

学習に基づくクオンタイズ： 発音時刻の楽譜上の位置の推定

浜中 雅俊 後藤 真孝 麻生 英樹 大津 展之
筑波大学 科学技術振興事業団さきがけ研究21「情報と知」領域 産業技術総合研究所
{m.hamanaka, m.goto, h.asoh, otsu.n}@aist.go.jp

本稿では、ジャムセッションなどで、テンポ一定の伴奏に合わせて演奏された発音時刻から、演奏者が意図する量子化された発音時刻を推定する手法について述べる。従来のビート・リズムの認識に関する手法の多くが拍位置の予測や推定に主眼を置いていたのに対し、本手法では、拍位置が既知という条件下で、発音時刻のゆらぎを取り除く問題を扱う。和音を含むジャムセッションのMIDI演奏をクオンタイズするため、我々は、発音時刻の遷移とゆらぎを隠れマルコフモデルでモデル化する手法を提案する。演奏記録を統計的に学習させて実験した結果、市販のシーケンスソフトウェアの機械的なクオンタイズより性能が良く、モデルが有効に機能したことが示された。

A Learning Based Quantization: Estimating the Position of Onset Time in a Score

Masatoshi Hamanaka Masataka Goto Hideki Asoh Nobuyuki Otsu
Univ. of Tsukuba “Information and Human Activity”, PRESTO, JST
National Institute of Advanced Industrial Science and Technology (AIST)
Mbox 0604, 1-1-1Umezono, Tsukuba, Ibaraki 305-8568 Japan

This paper describes a method for organizing onset times performed along a fixed-tempo jam-session accompaniment into the the normalized (quantized) positions in a score. Unlike most previous beat-tracking-related methods focusing on predicting or estimating beat positions, our method deals with the problem of eliminating the onset-time deviations under the condition that the beat positions are given. To quantize polyphonic MIDI recordings of jam session, we propose a method of using Hidden Markov Model for modeling onset-time transition and deviation. Experimental results showed that the model that statistically learned MIDI recordings was effective enough to surpass the performance of semi-automatic quantization of commercial sequence software.

1. はじめに

我々は、ジャムセッションシステムの研究[1]において、実在する人間の演奏者の振る舞いを模倣できる仮想演奏者を生成することを目指してきた。仮想演奏者は、事前に人間の演奏者から学習したモデルに基づき演奏意図を決定し、フレーズデータベースの中から仮想演奏者のそのときの意

図に対応したフレーズを接続することで演奏を生成していた。フレーズデータベースには、1小節から8小節の長さの演奏が収められているが、それらは手作業で作成していたため、演奏者固有のフレーズまでは模倣できなかった。

そのような模倣のためには、実際の即興演奏が

らフレーズを切り出して、データベースを自動作成する必要がある。その際、演奏の発音時刻には意識的・無意識的なゆらぎがあり、それを除去せずに切り出しをおこなうと、生成時に接続する際に、つなぎ目が不自然となる。これを解決するには、発音時刻のゆらぎ・ずれを除去する**クオンタイズ**(quantize)が必要となる。

本研究では、このような即興演奏のフレーズ分割を自動化するため、テンポ一定の演奏(例えばドラムスによる伴奏)に合わせて弾かれた**演奏の発音時刻**を、楽譜上の**正規化された位置**へとクオンタイズすることを目的とする。ただし、即興演奏の場合には楽譜を演奏するわけではないため、楽譜上の正規化された位置は、実際には演奏者が意図した小節・拍内の位置を意味する。

市販のシーケンスソフトウェアに搭載されているクオンタイズでは、ユーザーがグリッド間隔(分解能)を指定し(例えば、8分3連音符、16分音符等の一定間隔のグリッドを指定し)、発音時刻を最も近いグリッドへ整列させることができる。したがって、拍構造が既知で固定されている演奏に対しては有効だが、ジャムセッションのように、8分3連音符や16分音符が頻繁に入れ替わるような演奏を正しくクオンタイズするためには、演奏の部位ごとに人間が手作業で指定する必要があり、我々の自動化の目的には使えない。

従来、コネクショニストモデルによるクオンタイズの研究[2]、音響信号に対するクオンタイズの研究[3]、ビート・リズムの認識に関するビートトラッキングの研究[4]-[8]がなされてきた。文献[2]の研究は、隣接する発音時刻間隔(あるいはその和)の比が整数比になれば安定するエネルギー関数を定義し、繰り返し計算で発音時刻のずれを修正していたが、エネルギー関数が固定されていたため、多様な演奏へ対処することができなかった。また、文献[3]の研究は自動採譜のための拍の推定のためのものであり、文献[4]-[8]の研究は拍位置を予測することに主眼が置かれていた。いずれも、問題設定が本研究(テンポ一定の伴奏に合わせて弾かれた演奏のクオンタイズ)とは異なっており、直接それらの手法は適用できない。

このような問題設定でも、単純な閾値処理は有効でなく(ゆらぎによるずれを誤ってクオンタイズする等) 難しい課題である。

一方、文献[9]では、本研究に近い問題設定が扱われている。この研究は、連続音声認識の方法論が、音符推定・テンポ推定・拍子推定にも有効であることを示した点が優れており、音符長の遷移と変動を隠れマルコフモデル(HMM)でモデル化することで、市販のシーケンスソフトウェアのクオンタイズを凌ぐ性能を得ることに成功している。しかし、単音の音符列のみにしか対応していなかったため、今回対象とするジャムセッションの即興演奏をクオンタイズすることはできない。

本研究では、今後さらに発展が期待される文献[9]のアプローチを参考にし、和音のクオンタイズも可能なモデルとして、発音時刻に焦点を絞った**発音時刻の遷移とゆらぎのモデル**を提案する。以下、2節ではクオンタイズの問題を逆問題としてとらえる立場から、隠れマルコフモデルを用いた問題の定式化と実際のクオンタイズの方法について述べる。次に3節では、クオンタイズの正解がラベリングされている演奏データを用いて学習することで、提案モデルがシーケンスソフトウェアのクオンタイズを越える性能を持つことを示す。最後に、4節でまとめと今後の課題を述べる。

2. 学習に基づくクオンタイズ

人間の演奏者が同じ演奏をくりかえし弾いた場合でも、MIDIのレベルでまったく同じ演奏となることは稀であり、演奏動作の微妙な差や、演奏の表情づけの差などにより発音時刻がゆらぐ。これを、元々演奏者が意図した発音時刻(小節・拍内の正規化された位置)系列から、ゆらぎのある実際に演奏された発音時刻系列を求める順問題とすると、演奏された発音時刻系列から元の意図した発音時刻系列を求めるクオンタイズは、その逆問題となる(図1)。ここでは、前者のゆらぎの生じる過程を「発音時刻ゆらぎの順モデル」としてモデル化する。そして、この順モデルから求めた「発音時刻ゆらぎの逆モデル」を用いて逆問題を解き、クオンタイズを実現する。

このように、実際の演奏から、その拍構造、楽譜構造を推定する問題を逆問題として捉える考え方は、文献[4][5][6]で紹介された。文献[9]でも、同様に逆問題と捉えて定式化しており、本研究でもそれを参考にしている。

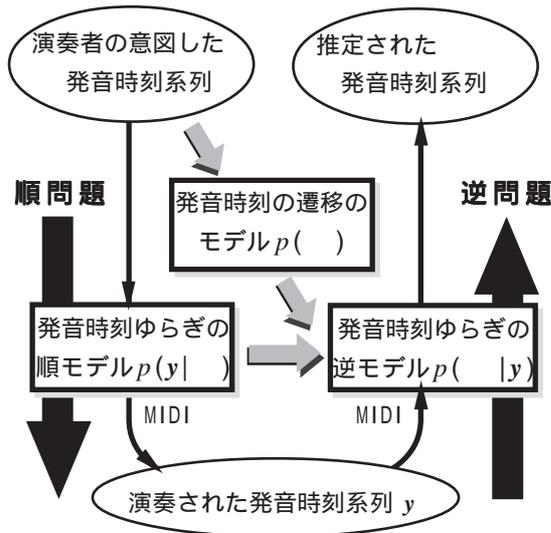


図1: クォンタイズにおける順モデルと逆モデル

2.1 発音時刻の遷移とゆらぎのモデル

発音時刻ゆらぎの順モデルは、意図した発音時刻系列を、演奏された発音時刻系列を y とした時、 $p(y|)$ で表される。このとき、逆モデルは、Bayesの定理より式(1)で表される。

$$p(|y) = \frac{p(y|)p()}{p(y)} \quad (1)$$

$p()$ は、に対する事前分布であり、演奏者がどのような発音時刻系列を弾きやすいかを表わす。 $p(y)$ は、に無関係であり、無視できる。逆問題の解は、式(1)を最大化するである(式(2))。

$$= \operatorname{argmax} p(|y) \quad (2)$$

2.2 隠れマルコフモデルによる定式化

本研究では、 $p(y|)$ と $p()$ を組み合わせたモデルを隠れマルコフモデル(HMM)で定式化する。隠れマルコフモデルはマルコフ的な隠れ状態遷移モデルと各状態における出力生成分布を組み合わせたモデルである。

(1) 演奏のモデル化

演奏をモデル化する前提条件として、以下のものを考える。

・何をモデル化するのか

本研究では、演奏の発音時刻すなわち音の立ち上がり時刻の分布をモデル化する。文献[9]では、音符の長さをモデル化していたが、

ギターのアルペジオや開放弦を含む演奏や、ピアノで右手と左手を同時に弾いているような演奏では、音符の長さのみで、演奏をモデル化することは困難であり、発音時刻でモデル化の方が有効であると考えられる。

発音時刻を用いて演奏をモデル化する場合、発音時刻の前後関係が保存されるようにすることで、音の順序が反転することがない妥当なクォンタイズを実現できる。

・演奏のモデル化の単位

本研究では、演奏をモデル化する単位として1拍の長さを採用した。1拍単位でモデル化を行う理由は、市販のクォンタイズを使った場合の主要な問題点である8分3連音符と16分音符の識別が1拍の長さで可能となるためである。(8分3連音符は1拍を3等分した位置、16分音符は1拍を4等分した位置に出現する。)実際に曲をクォンタイズする場合には、1拍のモデルを、次々とつなげあわせて曲全体を表現する。1拍という短い長さでモデル化することは、計算時間の節約と、学習の際により多くの演奏データを用意しやすいという利点をもつ。

・演奏の量子化単位

演奏の量子化単位を2種類定義する。1つは、実際に演奏された発音時刻(今後、イベント時刻と呼び、 k で表す)の量子化単位である。ここでは、シーケンスソフトウェアで採用されることが多い1拍の1/480を単位とする。もう1つは、演奏者の意図した発音時刻(今後、拍内での位置と呼び、 i で表す)の量子化単位である。ここでは、8分3連音符と16分音符の両方に対応できるように1拍の1/12を単位とした。本研究で述べるクォンタイズとは、イベント時刻から拍内での位置を求めることである。

(2) 1拍のHMMモデル

演奏の拍内での位置と、マルコフモデルの隠れ状態とを対応させることにより、演奏者の意図した発音時刻系列をマルコフモデルの隠れ状態の遷移に対応づけることができる。ここでは、拍内の12箇所的位置を、拍の頭から順に状態1から状態12までに対応させる。遷移は仮想的な状態であるStartからはじまって、Endで終了する。以下に、遷移の例をあげる。

- ・拍内に8分3連音符が3つ並んでいる場合
Start 1 5 9 End
- ・拍内に16分音符が4つ並んでいる場合
Start 1 4 7 10 End
- ・拍の頭で2音同時に鳴る和音の場合
Start 1 1 End
- ・拍内に音が無い場合
Start End

HMMの各パラメータを以下のように意味づけた。

- ・隠れ状態 i : 発音時刻の拍内での位置
- ・出力 k : イベント時刻
- ・状態 i から状態 j への状態遷移確率 a_{ij} : 拍内の i の位置で発音した後 j の位置で発音する確率。遷移は、常に拍内の前から後ろに向かって進むため、 $i > j$ のときは $a_{ij} = 0$ となる。
- ・出力確率 $b_i(k)$: 発音時刻の拍内での位置が i のときに、イベント時刻が k となる確率。状態 Start と End は出力を出さないため、対応する出力分布は存在しない。

状態遷移は常に S から始まるため、HMM初期状態分布 π_i は、 $\pi_S = 1$, $\pi_i (i = 1, 2, \dots, 12, E) = 0$ となる。図2に、本研究におけるHMMの概略図を載せる。図では、簡単のため12個の状態のうち、

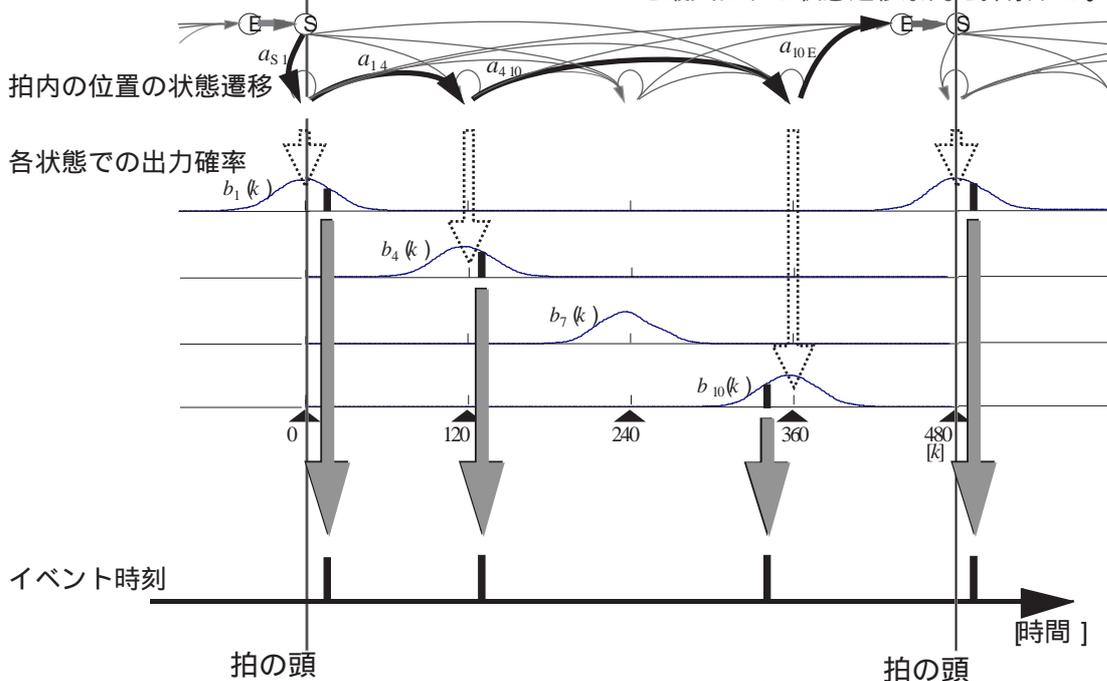


図2: 1拍の隠れマルコフモデルの概略図

1, 4, 7, 10以外の状態に関する遷移と出力確率は省略している。

(2) 複数のHMMによるモデル化

たとえば、演奏者が1拍の内部を8分3連音符で演奏しようとする場合と、16分音符で演奏しようとする場合とでは、拍内部の状態遷移確率は大きく変化することが予想される。すなわち、8分3連で演奏しようとする場合には、状態1, 5, 9への遷移確率が大きくなり、16分で演奏しようとする場合には、状態1, 4, 7, 10への遷移確率が大きくなるだろう。従って、1拍を一つのHMMによってモデル化するよりも、複数のHMMを切り替えて用いるほうが、より詳細なモデル化ができ、クオンタイズの性能がさらに向上すると期待される。そこで、本研究では、図3のように、16分音符で弾かれている16分遷移型、8分3連音符で弾かれている8分3連遷移型、拍の頭だけ弾かれている頭のみ型、何も弾かれていない音なし型、の4種類のHMMに分岐したモデルを用いる方法も試みた。

2.3 最適発音時刻系列の推定

1曲を通して、最適な状態遷移系列を推定するために、各拍ごとのモデルのStartとEndをつなぎ合わせたHMMを用いて、事後確率 $p(\cdot | y)$ を最大にする状態遷移系列を探索する。そのため

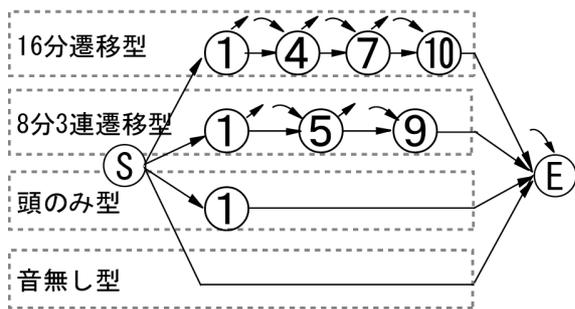


図3: 4種類のHMMに分岐したモデル

の方法としてはViterbi アルゴリズムを用いる。

T 個の音からなる演奏を考え、観測されたイベント時刻の系列を $y=(y_1, y_2, \dots, y_T)$ とする。これに対する最適状態遷移系列 $= (i_1, i_2, \dots, i_T)$ を求めるために、次の値を定義する。

$$f_t(i) = \max_{i_1, i_2, \dots, i_{t-1}} P[i_1, i_2, \dots, i_{t-1}, i_t = i; y_1, y_2, \dots, y_t] \quad (3)$$

$f_t(i)$ は、 t 個目のイベントまでを観測した時点で、状態遷移が状態 i で終わっている状態遷移系列中のベストスコア (最も高い確率を与える状態遷移系列の確率) である。また、 a_{ij} はモデルのパラメータを表す。ベストスコアの値は、次のような漸化式を満たす。

$$f_{t+1}(j) = \max_i [f_t(i) a_{ij}] \cdot b_j(y_{t+1}) \quad (4)$$

Viterbi アルゴリズムは、図4のようなトレリス上で、左から右に到る状態遷移系列の探索を行うアルゴリズムである。このトレリスは、横軸に

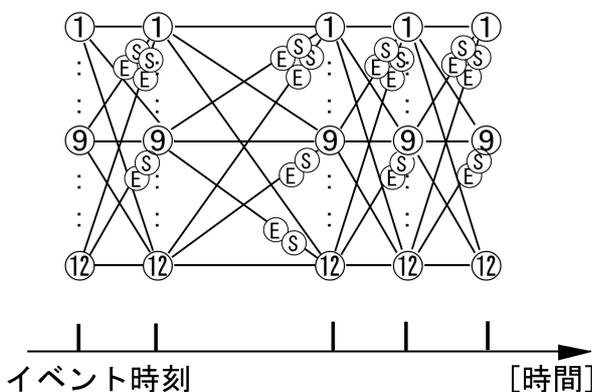


図4: 最尤な状態遷移を求めるトレリス

イベントを、縦軸に1から12の状態を配置しており、各イベントを1から12のいずれかの状態に対応づけてゆくことで、一つの状態遷移系列が得られる。拍と拍の境目をまたぐ状態遷移にはEndとStartを挿入する。このようなトレリス上で左から右に向かって式(4)を用いながら $f_t(i)$ を計算してゆくことにより、最も事後確率の高い系列を選択することができる。

EndとStartを適切に挿入するためには、二つのイベントの間で拍をまたいだか否かを判定する必要がある。ここでは次のような条件で判定した。

- $t_i > t_{i+1}$ の場合には、次の拍に移ったと判定し、End と Start を挿入する。

- $t_i \leq t_{i+1}$ の場合には、実際の演奏でのイベント間隔が、 q_t と q_{t+1} との距離に比べて半拍以上長い場合には、次の小節に移ったと判定し、End と Start を挿入する。

このような条件でEndとStartを挿入すると、実際の演奏イベントの位置とトレリス上の状態とのずれを半拍以内に行うことができる。実際の演奏でも、意図した発音時刻と実際の発音時刻が半拍以上ずれることはほとんどないため、この判定条件は妥当であるといえる。

4つのHMMに分岐したモデルの場合には、各拍で4つのHMMのどれを用いるかを、図5のような木構造を用いて探索していかなければならないが、プログラムとしては図4と同様のトレリス上の探索に帰着させることができる。その場合、トレリスの縦軸にあと2行を加え、16分遷移型の状態1と、8分3連遷移型の状態1と、頭のみ型の状態1とをそれぞれの別々の状態として扱う。また、音なし型の場合には、トレリス上にStartとEndを1つずつ挿入する。

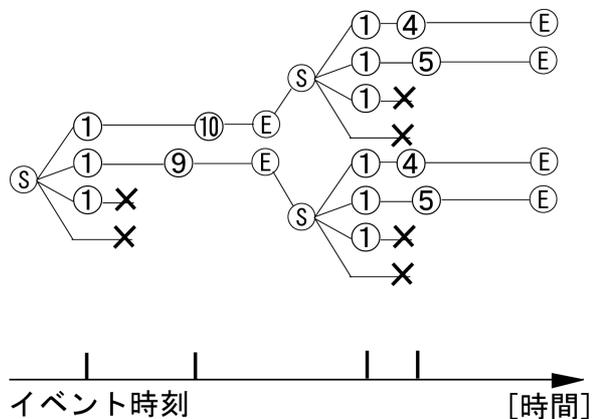


図5: 4つのHMMによる探索木の例

3. モデルパラメータの学習

本学習では、意図した発音時刻系列（正解データ）と、演奏された発音時刻系列 y の両方を学習用データとして用意することにより、発音時刻の遷移のモデル $p(\cdot)$ すなわち a_{ij} と、発音時刻ゆらぎの順モデル $p(y|\cdot)$ すなわち $b_j(k)$ を学習する。学習によって得られたモデルパラメータを使ってクオンタイズすることにより、モデルが正しい挙動をしていることを確認する。

3.1 学習用データ

学習用のデータセットとして、人工的に生成したデータと、人間の演奏者のデータとの2通りを用意した。人工データはプログラムの動作の検証を意図したものである。

(1)人工データ

人工データの生成では、まずその拍を8分3連音符の演奏にするか、16分音符の演奏にするかを乱数で決定し、1拍に含まれる音数を1から6までの乱数で決定する。次に、それぞれの音が、どの位置に割り振られるかを乱数で決定する。8分3連音符では3通り、16分音符では、4通りのうちいずれかの位置となる。以上のようにしてできたランダムデータ a の発音時刻を、平均0、標準偏差の正規分布の乱数を用いて、ゆらぎをつけたものを人工データ y_a とする。 $\sigma=10, 20, 30$ (1拍=480)とした3セットを作った。

(2)実演奏データ

3人の演奏者A,B,Cが、テンポ一定の伴奏に合わせてMIDIギターで即興演奏した記録を、実演奏データ y_h とする。演奏の長さは12コーラス(1コーラス=12小節)である。各演奏者に2回ずつ、一方はテンポ120、もう一方は演奏者自身が設定したテンポで演奏してもらい、合計6セットのデータを作成した。

(3)正解データ

学習を行うためには、正解データ、すなわち、演奏者が意図した発音時刻を教師データとして与える必要がある。人工データの場合には、ゆらぎを加える前のランダムデータ a が正解データとなる。

一方、人間の即興演奏データの場合には、正解データが存在しないので、人間が手作業でラベル

づけをすることにより正解データ y_h を作成した。具体的には、市販のシーケンスソフトウェア(Twelve Tone System, Cakewalk Pro Audio 9)を用いて、ピアノロール表示した音符の位置と長さを視覚的に確認しながら、1音ずつ適切な位置にラベルづけしていった。発音時刻のゆらぎが大きく、いくつかの正解候補が考えられた場合には、すべての候補のクオンタイズ結果を作って聴き比べ、より適切だと判断したものを採用した。

3.3 モデルパラメータの推定

正解データと演奏された発音時刻系列 y から、 a_{ij} 、 $b_j(k)$ を推定した。図6中の(a)は、人工データを作るときのゆらぎの標準偏差 $\sigma=20$ (1拍=480)とした場合の人工データから学習された、 $b_5(k)$ の分布である。(b)は、演奏者Cの2回目の演奏(演奏C2と呼ぶ)から求めた $b_5(k)$ の分布である。

人工データは、もともと正規分布の乱数でゆらぎを与えているので、 $b_5(k)$ の分布(a)も、正規分布となった。一方、実演奏データから求めた、 $b_5(k)$ の分布(b)もほぼ正規的な分布となったが、分布の中心がやや後ろにずれていた。このことは、状態5の位置の音符が、意図した発音時刻よりも遅れて演奏される傾向があることを示している。実際、演奏C2を調べたところ、8分3連音符の頭の音が通常より長く弾かれ、次の音の発音時刻が遅れる傾向が認められた。

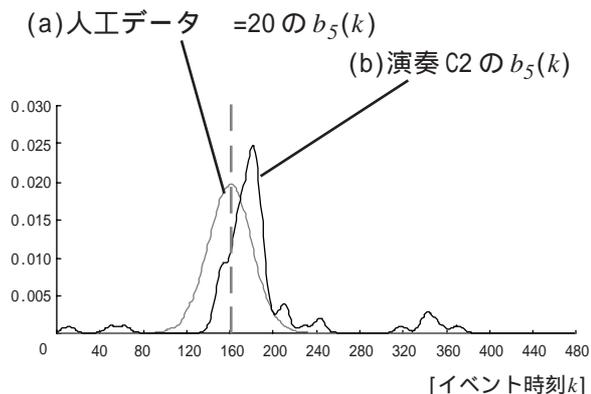


図6：人工・実演奏データから求めた $b_5(k)$

ピアノロール表示とは、縦軸を音の高さ、横軸を時間とし、音の出るタイミングと鳴り続けている長さを表示するものである。

3.4 クオンタイズ結果

クオンタイズの評価を、音単位と拍単位の2通りで行った。音単位の評価は、各音の発音時刻が正解と一致している割合である。拍単位の評価は、拍の種類を、16分遷移型、8分3連遷移型、頭のみ型、音なし型の4種類に分けたとき、各拍の種類が正解の種類と一致している割合である。この4種類への分け方は、1つのHMMでのモデルの場合は、クオンタイズ結果から求め、4つのHMMに分岐したモデルの場合は、クオンタイズしたときに選択されたHMMの型から求めている。

表1に正解データに含まれる4種類の拍の割合を示す。演奏者A,Bが2回の演奏でそれぞれ似た傾向の演奏をしているのに対し、演奏者Cは、1回目は16分音符を多用した演奏、2回目は16分音符のまったくない演奏であった。

本研究では、一致率を以下のように定義し、クオンタイズの性能を評価する。

$$(\text{音単位の一一致率}) = \frac{(\text{拍内の位置が正解と一致した音数})}{(\text{正解データ中の音数})} \quad (5)$$

$$(\text{拍単位の一一致率}) = \frac{(\text{拍の種類が正解と一致した拍の数})}{(\text{正解データの拍の数})} \quad (6)$$

(1)機械クオンタイズ

まず、閾値処理による機械的なクオンタイズ(機械クオンタイズと呼ぶ)の性能を評価するために、8分3連、16分、16分3連の3種類の分解能で機械クオンタイズを実行し、音単位での一致率を求めた(表2)。尚、これ以外の分解能では一致率はさらに悪くなる。

その結果、A1やC1、人工データのように16分の拍が多く含まれる演奏では、低い一致率しか得られなかった。ただし、C2のように16分で弾かれた拍がまったく無い演奏では、8分3連の分解能での機械クオンタイズが高い一致率を示すこともあった。

(2)学習に用いた演奏のクオンタイズ

学習結果のモデルパラメータを用いて、学習に使ったのと同じデータについてクオンタイズを実行し評価した。

その結果、人工データでは、1つのHMMでのモデルと4つのHMMでのモデルの両者とも、機械クオンタイズでの一致率に比べ、2割以上高い一致率を示し、また実演奏データの多くでも機械クオン

タイズより高い一致率を示していることから、HMMによるモデルが有効であることが確認できた(表2)。

1つのHMMによるモデルと4つのHMMによるモデルの音単位の一一致率を比較すると、実演奏データでは、4つのHMMによるモデルの方が、一致率が高かった。これは、状態1を、16分遷移型、8分3連遷移型、頭のみ型の3状態に分けることにより、各々の場合の $b_1(k)$ の分布を別々に学習し、より詳細にモデル化されたためであると考えられる。実際、拍単位の評価でも4つのHMMによるモデルのほうが、1つのHMMによるモデルより高い一致率を示していることから16分遷移型と8分遷移型の識別性能が向上したことが確認できた(表3)。

一方、人工データでは、1つのHMMによるモデルを使った場合と4つのHMMによるモデルを使った場合とが同じ一致率となった。これは、状態1を分けた場合の3つの $b_1(k)$ も、分ける前の $b_1(k)$ もいずれも同じ正規分布となるためである。

(3)他の演奏のモデルパラメータでのクオンタイズ

各実演奏データに対して、それ以外の演奏で学習したモデルパラメータを用いて、クオンタイズを実行し評価した。評価に用いたモデルパラメータは、1)同じ演奏者の別の演奏で学習したもの、2)自分以外の2人の演奏で学習したもの、3)ランダムデータ($n=20$)で学習したものの、3通りである。1)は、ある演奏から得られたモデルパラメータが、同じ演奏者のほかの演奏に適用できるか調べるためのものである。1)の結果、演奏者A,Bは、その曲から学習したパラメータを使ってクオンタイズするのと比べて、一致率が5%程度しか劣っていなかった。一方、演奏者Cは、20%近く劣っていた。これは、2回の演奏をしたときに、演奏者A,Bはそれぞれ似た傾向の演奏を2度したのに対し、Cは、大きく異なる傾向の演奏をしたためだと考えられる。このことから、A,Bのように、演奏傾向が似ている場合には同じモデルで適用できるが、大きく異なる場合には、適用できないということがわかった。2)3)は、モデルパラメータにいくつかの演奏の平均的なモデルや、正規的な分布を与えることにより、クオンタイズの性能がどう変化するかを調べたものであるが、いずれ場合も、A、Bでは1)より性能が劣っており、またCでも、1)より高くなったものの、その曲から学習したパラメータを使う場合と比べて一致率が低下していた。

表1: 正解データに含まれる4種類の拍の割合

	人工データ			演奏者A		演奏者B		演奏者C	
	=10	=20	=30	A1	A2	B1	B2	C1	C2
16分の拍の割合	38.8%	38.8%	38.8%	21.4%	5.5%	6.5%	2.7%	37.2%	0.0%
8分3連の拍の割合	37.8%	37.8%	37.8%	54.2%	68.7%	58.2%	85.8%	42.1%	39.7%
頭だけの拍の割合	6.9%	6.9%	6.9%	6.8%	10.1%	11.9%	6.4%	5.9%	26.7%
音がない拍の割合	16.4%	16.4%	16.4%	17.6%	15.7%	23.3%	5.1%	14.8%	33.8%

表2: 機械クオンタイズと本手法との比較 (音単位の一一致率)

	人工データ			演奏者A		演奏者B		演奏者C	
	=10	=20	=30	A1	A2	B1	B2	C1	C2
機械クオンタイズ(8分3連)	65.6%	58.2%	62.0%	67.6%	85.6%	79.4%	88.6%	57.0%	97.7%
機械クオンタイズ(16分)	63.9%	67.9%	65.9%	54.5%	37.3%	36.8%	34.7%	70.7%	45.5%
機械クオンタイズ(16分3連)	77.1%	70.7%	60.8%	57.7%	48.4%	57.8%	51.3%	56.1%	82.5%
1つのHMMでのクオンタイズ	99.6%	95.9%	86.5%	75.9%	84.8%	80.0%	90.5%	85.1%	95.0%
4つのHMMでのクオンタイズ	99.5%	95.9%	86.5%	82.3%	89.8%	81.4%	92.8%	85.5%	95.7%

機械クオンタイズに対して、本手法の方が性能が上がった箇所に「」をつけた。

表3: 1つのHMMによるモデルと4つのHMMによるモデルの比較 (拍単位の一一致率)

	人工データ			演奏者A		演奏者B		演奏者C	
	=10	=20	=30	A1	A2	B1	B2	C1	C2
1つのHMMでのクオンタイズ	98.6%	93.8%	81.3%	79.0%	85.4%	84.2%	91.9%	83.9%	94.9%
4つのHMMでのクオンタイズ	98.6%	93.8%	81.3%	84.9%	89.2%	87.5%	93.5%	84.7%	97.3%

表4: 他の演奏のモデルパラメータでのクオンタイズ

	人工データ			演奏者A		演奏者B		演奏者C	
	=10	=20	=30	A1	A2	B1	B2	C1	C2
同じ演奏者の別の演奏で学習したモデル	/			75.9%	84.8%	80.0%	90.5%	55.5%	77.9%
自分以外の2人の演奏で学習したモデル				70.2%	79.0%	59.2%	73.5%	76.9%	93.1%
ランダムデータで学習したモデル				73.3%	53.4%	60.8%	55.2%	82.5%	86.1%

このことは、本研究で提案したモデルが演奏者の発音時刻の特徴や癖を獲得しており、ある演奏者のモデルを一度作ればその演奏者が似た傾向の演奏をしたときには、同じモデルでクオンタイズが可能であることを示している。

4. まとめ

本稿では、テンポ一定の演奏に合わせて弾かれた和音を含む即興演奏の発音時刻から、演奏者が意図した発音時刻を推定するため、発音時刻の遷移とゆらぎのモデルを提案した。そして、正解データを用いた学習をすることにより、提案モデルが市販のシーケンスソフトウェアのクオンタイズを超える性能を持つことを示した。今後、フリーズデータベースの自動作成を実現するため、正解データがなく、実演奏データのみしか与えられない場合の、モデルパラメータの推定法について検討していく。

謝辞

演奏者として実験に協力して頂いた、橋本大輔氏、千田真一氏、斎田康弘氏に感謝いたします。

参考文献

- [1] 浜中, 後藤, 大津: "学習するジャムセッションシステム: 演奏者の振る舞いのモデルの獲得," 情報処理学会音楽情報科学, 99-MUS-34-5, pp.27-34, 1999.
- [2] P.Desain, H.Honing: "The Quantization of Music Time: A Connectionist Approach," Computer Music Journal, Vol.13, pp.56-66, 1989.
- [3] 片寄, 井口: "自動採譜システム," 人工知能学会誌, Vol.5, No.1, pp.59-66, 1990.
- [4] 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," 人工知能学会研究会資料 AIチャレンジ研究会, SIG-Challenge-9801-2, pp.7-14, November 1998.
- [5] Masataka Goto and Yoichi Muraoka: "An Audio-based Real-time Beat Tracking System and Its Applications," Proc. of ICMC, pp.17-20, 1998.
- [6] 後藤 真孝: "拍節認識(ビートトラッキング)," bit別冊 コンピュータと音楽の世界 ---基礎からフロンティアまで---, pp.100-116, 共立出版, August 1998.
- [7] R.Dannenberg, B.Mont-Reynaud: "Following an Improvisation in Real Time," Proc. of ICMC, pp.241-248, 1987.
- [8] P.Allen, R.Dannenberg: "Tracking Musical Beats in Real Time," Proc. of ICMC, pp.140-143, 1990.
- [9] 斎藤, 中井, 下平, 嵯峨山: "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," 情報処理学会音楽情報科学, 99-MUS-33-5, pp.27-32, 1999.