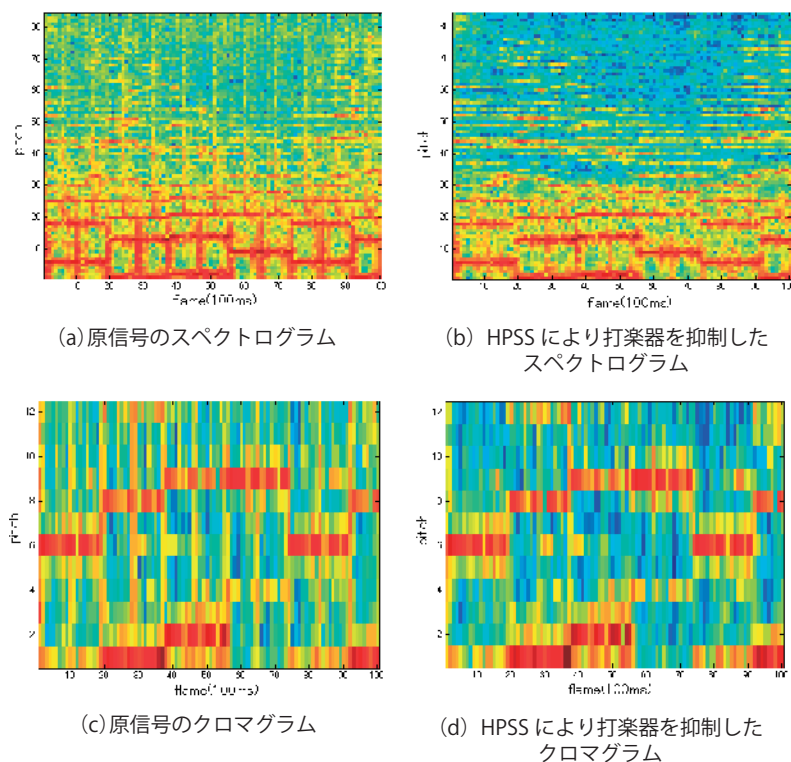


図-6 ポピュラー音楽のスペクトログラムとクロマ特徴量の HPSS の効果

旋律への自動和声づけ

■ 確率モデルによる自動和声づけ

次に、機械に旋律を与えて適切な和声づけをさせることはできるだろうか？ それが自動和声づけ問題である。和声づけは編曲の重要なステップである。音楽の愛好者は多く、旋律の作曲には多くの人が興味を持つが、それに適切な和声づけをし編曲するにはやや専門的な音楽

の知識や経験が必要なので、旋律に自動和声づけをして編曲ができる支援ツールには広い用途がある。作曲・編曲支援ツールとしては、楽曲データを入力すると、自動演奏や楽譜を清書する研究がされており、市販されているツールも多数ある。これらの従来の研究では、発見的な原理や規則に基づいたアプローチが多かった。

この問題は、前章までの問題の逆方向のように見えるが、実は同じ問題である。西洋音楽（特に古典派から近

代まで)やポピュラー音楽の大多数では、旋律の作曲時点で同時に、あるいは事前の和声設計の段階で、作曲家はその背後の和音を想定していると仮定できる場合が多い。分散和音形の旋律の場合はその典型である。これには「いや、私は和音など考えずに旋律を先に思いつく」などの異論があるだろう。その場合は、同じ旋律を生み出すときに、専門家ならば頭の中でどんな和音を想定しているはずか、と考えてもよい。和声と旋律の関係をHMMでモデル化すると、和声が隠れ状態であり、出力が旋律ということになる。図-3の概念はここでも同じである。

こう考えると、与えられた旋律への和声づけは、もととその旋律を生み出したはずの隠れた和声を推定する問題と言い換えられる。与えられた旋律を処理して適合する和声を決定するのではなく、その旋律を生成したと考えられる(事前に存在した)和声を、遡って推定する逆問題と考えられる。

すなわち、ある和声進行から旋律が生成される過程をHMMによって確率モデル化し、尤度最大の原理によって与えられた旋律を生成する遷移系列の中で最も尤度の高い系列をViterbi探索によって求めれば、その旋律を生んだ背後の和声進行を尤度最大の原理で推定することができる。候補となる和声進行は、最も単純にergodic HMMとすれば和音のbigram確率のみを与えることになるし、音声認識における語彙のように和声進行語彙を用意して連続音声認識のように尤度最大の単語列を求める方法も取れる。いずれの場合も、旋律が入力されたとき、この旋律を生成した尤度が最も大きい和声列を、隠れ状態列のViterbi経路探索により求めることができる。

■ 非和声音の種類を考慮した和声づけ

図-3に示したように、旋律は必ずしも和声内音だけで構成されるわけではない。さらに、旋律と和声の関係は、和声内音が使われる確率が高いだけという単純なものではない。非和声音を活用することで旋律が魅力的になることも多い。すでに述べたように非和声音にも一種の文法があり、和声学ではそれらの使用の仕方が理論付けられている。これらの生成確率は前後関係などを見なくてはならず、単純ではないが、それらを考慮したHMM出力確率が計算できればより精緻な和声づけができると期待される。また、各声部間では声部進行の禁則があり(ポップスでは守られていないことが多いが)、それも考慮に入れて和声を生成する必要がある。和声づけのツールとしては、和声づけ候補を複数提供し、使用者が好みの和声進行を選択することのできるものが望ましく、音声認識におけるN-bestアルゴリズム(Soong 1991)を用いて、尤度の高い複数の和声進行を得ることがで

きる。

■ HMMによる旋律の調性推定

旋律だけが与えられた場合の調推定も、転調の検出も、概念は和声推定と同様で、どのような調として旋律の背後の和声進行がうまく説明できるかが調性である。転調を検出することは、旋律に対して転調を含む和声づけをすることにほかならない。図-4に示すように調ごとのモデルを作成し、各調のモデル間(各調間)に遷移確率(転調確率)を持つような大きなHMMを用いることで、転調を含む旋律に対する和声づけが可能になる。すなわち、一定調の旋律に対しては、各調のモデルに旋律を入力し、最も尤度が高いモデルの調を旋律の調性として推定することができる。また、転調を含む旋律に対しては、転調モデルを用いて和声づけをした結果から、転調の検出を含めた曲の調の推定をすることができる。

リズム認識と自動採譜

■ 自動採譜におけるリズム認識問題

人の演奏を自動的に楽譜化できれば有用だろう。それを自動採譜と呼ぶ。音声分野における音声認識に相当するもので、音楽情報処理の究極の課題でもある。たとえば、いままで耳コピーに頼っていたジャズの即興演奏を楽譜化して自分で演奏することができる、大量の音楽データを楽譜データにして検索が高速で容易になる、移調されていても大丈夫、などさまざまな可能性が広がる。

自動採譜を行うためには、音楽演奏の1つ1つの音に対して音高(ピッチ)と音価(音符の長さ)の情報を得なければならない。音響入力の場合はまず多重音ピッチ解析が必要であるが、それについては本特集の亀岡氏による解説に譲ることにして、多重音解析の結果としてのMIDIデータ、あるいは人がMIDIキーボードなどを演奏して得られたMIDIデータから自動採譜を行うことを考えよう。

人間は意図的あるいは無意識にさまざまな要因で音長を変動させて演奏するので、一般に音長から音符の種類を一意に決定できない。このため、音長から楽譜にふさわしい音符の種類を決めるリズム認識が必要である。MIDIデータからの楽譜化では、市販のソフトでは、音長を「量子化する(quantize)」方法が従来用いられている。この場合は、自由に演奏した演奏データの楽譜化は難しい。一口で言えば、人は、全音符の長さを正確に16分音符の16倍の長さで弾けないからである。図-7はその一例である。図-8は、テンポをMM=96に保ったのべ50人の演奏データから得た、4分音符、8分音符、符点4分音符の音長ヒストグラムの例だが、このように音符



(a) 意図した楽譜 (Brahms 交響曲 2 番 3 楽章の冒頭の旋律)



(b) 単純な量子化による上記演奏の楽譜化



(c) HMM による楽譜復元

図-7 量子化によって得られる楽譜。演奏を正確に楽譜化しているが、欲しい楽譜ではない。確率モデルによりほぼ正確に復元できた。

長は分布する。テンポが自由である場合はこれらの分布は大きく重なり合う。

意図した音価に対応する音長からの揺らぎに対して補正する研究はいくつか報告されている。閾値処理をベースとして、ヒストグラム処理による基準拍を設定し、さらにフレーズの終わりは長めになるという音楽的なヒューリスティックルールを付加し、強制を行う手法 (片寄ら 1990) や、各音符長が比例関係にあることに着目した制約として、閾値処理に加えリズムを構文木と捉え文法的な強制を行う手法 (Ronguet-Higgins 1987) や、隣接する音長の比が有理数になれば安定するエネルギー関数により、安定するまで処理を繰り返す手法 (Desain 1989) などが報告されている。音響入力の場合は、曲のビートを解析するビートトラッキングをマルチエージェントによりモデルベースで音楽的解析を行う報告もされている (後藤 1997)。

■ HMM とリズム文法によるリズム・拍子・拍節の認識

しかし人間ならば、多少の揺らぎがあっても、意図されたリズムは感じとれる。多少訓練された人ならば、もっともらしく楽譜化できる場合が多い。これは、人間は常識的なリズムを知識として持っており、それを top-down 的に活用しているからであろう。そのような観点から、齋藤ら (1999) は HMM を用いた。つまり、当該音符の演奏が意図している音価を隠れ状態にし、実際に観測される音長を観測とする HMM によりモデル化した^{3), 4)}。まずは単旋律について考えると、音価の列はランダムではなく、たとえば符点四分音符の次には八分音符がくることが多いなど、リズムはいわば語彙を構成

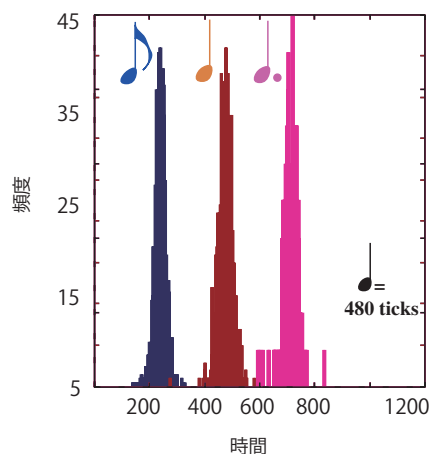


図-8 テンポ指定時の演奏の音長分布 (1/960 秒単位)

していると考えられる。それならば、連続音声認識と同様に、単語が接続してリズムのネットワークがあつて、音楽演奏はそのネットワーク上のある経路としてつながれた隠れ状態の列から、確率変動を伴った音長が出力され観測されていると定式化できる。したがって Viterbi アルゴリズムによって尤度最大の音価列を求めることで、演奏が意図していた音符が推定できることになる。

図-9 に示すのは連続単語音声認識に対応づけられるモデルで、出現する可能性のある短い音価系列、すなわち表-1 に示すような「リズム単語」を定義し、リズム単語の連鎖により曲が成立していると仮定するモデルである。モデルに含まれているリズム単語の連鎖のみ認識できる点で、モデルとしての拘束力は強いが、未知のリズム単語は扱えないという欠点もある。

以上の確率モデルのモデルパラメータは、多量の楽曲演奏データから学習することができる。もしテンポが一

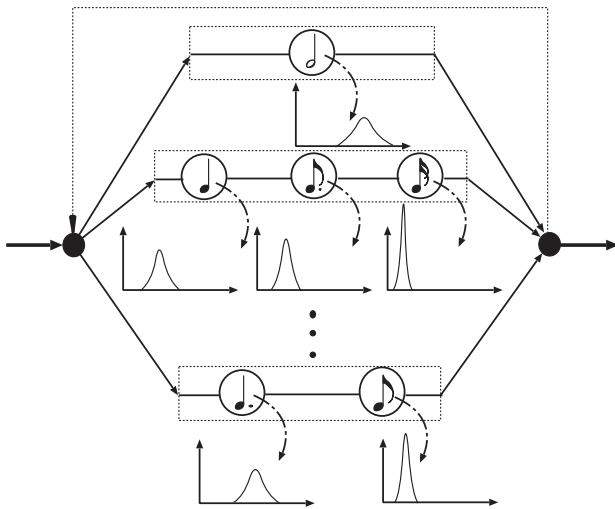


図-9 2拍単位のリズム単語モデル例

頻度順	1 小節単位	[%]	2 拍単位	[%]
1 位	♪ ♪ ♪ ♪	4.9	♪ ♪	16.7
2 位	♪. ♪ ♪ ♪	4.7	♪. ♪	12.4
3 位	♪. ♪ ♪ ♪ ♪	4.2	♪ ♪ ♪	11.9
⋮	⋮	⋮	⋮	⋮
10 位	♪. ♪ ♪ ♪	4.2	♪ ♪ ♪	11.9
⋮	⋮	⋮	⋮	⋮

表-1 音価列パターンの出現頻度例(4/4 拍子)

定で既知であるならば、以上のようにリズム語彙を揃えておけばそれらによるリズム解釈が得られることになる。しかし、実際はテンポは未知で変動する。それに対しては、前章の転調のモデルと同様に、図-10 のようなさまざまなテンポの語彙セットを用意することで、尤度最大の Viterbi 経路がリズムとテンポの両方の推定になる。

■ テンポ推定、拍子／開始拍／小節線位置推定

人間は変動テンポに追従して音価列を正しく理解できることが多い。HMM の場合は、テンポの異なるリズム語彙モデルを並列に持つことで、尤度最大の経路を Viterbi 探索することで、変動テンポに追従できる。

曲を聴いて楽曲の拍子と小節線の位置を推定することは、必ずしも容易ではない。2/4 拍子と 4/4 拍子、3 連符のみの 2/4 拍子と 6/8 拍子など、原理的に演奏からは区別できない場合も多く、さらに意図的に予想を覆すような楽譜も可能である。しかし人間は多くの場合、2 拍子系と 3 拍子系との区別や、上げ拍（アウフタクト、弱起）かどうか程度の推定は、比較的正確に行える。上げ拍も含む複数の拍子のリズム語彙の問題と考えて、尤度最大の Viterbi 経路探索を行えば、リズム語彙境界に小節線があると考えて拍子、拍節、小節線位置の推定も同時に行える。図-11 に興味深い誤りの例を示す。リズム

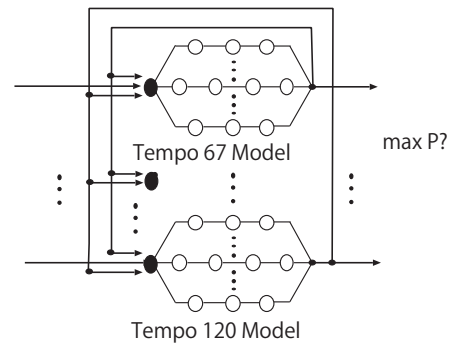


図-10 変動するテンポとリズム語彙のモデル

パターンからは 4 拍子と考えるのがより自然に見えるケースである。これの正解を得るには、音楽の 1 フレーズは 4 小節であることが多い、などの高レベルの知識が必要になる。

■ 多声部楽曲のリズム認識

以上は、単旋律の MIDI データから楽譜へ変換する問題と HMM による解法を述べたが、単旋律でなくピアノ曲などの多声部楽曲も扱う必要がある。武田ら(2003)は多声部の MIDI データの声部間 IOI (Inter-Onset Interval) を観測し、それもまた HMM の語彙モデルで表現できると仮定して、多声部の各音符の開始時刻の間隔パターンを HMM で処理する方法を取った。また、テンポの変動を許すために、連続する 3 つの音長の比を成分とし、成分の和を 1 に規格化したベクトルをリズムベクトルと定義することによって、複数のテンポの HMM を持たずに済ませる方法も取られている。

武田らはさらに、HMM によるリズム認識結果から抽出されるテンポの変動を滑らかな多項式の曲線でモデル化し、そのモデルに基づいて HMM を制御してリズム認識し、その結果によりテンポ曲線を更新するという反復解法を導いた。また、テンポが滑らかな曲線で表せない場合は区分化してテンポ曲線でモデル化し、そのモデルに基づいて再びテンポ区分を更新するような手法で、多声部のリズムとテンポ曲線とテンポ変化時点とを尤度最大の原理で同時推定する手法を研究した。

しかし、多声部楽曲のリズムの認識には、声部の動きや和声進行まで関係する。さらに、自動採譜では、調号や各音をシャープ（＃）かフラット（♭）のどちらで表すかなどのピッチスペリング問題なども調性や和声の推定と統合して解決する必要がある。さらには表情記号やその他の記号も推定しなければ、元の楽譜を復元することはできない。このように、自動採譜は総合的な奥の深い



(a) 正解の楽譜(「赤とんぼ」, 3/4 拍子)



(b) 拍子の誤認識例(上げ拍の 4/4 拍子として)

図-11 拍子推定における誤認識例。リズムパターンの観点からは妥当な解。

問題である。

リズムによる音楽音響入力の小節分割

■ One-pass DP 法を用いた小節分割

多くの音楽には拍子があり、人間は拍を感じながら音楽を聴く。そして小節を単位とした拍のサイクルを追うことができる。これを機械にさせられれば、たとえば自動ジャンル分類において、小節単位のミクロな音楽の特徴などを捉えて、解析精度が向上できる。それを行うにはどうすればよいか。

従来は、これは拍を追跡する問題であると考えられ、ビートトラッキングと呼ばれ、多数の研究者が取り組み、マルチエージェントや引き込みによる同期などが用いられた。しかし、局所的なボトムアップ処理は速度変動などに弱い。これらとは対照的に、角尾ら (2008)⁶⁾ は全体を小節へ最適単位分割する方法を取った。

ポピュラー音楽等の打楽器を含む音楽には、ほとんどの場合、小節を単位とする打楽器パターンが複数種類、繰り返し使われる。打楽器成分だけに注目すると、もし、それらの打楽器パターンが標準パターンとして既知ならば、連続音声認識と同型の問題になる。たとえば、8桁連続数字の音声認識では、十数字を標準パターンとし、あらゆる数字の標準パターンの組合せの接続を入力音声と非線形時間伸縮して整合する。それを効率的に行うには One-Pass DP アルゴリズム (Ney 1984) が利用できる。音楽の場合は、たとえば 4 種類のリズムパターンの連鎖で 50 小節の音楽ならば、1, 2, 3, 4 の数字を 50 桁発声した場合に相当する。曲全体を複数種類の小節単位の標準パターンの接続として認識し、時間伸縮して整合もとることができる。

■ 反復学習による打楽器パターン抽出法と小節分割

しかし実際は、標準パターンは未知である。入力 of 音楽から複数の標準パターンを取り出したいが、それには小節境界(小節線位置)が必要で、それも不明であり、一種の鶏と卵の問題になっている。そこで、何らかのランダムな複数の種パターンから出発し、種パターンのあらゆる接続との非線形時間整合を One-Pass DP アルゴリズムで効率よく行うことで、仮の小節分割と小節ラベルづけ(つまり小節パターン認識)を行う。この仮の結果から、それぞれの種パターンに対応するラベルづけされた小節を集めて平均して種パターンを改善する。これを用いて、小節分割と小節パターン認識を行う。これを繰り返せば、誤差を最小化するような小節分割、小節ラベルづけ、複数の平均パターンが同時に最適化される。数学的には、種パターンによっては最適化は保証されないが、誤差極小の解に収束することは保証される。

実際には、音楽信号には打楽器以外に旋律や和音が多く含まれるから、角尾らは、HPSS を用いた打楽器成分強調の前処理を行い、上記の手順により小節分割を行った。図-12 に小節分割とリズム分類の結果を示す。横軸は入力楽曲の時間を表し、縦軸はパターン学習された標準パターンの番号とその中の時間である。図-13 はスペクトログラム領域で反復学習された 4 つのリズムパターンである。

以上は DP (Dynamic Programming, 動的計画法) を基本にして説明したが、各小節のリズムパターンの観測を標準パターンからの確率的変動と見なせば、確率最大の Viterbi 分割を求める HMM の問題として理解してよい。

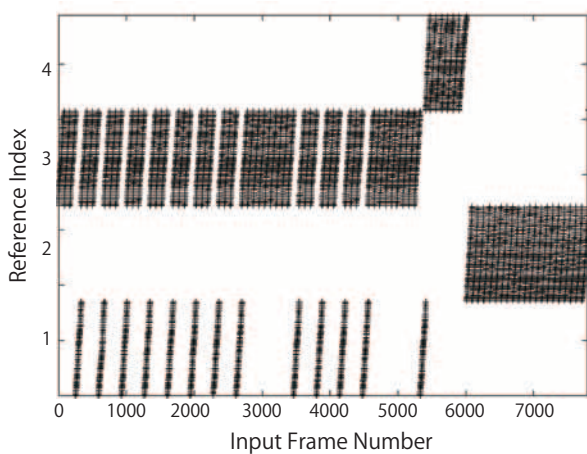


図-12 小節分割とリズム分類の結果（横軸：時間，縦軸：パターン番号）

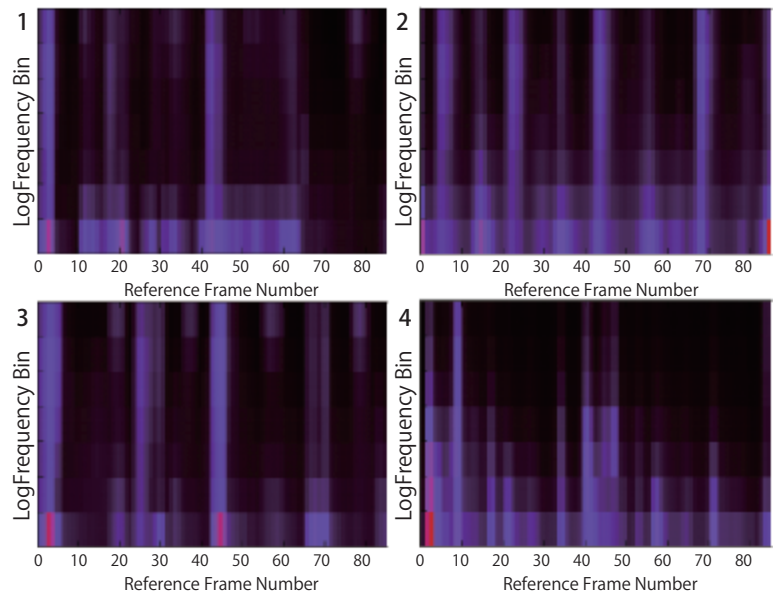


図-13 学習された4種類のスペクトルパターン

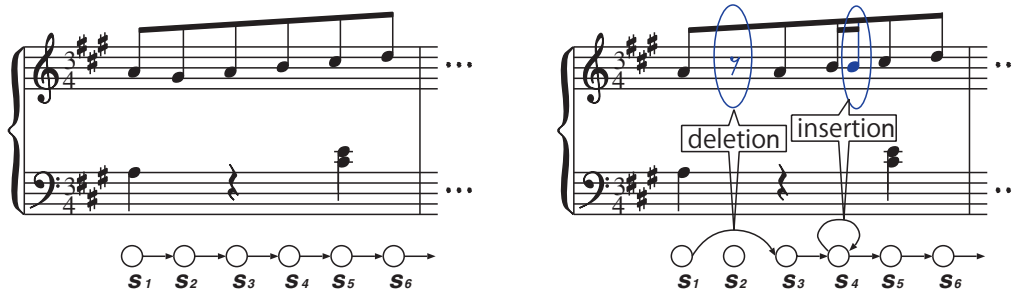


図-14 弾き誤りを含む演奏の例

楽譜追跡と自動伴奏

■ HMM による楽譜追跡

演奏を観測して現在楽譜のどこを演奏しているかを推定し追跡することを楽譜追跡 (score following) という。譜めくりや字幕同期のみならず自動伴奏(伴奏同期)にも有用である。この技術は、ピアノの片手ずつの練習、自動連弾、コンチェルトの練習、仮想的アンサンブル演奏など広い応用が見込まれる。さらに演奏者が拍打ちのみを行うならば、仮想指揮システムともなり得る。

もしテンポが既知で一定ならば楽譜追跡はまったく問題にならない。もし弾き誤りがなく、演奏が MIDI 信号として観測される場合も問題はやさしい。しかし実際は、演奏のテンポは変動するし、図-14のように、シャープやフラットを忘れるなどの弾き誤りのために別の音が弾かれたり、音が抜けたり、ミスタッチで余分な音が加わることが多い。

局所的な誤りを含む可能性のある演奏の楽譜追跡としては、楽譜と実演奏を音高について DP マッチングを行

う手法 (Dannenberg 1984) が知られている。また、拍の情報に注目した自動伴奏 (Raphael 2001) の研究も行われている。武田ら (2004) は弾き直しやスキップのように演奏中に演奏個所が大きく移動する演奏にも対応した⁵⁾。

楽譜にある音符を隠れ状態、実際に弾かれた音符の MIDI 信号を隠れ状態からの出力と考えれば、図-14のように、局所的な音符の脱落は状態のスキップ、余分な音符の挿入は同一の拍位置への遷移する過程として扱うことができる。また、さらに、一度弾いた部分の弾き直しは以前の状態への遷移、途中の部分スキップは以後の状態への遷移として扱える。和音の挿入・脱落誤り、弾き直し・スキップはすべて状態遷移として記述できる。和音の場合、和音の構成音は、現実の演奏では必ずしも同時に発音されるとは限らず、和音を構成する各音は拍位置の付近に、順不同で発音される可能性がある。あらゆる順序を考慮して拍位置に対応づけるには組合せが多いが、武田らはそれを拍位置の状態の内部に設けた状態の自己遷移として和音構成音が発音するようにモデル化し、単純化した。すなわち、1つの状態から順不同で和

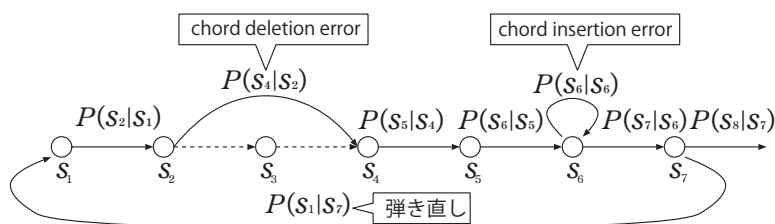


図-15 弾き誤りを含む演奏の確率的なモデル（各 S は楽譜中のどの音符を弾いているかを表す状態を意味する）

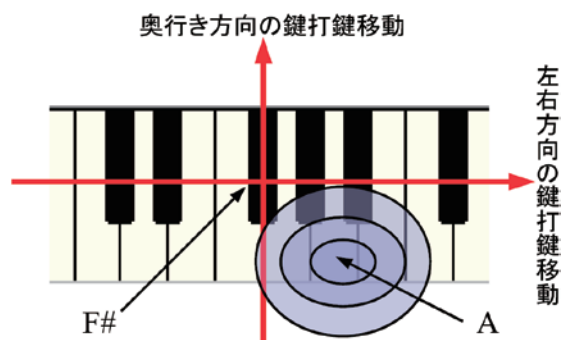


図-16 状態遷移に伴って鍵を押す位置の確率分布の概念

音の各音およびミスタッチの音が飛び出してくる確率的なプロセスとしてモデル化した。

演奏者のスキルを反映した遷移確率や出力確率分布をモデルに与えれば、Viterbi アルゴリズムにより、演奏の MIDI 信号から今までの遷移経路と現在の隠れ状態を最尤推定することができる。

■ ジャンプを許した自動伴奏

楽譜追跡ができれば、ピアノ協奏曲の自動オーケストラ伴奏、室内楽の自動合奏、自動連弾、片手を弾いてもう一方を自動演奏するなどが可能になる。実際、誤りなく弾けばそのような機能を持つ家庭用のキーボードも市販されている。しかし、弾き誤ったらそこから伴奏がついてこないシステムでは、少なくとも本番では使えないし、練習中でも楽しみ方は限られる。また、練習では間違えずに弾けるようになるまで同じ個所を何度も繰り返すことも多い。伴奏もそれについてきて欲しい。

そこで、前述の手法により、図-15 のようなモデルを用いてジャンプも許した楽譜追跡を行い、推定された演奏個所に対応する伴奏パートを MIDI などにより演奏する。楽譜追跡の結果から、演奏者が想定している現在のテンポを推定して、あたかも意志を持った伴奏者のように、演奏者が出遅れても先に進むなどの実際的な工夫も必要である。武田らは過去に弾かれた数個の音符からテンポを推定し、それを演奏者が弾く音符のタイミングと折衷して伴奏のタイミングとした。

自動運指決定

■ 自動運指決定の効用

最後に、楽器演奏の運指決定も同様の枠組みで考えられることを述べよう。運指の自動決定は、ピアノ自習者への模範演奏の提示、曲の難易度判定や難易度をクエリーとした練習曲検索など、効率的な演奏上達を促す支援システムへの応用のほか、手の大きな／小さな演奏者、

指を欠損／故障した演奏者など、さまざまなケースへ対応可能な運指決定などが考えられる。運指をピアノ演奏ロボットの動作計画と考えれば、ピアノ演奏に限定せず一般にロボット指／脚を使う作業全般への展開も期待できる。また、自動編曲において演奏可能かどうか、あるいは難易度かどうか、などを判定する上で重要である。たとえば既存曲のギター用編曲などでは、運指可能な解を見出すことが必要である。

従来のピアノ運指決定の研究の多くは片手の単旋律を対象とし、 n -gram 統計を用いた手法（野口ら 1996）、隣接する音符間の運指コスト関数を用いる手法（Kasimi ら 2005）、rule-based な手法（林田ら 2003）などが見られるが、米林ら（2007）⁷⁾ はピアノの運指決定問題に、HMM を用いた独特のアプローチを取った。従来は楽譜が入力として与えられたらそれを処理して運指を決定して出力する知能的なシステム、という考え方が多かったが、ここでは逆に運指から楽譜どおりの音楽が生み出されるには、という考え方をする。

■ HMM による自動運指決定

簡単に言えば、指をランダムに動かすとして、そうやって弾かれた結果がたまたまターゲットとしている曲に一致するならば、そのときの指の動きはどうであったと考えるのが最も確率が高いか、という確率的逆問題として最適運指決定を定式化した。言い換えると、（間違えずに弾かれた）理想的なピアノ演奏を「結果」とし、それをもたらした「原因」である運指を最大事後確率推定（MAP 推定）するような確率論的逆問題として運指決定を捉える。

ピアノ演奏とは、適切な手指状態の遷移系列から、現在の鍵から次の鍵へと打鍵位置の物理空間内の移動（以後、打鍵移動と呼ぶ）を（楽譜に合致するように）生み出すことであると考えられる。そのときに音楽演奏が楽譜どおりでありそのような遷移が未知である場合に、最も適切な解を得るのが自動運指決定である。図-16 は現在



図-17 自動運指決定結果例(J.S.Bach「二声のインヴェンション」より)。ほぼ妥当な結果が得られている。

の指の形に対応する隠れ状態から次の指の形に対応する隠れ状態へ遷移するとき、鍵盤上でどれだけの相対移動をするかという確率分布モデルの例である。親指などは指くぐりがあり得るため単峰性の分布でないで、混合正規分布でモデル化する。その他、親指や小指は他の指に比べて短いので黒鍵は打鍵しにくいことも打鍵移動の確率分布に含ませることができる。図-17に結果のサンプルを示す。

今後は、単旋律でなく和音も考慮して、両手で 2^{10} 通りの指の状態の間の遷移としてピアノ曲の運指を決定する必要がある。また、音符長のほか休符長、音の強さ、音色など、楽譜で明示的／暗示的に示される広義の演奏情報をHMM出力空間に含めるべきである。さらに、指だけで状態を考えるのではなく、掌の広げ程度、角度などを状態に含め、より精密にする必要がある。また、ここではピアノの運指を説明したが、今後はギターなどの運指およびそれを満たした自動編曲などがより重要になるだろう。

まとめ：確率的アプローチの効用

ここで多少とも扱った話題は、和声解析、和音推定、調性推定、転調検出、旋律への和声づけ、リズム認識（音価列推定）、拍子推定、拍節推定、テンポ推定、小節分割、楽譜追跡、自動伴奏、自動運指決定などであった。これらさまざまな音楽情報処理が、ほぼ同型の確率モデルにより扱えることを、筆者らの研究に基づいて述べた。確率モデルによるアプローチは、従来よく行われたボトムアップの処理、あるいはルールベースの処理に比べて、モデルの学習が可能であり、今後の高い可能性を持つ。

かつて音声認識分野では、20年前にその方向へ転換

してこれら確率的なアプローチを発展させた。他の分野もそうなる可能性がある。冗談に聞こえるかもしれないが、音声認識技術の最適の適用先は音声認識以外にあるのかもしれない。

謝辞 最後に、本稿で述べた研究を推進した下平博、小野順貴、西本卓也、川上隆、齋藤直樹、大槻知史、武田晴登、宮本賢一、内山裕貴、角尾衣未留、米林裕一郎（敬称略）、その他の関係者に感謝する。

参考文献

- 川上隆他：隠れマルコフモデルを用いた旋律への和声付け，平成11年電気関係学会北陸支部大会講演論文集，F-61，p.361（1999）。
- 内山裕貴，宮本賢一，西本卓也，小野順貴，嵯峨山茂樹：調波音・打楽器音分離手法を用いた音楽音響信号からの自動和音認識，情報処理学会研究報告，2008-MUS-76，23，pp.137-142（Aug. 2008）。
- 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，情報処理学会研究報告，99-MUS-33，pp.27-32（Dec. 1999）。
- 大槻知史，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，Vol.43，No.2，pp.245-255（Feb. 2002）。
- 武田晴登，篠田浩一，嵯峨山茂樹：確率モデルによる多声音楽演奏のMIDI信号のリズム認識，情報処理学会論文誌，Vol.45，No.3，pp.670-679（Mar. 2004）。
- 角尾衣未留，小野順貴，嵯峨山茂樹：リズムマップ：音楽音響信号からの単位リズムパターンの抽出と楽曲構造の解析，情報処理学会研究報告，2008-MUS-76，25，pp.149-154（Aug. 2008）。
- Yonebayashi, Y., Kameoka, H. and Sagayama, S.: Automatic Decision of Piano Fingering Based on Hidden Markov Models, Proc. of International Joint Conference on Artificial Intelligence (IJCAI), pp.2915-2921 (Jan. 2007).

（平成21年7月6日受付）

嵯峨山 茂樹（正会員） sagayama@hil.t.u-tokyo.ac.jp

1974年東京大学大学院工学系研究科修士課程修了，NTT，ATR自動翻訳電話研究所，北陸先端科学技術大学院大学などを経て，現在，東京大学大学院情報理工学系研究科教授，博士（工学）。