音 楽 情 報 科 学 49－3
（２００３．２．２１）

複数のモデルを利用した重回帰分析による演奏表現の学習

福 井 浩 司† 堀 内 靖 雄† 市 川 熹†

本研究では伴奏システムが独奏者の演奏表現を利用した制御を可能とするため、独奏者の表情豊かな演奏を学習することにより、その独奏者の演奏表現を考慮した独奏演奏の予測を行なう手法を提案する。

人間の演奏はさまざまな演奏表現がされる。伴奏システムは演奏表現を予測することでより良い適応が可能になると考えられる。本研究ではソロ演奏に焦点を絞り、全て単音の楽譜を用いて、人間の表情豊かな演奏を収録した。収録した人間の演奏履歴に対して複数のモデルで重回帰分析をおこない、楽譜上の各楽音で人間の演奏を予測するモデルを作成した。収録された演奏から得られたモデルの種類、予測誤差などについて検証し、伴奏システムの自動学習の可能性について検討を加えた。

Expectation of Musical Expression with Multiple Regression Analysis Based on using Plural Models

Koji Fukui,† Yasuo Horiuchi† and Akira Ichikawa†

In this paper, to manage with musical expression in the accompaniment system, we will introduce an expecting method of expressive human performance.

Humans perform several musical expressions. Therefore, better adaptation will be accomplished with such expected expression in the accompaniment system. Expressive human performances, where a performer played a monophonic piece, were recorded. These recorded performances were analyzed with regression analysis using plural models in order to generate performance expecting models at each note. We will discuss the possibility of auto learning of human performance by examining the type of model applied and the precision of estimation.

1. はじめに

合奏において人間は自由な演奏表現をおこなう。その場合の演奏は楽譜に忠実に一定のテンポ・音量で演奏されるわけではなく、楽譜に記述されていないテンポの揺らし、アクセントなどを含む演奏である。図 1 のような例では楽譜には演奏記号が全く記述されていないためテンポ一定で演奏されるように思えるが、実際に演奏されたテンポを見てみるとフレーズの初めで加速し終わりで減速するという形になっており、テンポは一定ではなくその速度には実に 1.5 倍程度の差がでている。

演奏表現は一見楽曲の構造から予測することもできそうだが、楽曲の解釈は演奏者ごとに異なるため演奏表現も変わってくる。この演奏表現によって曲に深みがでたり個性が現れたりするのである。これらの演奏表現を伴奏システムが獲得していなければ適切な追従ができず、演奏表現部分で演奏が破綻したり、そこまではいかなくても独奏者の演奏表現の幅を狭めてしまう。それゆえこの演奏表現を獲得して合奏時に利用することが重要である。

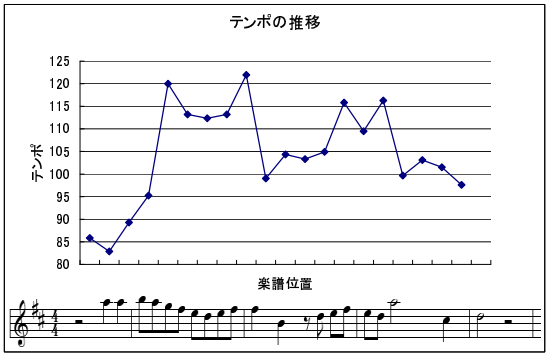

図 1 テンポの推移

演奏表現を獲得する手段としては独奏者があらかじめ伴奏システムに対して演奏表現を入力しておくことも考えられるが、全ての演奏表現を手動で入力するには大変な労力が必要であるということと、独奏者自身

† 千葉大学大学院自然科学研究科
Chiba University

が意識していない演奏表現を獲得できないといったことがあるため、演奏表現の自動学習が望まれる。

2. 予測方法

分析には楽譜上の楽音ごとに多数の予測モデルを作成して一番良い結果を出したモデルを採用し、次回の演奏で使用する、という方法をとった (図 2)。

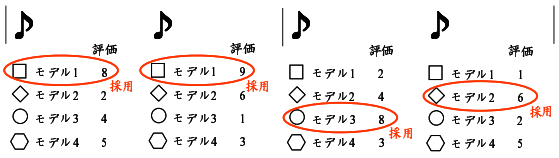

図 2 分析方法のイメージ

このように楽譜位置ごとに予測モデルを変更することで独奏者が各楽譜位置で異なることを基準として演奏生成していたとしてもそれに対応した予測モデルが採用され、学習できると考えられる。そしてより正確な演奏表現の獲得が期待できる。しかしもちろん演奏表現に対応したモデルが用意されていなければ学習することはできない。

2.1 独奏予測モデルの形と種類

独奏予測モデルには重回帰分析を利用した。重回帰分析によって作成される予測式は多項式であり計算時間はほとんどかからない。また、たとえ学習用のデータが 1 つしかないような少数のデータでも学習が可能である。但し学習結果が予測に役立つとは限らないので注意が必要である。

本研究では演奏者が発音時刻を決定するときに基準とすることは予測したい楽音の近辺の「楽音の発音時刻」と「拍あたりの時間長」であると考えた。「拍あたりの時間長」はテンポの逆数に相当するものであり、1 拍分の楽譜位置を進むために何秒かかるかという意味である。そして発音時刻の予測は

$$Y = T + \alpha s L$$

の形で予測をする (図 3)。T が「楽音の発音時刻」で L が「拍あたりの時間長」である。また s は予測したい楽音と T を得た楽音との楽譜上の距離を表した、楽譜から決まる定数で単位は [拍] である。T、L は演奏から得られ、s は楽譜から得られる。α は演奏履歴からの学習により得られた表情付けの度合いを表すパラメータである。

「楽音の発音時刻」と「拍あたりの時間長」についてはそれぞれ何通りかのバリエーションをもたせ、複数の予測モデルを作成した。

まず「楽音の発音時刻」は以下の 3 種類を候補とした。

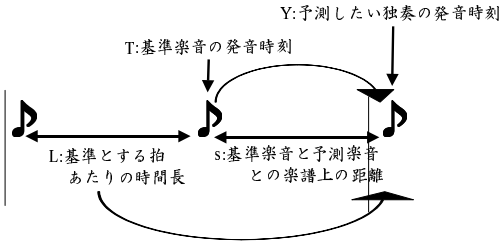

図 3 発音時刻の予測モデルの形

- 直前の楽音の発音時刻
- 直前の拍の発音時刻
- 直前の楽音の消音時刻

「直前の楽音の発音時刻」は発音時刻のうちで予測したい楽音に一番近いものであるから予測に役立つであろうと考えた。それから「直前の拍の発音時刻」も候補としたが、これは人間は全ての楽音を感じるわけではなく拍だけを感じていて、その拍に各楽音を乗せているということも考えられたからである。そして最後に「直前の楽音の消音時刻」であるが、これは予測したい楽音に一番近い時刻の情報ということで候補に加えた。消音時刻は発音時刻と比べると非常に揺らぎが大きいため普段は予測に役立たないと考えられるが、例えば長い音符を演奏した直後の音などでは発音時刻は時間的に遠いため消音時刻のほうが役立つということもあり得ると考えた。

「拍あたりの時間長」も同様に以下の 4 種類を候補とした。

- 直前の楽音の拍あたりの時間長
- 直前の拍の拍あたりの時間長
- 楽音の過去 1 小節の平均の拍あたりの時間長
- 拍部分だけの過去 1 小節の平均の拍あたりの時間長

「直前の楽音の拍あたりの時間長」は予測したい楽音に一番近い時間長であるため候補とした。過去の情報を多数使用せずに一番近いデータだけを使用するため揺らぎには弱くなると考えられるがテンポ変化部のように過去の情報が利用できない部分では良い性能を出すと考えた。「直前の拍の拍あたりの時間長」も同様にテンポ変化部を意識したものだが人間が拍を感じて演奏している場合を考えて用意した。逆に「楽音の過去 1 小節の平均の拍あたりの時間長」や「拍部分だけの過去 1 小節の平均の拍あたりの時間長」は過去 1 小節分の平均をとることで演奏の揺らぎになるべく影響されずに予測を行う目的で用意した。

これらのバリエーションを考えると「楽音の発音時

刻」が 3 種類、「拍あたりの時間長」が 4 種類の組み合わせとなり、12 種類の予測モデルとなる。これらにさらに「直前の楽音の消音時刻」と「直前の発音時間長」から予測するモデルを 1 つ加えた。発音時間長とは鍵盤を叩いた時刻から離した時刻までの時間感覚である。消音時刻を利用した場合は「直前の発音時間長」が一番近い時間長情報であり、予測に役立つ可能性もあり得るので特別に追加した。

結局合計 13 種類のモデルを発音時刻の予測モデルとして分析で使用した。

予測を行う時点でもし説明変量の式の中の楽音情報が得られない場合は、その楽音の情報を再帰的に予測し全ての情報が得られる時点にまで遡って予測を行えるようになっている。このようにすることで演奏ミスなどがあっても予測が破綻することなく正常にされる。

2.2 モデルの回帰係数の学習方法

ある楽音での予測モデルの作成方法は、まず演奏履歴を見て予測対象の楽音を探し、教師データ（目的変数）とする。そして次にその目的変数の回帰（予測）に必要な情報（説明変数）が全て履歴に存在するか調べる。しかしながら説明変数（例えば直前の発音時刻）が演奏ミスなどにより履歴に存在しない場合がある。そのような場合には履歴のその部分は学習に使用できない。履歴から利用できる全ての目的変数と説明変数を集めた段階で重回帰分析を行って回帰係数を学習する。

回帰係数を学習するときに学習に使用できるデータが少なく、モデルの自由度以下しかない場合、通常は回帰係数の学習をしないものである。これは学習データ数がモデルの自由度以下の場合、いかなる関係のないモデル式でも学習データの残差を 0 にするような回帰係数を学習することができてしまう。しかしモデル式が関係のないものであればそのとき学習された回帰係数も通常は無意味なものとなる。そのため通常は回帰係数の学習をしないのである。

しかし音楽においては 1 回の演奏をするだけでも時間がかかり、ユーザの負担が大きい。1 回でも少ない演奏回数で学習されることに意味がある。そこで本研究では学習に使用できるデータが自由度と等しいときでも回帰係数を学習した。

2.3 独奏予測モデルの評価方法

この段階で複数の予測モデルが作成されたが、これらのモデルを評価して一番良いモデルを採用する。評価にはいろいろな方法が考えられるが今回は AIC を評価の指標にした。

AIC とは赤池情報量規準のことで、重回帰モデルのあてはまりの良さを評価するためのものである。重回帰モデルでは

$$AIC = N\left(\log\left(2\pi\frac{S_E}{N}\right)+1\right)+2\,(p+1)$$

と定義される。ただし

$$S_E = \sum_{i=1}^{N}(y_i - Y_i)^2, y_i = \text{実測値}, Y_i = \text{予測値}$$

$p =$ 自由度 (説明変量の個数)

$N =$ サンプル数 (学習に利用した履歴の数)

AIC は値が小さいほど重回帰モデルのあてはまりよいということになる。式の初めの項はデータの数に対して残差が少ないほど小さな値となり、モデルによる予測の性能を評価している。式の 2 番目の項はモデルの自由度が小さいほど小さな値となり、モデルの複雑さを評価している。AIC を使うことによってモデルの自由度を考慮してモデルの当てはまりのよさを評価し、複数のモデルを比較することができる。

それぞれのモデルで AIC の値を計算し、最小の AIC の値を持つモデルを最適モデルとして採用した。

3. 収　　録

3.1 演 奏 曲 目

楽譜は A.Simonetti 作曲の「Madrigal」を利用した (図 4)。この曲は甘く流麗な旋律を持つため表情付けがされると考えた。本研究では演奏表現の獲得を目的とするため、演奏が表情付けをしやすいような曲を選曲した。

今回の楽譜では音量記号やアクセントなどの楽譜記号は演奏者の演奏表現の幅を狭める恐れがあるため全て削除した。これは本研究では万人共通のルールを抽出したいわけではなくある個人への適応をすることなので、なるべく演奏者ごとに異なる演奏を収録したかったためである。それから繰り返しの部分についても省略した。繰り返しの部分を演奏してもらうことは冗長であるので、それよりは繰り返しを省略して演奏時間を短くすることでよりたくさんの演奏回数のデータを収録できるようにした。また本来は伴奏パートなども存在するがこれも伴奏の楽曲構造から独奏の演奏表現の幅を狭める恐れがあるということで演奏者には提示せず、独奏者パートのみ記述された楽譜を使用した。

3.2 演　奏　者

演奏者はピアノの経験が 10 年以上の計 8 名である。うち 5 名は音楽大学ピアノ科卒業生である。過去にこの曲を聴いたことがある人は 4 名で、うち 3 名は演奏

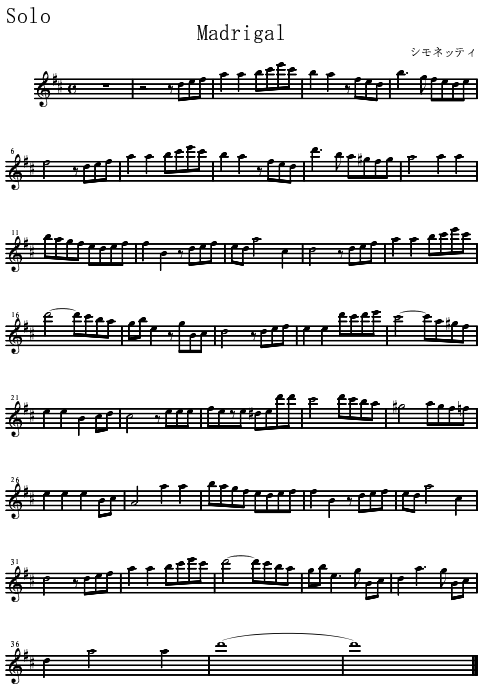

図 4 収録に使用した楽譜

をしたこともあった。

演奏者には予め実験の概要書と楽譜を渡しておき、おおまかな実験の内容を伝えるとともに若干の練習をしていただいた。ゆえに実際の収録が始まる時点では初見演奏ではなく、各自がある程度の自分なりの曲のイメージを作っていたことが期待できる。

3.3 収録したデータの内容

データは電子ピアノから出力される MIDI 信号のうち、NoteOn と NoteOff について収録した。即ち鍵盤を叩いた時刻と離された時刻、そしてその音程とベロシティの情報を記録する。時刻は $\frac{1}{1000}$ 秒、つまりミリ秒の単位で記録した。但しこれには機材の関係で最大± 2 ミリ秒程度の誤差が含まれていることを特記しておく。

3.4 収 録 環 境

収録は外部からの雑音のない静かな環境で集中して演奏できるように防音室内で行われた。演奏者は MIDI ピアノで演奏を行った。MIDI ピアノは音源およびスピーカを内蔵しているが今回は使用せず、一度外部で信号を記録しながら音を合成した。これには信号の観測と発音の時刻をなるべく合わせるようにしたいという意味や、MIDI ピアノの音源の時間精度の影響を排除したい、あるいは演奏者が聞いた音と全く同じ音を記録したデータから再現したい、といった意味がある。

MIDI ピアノから演奏された楽音の MIDI 信号は MIDI パッチャーへと送られる。MIDI パッチャーは信号を PC へ送り記録すると同時に MIDI 音源へも信号を送信し楽音信号を音響信号へと変換する。音色は違和感がないようにピアノの音を使用した。音響信号は２つのミキサーを通過して最終的にヘッドホンへ入り、ここで音に変換され、初めて演奏者の耳に入るようになっている。

3.5 収 録 手 順

収録ではまず演奏者は 10 分程度自由に練習を行った。ここで MIDI ピアノでの演奏に慣れてもらうとともに、自分なりの演奏表現を考えていただいた。それから楽譜への書き込みもこの練習時間に行った。ただしこの楽譜への書き込みとは意図した演奏表現を忘れないようにメモする程度のものであり、意識した演奏表現を全て記述するような類のものとはしなかった。

その後の収録では時々休憩を入れながら合計 12 回の演奏を行った。演奏回数はもっと多くの回数を確保することも可能であったが実際の伴奏システムで学習をする場合を考えたときに何十回もの演奏履歴が必要では実用にならない。そのため最大でも 12 回程度の履歴から学習がされなければ意味がないと判断し、12 回分の演奏しか収録を行わなかった。演奏では各個人の解釈に従って表情豊かに演奏するようにお願いした。そしてその演奏表現はできるだけ途中で変更せず、なるべく 12 回とも同じになるように心がけてもらった。実際の演奏では演奏表現が途中から変更されてゆくことも少なくはないが、本研究はまだ学習の初期のステップであり演奏表現の変化に対応できる段階ではないためそのような状況は排除した。また、演奏を間違えた場合でも気にせずにそのまま最後まで止まらずに演奏してもらった。演奏ミスと考えられる部分も含めて収録データとした。

4. 結果と考察

4.1 履歴の個数と標準誤差の関係

重回帰分析による予測で予測の誤差が十分に減少していくのかを見ていく。図 5 が演奏者 8 名の本手法で作成した予測モデルの標準誤差の変化である。横軸は学習に使用した履歴の個数、縦軸は次回の演奏を予測したときの標準誤差で単位はミリ秒である。履歴は実際に演奏された順番に 1 つずつ増やしていった。

この図を見てわかるように履歴を使用しないで楽譜の情報だけから予測した場合は予測誤差が大きく、個

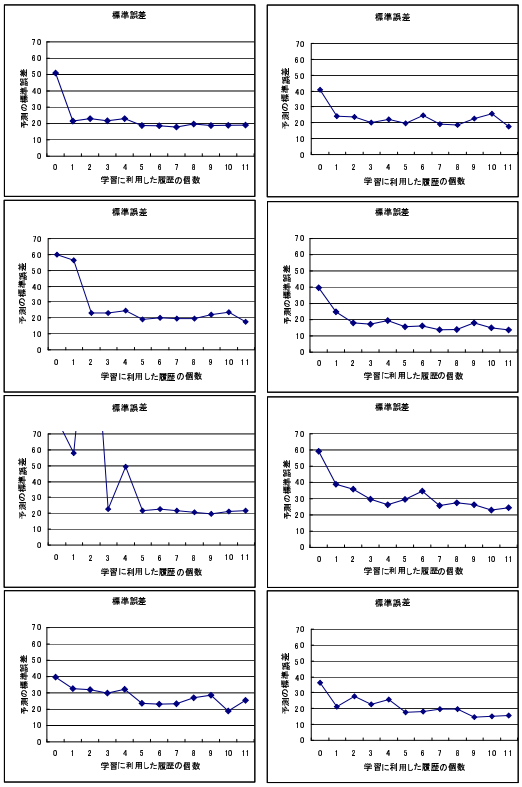

図 5　学習に使用した演奏履歴の個数と標準誤差の関係

人差はあるものの標準誤差が 60ms 程度の人もいる。これは演奏の揺らぎ以外にも楽譜からの逸脱である表情付けがあるために楽譜の情報だけからの予測では誤差が大きくなっていると考えられる。通常は人間の演奏の揺らぎの標準偏差はテンポなどにもよるが 25ms 程度であるし、演奏を学習していった場合に標準誤差が減っていくということは表情付けなどの予測によって解決できる何かしらの楽譜からの逸脱があるということである。

学習が進むにつれて予測の標準誤差は全ての演奏者で減少する傾向にある。本研究の予測モデルが予測に役立っているといえる。

今回のデータを見ると履歴を 11 個使用して十分に学習した場合、予測の標準誤差は 1 名を除いて 26ms 以下におさまっている。この程度の標準誤差ならば演奏の揺らぎの程度であるから、平均してみた場合に予測は十分にされているといっていいだろう。

4.2　採用されたモデルの種類

本研究では独奏予測モデルを複数作成し、一番評価の良かったものを採用した。表 1 はある演奏者で 12 回の演奏履歴を使って学習したときに採用された予測モデルの種類と数である。楽譜の総音符数は 174 であり、そのうち先頭の音符は直前楽音といった情報が使えないため予測が行えないので合計 173 箇所で予測モデルが作成された。横方向はモデルの「楽音の発音時刻」のバリエーション、縦方向はモデルの「拍あたりの時間長」のバリエーションである。表のうちカッコ内は自由度 0 のモデル式、つまり $\alpha = 1$ としてそのままの時間長で予測したほうが良いと判断されたものの数で、表情付けがされていないと考えられる楽音のことである。

表 1　採用された予測モデルの種類と数

	直前楽音の発音	直前拍の発音	直前楽音の消音
直前楽音	36(3)	0	0
直前拍	0	0	0
楽音の 1 小節の平均	68(12)	1	0
拍の 1 小節の平均	68	0	0

表を見ると「直前楽音の消音時刻」の情報はまったく使われていない。この傾向は全演奏者で共通であり、全データのうちで消音時刻での予測が採用された部分は 1 箇所のみであった。人間の消音時刻への感覚は発音時刻と比べると敏感ではない。発音時刻に対して消音時刻の揺らぎはそれほど気にならないため揺らぎが大きく、有効な情報となっていないと考えられる。あるいはピアノという減衰系の音色を用いたため消音時刻には音が減衰してしまいあまり演奏表現に用いられなかったということも考えられ、ヴァイオリンのような音色ではまた違う結果となることもありえる。

次に「直前の拍の拍あたりの時間長」をみるとこれも全演奏者でまったく利用されなかった。「直前の楽音の拍あたりの時間長」の採用された数も過去 1 小節の時間長の平均を使ったモデルと比べると半分程度と少ないことから過去 1 回の時間長で予測するよりは過去 1 小節程度の時間長で平均化して予測したほうが良い結果を出すといえる。

多くの部分の予測は過去 1 小節の平均の時間長で十分に当たっていた。テンポがそれほど変化しない部分では時間長を 1 データ使って予測するよりも 1 小節程度の複数の時間長データを平均して使ったほうが揺らぎの影響を受けにくく安定した予測をすることができると考えられる。

それから「直前拍の発音時刻」が予想よりもあまり基準とされていないということもいえる。この演奏者は拍の時刻を基準とした部分が特に少なく 1 箇所しか採用されなかった。但し直前拍が直前楽音の場合には「直前楽音の発音時刻」にカウントされているため分

析対象とできる楽音位置は他の発音時刻を基準とした予測モデルと比べると少なく、174 箇所中 50 箇所しかないということはある。他の演奏者では平均 10 箇所程度採用されているが、それでも分析対象のうち $\frac{1}{5}$ 程度で、その他の部分は「直前楽音の発音時刻」を基準としたモデルとなった。独奏演奏の場合、拍の位置の時刻情報よりも直前の位置の時刻情報を基準としていることが多いといえる。

5. おわりに

本研究では演奏表現のされた演奏を学習して次回の演奏で有益となる情報を獲得するために人間による表情豊かな演奏を収録し、「楽音の発音時刻」と「拍あたりの時間長」を説明変量に用いた複数のモデルによって重回帰分析した。分析によって独奏者の演奏を予測するモデルを作成し独奏演奏の予測をおこなった。また予測結果の標準誤差などを見てモデルの有効性を確認した。そして生成されたモデルの種類から人間の演奏生成に対する考察をおこなった。楽音はほとんどの場合直前楽音を基準とする、拍あたりの時間長は過去の平均をとったほうが予測が当たる、といったことを確認した。

今後の課題としては予測の難しかった「間」に相当する部分の予測が挙げられる。このような表現は曲によっては多用されるため予測できれば適応できる楽曲が増えるだろう。それから演奏ミスと考えられる部分についても何らかの基準で判別をおこない演奏ミスをした楽音を学習に使わないといったことも考えられる。ただし演奏ミスかどうかはそう簡単には判別できないため演奏ミス候補となる部分をユーザに提示し、手動で削除してもらうといったことでも解決できるだろう。それから今回は独奏演奏のみの履歴から学習をしていたがこれを合奏へと拡張することが考えられる。合奏時には伴奏者の発音時刻などの情報が加わるため相手とのずれといった相互作用的な情報を得ることができる。このようなことも考慮した独奏予測にしなければ伴奏システムへと実装するうえで中途半端なものとなるだろう。

また今後の展望としては独奏演奏の予測に限らず他の場面、たとえば自動演奏システムに応用することも考えられる。本研究の段階では伴奏システム向きの表情付けの学習であり、自動演奏システムの表情付けの生成とは違いがある。自動演奏システムでは楽曲構造や人間の演奏から普遍的な演奏生成ルールを抽出するため未知曲に対しても適用できる。しかし本手法は演奏者ごとに変わる表情付けを学習し適応することで、伴奏システムが追従を行う際に利用するというアプローチである。そのため既知曲で演奏されたものしか予測していない。しかしここへ例えば楽曲構造といったことを導入し、楽曲構造と演奏予測モデルの関係といったことから演奏生成のルールを抽出し人間の演奏生成のメカニズムに迫ることで、自動的な演奏の表情付けのシステムといった研究につながる可能性もある。

参考文献

1) Barry Vercoe Miller Puckette, Synthetic Rehearsal:Training the Synthetic Performer, International Computer Music Conference '85, pp.275〜278,1985
2) Roger B.Dannenberg Kenneth Bookstein, Practical Aspects of a Midi Conducting Program, Proceeding of International Computer Music Conference '91, pp.537〜540,1991
3) 井上渉 橋本周司 大照完, 音声認識を導入した歌声自動伴奏, 情報処理学会第 48 回 (平成 6 年前期) 全国大会,pp.383-384,1994
4) 五十嵐滋 小川大典 松浦陽平 水谷哲也, 楽曲構造や演奏表情の表現と自動協調システム, 人工知能学会全国大会 (第 9 回),pp.627-630,1995
5) 井上渉 橋本周司 大照完, 適応型歌声自動伴奏システム, 情報処理学会論文誌,pp.31-38,1996
6) 小川大典 五十嵐滋 戴岡, 計算機によるピアノ伴奏, 音楽情報科学 16,pp.37-42,1996
7) 野池賢二 野瀬隆 小谷善行 乾伸雄 西村恕彦 関本陽子, 演奏情報に関する楽曲の特徴抽出システムの作成, 第 21 回 音楽情報科学研究会,pp.1-6,1997
8) 野池賢二 野瀬隆 小谷善行 乾伸雄, 演奏情報と楽譜情報の対からの演奏表情規則の獲得とその応用, 第 26 回 音楽情報科学研究会,pp.109-114,1998
9) 堀内靖雄 坂本圭司 市川熹, 合奏における人間の発音時刻制御モデルの推定, 情報処理学会論文誌 Vol.43 No.2,pp.260-267,2002
10) 石川修 片寄晴弘 井口征士, 重回帰分析のイタレーションによる演奏ルールの抽出と解析, 情報処理学会論文誌 Vol.43 No.2,pp.268-276,2002
11) 大照完 橋本周司, 演奏履歴を考慮した自動伴奏, 仮想音楽空間,pp.80〜92, オーム社, 1994
12) 堀内靖雄,『自動伴奏』, 長嶋洋一 橋本周司 平賀譲 平田圭二（編）,「コンピュータと音楽の世界」,pp.252〜269, 共立出版, 1998
13) 五十嵐滋, 演奏を科学する, ヤマハミュージックメディア, 2000
14) 青柳龍也 小坂直敏 平田圭二 堀内靖雄（訳・監修）, コンピュータ音楽－歴史・テクノロジー・アート－, 東京電機大出版, 2001