

隠れマルコフモデルによる音楽演奏の音符列の推定

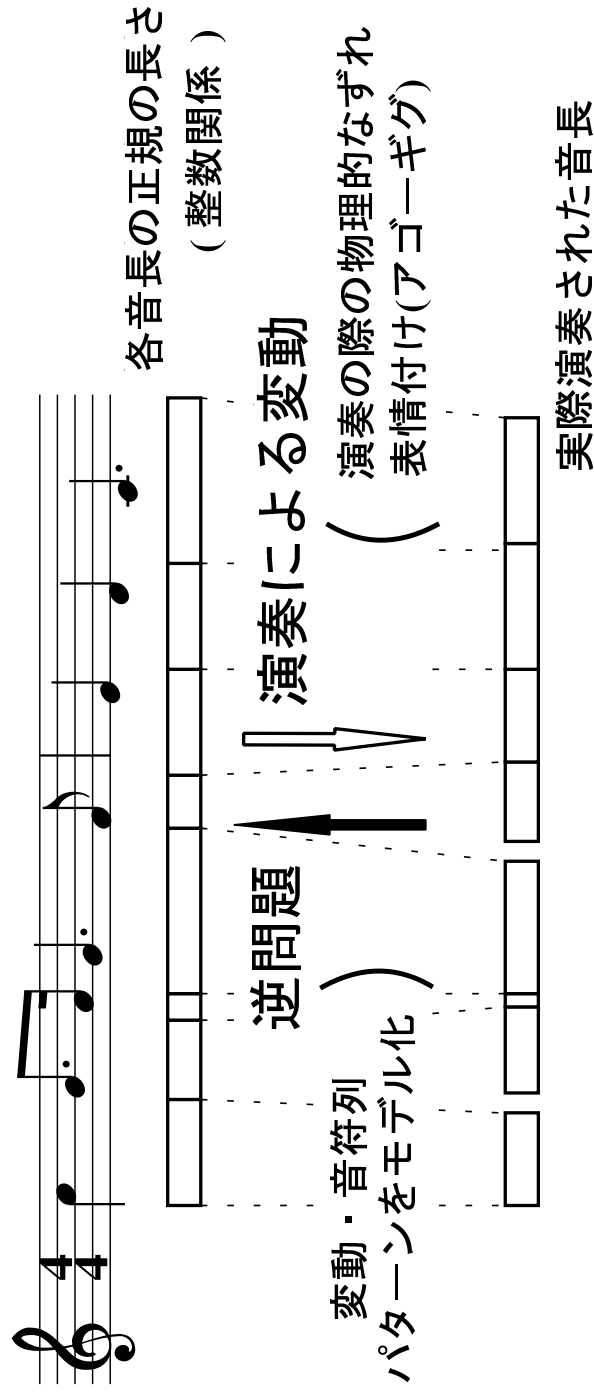
北陸先端科学技術大学院大学 情報科学研究科
知能情報処理学講座 嵯峨山・下平研究室

齋藤 直樹 中井 満 下平 博 嵯峨山 茂樹

1. 研究の目的と意図
2. 音符列推定に関して
3. HMM を用いた本手法のモデル化
4. 音楽的解析におけるモデル化と各種実験
5. まとめと今後の課題

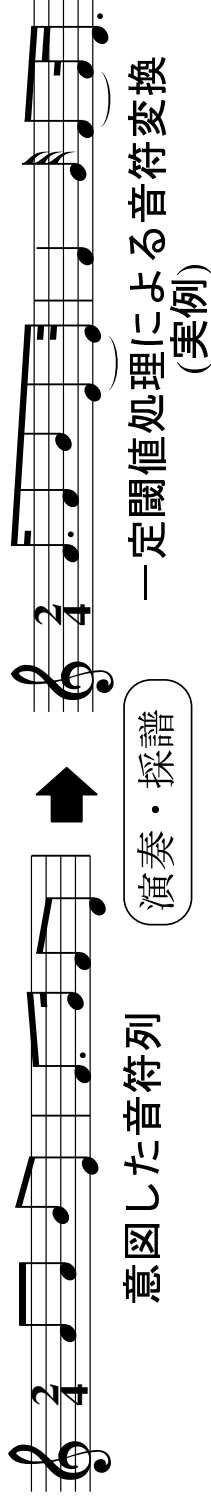
本研究の目的

- 本研究の目的
 1. 演奏の音長系列から音楽的意図に基づく音符列を確率モデルを用いて推定しその精度の向上を目指す
 2. 楽譜投入・自動採譜に対し意図した楽譜を作成
- 逆問題としての音符列推定



音符列推定について

殆どの市販品：一定閾値による量子化 (quantize)



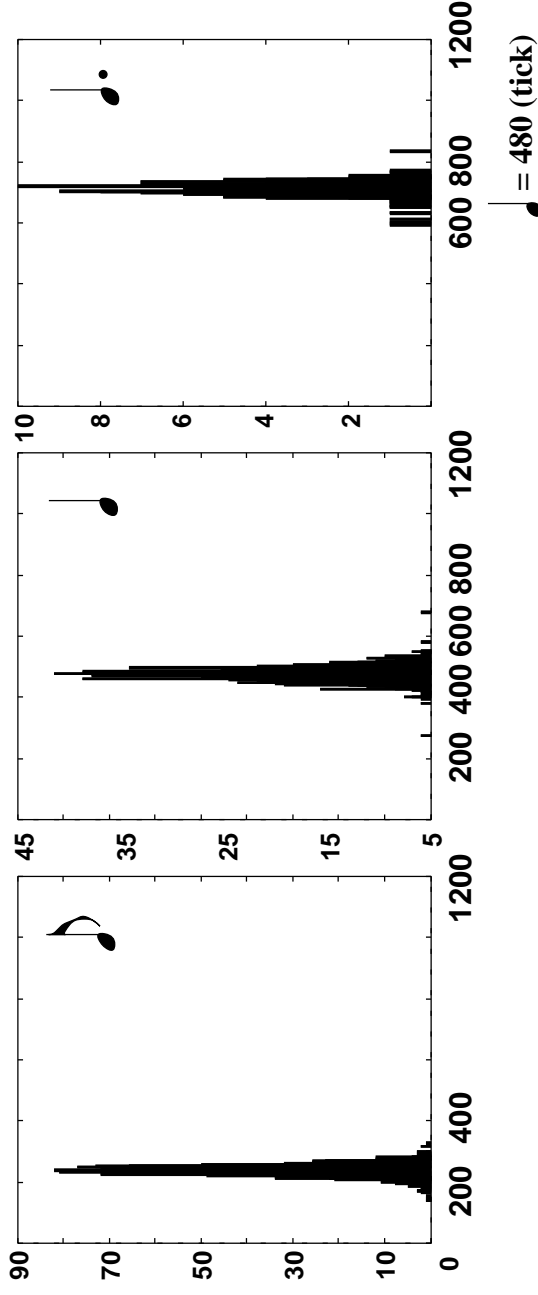
{ 演奏に忠実な音符変換だが、意図した音符列ではない
演奏の裏にある演奏者の意図を推定する必要あり



トップダウン(モデルベース)によるリズム理解
音声認識において有効性が示されている手法

	連続音声認識	音楽リズム認識
入力単位	文音声	楽曲
語彙	単語	リズムパターン
隠れ状態	音響イベント	音符
観測値	スペクトル列	物理的音符長列

モデル化（音長の変動モデル）：1



1. 音長の変動：確率分布としてモデル化する

2. 確率分布：正規分布による近似

平均： $\mu =$ 各音符の正規の長さ $\left\{ \begin{array}{l} A: \text{正規の音符長} \mu \text{ に対する変動率} \\ B: \text{音符長に依存しない固定分のずれ} \end{array} \right.$
 標準偏差： $\rho = A \times \mu + B$



最小自乗法により係数の決定

モデル化（音長の変動モデル）：2

モデルの想定テンポとの差、演奏者による違い



標準偏差をひろめに設定 $A : 0.06$, $B : 0.0114(S)$

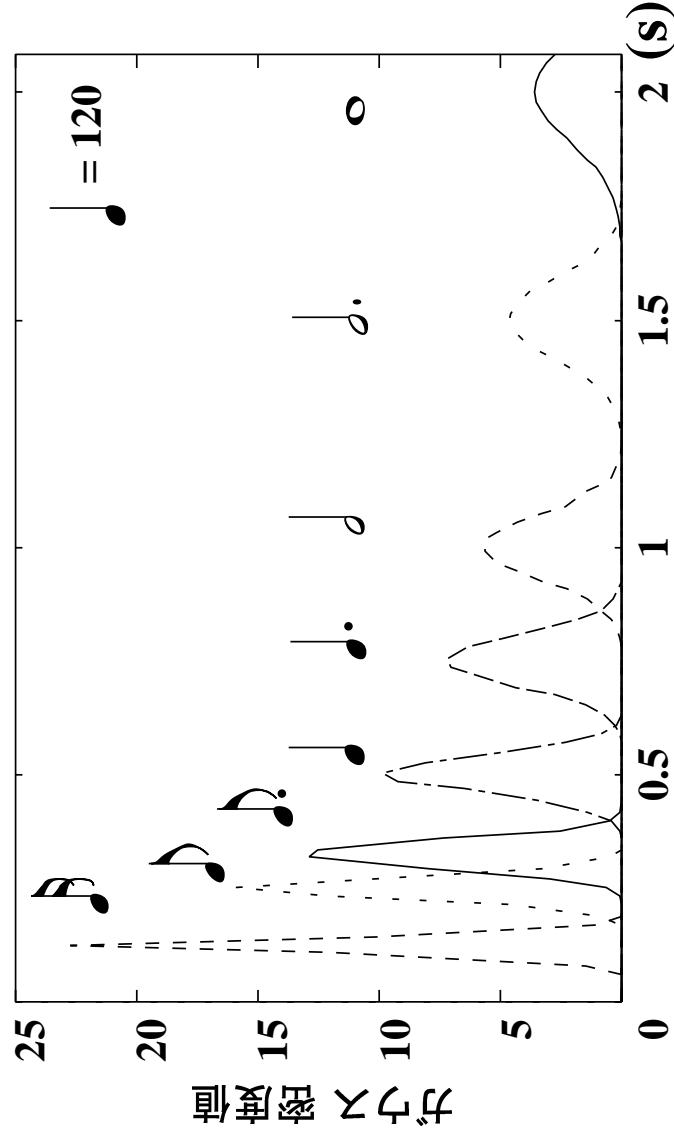
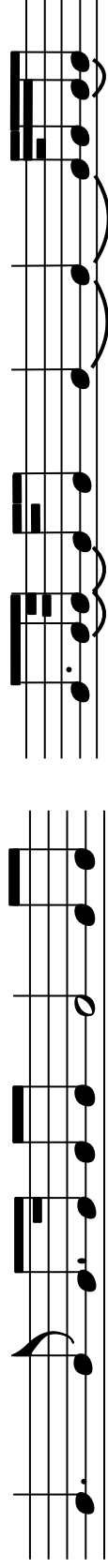


図 1. $\rho = 0.06\mu + 0.0114(s)$

ある音符 j が音長 x で演奏される確率： $b_j(x)$

音符列モデル a_{ij}

揺らぎがある演奏でも人間には意図した音符列が伝わる
 → 出現しうる音符列を常識的な知識として持つ



ありえる音符列



まずありえない音符列

音声認識の言語(文法)モデルを考慮し音符の遷移を確率モデル化

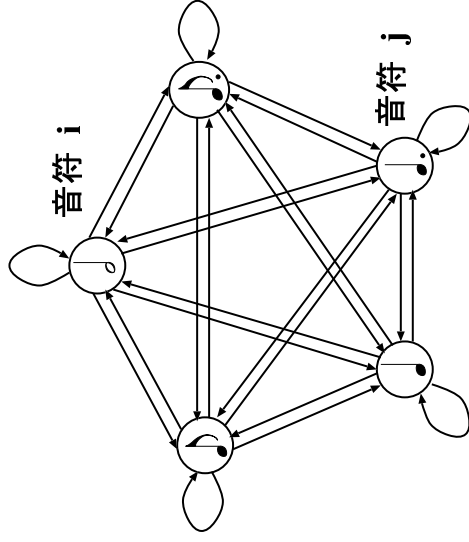


図 2. 2 音符連接モデル

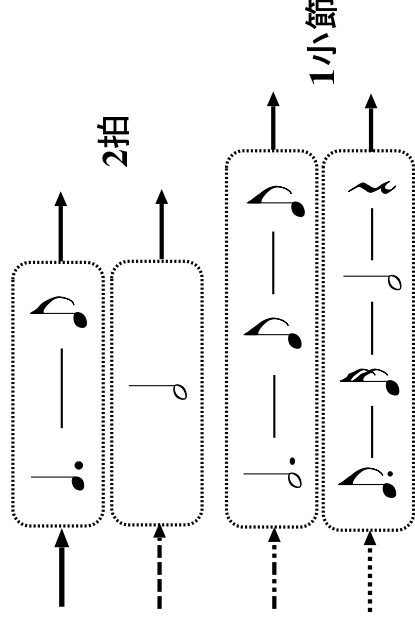


図 3. リズムパターン例

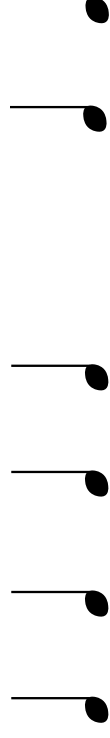
リズムパターンの統計結果

童謡・民謡・歌曲を対象に 4/4 拍子の曲 88 曲からの音符列の統計

- リズムパターン統計結果

	1 小節 1 パターン	2 拍 1 パターン
出現頻度	267 種類	137 種類

一番多い音符列

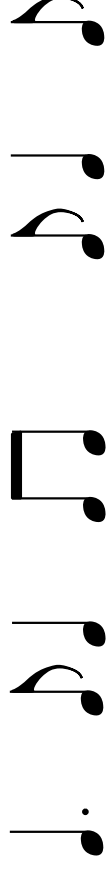


・

・

・

10 番目



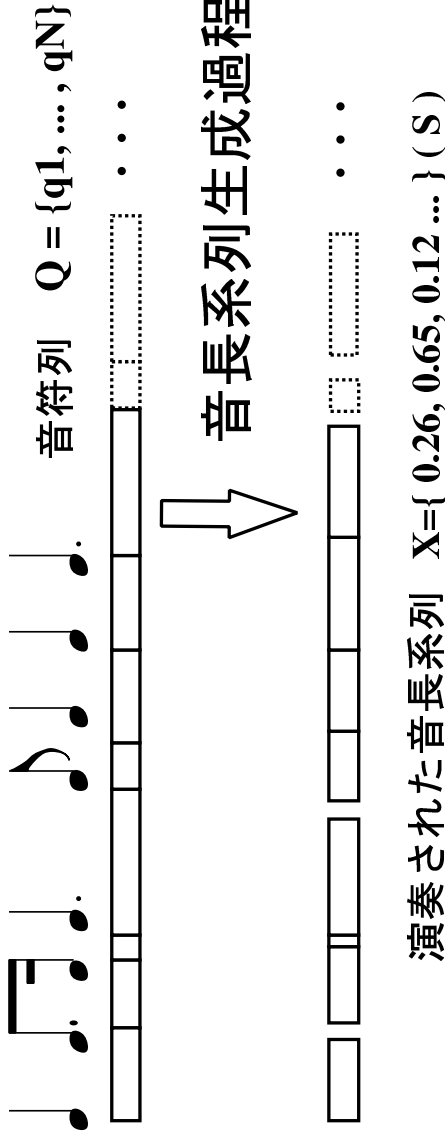
音符列モデルから得られる確率

ある音符 i からある音符 j へ遷移する確率： $a_{i,j}$

音長系列生成確率 $P(X|Q)$

2つの確率モデル { ある音符 j が音長 x で演奏される確率: $b_j(x)$
 ある音符 i からある音符 j に遷移する確率: $a_{i,j}$

$P(X|Q)$: 音符列 Q が音長系列 X で演奏される確率



音長系列 $X = \{x_1, \dots, x_N\}$ が生成される確率値

$$P(X|Q) = \prod_{t=1}^N a_{q_{t-1}, q_t} \cdot b_{q_t}(x_t)$$

音符列推定の定式化

$P(X|Q)$: 音長生成確率

音符列推定の定式化

$Q^* = \arg \max_Q P(Q|X)$: 音長系列 X が与えられた時仮説 Q の中で
確率が最大となる音符列 Q^* を求める

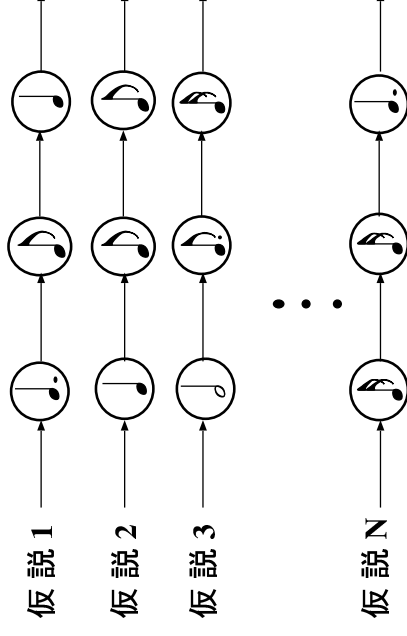
バイズの定理による逆問題の定式化

$$P(Q|X) = \frac{P(X|Q)P(Q)}{P(X)}$$



$$Q^* = \arg \max_Q P(X|Q)$$

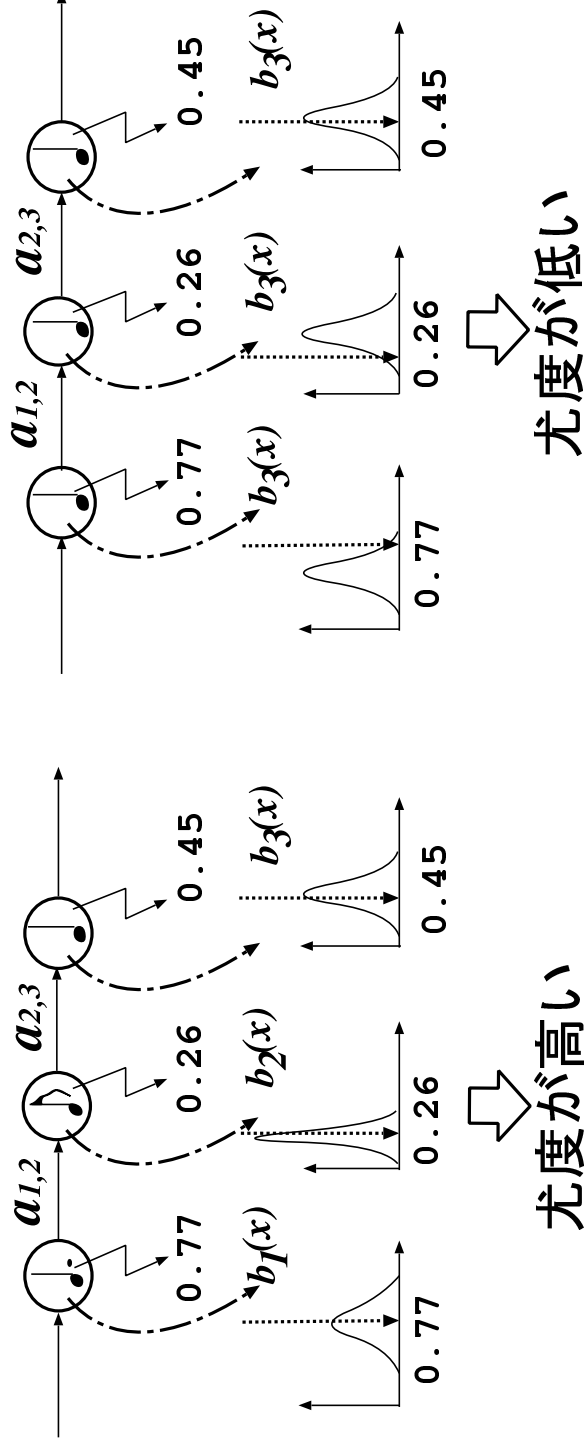
$$\mathbf{x} = \{0.77, 0.26, 0.45\}$$



HMM (Hidden Markov Model) による音符列推定 : 1

{ 音長の伸縮変動モデル → HMMによるモデル統合
 { 音符列モデル

入力系列 $X = \{ 0.77, 0.26, 0.45 \}$ (s)



Viterbi 探索により尤度最大の遷移系列を求める

HMM による音符列推定 : 2

HMM の各パラメータ

- 状態 s_i : 音符 i
- 初期確率 π_i : ある音符 i から曲が始まる確率
- 遷移確率 $a_{i,j}$: 音符 i から音符 j へ遷移する確率
- 出力確率 $b_j(x)$: 音符 j が音長 x で演奏される確率
- 入力系列 X : 演奏された音長系列 $X = \{x_1, \dots, x_n\}$

最小単位となるリズムパターンモデル

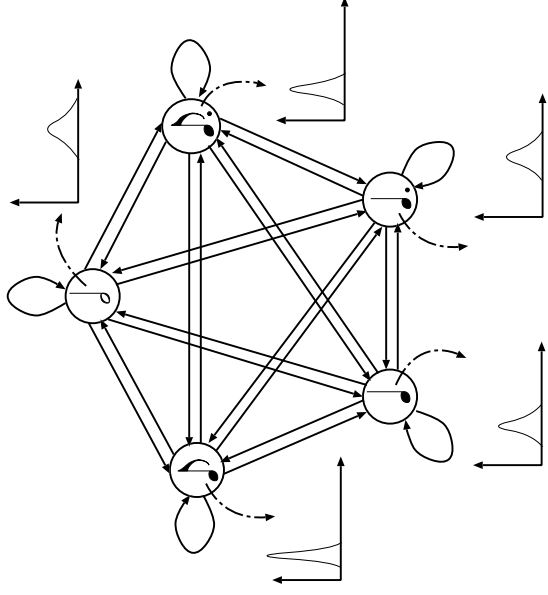


図 4. 2 音符接続 HMM

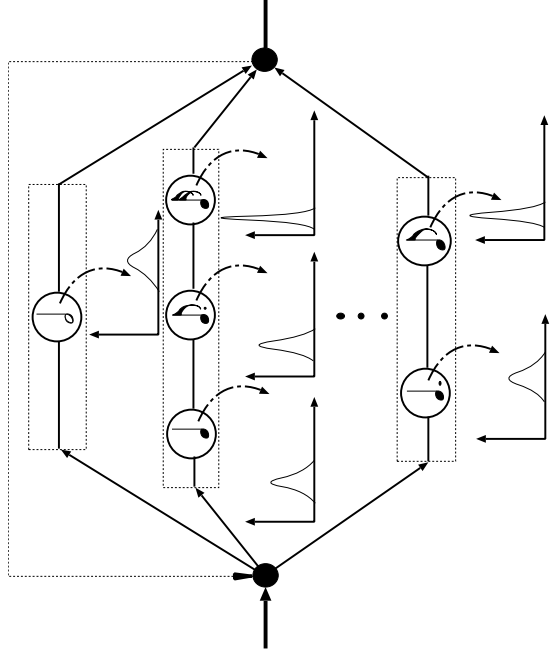


図 5. 2 拍単位リズム HMM

音楽的解析における推定実験

HMM を用いて推定実験を行う音楽的要素

音符列、演奏されたテンポ、拍子、小節線を HMM を用いて推定する

音符列の推定をベース



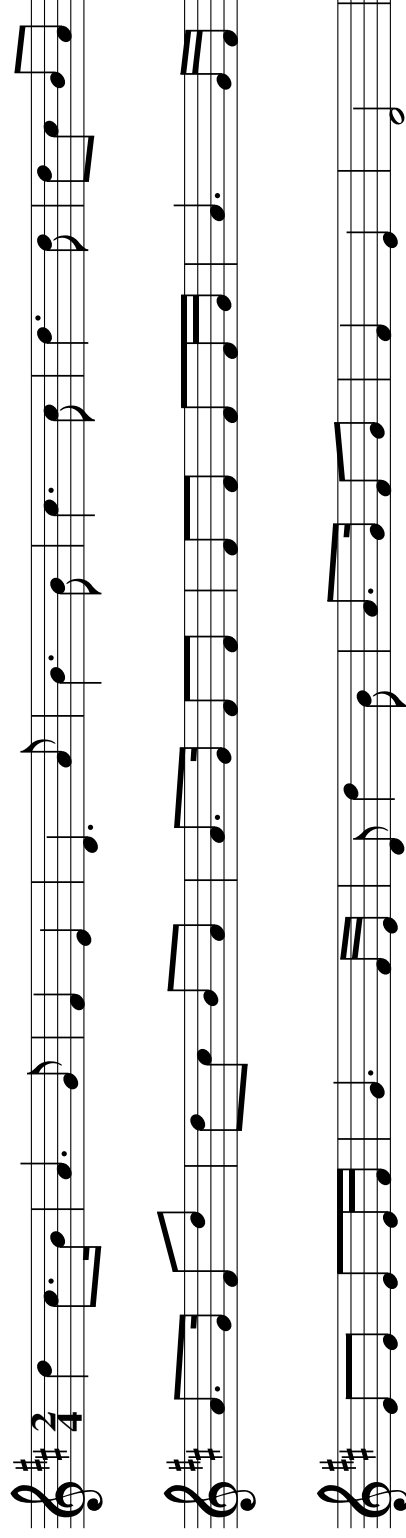
演奏したテンポの推定

拍節推定

これらと同じ原理により推定する

音符列推定実験

- 入力曲： 毘ろびとこぞりて (ニ長調)」
 - 出現する音符の種類が豊富、よく知られている、比較的短い
- 条件 1 (メトロノームありのテンポ一定で演奏)
- 被験者 10 人 (合計 10 曲)



音符列推定実験結果例：閾値処理

YAMAHA XGworks

Three staves of musical notation in treble clef, key of D major (one sharp), and 2/4 time signature. The first staff begins with a treble clef, a key signature of one sharp (F#), and a 2/4 time signature. The music consists of eighth and sixteenth notes, some beamed together, and rests. The second and third staves continue the melody with similar rhythmic patterns and note values.

Finale

Three staves of musical notation in treble clef, key of D major (one sharp), and 2/4 time signature. The first staff begins with a treble clef, a key signature of one sharp (F#), and a 2/4 time signature. The music continues with eighth and sixteenth notes, some beamed together, and rests. The second and third staves continue the melody with similar rhythmic patterns and note values.

音符列推定実験結果例：2beatRhythmHMM

2拍 RhythmHMM(休符挿入を認めるモデル)

Three staves of musical notation in 2/4 time, key signature of one sharp (F#). The first staff shows the original melody with rests. The second staff shows the model's prediction, which correctly identifies and inserts rests. The third staff shows the model's prediction, which incorrectly identifies some notes as rests.

2拍 RhythmHMM(休符挿入を認めないモデル)

Three staves of musical notation in 2/4 time, key signature of one sharp (F#). The first staff shows the original melody with rests. The second staff shows the model's prediction, which does not insert rests. The third staff shows the model's prediction, which incorrectly identifies some notes as rests.

評価方法

MIDI 信号を対象：各音長が正しく推定されているかに着目
音声認識で用いられている評価尺度

正解精度

$$accuracy = \frac{N - sub - del - ins}{N} \times 100 (\%)$$

- N : 未知入力の総音符数
- sub : 誤った音符に置換された数
- del : 正しい音符が脱落した誤り数
- ins : 異なる音符が挿入された数



休符挿入は閾値処理により判定 { 休符を考慮(挿入)した結果
休符を挿入しない結果

評価尺度に基づいた認識率

- 条件 1 (メトロノームありのテンポ一定で演奏)
- 被験者 10 人 (合計 16 曲)

表 1 : 音符列推定精度 : 16 曲の音符認識率 (%)

method	休符を認める (%)	休符を認めない (%)
閾値処理 (XGworks)	40.70	(85.86)
閾値処理 (Finale)	38.79	(87.20)
音符連接 HMM	53.73	87.39
2 拍 Rhythm HMM	59.65	97.26

HMMによるテンポ推定と変動テンポのモデル化

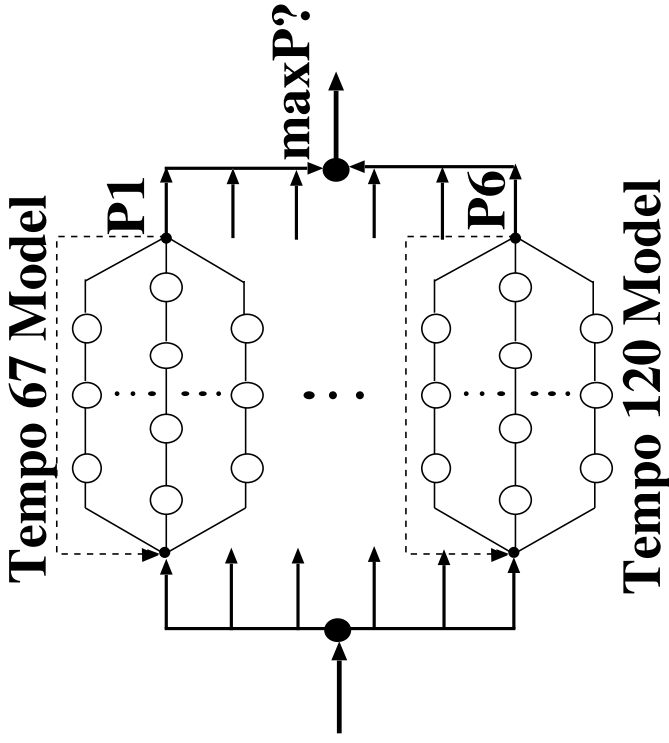


図6. 一定テンポ HMM

- 複数のテンポのリズムモデルを作成
- 尤度が最も高い遷移系列のテンポを推定

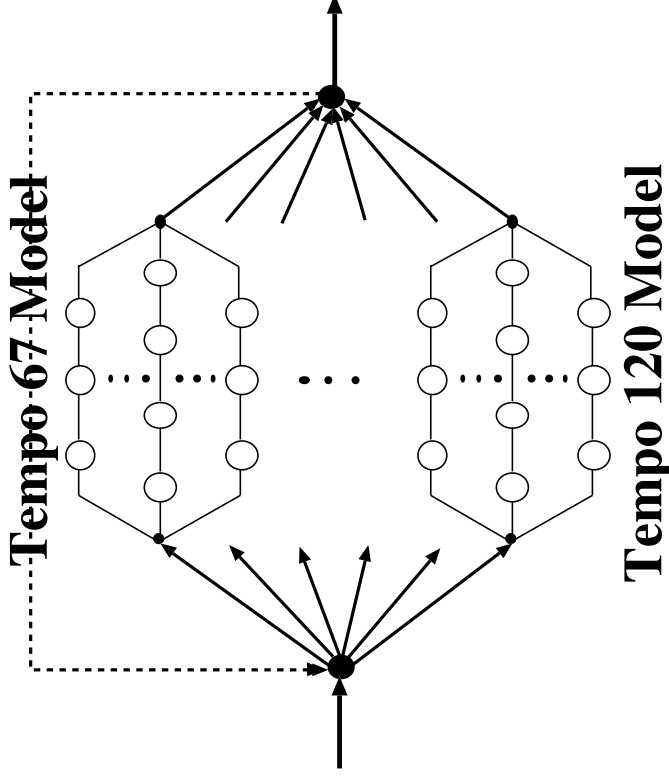


図7. 変動テンポ HMM

- 各テンポモデル間に遷移確率を設ける
- 激しいテンポ変動に対し追従した解析を行う

テンポ推定実験結果

- 入力曲： 毬ろびとこぞりて (ニ長調)
- 条件 2 (メトロノームなしで自由なテンポを一定で演奏)
- 被験者 10 人 (10 曲)
- 休符挿入なし

{ 実測値：曲の平均テンポ = $BEAT$ 数 (38 個) / 演奏時間 (分)

{ 推定値：尤度が高かった一定テンポ HMM

player#	実測値	推定値	player#	実測値	推定値
1	98.35	95	6	116.41	120
2	93.31	95	7	111.74	107
3	99.20	95	8	99.88	95
4	127.06	120	9	109.25	107
5	106.34	107	10	65.16	67

テンポ変動モデル実験結果

テンポが前後に変動する入力曲に対し、変動に追従した解析を行い、
音符列推定精度の向上を目指す

♩ = 95

120 => 107 => 95 => 107 => 95

85 => 120 => 95 => 85 => 76 => 67

図 8. テンポ変動に追従した推定結果（休符挿入なし）

現在テンポモデル間の遷移確率を等確率に設定

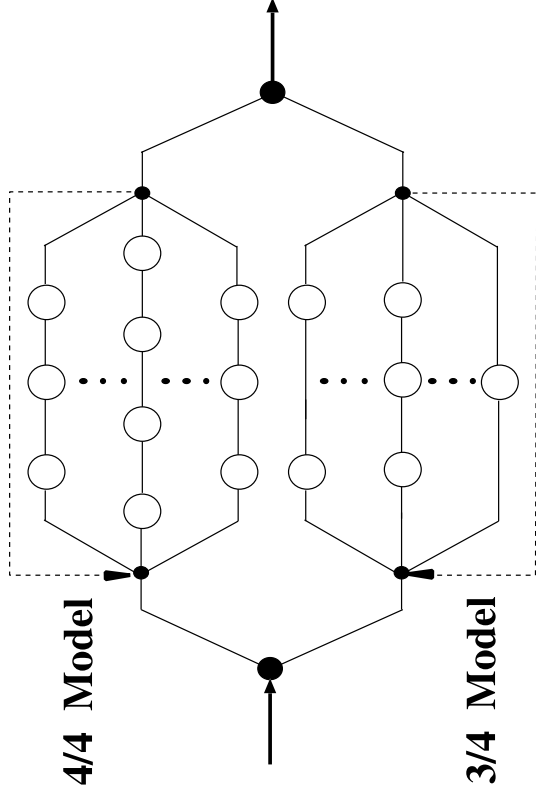


今後この遷移確率を検討し改善していくことが必要

IHMM による拍節推定問題のモデル化

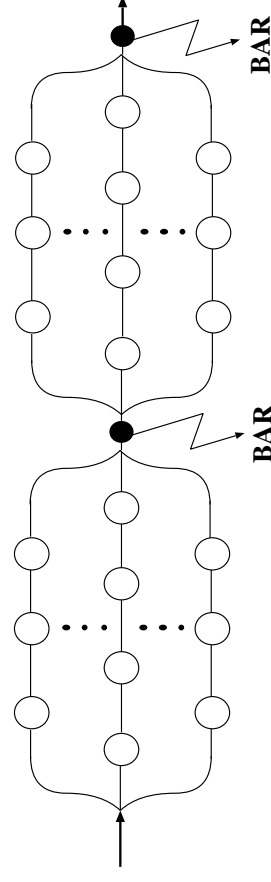
拍子推定問題

- 仮定：拍子特性 =
1 小節中の音符数とその組み合わせ
 - 3/4 拍子、4/4 拍子の 1 小節 1 パターンの統計
- ➡
- 求めた遷移系列が何拍子かを判定



小節線の挿入

- 1 小節 1 パターンを採用
- ➡
- リズムの最終状態の後、小節線挿入



拍節推定：アウフタクトの認識

アウフタクトは別パターンとして、初期確率のみを持たす



図9. アウフタクトパターン例



{ 曲の最初にでてくるパターンとしてモデル化
 { 曲の途中で採用されることはない

♩ = 90

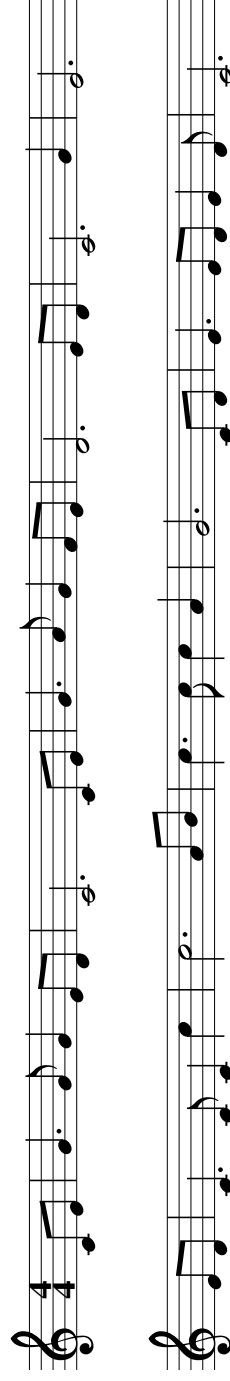


図10. アウフタクトの認識 若者たち」

拍子推定実験結果

- 入力曲と演奏条件
 - 4/4 拍子 10 曲と 3/4 拍子 10 曲の計 20 曲を入力
 - 演奏条件 1 (メトロノームありテンポ一定)
 - 休符挿入あり

● 実験結果

入力曲数	4/4 拍子	3/4 拍子
4/4 拍子 10 曲	10 曲 (4/4)	—
3/4 拍子 10 曲	2 曲 (4/4)	8 曲 (3/4)

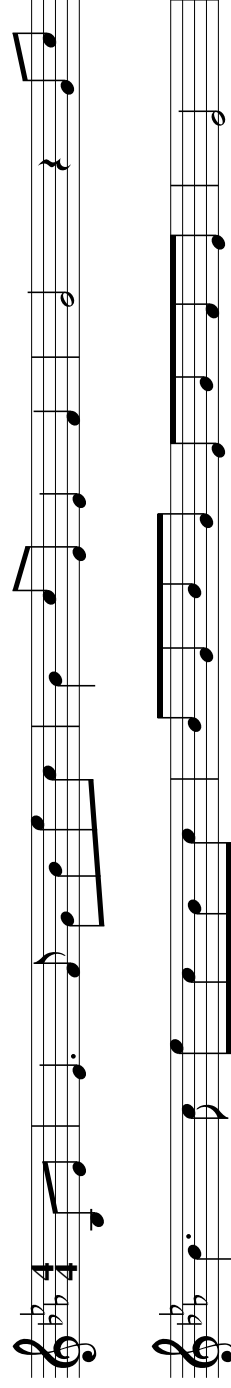


図 11. 4/4 拍子として誤認識した例 赤とんぼ (3/4 拍子)」

まとめと今後の課題

隠れマルコフモデルを用いた音符列・テンポ・拍節推定を提案

統一的な原理で各推定が可能であることを確認した

今後の課題

- 音響信号を対象とする自動採譜への拡張
- 様々なジャンルやスタイル毎の学習
- フレーズなどの大きな曲構造の反映
- ユーザ適応技術（音声認識における話者適応）
- リズムパターンによる音長変動の検討（文脈依存モデル）

などの検討を行い、適応可能性を広げることが課題である