

# 個性を模倣した 仮想演奏者の 実現

## 7



浜中 雅俊

m.hamanaka@aist.go.jp  
科学技術振興機構 さきがけ研究員

音楽情報処理の研究分野の最終的な目標の1つは、実在する音楽家のように動作する仮想演奏者を計算機上で実現することである。そのような仮想演奏者を可能とするためには、作曲・編曲能力や、即興演奏中にテンポやコード進行などを予測する能力など、さまざまな音楽的能力を計算機上で実現していく必要がある。このように、工学的手法を用いて人間のさまざまな能力を計算機上で実現しようとする試みは、その成果の実応用が期待されるばかりでなく、人間の仕組みを解明していく上でも重要である。

本稿では、筆者が研究遂行中の、計算機上に仮想演奏者を実現する試みについて紹介する。

### 人工個性とは

人間の能力の計算機上での実現は、まず、人工知能の研究分野で試みられた。そこでの目的は、あたかも人間のように認識や判断などをする計算機システムの実現である。実際に実用化されたものには、パターン認識による人間の知覚の代行や、エキスパートシステムによる人

間の判断の代行等があるが、そのようなシステムでは客観性が求められるため、感性や個性などの主観的な要因は排除されてきた<sup>1)</sup>。

一方、感性工学の分野では、これまで、感覚器でのセンシングから心理的認知的現象としてのイメージまでの過程の工学的モデル化が進められており、その過程における個別性について検討されてきた。そこでの主な目的は、人間と計算機とのよりよいインタフェースを実現するために、計算機上に人工感性を構築することである。実用化されたものとしては、類似検索や形容詞(印象語)検索による商品検索システム等がある<sup>2)</sup>。しかし、そこで扱われた個別性とは、商品の物理的特徴と検索空間とのマッピングの個人差のみであった。

あたかも実在する人間のような個性を持った計算機システムを構築するためには、人工知能や人工感性を実現するだけでなく、各人の固有の癖をなんらかの形で獲得し、それをを用いて人工的に個性を再現する必要がある。表-1に人工知能と人工感性、人工個性のそれぞれの対象、目的、用途について筆者の考えをまとめる。

領域	対象	目的	実用化用途
人工知能 (artificial intelligence)	知能 (知識)	知覚, 判断, 思考, 学習, 記憶などの過程の 計算機上での実現	パターン認識 エキスパートシステム
人工感性 (artificial Kansei)	感性	人間と計算機とのよりよいインタフェースの実 現	印象語による商品検索 デザイン支援システム
人工個性 (artificial personality)	個性	各人の固有の癖の獲得とその計算機上での再 現	仮想音楽家 仮想デザイナー

表-1 人工知能と人工感性, 人工個性

## 音楽分野における人工個性

人工個性の技術の最終的な目標は、実在する音楽家の個性を模倣した仮想音楽家を実現したり、工業デザイナーの個性を模倣した仮想デザイナーを実現することである。もしそれが実現すれば、獲得した個性をネットワーク上で配信しユーザがそれを利用することが可能となる。また、専門家が作品や製品を「創造」し、ユーザがそれを「鑑賞・使用」という従来の枠組みに対して、ユーザが制作に「参加」という新たな枠組みの提供が可能となる。たとえば、著名な作曲家の個性を持った仮想作曲家を生成し、自分の好みの曲を作曲させることや、ある演奏家の個性を持った仮想演奏者を生成し、それと共演することも可能となる。また、音楽の知識や技能を持たない一般のユーザは、自分の好みの仮想演奏者を集めて演奏させることで、あたかもバンドオーナーのような立場で気軽に演奏を楽しむことができるようになる。このような研究は音楽初心者にも音楽を楽しませるアミューズメント性を持つだけでなく、プロの演奏家にとっても、音楽制作の生産性を上げる技術の1つとして期待がもたれる。具体的には、ギターしか弾けない演奏者は、ギター以外のパートを、さまざまな個性を持った仮想演奏者に演奏させることにより、バラエティに富んだ曲を短時間で制作できるようになる。

筆者は、上記のようなさまざまな効果が期待されている音楽分野における人工個性に着目し、音楽分野における人工個性実現の第一歩として、ジャムセッションにおける演奏者の個性を扱ってきた<sup>3)</sup>。ジャムセッションとは、コード進行など最小限の約束ごとのみを決め、順次ソロパートを自由に交代しながら進めていく即興演奏であり、その特徴は、演奏者に固定した主従関係が存在せず、すべての演奏者が対等な立場でインタラクションできる点である。したがって、ある演奏者の演奏がそこで演奏している演奏者全員に影響を与え、他の演奏者の新たな演奏として返ってくるというインタラクションが頻繁に起こる。そして、そこでのインタラクションの仕方は、演奏者の個性によって大きく異なる。このようなジャムセッションにおける個性のモデル化をまず実現していくことで、今後、作曲や編曲をはじめとするさまざまな人間の音楽的活動を総合的に行うことのできる仮想演奏者の実現を目指している。

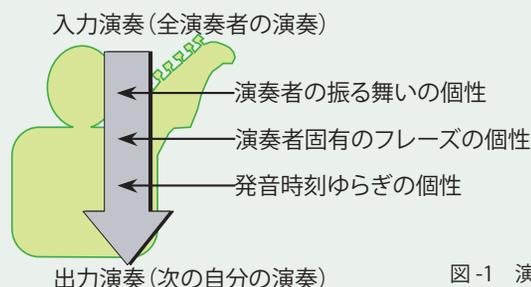


図-1 演奏者の3種類の個性

## Guitarist Simulator : 個性を模倣するジャムセッションシステム

演奏者の個性にはさまざまな側面があるが、たとえばジャムセッションでは、相手の演奏に対してどのような演奏で反応するかによって演奏者の個性が表れる。人間の演奏者は自分がソロを弾いているときに、相手が同時にソロを弾き出したら、次の小節からソロをあきらめたり、逆にもっと激しいソロを弾いたりする。そのような個性をここでは、演奏者の振る舞いと呼ぶ。演奏者の振る舞いを実在する人間の演奏者から獲得することができれば、仮想演奏者はあたかもその演奏者のように、相手の演奏を聴きそれに対してソロや伴奏でインタラクションすることが可能となる。また、ジャムセッションではグルーブ感と呼ばれるような演奏の心地よいゆらぎや、フレーズや音自体にも個性が表れる。演奏者が実際にどのようなフレーズを弾くかは、演奏者固有のフレーズによって決まる。すべての演奏者は、手癖と呼ばれるような自分の得意とするフレーズをいくつも持っている。そのようなフレーズを獲得することができれば、仮想演奏者は模倣したい演奏者のフレーズで演奏することができる。実際に演奏を出力する際には、演奏の発音時刻にゆらぎが生じる。このようなゆらぎにも演奏者の個性が反映されている。発音時刻ゆらぎの個性を獲得することができれば、あたかも模倣したい演奏者のように発音時刻にゆらぎのある演奏を生成することができる。

以上述べてきた、演奏者の3つの個性は図-1のような関係にあると考えられる。Guitarist Simulatorは、調はA、テンポは120で固定、コード進行は12小節単位で繰り返す循環コードという条件でギタリスト3人がジャムセッションした5分間のMIDI演奏記録から、これら3種類の個性の学習を可能とした。MIDIとはMusical Instrument Digital Interfaceの略で、電子楽器の演奏情報を送受信するための共通の規格である。



の切れ目を見つけて切り出しを行うかが問題となる。フレーズの分割手法として我々はこれまで、2種類の手法を提案してきた。

1つ目の手法はボロノイ線図を用いたものである(図-5)。これは、ピアノロールと呼ばれる縦軸をピッチ、横軸を時間とした楽譜上にボロノイ線図を描き、面積の小さい領域を近くの領域に再帰的に結合させていくことにより階層的なフレーズ構造を獲得するものである。この手法はどんな曲でも分割が可能な反面、音楽的根拠に乏しく曲によっては適当でない位置で分割される場合があった。

2つ目の手法は、音楽理論GTTM (Generative Theory of Tonal Music)<sup>4)</sup>に基づく手法である。GTTMは、20年以上前に考えられた音楽の構造を分析するための理論であるが、これまで計算機上への実装ができていなかった。我々は、GTTMに基づく音楽分析システムATTA (Automatic Time-span Tree Analyzer) を計算機上へ実装し、GTTMに基づくフレーズ分割を可能とした(図-6)<sup>5)</sup>。ATTAは現在<http://staff.aist.go.jp/m.hamanaka/atta/>にて、インターネット上に公開している。ただし、現在扱えるのは単音旋律のみで、和音を含むギター演奏のフレーズ分割を実現するためには、今後の改良が必要である。

### ❖ 発音時刻ゆらぎの模倣

演奏者が同じ演奏を繰り返し弾いた場合でも、MIDIのレベルでまったく同じ演奏となることは稀であり、演奏の表情づけや演奏動作の微妙な差などにより発音時刻がゆらぐ。我々は、そのような発音時刻ゆらぎを隠れマルコフモデルを用いてモデル化し、個性による発音時刻ゆらぎの違いを隠れマルコフモデルのモデルパラメータの形で獲得することを可能とした(図-7)。

隠れマルコフモデルのモデルパラメータを教師なしで学習する手法としては、Balm-Welchアルゴリズムという手法があるが、5分間のMIDI演奏記録では学習データのサンプル数が不足してうまくパラメータを推定できない。そこで、Balm-Welchアルゴリズムによるパラメータの推定と、ヘルドアウト補完法によるパラメータの補完を交互に繰り返すことによって、限られた長さの演奏データからモデルパラメータの推定を可能とした<sup>6)</sup>。

### ❖ 音楽版チューリングテストの実施

以上述べてきたようなGuitarist Simulatorを用いて5人のギタリストに、人間1人と仮想演奏者2人の組合せでジャムセッションを実施したところ、仮想演奏者は1人

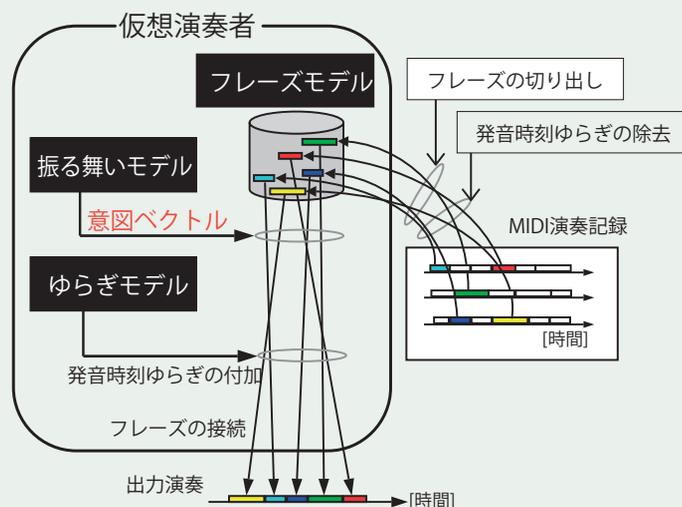


図-4 フレーズデータベースによる出力演奏の生成

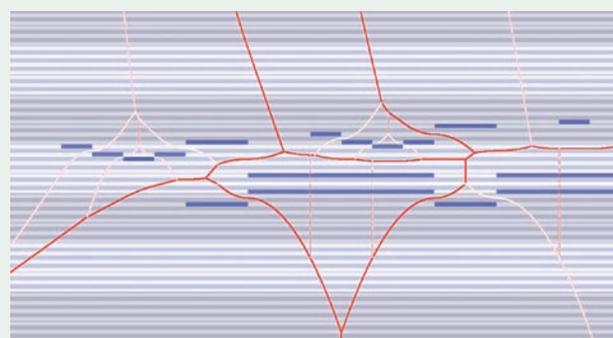


図-5 ボロノイ線図に基づくフレーズ分割

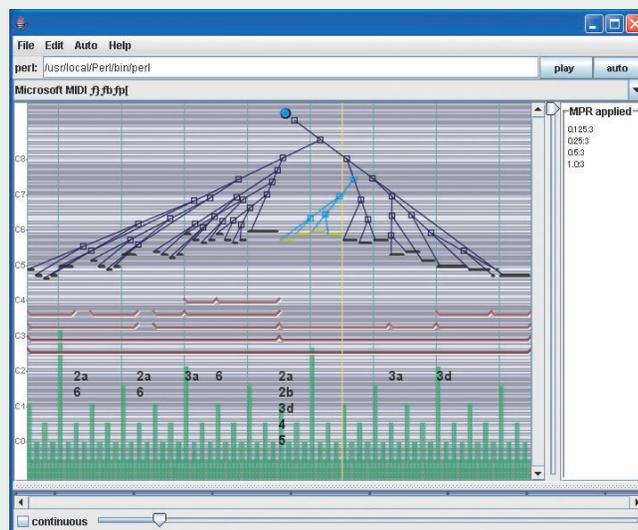


図-6 GTTMに基づく音楽構造分析

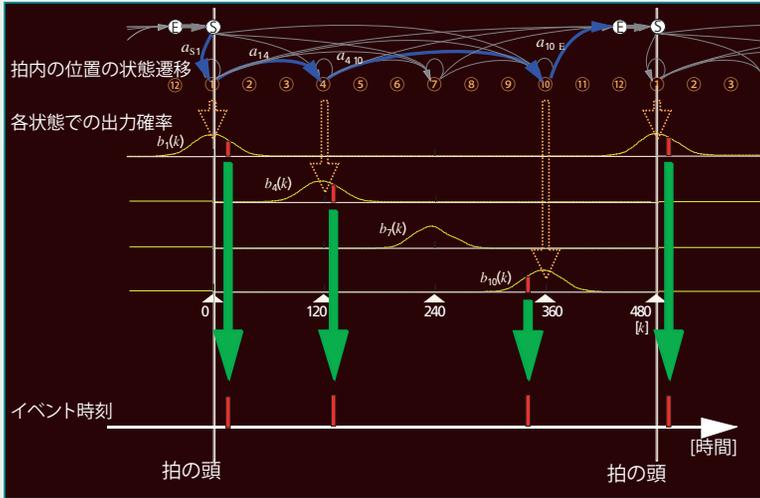


図-7 隠れマルコフモデルによる発音時刻ゆらぎのモデル化

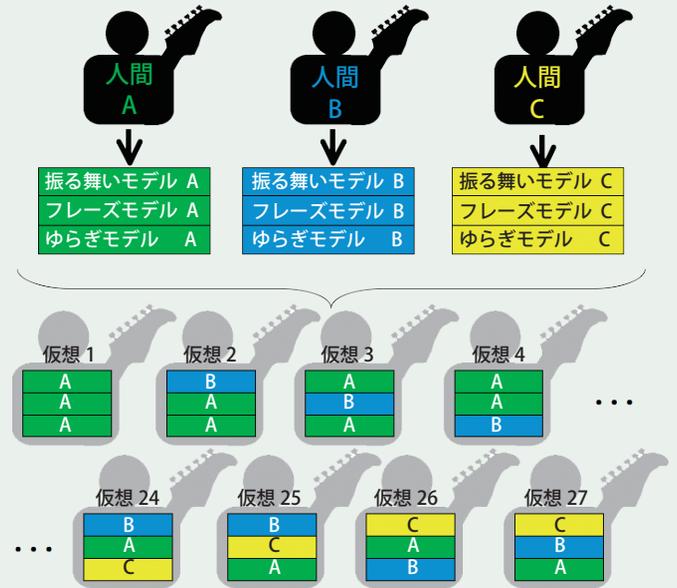


図-8 27種類の仮想演奏者の生成

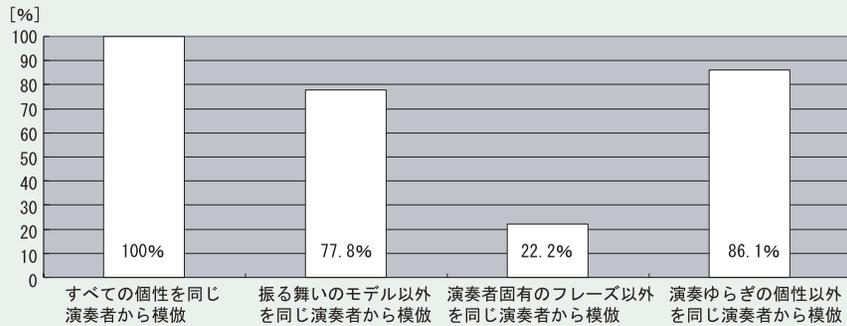


図-9 評価結果と模倣した演奏者が一致した割合

1人个性的であるとの意見が多かった。特に、自分自身を模倣した仮想演奏者とジャムセッションしたギタリストは、仮想演奏者が自分の真似をしているようで演奏しにくいと感じた者が多かった。

音楽システムの評価の方法については従来からさまざまな議論がなされているが、絶対的な評価の枠組みはまだ確立されていない。ここでは、チューリングテストからヒントを得た心理実験を行った。チューリングテストでは、相手が人間であるか機械であるかを判断するが、今回は仮想演奏者がどの演奏者の個性を模倣しているかを判断する。

実験ではまず、同じバンドでの演奏経験がありお互いに相手の演奏の個性をよく知っている3人の演奏者A、B、CのMIDI演奏記録からそれぞれの振る舞いモデル、フレーズモデル、ゆらぎモデルを獲得する。次に獲得したモデルを組み合わせることで27種類の仮想演奏者を生成し

た。そして、仮想演奏者がA、B、Cのうち誰の個性を模倣しているか、評価させた(図-8)。

図-9はその結果をまとめたもので、3種類の個性モデルを同じ演奏者から模倣した場合には、100%の割合で模倣元の演奏者を正しく評定できていた。一方、どれか1つでも他の演奏者の個性を用いている場合には正しく評定されている割合が減っていた。このことは、獲得した3つの個性モデルがいずれも必要であったことを表している。

### 個性の違いを聴き分けるためのインタフェース Sound Scope Headphone

Guitarist Simulatorは、ギタリスト向けのシステムであるが、現在、演奏初心者でも楽しめるように拡張を行っている。そこで問題となるのは、初心者が複数のパート

の演奏を聴き分ける能力である。たとえ、複数の仮想演奏者がそれぞれ人間の演奏者の音楽的な個性を獲得していたとしても、ユーザがそれを聴き分けることができなければ、個性の違いを感じることは困難である。

複数パートを聴き分けるためには、パートごとに複数トラックに録音された音源を用意して、音楽用ミキサーで各パートのボリュームや定位（左右の音量比）を調節すればよい。しかし、市販の音楽用ミキサーは操作が煩雑で初心者が直感的に扱うことは困難であった。そこで、音楽初心者でも複数パートの聴き分けが容易となるよう、直感的な操作で各パートのミキシングが変更できるヘッドフォンSound Scope Headphoneを制作した。

Sound Scope Headphoneの特長は、頭を上下左右に振ったり、手を耳に近づけて耳を澄ませるようなポーズをするなど、人間が音を聴くときに自然に行う動作を、ヘッドフォンに搭載した電子コンパス、傾斜センサ、距離センサの3種類のセンサで検出することで音楽ミキサーのコントロールを可能にした点である(図-10)。

#### ❖ 電子コンパスによる顔の方向の検出

通常ヘッドフォンでCDなどの曲を聴く場合、右側から聴こえている音は、右を向いてもやはり右側から聴こえることになる。一方、Sound Scope Headphoneでは、ヘッドフォンのアーク部に取り付けてある電子コンパスを用いてユーザが向いている方向を検出し、右を向けば右から聴こえていた音が正面から聴こえるようにしている。同様に、左を向けば左から聴こえていた音が正面から聴こえるようにしている。ユーザは首を左右に動かすという自然な動作により、聴きたいパートを探し、そのパートが正面から聴こえるようにすることができる。

#### ❖ 傾斜センサによる頭の上下動の検出

首を左右に振り、聴きたいパートを正面から聴こえるように操作しても、まだ正面に複数のパートがあり、目的とするパートが聴きとりにくい場合には、今度は頭を上下に動かすことにより、ミキシングを変更することができる。ヘッドフォンのアーク部に取り付けた傾斜センサで頭の上下動を検出し、上を向けばより遠くに配置したパートの音量を大きくし、下を向けば近くに配置したパートの音量を大きくしている。このとき、各パートの配置は図-11のようなGUI (Graphical User Interface) を用いて手動で変更することができる。中心にある円が聴者のアバタの位置と頭部の方向を表し、その周囲にある丸数字が各パートの配置を表している。



図-10 Sound Scope Headphone



図-11 パートを配置する GUI

#### ❖ 距離センサによる耳に手をかざす動作の検出

耳に手をかざして正面から聴こえる音や声を注聴するような動作を、ヘッドフォンのスピーカ部に取り付けた距離センサで検出する。そして、距離センサと手の距離によって、聴こえるパートの範囲を設定している。耳に手を近づけると、真正面から聴こえるパートの音のみが聴こえるようになり、離すと真後ろ以外すべての音が聴こえるようになる。また、その中間の場合には、前半分から聴こえる音のみが聴こえるようになる。

図-12は、図-11のように各パートを配置し正面を向いているときに手を耳から遠ざけた時と近づけたときのミキサー画面の変化を示したものである。縦に並んだスクロールバーがミキシング時の増幅率を表し、スクロールバーのつまみが高い位置にあるほど、増幅率が高いことを表す。手を耳から遠ざけた時(a)には、正面にある6番のパートの増幅率が一番高く、後ろ方向のパート



図-12 距離センサの操作による増幅率の変化



図-13 ライトによる視覚的効果を付加した Sound Scope Headphone

になるにつれてなだらかに増幅率が減少する。一方、手を耳に近づけると (b), その距離に応じて増幅率が減少する。このとき正面のパートの増幅率の減少よりも、後ろ方向のパートの増幅率の減少のほうが大きくなるため、正面の音が相対的に大きな音で聴こえるようになり、あたかも正面のパートにフォーカスしたように聴くことができる。

### ❖ 音楽教育などへの利用

図-13は、パートの個数分の譜面台に楽器の絵を載せ、各パートの音量に応じてライトの明るさを変化させるようにしたものである。このようにすると、聴者が耳を澄ませるポーズをして、ある特定のパートを聴いている場合、その楽器がライトで明るく照らされ、視覚的にも鳴っている楽器が確認できる。Sound Scope Headphoneは、元々演奏の個性の違いを聴き分けるために制作したものであるが、このような視覚的な効果を付加することによって、あたかも指揮者のように各パートの音量を積極的にコントロールして楽しむエンタテインメントとして使うこともできる。また、楽器の音色に親しみのない子供たちに対しては、楽しみながら楽器と音色との対応関係を覚える音楽教育などへも利用できる。

### 今後の展望

現状のGuitarist Simulatorは、調やコード進行、テンポが固定であるなどある一定の制約下で動作するもので

あるが、今後、その制約を1つずつ外していくことにより、人間の音楽的活動を総合的に行うことのできる仮想演奏者の実現を目指していく。今後システムと演奏者モデルを収めた個性データベースを公開し多くのユーザに試用してもらい、システムを使って演奏した人のモデルをデータベースに登録できるようにして、さまざまな演奏者モデルを収集できるようにしていく予定である。そして近い将来、演奏者の個性モデルがテキストや音楽、画像、動画、音声など、さまざまなコンテンツと同じようにネットワーク上で配信され、一般のユーザがそれを利用できるようにしていきたい。

### 参考文献

- 1) 大津展之：パターン認識における特徴抽出に関する数理的研究，電子技術総合研究所研究報告，No.818，pp.177-180 (1981).
- 2) 栗田多喜夫，加藤俊一，福田郁美，板倉あゆみ：印象語による絵画データベースの検索，情報処理学会論文誌，Vol.33，No.11，pp.1373-1383 (Nov. 1992).
- 3) Hamanaka, M., Goto, M., Asoh, H. and Otsu, N.: A Learning-Based Jam Session System that Imitates a Player's Personality Model, Proceedings of the 2003 International Joint Conference on Artificial Intelligence (IJCAI2003), pp.51-58 (2003).
- 4) Lerdahl, F. and Jackendoff, R.: A Generative Theory of Tonal Music, MIT Press, Cambridge, Massachusetts (1983).
- 5) 浜中雅俊，平田圭二，東条 敏：GTMグルーピング構造分析の実装：ルールを制御するパラメータの導入，情報処理学会 音楽情報科学研究会研究報告 2004-MUS-55-1, Vol.2004, No.41, pp.1-8 (2004).
- 6) Hamanaka, M., Goto, M., Asoh, H. and Otsu, N.: A Learning-Based Quantization: Unsupervised Estimation of the Model Parameters, Proceedings of the 2003 International Computer Music Conference (ICMC2003), pp.369-372 (2003).

(平成18年2月8日受付)