Vol. 46 No. 4 情報処理学会論文誌 Apr. 2005

テンポ曲線と隠れマルコフモデルを用いた多声音楽 MIDI 演奏のリズムとテンポの同時推定

武田晴登[††] 西本卓也[†] 嵯峨山茂樹[†]

本稿では人間の音楽演奏を記録した MIDI (Musical Instrument Digital Interface) 信号から，演奏されたリズムとテンポを推定する手法を議論する．我々は，音楽演奏には次の 2 つの傾向が見られることに注目して最も尤もらしいリズムとテンポを推定する．(1) 演奏されるテンポは時間について連続で滑らかに変動する．(2) 演奏される曲のリズムは典型的なリズムパターンの組合せで表現される．テンポ曲線を仮定したとき，HMM(Hidden Markov Model, 隠れマルコフモデル) を用いて事後確率を増加させる音価列を推定することができる．また，リズムを仮定したとき，区分的に連続であるテンポ曲線を事後確率を増加させるように更新することもできる．本手法は，このようにリズムとテンポの推定を交互に行う反復アルゴリズムであり，適切な初期値から出発すれば，事後確率最大化の意味で最適解に収束し，さらにテンポが不連続な変化を伴う場合も扱うことができる．本手法を用いて，テンポが変動する人間の実演奏を記録した MIDI データ 37 曲に対して，81.9%〜85.5% の音価正解率を得た．

Joint Estimation of Rhythm and Tempo of Polyphonic MIDI Performance Using Tempo Curve and Hidden Markov Models

Haruto Takeda,[†] Takuya Nishimoto[†] and Shigeki Sagayama[†]

This paper discusses joint estimation of rhythm and tempo from a given musical performance recorded in the MIDI (Musical Instrument Digital Interface) format performed by a human player. We estimate the most likely pair of rhythm and tempo using 2 clues derived from general characteristics of musical performance: (1) tempo of musical performance changes smoothly over time, and (2) note sequence of performed rhythm consists of concatenation of typical rhythm patterns. Given a tempo curve, note values are estimated with HMM (Hidden Markov Model). Given note values, *a posteriori* probability is increased by reestimating the tempo curve. It is notable that our iterative algorithm converges to the optimum solution to maximize the *a posteriori* probability by alternately estimating the rhythm and tempo given an appropriate initial value, and also it allows discontinuous changes in tempo. Experimental evaluation gave a note value accuracy of 81.9–85.5% for 37 MIDI data performed by human players.

1. はじめに

本論文では，音楽演奏から，対応する楽譜の音価と演奏のテンポを推定する手法について論じる．主目的は自動採譜や音楽情報検索であるが，その派生として音色変換，自動伴奏，自動編曲，ジャンル分類，楽曲類似度計算，作曲家推定など極めて広い応用が期待される．入力としては，人間が演奏した MIDI(Musical Instrument Digital Interface) 規格の電子楽器出力やそれを記録した標準 MIDI ファイルを想定する．近年，音楽演奏のオーディオ信号から MIDI データへ変換する研究が進んでいる[1] ので，将来はオーディオ入力からの自動採譜も想定できる．

MIDI データとして与えられる音楽演奏のテンポは，未知で，かつ曲中で変動することが多い．さらに，フレーズ内など局所的には連続的に緩やかにテンポが変化する傾向を持つものの，フレーズ境界などでは不連続にテンポが変化することもある．後に述べるように既に自動採譜の研究は多くなされているが，このような実際的なテンポ変動への対処は十分ではなく，解

† 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo
†† 関西学院大学理工学部
Faculty of Science and Technology, Kwansei Gakuin University

の最適性への配慮も不足していた．そこで，本論文では，テンポが連続的あるいは不連続に変化する実演奏の MIDI データから，楽譜中の各音符の音価と共に，変動するテンポを同時に最適推定する問題を扱う．

ここで，楽譜の音価 (time value; 時価ともいう) とは，楽譜上の各音符には，四分音符，符点八分音符など，論理的な音の長さを指す．実際に楽譜が演奏されると，音価はあるテンポのもとで MIDI データの中で発音 (note-on) と消音 (note-off) の時刻に反映されて観測される．各音符の音高 (pitch) はすでに MIDI データに含まれるので，逆に，MIDI データから楽譜を復元するには，各音符の時刻情報から各音符の音価を推定すれば良い．本論文では，この問題を「リズム認識」と呼ぶことにする．ただし，強弱記号，表情記号，調号，楽器指定，声部の分離など，より広義の楽譜の復元は別の機会に譲って，ここでは音符の復元に限定して議論することとする．

演奏のテンポが一定で既知である場合，演奏される音長は変動するので音価の推定は自明な問題ではないが，過去の研究で，音長の変動を確率的に扱うことで音価が推定できることが報告されている．齋藤や大槻らは，音声認識[2] とリズム認識の同型性に基づいて HMM を用いて単旋律の MIDI 演奏の音長を変動をモデル化し[3),4)]．また，浜中らは多重音を含む演奏を扱った[5)]．これらの手法は，テンポが一定で既知である場合に限定されていた．

実際の音楽演奏においてはテンポが変動することが多いが，このテンポの変動を扱う研究として，著明な演奏家の演奏の比較を目的とした演奏解析の研究がなされている．このような研究では，動力学モデルによるテンポの記述[6)] や 2 次曲線へのテンポのフィッティング[7)] 等，時間の連続関数として扱われることが多い[8)]．また，時間に対して緩やかなテンポを仮定したときの拍打が厳密な拍の位置での拍打より自然と感じるという報告もある[9)]．しかし，これらの知見は自動採譜技術に活用されるまでには至っていない．本研究では，これらの先行研究にあるようにテンポを連続関数として扱うことにする．

さて，本論文のテーマと同様にリズムとテンポの両者を推定する過去の研究では，テンポ変動に関しては連続的変化のみが対象とされ，不連続変化は扱われて来なかった．いずれも，局所的にはテンポがほぼ一定で変化しないことを仮定して，テンポの推定を行なった．演奏の発音時刻間隔から規則によりテンポの仮説を求め追跡を行なう手法[10)] が提案されている．また，発音時刻毎のテンポをランダムウォーク[11)] やカルマンフィルタ[12),13)] でモデル化したり，あるいは，一小節毎に変動するテンポをマルコフモデル[3),4)] でモデル化する研究もある．更に，我々も過去にテンポが局所的にほぼ一定であると仮定され場合に，音長比 (リズムベクトル) を用いた HMM でリズム認識を行なう手法も議論した[14)]．テンポの変化が常に微小であることを用いたが，時間に対する不連続な変化をモデル化できていなかった．この中で，特に Cemgil[13)] と Raphael[11)] らの研究は，我々と同様に観測した演奏に対して尤もらしいテンポとリズムを求める確率的逆問題を扱っているが，モデルにおいてテンポとリズムの 2 変数についての最適化もしくは局所最適化が行なえなかった．Cemgil はリズムとテンポの同時確率をテンポについて積分して得られる周辺分布 (marginal distribution) を用いてリズムを求めたが，この周辺化の計算に Monte Carlo 法を用いているため，パラメータ数の増加に対して計算量が指数的に増える上に計算そのものが近似的であった．Raphael はテンポは実数値でモデル化しており，DP(Dynamic Programming, 動的計画法) を用いた計算を行なっているがテンポについての局所最適性は保証されていない．この他に，音価の推定に関しては規則に基づく手法[15)~17)] が報告されているが，実際に使用するには規則の整合性を保つために規則やコストの与え方の調整が必要であるため，その適用範囲が限られていた．

本論文ではこれまで著者らの研究グループが行なってきた HMM を用いたリズム認識を発展させ，音価を推定すると同時に時間に対して連続的に変化するテンポも併せて推定する手法について議論する．この中で，音価とテンポの反復推定が (局所) 最適解に収束することも示す．以下，第 2 章でテンポ曲線を用いた音長生成モデルを，第 3 章でテンポとリズムの同時推定を議論し，第 4 章で評価結果を報告する．

2. テンポ曲線と HMM に基づく音長の生成モデル

2.1 本論文の着眼点

まず，本論文の基本的な考え方を議論する．各音符が演奏された音長 (note length) を観測して，その記譜上の音符の長さである音価 (note value) と，演奏時のテンポ (tempo) の両者を推定する問題は，両者が相補的な関係にあるために解を一意に決定できない優決定問題である．テンポが分からなければ音符の音価は決定できないし，各音に意図された音価が分からなければテンポも決定できない．

例えば，**図 1**(a) に示すリズムパターンを徐々に遅

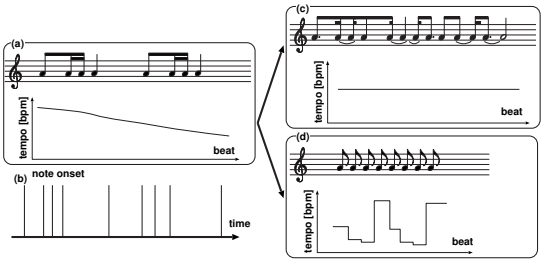


図 1 (a) と意図した演奏の発音時刻 (b) の別のリズムとテンポの解釈の例：不自然なリズム (c)，不自然なテンポ (d)
Fig. 1 Examples of interpretation of onset time patterns (b) intended as (a); unnatural rhythm in (c) and unnatural tempo in (d)

くなるテンポで演奏したときの発音時刻データ図 1(b) が与えられた場合を考えよう． これに対して，テンポは一定で図 1 (c) のような音価列として解釈することは可能であるが，符点リズムや短い音符がタイでつなげられたシンコペーションを多く含む楽譜は，通常は見かけない不自然なリズムである．また，逆に一音ごとにテンポが大きく変動してもよいとして図 1 (d) に示すように各音の音価がすべて等しいとする解釈も論理的には考えられるが，一音ずつテンポが大きく変動することは，我々が普段感じるテンポとして不自然である．

以上の例からも理解されるように，我々が図 1 (b) のような演奏を聴いて，演奏に対応するリズムとテンポの解釈として図 1 (c) や図 1(d) に比べて，図 1 (a) の方が妥当であると感じるのは，テンポは時間に対して緩やかに変動するものと了解し，かつ，音楽知識として持っている典型的なリズムパターンに当てはめてリズムを認識するからであると考えられる．すなわち，2 つの仮説:

仮説 1: 音楽リズムは常識的なパターンとして理解される

仮説 2: 音楽テンポは (ある区間内では) 緩やかに連続的に変化する

に基づけば，論理的には無限な可能なリズムとテンポの解釈の組み合わせから，最も妥当なリズムとテンポを同時に最適推定する問題として解決できる可能性がある．

但し，作曲者には如何様にも不自然な楽譜を書く自由があり，また演奏者も極端にテンポや音長を変動させて演奏する可能性があるので，リズム認識にはそのために原理的困難が残ることには，評価法や応用において留意する必要がある．

以下では，この着眼点からリズムとテンポから音長が生成される過程を確率モデルで定式化し，それに基づいて両者を推定する手法を論じる．将来は旋律等の音高情報もリズム認識に利用できるはずであるが，この論文の議論からは切り離しておく．

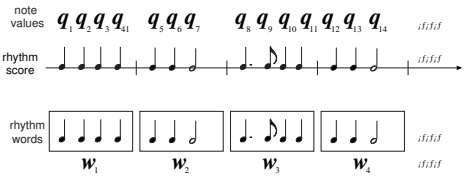


図 2 リズム譜のリズム単語によるモデル化
Fig. 2 Modeling the rhythm score with rhythm words

2.2 音価列生成の確率モデル

まず，多声楽曲のリズムの音価生成のモデルを議論する．我々の既発表の研究[14] と同様に多声楽曲のリズム譜を考えることで，単旋律のリズム認識手法[3] を応用できる．

2.2.1 音価とリズム譜

音価は，例えば四分音符を 1 拍として，拍と整数比関係にある離散的な量として扱うことができる．**図 2** に示すように，楽譜から音高情報を取り除いて音価のみの情報にした「リズム譜」において，第 n 番目の音符 (以後,「音符 n」と呼ぶ) の音価を q_n[拍] と表すと，このリズム譜は音価列 $Q = \{q_1, q_2, \cdots, q_N\}$ として表すことができる．この音価の並びはリズムパターンとして認知されるので，本稿では便宜的に音価列と「リズム」を同義として扱う．

2.2.2 リズム単語とリズム語彙のモデル化

対象とする音楽が音価列 Q からなる事前確率を $P(Q)$ とする．自然なリズムや不自然なリズムは $P(Q)$ の大小として表現できる．あり得るあらゆる Q に対して $P(Q)$ を個別に求めるのは困難だが，図 2 に示すように音価列 Q を小節などの単位に区切った「リズム単語」$w_m (m = 1, 2, \cdots)$ の列に分解し，それらの間の N-gram 確率の積により $P(Q)$ を近似することができる[3]．リズム単語の集合を「リズム語彙」と呼ぶ[3]．リズム語彙とその N-gram 確率は，既存の楽曲の楽譜から得られるリズム譜の統計から学習できる．音声認識における未知語の問題と同様に，学習に使用した楽曲に含まれないリズム単語は学習して得られるリズム語彙には含まれない．限られた量の楽譜からリズム語彙の確率モデルを頑健に学習する問題は，音声認識における言語モデルの学習と同様に扱えるが，その検討は本論文の主題とはせず機会を改めて論じることとし，本稿では未知語のない場合について議論する．

2.2.3 多声楽曲の音価とリズム語彙

多声音楽のリズムの情報は，**図 3** に示すように楽

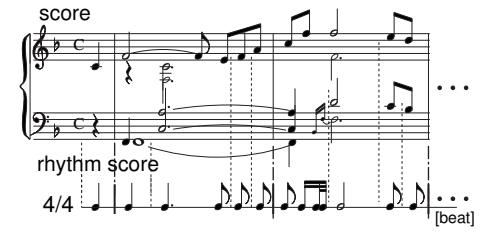


図 3 多声音楽のリズム譜
Fig. 3 Rhythm score obtained from multiphonic music

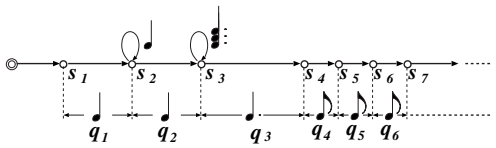


図 4 拍位置のマルコフ遷移によるリズム譜のモデル化
Fig. 4 Modeling rhythm score with Markov transition of beat position

譜中の音価情報を 1 次元の時系列に射影して得られる音価の時系列をリズム譜[14] として扱うことができる． このリズム譜に 2.2.2 節で議論したリズム語彙モデルを適用することにより，多声音楽についてリズムを学習し，リズムの事前確率 $P(Q)$ を与えることができる．

2.2.4 マルコフモデルによる拍位置の遷移のモデル化

ここで，楽譜における音符 n の位置 (単位:拍) は，音価の累積である拍位置 $s_n = \sum_{i=1}^{n-1} q_i$ で表わすことができる．リズム譜 Q に含まれる N 個の音符の拍位置をノードとする状態遷移ネットワークを考えると，リズム譜は拍位置のノードの遷移する時系列とモデル化することができる．過去の研究[5),11)] と同様に， 和音はその拍位置のノードにおける自己遷移に対応させることができる．例えば，図 3 に示す楽譜の冒頭部は**図 4** に示す状態遷移系列とモデル化できる．更に，拍位置の遷移をマルコフモデルと仮定すると，状態遷移確率によってその拍位置において同時に発音される音の個数の大小を表すことができる．

2.3 テンポ曲線によるテンポのモデル化

2.3.1 テンポ変動の曲線モデル

次に演奏のテンポの定式化を行なう．楽譜上で拍位置 s にある音符が 1 拍あたり $R(s)$[秒/拍] の音長で演奏されるものとする．一定速度の演奏では $R(s)$ は定数関数であるが，テンポが変動する一般の場合は $R(s)$ は曲線であり，これを「テンポ曲線 (tempo curve)」と呼ぶことにする．この定義は，テンポのメトロノーム表記 (bpm; beats per minute, 毎分の拍数) とは反比例の関係にある．$R(s)$ は，楽譜中のテンポ指定などに基づいて区分的に滑らかな曲線で表現され得るものと仮定する．テンポの緩やかな時間変化は時間の連続関数としてモデル化される．また，急激なテンポの変化や滑らかでないテンポの変化は連続関数の切替えによってモデル化することができ，その変化の起きる拍位置ではテンポ曲線は不連続となる．

以下の議論では，テンポ曲線の例として対数スケールでの単一多項式と区分多項式の 2 種類を扱うことにする．単一多項式

$$\log R(s) = \sum_{d=0}^{D} a_d s^d \tag{1}$$

は，演奏者はテンポを対数的なスケールで変動させるという仮定において，緩やかなテンポ変化を表すことのできる連続関数で，かつ最適化が容易に行なえるので，テンポ曲線として用いることにする．区分多項式は，この単一多項式が K 個に分けられた各区間で連続関数 $R^{(k)}(s)$ が切り替わるものと拡張したもので，区分的に連続であり不連続点を含むテンポ曲線を表現することができる．

$$\log R(s) = \log R^{(k)}(s) = \sum_{d=0}^{D} a_d^{(k)} s^d \tag{2}$$
$$s_{N^{(k-1)}} \le s < s_{N^{(k)}} \quad k = 1, \cdots, K$$

ここで，$N^{(k)}$ は k 番目の区間の開始となる音のインデックスを表し $a_d^{(k)}$ は k 番目の区間における多項式の係数を表す．ただし，$N^{(0)} = 1$, $N^{(K)} = N$ とする．

2.3.2 テンポ曲線の事前知識

テンポ曲線に対する事前知識として，R が出現する確率 $P(R)$ を考えることができる．これは大量のテンポ曲線の統計から学習して求めるか，あるいは，与えられた演奏曲のテンポについて事前知識が与えられいる場合に活用することができる．

2.4 音長と HMM

2.4.1 多声部間 IOI

音符 n に対応して音楽演奏中で観測された音の物理的長さを音長と呼び x_n [秒] と表記する．音長 x_n は，より正確には音の長さとして認知されるような物理的な連続時間量であり，音符の発音時刻の間隔 (IOI, inter-onset interval) として観測できる．たとえば同一音符のスタッカート演奏とレガート演奏では，音符の発音時間自体は異なるが，次の音符までの時間間隔は同一の音価を反映した長さになる．多声音楽の演奏の場合には複数の旋律 (声部) が同時に演奏されるが，

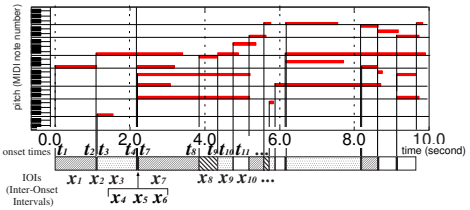


図 5 多声楽曲の演奏の声部間 IOI 系列
Fig. 5 An IOI sequence of performed music with a polyphonic structure

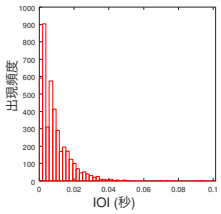


図 6 Burgmüller 「練習曲」全 25 曲の実演奏における同時発音の IOI の統計
Fig. 6 Statistics of IOIs of simultaneous note onset included in real performance of Burgmüller's "Etudes" 25 pieces

このときは**図 5** に示すように声部を区別せずに全ての音の発音時刻の間隔を取ることにする．このときの IOI を「声部間 IOI」と呼ぶことにする．声部間 IOI は，単旋律の IOI と異なり 2 音が同時に発音される場合に理想的には 0 となる IOI が含まれることがある．

2.4.2 IOI 変動の確率モデル

音符 n が拍位置 s_n に演奏されるとき，その音長 x_n は，音価 q_n とテンポ $R(s_n)$ の積になることが期待されるが，実際には演奏者のスキルを始め様々な要因によりそれから変動した値として観測されるであろう．ここでは，正規の音長が長くなる程その音長の伸縮の幅も大きくなると考えられるので，IOI は対数的スケールで変動すると仮定して，$\log x$ は平均 $\log(q_n \cdot R(s_n))$，分散 σ^2 の正規分布に従うと仮定する．すなわち，IOI $x \geq 0$ の変動の音価とテンポの条件付確率密度関数を次の対数正規分布

$$p(x_n|q_n, R(s_n)) = \frac{1}{\sqrt{2\pi\sigma^2}x_n} \exp\left(-\frac{(\log x_n - \log(q_n \cdot R(s_n)))^2}{2\sigma^2}\right) \quad (3)$$

で与えることにする．また，

$$x_n[\text{秒}] = r_n[\text{秒/拍}] \times q_n[\text{拍}] \quad (4)$$

として表して，r_n を音符 n の「瞬時テンポ (instantaneous tempo)」と呼ぶことにすると，式 (3) は，瞬時テンポ r_n の $R(s_n)$ からの確率的変動を与えるとして，

$$p(x_n|q_n, R(s_n)) = \frac{1}{\sqrt{2\pi\sigma^2}x_n} \exp\left(-\frac{(\log r_n - \log R(s_n))^2}{2\sigma^2}\right) \quad (5)$$

と書き直し，瞬時テンポのテンポ曲線からの変動確率と言い替えることもできる．

実際の音楽には，以上の音長変動のモデル化についての仮定が当てはまらない例外がある．例えば，フェルマータ (fermata) や極端なテヌート (tenuto)，あるいは一音のみを Adagio に指定した楽譜などは，演奏された音長から楽譜を復元する上で多様な解釈が可能で，多義性を解消することは困難である．本稿ではそのような例外は除外せざるを得ないが，上記の音楽演奏に関する仮定はかなり広い範囲で妥当と考えられる，

2.4.3 同時発音の IOI の確率モデル

和音の発音時刻は理想的は同時刻であるが，実際には厳密には同時に演奏されないので，和音を構成する音の間の IOI は必ずしも 0 でなく，**図 6** に示すように比較的小さい値に分布する．同時に発音されるべき 2 音が意図した時刻から独立に分散 ν^2 の正規分布に従って変動した時刻に発音されると仮定すると，この 2 音間の IOI は片側正規分布

$$p(x) = \begin{cases} 0 & x < 0 \\ \frac{2}{\sqrt{2\pi\nu^2}} \exp\left(-\frac{x^2}{2\nu^2}\right) & x \geq 0 \end{cases} \quad (6)$$

に従うことになる．この仮定のもとで 3 個以上の音からなる和音の各 IOI はより複雑な分布に従うが，簡単のためここでは式 (6) を和音の構成音の個数にかかわらず同時発音の IOI の確率分布を上式で与えることにする．

2.4.4 HMM によるリズム単語のモデル化

以上から，拍位置の遷移を表すマルコフモデル，IOI 変動の確率モデル，同時発音の IOI の確率モデルは，拍位置を各状態とし，IOI を出力信号とする HMM として統合することができる．**図 7** に示すようにリズム単語 w_m から声部間 IOI が生成される過程を表すことができる．この HMM により，リズム単語列によって表される音価列 Q をテンポ R で意図した演奏の IOI 系列 X である確率 $P(X|Q, R)$ を与えることができる．

2.5 演奏の生成確率

以上の生成モデルを統合することにより，あるリズム譜 Q がテンポ曲線 $R(s)$ で IOI 系列 X のように演奏される確率 $P(X|Q, R)P(Q)P(R)$ が与えられ

表 1 演奏生成の確率モデルの要素と HMM との対応
Table 1 Corresponding elements in a probabilistic model of music performance generation and HMM

音楽的要素	HMM の要素
IOI の変動	状態間遷移確率
和音の発音時刻変動	自己状態遷移出力確率
和音の発音数	状態遷移確率
リズムの出現	HMM 間の遷移確率 (N-gram 確率)

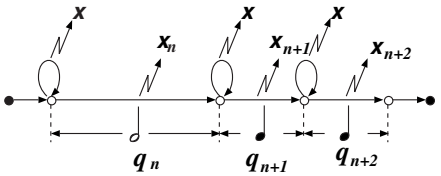


図 7 HMM によるリズム単語のモデル化
Fig. 7 Modeling a rhythm word with HMM

る．このときのモデルの各要素は，表 1 に示すように HMM のネットワークモデルと対応する．これにより，リズムが演奏される過程は，N-gram 文法で接続された HMM のネットワーク内で拍位置の状態を遷移しながら IOI を出力する確率過程とモデル化できる．

3. テンポ曲線と HMM を用いたリズムとテンポの推定

3.1 リズムとテンポの最大事後確率反復推定

音価とテンポの同時推定は，実演奏の IOI 系列 $X = \{x_1, x_2, \cdots . x_N\}$ から音価列 Q とテンポ曲線 $R(s)$ を推定する問題として定式化できる．音価列 Q とテンポ R が互いに独立と仮定すれば，事後確率は Bayes の定理により，

$$P(Q, R|X) \propto P(X|Q, R)P(Q)P(R) \qquad (7)$$

と書ける．これを最大化する Q と R の組を求めればよいが，これは一般には容易ではない．しかし，事後確率を単調増加させる音価 Q とテンポ R を交互に推定することができれば，事後確率は有界なので収束し，音価とテンポの両方に関する事後確率を極大化する解を求めることができる．従って，適切な初期条件を与えれば MAP 推定の最適解が得ることができる．

但し，この問題は本質的に多義性を持つ．例えば，ある楽譜の音価をすべて倍にし演奏速度も倍に指定した楽譜は，元の楽譜と各音の音長は等しいので，音楽としては同等である．このような楽譜の本質から，事後確率値には複数の極値を持ち得る可能性がある．

以下で，この反復推定として行なうリズム推定とテンポ推定について論じる．

表 2 音声認識とリズム推定の対応
Table 2 Analogy between CSR and rhythm estimation

モデル	連続音声認識	リズム認識	定式化
シンボル列	文	リズム譜	Q
N-gram	言語モデル	リズム語彙	$P(Q)$
HMM	音響モデル	IOI の変動	$P(X\|Q, R)$
特徴量	MFCC	IOI	X

3.2 テンポを仮定したときのリズム推定

まず，テンポ曲線 $R(s)$ が仮定されたとき，各音符の IOI 系列 X に対応する音価 Q を求めるリズム推定について議論する．

ここで扱う問題は，同時発音を除くと著者らの研究グループが以前に扱った単旋律のリズム認識と等価になる．この研究では，リズム認識の問題と連続音声認識の問題の同型性に着目し，音符を隠れ状態とし IOI を音符からの出力とする HMM を用い，観測された IOI 系列に対して最も尤もらしい音符列を推定する手法を提案した[3]．

同時に複数の音が発音される場合を含む多声音楽の場合のリズム認識も，HMM で構成される状態遷移ネットワークの経路探索問題となる．経路を探索することにより音価だけでなく同時発音として和音を構成する音のグルーピングも同時に求められる．連続音声認識と同様に one-pass DP 法 (Dynamic Programming, 動的計画法) による効率的な探索 (時間同期 Viterbi 探索) を行なうことができる[2] ので, 固定したテンポ R のもとで事後確率を最大にする音価列 Q を求めることができる．

以上により，テンポ曲線を与えられたときに音価列の更新により事後確率を単調増加させることができる．

3.3 リズムを仮定したときのテンポ推定

次に，リズム認識の結果として音価列 Q が与えられたときに，事後確率を増加させるテンポ曲線 $R(s)$ の推定について議論する．

事後確率においてテンポ曲線に関係する因子 $p(x_n|q_n, R(s_n))$ のみに注目して式 (5) を用いると，事後確率の対数は，

$$\begin{aligned}&\log P(X|Q, R)P(Q)P(R)\\ &= -\frac{1}{2}\sum_{n=1}^{N}\frac{(\log r_n - \log R(s_n))^2}{\sigma^2} + \log P(R)\\ &+\text{const.}\end{aligned} \qquad (8)$$

のように書き直せる．式 (8) から最大化する曲線の多項式の係数 a_k の最尤推定値を求める問題は最小二乗法と等価である．ここでは，テンポ曲線に関する事前知識 $P(R)$ が与えられない場合を想定し，$P(R)$ は一

様分布であると仮定し，以後の計算では用いないことにする．

式 (8) を増加させる単一多項式の更新は，式 (8) の a_d による偏微分を 0 とおいて得られる正規方程式

$$\sum_{d'=0}^{D}\sum_{n=1}^{N}\frac{{s_n}^{d'+d}}{\sigma^2}a_{d'}=\sum_{n=1}^{N}\log r_n\frac{{s_n}^{d}}{\sigma^2} \qquad (d=0,\cdots,D) \tag{9}$$

を解いて係数 a_d を求めることで行なうことができる．連続関数として多項式以外の関数を使用することは可能であり，冪乗以外に任意の基底関数の線形和で展開した場合も同様に正規方程式を解くことで最適化を行うことができる．

また，式 (8) を増加させる区分多項式の更新には，セグメンタル k-means 学習法[18] に準じた方法で行なえる． すなわち，K 個の区間の適当な初期境界から出発すれば，各区間内でのテンポ曲線 $R^{(k)}(s)$ の最適化と，それらに基づく最適な区分境界 $\{N^{(k)}\}_{k=1}^{K}$ を求めるセグメンテーションとを交互に繰り返せば，事後確率が単調増加するので，最適解に達する．但し，区間数 K は事前に知られているものとし，その自動推定問題は別の機会に論じたい．初期境界としては，たとえば $N^{(k)}$ を等間隔に配置して与えるなどが容易であり，以後の実験ではそのようにした．

以下に区分的なテンポ曲線推定の具体的なアルゴリズムを示す．

(1) 区間の初期配置：何らかの方法で初期境界を与える．
(2) 区間のテンポ曲線最適化: 境界を固定して，各区間のテンポ曲線のパラメータ値を事後確率最大化により推定する．具体的には (式 (9) による正規方程式を解く．)
(3) 最適境界位置の探索: 与えられた各テンポ曲線に対し，最適な境界を one-pass DP 法により求めて，修正する．
(4) 収束判定: 以上の事後確率の増加が，あらかじめ設定した値より小さければ，収束したものとして手順を終了する．そうでなければ (2) へ戻る．

以上により，音価列が与えられたときにテンポ曲線の更新により事後確率を単調増加させることができる．

3.4 リズムとテンポの同時推定のアルゴリズム

以上から，3.2 節で議論したリズム推定法と 3.3 節で議論したテンポ推定法を組み合わせて 3.1 節で述べたリズムとテンポの同時推定を最適に行なうことができる．すなわち，与えられた MIDI 演奏に対して以下の手順で推定できる．

(1) MIDI 演奏から多声部間 IOI を求める．
(2) 初期条件として，適当な初期テンポ曲線 $R(s)$ を与える．演奏曲のテンポに関する事前知識がない場合は，適当な一定テンポ R_c を与えることにする．
(3) **リズム認識**: テンポ曲線 $R(s)$ の仮定のもとにリズム語彙 HMM 中の Viterbi 経路を探索することにより，IOI 時系列 X に対して事後確率が最大になるような音価列 Q を求める．
(4) **テンポ推定**: 上で求めた音価列 Q の仮定のもとに，瞬時テンポ X/R に対する事後確率が最大になるようなテンポ曲線 (多項式あるいは区分多項式) を求め $R(s)$ を更新する．
(5) **収束判定**: 以上の事後確率の増加が，あらかじめ設定した値より小さければ，収束したものとして手順を終了する．そうでなければ (3) へ戻る．

なお，この反復アルゴリズムは，テンポと音価の確率分布 $p(x_n|q_n,R(s_n))$ によって与えられる x_n と $q_n\cdot R(s_n)$ の間の確率的距離尺度を，リズム語彙とテンポ曲線の制約のもとで小さくする操作をくり返すということに相当する．リズム推定とテンポ推定で共通の確率モデル $P(X|Q,R)$ を用いているので，各推定で全体で事後確率 $P(X|Q,R)P(Q)P(R)$ を単調増加させることができ，その結果，反復計算の収束性が保証される．すなわち，事後確率最大化の定式化の意味で，適切な初期値とテンポモデル区間数が与えられれば，この反復計算によってその最適解が得られる．

3.5 本手法の原理的限界

すでに 2.1 節で触れたように，ある楽譜の演奏からその楽譜を復元することには本質的に限界がある．たとえば，ある演奏がいかにも下手に聴こえても，実は故意にそのように書かれた楽譜を正確に演奏しているのかも知れない．両仮説を正確に区別することは原理的に不可能である．このことはパターン認識全般に共通する問題であり，たとえば音声認識において，言い間違いと意図的な発話を明確に区別することはできない．

我々は確率モデルを導入することにより，楽譜があるべき常識的なリズムパターンと演奏者による音長変動を統合して事後確率最大化の問題として定式化したので，このような曖昧な問題に対しても常識的な解を得る可能性を持つが，IOI 変動が，異なる音価を意図して演奏されたものか，演奏の表情付けやスキルの不足によって変動したものであるか正しく判断することは原理的に限界がある．これは，同時発音なのか短い音価を意図して演奏されたのかの区別でも同様である．

表 3 反復推定で使用する確率モデルのパラメータ値
Table 3 Model parameters in iterative estimation

パラメータ	値
テンポ曲線の多項式の次数 D	2
テンポ曲線のセグメント数 K	4
音長変動の対数正規分布 σ^2	0.1
同時発音の確率分布 ν^2	0.001 [sec^2]
反復推定の収束判定に用いる閾値	0.0001
テンポ曲線の初期値の一定テンポ R_c	120 [bpm]

表 4 演奏データ生成に用いたテンポ曲線の多項式の係数
Table 4 Coefficients of piecewise polynomial tempo curve for generation of experimental data

区間 k	係数 $a_0^{(k)}$, $a_1^{(k)}$, $a_2^{(k)}$
1	0.15, -0.2, -0.0001
2	-0.1, -0.1, 0.02
3	0.15, -0.2, -0.0001
4	-0.1, -0.1, 0.05

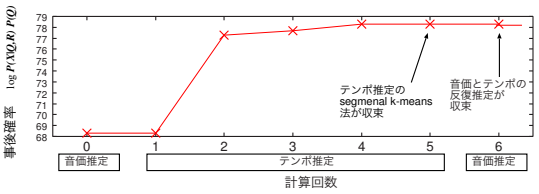


図 8 反復推定による事後確率の収束の例
Fig. 8 An example of converging MAP probability

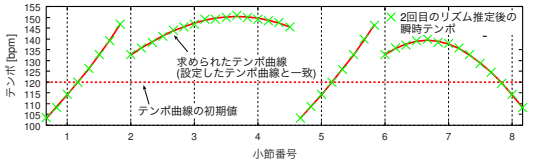


図 9 推定されたテンポ曲線
Fig. 9 Estimated piecewise polynomial tempo curve

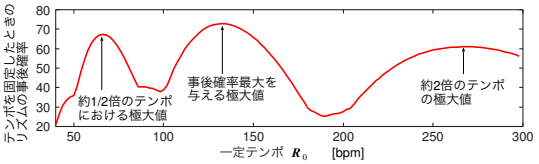


図 10 一定テンポを仮定したときの音価列推定後の事後確率値
Fig. 10 MAP probability of rhythm for given constant tempo

例えば，ショパンの幻想即興曲のように，左手で 8 分 3 連符を，右手で 16 分音符を同時に演奏する場合は，1/4, 1/12, 1/4 拍に相当する短い音価に相当する IOI が変動を持って観測される．このような複数の声部間で同期しない音を多く含む楽曲の演奏では，スキルの不足で不均一な IOI で演奏したのか，実際に不均一な音価を意図して演奏したのか，区別は難しい．これは，音高情報や旋律線に着目することにより解決できる可能性があるが，別の機会に論じたい．

4. 評価実験

提案手法を実装し，人工的に作成した MIDI データと人間の演奏を記録した MIDI データを用いてアルゴリズムの検証と性能評価を行なった．

4.1 テンポ曲線推定の実験検証

4.1.1 実験目的と実験条件

まず，テンポ曲線が本稿で述べたアルゴリズムで正しく推定できることを検証するために，テンポ曲線 $R(s)$ を与えた MIDI データからもとのテンポ曲線を推定し，設定したテンポ曲線に一致することを確認した．

ここでは，ベートーヴェン作曲「エリーゼのために」(WoO. 59) の冒頭 8 小節の楽譜を対象に，テンポ曲線のパラメータを音楽的に自然な演奏に聴こえるように表 4 のように設定し，自動生成した MIDI データを用いた．MIDI データの発音時刻は与えたテンポ曲線の情報を直接に反映し，音長の変動や和音の発音時刻のずれを含まないように，IOI が $x_n = R(s_n) \cdot q_n$ となるように生成した．

ピアノ曲 133 曲からリズム語彙を学習し，表 3 に示す条件で，一定テンポのテンポ曲線を初期条件として反復計算を開始し，提案手法により音価とテンポ曲線の推定を行なった．

4.1.2 テンポ曲線推定の実験結果

音価とテンポの反復推定は，**図 8** に示すように 2 回目で事後確率の収束が見られ，**図 9** に示すように作成したもとのテンポ曲線の係数と一致するテンポ曲線が推定できた．

また，3.1 節で言及した有理数倍のテンポ違いに相当する事後確率の極大値の存在を調べるために．反復推定の最初の音価列推定後の事後確率 $\max_Q \log P(X|Q, R_c)P(Q)$ を様々な一定テンポ R_c のもとで求めた．その結果得られた**図 10** は，事後確率を最大にするテンポの 2 倍や 1/2 倍のテンポにおいても事後確率は極大値となっている．このことは，およそ 2 倍や 1/2 倍のテンポでの極大値を与える解が存在し，初期条件によっては反復推定はそれらへ収束することを示している．

4.2 テンポトラッキング

4.2.1 実験目的と実験条件

次に，テンポトラッキングとしての性能を既存の研究と比較するために，Cemgil らによって公開されて

表 5 ベンチマークを用いたテンポトラックの評価結果]
Table 5 Accuracy of note value recovering and simultaneous onset detection with proposed method

	本手法	従来法 1	従来法 2
All Perfs.	90	92	90
Jazz	82	95	94
Amateur	95	92	92
Classical	93	89	86
Fast	88	94	93
Normal	90	92	92
Slow	91	90	87

従来法 1: Tempogram (non-casual)[12)]
従来法 2: Tempogram (casual)[12)]
従来法の評価値は文献[12)] による

表 7 リズム語彙に学習データ
Table 7 Training data for rhythm vocabulary

語彙名	単語数	曲数	学習に使用する曲
Open	622	100	ピアノ作品 100 曲
Closed	681	137	上記 + 評価に使用する 37 曲

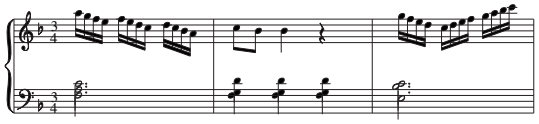


図 11 推定結果の例 (Burgmüller 作曲 「練習曲」より第 5 曲)
Fig. 11 An example of estimated score (Etude Op.100-5 by Burgmüller)

表 6 評価に用いた MIDI 演奏の楽曲
Table 6 MIDI-recorded music pieces used for evaluation of the proposed method

作曲者	作品名	曲数
Burgmüller	練習曲 Op. 100	25
Schumann	Kinderszenen Op. 15 (抜粋)	8
Chopin	Mazurka (抜粋)	4

表 8 提案手法による音価と同時発音検出の正解率 [単位：%]
Table 8 Accuracy of note value recovering and simultaneous onset detection with proposed method

リズム語彙	音価		同時発音	
	収束前	収束後	収束前	収束後
Closed	85.4	85.5	98.6	98.7
Open	81.8	81.9	98.3	98.3

いるデータセット☆を使用した．先行研究が The Beatles の “ Yesterday” の演奏を対象に評価を行なっているので，我々も同じ演奏を評価データに用いた．未知語のない状態での性能を評価するために，演奏曲 “ Yesterday” のリズムを含む The Beatles の 5 曲からリズム語彙を構築し，それ以外のパラメータは表 3 と同様のものを用いた．

4.2.2 テンポトラッキングの実験結果

Cemgil らの研究と同様の評価を行なうために，文献[12)] に従って推定結果と正解の beat time のそれぞれの時系列の距離を用いて評価を行なった．その結果は表 5 で，文献[12)] による評価値を比較をしたところ，テンポトラッキングについてはほぼ同程度の性能が確認された．一方，リズム推定は Cemgil らは行っていないので性能比較はできない．

4.3 楽譜の音価推定

4.3.1 実験目的と実験条件

次に，テンポが変動する音楽演奏から音価とテンポが推定可能であることを検証する為に，人間の電子ピアノによる演奏を記録した MIDI データ 37 演奏を用いて評価実験を行なった．この場合，正解のテンポ曲線は未知であるが，音価は楽譜の記載から得られるので，音価の復元による評価は行える．表 6 に示した楽曲を，2 名の演奏者のうち，1 名が Burgmüller を演奏し，もう一名は Schumann と Chopin を演奏した．

表 7 に示すデータを用いてリズム語彙を学習した．ただし，楽曲の中の装飾音は，これと等価と考えた音価に置き換えて学習に用いた．2 種類のリズム語彙のうち，Closed モデルは未知語の影響を受けずに推定アルゴリズムの性能を評価するために，また，Open は一般に未知のリズムパターンがある場合でもその中で適切なリズムを求めることができることを検証するために用意した．そして，テンポ曲線の区間数を $K = 10$ とし，それ以外の値は表 3 と同様にした．

4.3.2 音価推定の評価結果

音価正解率と同時発音正解率は，音高情報を手掛かりに推定結果と原楽譜を和音単位でマッチングを行い，このマッチングに基づいて音価の脱落・挿入・置換を数え，楽譜全体での倍テンポ違いも正解と見なして評価した．表 8 に示す評価データの平均正解率に見るように，提案手法が演奏曲から演奏曲の音価推定手法として有効であることが確認された．反復推定による音価正解率の改善が見られた．反復推定開始時にもある程度の音価が認識できているのは，リズム語彙による音価列の推定が可能な音価列の絞りこむに働きをしているからであると考えられる．

リズム認識により得られた楽譜を**図 11** に示す．ただしここでは，推定結果を見易くするための声部の分離と調性は人手で与え，また，休符は音符の継続の音価の量子化により推定した．

4.3.3 音価推定誤りの傾向と今後の展望

推定結果には，同時発音のグルーピングの誤りで，

☆ http://www.nici.kun.nl/mmm/archives /index.php ?archive=beatles

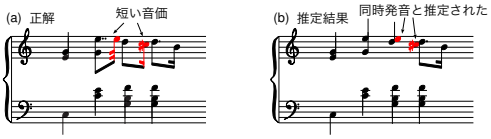


図 12 短い音価を同時発音と誤推定した例 (Burgmüller 作曲 “La chevaleresque”)
Fig. 12 Examples of short note values misrecognized into simultaneous onsets (“La chevaleresque” composed by Burgmüller)

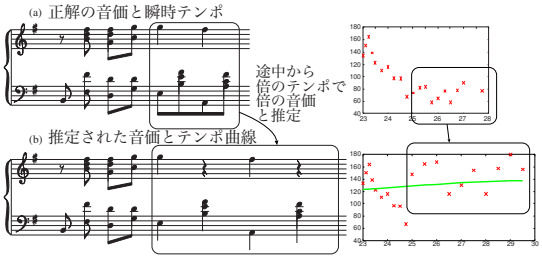


図 13 テンポを誤推定した例 (Schumann 作曲 “Fürchenmachen”)
Fig. 13 An expamle of tempo misrecognition (“Fürchenmachen” composed by Schumann)

図 12 に見られるように短い音価を同時発音と推定する傾向が見られた．また，いくつかの等しい長さの音価が連続する箇所では，3 連符の連続を 8 分音符の連続とする置換誤りのように局所的なテンポ推定誤りが観測された．また，長い区間に対して緩やかに変化するテンポ曲線を推定する結果として，局所的に大きくテンポが変化する場合に一致しないテンポ曲線を推定する場合が見られた．**図 13** のような例では，途中から有理数倍の異なる音価に誤推定されている．

以上の誤推定は，本研究で用いていない特徴を用いることで今後改善できると考えられる．装飾音は IOI が 40〜120 ミリ秒程度で演奏されており，同時発音か装飾音であるかの識別は難しいが，例えば継続時間長や後続音との時間的な重なりを調べること等を手掛かりに識別できる可能性がある．また，それ以外の音価の誤りも，テンポと音価の依存性や和音変化が拍や小節単位の境界に対応する可能性があることをリズム単語の特性としてモデル化することで改善できる可能性がある．

5. おわりに

本稿では，テンポが変動する MIDI 演奏からリズムとテンポを同時推定する手法を議論した．本研究では，演奏されるリズムは常套的なリズムパターンの組合せである確率が高く，かつ，テンポは時間に対して緩やかに変化するという音楽演奏に関する基本的な特性に着目し，確率モデルに基づいて事後確率を最大化する音価列とテンポを求める推定問題として定式化した．HMM を用いたリズム認識とテンポ曲線を用いたテンポ推定からなる反復計算により，リズムとテンポの同時最適推定できることを示した． また，テンポが変動する場合も，適切な区間数と初期解を与えれば，音価，区間ごとのテンポ曲線，区間境界が反復アルゴリズムにより同時に最適推定できることを示した． テンポが変動する実演奏を対象とした評価実験で 81.9%〜85.5% の音価正解率を得た．

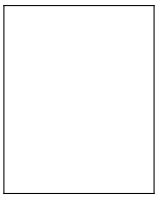
なお，今回は MIDI 演奏の各音符の発音時刻のみを用いた音価とテンポの推定に限定して論じたが，今後の課題として，旋律，和声，音量，くり返しパターン，等の情報も活用し，より高度な自動採譜の手法について検討して行きたい．

参 考 文 献

1) H. Kameoka, T. Nishimoto, S. Sagayama, “Audio Stream Segregation Based on Time-Space Clustering Using Gaussian Kernel 2-Dimensional Model,” Proc. of ICASSP, Vol.3, pp.5-8, 2005.
2) X.Huang, A.Acero, H-W.Hon, “Spoken Lanuage Processing: A Guide to Theory, Algorithm, and System Development,” Prentice Hall, 2001.
3) 齋藤, 中井, 下平, 嵯峨山, “隠れマルコフモデルによる音楽演奏情報からの音符列推定,” 平成 11 年度電気関係学会北陸支部連合大会講演論文集, F-62, pp.362, Oct. 1999.
4) 大槻, 齋藤, 中井, 下平, 嵯峨山, “隠れマルコフモデルによる音楽リズムの認識,” 情報処理学会論文誌, Vol.43, No.2, pp.245–255, 2002.
5) M.Hamanaka, M.Goto, H.Asoh, N.Otsu, “Machine Learning Techniques for Real-time Improvisational Solo Trading,” Proc. of ICMC, pp.439-446, 2001.
6) N. P. M. Todd, “The Dynamics of Dynamics: A model of Musical Expression,” Journal of Acoustical Society of America, Vol. 91, No. 6, pp.3540–3550, 1992.
7) B. H. Repp, “Diversity and Commonality in Music Performance: An Analysis of Timing Microstructure in Schumann’s “Träumerei”,” Journal of Acoustical Society of America, Vol.92, No.5, 1992.
8) H.Honing, “The Final Retard: On Music, Motion, and Kinematic Models,” Computer Music Jornal, Vol.27, pp.66-72, 2003.
9) S. Dixon, W. Goebl and E. Cambouropou-

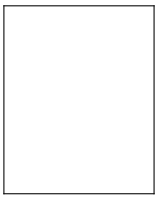
los,“Perceptual Smoothness of Tempo in Expressively Performed Music,” Music Perception, Vol.23, No.3, pp.195-214, 2006.
10) S. Dixon, “Automatic Extraction of Tempo and Beat from Expressive Performances,” Journal of New Music Research, Vol.30, No.1, pp 39-58, 2001.
11) C. Raphael, “Automated Rhythm Transcription,” Proc. of ISMIR, pp.99–107, 2001.
12) A. Cemgil, B. Kappen, P. Desain, H. Honing, “On tempo tracking: Tempogram Representation and Kalman filtering,” Journal of New Music Research, Vol. 28, No. 4, pp. 259–273, 2001.
13) A. Cemgil, B. Kappen,“Integrating Tempo Tracking and Rhythm Quantization by Sequential Monte Carlo,” Journal of New Music Research, Vol.18, pp.45–81, 2003.
14) 武田, 西本, 嵯峨山, “確率モデルによる多声音楽演奏の MIDI 信号のリズム認識,” 情報処理学会論文誌, Vol.45, No. 3, pp.670 – 679, 2004.
15) D.Rosenthal, “Emulation of Human Rhythm Perception,” Computer Music Journal, Vol.16, No.10, pp.64-76, 1992.
16) F.Lerdahl, R.Jackendoff, “A Generative Theory of Tonal Music,” MIT Press, 1983.
17) D. Temperley, “Cognition of Basic Music Structure,” MIT Press, 2001.
18) B.H. Juang and L.R. Rabiner, “The Segmental K-Means Algorithm for Estimating Parameters of Hidden Markov Models,” IEEE Trans. on Accoustics, Speech, and Signal Proc., Vol.38, No.9, pp.1639-1641, 1990.

(平成 16 年 11 月 28 日受付)
(平成 17 年 2 月 4 日採録)



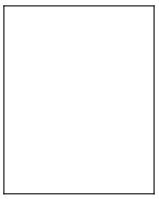
武田 晴登（学生会員）

2001 年慶應義塾大学理工学部卒. 2003 年東京大学大学院情報理工学系研究科修士課程了. 2006 年関西学院大学理工学部 CrestMuse プロジェクト研究員. 自動採譜, 自動伴奏の研究に従事. 情報処理学会, 日本音響学会各会員.



西本 卓也（正会員）

1993 年早稲田大学理工学部卒. 1995 年同大大学院理工学研究科修士課程了. 1996 年京都工芸繊維大学工芸学部助手. 2002 年東京大学大学院情報理工学系研究科助手. 音声インタフェース, 音声対話システムの研究に従事. 日本音響学会, 電子情報通信学会, 人工知能学会, ヒューマンインタフェース学会会員.



嵯峨山茂樹（正会員）

1974 年東京大学大学院工学系研究科計数工学専攻修士課程. 同年, 日本電信電話公社に入社, 武蔵野電気通信研究所にて音声情報処理の研究に従事. 1990 年 ATR 自動翻訳電話研究所音声情報処理研究室長として自動翻訳電話プロジェクトを遂行. 1993 年 NTT ヒューマンインタフェース研究所にて音声認識・合成・対話の研究開発に従事. 1998 年 北陸先端科学技術大学院大学情報科学研究科教授. 2001 年 東京大学大学院工学系研究科のち情報理工学系研究科教授. 博士 (工学). 1990 年発明協会発明賞, 1994 年 日本音響学会技術開発賞, 1995 年 情報処理学会山下記念研究賞, 1996 年 科学技術庁長官賞 (研究功績者表彰) および電子情報通信学会論文賞などを受賞. 日本音響学会, 電子情報通信学会, 情報処理学会, IEEE, ヨーロッパ音声通信学会 (ESCA) 各会員.