階層化された相互引き込みモデルに基づくアンサンブルシステム[†]

小林 洋平*・三宅 美博*

New Ensemble System Based on Hierarchical Mutual Entrainment Model[†]

Yohei KOBAYASHI* and Yoshihiro MIYAKE*

It is difficult problem to make a human like ensemble system. In this research, we analyzed the rhythm interaction between performers by measure the hitting time and hitting cycle of keyboard performance. From the analysis, we discovered importance of two effects, which are hitting lag and hitting lag variation between performers. We modeled its dynamics by non-linear phase oscillator, which has two kinds of interaction mechanism, which represents the effect from hitting lag and the effect from hitting lag variation. We evaluated this system with human performer and discovered that the coupling constant of each model determine the dynamics of performance and the balance of these two interactions is important to realize human like ensemble. Finally we show the best parameters of the models and realize human like ensemble system.

Key Words : mutual entrainment, ensemble system, phase oscillator, performance, phase difference

1. はじめに

近年の音楽機器の発達によって、アンサンブルに代表されるように人間同士の演奏の代替として機械との演奏が頻繁に用いられるようになってきている¹⁾。しかしながら、実際の人間同士の演奏と比較すると現状の機械の演奏は未だ及ばない部分が多く、それを改善することが今後の研究の課題である²⁾。人間らしい演奏を行なうアンサンブルシステムを構築するには、大きく3点の問題を解決する必要がある。それは、人間同士がどのように演奏を行なっているかを調べる解析的な問題、得られた解析結果のシステムへの実装についての問題、そして構築したシステムの人間らしさを評価する問題である。解析的な問題については、演奏者同士や演奏者と聴衆者の間の相互作用調べたもの、また機械の演奏にどのように人間が追従するかを解析したものがある^{3)~6)}。また、実装については機械的なスコアマッチングシステムとして特化したものや、人間の演奏者の発音タイミングを回帰分析でモデル化を試みたものがある^{6)~10)}。また、構築したシステムの評価としては、実際に人間と演奏することによって比較評価することが多く用いられている^{11)~14)}。

われわれもこれらの方法に従って人間らしい演奏を行なうアンサンブルシステムの構築を目指してきた¹⁵⁾。リズムの相互作用という視点から人間同士の2者間演奏を計測し解析した結果、打鍵ずれ変化量が打鍵周期変化に影響を強く及ぼしていることを発見した。われわれはこの作用関係を相互引き込みモデルによって再構成し、モデルの振動子と楽譜情報を

対応させることによってアンサンブルシステムとして実装した。

また、構築したシステムの評価には、人間同士の演奏と比較することによって行なった。その方法として、演奏の揃い具合などの客観的指標と演奏者自身の主観的評価によってシステムの評価を行なった。その結果従来のシステムよりもよい評価が得られたが、人間同士の演奏とは未だに差があることもわかった。特に主観的評価で違いがみられたが、その原因として演奏のダイナミクスに差があることが示唆され、それに則した評価とモデルが必要となった。そこで本論文では、まず第2章で人間同士の演奏をより詳細に解析する。これは人間の演奏するタイミングがどのように生成されているかを、相互作用の時間長に着目して行なう。そして第3章でその作用関係を相互引き込みモデルを階層化する形で記述する。そして第4章でモデルをシステムに実装する。最終的に第5章で各モデルのパラメータを変化させながら評価するとともに、第6章でその作用機構について考察する。

2. 人間の演奏の相互作用解析

本研究では人間同士の演奏の相互作用を調べるために、演奏のタイミングとその時間発展に関して解析を行なう。すでに先行研究でも人間の演奏機構の解析として、発音時刻のモデル推定がなされている。それは、テンポの変化があらかじめ定められた機械演奏に、どのように人間が追従していくのかを解析したものである⁹⁾。また、より最近の先行研究においては、人間同士の演奏を独奏者と伴奏者にわけ、より多くのパラメータを用いて重回帰分析によってモデル推定を行なっているものもある⁶⁾。ここでは演奏者同士のリズムのずれがリアルタイムに演奏に影響を及ぼしているという仮説のもと、特に打鍵直前の打鍵周期と打鍵ずれに注目し、時間スケール

[†] 第41回計測自動制御学会学術講演会で発表(2002・8)

* 東京工業大学大学院総合理工学研究科 横浜市緑区長津田町4259

* Tokyo Institute of Technology, 4259 Nagatsuta, Midori, Yokohama

(Received December 24, 2004)

(Revised April 18, 2005)

をかえて関係性を調べる。

いくつかの先行研究においては、一般的なポリフォニックな楽曲を用い、アコースティックな音源を含めたさまざまな音質による同調なども試みられている¹⁴⁾。しかし本研究の目的のひとつである、リズムの相互作用について解析を行なうためには、演奏者の音楽的解釈の影響をできるだけ少なくし、作用関係の同定が容易になるような演奏を行なう必要がある。そこでこの実験では、二人の演奏者がFig.1に示したような単純な曲を、音のたち上がりか最も明瞭な楽器の一つであるピアノを用いて2オクターブずらして同時に演奏する。この曲は打鍵周期が常に8分音符で構成されており、演奏者同士の演奏内容が同じであるため打鍵周期や打鍵ずれの計測も容易である。

それぞれの演奏者は直接的な相互作用が遮断された状況で演奏を行ない、設置されたスピーカーから両者の演奏内容が提示される。演奏実験はピアノ教育を10年以上受けたツェルニー100番練習曲程度の演奏をこなすことができる3人の20代学生(右利き)を被験者として行なわれた。3人の演奏者については組み合わせを変えて上下のパート1回ずつ計6回の演奏を行なわせた。それぞれの演奏者はFig.1の曲について演奏を行なう十分な技術があり、ミスタッチがないよう十分に練習を行なっている。ここではリズムの相互作用を調べることに重点を置き、演奏を同期させるように求め、音楽的解釈などが入らない範囲で自由に演奏を行わせた。試行ごとに最初に機械が150BPMのテンポを提示し、その速度を維持するよう求めた。

計測は、両演奏者の n 番目の打鍵時刻 $t_h(n), t_l(n)$ をPCを用いて1ms単位で計測する。ここで、 $t_h(n)$ は高音側、 $t_l(n)$ は低音側の打鍵時刻である。そこからFig.2にしたがって、打鍵間隔 $t_h(n+d)-t_h(n), t_l(n+d)-t_l(n)$ から打鍵周期(Hitting cycle) $c_h(n), c_l(n)$ を計算する。ここで d は周期を計算するための音符の間隔である。さらに両演奏者の打鍵の時間差(Hitting lag) $t_h(n)-t_l(n)$ を打鍵ずれ $l(n)$ として、打鍵ずれ変化量(Hitting lag variation) $l(n)-l(n-d)$ を $l_v(n)$ として、打鍵周期変化量(Hitting cycle variation) $c_h(n)-c_h(n-d), c_l(n)-c_l(n-d)$ を $c_{vh}(n), c_{vl}(n)$ として算出する。

まとめると、まず、打鍵ずれと打鍵ずれ変化量については以下のように定義される。

$$\begin{aligned} l(n) &= t_h(n) - t_l(n) \\ l_v(n) &= l(n) - l(n-d) \end{aligned} \quad (1)$$

また打鍵周期と打鍵周期変化量については以下のようになる。

Fig.1 Music score for ensemble

Fig.2 Definition of each indicator

Table 1 Correlation coefficient between hitting lag variation $l_v(n)$ and hitting cycle variation $c_{vh}(n), c_{vl}(n)$ in each d

d	1	2	3	4	5	6
c_{vh}	0.47	0.62	0.58	0.66	0.62	0.65
c_{vl}	0.43	0.63	0.53	0.59	0.49	0.52
Average	0.45	0.63	0.55	0.62	0.56	0.58

Table 2 Correlation coefficient between hitting lag $l(n)$ and hitting cycle variation $c_{vh}(n), c_{vl}(n)$ in each d

d	1	2	3	4	5	6
c_{vh}	0.34	0.53	0.50	0.58	0.53	0.55
c_{vl}	0.32	0.50	0.46	0.54	0.47	0.51
Average	0.33	0.51	0.48	0.56	0.50	0.54

Fig.3 Correlation coefficient of hitting lag and hitting lag variation

$$\begin{aligned} c_h(n) &= t_h(n+d) - t_h(n) \\ c_{vh}(n) &= c_h(n) - c_h(n-d) \end{aligned} \quad (2)$$

$$\begin{aligned} c_l(n) &= t_l(n+d) - t_l(n) \\ c_{vl}(n) &= c_l(n) - c_l(n-d) \end{aligned} \quad (2)'$$

このような定義に従って、演奏を40小節以上にわたって計測を行ない、最初10拍分のデータを削除し、そこから150拍を有効なデータとして $t_h(n), t_l(n)$ に使用することとした。

本実験での解析は、打鍵周期変化量 $c_{vh}(n), c_{vl}(n)$ に影響を与えている量として、打鍵ずれ $l(n)$ と打鍵ずれ変化量 $l_v(n)$ を考えた。そこで $l_v(n)$ と $c_{vh}(n), c_{vl}(n)$ の相関値、また $l(n)$ と $c_{vh}(n), c_{vl}(n)$ の相関値をそれぞれ計算し比較する。Table 1に打鍵ずれ変化量 $l_v(n)$ と打鍵周期変化 $c_{vh}(n), c_{vl}(n)$ の相関値を、またTable 2に打鍵ずれ $l(n)$ と打鍵周期変化 $c_{vh}(n), c_{vl}(n)$ の相関値を、打鍵間隔 d と対応させる形で示す。相関値の高音側は必ず負となったが、ここではその絶対値のみを示しAverageはその平均である。この平均を、横軸に打鍵間隔 d として比較した図をFig.3に示す。

われわれはこれまでの研究で、Fig.3から読み取れるように、より相関が高い打鍵ずれ変化量の影響のみを、相互引き込みとしてモデル化してシステムに実装してきた^{15), 16)}。その結果、人

間同士の演奏をある程度は再現できたものの、未だに差があることもわかった。そこで今回は、打鍵ずれの影響についても考慮することによって、より人間同士の近い演奏を目指す。しかし、それぞれがどのような時間長で影響をより強く及ぼしているかを考慮しなければならない。まず、打鍵ずれ変化量についての相関値を見てみると、1拍と2拍に対応する間隔2と間隔4で最も高い相関値を示している。一方打鍵ずれの方は、2拍に対応する打鍵間隔4で最も高い相関値を示し、ついで3拍に対応する打鍵間隔6で最も高い相関となっている。

このように、それぞれの作用は影響を及ぼす時間長が異なっていると考えられ、1拍や2拍程度の短時間では、打鍵ずれ変化量から打鍵周期変化量が強い影響を及ぼしており、それより長時間になるほど、打鍵ずれからの影響が強くなっていくことが考えられる。これらの記述についての詳しい方法は次章で述べるが、前者の影響は同じ次元にあるため、時間相関に関係ない作用関係によって記述することが可能である。一方後者の影響は違う次元にあるため、時間相関を考慮する必要がある。Fig.3をみると、長時間スケールについて影響が強いことから、長時間相関によって記述することが必要であると考えられる。

3. モデルの提案

本章では前章で得られた結果を、一般的な連続なモデルとして近似を行なう。先行研究においては離散的に回帰分析からモデル化を行なったものがあるが、ここではその中でも特に、第1の近似として打鍵ずれ変化量からの影響を、第2の近似として打鍵ずれからの影響をモデル化する^{5),6)}。

まず第1の近似については、高音側の打鍵ずれ変化量と打鍵周期の関係はある負の値 k を用いて

$$c(n) - c(n-d) = k\{(t_h(n) - t_l(n)) - (t_h(n-d) - t_l(n-d))\} \quad (3)$$

と表現することができる。つまり打鍵ずれ変化量と打鍵周期変化に比例関係が成り立つということである。これを変形すると

$$c(n) - k\{t_h(n) - t_l(n)\} = c(n-d) - k\{t_h(n-d) - t_l(n-d)\} \quad (4)$$

と表現することができ、この等式がすべての n について成り立つと仮定すると

$$\begin{aligned} c(n) - k\{t_h(n) - t_l(n)\} &= c(n-d) - k\{t_h(n-d) - t_l(n-d)\} \\ &= c(n-2d) - k\{t_h(n-2d) - t_l(n-2d)\} \\ &= c(n-3d) - k\{t_h(n-3d) - t_l(n-3d)\} \cdots \\ &= c(0) - k\{t_h(0) - t_l(0)\} \end{aligned} \quad (5)$$

となる。つまり、上式が成りたつようにするには、ある定数 a を用いて n 番目の音符について

$$c(n) = k\{t_h(n) - t_l(n)\} + a \quad (6)$$

と記述できればよい。(6)式は(5)式が成り立つための十分条件であり、(6)式には d が含まれないことから、任意の d について成り立つ関係式として表現できることがわかる。このような作

用関係を扱う方法は色々あるが、ここでは位相振動子による引き込み現象によってモデル化を試みる¹⁴⁾。そこで a を固有振動数 ω を用いて $2\pi/\omega$ 、 $c(n)$ をみかけの振動数 $\omega_p(n)$ を用いて $2\pi/\omega_p(n)$ 、打鍵ずれを位相差 $\phi(n)$ とみかけの振動数 $\omega_p(n)$ を用いて $2\pi\phi(n)/\omega_p(n)$ で記述すると、

$$\omega_p(n) = \omega - k\omega\phi(n) \quad (7)$$

となり、固有振動数を基準としてみかけの振動数が変化する現象として記述が可能となる。ここで k は負であるから $-k$ は正となり、固有振動数を中心としてみかけの振動数と位相差が比例関係にある。なお、適用可能な時間長が d に依存しないことから、すべての時間長に適用可能なモデル式であると考えられる。

つぎに第2の近似として打鍵ずれからの影響のモデル化を考える。ある負の定数 h を用いて打鍵ずれから打鍵周期への影響は以下のように記述することができる。

$$c(n) - c(n-d) = h\{t_h(n) - t_l(n)\} \quad (8)$$

これを(5)式と同様に展開して行くと

$$\begin{aligned} c(n) &= c(n-d) + h\{t_h(n) - t_l(n)\} \\ &= c(n-2d) + h\{t_h(n) - t_l(n)\} + (t_h(n-d) - t_l(n-d)) \end{aligned} \quad (9)$$

$$\cdots = c(0) + h \sum_{j=0}^{n/d} \{t_h(jd) - t_l(jd)\}$$

となり(ただし $n=jd$)、定数 a_0 を用いることによって以下のように記述することができれば上式は成り立つ。

$$c(n) = h \sum_{j=0}^{n/d} \{t_h(jd) - t_l(jd)\} + a_0 \quad (10)$$

これを(6)式と同様に a_0 を基準振動数 ω_0 を用いて $2\pi/\omega_0$ 、 $c(n)$ を固有振動数 $\omega(n)$ を用いて $2\pi/\omega(n)$ 、打鍵ずれを位相差 $\phi(n)$ と固有振動数 $\omega(n)$ を用いて $2\pi\phi(n)/\omega(n)$ で記述すると

$$\omega(n) = \omega_0 - h\omega_0 \sum_{j=0}^{n/d} \phi(jd) \quad (11)$$

となる。この(11)式の h についても k と同様に負であるため $-h$ は正となり、位相差によって固有振動数が変化するより長時間スケールの制御機構であることがわかる。

また、前章の結果より(7)式と(11)式のそれぞれの役割について考える必要がある。まず(7)式は位相差と見かけの振動数の関係が一次関数の関係にあることを示している。つまり(7)式はすべて n についての変数で記述できるため、Fig.3に示した打鍵ずれ変化量からの影響をすべての時間長について表現できているものと考えられる。一方(11)式は位相差の総和が振動数を決定するため、その数によって挙動が異なり、表現できる時間長が異なってくる。たとえば $d=1$ であるような1音符ごとの高頻度で集計すると、結果的に位相差に対して敏感に振動数を変化させることとなり、短時間相関が強くなる。逆に d を大きくすると振動数は変化しにくい。Table 2の結果を考慮すると、より低頻度で変化させ、長時間相関を表現することが妥当であると考えられる。つまり前章における結果の d との関連性を考慮すると(7)式では、 $\phi(n)$ は1音符ごとに対して計算されるの対

して、(11)式では1音符ごとだけでなく、より長い間隔で計算されることが妥当であると考えられる。

4. システムの実装

前章で得られた(7)式と(11)式の結果を併記する形で記述すると以下ようになる。

$$\omega_p(n) = \omega(n) - k\omega(n)\phi(n) \quad (12)$$

$$\omega(n) = \omega_0 - h\omega_0 \sum_{j=0}^n \phi(jd) \quad (13)$$

ここで $k < 0$, $h < 0$ であり, ω_0 は基準振動数, $\phi(n)$ は n 番目の音符の位相差, $\omega(n)$ が出力されるみかけの振動数である。このように, 上位の作用機構と下位の作用機構の併記によって, 前章までの結果を記述することができる。

人間との演奏は Fig.4 に示したように行なわれるものとする。Fig.4において ω_{ph} , ω_{pm} は人間と機械それぞれの演奏されるテンポに対応する。それぞれの演奏者の楽譜上の演奏位置は, 楽譜と1対1対応で定義された位相 θ_m と θ_h として提示され, 位相差 $\phi_m = \theta_h - \theta_m$ と $\phi_h = \theta_m - \theta_h$ が(12)(13)式で示された関係に従って ω_{ph} , ω_{pm} を変化させる。このうち, システムに実装される作用機構は, Fig.4の右側のみであり, 以下のような非線形な位相振動子の引き込みとして表現される¹⁶⁾。

$$\omega_{pm}(n) = \omega_m(n) \cdot \{1 + K \cdot \sin(\phi_m(n))\} \quad (14)$$

$$\omega_m(n) = \omega_0 \{1 + A \sum_{j=0}^n \sin(\phi_{sm}(j))\} \quad (15)$$

$$(\phi_{sm}(j) = \phi_m(jd), K > 0, A > 0)$$

このように相互引き込みを, (14)式のような下位モデルと, (15)式のような上位モデルで階層化することによって記述がなされる。これらのモデル式では相互作用項に $\sin$ を用いることによって, 振動子の一周期 2π を一拍として換算し, 曲の基本単位である一拍単位でのリズムの相互作用を実現する。ここでみかけの振動数 $\omega_{pm}(n)$ は機械の演奏されるテンポに対応し, 演奏される速度は, 基準振動数 ω_0 と位相差 $\phi_m(n)$ によって決定される。 K と A はそれぞれの方程式の結合係数であり, 正の値をとる。それぞれの演奏者の楽譜上の演奏位置は, 楽譜と1対1対応で定義された位相 θ_m と θ_h として提示される。その差 $\theta_h - \theta_m$ を位相差 $\phi_m(n)$ として計算し, 機械にフィードバックして演奏を変化させる。

この上位モデル, 下位モデルのうち, (14)式の下位モデルの $\phi_m(n)$ は1音符ごとに決定されるが, (15)式の上位モデルにおける $\phi_{sm}(j)$ は, $\phi_m(n)$ から特定の音符についてサンプリングして入力する必要がある, そのサンプリングの方法で大きく挙動が異なることは容易に推察される。2章の結果で示したように, (13)式, (15)式はより長時間スケールに関する相互作用を行なうものであるが, ここでは最も最近の6個の $\phi_m(n)$ から平均値を計算して, $\phi_{sm}(j)$ として(15)式に与え動作させる。よって, 特定周期の相互作用のみ扱うのではなく, Fig.3の複数のピークに示されるような, 複数周期での相互作用を行なえとえられる。

Fig.4 System structure

このようにして $\omega_{pm}(n)$ が決定されるが, そこから θ_m への変換の計算は, 実時間で動作するシステムであることを考慮し, 時刻 t において演奏される位相 $\theta_m(t)$ は以下のような式に従う。

$$\begin{aligned} \dot{\theta}_m(t) &= \omega_{pm}(n) \\ (t_\phi(n) \leq t < t_\phi(n+1), \theta_m(0) &= 0) \end{aligned} \quad (16)$$

ここで, $t_\phi(n)$ は n 番目の音符が入力された時刻であり, つぎの音符が入力されるまでみかけの振動数を一定にすることで, 音符間の演奏を補完することを意味している。同様に位相差 $\phi_m(n)$ も音符が入力された瞬間に計算された位相差で音符間を補完する。詳しい方法は参考文献を参照していただきたい¹⁵⁾。

(14)式の下位モデルに関しては以下のようなアルゴリズムにしたがって動作を行なう。

- 1: 最も最近に計測された人間の演奏から位相差を計算。
- 2: (14)式にしたがって $\omega_{pm}(n)$ を計算。
- 3: (16)式にしたがって θ_m を増加させ, 演奏を進行させる。
- 4: 1に戻る。

一方, (15)式は(14)式とダイナミクスを切り分けるために, 先行研究で提案している方法を用いる¹⁷⁾。その方法は, 6個の音符の位相差が一定以内に収まっていれば, その平均値で位相差 $\phi_{sm}(j)$ とする。このように, 位相差が一定以内に安定している場合では, 下位モデルの動作が少ないため, 相対的に上位モデルの作用が反映される。その結果, 上位モデルと下位モデルそれぞれのダイナミクスを切り分けることが可能になる。人間の演奏では位相差の標準偏差が 15° 程度であったので, 6個の音符では 90° が一般的な位相差ゆらぎと考えられる。ここではその半分の 45° より小さければ動作させることにする。よって, 以下のようなアルゴリズムにしたがって動作を行なう。

- 1: 音符ごとに計測された位相差を順番に記録。
- 2: 現在の音符と6個前までの音符との位相差を比較。
- 3: 上記の差の総和が 45° 以内であれば位相差の平均値を計算し(15)式の $\phi_{sm}(j)$ に新たなデータとして加える。
- 4: 差の総和が 45° より大きければ $\phi_{sm}(j)$ に変更は加えない。
- 5: 与えられた $\phi_{sm}(j)$ と(15)式にしたがって $\omega_m(n)$ を計算する。
- 6: 1に戻る。

このアルゴリズムは人間の演奏が入力されるたびに行なう。

5. 結果

5.1 解析方法

モデルの挙動は結合係数 K と A に強く依存するため, 適切

にパラメータを決定する必要がある。そこで2章で得られた人間同士の演奏をリファレンスとし、パラメータを決定する。ここでは(14)式の K が0.05, 0.1, 0.15, 0.2の場合, (15)式の A が0, 0.05, 0.1, 0.15, 0.2の場合について調べる。演奏を評価する指標としては、以下の3種類を用いる。それは演奏を通しての位相差 $\phi(n)$ の平均値の絶対値, $\phi(n)$ の標準偏差, そして $\phi(n)$ のゆらぎのスペクトル解析から得られる近似直線の傾き α である。ここで $\phi(n)$ は、一試行中における n 番目の音符の位相差である。

$\phi(n)$ の平均値の絶対値と $\phi(n)$ の標準偏差は演奏のそり具合をあらわす指標である。曲を通して各音符の位相差が $\phi(n)$ で与えられ、曲全体の音符数が N であった場合、平均値の絶対値 $\phi_{average}$ と標準偏差 ϕ_{stdev} は以下のように計算される。

$$\phi_{average} = \frac{\sum_{n=0}^{n=N} \phi(n)}{N} \quad (17)$$

$$\phi_{stdev} = \sqrt{\frac{1}{N-1} \sum_{n=0}^{n=N} (\phi(n) - \bar{\phi})^2} \quad (18)$$

この指標は一試行ごとに計算され、演奏の同調度合いを示す統計的指標であり多くの先行研究で用いられてきた典型的指標である^{5), 6), 10), 15)}。これに加えて本研究ではダイナミクスを比較する指標として、スペクトルの解析結果を用いる。このスペクトルの傾き α は、横軸を周波数、縦軸をパワーで両対数プロットした時の近似直線の傾きによって計算される。この値が大きいほど傾きが右下がりとなるため、低周波のゆらぎが大きいことを表わし、逆に α が0に近くなるほどホワイトノイズに近くなっていることを表わす。つまり新たにこの指標を用いることによって演奏がどのように時間発展するかも含めて比較できるようになると考えられる。

本実験では人間と機械の演奏に加えて、比較実験として2章と同様の環境で3人の演奏者がそれぞれ上下のパートを1回ずつ演奏を行ない、一つの条件につき計6回の演奏を集計した。また、機械の ω_0 は第2章の条件とあうように150BPMに設定した。

5.2 解析結果

まずFig.5に K と A を変化させた場合の結果を示す。グラフは上からそれぞれ、Fig.5aが位相差の平均値、Fig.5bが標準偏差、Fig.5cが位相差ゆらぎのスペクトルの傾き α についてであり、全6試行についての平均である。これらのグラフは K と A をXY平面上にプロットし、Z軸を各値に対応して三次元のグラフにしたものである。また、Fig.6は位相差の時間発展の典型例を示したものである。これらの結果より、まず位相差の平均値と標準偏差についてみてみると、 $K=0.1, A=0.1$ 付近で極小値をとるような谷型の傾向を示すことがわかり、演奏が最も安定する。また、ゆらぎのスペクトルの傾き α についてみてみると、 A が増加するほど、そして K が減少するにしたがって負の傾きが大きくなることわかる。Fig.6の時間発展をみてみると、 A が増加し K が減少する右上に行くほど大きなゆらぎが発生していることがわかる。

Fig.5 Result of difference in each K and A

そこで、この結果をより詳細に人間同士の演奏と比較するために $K=0.1, A=0.1$ を含む直線で、それぞれの指標において比較を行なう。Fig.7に人間同士の演奏として、各演奏者の組み合わせでの位相差の時間発展と、位相差スペクトルの結果を示す。このように人間同士の演奏では、スペクトルが負の傾きをもち、時間発展も大きなゆらぎと小さなゆらぎが混在しているのが特徴的である。

5.3 Aに対する依存性

Fig.8に K を0.1に固定し、 A を変化させた場合の結果を示す。Fig.8aが位相差の平均値、Fig.8bが標準偏差、Fig.8cが位相差ゆらぎのスペクトルの傾き α についてである。グラフ中でHumanとして示された部分は、比較のために人間同士の結果を示したものである。まず、位相差の平均値を見ると、 $A=0.1$ を最小とした左右対称の傾向が見られる。人間の演奏と比較す

Fig.6 Temporal development of phase difference in each K and A

Fig.7 Temporal development and spectrum of phase difference of humans

ると A が0.1に近づくほど人間同士の演奏とは異なってくる
ことがわかる。Fig.7を見ればわかるように、人間の演奏は0を
中心とした演奏ではなく、かなりずれを含んだ演奏が行なわれ
ることが特徴的である。一方、位相差の標準偏差については、
 A が0.0-0.1の間で小さな値をとり、人間同士の結果とほぼ同じ
値を得ることがわかり($p>0.5$)、さらに A の値が増大するにつ
れて標準偏差も増大することがわかる。また同時に、 A の値が

増大するとスペクトルの傾きも負の方向に増大することがわか
る。この傾きは A が0.1-0.15で人間同士の演奏に最も近づくこ
とがわかる($p>0.5$)。また位相差の時間発展を見ても、 A が増大
するにつれて大きなゆらぎが生成され、人間同士の演奏と比較
すると A が0.1で最も波形が類似することがわかる。

5.4 K に対する依存性

A を0.1に固定して、 K を変化させた場合の結果をFig.9に示

a) Average of phase difference b) S.D. of phase difference
c) spectrum of phase difference

Fig.8 Result of K=0.1

す。Fig.9aが位相差の平均値, Fig.9bが標準偏差, Fig.9cが位相差ゆらぎのスペクトルの傾き α についてである。グラフ中でHumanとして示された部分は、比較のために人間同士の結果を示したものである。まず位相差の平均値についてはKが0.1が最も小さいが、人間同士の値と比較すると、どのKにおいても有意に小さいことがわかる($p < 0.05$)。つまり、位相差の平均値についてはKの値によらず0付近に収束する。また位相差の標準偏差については、Kが0.1を最小とした谷型の傾向を示し、この値で人間と同様の結果を示した($p > 0.5$)。また、スペクトルの傾きについては増加するほど負の傾きが減少することがわかり、これもKが0.1で人間同士の値に近づく($p > 0.5$)。また位相差の時間発展を見ると、Kが小さい場合には大きな揺らぎが発生し、Kが大きくなると小刻みな揺らぎが発生することがわかる。Aが0.1で人間同士の演奏に最も波形が類似することがわかる。

5.5 目標位相差を含むモデルの提案と結果

ここまでの結果でKが0.1, Aが0.1-0.15で位相差の標準偏差とゆらぎのスペクトルの傾きについては、人間同士の演奏に近くなることがわかった。しかし位相差の平均値についてはどれも人間同士よりも有意に小さく異なっている。このような、位相差の平均値が0に収束してしまうという現象は、(15)式の上位モデルが動作することによって生じている。そこで、0ではなく任意の値で収束可能にする拡張モデルを以下に示す。

a) Average of phase difference b) S.D. of phase difference
c) spectrum of phase difference

Fig.9 Result of A=0.1

$$\omega_m(n) = \omega_0 \left(1 + A \sum_{j=0}^n \sin(\phi_{sm}(j) - \phi_0) \right) \quad (19)$$

ここで ϕ_0 は目標位相差であり、この値で位相差が収束することを期待している。この新しい上位モデルを用いた時、K=0.1, A=0.1で ϕ_0 を変化させた場合の結果をFig.10に示す。Fig.10aが位相差の平均値, Fig.10bが標準偏差, Fig.10cが位相差ゆらぎのスペクトルの傾き α についてである。この結果を見ると位相差の平均値が最小となる $\phi_0=0$ を中心とし、 ϕ_0 の絶対値が増加するにつれて平均値が増大する傾向が観察され、 ϕ_0 の絶対値が 10° 程度で人間同士の演奏に最も近づくことがわかる($p > 0.5$)。Fig.11は目標位相差を $-20^\circ, -10^\circ, 10^\circ, 20^\circ$ にした場合の位相差の時間発展である。このように波形の形状自体は大きな変化はないものの、位相差の中心が0よりも上下にずれていることがわかる。一方、位相差の標準偏差については、目標位相差 ϕ_0 が負側に大きくなるほど増大する傾向が観察された。同様に、位相差のスペクトルの傾き α についても、目標位相差 ϕ_0 を負にするほど傾きが大きくなることがわかる。このように新しいモデルを用いることによって、位相差の平均値についても人間同士の演奏に近づけることができ、同時に位相差の標準偏差やスペクトルの傾き α に変化があることもわかった。

a) Average of phase difference b) S.D. of phase difference
c) spectrum of phase difference

Fig.10 Result of each offset ϕ_0

6. 考察

6.1 結果のまとめ

われわれは人間同士の演奏を解析し、打鍵ずれ変化量から打鍵周期への影響と、打鍵ずれから打鍵周期への影響を比較した。このうち、時間長に関係なく前者の影響がより強いことがわかり、相対的に長時間長では後者の影響が強くなることがわかった。この二つの作用をモデル化するにあたり、前者は固有振動数を基準として見かけの振動数が変化する下位モデルによって、後者の作用はこの固有振動数が変化する上位モデルによって記述した。これらのモデルを人間同士の演奏に近くなるように結合係数を適切に調整した。人間の演奏との比較には演奏者間の位相差を計測し、平均値の絶対値、位相差の標準偏差、位相差ゆらぎのスペクトルの近似直線の傾きを用いた。最終的に適切なパラメータをとることによって、人間同士の演奏に近い値を再現することができた。結果的に、上位モデルと下位モデルのバランスが人間らしい演奏を実現するために重要であることがわかったが、それぞれの作用が演奏にどのように関わっているかを考察しておくことが必要である。

6.2 先行研究の問題点

人間同士の演奏では単純にお互いの演奏をあわせるだけでなく、そこに複雑な構造があることが特徴的である。たとえば、Fig.7のスペクトルの結果からもわかるように、人間同士の演奏では1/f性をもって行なわれることが特徴的である^{18),19)}。

Fig.11 Temporal development of phase difference of each offset ϕ_0

これは、人間らしい演奏は単純な線形フィードバックだけではなく、より階層化された複雑な作用の存在を示している。また、Fig.7の位相差の時間発展を見ても、複雑なゆらぎが観察されている。このような人間の演奏を再現するために、先行研究では(14)式に示されるような下位モデルのみを用いてきた¹⁵⁾。このモデルで人間に近いパラメータ $K=0.2$ で演奏を行なった場合では、Fig.12に示すような結果となる。この結果でFig.12aが位相差ゆらぎの時間発展の一例、Fig.12bが位相差ゆらぎのスペクトルの一例である。Fig.12aに示したように、下位モデルを用いた場合では、位相差の時間発展が小刻みなゆらぎになる。このようなゆらぎをスペクトル解析するとFig.12bに示したような結果となり、特定周波数にピークをもち、かつ背景にホワイトノイズ性が強い傾向を示している。

このような差がどれほど演奏者の主観に反映されているかを示しているのが、Fig.12cに示したようなチューリングテストの結果である²⁰⁾。ほかの先行研究ではより一般的かつ複雑な楽曲で聴取者による主観的な比較をしたものがあるが、ここでは本研究で使用するような単純な楽曲でも差が出るかということを演奏者の主観で比較している²¹⁾。Fig.12cは、人間同士の

演奏、下位モデルのみの演奏($K=0.2$)、単独演奏の録音との演奏($K=0$)を、いかに人間らしかったかという指標で順位付けを行なわせた結果である。 $K=0.2$ は $K=0$ の場合より改善はしているものの、人間との間には差が存在する。確率過程的な観点からも、人間同士のものの定常性が弱く、一方ここで用いたモデルは定常性が強いと考えられる²²⁾。先行研究で用いたような下位モデルは直前の位相差のみに従うため、マルコフ過程に近く、定常性が強くなっている可能性があり、人間らしい演奏との差となっていると考えられる。このようなダイナミクスの視点から評価を行ない、改善をすることが先行研究で残された課題であった。

6.3 本研究における改善点

本研究ではより詳細に人間同士の演奏を解析し、打鍵ずれ変化量から打鍵周期への影響だけでなく、先行研究では行なわれなかった打鍵ずれから打鍵周期への影響についてもモデル化を試みた。このモデルとして、新たに(15)式に示すような上位モデルを加え、システムに実装した。これらのモデルを用いて、人間同士の演奏の再現を試みたところ、適切なパラメータで人間同士の演奏が再現された。そこでシステムの挙動を調べたところ、各モデルの作用によって演奏のダイナミクスが変化することがわかった。下位モデルは低周波成分を減少させ高周波成分を増加させることがわかり、これはすでに述べたように、先行研究の問題点でもある。一方、上位モデルは低周波成分を増加させ高周波成分を減少させることがわかった。このように下位モデルは短時間スケールの作用が強く、対照的に上位モデルは長時間スケールの作用が強いことがわかる。つまり、下位モデルはすでに述べたように定常的な相互作用を行ない、上位モデルは非定常的な相互作用を行なうと考えられる。このように、それぞれのモデルの作用する割合によって、ダイナミクスが変化することがわかり、結果的に先行研究の問題点が改善された。

確率過程的な観点からも、今回用いたモデルの妥当性が裏付けられている²²⁾。新たに加えた上位モデルは、過去の位相差が蓄積されて現在の演奏に影響を及ぼすため、非マルコフ過程に近く、非定常性が強くなると考えられる。一方下位モデルはすでに述べたように、マルコフ過程に近く定常性が強い。このように、上位モデルと下位モデルの割合によってダイナミクスが変化することは、確率過程で考えれば必然であるといえる。また、主観的な違いにおいては、上位モデルを加えた演奏は演奏速度に大きな変化を与えるため、先行研究までのシステムとは違った感覚を演奏者に与えることもわかった。今回は演奏者にとっての主観的な人間らしさについては、体系的な評価は行なわなかった。しかし、今後は主観的な評価についても調べることが、今回提案した上位モデルの妥当性を示す上で必要な課題であると考えられる。たとえば先行研究で用いたチューリングテストなども、今回構築したシステムの評価に適用可能であると考えられる²¹⁾。

ただし、本論文で得られた知見は、テンポを150BPMに固定し、楽譜構造が半拍のみで構成された場合に限定される。ほか

のテンポでの演奏や楽譜構造が異なる場合においては、違った傾向が出ることも予想され、今回得られた知見がどれだけ適用可能かを調べる必要がある。特に今回明らかになった相互作用が、楽譜時間依存であるのか、もしくは実時間依存であるのかは重要な問題である。Fig.13は第2章と同様の実験を設定テンポをかせて行なった結果である。Fig.13aは打鍵ずれ変化量からの相関係数、Fig.13bは打鍵ずれからの相関係数を三次元グラフにしたものである。このように演奏テンポを変化させるにともなう、相関係数のピークが移動していくことがみてとれる。このことから人間の演奏者は楽譜時間が同じでも実時間スケールにも依存して演奏が変化することが推測される。つまり本論文でわれわれが提示した知見とモデルをFig.13に示したような結果にも対応させられるようにする必要がある。これをどのように対応させるかは本論文では残された課題として今後の展望としたいが、幸い、今回われわれが提案したモデルが連続的なモデルであるため、任意の実時間スケールに対応させたモデルに拡張させることは比較的容易であるように考えられる。

7. まとめ

人間同士の演奏を理想としたアンサンブルシステムの構築を行なうため、人間同士の演奏の相互作用を詳細に解析した。その結果、打鍵ずれ変化量と打鍵ずれの、二つの作用が打鍵周期に影響していることがわかり、それぞれ下位モデルと上位モデルによって記述した。それぞれのモデルの結合係数を調整し、人間同士の演奏に近づけることができた。モデルの挙動を比較したところ、下位のモデルは高周波、上位のモデルは低周波の作用が強く、ダイナミクスが異なることがわかった。つまり、人間らしい演奏には、それぞれのダイナミクスをもった作用が、適切に作用していることが示唆された。今後は演奏者の主観的な評価についても、モデルの妥当性を検証し、より一般的な楽曲やさまざまなテンポに対応できるモデルを構築していきたい。

参考文献

- 1) 矢口木三:マッキントッシュミュージック入門, 西東社 (1993)
- 2) 長島, 橋本, 平賀, 平田:コンピューターと音楽の世界, 共立出版 (1998)
- 3) 山本, 三宅:音楽の生演奏時における演奏者と聴取者の相互作用の解析, 計測自動制御学会論文集, 38-9, 800/805 (2002)
- 4) B.Veroce:The synthetic performer in the context of live performance, Proc. of ICMC, 199/200 (1984)
- 5) 堀内, 坂本, 市川: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, 43-2, 260/267 (2002)
- 6) 堀内, 坂本, 市川:合奏時の人間の演奏制御の分析・推定, 情報処理学会論文誌, 45-3, 690/697 (2004)
- 7) P.Desain,H.Honing and H.Heijink:Robust Score-Performance Matching:Taking Advantage of Structural Information, Proc. of ICMC, 337/340 (1997)
- 8) L.Grubb and R.B.Dannenberg:A Stochastic Method of Tracking a Vocal Performer, Proc. of ICMC, 301/308 (1997)
- 9) R.Rowe:Interactive Music Systems Machine Listening and Composing, The MIT Press (1993)
- 10) 井川, 直井, 大, 橋本:相互作用モデルによる実時間適応自動伴奏とその動作解析, 電子通信情報学会春季全国大会講演論文集, 7, 216/216 (1990)
- 11) 井上, 橋本, 大照:適応型歌声自動伴奏システム, 情報処理学会論文集, 37-1, 51/56 (1998)
- 12) 瀧陽, 橋本:一定の主従関係を持たないインタラクティブな自動伴奏システム, 情報処理学会第60回全国大会講演論文集, 243/244 (2000)
- 13) 東, 橋本:音声認識とピッチ検出を併用した歌声の自動伴奏, 音楽情報科学研究会例会, MUS-22-1 (1997)
- 14) Christopher Raphael:A Hybrid Graphical Model for Aligning Polyphonic Audio with Musical Scores, Proc. ISMIR, 171/178, (2004)
- 15) 小林, 三宅:相互引き込みモデルを用いたアンサンブルシステムの開発, 計測自動制御学会論文集, 40-9, 948/957 (2004)
- 16) 蔵本, 川崎, 山田, 甲斐, 篠本:パターン形成, 朝倉書房 (1991)
- 17) 三宅, 宮川, 田村:共創出コミュニケーションとしての人間-機械系, 計測自動制御学会論文集, 37-11, 1087/1096 (2001)
- 18) 井上, 宮本, 大橋, 上田, 斎藤 共著:ゆらぎの科学 1, 森北出版 (1991)
- 19) 土井, 秋山, 松田:楽器演奏時の1/f ゆらぎに関する考察, 94年音楽音響研究会予稿 (1994)
- 20) A.M.Turing:計算する機械と知性について, Mind, 59, 433/460 (1950)

a)

b)

a)hitting lag variation b)hitting lag

Fig.13 Correlation coefficient of hitting lag and hitting lag variation in each tempo

- 21) 野池, 橋田, 平田, 片寄, 平賀: NIME04 Rencon 開催報告と次回への課題, 情報処理学会研究報告 音楽情報科学, 2005-14 (2005)
- 22) 中川, 真壁:確率過程, 培風館, 75/113 (2002)

「著者紹介」

小林 洋平

2001年, 東京工業大学大学院総合理工学研究科知能システム科学専攻博士前期課程修了。2005年東京工業大学大学院総合理工学研究科知能システム科学博士後期課程修了(工学博士)。現在, 京都産業大学経済学部特定研究員。ヒューマンインタフェース, ヒューマンコミュニケーションの研究に従事。

三宅 美博(正会員)

1989年東京大学大学院薬学系研究科博士課程修了(薬学博士)。金沢工業大学情報工学科助手, 講師, 助教授を経て, 96年より東京工業大学大学院総合理工学研究科知能システム科学専攻助教授。99年よりミュンヘン大学客員教授併任。主として生命的自律性の研究に従事, 生物物理学会, IEEEなどの会員。