

画像処理を用いた MIDI 鍵盤演奏の 指使い認識法に関する研究

山川晃^{†1} 臼杵潤^{†1}

昨今ピアノを演奏する人が増えており、演奏練習を支援する研究も進められている。特に初心者の場合にはなめらかに指を動かせるようになるために練習時に指使い(運指)の配慮をすることが大切となる。指使いは練習曲などにおいて譜面上の一部の重要な箇所のみ指番号で記載されていることがあり、これを見ながら練習することになる。そこで、本研究では初心者の指使い練習の支援を行う方法を検討する。演奏の様子をビデオカメラで撮影し、着色した爪を追うことで打鍵した爪を認識する方法を示す。次に指使い認識に関する評価法を検討し、最後に提案手法を用いた指使い練習支援法を示す。

Recognition Method of Fingering in MIDI Keyboard Performance by Image Analysis

AKIRA YAMAKAWA^{†1} JUN USUKI^{†1}

The population of beginners of piano players has increased gradually in recent years. At the same time, the number of researches has increased that supports piano training. Especially, for the beginner, it is important to pay attention to fingering on the training to move the fingers smoothly. The fingering numbers are usually written in only an important part on the score. And the beginner practices the fingering technique by checking the fingering numbers. In the paper considers method of support of fingering practice of the beginner. At first step is taking a movie of the fingering movement with the video camera. This paper shows method of instead of the fingering movement, recognize nails. Also, the evaluation method for recognizing fingering is considering and the support of fingering practice using proposed method is considering.

1. はじめに

昨今ピアノを演奏する人が増えており、ピアノマスター [1]など演奏練習を支援するものや演奏上達を支援しようとする研究[2][3][4]がある。ここで、指使い(運指)は譜面上に指番号で指定されていることがあり、すべての音符に指番号が記述されていることはほとんどなく、練習曲などにおいて重要な箇所にとり込んで記述されているのが一般的である。演奏者は一部の音符に記述された指番号を見るだけで連続する鍵盤に対して自然に指を割り当てられるようになるが、初心者の場合はレッスンを受ける際に指導者に指使いの誤りを指摘してもらうことになる。この点に着目し、本研究では初心者が指使いを覚えてなめらかな指の動きができるようになる練習を支援する方法を検討する。

まず、鍵盤上部にカメラを設置し、撮影した映像の中から打鍵鍵盤を特定するために鍵盤のラベリングを行う。

続いて演奏中に撮影した映像から爪に色を塗った各指の打鍵時の動きを取得する。ここでは鍵盤のラベリング情報と MIDI 情報を用いる。これにより、どの指がどの鍵盤を弾いたかがわかるようになる。最後にこの方法を用いて演奏時に指使いを間違いやすい箇所を抽出し指摘することで練習を支援できるようにする。

2. 鍵盤認識方法と打鍵指検出について

2.1 鍵盤ラベリングと打鍵鍵盤検出

鍵盤情報を取得するため、2 値化による鍵盤範囲の抽出を行い、この抽出画像に隣接画素(8 近傍)の各集合を鍵盤として認識する。まず、黒鍵盤を抽出するため、図 1(a)のような画像を 2 値化し(図 1(b))、隣接する白要素(8 近傍)を集合として各鍵盤を抽出する(図 1(c))。この抽出の際、各鍵盤の画素数を記録し、ある閾値の範囲にある集合だけを鍵盤とすることで、すべての鍵盤を抽出する。このとき同時に鍵盤番号の割り当てを行なう。ここでは閾値の最大値を key_{max} 画素、最小値を key_{min} 画素と設定する。



図 1 黒鍵盤ラベリング

Figure 1 Labeling of black keys.

白鍵盤にも同様の処理を行うと図 2 のようになる。黄緑色部分が黒鍵盤($key_{max}=3500$, $key_{min}=2000$)、紫色部分が白鍵盤($key_{max}=6000$, $key_{min}=1500$)を抽出した結果である。



図 2 ラベリング結果

Figure 2 Result of labeling.

^{†1} 神奈川工科大学大学院
Department of information and computer sciences, Kanagawa Institute of
Technology

白鍵盤と黒鍵盤にはそれぞれ異なる値が連続して割り当てられており、これを鍵盤番号としている(図2では鍵盤の色を見分けやすくするため、全88鍵について白鍵盤は50~101, 黒鍵盤は120~155としている)。こうすることで、打鍵時に鍵盤番号を指定することで、打鍵鍵盤の範囲を検出できるようになる。そのため、打鍵時はMIDI情報を参照して鍵盤番号と照合することで打鍵鍵盤を特定する。

2.2 マハラノビスの汎距離[5][6]による指の判別

2.2.1 マハラノビスの汎距離を用いた色の抽出

特定した打鍵鍵盤上の指を検出することで、打鍵時の指情報を取得する。ここでは着色した付け爪を指の爪に取り付ける(図3)。これは、着色することで各指を認識しやすくなることと、装着してもあまり演奏の邪魔にならないことを重視したためである。また付け爪にすることで色を着色した際に発色を良くしやすくでき、練習時に着脱できる利点もある。



図3 付け爪を装着した各指

Figure 3 Using fake tips.

なお、付け爪は両手共に親指を除く4本とし、図4に示すように色相値が分散している4色を使用する。

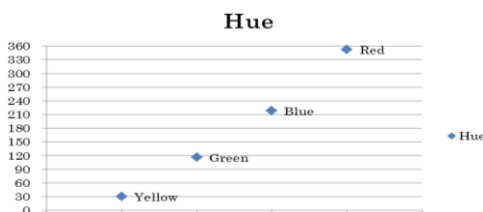


図4 色相値

Figure 4 Hue value.

両手の指については表1のように定義し、色と番号を割り当てている。この番号は、打鍵指の検出に使用する。

表1 指の色と指番号

Table 1 Colors and numbers.

	Thumb Finger	Index Finger	Middle Finger	Ring Finger	Little Finger
Color	-	Green	Red	Yellow	Blue
Group	1	2	3	4	5
Right Hand	1	2	3	4	5
Left Hand	6	7	8	9	10

図3と表1に示したように、人差し指、中指、薬指、小指の色には右手も左手も同じ色を割り当てている。色の抽出にはHSV値を用い、判別にはマハラノビスの汎距離を用いる。これは、他の情報を取得したい場合に色を用いる可能性を考慮し、なるべく少ない色数を使用することを狙ったためである。ここで4色の付け爪から取得したHSV値に

ついて算出した各グループの平均値を表2に示す。

表2 各群のHSV平均値

Table 2 Average of each HSV.

	H(sample 1)	S(sample 2)	V(sample 3)
Group 2	116.091154	0.364310	0.450538
Group 3	352.489014	0.772711	0.504901
Group 4	29.832368	0.657299	0.628751
Group 5	218.274474	0.451776	0.518522

次に、使用した全サンプルの分散・共分散行列 $S_{i,j}$ を式(1)で求める。 N は使用するサンプル数、 x は取得した値、 n はサンプルの種類でこの場合HSV値の3種である。

$$S_{i,j} = \frac{1}{N-1} \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}_i)(x_j - \bar{x}_j) \quad (1)$$

$$N = \text{total count of sample}$$

$$n = 3$$

続いて、算出した分散共分散行列の逆行列 S' を算出する。この結果を式(2)に示す。

$$S' = \begin{pmatrix} 0.003042 & -0.038224 & -0.042105 \\ -0.038224 & 77.336552 & -6.183917 \\ -0.042105 & -6.183917 & 247.420456 \end{pmatrix} \quad (2)$$

この結果を用いて式(3)によりマハラノビスの汎距離 D^2 を求める。 $x_{G,i}$ は各グループのサンプルで G はグループ番号、 i はサンプル番号である。

$$D_G^2 = \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}_{G,i}) S'_{i,j} (x_j - \bar{x}_{G,j}) \quad (3)$$

$$n = 3$$

実際に画像処理でマハラノビスの汎距離を用いた場合、図5のような画像が出力される。

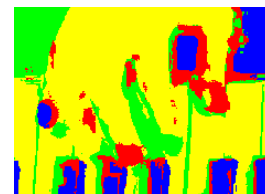


図5 マハラノビスの汎距離による指の色抽出結果

Figure 5 Application result of Mahalanobis' distance.

これは、式(3)の判別分析により4色のうち距離が最も短いグループを選択した結果である。そこで、爪のみを抽出するために表3のように閾値を設ける。

表3 各群の閾値

Table 3 Threshold of each group.

	Group 2	Group 3	Group 4	Group 5
Threshold	7.0	5.0	5.0	8.0

そして、選択されたグループの画素に対し表 3 の閾値未満のものにグループ番号 1~4 を付与し、閾値以上の領域はグループ番号 0 として破棄する(図 6).

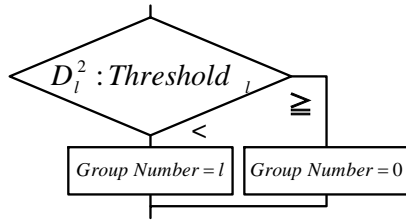


図 6 閾値による処理
 Figure 6 Extraction of nail area.

このようにして図 5 に対して閾値処理を行った結果を図 7 に示す.

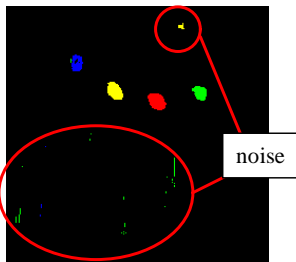


図 7 マハラノビスの汎距離と閾値による色抽出結果
 Figure 7 Results of nail extraction.

2.2.2 ノイズの削除

図 7 には指でない部分がノイズとして残留している. このノイズを除去するため, 図 8 に示すとおり, 各グループに対し 8 近傍処理を行う. まず, $k_{x,y}$ を注目画素, グループ番号($G=2,3,4,5$)とし, $L(k_{x,y})$ を $k_{x,y}$ のラベル番号とする. このとき, 8 方向の隣接する画素に対し, $k_{x,y}$ と同じグループ番号 G があつた場合はラベル番号 $L(k_{x,y})$ の集合として $k_{x,y}$ と同じラベル番号にするとともに, 各ラベルの画素数に計算する.

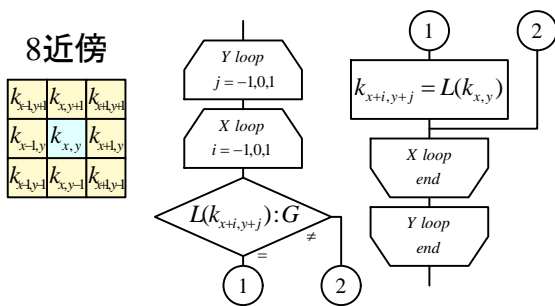


図 8 近傍処理
 Figure 8 Neighborhood processing.

このようにして各グループ G 内で複数得られるラベル 1 つずつを $Label_i (i = G+1, G+2, \dots)$, 各ラベルで算出した画素数を C_i とおく. このとき, C_i を降順にソートし, 3 番目以降の $Label_i$ の領域はノイズとして除去し, 残った 2 つのラベルを新たに $Label_{G+1}, Label_{G+5}$ とおく. この様子を図 9 に示す.

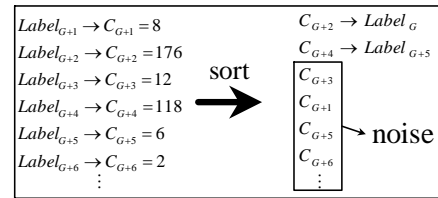


図 9 グループ $G(2\sim 5)$ 毎の色領域のラベリング
 Figure 9 Labeling of the color regions of each group.

このようにしてノイズの除去を行った結果を図 10 に示す.



図 10 ノイズ除去結果
 Figure 10 Results of image denoising.

2.2.3 左右 4 色の割り当てと指の検出

本研究では, 右手と左手に対応する爪に同じ色を使用している. そのため, 2.2.2 で抽出したラベル画像に対してそれぞれどちらの手の指かを判断する必要がある. そこで画像中の右側に存在する色と左側に存在する色領域 $Label_2 \sim Label_5, Label_7 \sim Label_{10}$ にそれぞれ指番号を割り当て, 色の座標を取得する. ここでは左側にあるラベルには右手の指番号を, 右側にあるラベルには左手の指番号を割り当てる (図 11). なお, 図 11 において x_G は $Label_G$ の領域の中心の X 座標値, G はグループ番号(2~4)であり, この座標を用いて左右の指番号の判定を行う.

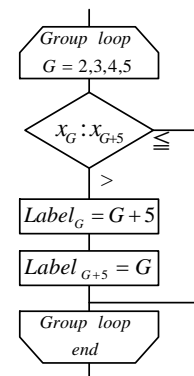


図 11 各領域の指番号の割り当て
 Figure 11 Assignment of finger number to each label.

このようにして各色を右手と左手の爪に割り当てた結果を図 12 に示す.

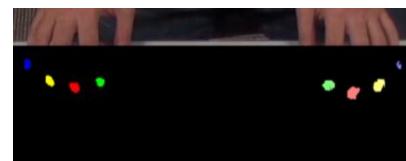


図 12 各爪の割り当て結果
 Figure 12 Result of assignment of each nail.

しかし、演奏中に指番号の割り当てを行う際、誤ってノイズを爪と判定してしまう場合がある(図 13). 図 13 の左の赤丸はノイズで右の赤丸は実際の爪の位置である.



図 13 ノイズを誤検出した例

Figure 13 Example of false detection.

2.2.2 のノイズ除去処理において本来検出するべき爪の画素数が他のノイズの画素数を下回っていた場合、このようなことが生じてしまう. そこで、異常な場所に爪があると判定された場合、これをノイズとして削除し、指番号の再割り当てを行う(図 14). まず、指座標 x_G, x_{G+5} は常に x_{G+1}, x_{G+6} の間に存在していると仮定し、これが満たされない領域はノイズと判断し、爪の色情報 $Label_G, Label_{G+5}$ の削除を行った後、再び x_{G+1} から x_{G+6} の X 座標範囲を走査し 2.2.1, 2.2.2 と同様の処理を行う. なお、爪の色と鍵盤のエッジ部のマハラノビスの汎距離が近い場合にノイズが発生しやすいことを確認し、このアルゴリズムは 2 番と 3 番の指にのみ適応すべきであることを実験により判断した.

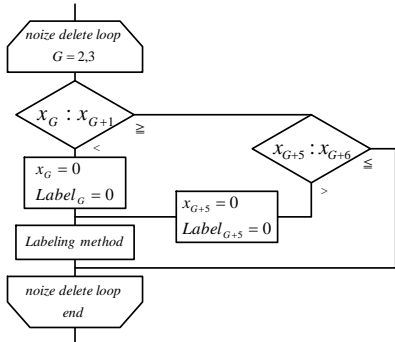


図 14 ノイズ削除処理と指番号の再割当て

Figure 14 Noise disposal process and reassignments of fingers number.

図 13 に対して再割当てを行った結果を図 15 に示す.

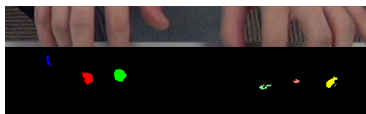


図 15 ノイズ削除と再割当て結果

Figure 15 Results of image denoising and reassignment

2.2.4 親指の判定

2.2.1 項の打鍵鍵盤特定の際に閾値を超えて破棄された(指番号 0 と判定された)ときに親指の特定を行う. このとき、打鍵鍵盤が右手と左手のどちらの人差し指に近いかを基準に右手と左手のどちらの親指かの判定を行う. この処理を図 16 に示す.

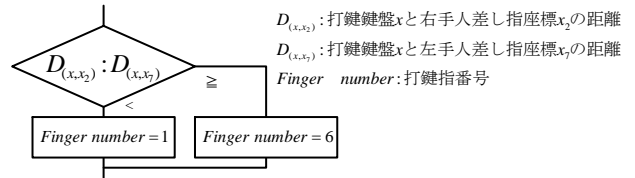


図 16 親指判定

Figure 16 Thumb detection.

2.3 システム評価法の提案

従来から画像処理等で用いられている評価法としての認識率算出方法には認識誤差の百分率から再現率や適合率 [7][8] などを含めいくつかあるが、鍵盤演奏時の認識精度を測るためにはより細かく評価できる手法が必要であると考える. ここに新たな評価法を提案する.

鍵盤演奏時に打鍵すべき鍵盤を注目鍵盤とする. このとき、注目鍵盤に対し、”正しい指で弾いている”, ”弾いていない”, ”異なる指で弾いている”の 3 種類が考えられる. また、認識に関しても同様の 3 種類が考えられる.

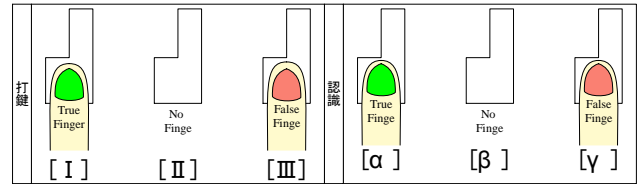


図 17 起こりうる打鍵と認識のパターン

Figure 17 Pattern of fingering and recognition.

この実際の打鍵を I, II, III, システム上の認識を α, β, γ としたとき、これらの関係を図 18 に示す.

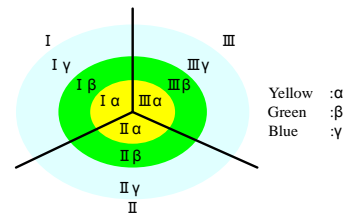


図 18 誤認識が起こるパターン

Figure 18 Pattern of false recognition.

図 18 において正しく認識を行っている領域は $I \alpha, II \beta, III \gamma$ の 3 つになる. この考えからシステム全体の認識率は式(4)で評価できることになる.

$$(\text{システム評価}) = \frac{(\text{楽譜通りに弾くべき数}) - (\text{誤認識数} : A)}{(\text{楽譜通りに弾くべき数})} \quad (4)$$

$$A = II \alpha \cup III \alpha \cup I \beta \cup III \beta \cup I \gamma \cup II \gamma$$

式(4)をシステム評価式と呼ぶことにする. この値は誤認識が少なくなる程 1.0 (=100%) に近づき、多くなるほど 0 に近づく.

システム評価式を誤認識の種類ごとに測る場合は式(5)のように a の内容をいずれかに絞り込めば算出することができる.

$$(\text{システム評価} a) = \frac{(\text{楽譜通りに弾くべき数}) - (\text{誤認識数} : a)}{(\text{楽譜通りに弾くべき数})} \quad (5)$$

$$a = II \alpha, III \alpha, I \beta, III \beta, I \gamma, II \gamma$$

また式(4), 式(5)ともに各指毎に評価することもできる.

3. 指使い練習支援法とその演奏評価法の提案

3.1 指使い練習支援法

演奏者の指使いの練習を支援するために, 演奏者が指使いを誤った箇所を次の3種類の方法で提示することを考える.

- (a) 何小節目にミスが多いか
- (b) 何小節目の何拍目にミスがあるか
- (c) ミスの多い指はどの指か

演奏情報を時系列データとして扱うことを考える. 初心者の演奏ミスには弾き忘れ, 弾き違い, 隣接する鍵盤の誤打鍵などがあり, これらはデータの過不足等で表現することができる. このように過不足や誤りのあるデータについて一致度合いを調べるため, 本研究では DP マッチング[9]を用いる. 判定する情報は打鍵指情報と打鍵鍵盤情報とする. 打鍵指情報は打鍵時に取得した指番号のことで, これをお手本データと比較することで誤りを検出する. 打鍵鍵盤情報は MIDI データから取得できる打鍵した鍵盤のことで, これをお手本データと比較することで誤りを検出する. ここでは式(6)の DP 漸化式を用いて経路探索し, 最短経路を導き出すことにより演奏情報から誤りを検出可能となる. 式(6)において $g(i,j)$ とは総合距離で(0,0)から(i,j)までの移動の和である. $d(i,j)$ は移動に対する重みで一致度により変化する.

$$g(0,0) = d(a_0, b_0) = d(0,0) \quad (6)$$

$$g(i,j) = \min \begin{cases} g(i-1, j) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i, j-1) + d(i, j) \end{cases}$$

お手本データは小節情報を付与しておくことで, DP マッチングの結果と追従して小節情報を取得できるようになる. 図 19 に誤りに検出例を示す.

小節数	1	1	2	2	2	3	3	3
お手本	1	2	3	1	2	3	1	2
打鍵	2	3	3	1	3	4	1	2
判定	X	X	O	O	X	X	O	O
miss=4								

図 19 DP マッチングによる指番号ミス検出

Figure 19 Result of fingering.

図 19 のようにマッチングを確認し, 誤り箇所を次のように提示する.

- (判定 a) 1 小節目: 2 回
2 小節目: 2 回
- (判定 b) 1 小節目: 1,2 拍目
2 小節目: 2,3 拍目
- (判定 c) 右手親指: 1 回
右手人差し指: 2 回
右手中指: 1 回

3.2 演奏評価法の提案

ここでは演奏者の上達度合いを判定する演奏評価式を提案する. 演奏ミスには図 20 に示すように, 弾き過ぎ, 弾

き忘れ, その他の間違いがある.

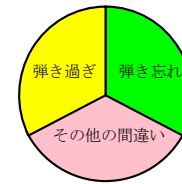


図 20 演奏ミスパターン

Figure 20 mistake pattern of keystroke..

特に指使いの練習をする際は, ”その他の間違い”は間違った指での打鍵になる. そこで, 弾きすぎ, 弾き忘れ, 指の間違いを右手左手の全指に対し検出すると, 演奏評価式の検出率は式(7)で表現できる.

$$(\text{演奏評価}) = \frac{(\text{楽譜通りに弾くべき数})}{(\text{楽譜通りに弾くべき数}) + (\text{弾き間違えた数} : B)} \quad (7)$$

ただし, $B = \sum_{i=1}^n R_i \cup \sum_{j=1}^m L_j$, R_i, L_j はそれぞれ左右の指の誤り回数
 ※間違った指で弾いてしまった場合は弾くべき指が誤ったとし誤り回数には加算する.

式(7)は弾き間違いが少なくなるほど 1.0(100%)に近づき, 多くなるほど値が 0 に近づく. なお, 各指に対して評価を行いたい場合は, 式(8)で表現できる.

$$(\text{演奏評価}b) = \frac{(\text{楽譜通りに弾くべき数})}{(\text{楽譜通りに弾くべき数}) + (\text{弾き間違えた数} : b)} \quad (8)$$

$$b = R_1, R_2, R_3, R_4, R_5, L_1, L_2, L_3, L_4, L_5$$

ただし, システム評価結果が不十分の場合, 演奏者の正しい評価は難しくなる.

4. 演奏練習による提案手法の実験

4.1 演奏曲と撮影時の環境について

練習曲は打鍵する鍵盤と指が指定されており, 本研究の目的と提案手法を評価するのに適していると考え, ハノン第 1 部 1 番を使用する. また, ハノンの練習方法に習い, 演奏時のテンポは 60~108 の中から 60, 80, 108 の 3 通りについて実験を行う. なお, 撮影環境は蛍光灯のある明るい部屋で直射日光の当たらない場所とした.

4.2 演奏結果による提案手法の評価

4.2.1 指の認識のシステム評価

実験結果として式(4)によるシステム評価を表 4 に, システム評価の内訳を表 5 に示す. 表 5 の II と III の 100%は演奏ミスが生じなかったことを示している.

表 4 システム評価結果

Table 4 Results of system evaluation.

Tempo	60	80	108	Average
システム評価	95.3%	94.4%	93.6%	94.4%

表 5 種類別の誤認識結果(3 種類テンポの平均)

Table 5 Result of false recognition(average).

	I	II	III
α	-	100.0%	100.0%
β	100.0%	-	100.0%
γ	94.4%	100.0%	-

