

DP マッチングのコストに視線情報を反映させた 楽譜追跡システムの構築

寺崎 栞里^{1,a)} 竹川 佳成^{2,b)} 平田 圭二^{2,c)}

受付日 2017年6月11日, 採録日 2017年12月8日

概要: 本稿では DP マッチングのコストに視線情報を反映させた楽譜追跡手法を提案する。楽譜追跡は自動伴奏や自動譜めくりの要素技術として多く研究されてきた。従来の楽譜追跡の多くは演奏箇所の情報を持つ演奏の音響信号や MIDI などの打鍵情報をもとに演奏箇所を推定している。しかし、直前とまったく違う箇所から演奏を再開したり、同じフレーズを含む楽曲を演奏したりしている場合には、打鍵情報のみからでは現演奏箇所や演奏再開箇所の推定が難しく、楽譜追跡に遅延や誤りが発生する。そこで、打鍵情報とは独立に演奏箇所情報を与える視線を楽譜追跡に統合させることで解決を試みる。視線情報を利用するにあたって、HMM によって視線の動きを予測するモデル（視線モデル）を作成することで、視線のノイズ低減の効果を狙う。また、推定結果から視線の尤もらしさ（視線尤度）を算出し、視線尤度を DP マッチングのコストに反映させることで誤打鍵に左右されない楽譜追跡を実現する。楽譜追跡における視線情報の有効性を検証するために、楽譜追跡に視線情報を反映させた場合と反映させない場合を比較したところ、楽譜追跡に視線を反映させると正解率が 1.2 倍（72.7%から 85.2%）、平均推定時間は 0.7 倍（1.37 秒から 0.94 秒）に短縮された。

キーワード：楽譜追跡, 譜読み, 視線, DP マッチング

Construction of Score-following System Reflecting Gaze Information on Cost of DP Matching

SHIORI TERASAKI^{1,a)} YOSHINARI TAKEGAWA^{2,b)} KEIJI HIRATA^{2,c)}

Received: June 11, 2017, Accepted: December 8, 2017

Abstract: The goal of our study is to build a score-following system employing keying information and gaze information. A score-following system, which estimates the user's current playing position on a musical score, is one of the fundamental technologies of real-time computer accompaniment. Conventional score-following systems have estimated the current performance position based only on keying information. However, this is difficult to do with a high degree of accuracy in the case where the player starts playing from a point on the score which is different from the previous playing position, and performs from a score including repeated phrases. In such cases errors and delays occur in score-following. Therefore, the proposed system solves this problem by using not only keying information but also gaze information, which expresses the player's thought. Noise is reduced by creating a model (gaze model) that predicts gaze movement. Also, we realize score-following which is not affected by errors, by reflecting the gaze likelihood calculated from the gaze model in the cost of DP matching. We compared the case of score-following using gaze information and not using gaze information. Consequently, the accuracy rate increased from 1.2 times, the average estimated time was reduced to 0.7 times.

Keywords: score-following, reading music, gaze, DP matching

¹ 公立はこだて未来大学大学院
Graduate School of Future University Hakodate, Hakodate,
Hokkaido 041-8655, Japan

² 公立はこだて未来大学
Future University Hakodate, Hakodate, Hokkaido 041-8655,
Japan

a) g2116028@fun.ac.jp

b) yoshi@fun.ac.jp

c) hirata@fun.ac.jp

1. はじめに

本研究の目的は、従来の打鍵情報による楽譜追跡結果に視線情報を反映させることで、譜面上で反復される音形の判別や演奏が再開される箇所（演奏再開箇所）の推定を可能とした楽譜追跡システムの構築である。楽譜追跡は自動伴奏や自動譜めくりの構成技術として多く研究されてきた [1], [2], [3], [11], [12], [13], [19]。従来の楽譜追跡の多くは楽譜追跡情報を持つ演奏の音響信号や MIDI などの打鍵情報をもとに演奏箇所を推定している。そのため、譜面上で反復される音形の判別や、演奏が一時中断された後に、直前の演奏箇所とまったく異なる箇所から演奏を再開された場合には楽譜追跡が難しい。このように打鍵情報のみからは演奏箇所や演奏再開箇所の推定が難しい状況においては楽譜追跡に遅延や誤りが発生する。そこで、本研究では打鍵情報とは独立に演奏箇所情報を持つ視線に着目し、打鍵情報による楽譜追跡結果に視線情報を反映させた楽譜追跡システムを提案する。

また、隠れマルコフモデル (HMM: Hidden Markov Model) によって、視線の動きを予測するモデル (視線モデル) を構築し、視線のノイズ除去を試みる。さらに、視線モデルから視線尤度を算出し、DP マッチングのコストに反映させることで、視線情報を反映させた楽譜追跡アルゴリズムを実装し、打鍵情報から演奏箇所の推定が難しい場合に対応した楽譜追跡を目指す。本研究では、楽譜追跡における視線情報の有効性について優先的に議論するため、打鍵情報からの楽譜追跡は比較的単純なアルゴリズムである DP マッチングを採用した。

本稿では文献 [17] のシステムを実装し、有効性や効果を検証する。

2. 関連研究

2.1 自動伴奏システム

Dannenberg の自動伴奏システムでは演奏と譜面との一致した音の数 (ベストマッチ) をコスト関数に設定し、Dynamic Programming (動的計画法, DP) マッチングによって単旋律演奏の楽譜追跡を実現している [2]。ベストマッチを求めていくことで 1 音程度の誤打鍵や弾き飛ばしに対応した楽譜追跡ができる。しかし、DP マッチングでは現在の推定位置から未来の譜面との照合を行うため、間違っただけのマッチングをした場合は楽譜追跡の回復が難しい。つまり、演奏の繰返しや大幅な跳躍、演奏箇所の弾き戻りに対応することが難しい。

そこで、Dannenberg らは複数の候補で DP マッチングを行う手法を提案している [3]。この手法では、楽譜追跡の結果が疑わしいとき、複数の候補を用意し、並列にマッチングを行う。そして、ある 1 つの候補で安定したマッチングを得ると複数の候補は除外され、最も確からしい候補を演奏

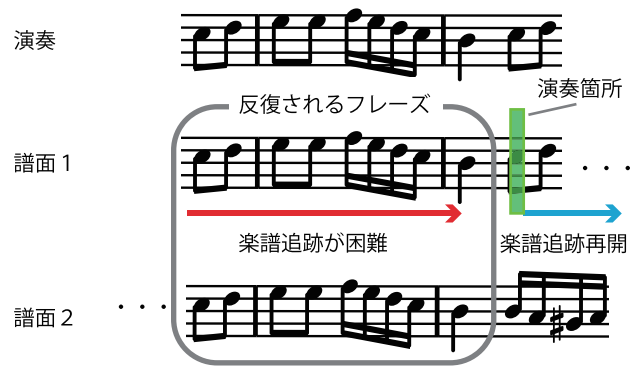


図 1 反復される音形を含む譜面に対して楽譜追跡が困難である例
Fig. 1 Example of difficult score-following in a musical score including repeated phrases.

箇所とする。このように、譜面全体を探索することで、演奏箇所の大幅な跳躍や繰返しへの対応を可能としている。しかし、この手法では反復される音形の判別は考慮されていない。

また、中村らの Eurydice は演奏生成を HMM によってモデル化し、Viterbi アルゴリズムを改良して最尤推定の計算コストを抑えて演奏箇所を探索することで、大幅な跳躍や弾き直しにも対応した高速な楽譜追跡を実現している [10]。さらに、中村らは同一の音形を含む楽曲において楽譜追跡の誤り率を評価している。演奏の跳躍を含む場合の誤り率について、同一の音形を含まない楽曲と比較すると、同一の音形を含む楽曲に対する誤り率が 3% 増加している。

このように、従来の自動伴奏システムは打鍵情報から演奏箇所を探索するため、譜面内に複数同一の音形が出現している場合、譜面に違いが現れるまで、どのフレーズが演奏されたかを即座に特定することが難しい (図 1)。同一の音形を含む楽曲は数多く存在し、練習環境においては譜面の最初から順に演奏されるとは限らないため、この問題を解決することが望ましい。

2.2 演奏情報を補うために映像情報を活用する研究

前節で述べたように、演奏情報のみを利用する楽譜追跡は同一の音形の特定や演奏再開箇所の推定が困難であるという問題がある。そこで、このような演奏情報の不足を補うために、映像情報を活用する研究がいくつか存在する。Maezawa ら [9] は、コンピュータと人間の合奏において、人間の演奏における発音タイミングの予測を試みている。演奏者のカメラ映像を利用することで、演奏を開始するときの予備動作を検出し、演奏の発音タイミングを予測している。また、Itohara ら [5] は人のギターの伴奏のような演奏テンポの揺らぎや裏拍ビートを含む多様なリズムに対するビートトラッキングが難しいことから、ギター演奏者の手をトラッキングすることで、ビート推定を行っている。

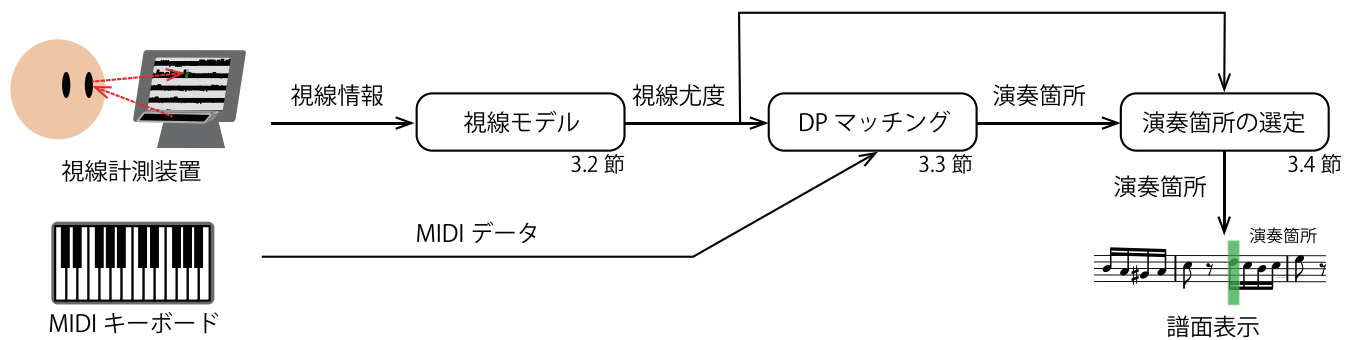


図 2 システム構成

Fig. 2 System configuration.

これらの研究から発音タイミングの推定に視覚情報が有効であることが示されている。しかし、これらの研究は合奏を前提としているため、譜面の弾き飛ばしや繰返しは考慮していない。

2.3 演奏と独立に演奏箇所の情報を持つ視線

安価な視線計測装置の出現により、手軽に視線情報が得られるようになり、視線を活用した研究がさかんに行われている。その中でも、演奏中の視線を対象とした研究がある [6], [7], [8]。川崎は演奏者の楽譜の認識の個人性を調査するため、演奏中の眼球運動を分析した [6]。この研究では読譜中の譜面上での演奏行為と注視点の位置、停留時間の関連性に着目している。その結果、初級者が初見で譜面を見ながら演奏する場合は、譜面上の音符を逐次的に1つずつ注視するが、練習を重ね、楽譜の構造を理解すると、初級者は音符をまとまりとして認識できるようになることが示されている。したがって、一般的に演奏者の視線は、小節単位あるいはフレーズ単位で遷移するといえる。しかし、音符単位でより細かく遷移する場合もある。

3. 視線を活用した楽譜追跡システム

3.1 システム構成

本提案は、DP マッチングのコストに視線情報を反映させた楽譜追跡の実現である。提案システムは実時間の演奏に対して、誤打鍵に頑健に対応する。

提案システムの構成を図 2 に示す。提案システムは画面上に楽譜を表示し、演奏者は提示された譜面を見ながら演奏する。入力として、MIDI データ (打鍵位置や打鍵強度) と視線情報 (位置座標) を取得し、楽譜追跡を行う。視線情報が得られている場合は、視線モデルを利用し、視線の生データから演奏者が見ている音符を推定することで、視線のノイズを除去する (3.2 節)。また、視線モデルの出力確率から算出される視線尤度を反映させた DP マッチングを行い、演奏箇所を推定する (3.3 節)。ただし、視線情報、打鍵情報のどちらか、または両方が検出されない場合は上記の楽譜追跡が難しいため、それぞれ取得している情報が

ら演奏箇所を推定する (3.4 節)。以上の過程で推定された演奏箇所を譜面に提示する。

視線計測装置として、EyeTribe [18] を用いる。EyeTribe は小型かつ据え置き型の視線計測装置であるため、演奏者の演奏を妨げない。また、EyeTribe は 30 Hz で視線情報のサンプリングを行う。

3.2 視線モデルの構築

視線の動きを予測する視線モデルの構築について述べる。人間はサッケードと呼ばれる高速な眼球の運動により、視対象を認識している。そのため、サッケードが発生するときや視線が一瞬視対象から外れたときは、ノイズのためにシステムが視対象を誤認識し、楽譜追跡の性能が低下してしまう。

そこで、視線のノイズを含んだ視線情報から実際の視対象を推定する手法として、画面領域を分割し、各分割領域を対象に割り当て、一定のフレーム内の連続したデータの注視点のうち、最も多く注視点が存在している領域を視対象とする方法が多くとられている [4], [14]。提案手法では、この手法を確率的に考えることで、視線のノイズの影響を緩和するとともに、ある音符を見ている尤もらしさを算出し、楽譜追跡に活用する。

演奏中の視線は最も細かく動く場合 (初心者が演奏する場合) では、音符単位で遷移する [6]。また、演奏者が熟達するにつれて、小節単位、あるいはフレーズ (複数の音符のまとまり) 単位で視認する [6]。つまり、演奏中の視線は離散的に遷移すると仮定できる。そこで、視線のノイズを含む視線の生データから、離散状態モデルである HMM によってノイズを除去した視線を予測するモデル (視線モデル) を構築する (図 3)。演奏者が視対象を視認しているとき、視線は視対象付近に多く出現する。ここで、視線が出現する位置座標の確率分布は明確になっていないため、正規分布に従うとする。ゆえに、正規分布に従う確率密度関数を出力確率とする連続型 HMM によって視線モデルを構築する。

文献 [17] では演奏者が音符をまとまりで読むという特

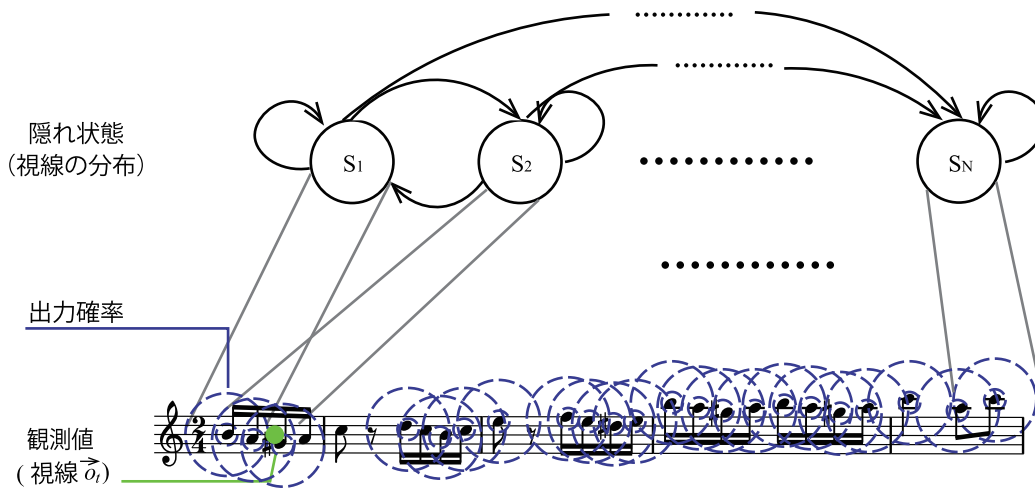


図 3 視線モデル
Fig. 3 Gaze model.

性を持つことから、音符のまとまりごとに状態があると仮定した。しかし、ヒューリスティクスにより視線情報と打鍵情報を組み合わせた楽譜追跡 [16] と比較し、正解率が 19.1% 低下してしまうという結果になった [17]。原因として、ある一定数の音符グループごとに状態を与えていたが、その音符の個数やグループの重なり程度が最適値ではなかったことが考えられる。これらの最適値は、楽譜に含まれる音符の個数や並び方、演奏者の技量などに関わってくるので、一概に決めることは現実的でない。そこで本提案では、主として初心者利用を想定し、最も細かい視線の動きに追従できるように、そして音符の個数や並び方に左右されないように、個々の音符ごとに状態を持たせることとする。

各隠れ状態 s_m から視線 $\vec{o}_i = [x_i, y_i]$ が観測される出力確率 $b_m(\vec{o}_i)$ は以下の式で定義される。

$$b_m(\vec{o}_i) = \frac{1}{(\sqrt{2\pi})^D \sqrt{|\Sigma_m|}} e^{-\frac{1}{2}(\vec{o}_i - \vec{\mu}_m)^T \Sigma_m^{-1} (\vec{o}_i - \vec{\mu}_m)} \quad (1)$$

ただし、 D は正規分布の次元数であり、視線モデルは視線の位置座標 (x, y) の二変数を出力変数とするため、 $D = 2$ とする。また、 $\vec{\mu}_m, \Sigma_m$ はそれぞれ状態 s_m から出力される位置座標の正規分布の平均ベクトルと共分散行列を表す。視線モデルは各音符ごとに状態を持つ HMM であるため、平均値は各音符の中心座標を与える。また、すべての状態に恣意的に同じ共分散を与えることで、推定の偏りを回避する。さらに、漏れなくノイズを吸収するため、分散は広く設定する。HMM の状態数を N とすると、各状態は各音符に対応するため、 N は譜面上の音符の数と等しい。このとき、初期状態確率 π 、状態遷移確率 $A = \{a_{ij}\}_{i,j=1}^N$ 、出力確率 $B = \{b_i(\cdot)\}_{i=1}^N$ により、HMM のパラメータは $\lambda = (\pi, A, B)$ と表せる。すべての状態を $S = \{s_1, s_2, \dots, s_N\}$ とする。また、重複を許す S の部分集合を S' とする。視線の位置座標 $O = \{\vec{o}_1, \vec{o}_2, \dots, \vec{o}_T\}$ が観

測されたときの最尤状態 \hat{s}'_T は以下の式で計算される。

$$\hat{s}'_T = \arg \max_{s'} P(O, S|\lambda)$$

$$P(O, S|\lambda) = \prod_{i=1}^T a_{s_{i-1}s_i} b_{s_i}(o_i) \quad (2)$$

ここで s_i は状態変数である。計 40 分のサンプル演奏における視線の動きを与え、Baum-Welch アルゴリズムによって学習し、状態遷移確率、初期状態確率を決定する。このとき、サンプル演奏には跳躍や繰返しが含まれた。

予測フェーズでは、実際の練習演奏における視線の動きを与え、Viterbi アルゴリズムによって隠れ状態（各音符）の遷移列を求める。したがって、跳躍や繰返しが生じたときにはその視線の動きからどの音符を演奏しているかを推定できる。

3.3 視線情報を反映させた DP マッチングによる楽譜追跡

視線情報を反映させた DP マッチングによって楽譜追跡を行う方法について述べる。一般の DP マッチングでは照合結果を一意に決めるため [2]、大幅な跳躍や弾き戻りや繰返しへの対応は困難である。そこで、譜面音列から打鍵された音高をすべて探索し、すべての譜面箇所候補を作成する。複数の候補に対してマッチングを同時に進める。演奏箇所の候補の信頼度が低くなってしまった場合、最も信頼度の高い候補を演奏箇所として選択すれば、即座に楽譜追跡の修正が可能である。

たとえば、各候補のマッチング結果が大幅に跳躍したとき、演奏者が意図する打鍵位置と近い鍵盤を誤打鍵した可能性があり、マッチング結果の信頼度は低い。そこで、マッチング結果の信頼度が低いとき、視線情報を DP マッチングに反映させる。DP マッチングのコストに視線情報を反映させることで、マッチングの際にスキップした音符

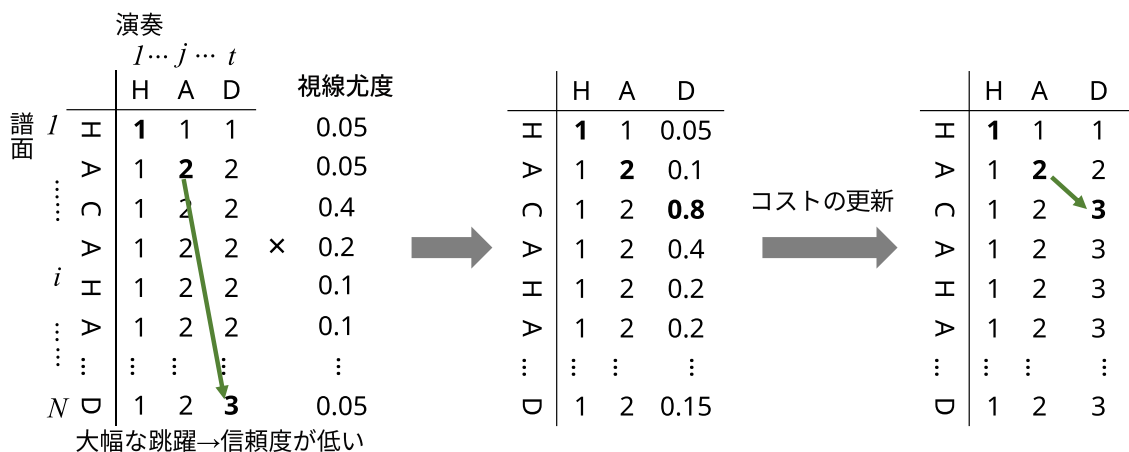


図 4 視線情報を反映させた DP マッチングの例

Fig. 4 Example of DP matching reflecting gaze information.

中に演奏箇所が存在した場合でも楽譜追跡が可能になる。

DP マッチングのコストに視線情報を反映させるために、視線モデルの出力確率 (式 (1)) をある座標を注視している尤もらしさ (視線尤度) と見なす。ここで、視線尤度の変化量が大きいと、DP マッチングのコストに掛け合わせたとき、視線による影響が大きくなりすぎてしまい、打鍵情報があまり反映されなくなる恐れがある。そこで、定数 c を式 (1) の指数部に掛け合わせて、各音符の中心座標における視線尤度 $\{L_i\}_{i=1}^N$ の変化量を調節する。定数 c は $0 < c \leq 1$ の値をとり、値が小さければ小さいほど、楽譜追跡における視線情報の優先度が低くなる

$$L_i = \frac{1}{(\sqrt{2\pi})^D \sqrt{|\Sigma_m|}} e^{-\frac{1}{2} c (\hat{\sigma}_i - \mu_m)^T \Sigma_m^{-1} (\hat{\sigma}_i - \mu_m)} \quad (3)$$

DP マッチングのコスト (視線尤度を反映させていない) を $\{C_{ij}\}_{i,j=1}^{i=N,j=t}$ (t はその候補が出現してから、現在まで入力された音の個数)、直前に入力された視線に対する視線尤度を掛け合わせた DP マッチングのコストを $\{\gamma_i\}_{i=1}^N$ とすると視線尤度を掛け合わせた DP マッチングのコストは以下の式で表せる。

$$\gamma_i = C_{it} \cdot L_i \quad (4)$$

視線尤度を掛け合わせたコストをそのまま採用すると、値が極端に小さくなってしまい、視線尤度を掛け合わせなかった場合の DP のコストと比較することが難しい。そこで、 $\max(\gamma_i)$ をベストマッチとして、コストを更新する。ただし、DP マッチングの結果が信頼できる場合には、視線尤度を反映させず、DP マッチングの結果をそのまま用いる。

DP マッチングの結果を信頼する条件を説明する。本研究における DP マッチングでは、1 音から 2 音の跳躍や繰返しには DP マッチングのみで十分に信頼できる楽譜追跡ができる。そのため、隣接する次の音が正しく打鍵された場合のみ DP マッチングによる楽譜追跡の結果を信頼する

のではなく、ある程度のスキップを許容する必要がある。そこで、任意で 2 音から 10 音の跳躍を許容し、DP マッチングの結果を信頼する。

図 4 に視線尤度を反映させた DP マッチングの例を示す。譜面 H - A - C - A - H - A - ... - D に対して、H - A - D が演奏された状況を考える。演奏音から DP マッチングを行うと、3 音目の D が演奏されたとき、大幅な跳躍が発生する。このとき、演奏が進行しなければ、マッチングの確かさを判定することは難しい。そこで、ベストマッチを求める行列に視線尤度を掛け合わせる。すると、視線尤度の最も高い (演奏者が演奏している可能性が高い) 3 音目の C の音が尤もらしいと推定される。そこで、3 音目の C の音をベストマッチとしてコストを更新することで 3 音目の C の音を演奏箇所として選ぶことができる。

このように候補のマッチング結果の信頼度が低い場合には視線尤度を掛け合わせることで打鍵情報のみからでは楽譜追跡が難しい場合でも、頑健に楽譜追跡ができる。

3.4 演奏箇所の選定

視線情報と打鍵情報がともに検出されている場合には 3.3 節で述べたアルゴリズムによって楽譜追跡できる。しかし、視線情報、打鍵情報のどちらか、または両方が検出されなかった場合には上記アルゴリズムの適用は難しい。そこで、上記アルゴリズムが適用できない場合の処理について考える。それぞれの処理について以下に示す。

- (1) 打鍵はされているが、視線が取得できない場合、手元を見ながら譜面を見ずに演奏している可能性が高い。この場合、視線尤度を無効化すれば、3.3 節のアルゴリズムを適用できる。しかし、視線情報が反映できないため、演奏箇所が 2 カ所以上推定される可能性がある。このとき、演奏箇所を一意に確定するのに十分な根拠がないとして、演奏箇所を 1 つに選定することはしない (図 5 中 (1))。演奏者が再び譜面を視認すれば、

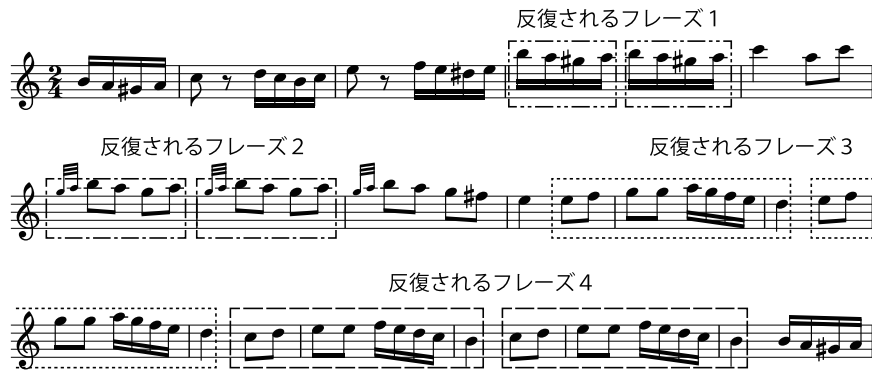


図 6 モーツァルト作曲トルコ行進曲 K.331 第 3 楽章の冒頭 16 小節
 Fig. 6 The first 16 bars from the opening of Turkish March K.331.

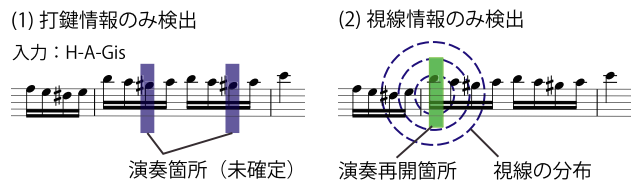


図 5 視線情報、打鍵情報のどちらかが検出されないときのシステムの動き

Fig. 5 System behavior when either gaze information or keying information is not detected.

即座に演奏箇所を決定できる。

- (2) 打鍵がなく、視線のみを取得している場合、打鍵と打鍵の合間であり、直後に打鍵される可能性が高いため、打鍵情報が入力されるまで待機する。ただし、1 秒以上打鍵されなかった場合は、演奏者が演奏を一時中断している可能性が高い。そこで、演奏再開に備えて、視線によって演奏再開箇所を推定する。視線尤度の最も高い位置にある音符を見ているとして、その音符を演奏再開箇所と推定する (図 5 中 (2))。
- (3) 視線と打鍵がともにない場合、演奏者が休符の間に一瞬手元を確認していたり、演奏以外のことに気をとられている可能性が高い。また、演奏箇所を補正する根拠が得られないため、視線情報か打鍵情報が入力されるまで待機する。

4. 提案アルゴリズムの妥当性の検証

本章では提案アルゴリズムの有効性を段階的に評価する。

4.1 実験

楽譜追跡における視線情報の有効性を検証するために、楽譜追跡に視線情報を反映させた場合 (以後、提案手法と呼ぶ) と反映させない場合 (つまり、DP マッチングのみ、以後、比較手法と呼ぶ) の正解率と平均推定時間を比較する。ここで、正解率は、連続するすべてのオンセット時間に対し、1 ミリ秒単位で一致している時間の割合とする。また、平均推定時間は誤打鍵が発生してから、正しく演奏

箇所を推定するまでの時間とする。平均推定時間が短いほど、高速に演奏箇所の推定ができたといえる。

入力とするデータセットを構築するために、譜面を読みながら演奏できるピアノ初級者以上 7 名に協力を依頼した。視線計測装置を取り付けたディスプレイに、モーツァルト作曲トルコ行進曲 K.331 第 3 楽章の冒頭 16 小節を全体表示し、主旋律を単音で演奏させた。図 6 に示すようにトルコ行進曲は冒頭 16 小節中に反復される音形が 4 組存在し、演奏箇所の同定が難しい譜面である。このときの打鍵情報 (打鍵位置、打鍵強度) と視線情報 (位置座標) を記録し、データセットを構築した。実験中に被験者が外部の刺激や興味が演奏以外に反れてしまうことを極力抑制するために、実験を行う空間には MIDI キーボード、視線追跡装置を取り付けたディスプレイ、パソコン、実験者のみが存在する状況にした。実験は、基本的に譜面を読みながら演奏してもらうが、手元が不安な場合は確認してもよいこととした。

繰返し、跳躍、間違いがなく、正しく演奏された場合、演奏情報のみから十分に打鍵追跡が可能である。そのため、システムは視線を利用せず、打鍵情報のみから楽譜追跡を行う。したがって、正しい演奏においては比較手法と変わらない性能で追跡できることが明らかであるため、本実験では誤打鍵・跳躍・繰返し (以後、これらをまとめて異打鍵と呼ぶ) を含んだ演奏に対する楽譜追跡の性能を検証する。本実験では異打鍵を含んだ演奏に対する楽譜追跡の性能を検証するため、演奏には意図的に異打鍵を含めるように教示した。ここで、誤打鍵とは音の置換や挿入、1 音から 2 音の欠落のことを示し、跳躍は小節をまたぐ跳躍、繰返しは譜面を戻って演奏することとする (図 7)。演奏は意図的な異打鍵に関する指示を変えて、6 回演奏させた (表 1)。指示は毎回、演奏の直前に口頭で伝えた。演奏に含めることを指示した異打鍵も含め、結果的に表 2 に示すようなデータセットが得られた。

構築したデータセットを利用し、提案手法と比較手法で楽譜追跡の結果を得たあと、それぞれ正解率と平均推定時



図 7 (a) 誤打鍵, (b) 跳躍, (c) 繰り返しの例
Fig. 7 Example of mistake, skip, and repeat.

表 1 演奏への指示
Table 1 Instructions for performance.

回数	演奏への指示
1	通常の演奏
2	誤打鍵 (8 カ所)
3	誤打鍵 (8 カ所)
4	跳躍 (4 回, 同じ音形の同定を要する)
5	跳躍 (4 回, 同じ音形の同定を要さない)
6	繰返し (3 カ所 × 5 リピート)

表 2 各演奏に含まれた異打鍵の数 (音符数)

Table 2 The number of uncertain factors included in each performance.

No.	正打鍵	誤打鍵	跳躍	繰返し
1	564	14	10	12
2	599	54	7	31
3	492	69	8	27
4	552	21	9	11
5	505	31	8	17
6	644	30	12	20
7	550	25	11	16

間を算出した. このとき, 定数 c (3.3 節) は 0.6 に設定した.

4.2 結果と考察

4.2.1 正解率と平均推定時間の評価

表 3 に提案手法と比較手法の正解率と平均推定時間の比較を示す. 演奏者が譜面を読みながら演奏しているか確認するために表 3 中に視線認識率を示した. 視線認識率は 1 回の演奏が開始されてから終了するまでの間で画面上に視線が存在した時間の割合である. 演奏者が演奏中に譜面を視線認識率の平均は 66.1% であり, 演奏者が演奏中に譜面を見ていることが分かる. また, 楽譜追跡に視線を導入すると正解率が約 1.2 倍 (72.7% から 85.2%) 上昇するこ

表 3 提案手法と比較手法における正解率と平均推定時間の比較
Table 3 Accuracy rate and average estimation time of score-following in the comparison method and the proposed method.

No.	視線認識率 (%)	正解率 (%)		平均推定時間 (秒)	
		提案	比較	提案	比較
1	73.0	90.3	78.4	1.05	1.46
2	72.8	81.4	66.7	0.89	1.52
3	38.1	84.1	64.8	1.07	1.71
4	70.5	86.4	75.2	0.68	1.14
5	83.4	81.1	73.5	1.01	1.22
6	68.6	89.4	75.9	0.8	1.25
7	86.3	84.7	76.8	1.06	1.27
平均	66.1	85.2	72.7	0.94	1.37
標準偏差	14.5	3.3	4.8	0.1	0.2

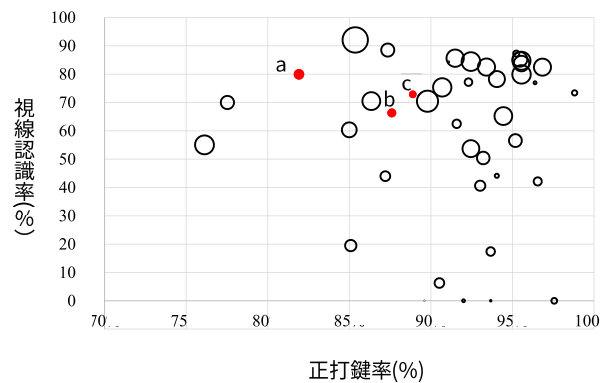


図 8 提案手法と比較手法における視線認識率と正打鍵率に対する正解率の差分. 丸の大きさがその差分を示す

Fig. 8 Difference in accuracy rate between the proposed and comparison method.

とが分かる. 分散の等質性を満たすために正解率について角変換を行い, 有意水準両側 1% において Welch の t 検定を適用したところ, 有意な差が認められた ($t(40) = 7.31$, $p < .01$). また, 平均推定時間は約 0.7 倍 (1.37 秒から 0.94 秒) に短縮されることが分かる. 同様に平均推定時間について有意水準両側 1% における Welch の t 検定を適用したところ, 有意な差が認められた ($t(40) = -5.56$, $p < .01$). このことから楽譜追跡に視線を導入することでより頑健で高速な楽譜追跡が可能になったといえる.

4.2.2 視線認識率と正打鍵率に対する正解率の差分の評価

この結果を詳しく分析するために, 視線認識率と正打鍵率に対する正解率の差について考察する. 図 8 に視線認識率と正打鍵率に対する正解率の関係性を示す. 図 8 中の円の直径は提案手法と比較手法の正解率の差を表す. つまり, 大きい円であるほど視線情報によって楽譜追跡精度が向上している. すべての実験データにおいて比較手法より提案手法の正解率が高かった. 視線認識率と正解率の関係を見るために, ピアソンの相関分析を行ったところ, 正の相関が認められた ($r = .52$, $p < .01$). つまり, 視線認識

表 4 図 8 中の赤い丸の提案手法における正解率と平均推定時間

Table 4 Correct keying rate and the average estimated time of the data represented by red circles in Figure 8 in the proposed method.

	正解率 (%)	平均推定時間 (秒)
a	90.5	0.83
b	88.3	0.85
c	85.1	0.67

表 5 各異打鍵における演奏箇所の推定に要した音符数

Table 5 The number of notes to estimated the performance position in each uncertain factor.

異打鍵の種類	提案		比較	
	平均	標準偏差	平均	標準偏差
誤打鍵	2.0	2.1	2.6	2.8
跳躍	2.5	1.9	4.7	3.6
繰返し	2.1	1.4	2.7	2.0

率が高いほど、正解率の差（単位はパーセント）が大きくなった。

図 8 中に赤い丸で示したように、視線認識率が平均以上、かつ、正打鍵率が 90%以下で、正解率の差分が 10%以下のデータがある。これは、視線によって楽譜追跡精度の向上があまり見られなかったことを示す。これらの赤い丸のデータを見ると、演奏情報のみから十分に推定が可能である装飾音符などの 1 音から 2 音程度の跳躍が頻出していた。また、表 4 に示すとおり、これらのデータの正解率は平均以上である。つまり、視線情報を楽譜追跡結果に反映させなくても、DP マッチングのみで十分な楽譜追跡結果が得られているため、視線情報による楽譜追跡精度の差があまり見られない。

4.2.3 異打鍵が発生時における推定時間の評価

次に、異打鍵の種類ごとに楽譜追跡精度を検証する。表 5 に、異打鍵が発生してからシステムが正しく演奏箇所を推定するまでに要した音符数を示す。このとき、異打鍵の種類ごとの特徴を分析するために、次の異打鍵が発生するまでに、演奏箇所を正しく推定した場合のみを対象とした。表 5 から、比較手法では跳躍が発生したとき、他の異打鍵と比較して推定に時間を要することが読み取れる。これは、反復される音形への跳躍が多く含まれ、演奏箇所の推定が困難になったためである。一方で、提案手法では比較手法よりも異打鍵の種類ごとの差が小さいため、異打鍵の違いにほとんど左右されない楽譜追跡が実現されている。比較手法と提案手法における跳躍が発生したときの演奏箇所の推定に要した音符数を比較すると、比較手法では 4.7 音要するのに対して、提案手法では 2.5 音で推定しており、2.2 音早く推定している。この差について有意水準両側 1%における Welch の t 検定を適用したところ、有意な差が見られた ($t(96) = 3.92, p < .01$)。つまり、比較手法では反復

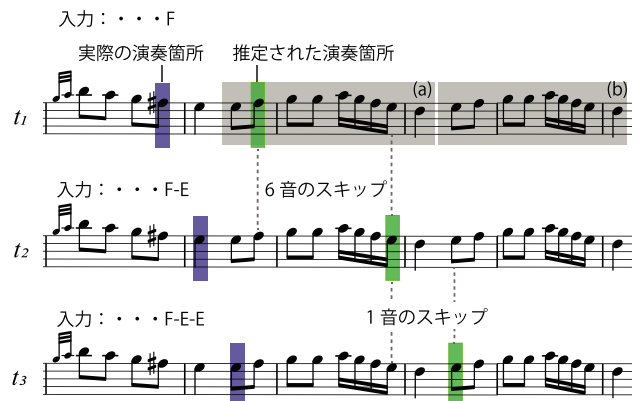


図 9 提案手法において楽譜追跡がうまく動作しない例

Fig. 9 Case in which the proposed method fails.

される音形の判別は難しいが、提案手法では反復される音形の判別が容易になったといえる。

また、表 5 中の誤打鍵と繰返しはともに 0.6 音差であり、比較手法と大きな差が見られなかった。つまり、誤打鍵と跳躍に関しては視線情報による楽譜追跡の向上があまり大きくないことを示す。しかし、表 5 では連続して異打鍵が発生したときや、ある異打鍵が発生して正しく演奏箇所が推定される前に次の異打鍵が発生した場合は除外している。そのため、表 5 の計算から除外された場合については検証していない。

4.2.4 提案手法がうまく動作しなかった状況

演奏箇所の推定が難しかった異打鍵の処理例を時系列で示す(図 9)。図 9 のように、反復される音形の直前で挿入や置換された音が、1 回目の反復される音形 (a) にマッチングしてしまうと、演奏者が続けて 1 回目の反復される音形 (a) を演奏したとき、2 回目の反復されるフレーズ (b) とマッチングしてしまう場合があった。本研究における DP マッチングは 1 音から 2 音の跳躍について対応できるほか、繰返しについても DP マッチングのみから信頼度の高い楽譜追跡結果が得られる場合がある。このように、視線よりも DP マッチングが信頼度が高い場合は DP マッチングの結果を信頼し、そのまま利用したい。そこで、任意の音符数のスキップを許容している。しかし、図 9 に示した例のように、スキップの許容によって、視線情報を反映させると正しく演奏箇所の推定ができる場合でも視線情報が楽譜追跡結果に反映されず、提案手法がうまく動作しない場合がある。この問題を解決するために、スキップを許容する音符数を小さくすると、DP マッチングの結果が信頼できるにもかかわらず、視線情報を楽譜追跡結果に反映させてしまい、楽譜追跡の精度が低下する可能性がある。このように、打鍵情報と視線情報の優先順位はトレードオフの関係にあり、一意に優先順位を決定することは難しい。今後、この優先順位の課題に取り組むことが重要である。

提案手法がうまく動作しなかったのは以下のような場合

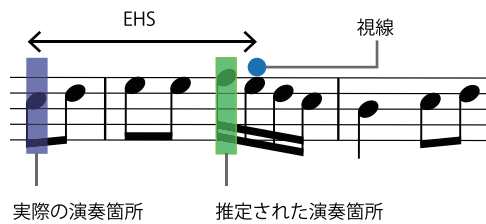


図 10 EHS の影響により、数音符先の譜面にマッチングしてしまう現象

Fig. 10 Phenomena matching several notes ahead of the performance position because of EHS.

である。提案手法では、打鍵情報から導かれる楽譜追跡結果の信頼度が低いとき、楽譜追跡結果に視線情報を反映させるが、実際の演奏箇所よりも未来の譜面箇所でもマッチしてしまうという現象が見られた。DP マッチングは過去の譜面とのマッチングをしないアルゴリズムであるため、誤って未来の譜面でもマッチングしてしまった場合、楽譜追跡の回復が難しい。この原因として、演奏者の譜面の先読み幅（EHS：Eye-Hand-Span）の未考慮が考えられる。演奏者は熟達度が高くなればなるほど、EHS が大きくなる [15]。しかし、提案手法は個人の EHS の違いを考慮していないため、EHS の大きい演奏者の場合には、実際の演奏箇所よりも未来の譜面箇所とのマッチングが頻発し、楽譜追跡精度が下がってしまうことが考えられる（図 10）。そこで、EHS を考慮した楽譜追跡を実現するためには、演奏者の熟達度や譜面の難易度に対応して視線モデルのパラメータを設定することが必要である。ここでパラメータとしては、ひとまとまりに含まれる音符の個数、注目している音符のまとまりが視線の動きに沿って移動する際の重なりなどが考えられる（3.2 節）。

5. おわりに

本稿では跳躍や繰返しに対応するとともに反復される音形の判別や演奏再開箇所の推定を可能とするために、DP マッチングのコストに視線情報を反映させた楽譜追跡システムを提案した。課題として、視線のノイズの低減や視線情報を反映させる手法の検討があり、HMM によって視線の動きを予測する視線モデルを構築し、出力確率を視線尤度として DP マッチングのコストに掛け合わせることで解決した。楽譜追跡における視線情報の有効性を示すために、視線情報を導入した場合と導入しない場合で楽譜追跡の誤り率を評価したところ、正解率は 1.2 倍（72.7% から 85.2%）に上がり、平均推定時間は 0.7 倍（1.37 秒から 0.94 秒）に短くなった。また、異打鍵の種類ごとに正しく演奏箇所を推定することに要した音符数を検証し、視線情報を反映させた楽譜追跡が大幅な跳躍において有効であることを示した。

さらに楽譜追跡の性能を向上させるために演奏者の EHS

の考慮が考えられる。演奏者は熟達度によって EHS が異なる [15]。しかし、現状では各演奏者に合わせた EHS を考慮していないため、EHS の大きい演奏者が誤打鍵を頻発した場合、演奏箇所の視線尤度が小さくなり、楽譜追跡の性能が低下することが考えられる。そこで、EHS を考慮した楽譜追跡を実現するためには、演奏者の熟達度や譜面の難易度に対応して視線モデルのパラメータを設定することが必要である。ここでパラメータとしては、ひとまとまりに含まれる音符の個数、注目している音符のまとまりが視線の動きに沿って移動する際の重なりなどが考えられる（3.2 節）。

また、打鍵情報と視線情報の優先度はトレードオフの関係にあり、一意に優先順位を決定することは難しいため、優先順位の課題に取り組む。

謝辞 評価法について適切なアドバイスをしてくださった寺井あすか先生（公立はこだて未来大学）に感謝いたします。本研究は JSPS 科研費 JP16K12560, JP15K00279 の助成を受けたものです。

参考文献

- [1] Cont, A.: ANTESCOFO: Anticipatory Synchronization and Control of Interactive Parameters in Computer Music, *Proc. International Computer Music Conference (ICMC)*, pp.33–40 (2008).
- [2] Dannenberg, R.B.: An On-Line Algorithm for Real-Time Accompaniment, *Proc. 1984 International Computer Music Conference (ICMC)*, pp.193–198 (1984).
- [3] Dannenberg, R.B. and Mukaino, H.: New Techniques for Enhanced Quality of Computer Accompaniment, *Proc. International Computer Music Conference (ICMC)*, pp.243–249 (1988).
- [4] 伊藤和幸, 数藤康雄, 伊福部達: 重度肢体不自由者向けの視線入力式コミュニケーション装置, *電子情報通信学会論文誌*, Vol.83, No.5, pp.495–503 (2000).
- [5] Itoharu, T., Nakadai, K., Ogata, T. and Okuno, H.G.: Improvement of Audio-Visual Score Following in Robot Ensemble with Human Guitarist, *Humanoids 2012*, pp.574–579 (2012).
- [6] 川崎智子: ピアノ演奏時における読譜の為の眼球運動, *三重大学教育学部研究紀要教育科学*, Vol.33, pp.49–66 (1982).
- [7] Kawase, S.: An Exploratory Study of Gazing Behavior During Live Performance, *Proc. Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM)*, pp.227–232 (2009).
- [8] Kobori, S. and Takahashi, K.: Cognitive Processes During Piano and Guitar Performance: An Eye Movement Study, *Proc. International Conference on Music Perception and Cognition (ICMPC)*, pp.748–751 (2008).
- [9] Maezawa, A. and Yamamoto, K.: MuEns: A Multimodal Human-Machine Music Ensemble for Live Concert Performance, *Proc. 2017 Conference on Human Factors in Computing Systems (CHI)*, pp.4290–4301 (2017).
- [10] 中村栄太, 武田晴登, 山本龍一, 齋藤康之, 酒向慎司, 嵯峨山秀樹: 任意箇所への弾き直し・弾き飛ばしを含む演奏に追従可能な楽譜追跡と自動伴奏, *情報処理学会論文誌*, Vol.54, No.4, pp.1338–1349 (2013).
- [11] Nakamura, E., Nakamura, T., Saito, Y., Ono, N. and

- Sagayama, S.: Outer-Product Hidden Markov Model and Polyphonic MIDI Score Gollowing, *Journal of New Music Research*, Vol.43, No.2, pp.183–201 (2014).
- [12] Orio, N. and Dechelle, F.: Score Following Using Spectral Analysis and Hidden Markov Models, *Proc. International Computer Music Conference (ICMC)*, p.1 (2001).
- [13] Raphael, C.: Music Plus One: A System for Expressive and Flexible Musical Accompaniment, *Proc. International Computer Music Conference (ICMC)*, pp.487–512 (2001).
- [14] 下西 慶, 石川恵理奈, 米谷 竜, 川嶋宏彰, 松山隆司: 視線運動解析による興味アスペクトの推定, *ヒューマンインタフェース学会論文誌*, Vol.16, No.1-4, pp.103–114 (2014).
- [15] Sloboda, J.: The Eye – Hand Span – An Approach to the Study of Sight Reading, *Psychology of Music*, Vol.2, No.2, pp.4–10 (1974).
- [16] 寺崎栞里, 竹川佳成, 平田圭二: 視線情報を活用した楽譜追跡システムの構築, *研究報告音楽情報科学*, Vol.2016-MUS-112, No.11, pp.1–8 (2016).
- [17] 寺崎栞里, 竹川佳成, 平田圭二: DP マッチングのコストに視線情報を反映させた楽譜追跡手法の提案, *研究報告音楽情報科学*, Vol.2017-MUS-114, No.2, pp.1–6 (2017).
- [18] The Eye Tribe, available from (<https://theeyetribe.com/>) (accessed 2017-04-14).
- [19] Vecoe, B.: The Synthetic Performer in the Context of Live Performance, *Proc. 1984 International Computer Music Conference (ICMC)*, pp.199–200 (1984).



平田 圭二 (正会員)

1987年東京大学大学院工学研究科情報工学専門課程博士課程修了。同年NTT基礎研究所入社。1990～1993年(財)新世代コンピュータ技術開発機構(ICOT)に出向。2011年より公立はこだて未来大学教授, 現在に至る。博士(工学)。1993～1995年情報処理学会音楽情報科学研究会初代主査, 2010～2015年デジタルプラクティス誌編集委員長。音楽情報処理, 実時間デマンド型公共交通SAV, うつ病家族看護者のICT支援研究に従事。本会フェロー。



寺崎 栞里 (学生会員)

2016年公立はこだて未来大学システム情報科学部情報アーキテクチャ学科卒業。現在, 同大学大学院博士前期課程システム情報科学研究科2年に在学中。音楽情報処理に関する研究に従事。



竹川 佳成 (正会員)

2007年大阪大学大学院情報科学研究科博士課程修了。同年より神戸大学自然科学系先端融合研究環重点研究部助教。2012年より公立はこだて未来大学システム情報科学部助教。2014年より公立はこだて未来大学システム情報科学部准教授, 現在に至る。博士(情報科学)。ヒューマンコンピュータインタラクション, エンタテインメントコンピュータティング, 音楽情報科学の研究に従事。