

# ハンドジェスチャーによる身体的動作を用いたシンセ演奏表現の拡張

福知山公立大学 情報学部情報学科

32245020 小野寺輝人

指導教員 橋田光代 准教授

提出日 2026年1月30日

<b>— 目次 ——</b>	<b>6</b>
<b>1 はじめに</b>	<b>1</b>
<b>2 電子楽器とその演奏について</b>	<b>1</b>
2.1 シンセサイザー . . . . .	1
2.2 MIDI コントローラー . . . . .	1
2.3 インタラクティブ楽器 . . . . .	1
2.4 研究の方針 . . . . .	2
<b>3 システム構成</b>	<b>2</b>
3.1 MediaPipe . . . . .	2
3.2 Max . . . . .	2
3.3 MIDI . . . . .	2
3.3.1 ノートオン/オフ・ノートナンバー	
3.3.2 コントロールチェンジ(CC)	
<b>4 ハンドジェスチャーによる MIDI の制御</b>	<b>3</b>
4.1 ユーザーが取るジェスチャー . . . . .	3
4.2 ハンドサイン . . . . .	3
4.3 カメラからの距離 . . . . .	3
4.3.1 手の開き具合	
4.4 ノブを回す動作 . . . . .	3
4.5 スワイプ . . . . .	4
<b>5 システムマニュアル</b>	<b>4</b>
5.1 基本設定画面 . . . . .	4
5.1.1 セクション説明	
5.1.2 ジェスチャーアイコン説明	
5.1.3 基本的な実行の流れ	
5.2 基本姿勢 . . . . .	4
5.3 MIDI 送信設定 . . . . .	4
5.3.1 MIDI 送信先	
5.3.2 CC ナンバー・ノートナンバー	
5.3.3 MIDI チャンネル	
5.3.4 MIDI CC 範囲	
5.4 トリガージェスチャーとコントロールジェスチャー . . . . .	5
5.5 トリガージェスチャー . . . . .	5
5.5.1 MIDI CC 機能	
5.5.2 ピッチ機能	
5.5.3 タイム機能	
5.5.4 DISTANCE 機能	
5.6 コントロールジェスチャー . . . . .	5
5.6.1 対応ジェスチャー	
5.6.2 MIDI CC 機能	
5.6.3 ピッチ機能	
5.7 機能説明 . . . . .	6
5.7.1 スムージング機能	
<b>6 実装</b>	<b>6</b>
<b>7 考察</b>	<b>6</b>
<b>8 おわりに</b>	<b>6</b>
<b>— 図目次 ——</b>	
1 ノブ・つまみ、フェーダー、ボタンの例 . . . . .	1
2 MIDI コントローラーの例 (AKAI MPK mini MK3、Novation Launchpad) . . . . .	1
3 テルミン . . . . .	1
4 MediaPipe での手情報取得の例 . . . . .	2
5 位置情報を取得する 20箇所 . . . . .	2
6 4種類のハンドサイン . . . . .	3
7 カメラからの距離による値の変化 . . . . .	3
8 手の開き具合による値の変化 . . . . .	4
9 つまみを回す動作による値の変化 . . . . .	4
10 MIDI 送信設定画面 . . . . .	5
11 CC 範囲設定つまみ . . . . .	5
12 CC 範囲設定の詳細 . . . . .	5
<b>— 表目次 ——</b>	
1 スケール、ルート音、オクターブ範囲 . . . . .	6

## 1. はじめに

小型のハードウェアシンセサイザー等電子楽器を複数個用いた演奏形態であるマシンライブは、ノブやフェーダー、ボタンを用いた微細かつ連続的なパラメータ制御、ならびに時間的に安定したトリガー入力を可能とする点で高い操作精度を有している。一方で、そのような入力インターフェースでは、演奏行為自体が視覚的には単調に見えやすく、身体動作と音響変化との関係性が直感的に伝わりにくい。筆者は、この身体動作と音響変化の関係性の伝わりにくさは、電子楽器演奏における表現の幅を制限している要因の一つであると考えており、その幅を広げることで、これまでの電子楽器演奏のクールな印象に加えて、新しい印象やイメージ、価値の持ったライブ体験を届けることが可能になると考へた。そして、電子楽器を用いた演奏がもっと視覚的に楽しめるようにするには、これまでの操作に加えて、身体動作と音響変化が一致して伝わる、つまり身体的ジェスチャーを用いた直感的な操作で音を操るといった音響的制御が大切であると考えた。そこで本研究では、カメラベースのモーションキャプチャ技術である MediaPipe を用い、演奏者の身体動作を直接 MIDI 制御へと変換するインタラクティブ演奏システムを提案する。これにより、演奏行動そのものを、視覚的・音響的な表現として捉え直し、電子楽器演奏における没入感の増幅と、既存の枠組みを超えた新たな印象および価値の創出を目指す。

## 2. 電子楽器とその演奏について

電子楽器の歴史はこの章では、

### 2.1 シンセサイザー

### 2.2 MIDI コントローラー

MIDI は、シンセサイザー等の電子楽器やコンピュータ間で演奏情報（音の高さ、長さ、強さなど）をやり取りするための共通規格である。そして MIDI を使い、電子楽器やソフトウェアをコントロールするためのデバイスを、MIDI コントローラと呼ぶ（図 2）。主に、鍵盤やパッド、ノブ、フェーダー等のいずれかを組み合わせて構成されることが多い、これらによって、MIDI を用いた音の発音やパラメータの調整が可能になる。音楽制作における便利ツールとしてよく使用されるが、最近ではリアルタイムでの演奏にも使用される。

MIDI コントローラーの利点といえば、やはり

### 2.3 インタラクティブ楽器

電子楽器の分野では、鍵盤を弾くというような、一般的な入力方法のほかに、センサ技術やコンピューターの発展とともに、身体動作を音響制御に用いるインタラクティブ楽器が数多く提案してきた。

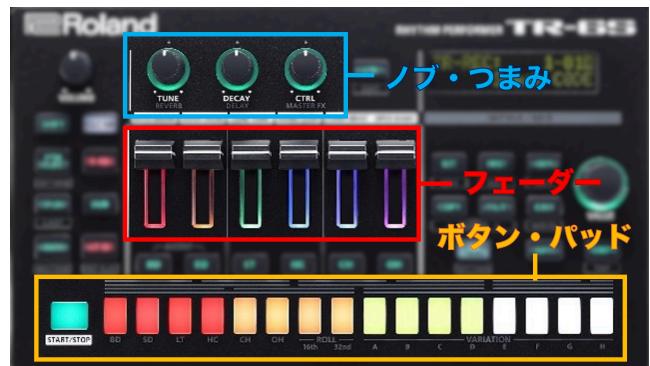


図 1: ノブ・つまみ、フェーダー、ボタンの例



図 2: MIDI コントローラーの例 (AKAI MPK mini MK3、Novation Launchpad)



図 3: テルミン

その1つとして有名であるのが「テルミン」である（図 3）。テルミンは 1920 年にロシアで生まれた電子楽器で、アンテナと手の間に生まれる電磁場の変化を感じて、音を制御する電子楽器である。そしてなんといっても 1 番の特徴は操作方法である。本体から伸びる 2 本のアンテナに手を近づけたり離したりすることで、音の高さと大きさの制御が可能であり、鍵盤のように音の場所が決まっていないため、手を小さく素早く振ることでビブラートを自由に表現できるというように、演奏者の身体動作が音響変化に反映される。このような楽器本体に触れずに、音を空中で操るといった演奏姿は、魔法使いのようにも感じられる。

また、日本でも YAMAHA が 1995 年に「Miburi」を開発・販売している。Miburi（ミブリ）は、指、手首、肘、肩、

足にセンサーが付いたウェアを着た演奏者の身振り手振りといった身体動作を音楽の出力に関連付けることで、音を直接コントロールする電子楽器である。

そして、近年では、Arduino や Raspberry Pi といった、安価で応用用途の幅広いマイコンが開発され、一般ユーザーでもセンサ等を用いた、電子工作による新しい電子楽器の開発が比較的容易になった。

しかし、これらの多くは専用デバイスの装着や特定環境への依存を前提としているほか、身体動作と音響変化が一致しても、演奏操作には一定の習熟が求められることから、ライブパフォーマンスへの導入には依然として制約が存在する。

## 2.4 研究の方針

このような既存の電子楽器の特徴から、本システムでは以下のような要素を持つように方針を定めた。

### (1) カメラ映像設定

このセクションでは、入力映像と MediaPipe における設定を行う。

### (2) 全体設定

### (3) ジェスチャー設定(トリガー・コントロール)

## 3. システム構成

ユーザーのハンドジェスチャーに対する、電子楽器演奏における没入感の増幅と、既存の枠組みを超えた新たな印象および価値の創出の有効性を検証するため、ハンドジェスチャーによって直感的な MIDI 制御が可能なエンタテイメントシステムを構築した。ユーザは片方どちらかの手全体が Web カメラに映る状態を取り、システムに用意されているジェスチャーを行うと、各ジェスチャに対応した MIDI 情報がシンセサイザー等の MIDI 対応機器に送られ、ユーザーは直感的な操作による演奏を楽しむことができる。

### 3.1 MediaPipe

手指の骨格情報と位置、動きの検出のため、MediaPipe を使用することにした。MediaPipe は、画像や動画データをリアルタイムで処理し、顔認識、姿勢推定、手の検出といった身体動作検出が可能な、Google 社が提供するオープンソース機械学習フレームワークである。図 4 は、MediaPipe での手情報取得の例である。

本システムでは、Web カメラを使用してユーザーの手を撮影し、得られる手指の関節 20箇所の位置情報を映像から取得する。手指の関節 20箇所の詳細を図 5 に示す。そして、それぞれの箇所の映像中での x 座標、y 座標を

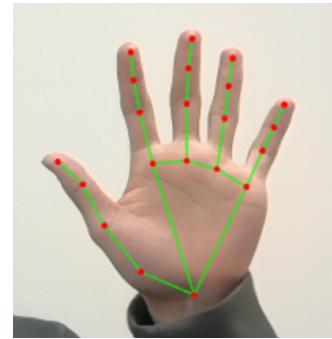


図 4: MediaPipe での手情報取得の例



図 5: 位置情報を取得する 20 箇所

### 3.2 Max

本システムは、システム構築環境として、ビジュアルプログラミング環境である Max(Cycling '74) を用いた。Max は、音響信号処理や映像処理をリアルタイムに統合して扱うことが可能な開発環境である。ユーザは、それぞれ特定の処理を行うオブジェクトをパッチャ上に配置し、それらを接続することで処理の流れを構築する。今回は、外部入力デバイスから取得した情報を音楽的制御信号へ変換する処理や構造を直感的に設計できることに加え、視覚的に把握することができるといった点から、Max を使用してシステム構築をすることにした。

### 3.3 MIDI

本システムでは、MIDI を用いて電子楽器を制御する。

MIDI は、シンセサイザー等の電子楽器やコンピュータ間で演奏情報(音の高さ、長さ、強さなど)をやり取りするための共通規格である。MDI で扱うデータ(メッセージ)には、いくつかの種類が存在するが、本システムでは主にノートオン/オフ・ノートナンバー、コントロールチェンジ(CC)の 3 つのメッセージを扱う。それぞれのメッセージについて詳しく説明する。

#### 3.3.1 ノートオン/オフ・ノートナンバー

ノートオンは鍵盤を押したときに送られる音を鳴らす信号、ノートオフは鍵盤を離したときに送られる音を止める信号のことである。また、ノートナンバーは、音の高さを 0 から 127 までの数字で表現したものである。これらのメッセージを組み合わせることで、どの高さの音を鳴らす、止めるといった制御が可能になる。

### 3.3.2 コントロールチェンジ (CC)

コントロールチェンジ(以降 CC と呼ぶ)は、音量、定位(パン)、音色、モジュレーションなど、音楽的な表現やパラメーターを制御するための MIDI 信号であり、シンセサイザー等のノブやフェーダーの制御量を 0~127 の数値として送信し、音色やエフェクトなどをリアルタイムに制御できる。

CC は、どのパラメーターを制御するかを示す 0~127 の番号であるコントロールナンバー(CC ナンバー)と、そのパラメーターの制御量 0~127 の主に 2 つの要素で構成される。

## 4. ハンドジェスチャーによる MIDI の制御

この章では、本システムにおいて値の制御に用いるジェスチャーの取り方と、その仕組みについて述べる。

### 4.1 ユーザーが取るジェスチャー

このシステムのジェスチャーは主に、単発的な信号を送信するものと、連続的な信号を送信するものの 2 種類に分かれる。これは既存の電子楽器のように、鍵盤やボタン、パッドといった単発的な出力を得意とする入力インターフェースと、ノブやつまみといった連続的な出力を得意とする入力インターフェースの 2 つを備えることで、それらの操作を非接触の身体動作によって実現することが可能としている。さらに、入力デバイスを身体動作そのものに置き換えることで、柔軟な演奏インターフェースを提供し、演奏者の直感的な表現を拡張することを目的としている。

実装されているジェスチャーと信号送信の種類は、以下のようである。

- ハンドサイン(単発)

(Closed Fist, Pointing Up, Victory, ILoveYou)

- カメラからの距離(連続)

- 手の開き具合(連続)

- つまみを回す動作(連続)

- スワイプ(単発)

これらのジェスチャー 1 つ 1 つについて、詳しく説明する。

### 4.2 ハンドサイン

本システムでは、Closed Fist、Pointing Up、Victory、ILoveYou の 4 種類のハンドサインを自動認識し、これらのジェスチャーを取ったら MIDI 情報を送信するといった単発的なトリガーとして使用することができる(自動認識の様子)。また、これらのジェスチャーを取ったときと、ジェスチャーをやめた時の 2 回、信号を送ることが可能である。



図 6: 4 種類のハンドサイン



図 7: カメラからの距離による値の変化

### 4.3 カメラからの距離

本システムでは、カメラと手の距離に応じて、図 7 のように連続的に送信値を変化させることができる。

カメラからどれだけ離れているかの測定には、MediaPipe によって得られた手首と小指の付け根の位置座標間の距離を使用している。つまり、距離が短くなる=カメラから手が離れているとなるわけである。

今回は、それぞれの座標間の距離の中で、カメラからの距離を変えずに様々なハンドジェスチャーを取った際に、距離の変化が少なく、測定箇所が他の指で隠れるようなことも少ないため、この座標間の距離を使用した。また、MediaPipe は z 軸座標の取得もできるが、測定誤差が大きいため、本システムでの使用は断念した。

座標間の距離は、手首と小指の付け根それぞれの x 軸、y 軸座標を、Max 内での三平方の定理の計算に用いることで出力している。

### 4.3.1 手の開き具合

本システムでは、手の開き具合に応じて、図 8 のように連続的に送信値を変化させることができる。

開き具合の測定には、手首と中指先端の位置座標間の距離を使用しており、手首と中指先端それぞれの x 軸、y 軸座標を、Max 内での三平方の定理の計算に用いることで距離を出力している。

また、この出力された値は、単にカメラ画角内の 2 座標間の距離によるもので、この状態ではカメラと手の距離によっても出力される値は変化してしまう。今回は手の開き具合のみの値を出力したいため、「手首と中指先端間の距離 ÷ カメラからの距離」の計算をすることで、この問題を解決し、ユーザーの手がカメラ画角内のどの位置にいても、出力値が同じ値の幅になるようにした。

### 4.4 ノブを回す動作

本システムでは、図 9 のようにシンセサイザー等についているノブ・つまみを回すようなジェスチャーをカメラに

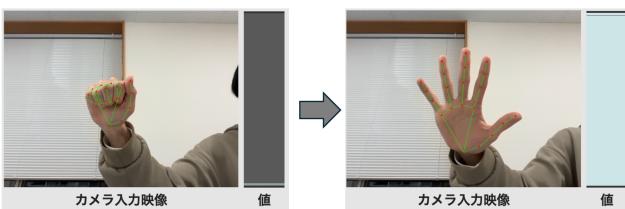


図 8: 手の開き具合による値の変化



図 9: つまみを回す動作による値の変化

向けて取ることで、連続的に送信値を変化させることができる。

送信値は、親指先端と人差し指先端の 2 点間の角度を元に変化する。本システムでは、2 点の x 座標、y 座標それぞれの座標差を使い、アーカタンジェント 2 関数でラジアン値を得て、それを度数に変換することで、角度を出力している。以下はその式である。

$$\theta_{\text{deg}} = \arctan\left(\frac{f_1}{f_2}\right) \times \frac{180}{\pi}$$

$f_1$  : y 座標差,  $f_2$  : x 座標差

#### 4.5 スワイプ

本システムではカメラ画角内で、上下左右に指先を素早く振るジェスチャーをトリガーとして、単発的な値の送信に扱うができる。

### 5. システムマニュアル

この章では、前述したジェスチャーで MIDI を制御するための、ユーザーが Max 内で行う設定について述べる。

#### 5.1 基本設定画面

ユーザーは、Max でカメラ映像設定やそれぞれの値出力パターンに対応するジェスチャーの選択、MIDI 送信のための設定などを行う必要がある。

そして、図〇は実際の設定画面である。これらの設定を行うことで、ジェスチャーを取った際に適切な電子楽器への MIDI 送信が実行され、ユーザー好みの直感的な操作による演奏が可能になる。

#### 5.1.1 セクション説明

##### (1) カメラ映像設定

このセクションでは、入力映像と MediaPipe における設定を行う。

##### (2) 全体設定

このセクションでは、システム全体

#### (3) ジェスチャー設定(トリガー・コントロール)

このセクションでは、単発的な値の送信(トリガー)と連続的な値の送信(コントロール)のそれぞれで、どのような動きの値をどの電子機器へどのようなジェスチャーで送信するのか決めるため、ジェスチャーの選択や MIDI 送信における設定を行う。

##### 5.1.2 ジェスチャーアイコン説明

ジェスチャー設定セクションにて、使用するジェスチャーを選択するためのアイコンを用意しているため、これについての説明を行う。

このアイコンは、クリックすることでアイコンの縁が光り、ON の状態になったことを示す。つまり、ON にしたジェスチャーによって値の制御を行うことを示す。

##### (1) ハンドサインアイコン

ハンドサイン (Closed Fist,Pointing Up,Victory,I Love You) が制御ジェスチャーとして選択されるアイコン。

##### (2) スワイプアイコン

このセクションでは、システム全体

##### (3) 距離アイコン

このセクションでは、単発的な値の送信(トリガー)と連続的な値の送信(コントロール)のそれぞれで、どのような動きの値をどの電子機器へどのようなジェスチャーで送信するのか決めるため、ジェスチャーの選択や MIDI 送信における設定を行う。

##### (4) 開閉アイコン

##### (5) ノブアイコン

#### 5.1.3 基本的な実行の流れ

#### 5.2 基本姿勢

今回はシステムの仕様上、手掌面が撮像デバイスに対して正面を向いている姿勢を基本姿勢とする。ユーザーはこの姿勢からハンドジェスチャーを取ることで、トリガーとしての役割や、MIDI 送信値の連続的な変化を、適切に実行することができる。

#### 5.3 MIDI 送信設定

シンセサイザー等へ MIDI を送信するためには、図 10 のように MIDI 送信先、ノートナンバー・CC ナンバー、MIDI チャンネル、MIDI CC 範囲の 4 種類の項目をユーザが設定する必要がある。それぞれの項目について詳しく説明する。

##### 5.3.1 MIDI 送信先

この項目では、MIDI 情報の送信先を決める。Max の起動時または MIDI 機器の検索ボタンを押した際に、PC につながっているシンセサイザー等の MIDI 機器がポップアップメニューに表示され、選択した MIDI 機器に情報が送信される。



図 10: MIDI 送信設定画面

### 5.3.2 CC ナンバー・ノートナンバー

トリガージェスチャーを扱う際に設定が必要になる項目であり、単発的に送る CC ナンバー

### 5.3.3 MIDI チャンネル

MIDI 機器によっては、送信先の設定のほかにも、MIDI チャンネルを設定する必要がある。

MIDI チャンネルは、1 本の MIDI ケーブルで複数の楽器(パート)の演奏情報を区別して送受信するための 1~16 の番号で、各パートにチャンネルを割り当てることで、異なる楽器を同時にコントロールすることができる。

本システムでも、1~16 の番号を指定することで、1 つの PC で複数の MIDI 機器を使い分けることが可能である。

### 5.3.4 MIDI CC 範囲

MIDI コントロールチェンジを使ってシンセサイザー等 MIDI 機器を制御する際に、どの値の範囲のみジェスチャーで変化させるかを設定することができる。

図 11 は設定つまみである。このつまみ両端の白と青の丸を動かすことで範囲を設定することができる。この範囲は図 12 のように、MIDI で制御するシンセサイザーのつまみ等のパラメータと模しているものになっており、どの範囲のみパラメータを変化させるかが視覚的に理解しやすく設定できる。

つまり、カメラからの距離でシンセサイザーを操る場合、図 11 の A の時は、カメラと手の距離が 1 番短い場合に CC 値 30 を出力し、1 番長い場合に CC 値 100 を出力する。

また図 11 の B のように、両端の白と青の丸が逆転している場合は、先ほどと逆方向に値が変化する。つまり、カメラと手の距離が 1 番短い場合に CC 値 100 を出力し、1 番長い場合に CC 値 30 を出力する。

### 5.4 トリガージェスチャーとコントロールジェスチャー

このシステムには、単発動作をトリガーに、指定した単発的な MIDI 情報を電子楽器等へ送信する [単発モード] と、細かな手指の動きによる連続的な値の変化によって MIDI CC のコントロールが可能な [連続] の 2 つのモードが存在する。



図 11: CC 範囲設定つまみ



左の画像のような CC 範囲設定は、右の画像(ソフトシンセのつまみ)の黄色の部分の範囲のみのパラメーターを変化させることを意味している。

図 12: CC 範囲設定の詳細

### 5.5 トリガージェスチャー

指定した MIDI CC とノートナンバー、そしてこれら 2 つのどちらかを時間的

#### 5.5.1 MIDI CC 機能

#### 5.5.2 ピッチ機能

#### 5.5.3 タイム機能

#### 5.5.4 DISTANCE 機能

最大 4 種類の MIDI CC を送ることが可能であり、1 つのジェスチャーで複数の MIDI CC を送信することもできる。

### 5.6 コントロールジェスチャー

指定した MIDI CC とピッチ範囲で、連続的な値変化

#### 5.6.1 対応ジェスチャー

#### 5.6.2 MIDI CC 機能

操作値が大きくなればなるほど、MIDI CC 範囲に従って、送信値が変化する。カメラ画角内で手が認識されなくなった場合、送信値は MIDI CC 範囲の最低値(白丸の値)に戻る。

また、ハンドサインアイコンを選択した場合は、選択したハンドサインを取った状態でカメラからの距離を変えることで、操作値が変化するようになり、ハンドサインを取っていない場合は、値の変化は起きない。カメラ画角内で手が認識されなくなった、もしくはハンドサインを取るのをやめた場合、送信値は MIDI CC 範囲の最低値(白丸の値)に戻る。

#### 5.6.3 ピッチ機能

操作値が大きくなればなるほど、スケールにしたがって高い音が出る、モノフォニック・シンセサイザーのように

表 1: スケール、ルート音、オクターブ範囲

変更する値	選択肢・値の範囲
スケール	major minor pentatonic minor-pentatonic harmonic-minor dorian mixolydian lydian phrygian locrian
ルート音の MIDI ノート番号	0 127
オクターブ範囲	1 3

使用することができる機能である。スケールとルート音、オクターブ範囲の設定によって、発音される音の高さと範囲が決められる。カメラ画角内で手が認識されなくなった場合は音が止まる。

また、ハンドサインアイコンを選択した場合は、選択したハンドサインを取った状態でカメラからの距離を変えることで、操作値が変化するようになり、スケールに沿って音が発音される。ハンドサインを取っていない場合は、値の変化や音の発音は起きない。カメラ画角内で手が認識されなくなった、もしくはハンドサインを取るのをやめた場合は音が止まる。

実装しているスケールと、入力できるルート音の MIDI ノート番号範囲と、オクターブ範囲は表 1 のとおりである。

## 5.7 機能説明

### 5.7.1 スムージング機能

MediaPipe から得られる連続値を扱う際に、測定誤差や環境要因によって、急激な値変化が起きてしまうことがある。これは、結果的に出力される音の音切れや不自然な音色変化につながってしまうため、どうにか対処する必要があった。そこで、本システムでは、値の時間的変化を平滑化するスムージング機能を導入することで、急激な値変化の抑制と、より自然かつ安定した MIDI 制御を実現した。

この機能は、Max 内の line オブジェクトを使用することで成り立っている。今回は、「100ms かけて前の値から新しい値へスムーズに変化」とすることで、スムージングを可能にしている。

## 6. 実装

## 7. 考察

## 8. おわりに

## 参考文献