

音楽のリズム認知過程の計算論的モデル

— テンポの影響を考慮したモデルによるコンピュータシミュレーションと
モデルの心理学的妥当性の検討 —

後藤 靖 宏

目 次

1. はじめに
 - 1.1 背景と目的
 - 1.2 音楽の時間的側面の処理に関する音楽情報処理および認知心理学的アプローチ
 - 1.3 拍節構造知覚過程に関する既存の心理学的モデルの問題点
2. テンポの影響を考慮に入れた拍節構造知覚過程のモデルの構築
 - 2.1 Longuet-Higgings and Lee (1982) によるモデル
 - 2.2 新たに構築するモデルに加えられた心理学的制約
 - 2.3 モデルのアルゴリズム
3. モデルの評価
 - 3.1 拍節構造の知覚過程に関する認知心理学的実験
 - 3.2 実験結果とシミュレーション結果との比較
4. まとめ

1. はじめに

1.1 背景と目的

本研究の目的は、人間がメロディの“拍節構造 (metrical structure)”を知覚する過程について、認知心理学的に妥当性の高い計算論的モデルを構築すること、および、そのモデルの心理学的な妥当性について評価を行うことである。

音楽と、それに対する人間の関わりについては、非常に幅広い領域で研究が行われてい

る。たとえば、音楽学 (musicology) の領域では、音楽そのものに関する歴史的・理論的研究が行われてきたし、音楽美学 (aesthetics of music) の領域では、「音楽の本質に関わる美学的な諸問題」が扱われてきた¹⁾。音楽社会学 (sociology of music) の領域では、音楽が種々の社会現象と絡み合った時に提起される多様な問題が研究されているし、音楽教育学 (pedagogy of music) の領域では、音楽をいかに教授するかといったことを主眼において行われてきた研究がある。

このような中で、音楽情報処理と認知心理学の各分野において、人間が音楽をどのように認知するのかといった側面について、それぞれ独自の努力が継続されている。音楽情報処理の分野では、音源分離や楽曲分析、演奏・音響分析などを行い、計算機上で音楽家のもつ様々な能力を実現しようとする研究が行われている。一方、認知心理学の分野では、人間が音楽を認識するために必要な心内処理そのものに焦点をあて、心理学的実験やモデル化などの作業を通じて、その過程を明らかにしようとする研究が行われている。

本論文では、特に音楽の時間的側面に焦点を絞り、その知覚過程を計算論的モデルとして構築する。同時に、心理学的実験を行って、そこから得られた結果とモデルのシミュレーション結果と比較することにより、人間の音楽知覚過程の基本的な性質を明らかにしていくことを目標とする。

1.2 音楽の時間的側面の処理に関する音楽情報処理および認知心理学的アプローチ

先行研究²⁾によれば、人間が知覚する音楽の時間構造には、“グルーピング構造 (grouping structure)”と“拍節構造 (metrical structure)”という異なった2つの側面があるとされている。前者は、音楽用語でいうところの“フレーズ(phrase)”などにあたるものであり、たとえば演者が息つきなどをする箇所^{*}で区切られるまとまりのことである。一方、後者は、音楽用語でいえば“拍子(meter)”や“小節 (measure)”などにあたるものであり、一定の時間間隔を単位として区切られるまとまりのことをいう。本論での考察対象は拍節構造についてであり、以後はその知覚に限定して論を進める。

音楽情報処理の分野では、拍節構造の処理について、いわゆる“ビートトラッキング (もしくはフットタッピング)”問題を中心として、多くの研究が報告されている(文献3, 4)など)。たとえば、文献3)では、発音時刻やコード変化、ドラムパターンなどを手がかりとしたビートトラッキングのモデルが構築され、並列計算機上に実装されている。構築されたモデルは、認識率やトラッキングの開始時刻、正確さなどの諸点から評価され、高い精度をもつシステムであることが証明されている。これらの研究は、人間の拍節構造の知覚過程をモデル化したものではないが、主にボトムアップ的に音列を処理することによって、人間が認識するのと同様の解釈を得ることを目的としたものとして、非常に興味深いといえることができる。

一方、認知心理学では、人間の拍節構造知覚過程そのものをモデル化し、その処理の特徴を解明しようとする研究が続けられている。人間が物理的な音の系列を“旋律(melody)”として認識できるためには、“スキーマ(内的な知識や処理の枠組みのこと)”に合致するような処理を行う必要があるという考え方

がある^{5),6)}。たとえば西洋音楽に慣れ親しんだ聞き手は西洋音楽に関するスキーマを持っており、そのスキーマに適合するように音列を知覚的に体制化することによってメロディを認知すると考えるのである。

このような観点に基づき、音楽の時間構造の処理について心理学的な面からの考察が進められている⁷⁾⁻¹⁰⁾。具体的には、人間はどのようにして入力音列を分節化して拍を知覚し、また、どのようにしてそれらを構造化して拍子を知覚するのかということ、つまり、こうした“拍節的体制化(metrical organization)”の過程の性質を明らかにすることを目的として、実験研究とモデル研究を相補的・循環的に行いながらそのアルゴリズムの基本原理が模索されている¹¹⁾⁻¹³⁾。

1.3 拍節構造知覚過程に関する既存の心理学的モデルの問題点

これまでに、拍節構造の知覚過程に関する心理学的なモデルがいくつか提案されてきている¹⁴⁾⁻¹⁸⁾。これらのモデルは、それぞれが独自の特長をもっているが、いずれも、人間が拍節構造を知覚する過程を統合的に説明するためのアルゴリズムを、明示的に定式化しようとしている点において一定の評価をすることができる。

しかしながら、これらのモデルは全て“テンポ”を考慮していないという本質的な問題点がある。多くの先行研究で指摘されているように、人間がメロディの拍節構造を知覚する場合には、そのテンポに大きな影響を受ける^{7),19),20)}。すなわち同じ音列を聴取したとしても、そのテンポが異なると拍として知覚する時間単位も変化する場合があるため、それにつれて拍より上位の階層の時間単位(たとえば小節)の知覚の仕方も変わってくるというケースもあるのである。既存の拍節構造知覚過程のモデルでこの点について考慮されているものはなく、心理学的妥当性は低かった。

人間の拍節的体制化の過程をモデル化するためには、聞き手の知覚に観察されるテンポの影響を明確に反映させる必要がある。

2. テンポの影響を考慮に入れた拍節構造知覚過程のモデルの構築

2.1 Longuet-Higgings and Lee (1982)によるモデル

これまでに提案されているモデルの中で、Longuet-Higgings and Lee (1982) (以下“LHL”と略記する) のモデルは、人間の拍節構造の知覚過程をうまく模しているという点で高く評価されている。このモデルは、それ以前の理論的研究^{19),20)}では必ずしも的確に実現されなかった人間の拍節構造知覚過程を、忠実にモデル化しようとしたものであった。このモデルのもっとも基本的な特徴は、音価の大きな音が群化の開始点になるような解釈を、一音入力される度に出力していく、という点にある。異なったレベルの群化をうまく説明できないなど改善すべき点も少なくはないが、その後の理論的研究に大きな影響を与えており、現在までに提案されている有力なモデルの多くがこのモデルの考え方を基盤にしている。本論文で提案するモデルも、基本的なアルゴリズムとしてこのモデルのそれを採用し、妥当性の低い部分について改善を図っている。

そこで、以下ではまずLHLのモデルのアルゴリズムを詳述し、その後、新たに構築するモデルに実装した改善点を説明する。

2.1.1 基本的な仮定

LHLのモデルでは、「人間は最初の2音を聞いた時点で心内に第1拍と第2拍を知覚し、その時間間隔(“ユニット”)に基づいて次の拍の発生を予測する」ということを大前提としている。そして、その予測の正誤に従って、随時解釈を確定し、あるいは変更してゆくとする。

このモデルのもっとも大きな特徴の一つとして、“処理の漸進性”をあげることができる。人間のメロディの知覚は原則として時間的な流れに制約を受けており、音列の進行につれて大きな時間的遅延なしにそのリズムを知覚していくという大きな特徴がある²²⁾。このモデルでは人間のそのような聴取の特徴が忠実にモデル化されている。

第2の特徴として、相対的に音価の大きな音が、より高次の拍節構造の開始点となるように処理を行う点を指摘することができる。精神物理学の実験によって、相対的に音価の大きな音は知覚的により顕著であることがわかっている(たとえば、文献23)など)。このモデルでは、人間のこのような知覚的傾向を処理に反映したものとなっている。

さらに第3の特徴として、このモデルはある階層の時間単位を認定すると、その時間単位を「2倍化」することによってより高次の階層の時間単位を設定していく、という点を指摘することができる。これは、人間の拍節構造知覚の特徴である「2倍型の拍節構造知覚への偏好性(ある階層の時間単位とその一つ上の階層の時間単位とが、長さの点でべき乗倍の関係になるように知覚する傾向があること^{19),24)})」を意識したものである。

2.1.2 処理の流れ

モデルが起動すると、第1音に記号“ t_1 ”を、第2音に記号“ t_2 ”を、それぞれ設定する。また、 t_2 より後に、 $t_2 - t_1$ となる時点に記号“ t_3 ”を設定する。これによって、聞き手が予測する次の拍の生起時点を表現する。実際の処理では、「予測の正誤」、つまり音列上で予測時点(t_3)に実際に音が発生するかどうか、を評価することによって処理が進行する。モデルのアルゴリズムをプロダクショナル的に記述すると以下ようになる。

(1) 第1音に t_1 、第2音に t_2 をそれぞれ設

定した後、それらの時間間隔 ($t_2 - t_1$) を算定し、“ユニット”とする。

(2) 第2音以後に、そのユニット分の時点(t_3)で音が生起することを予測する。

(a) 予測時点 (t_3) に音の onset が存在する場合、“CONFLATE”と呼ばれる処理を行う。

(b) t_2 と t_3 の間に、 t_2 を設定した音よりも音価の大きな音が出現した場合、“STRETCH”と呼ばれる処理を行う。

(c) t_2 から始まる音が t_3 を越え、次音が登場しない場合、“UPDATE”と呼ばれる処理を行う。

(i) UPDATE適用後、 t_1 から t_3 の間に音の onset が全く存在しない場合、“LONGNOTE”と呼ばれる処理を行う。

(3) (2)に戻り処理を繰り返す。どの処理も行われなくなった場合に処理を終了する。

“CONFLATE”、“STRETCH”、“UPDATE”および“LONGNOTE”と命名された処理の具体的な内容は以下の通りである。

• CONFLATE

予測時点に実際に音が発生した場合に適用される処理である (図 1)。 t_3 を予測した時点で音の onset が存在する場合、 t_2 を t_3 に移動し、当該のユニットを2倍化する。これは、拍の長さが2倍に変化することを意味し、より高次の階層のユニットの認定へとつながる。



図1 CONFLATEルーチンの働き

• STRETCH

予測時点よりも実際の音の発生時点が早い場合に適用される処理である (図 2)。予測時点 t_3 よりも早く次音の onset があり、かつ t_2 を設定した音よりも音価が大きい場合に適用される。 t_2 を次の音の onset に移動させ、ユニットを再計算して予測時点 t_3 を設定し直す。これにより当該のユニットが伸長化される。

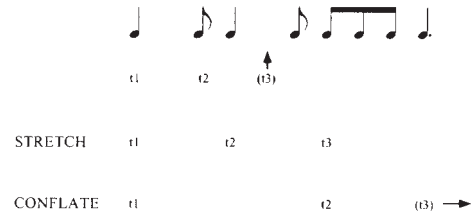


図2 STRETCHルーチンの働き

• UPDATE

予測時点よりも実際の音の発生時点が遅い場合に適用される処理である (図 3)。予測時点を越えても音の onset がない場合、 t_2 を設定してあった音に t_1 を移動させる。当該ユニットの開始時点が変化することになり、拍の開始時点の変更を意味する。ユニットの大きさは変化しない。

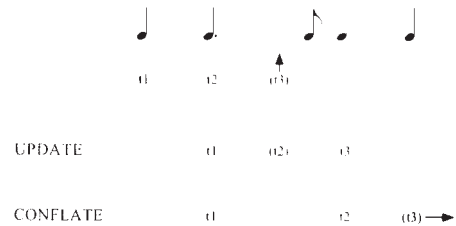


図3 UPDATEルーチンの働き

- LONGNOTE

UPDATE を適用した後にのみ適用される処理である (図 4)。この処理は、UPDATE を適用した後、 $t_1 \sim t_3$ に音が全く存在しない場合、 t_2 を次の音に移動させる働きをする。

表1 新たに構築するモデルとLHLとの違い

	新たに構築するモデル	LHL	
改良した点	階層化の上限	上限あり	上限なし
	複層性	複層	単層
	テンポの影響の考慮	あり	なし
維持した点	処理形式	漸進的処理	
	拍節的単位の開始時点	相対的に音価の大きな音	
	拍節構造の偏好性	2倍型への偏好	

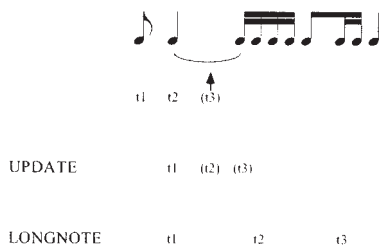


図4 LONGNOTEルーチンの働き

モデルでは、これらの処理を適宜作動させながら、一音ずつ音列を処理し、最終音まで達した段階で処理が終了する。

2.2 新たに構築するモデルに加えられた心理学的制約

2.2.1 LHL のモデルとの違い

LHL の提案は、人間の拍節構造知覚過程の基本的な部分をモデル化しているが、一方でいくつかの本質的な問題点がある。新しく構築するモデルは、以下の各点についてLHL のモデルに制約を加えて改善した。

制約 1 : 出力時間単位の階層の制限

LHL のモデルでは、ユニットとして認定する時間単位は、原則として無限に階層化できることになっている。一方、人間が知覚できる拍節的な時間単位には上限がある。人間の拍節的体制化の過程をモデル化するためにはこういった制約を設けなければならない。

制約 2 : 時間単位の“複層性”

LHL のモデルでは、出力される時間

単位のうち、どの時間単位を拍とし、どの時間単位を小節とするかといった明確な規定がない。これは、拍子の概念に相当する単位も出力することができないことにもつながっている。一方、人間は拍や小節などといったものを異なったレベルの時間単位として知覚することができ、それらを階層化して複層的に知覚している。人間のモデルとしては、こういった複層性を反映している必要がある。

制約 3 : 拍として認定する時間単位の制限

LHL のモデルでは音価の比率だけを対象に処理を進めるために、テンポの影響が全く反映されないことになる。前述のように、人間は、同じ音列でもテンポが異なると知覚結果が異なることがある。人間のモデルとしては、こういったテンポによる知覚の変化についても説明できる必要がある。そのためには、人間の知覚できる拍の時間的長さについて、処理アルゴリズムの中に明示的に組み込む必要がある。

一方、LHL で定式化されている“漸進的処理”や“拍節単位の開始時点”、“(3倍型解釈に対する)2倍型解釈への偏好性”については、心理学的に妥当性が高いと考えてそのまま引き継いだ(表1参照)。

2.2.2 聞き手が知覚する時間単位—“tactus” と“メトリカル・ユニット”の定義

前節の制約を充足するモデルを構築するためには、聞き手が知覚する時間単位を特定しておく必要がある。ここでは、“tactus”および“メトリカル・ユニット”という用語を定義する。

拍とは、一般的には、聞き手が知覚する、もともと顕著な一定周期の時間単位のひとつのことを言い、心理物理学の実験研究によれば、実時間にして約 250 ms ~ 1500 ms 程度を中心に分布することがわかっている^{19),20),23),25)}。しかし、拍子もしくは小節などといった、より高次の階層の時間単位についてもまた、同じ拍という語によって表現されることがある。本論文で構築するモデルでは、これらの混乱を避けるために、聞き手が知覚するある実時間の範囲に収まる時間単位を、文献2)にならって“tactus”という語を用いることとする。

一方、“メトリカル・ユニット”という用語を用いて、周期的な一定の長さを持つ時間の単位を表現することとする。つまり、聞き手が知覚する *tactus* や拍子などといった時間単位全般をメトリカル・ユニットと総称する。このことは、モデルが音列を処理した結果、メトリカル・ユニットを認定することができれば、それが聞き手の知覚した拍節構造を特定したことになると仮定していることになる。また、構築するモデルでは、これらのメトリカル・ユニットを階層的に積み上げていくことで、より上位の拍節単位を特定する。これは、複数の拍節単位の間接関係を特定することになり、結果として拍子に当たる概念を出力できるということを意味する。

2.3 モデルのアルゴリズム

以上のような定義に基づき、構築したモデルのアルゴリズムは以下のような処理手順で記述することができる。

2.3.1 処理の流れ

モデルは、音列の構成音の音価比を整数比で入力することによって処理を開始する。

- (1) 音列の第1音の onset を認定した段階で、その音の onset にマーク “P1(Position of the 1st tone)” を置く。
- (2) 音列の第2音の onset を認定した段階で、その音の onset にマーク “P2(Position of the 2nd tone)” を置く。
- (3) 第1音と第2音の onset 間時間間隔 ($P2 - P1$)、すなわち、第1音の長さを計算し、この時間間隔を “ユニット(U)” の長さとする。
- (4) 時間軸上で、P2からユニットの長さ分の群化を予測し、マーク “Pe(Position of the expected tone)” を置く。
 - (a) P2 と次の拍節開始の予測時点 Pe との間に、P2 の位置に onset がある音より音価の大きな音の onset がない場合は、その時点のユニットを “メトリカル・ユニット(M.U.)” とする。
 - (i) 初めてメトリカル・ユニットを認定した場合には、そのメトリカル・ユニットが *tactus* として適切かどうかを評価する。*tactus* の評価は以下の式によって行われる。

$$tactus = \begin{cases} 2M.U. & (M.U._{1st} < T_{mini}) \\ M.U. & (T_{mini} \leq M.U._{1st} \leq T_{max}) \\ M.U./2 & (T_{max} < M.U._{1st}) \end{cases}$$

ここで、 T_{max} は人間が認定する *tactus* の最大値、 T_{mini} は最小値をそれぞれ表し、以下の式によって計算される。

$$T_{max} = \frac{tempo}{(length_{the-quarter-note})} k_1 \quad (1)$$

$$T_{mini} = \frac{tempo}{(length_{the-quarter-note})} k_2 \quad (2)$$

ここで, k_1 , k_2 は,

$$k_1 = \frac{1}{15}: (\text{定数}) \quad (3)$$

$$k_2 = \frac{1}{60}: (\text{定数}) \quad (4)$$

に固定されている^{*2}。なお
tempo ~ 音列の“テンポ”

T_{max} ~ tactusとして知覚できる範囲の
最大値

T_{mini} ~ tactusとして知覚できる範囲の
最少値

を示す。

- (b) P2 と次の群化開始の予測時点 Pe との間に, P2 の位置に onset がある音より音価の大きな音の onset がある場合は, その onset に P2 を移し, ユニットの長さを伸長させて, 新しい群化開始時点 Pe を予測する。
- (c) これらの処理を, メトリカル・ユニットが確立するまで繰り返す。
- (5) ある時点のメトリカル・ユニットよりも長い音の onset が, 群化開始予測時点, すなわち, 時間軸上でのメトリカル・ユニットの整数倍の位置にあるならば, その音の上に P1 を移動し, 群化開始時点の位相を変える。また, その位相変化とともに, P2 を時間軸上を遡らない, もっとも近い群化開始予測時点, すなわち, 時間軸上でのメトリカル・ユニットの整数倍の位置に置く。
- (6) 次音を読み込み, 再び(4)に戻り同様の処理を繰り返す。
- (7) (4) の処理をその必要がなくなるまで繰り返す。そして, その時点でのメトリカル・ユニットを最高次のユニットとして認定する。

2.3.2 音列のシミュレーション例

図5はモデルによるシミュレーション例を示している。

音列の第1音目, 第2音目が入力されるとそれぞれの音の onset 上に P1, P2 を置き, $(P2 - P1)$ を計算して“ユニット”として仮定する。そして, その後時間軸上で, P2 からユニット分の長さの群化を位置 Pe に予想する。P2 と Pe の間にはマーク P2 上にある音よりも音価の大きい音 (音符番号: 3) があるため, そこに P2 を移動し, 新たに Pe の位置を計算し直す (第4音段階)。今度は P2 と Pe の間には P2 よりも音価の大きい音は存在しない。ここで $P1 - P2$ ($P2 - Pe$) はこの時点で“メトリカル・ユニット”として確立される。このメトリカル・ユニットは初めて認定されたものであるので, tactus となり得るかどうかを評価される。この場合は, $P1 - P2$ の時間的大きさが tactus として認定される。

次に, この時点でのメトリカル・ユニット (すなわち, 4分音符メトリカル・ユニット) よりも大きい音 (音符番号: 3) が拍節構造の開始予測時点, つまり時間軸上でメトリカル・ユニットの整数倍の時点にあるので, その音の上に P1 を移動し, 拍節構造の開始時点の位相を変えるとともに, P2 をもっとも近い拍節構造の開始予測時点 (音符番号: 5) に置き, 再び $(P2 - P1)$ を計算して Pe を予測する。この時, P2 と Pe の間に P2 よりも音価の大きい音 (音符番号: 7) が存在するので, そこに P2 を移動し, 再度 Pe の位置を計算する。今度は P2 と Pe の間には P2 よりも音価の大きい音は存在せず, ここで“付点2分音符メトリカル・ユニット”が確立する。残りの音列には“付点2分音符ユニット”よりも音価の大きい音は存在せず, これ以上高次のメトリカル・ユニットを推論しない。最終的な音列の解釈は, 図の最下段に示したように, 3音目に第1拍を置く, 3/4拍子となる。

1 2 3 4 5 6 7 8 9 10

処理音列

(1) 第3音段階
P1 P2 Pc

(2) 第4音段階
P1 P2 Pc metrical unit: 4

(3) 第5音段階
P1 P2 Pc metrical unit: 4

(4) 第6音段階
P1 P2 Pc metrical unit: 4

(5) 第7音段階
P1 P2 Pc metrical unit: 4

(6) 第8音段階
P1 P2 Pc metrical unit: 4

(7) 第9音段階
P1 P2 Pc metrical unit: 4
12

(8) 第10音段階
P1 P2 Pc metrical unit: 4
12

最終的な解釈 3/4

図5 構築したモデルによるシミュレーション例。

3. モデルの評価

3.1 拍節構造の知覚過程に関する認知心理学的実験

著者は人間の拍節構造の知覚過程について心理学的実験を行った¹⁹⁾。実験の目的は、拍節的体制化の特徴を明らかにし、人間が拍や拍子をどのように認定するのかを調べることであった。

3.1.1 方法

3.1.1.1 被験者

被験者は20人であった。被験者の平均音楽経験年数は16.8年で、いずれも西洋音楽に熟達していた。全員拍節構造の概念をよく理解しており、記譜法にも習熟していた。

3.1.1.2 材料

材料は、6音～17音から成る30種類の音列をもとにして作成した、262種類の音列パターンであった。30種類の音列のうち、26種類の音列は既存の楽曲の主旋律から小節単位で採つ



図6 実験における材料提示の流れ。

て加工したものであった。これらの音列の拍子は特定の拍節構造のものに偏らないように配慮した（原曲が4/4拍子の音列8種類，3/4拍子の音列8種類，6/8拍子の音列8種類，“変拍子の音列”2種類）。残りの4種類の音列は実験者が作成したものであった（等音価の音列3種類，ランダム音列1種類）。表2内の“実験材料”に，使用した材料を音価比（1=16分音符）で示す。

音列パターンは以下の手順で作成した。まず“音価のみ”の音列を作成した。これは個々の音符の音価のみを残したものであり，それ以外の全ての物理的特徴（たとえば，音高や強さなど）は全て均一になっていた。次に，それら30種類の音列を，1音目から3音分までの音列パターン（第3音段階），1音目から4音分までの音列パターン（第4音段階），1音目から5音分までの音列パターン（第5音段階）...，というように分解し，一音ずつ最終音まで加えていくことによって，実験で使用する音列パターン合計262個を作成した。全ての音列の構成音の周波数は440Hz，音色はピアノ音，レベルは被験者の快適聴取レベルとした。音列は4分音符 = 500ms の速度で提示した。

3.1.1.3 手続き

実験は一人ずつ防音室内で行なった。材料は被験者ごとにランダムに提示した。被験者には，各音列について，第3音段階の音列パターン，第4音段階の音列パターン，第5音段階の音列パターン...，というように1音ずつ増加させた音列パターンを最終音段階まで順次提示し，課題を行なわせた（図6）。

被験者の課題は，それぞれの音段階の音列パターンに対する解釈を記譜し，その解釈に対する確信度評定を行なうことであった。ここで“記譜”とは「リズム譜」を書くこと，すなわち，「音価」，「拍子記号」および「小節線」を記すことであり，“確信度評定”は，記譜した音価，拍子記号および小節線それぞれに対する自身の反応の確信度を9段階（1：「全く自信がない」～9：「非常に自信がある」）で評定することであった。第3音段階から最終音段階まで課題を行なうことを1試行とした。各試行の間には，直前の試行がその直後の試行に与える影響を最小限に抑えるために，歌謡曲やジャズ，ロックなど様々なジャンルから引用した楽曲を約30秒提示して印象評定を行なわせた。

表2 被験者の解釈とモデルの出力

SEQ	実験材料 (16 分音符= “1”)	被験者		モデル				
		拍子	第 1 拍	新モデル		LHL		
				拍	拍子	第 1 拍	ユニット	第 1 拍
01	2-2-2-10-2-2-2-10-2-2-2	1/4	1 音目	4	4/4	4 音目	12	4 音目
02	4-10-1-1-2-1-1-4-4	1/4	1 音目	4	-	2 音目	12	2 音目
03	2-2-2-2-1-1-6-1-1-2-2-2-1-1-6-1-1	1/4	2 音目	4	4/4	1 音目	16	1 音目
04	2-2-2-8-4-3-1-12	4/4	1 音目	4	4/4	4 音目	12	4 音目
05	4-4-12-2-2-4-2-2-4-2-2-4-4	4/4	3 音目	4	4/4	3 音目	10	3 音目
06	8-1-1-6-1-1-2-2-2-8	4/4	1 音目	4	-	1 音目	8	1 音目
07	8-4-4-6-2-8-8-4-2-2-8	1/4	1 音目	4	4/4	1 音目	32	1 音目
08	6-2-6-2-6-1-1-6-2	4/4	1 音目	4	-	1 音目	6	5 音目
09	4-6-2-4-4-4-8-2-2-2-2	3/4	1 音目	4	2/4	1 音目	14	2 音目
10	4-8-4-4-8-4-12-4-4	4/4	1 音目	4	2/4	1 音目	16	5 音目
11	2-2-2-4-6-2-6-2-2-2	4/4	1 音目	4	(2,4,6)	3 音目	6	3 音目
12	4-6-2-2-2-4-4-4-4-2-2-3-1	3/4	2 音目	4	3/4	2 音目	16	2 音目
13	6-1-1-1-1-1-2-1-1-6-1-1	4/4	1 音目	3	6/8	1 音目	12	1 音目
14	2-2-2-2-4-16-4-4-4-2-2-4	4/4	1 音目	4	(2,4)	6 音目	12	6 音目
15	12-6-2-2-2-12-6-2-2-2	3/4	1 音目	6	(6,12)	1 音目	12	1 音目
16	4-3-1-1-1-1-1-2-2-2-2	4/4	1 音目	4	4/4	1 音目	8	1 音目
17	2-2-3-1-4-2-2-3-1-4-2-2	3/4	1 音目	4	2/4	4 音目	8	1 音目
18	2-4-2-8-2-12	4/4	1 音目	4	4/4	1 音目	6	4 音目
19	2-2-2-2-2-2-2-2-1-1-2-2	2/4	1 音目	4	-	1 音目	8	1 音目
20	2-2-2-4-3-1-2-6	4/4	1 音目	4	4/4	1 音目	4	1 音目
21	6-2-2-2-6-8-2-2-2-2-2	6/8	1 音目	3	6/8	1 音目	6	1 音目
22	8-2-1-1-7-1-1-1-1-1	3/4	1 音目	4	2/4	1 音目	12	1 音目
23	3-1-2-4-2-3-1-2-4-2	6/8	1 音目	3	6/8	1 音目	6	1 音目
24	6-4-2-12-2-2-2-4-2-12	6/8	1 音目	3	6/8	1 音目	12	1 音目
25	8-4-4-4-4-8-4-4-4-8	3/4	1 音目	4	2/4	1 音目	28	1 音目
26	4-2-1-1-4-2-1-1-4-2-2-4	2/4	1 音目	4	-	1 音目	8	1 音目
27	4-4-4-4-4-4-4-4-4-4-4	4/4	1 音目	4	-	1 音目	16	1 音目
28	2-2-2-2-2-2-2-2-2-2-2	4/4	1 音目	4	(2,4)	1 音目	8	1 音目
29	1-1-1-1-1-1-1-1-1-1-1	4/4	1 音目	4	(1,4)	1 音目	4	1 音目
30	12-6-3-12-4-8-2-3-6-1	-	-	4	(6,12)	1 音目	-	1 音目

3.2 実験結果とシミュレーション結果との比較

文献 2)によれば、メロディの拍節単位は、「拍」、「小節」および「第 1 拍の位置 (第 1 拍を何音目に認定したか)」を決定することによって特定化される。被験者の記譜結果について、各音列音段階ごとに、「拍」、「拍子」(拍子は“小節”に相当する時間単位と拍と関係から決定される) および「第 1 拍の位置」についてそれぞれ分析した。その結果、被験者の回答はおおよそ一意に収束することが明らかになった¹⁹⁾。そこで、もっとも多かった被験者の解釈結果と、モデルのシミュレーション結果を比較した。

表 2 に、被験者の回答および構築したモデルと LHL のモデルで行ったシミュレーション結果を示す。構築したモデルと被験者の最

終的な解釈結果とが、拍、拍子および第 1 拍の位置全てにおいて完全に予測できたのは、全 30 音列中 9 音列であった。以下では、数量的な観点および質的な観点からモデルの妥当性を考察する。

3.2.1 数量的分析—拍、拍子および第 1 拍の位置についての予測結果

「拍」と「第 1 拍」の位置についての構築したモデルによる一致率を図 7 に示す。図 7 には LHL のモデルの一致率も併せて示してある。

・拍

被験者については、各音段階ごとにもっとも多かった拍子解釈から推測される時間的大きさを「拍」とした。たとえば、被験

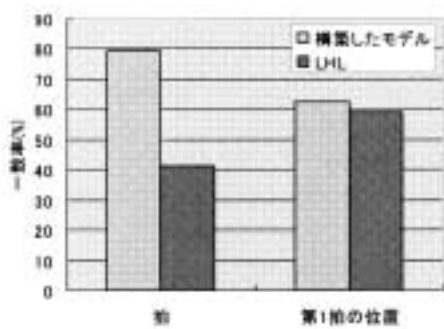


図7 構築したモデルおよびLHLのモデルと実験結果との一致率。

者の最終的な拍子解釈が「4/4拍子」の場合には、「4分音符1つ分の大きさ」を拍とした^{*4}。一方、構築したモデルについては、tactusとして出力されたメトリカル・ユニットを拍とした。LHLのモデルは、出力されたユニットそのものを拍とみなして計算した。

図7を見ると、LHLのモデルと比較して、構築したモデルは拍についてより高い予測力をもっていることがわかる。音列ごとに、両モデルの各音段階の正答率を平均した値を計算した結果、構築したモデルの正答率の方が有意に高い値を示した ($t(58)=2.96, p < .01$)。これは、構築したモデルはテンポの影響を考慮しており、テンポから計算したある一定の時間単位をtactusとして認定するというアルゴリズムが有効に働いたからだと考えられる。一方、LHLのモデルはテンポの影響を全く考慮しておらず、その結果構築したモデルと比較して予測率が低くなったと考えられる。

・拍子

構築したモデルは、あるメトリカル・ユニットをtactusとして特定化する。tactusとして認定されたメトリカル・ユニットと、それ以外のメトリカル・ユニットのうちの最大のものとの大きさの関係によって、その音列を何拍子として解釈したかを出力す

ることになる。

構築したモデルは、全30音列中10音列について、被験者が認定した拍子と同じ拍子を出力した。一方、LHLのモデルについてはtactusを特定できず、したがって、拍子も出力できなかった。

・第1拍の位置

第1拍を何音目に認定するか、という点について、被験者の解釈と構築したモデルおよびLHLのモデルの予測の一致率を図7に示す。ここで的一致率とは、各音段階での一致率を平均したものである。

図7を見ると、構築したモデル、LHLのモデルともに約65%前後の予測率となっており、両者の間には顕著な違いは観察されなかった。音列ごとに、両モデルの各音段階の正答率を平均した値を計算した結果、両者の間の差は有意傾向であった

($t(58)=1.64, p = .1$)。

3.2.2 質的分析—漸進的処理の精度に関する考察

人間の漸進的な知覚過程と、モデルの漸進的な処理過程とを、各音段階ごとに比較した。以下では、いくつかの音列の具体例を示しながらモデルのアルゴリズム上の問題点を考察する。

3.2.2.1 拍と拍子の認定

図8および図9に示す音列は、モデルの処理過程と被験者の知覚過程がほぼ一致した例である。図中には、被験者の反応として拍子(上段の数値)と第1拍の位置(下段のカッコ内の数値)、モデルの予測としてメトリカル・ユニットの大きさ(上段の数値)と第1拍の位置(下段のカッコ内の数値)が、それぞれ示されている。

図8の音列(SEQ. 05)では、被験者は最終的に第3音に第1拍をおき、4/4拍子として解釈している。一方、モデルは、長さが4および16の2つの階層のメトリカル・ユニッ

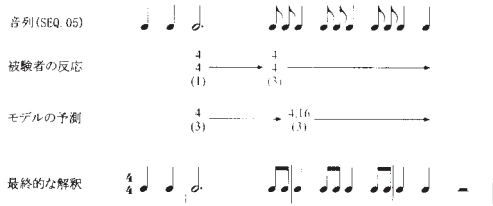


図8 モデルが聞き手の解釈を適切に予測することのできた音列の例・1。

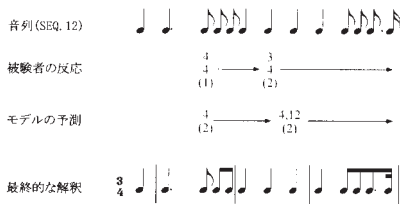


図9 モデルが聞き手の解釈を適切に予測することのできた音列の例・2。

トを出力している（これは4/4拍子を表している）。最終的に、第1拍を第3音においた4/4拍子として解釈したことになる。図9の音列（SEQ. 12）では、被験者は当初第1音に第1拍をおいた4/4解釈をしているが、音列の進行につれて解釈を変更し、最終的には第11拍の位置を第2音に移動して拍子解釈も3/4拍子に変更している。一方モデルは、長さが4および12の2つの階層のメトリカル・ユニットを出力している（これは3/4拍子を表している）。最終的に、第1拍を第2音においた3/4拍子として解釈したことになり、被験者の最終的な解釈と一致している。解釈のタイミングにはわずかなズレがあるが、総じて被験者の知覚過程をうまくシミュレートできていたと言える。

一方、図10に示す音列は、拍子の認定に失敗した例である（SEQ.09）。この音列に対し、人間は4/4拍子として解釈を進めるものの、途中（第6音段階）で3/4拍子に解釈を変更し、最終音までその解釈を維持する。これに対し、モデルでは、認定する tactus は正しいものの、“2分音符”分の長さを最大のメト

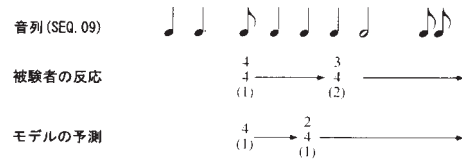


図10 拍子の認定に失敗した音列の例。

リカル・ユニットとして認定するために、結果として2/4拍子解釈を出力することになってしまう。この音列では、時間軸上でモデルが音の発生を予測する時点(Pe)には必ず音の onset がある。モデルの基本的なアルゴリズムとして予測時点における音の onset の有無を判定基準として採用しているため、解釈の変更が行われないのである。

3.2.2.2 第1拍の位置の認定

聞き手の解釈とモデルの予測が一致しなかった音列の例を図11に示す（SEQ.01）。



図11 第1拍の位置の認定に失敗した音列の例。

この音列に対して、聞き手の70%が第1音に第1拍を置いて解釈したのに対し、モデルは第4音目が入力されるとその音に第1拍を置き、冒頭の3音はアウフタクトとして解釈した。

この音列をうまく予測できなかった理由として、聞き手とモデルとでは相対的に音価の大きい音の処理の仕方が異なっているということが考えられる。モデルは相対的に音価の大きい音が入力されると、即その音から高次の拍節単位を開始しようとするのに対し、聞き手はそのような音を聴取してもすぐには解釈を変更しない。このような、相対的に音価の大きな音に対する、いわば“敏感さ”の違いによって、第1拍の位置をうまく予測できない音列があるのであろう。

4. まとめ

本論で構築したモデルは、テンポの影響を考慮しているという点について、既存のモデルと比較して心理学的妥当性が高いと言える。その他にも、拍や小節、拍子などといった、聞き手が知覚する複数の時間単位について、別々にかつ具体的に予測することができていた点、漸進的処理様式のアロリズムを想定し、音列の進行とともに解釈を生成していくという点、などについては、人間の拍節的体制化の過程のもっとも基本的な部分を適切にモデル化していると言える。

一方、いくつかの問題点も明らかになった。第1に、いくつかの音列では単層のメトリカル・ユニットしか認定できない場合があった。ひとたび確定したメトリカル・ユニットが *tactus* として認定できる時間範囲内に入っており、かつ、その後の音列のなかの音符の長さの変化がない場合には、認定した以上の時間単位を構築することができない。単層のメトリカル・ユニットしか認定できないということは、“拍子”の概念に相当する単位を出力しないということを意味する。つまり、人間がこのような音列に対して拍よりも高次の拍節単位を知覚できるということを説明できていないことになる。

第2に、音価の大きな音に対して“過敏”な場合があった。構築したモデルは、相対的に音価の大きな音が出現すると、即座に処理を開始し、より高次の拍節構造を開始しようとする。しかし、人間はそれまでの解釈を維持しようとする傾向があったり、その他の要因（たとえば、同型の繰り返しパターンなど）の影響で一定時間の後に解釈を変更するという場合があったりする。構築したモデルはこの点の見究めが不十分であり、改善の余地が残る。

第3に、構築したモデルは、ある実時間的範囲内に収容できないメトリカル・ユニット

を2倍化もしくは1/2倍化するという振舞いをする。これは、聞き手の2倍型解釈への偏好性を意識したものであり、実験結果をよく説明することができていた。しかし、聞き手は、たとえば3倍型の解釈で拍節構造を認知することもできるが、現在のアルゴリズムではそのようなパフォーマンスを十分に説明できない。この点については、モデルのアルゴリズムの本質的な問題であるため解決は容易ではないが、現在のモデルに反映されている2倍型解釈への偏好性を基本にしつつ、その他の解釈可能性をも説明できるように改良する必要がある。

今後は、こういった問題点を克服すると同時に、人間の持つさらなる認知機能（たとえば、音列に対する顕在・潜在記憶^{27),28)}や注意など）をも考慮に入れた計算論的モデルの構築を行っていく必要がある。

[参考文献]

- 1) 朝香淳(編): 標準音楽事典, 音楽之友社 (1992).
- 2) Lerdahl, F. and Jackendoff, R.: *A Generative Theory of Tonal Music*, MIT Press, Massachusetts (1983).
- 3) 後藤真孝: 音楽音響信号を対象としたリアルタイムビートトラッキングに関する研究, 博士論文, 早稲田大学大学院理工学研究科 (1998).
- 4) 後藤真孝, 村岡洋一: 音響信号を対象したリアルタイムビートトラッキングシステム - コード変化検出による打楽器音を含まない音楽への対応 -, 電子情報通信学会論文誌D-II, Vol. J81-D-II, pp. 227-237 (1998).
- 5) 阿部純一: 旋律はいかに処理されるか, 音楽と認知 (波多野諺余夫(編)), 東大出版会, chapter 2 (1987).
- 6) 阿部純一: メロディの知覚的体制化の過程, 日本認知科学会第12回大会発表論文集, pp. 16-19 (1995).
- 7) 阿部純一: リズム的体制化の過程 - そのモデ

- ル化への試論 - , *Hokkaido Behavioral Science Report, Series*, Vol.66 (1993). 北海道大学文学部行動科学科.
- 8) 阿部純一: リズム的体制化の過程: そのモデル化についての考察, 音楽認知における“体制化”の過程. 平成4年度科学研究費補助金(一般研究c時限)研究成果報告書, pp. 1-16 (1992).
- 9) Goto, Y.: Implicit memory for rhythm. An effect of pitch height and timbre for repetition priming of musical tone sequence., *Proceedings of 7th International Conference of Music Cognition and Perception*, in CD-ROM (2002).
- 10) 後藤靖宏: 異なる拍子のメロディに対する拍節構造解釈の変化 - 音楽非熟達者の拍節構造知覚過程からの考察 -, 音楽心理学音楽療法研究年報, Vol. 28, pp. 13-22 (1999).
- 11) Goto, Y. and Abe, J.: On cognitive models of rhythmic interpretation, *International Journal of Psychology*, p. 51 (1996).
- 12) 後藤靖宏, 阿部純一: “リズム”認知過程のモデル化: “メトリカル・ユニット階層化モデル”と“内的クロック生成モデル”の比較と今後の方向性, 日本認知科学会第12回大会発表論文集, pp. 152-153 (1995).
- 13) 後藤靖宏, 阿部純一: 拍と拍子の認知過程・2 - 実験とシミュレーション及び複層性の導入 -, 日本心理学会第59回大会発表論文集, p. 625 (1995).
- 14) Steedman, M.J.: The perception of musical rhythm and meter, *Perception*, Vol. 6, pp. 555-569 (1977).
- 15) Longuet-Higgins, H.C. and Lee, C.S.: The perception of musical rhythms, *Perception*, Vol. 11, pp. 115-128 (1982).
- 16) Longuet-Higgins, H.C. and Lee, C.S.: The rhythmic interpretation of monophonic music, *Music Perception*, Vol. 1, pp. 424-441 (1984).
- 17) Lee, C.S.: The rhythmic interpretation of simple musical sequence: Towards a perceptual model, *Musical Structure and Cognition* (Howel, P., Cross, I. and West, R.(eds.)), Academic Press, London, pp. 53-69 (1985).
- 18) Povel, D.J. and Essens, P.: Perception of temporal patterns, *Music Perception*, Vol. 2, No. 4, pp. 411-440 (1985).
- 19) 後藤靖宏, 阿部純一: 拍子解釈の基本的嗜好性と漸進的確立, 音楽知覚認知研究, Vol. 2, pp. 38-47 (1996).
- 20) Lee, C.S.: The perception of metrical structure: Experimental evidence and a model, *Representing Musical Structure* (West, R., Howell, P. and Cross, I.(eds.)), Academic Press, London, pp. 59-128 (1991).
- 21) Simon, H.A. and Sumner, R.K.: Pattern in music, *Formal representation of human judgment* (Kleinmuntz(ed.)), Academic Press, New York (1968).
- 22) 平賀讓: 音楽認知のための知識表現, 音楽と認知(波多野誼余夫(編)), 東大出版会, chapter4 (1987).
- 23) Woodrow, H.: A quantitative study of music, *Archives of Psychology* 14 (Woodworth, R.S.(ed.)), New York: Science Press (1909).
- 24) Drake, C.: Reproduction of musical rhythms by children, adult musicians and adult nonmusicians, *Perception & Psychophysics*, Vol. 1, pp. 25-33 (1993).
- 25) Essens, P.: Hierarchical organization of temporal patterns, *Perception & Psychophysics*, Vol. 2, No. 40, pp. 69-73 (1986).
- 26) 石桁 真礼生 丸田 昭三 金光 威和雄 末吉 保雄 飯田隆, 飯沼信義: 楽典, 音楽之友社 (1965).
- 27) Goto, Y.: Memory for metrical time units of musical rhythm pattern, *International Journal of Psychology*, p. 53 (2000).
- 28) Goto, Y.: Implicit memory for rhythmic tone sequence: A pilot study on perceptual priming for short temporal pattern, *Acoustical Science and Technology*, Vol. 22, No. 3, pp. 219-225 (2001).

[注]

- * 1 本論文で“西洋音楽”と言った場合、バロック、古典派からロマン派までの期間の音楽に限定することにする。西洋音楽の中には、中世の音楽や現代音楽のように、“調性”や“拍節構造”の意識が希薄なものもあるが、そういったものはとりあえずここでは考察の対象外としておく。
- * 2 これは、人間が自然な拍として知覚しやすい時間的長さを375 ms ~ 1500 ms としたことになる。
- * 3 表2に示されている被験者の解釈結果とは、音列を最終音段階まで聴取した時点でのものであり、拍子と第1拍の位置について掲載した。構築したモデルのシミュレーション結果としては、モデルが *tactus* として認定したメトリカル・ユニットを“拍”として、処理の結果出力された複数のメトリカル・ユニットのうち最大のものを拍によって割ることにより計算されるものを“拍子”として、それぞれ掲載した。また、第1拍の位置についても掲載した。LHLのモデルについては、出力したユニットと第1拍の位置について掲載した。なお、構築したモデルの結果に関して、単数のメトリカル・ユニットしか出力しなかった場合は「拍子」の項は「-」、複数のメトリカル・ユニットを出力するが、そこから拍子を推測できない場合には、括弧内に出力したメトリカル・ユニットのすべてを記した。
- * 4 6/8拍子は、1拍として「⁽³⁶⁾(8分音符)3個分」を1拍として計算した。

[Abstract]

**Computational Model of the Process of Musical Rhythm:
A Computer Simulation by the Model Considering the Effect of Tempo and
an Examination of its Psychological Validity**

Yasuhiro GOTO

A computational model of the process of perception of metrical structure was constructed and its algorithm was implemented as a computer program. This model takes into account the influence of "tempo," which was ignored in the previously proposed models. The algorithm of the model reflects the psychological findings that listeners perceive time units in a certain time range, from about 250ms to 1500ms. As a result of this constraint, the model can predict adequately the metrical structure listeners will perceive as "beat" or "meter." The model can be said to be psychologically valid in that it considers the incremental perception process of listeners and the multiple units listeners can perceive. In order to evaluate the validity of the model, computer simulations were performed for the tone sequences that were used in our previous psychological experiment. The results of the simulations showed that the model could predict appropriately which metrical time units listeners perceived, especially in "beat." On the other hand, the model could not predict correctly listeners' interpretation of metrical structure. Some problems, such as the model being more sensitive than listeners to tones that have a relatively high note value, remain to be solved in the future.