

# マルチモーダル情報の統合により技能差に適応する 楽譜追跡システム

能登 楓<sup>1,a)</sup> 竹川 佳成<sup>1,b)</sup> 平田 圭二<sup>1,c)</sup>

**概要：**本研究の目的は、任意箇所への跳躍を含む演奏に対して頑健な楽譜追跡システムの実現である。従来の楽譜追跡の多くは、打鍵情報をもとに演奏箇所を推定しているため、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、直前の演奏箇所と全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。我々の先行研究では視線情報の導入によってこの問題の解決を試みたが、演奏者の技能差を考慮していないため誤った推定をしてしまう例が見られた。そこで、本研究では演奏者の技能の指標とみなされている視線の先読み幅 (EHS: Eye-Hand-Span) と視線情報、打鍵情報に対する優先度をパラメータとして導入した。実験の結果、比較手法に比べ追跡精度は約 1.3 倍 (0.76 から 0.96) に向上した。また、パラメータ導入による効果の検証も行った。

**キーワード：**楽譜追跡システム, 視線情報, DP マッチング

## Score Following System Adapting to Skill Levels by Integrating Multimodal Information

### 1. はじめに

本研究の目的は、ミス打鍵、跳躍、繰り返しを含んだ演奏に対して頑健に動作する楽譜追跡システムの開発である。従来、楽譜追跡は、自動伴奏、協奏、譜めくりシステムの要素技術として開発されてきた [1, 4, 5, 6, 8, 9]。一般に、楽譜追跡システムが解決すべき課題は、演奏に対応する譜面上の箇所を推定することである。演奏と譜面が完全に一致する場合には、ミスなく楽譜追跡できるが、人間の演奏には多くの不確定要素が含まれる。

一般に、楽譜追跡システムが想定している入力、時系列に沿って演奏されるような、所謂、本番演奏における MIDI や音響といった打鍵情報である。そのため、このような入力に含まれる不確定要素は、ミス打鍵や弾き飛ばし、テンポの揺らぎのような事象であった。一方、入力が練習環境における演奏のような状況を想定する。練習環境のような状況に

おいて、不確定要素は打鍵情報から観測される事象に加え、演奏位置に関する事象がある。練習環境では、上手に弾けない箇所に対する繰り返し練習や、演奏箇所の跳躍というように、演奏位置は必ずしも時系列に沿わない。譜面上に同一のフレーズが存在しない場合に限り、従来手法を用いた楽譜追跡は頑健に動作する。しかし、多くの楽曲には同一のフレーズが複数存在するため、楽譜追跡は困難である。

そこで、我々の先行研究 [9] では打鍵情報に加え視線を用いることによって、これらの問題の解決を試みた。演奏中の視線から大まかな演奏位置を獲得することができるため、演奏箇所の推定に有用である。寺崎らは視線情報を利用するにあたって視線の動きを HMM によってモデル化し、打鍵情報による楽譜追跡システムへ反映させた。楽譜追跡における視線情報の有用性を検証するために、従来の楽譜追跡手法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、視線を導入した場合の正解率は約 1.2 倍 (72.7%から 85.2%) に向上した。

しかし、視線情報を反映させる際、大幅な跳躍に対応させるために、優先度を固定値で与えていた (打鍵情報への重みを 0.4、視線情報への重みを 0.6)。そのため、打鍵情報から信

<sup>1</sup> 公立はこだて未来大学大学院  
Graduate School of Future University Hakodate

a) g2118030@fun.ac.jp

b) yoshi@fun.ac.jp

c) hirata@fun.ac.jp

頼度の高い楽譜追跡結果が得られるにもかかわらず、鍵盤の確認や先読みといった演奏位置から離れた箇所を注視してしまった場合には十分な追跡ができなかった。この理由として寺崎らは、Eye-Hand-Span (EHS), 打鍵情報と視線情報のトレードオフの未考慮を挙げている。EHS は演奏中の譜面における先読み幅を指す。この先読み幅は演奏者の技能やジャンルに応じて変化するもの [6, 7, 12] であるため、上級者と初級者では演奏中に注視している箇所は異なる。

演奏者の技能に応じて演奏中の視線、打鍵の振る舞いは異なる。そのため、打鍵情報と視線情報の優先度は演奏者の技能に適合させる方が良く考えられる。上級者は演奏経験から鍵盤の位置を憶えているため、打鍵時に鍵盤上の位置を確認することが少ない。そのため、上級者の視線情報は安定して獲得できるため信頼度が高く、重みを増すと良いと思われる。

本研究の目的は、技能差に適応する楽譜追跡システムの実現である。演奏者のジャンルはわずかな時間で変化するため [2], 技能差を扱うための必要条件として、少量の学習データを用いた技能量の推定があげられる。本研究では、EHS の推定にベイズ推定、優先度を考慮した情報情報の統合方法に加重平均を採用し、技能差に適応する楽譜追跡アルゴリズムを実装した。

## 2. 提案手法

### 2.1 演奏者の技能差を学習した楽譜追跡システム

本システムは、打鍵情報(打鍵音のノートナンバ)と視線情報(譜面上の  $x, y$  座標)を入力とし、楽譜追跡を行う。楽譜追跡を行う前に、本システムではユーザの技能差を学習する。本システムでは、ユーザの技能差を表すパラメータとして EHS, 打鍵情報, 視線情報の優先度を用いる。楽譜追跡を行うにあたって、事前に各ユーザの EHS 幅の推定 (2.2 節), 打鍵情報, 視線情報の優先度の決定 (2.4 節) を行う。これらのパラメータを推定した後に、DP マッチングへの反映を行う (2.3 節)。以上の過程から推定された演奏箇所を譜面上に提示する。

### 2.2 EHS を学習した結果

演奏者の技能差を表すパラメータとして、EHS の推定を行う。既に述べたように、演奏中の EHS は演奏者の技能や楽曲への理解に応じて変化する。また、演奏者の技能や楽曲への理解は楽器練習を通して、徐々に変化するものである [6, 7]。これらの特徴を踏まえた上で、本研究では EHS の推定にあたり、少量の学習データから解析的に計算可能な手法であるベイズ推定を用いる。演奏者が演奏音を注視する際の視線は演奏音を中心とした釣鐘上に分布することが想定される。そのため、ベイズ推定を行うために EHS の分布は正規分布に従うと仮定した。また、事後分布は共役事前分布である正規分布を用いることによって、EHS のパラメー

タは解析的に計算できる。

EHS をディスプレイ上の注視点 ( $g_x, g_y$ ) と演奏音の座標 ( $i_x, i_y$ ) の pixel の差を平均とした確率分布として定義する。譜面上の音符を注視しながら演奏した場合には、EHS は小さくなり  $i_x - g_x$  は 0 に近い値をとる。一方、演奏者が先読みをしながら演奏する場合には、 $i_x - g_x$  の値が大きくなる。譜面上の何段目の何小節を見ているのか推定するために、システムでは水平、垂直方向に対して、それぞれの EHS を作成した (式 1, および 2)。

$$P_x(i_x - g_x) \sim \mathcal{N}(i_x - g_x | \mu_{g_x}, \sigma_{g_x}) \quad (1)$$

$$P_y(i_y - g_y) \sim \mathcal{N}(i_y - g_y | \mu_{g_y}, \sigma_{g_y}) \quad (2)$$

これらの確率分布のパラメータ  $\mu_g$  と  $\sigma_g$  は、それぞれ EHS の大きさおよび分散を表す。

これらの分布はのちに、打鍵情報による楽譜追跡結果へ反映する。そのため、これらの分布の確率変数を pixel から音符番号へと変換する。音符番号とは譜面に記載されている音符の曲先頭からのインデックスを指す。音符が記載されている座標 ( $x, y$ ) と音符番号は 1 対 1 対応をとるため、求めた分布の特性を失わずに、変換することができる。この変換は水平方向、垂直方向の EHS それぞれに対して行い、変換後の分布は  $P_x(i_x - g_x)$  を  $P_x(i)$ ,  $P_y(i_y - g_y)$  を  $P_y(i)$  と表記する。

水平方向、垂直方向の EHS を推定したのちに、これらを統合する (図 1)。図 1 の分布は、図 3 に記載されている楽譜の 2 段目 8 小節 (図中\*マーク) を注視した際の  $P_x(i)$ ,  $P_y(i)$  である。垂直方向の EHS である  $P_y(i)$  からは、何段目の音符を見ているのか、水平方向の EHS である  $P_x(i)$  からは、各段における何小節目の音符を見ているのかを推定することができる。そのため、図 3 のように 3 段ある楽曲に対する水平方向の EHS である  $P_x(i)$  には 3 つの峰が立つ。同様に、垂直

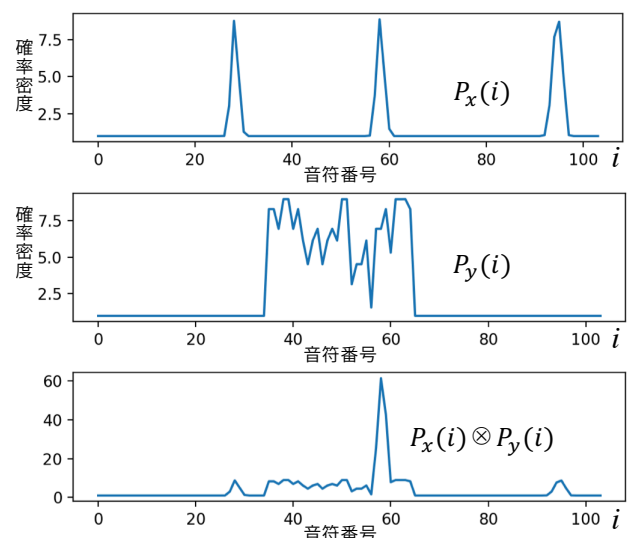


図 1 視線情報による演奏箇所の推定

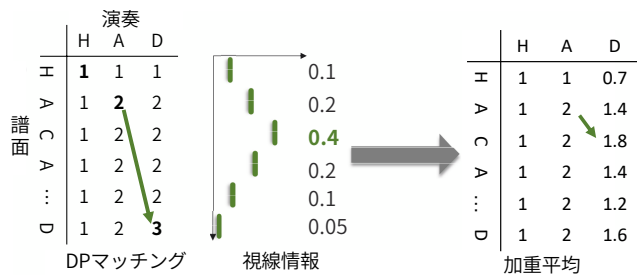


図2 加重平均による情報の統合

方向の EHS である  $P_y(i)$  からは各段における何小節目の音符を見ているのか推定することができる。視線情報のみを用いた楽譜追跡の際には、 $P_x(i)$  と  $P_y(i)$  の要素積を計算し、値が最大となる確率変数に相当する音符を現在の演奏音として推定する (式 3)。

$$P_g(i) = P_x(i) \otimes P_y(i) \quad (3)$$

### 2.3 視線情報と打鍵情報の統合

2.2 節で推定した視線情報を打鍵情報による楽譜追跡手法へ反映させる。打鍵情報による楽譜追跡手法としては Dannenberg によって提案された DP マッチングを採用する [4]。DP マッチングでは、演奏入力音と譜面上の音符との一致数をコストとして求める。コストが最も大きな値となる音符を演奏箇所として推定するため、少量の弾き飛ばしやミス打鍵を含む演奏に対して頑健な楽譜追跡が可能である。しかし、打鍵情報のみによる楽譜追跡手法では、同一フレーズの識別や大幅な跳躍を含む演奏に対する楽譜追跡は困難である。この問題を解決するために、視線情報を DP マッチングのコストに反映させる。

視線情報と打鍵情報の統合には加重平均を用いた。加重平均では、より重要なデータに重みを加えることによって優先度を考慮した平均を算出することが可能である。視線情報、DP マッチングにおけるコストをそれぞれ  $x_i$  (つまり  $N = 2$ )、重みのパラメータを  $w_i$  とすると各音符における情報の統合結果は以下の式で表される (式 4)。

$$\frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i} \quad (4)$$

重みの求め方については 2.4 節で述べる。

対応する音符における加重平均を計算することによって、演奏箇所を推定する。図 2 に視線情報を反映させた DP マッチングの例を示す。譜面 H-A-C-A-⋮-D に対して、演奏 H-A-D が入力された場合を想定する。打鍵情報のみによる楽譜追跡では、3 音目の演奏音である D がシステムへ入力された場合、大幅な跳躍が発生する。この時、継続する演奏音が入力されない限り、演奏者が意図的に跳躍したのか、入力がミス打鍵だったのか判別することは難しい。一方、図中央のように視線情報を考慮した場合、この判別は容易である。そこで、DP マッチングのコストと視線情報の加重平均

をとった際の値を求めることによって、入力 D が最も値が大きくなる 3 音目の演奏音であると推定できる。

### 2.4 視線情報を打鍵情報へ反映させる重み

視線情報と DP マッチングのコストとの加重平均を求めるにあたって、重みを推定する。この重みを奏中の視線情報と打鍵情報の優先度として解釈し、重みの推定を行う。優先度に応じて、楽譜追跡の精度は大きく変化する。例えば、視線情報を最優先させた場合は任意箇所への跳躍を含む演奏に対して有効に動作することが考えられる。しかし、視線情報を優先しすぎてしまう場合には、鍵盤やそのほかの箇所へ視線を移した場合に楽譜追跡の精度が低下する恐れがある。このように、視線情報と打鍵情報の優先度はトレードオフの関係にあるため、適切な優先度を求める必要がある。

本研究では、システム利用前に演奏者に適した重みの算出を行う。重みの算出のために、演奏者にはシステム利用前に課題曲を 1 度、先頭から末尾まで順番に演奏するよう指示する。正確な演奏位置を把握するために、演奏者にはミス打鍵した場合でも弾き直すことはせず、追従する音符を演奏するように指示する。この際に得られた視線情報、打鍵情報それぞれの情報による楽譜追跡を行う。視線情報による楽譜追跡は (式 3) を、打鍵情報による楽譜追跡は DP マッチングを用いる。楽譜追跡の結果、得られた正解数の比率を加重平均における重みとする。

## 3. 実験

演奏者の技能差を考慮した楽譜追跡の有用性を検証するために評価実験を行った。検証を行うにあたり、打鍵情報のみを用いた楽譜追跡、視線情報のみを用いた楽譜追跡、寺崎らによって提案された手法 (以降、比較手法)、提案手法の 4 つの方法を用いて楽譜追跡を行い、正解率を比較した。いずれの手法においても同一の演奏から得られる視線情報 (画面上の注視点  $x, y$  座標)、打鍵情報 (ノートナンバ) を入力データとした。打鍵情報は MIDI キーボードから、視線情報は Tobii 4C [11] から獲得した。Tobii 4C は小型かつ軽量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、Tobii 4C は 90Hz で視線情報のサンプリングを行う。

### 3.1 実験データ

課題曲として被験者に提示した楽曲は、ヤマハ音楽能力検定 (グレード 5 級) 初見演奏・即興演奏範例曲集に収録されている a-moll より冒頭 27 小節 (103 音) における右手単旋律譜面である [3]。本研究において、演奏者の技能差を扱うパラメータである EHS と各情報の優先度は演奏者の事前の知識に影響を受ける。そのため、課題曲はすべての被験者にとって初見の楽曲である楽曲を課題曲として選ぶことによって、条件をそろえた。また、図 3 に示すように、課題曲には同一のフレーズが 3 箇所、同音の連続が 11 箇所含まれ



図3 反復される音形が3組存在し、同音の連続が11箇所存在する課題曲

る楽曲である。そのため、演奏箇所への推定が困難である。

システムへの入力データを作成するために、譜面を読みながらピアノ演奏が可能な学生7名(男性5名, 女性2名)に協力を依頼した。そのうち、4名はピアノ学習経験がある者、その他の3名はピアノ以外の楽器を演奏したことがあるが、ピアノ学習経験がない者であった。

跳躍, 繰り返しを含む演奏に対する楽譜追跡性能を検証するために、被験者には第20小節から第24小節を演奏した後に、第6小節から第9小節へと跳躍するという試行を2度連続で演奏するように指示した。この際、ミス打鍵に関しては作適的な検証を避けるために、自然に発生したミス打鍵のみを用いた。そのため、演奏指示箇所を正確に演奏した場合には、ミス打鍵数を0とし評価を行なった。

### 3.2 正解率の算出方法

楽譜追跡の性能を測るための指標として、正解率を算出した。正解率は、演奏にミス打鍵が含まれる場合と含まれない場合では算出方法が異なる。被験者が演奏指示箇所を正確に演奏した場合、正解率は単純に楽譜追跡結果の正解数/総打鍵数として算出した。一方、被験者がミス打鍵をした場合には、ミスした音符をもう一度演奏しようとする、または、次の音符を演奏しようとするという2通りが考えられる。このような場合、被験者が意図する音符を正確に把握することができないため、正解率は、正解数/(総打鍵数 - ミス打鍵数)として算出した。

### 3.3 実験結果

表1に打鍵情報のみ、視線情報のみ、比較手法、提案手法による楽譜追跡の精度、および、実験におけるミス打鍵数と総打鍵数を示す。表1から、打鍵情報のみを用いた平均追跡精度は0.42、視線情報のみを用いた平均追跡精度は0.15であり、ほぼ追跡ができていないことがわかる。比較手法では平均追跡精度が0.76であり、それぞれの情報によって楽譜追跡を行なった場合に比べて大幅に向上したことがわかる。また、提案手法では平均追跡精度が0.96であり、比較手法に比べ約1.3倍(0.76から0.96)に向上した。

表1 提案手法と比較手法における正解率の比較

	打鍵	視線	比較	提案	ミス打鍵	総打鍵数
P	0.43	0.27	0.83	1.00	0	60
Q	0.43	0.07	0.80	0.93	8	68
R	0.35	0.06	0.70	0.90	15	75
S	0.43	0.23	0.73	0.97	8	68
T	0.43	0.23	0.76	0.97	0	60
U	0.43	0.03	0.80	0.97	2	62
V	0.35	0.28	0.68	0.93	7	67
平均	0.42	0.15	0.76	0.95	-	-

以上の結果より、打鍵情報のみ、視線情報のみでは楽譜追跡が困難な場合でも、情報を統合することによって楽譜追跡が可能であることいえる。また、比較手法と提案手法ではt検定の結果、有意水準1%で有意差が見られた( $p = 0.001$ )。

## 4. 考察

### 4.1 誤打鍵に対する頑健さ

誤打鍵の多寡による正解率への影響について議論する。ミス打鍵が多い被験者R, S, Vの正解率は、打鍵情報のみでは0.35, 0.43, 0.35、比較手法では0.70, 0.73, 0.68であり、提案手法では0.90, 0.97, 0.95であった。打鍵情報のみによる楽譜追跡では、被験者R, Vは平均を下回る追跡精度であった。一方、被験者Vと同程度のミス打鍵数である被験者Sの正解率は平均と同様であった。楽譜追跡は一度誤った箇所を推定してしまった場合、より演奏箇所である可能性の高い音符を推定するまで回復が困難であるという特性がある。そのため、ミス打鍵発生が序盤であるか、終盤であるかといったミス打鍵の発生タイミングによって最終的な正解率が変化する。このことから、ミス打鍵数がほぼ同じ被験者SとVの正解率に差があった理由は、ミス打鍵のタイミングであると考えられる。

いずれの手法においても被験者R, S, Vは概ね平均を下回る正解率であった。このことより、誤打鍵の多寡は正解率の精度に影響を与えるが、手法によってミス打鍵に対する頑健さが異なった。

表2 シャピロウィルク検定による検定結果

	W	p 値
P <sub>x</sub>	0.67	9.10e-14
Q <sub>x</sub>	0.92	1.37e-05
R <sub>x</sub>	0.48	2.96e-17
S <sub>x</sub>	0.29	7.00e-20
T <sub>x</sub>	0.33	2.09e-19
P <sub>y</sub>	0.92	1.96e-05
Q <sub>y</sub>	0.97	0.12
R <sub>y</sub>	0.77	4.55e-11
S <sub>y</sub>	0.70	6.08e-13
T <sub>y</sub>	0.53	2.39e-16

#### 4.2 EHS が正規分布に従うという仮定の妥当性

本研究ではベイズ推定の枠組みで視線情報を扱うために、EHS が正規分布に従うことを仮定した。演奏者の EHS の分布が正規分布に従うという仮定の妥当性を検証するために、シャピロウィルク検定を行った。シャピロウィルク検定では、標本が正規分布に従う母集団からサンプリングされたものであるという帰無仮説を検定する。標本は 2.4 節で述べた事前の演奏データから作成した。演奏データからは打鍵時の視点を取得し、演奏音の譜面上の座標との各軸 (x, y) における距離を標本とした。この際、音価の差は考えなかった。被験者 P から V の視線について、シャピロウィルク検定を行なった際の検定統計量 W, p 値を表 2 に示す。

表 2 より Q<sub>y</sub> を除く全ての視線分布において、有意水準 1% で有意差が見られた (p < .01)。また、Q<sub>y</sub> においても有意水準 5% で有意差が見られた (p < .05)。このことより、演奏中の視線の振る舞いは正規分布に従うと考えて良いだろう。

#### 4.3 技能差を学習することの有用性

本研究で技能差を扱うパラメータとして EHS の距離、優先度を挙げた。これらの値が楽譜追跡精度に与える影響について考察する。

まず、表に被験者の水平方向、垂直方向の EHS を示す。水平方向の EHS の平均において、最も小さかった値は被験者 V の -21.5pixel、最も大きかった値は被験者 S の 9.4pixel であり差分は 30.9pixel であった。水平方向の EHS の平均が負になるということが意味するのは、演奏者が演奏音以前の箇所を見ながら演奏しているということである。対して、正

表3 全被験者の EHS(pixel)

	水平平均	垂直の平均	水平の分散	垂直の分散
P	5.0	-15.8	64.9	25.4
Q	-3.6	93.4	32.6	42.5
R	-14.8	-106.2	269.9	141.8
S	9.4	27.0	126.4	42.4
T	8.4	10.3	141.8	79.1
U	0.9	1.8	94.8	43.0
V	-21.5	-0.4	151.8	67.1
平均	-2.3	1.5	126.0	63.0

表4 譜面上における各音符間の距離の平均 (pixel)

	付点 4 分音符	4 分音符	8 分音符	16 分音符
距離 (pixel)	53.3	28.8	31.2	18.9

になる場合には演奏者が先読みしながら演奏していることを表している。分散に関しては最も小さい値は被験者 Q の 32.6pixel、最も大きな値は被験者 R の 269.9pixel であった。

譜面上におけるこれらの数値を解釈するため、表 4 に音符間の距離の平均を示す。これらの値は対象の音符の符頭から、後続の音符の符頭までの距離を表す。図 3 のように課題曲で用いた楽曲における音価は付点 4 分音符、4 分音符、8 分音符、16 分音符の四種類であった。

最も EHS 長の大きな被験者 V は 21.5pixel であり、これは 16 分音符 1 つ分の距離に相当する。このことより、演奏者間の EHS 長は楽譜追跡に大きな影響を与えないと考えられる。一方、分散については最も小さな値は被験者 Q の 32.6 であり 8 分音符に相当する距離、最も大きな値は被験者 R の 269.9 であり、およそ 4 小節に相当する距離だった。このことから、演奏者の技能差を扱うパラメータは EHS の距離ではなく、EHS の分散であることが示唆される。この結果は EHS の大きさに演奏者の技能差が現れるという先行研究 [7, 10] とは異なる。

先行研究との違いが生じた理由として実験方法の違いが考えられる。まず、先行研究において EHS は、音符を注視した時刻とその音符を打鍵した時刻の差から推定しているが、本研究では、打鍵時の視線の位置から EHS を推定している。また、EHS の推定にあたって、先行研究では演奏テンポの指定がされているが、本研究では演奏者は自由なテンポで演奏した。これらの条件の違いから、本研究における EHS の平均幅には技能差が現れなかったと考えられる。

#### 4.4 優先度に関する考察

実験で使用した加重平均における重みを表 5 に示す。最も視線情報と打鍵情報の重みの差が小さな値となったのは被験者 P であり、重みの差は 0 だった。一方、最も差が大きな値となったのは被験者 V で、重みの差は 0.25 だった。また、被験者によっては視線情報と打鍵情報の重みの大小が逆転した。

表5 加重平均における重み

	視線情報	打鍵情報
P	0.50	0.50
Q	0.61	0.39
R	0.38	0.62
S	0.41	0.59
T	0.42	0.58
U	0.50	0.50
V	0.37	0.63
平均	0.46	0.54

表 6 提案手法と重みの平均パラメータによる楽譜追跡における正解率の比較

	提案手法	平均値を用いた楽譜追跡
P	1.00	1.00
Q	0.93	0.93
R	0.90	0.93
S	0.97	0.97
T	0.97	0.97
U	0.97	0.97
V	0.93	0.93
平均	0.95	0.95

演奏者ごとの優先度を推定することの有用性を検証するために、重みの平均値を用いた楽譜追跡結果との比較を行う(表 6)。比較に用いたデータは全て本実験で使用した視線情報、打鍵情報であり異なる箇所は優先度のパラメータのみである。

表 6 より、被験者 R のみ正解率が上昇し、そのほかの被験者では正解率が変わらなかった。この理由として、被験者 R の視線情報の重みは 0.38 であったのに対し、平均は 0.46 と視線情報の重みが大きく上昇したことが挙げられる。実験に用いた課題曲では、演奏箇所の跳躍が含まれていた。そのため、このような場合には打鍵情報よりも視線情報が優先されることが望ましく、被験者 R の正解率が上昇したと考えられる。しかし、ほぼ全ての被験者において同等の追跡性能であった。

このことより、楽譜追跡結果に大きく起因したものは、視線情報の重みではなく、視線情報から推定される演奏箇所であると推測される。その理由としては、課題曲に含まれる音符が考えられる。今回、実験に用いた課題曲には 8 部音符が多く含まれていた。8 部音符間の平均距離は 31.2pixel であり、水平方向における EHS の絶対値の平均 9.07pixel と比較して著しく大きい。このことより、演奏者が演奏音を注視しながら演奏していた場合、EHS の分布が最も高い値をとるのは推定すべき演奏音となる。そのため、DP マッチングの推定結果へ視線情報を反映させた際、視線情報の重みが 0 でなければ、演奏位置推定結果は重みに関係なく決まる。

しかし、表 1 に示したように、視線情報のみによる楽譜追跡の正解率は平均して 0.15 であった。この結果が示すように、演奏者が注視している音符は多くの場合、推定すべき演奏音ではなかった。このことより、提案手法は視線情報から大まかな演奏位置を推定し、打鍵情報によって詳細な楽譜追跡を行っていたと解釈できる。そのため、優先度に関する議論をするためには、演奏音が密集している楽譜に対して楽譜追跡を行う必要があった。

より高い精度での楽譜追跡を行うには、打鍵情報と視線情報の優先度ではなく、優先するタイミングに関するパラメータの導入が必要であると考えられる。タイミングとして考えられる事例は、ミス打鍵が増えた場合や鍵盤位置の確認

など、どちらかの情報の信頼度が著しく下がった場合などが挙げられる。

## 5. おわりと今後

本研究の目的は、ミス打鍵、跳躍、繰り返しを含んだ演奏に対して頑健に動作する楽譜追跡システムの開発である。技能差を扱うパラメータとして、EHS、視線情報と打鍵情報の優先度を導入した。実験の結果、比較手法に比べ追跡精度は約 1.3 倍 (0.76 から 0.96) に向上し、t 検定の結果、有意水準 1%において有意差が見られた ( $p = 0.001$ )。しかし、技能差を扱うパラメータに関しては重みの平均による楽譜追跡との差はほぼなかった。今後の予定として、演奏音が密集している楽譜など課題曲の追加が挙げられる。

## 謝辞

本研究を行うにあたり、ご指導・ご教授を賜りました五十嵐 健夫教授に心より感謝申し上げます。本研究は JST CREST JPMJCR17A1, JSPS 科研費 16H01744 の支援を受けたものである。

## 参考文献

- [1] Arzt, Andreas, Gerhard Widmer, and Simon Dixon. "Automatic Page Turning for Musicians via Real-Time Machine Listening", ECAI, 2008.
- [2] 石垣 愛美, 竹川 佳成, 平田 圭二, 富永 敦子, グラウンデッド・セオリー・アプローチの方法による ピアノ演奏習得におけるチャンク形成要因の予備分析, 研究報告音楽情報科学 (MUS), pp.1-8, 2017.
- [3] 浅野 義弘, ヤマハ音楽能力検定 (グレード 5 級) 初見演奏・即興演奏範例曲集 ('90 年度改訂版), 財団法人ヤマハ音楽振興会, 1997.
- [4] R. B. Dannenberg, and H. Mukaino, "New techniques for enhanced quality of computer accompaniment", ICMC, 1988.
- [5] M. Dorfer, F. Henkel, and G. Widmer, "Learning to Listen, Read, and Follow: Score Following as a Reinforcement Learning Game", ISMIR, 2018.
- [6] F. Henkel, S. Balke, M. Dorfer, "Score Following as a Multi-Modal Reinforcement Learning Problem", ISMIR, 2019.
- [7] S. Kobori, K. Takahashi, "Cognitive Processes During Piano and Guitar Performance: An Eye Movement Study", ICMPC, pp.748-751, 2008.
- [8] S. Sagayama, T. Nakamura, E. Nakamura, Y. Saito, H. Kameoka, and N. Ono, "Automatic Music Accompaniment Allowing Errors and Arbitrary Repeats and Jumps", POMA, vol. 21, no.1, 2014.
- [9] S. Terasaki, Y. Takegawa, K. Hirata, "Proposal of Score-Following Reflecting Gaze Information on Cost of DP matching", ICMC, pp. 144-149, 2017.
- [10] J. Sloboda, "The eye-hand-span an approach to the study of sight reading", Psychology of Music, 2(2), pp. 4-10, 1974.
- [11] Tobii Eye Tracker 4C, <https://gaming.tobii.com/tobii-eye-tracker-4c/> (参照 2020-1-31).
- [12] Weaver, H. E., "Studies of Ocular Behavior in Music Reading", Psychological Mono-graphs, 55(1), pp.1-29, 1943.