

ハンドジェスチャーで制御するシンセ演奏システムの制作

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

— 目次 ——		
1	はじめに	1
2	インタラクティブ楽器	1
2.1	インタラクティブ楽器の歴史（仮）	1
2.2	本研究の狙い	1
3	システム構成	1
3.1	MediaPipe	1
3.2	Max	1
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）	
4	ハンドジェスチャーによる MIDI の制御	2
4.1	ユーザーが行う動作	2
4.1.1	基本姿勢	
4.1.2	ハンドサイン	
4.1.3	カメラからの距離	
4.1.4	手の開き具合	
4.1.5	つまみを回す動作	
5	システムマニュアル	3
5.1	TRIGGER モードと MOTION モード	3
5.2	TRIGGER	3
5.2.1	基本操作	
5.2.2	PITCH 機能	
5.2.3	DISTANCE 機能	
5.3	MOTION	3
5.4	機能説明	3
5.4.1	スムージング機能	
5.5	MIDI 送信設定	3
5.5.1	MIDI 送信先	
5.5.2	ノートナンバー・CC ナンバー	
5.5.3	MIDI チャンネル	
5.5.4	MIDI CC 範囲	
6	考察	4
7	おわりに	4
— 表目次 ——		
7	MIDI 送信設定画面	3
8	CC 範囲設定つまみ	4
9	CC 範囲設定の詳細	4

— 図目次 ——

1	MediaPipe での手情報取得の例	1
2	位置情報を取得する 20 箇所	1
3	4 種類のハンドサイン	2
4	カメラからの距離による値の変化	2
5	手の開き具合による値の変化	3
6	つまみを回す動作による値の変化	3

1. はじめに

現代のシンセサイザー等電子楽器を用いた演奏形態（マシンライブや DJ）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ため、ギターやドラム演奏に比べると、ライブ体験において演奏の印象や没入感が弱く感じる要因となり得る。

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えた。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增幅と、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. インタラクティブ楽器

2.1 インタラクティブ楽器の歴史（仮）

電子楽器の分野では、テルミンに代表される初期の非接触型楽器に始まり、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されてきた。特に近年では、加速度センサや深度センサを用いたジェスチャー入力により、演奏者の身体性を音楽表現へ拡張する試みが盛んに行われている。一方で、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、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。

2.2 本研究の狙い

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

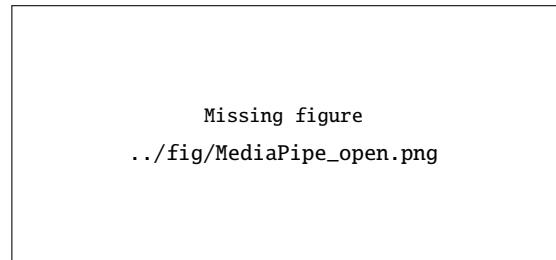


図 1: MediaPipe での手情報取得の例

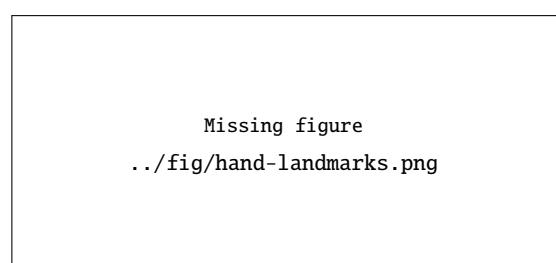


図 2: 位置情報を取得する 20 箇所

スチャーによって直感的な MIDI 制御が可能なエンター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、各ジェスチャーに対応した 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 1 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 2 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

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) [1] 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。MIDI 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

4.1 ユーザーが行う動作

本システムにおいて、ユーザーはハンドサイン、カメラからの距離、手の開き具合、つまみを回す動作といった動作を取ることで、MIDI を操ることができる。

4.1.1 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。

4.1.2 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（図 3）を自動認識し、MIDI 情報を送信するトリガーとして使用することができる。

4.1.3 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 4 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

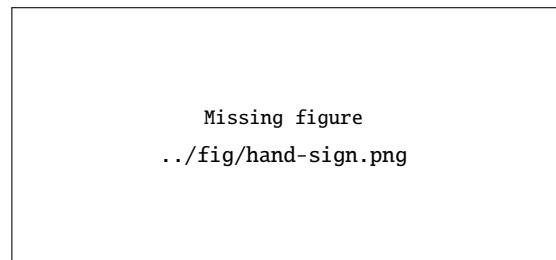


図 3: 4 種類のハンドサイン

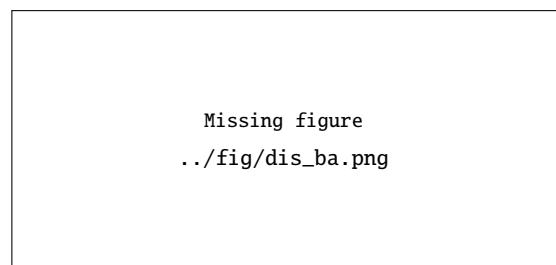


図 4: カメラからの距離による値の変化

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。

4.1.4 手の開き具合

本システムでは、手の開き具合に応じて、図 5 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。

4.1.5 つまみを回す動作

本システムでは、図 6 のようにシンセサイザー等についているつまみ・ノブを回すようなジェスチャーをカメラに向けて取ることで、連続的に送信値を変化させることができる

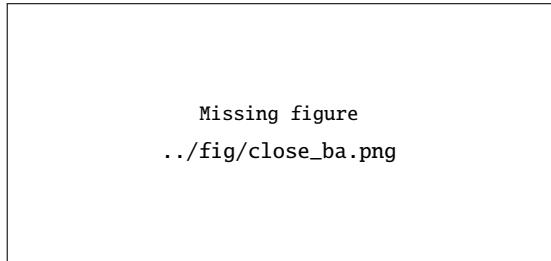


図 5: 手の開き具合による値の変化

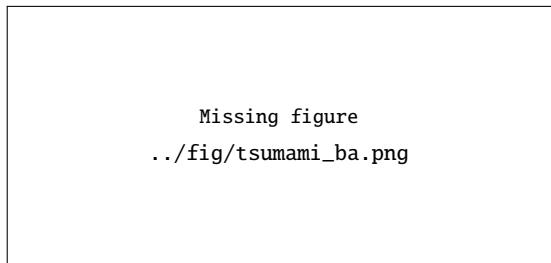


図 6: つまみを回す動作による値の変化

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。

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

$(f_1 : y \text{ 座標差}, f_2 : x \text{ 座標差})$

5. システムマニュアル

5.1 TRIGGER モードと MOTION モード

このシステムには、ハンドサインをトリガーに、指定した MIDI 情報を電子楽器等に送信する [TRIGGER] と、細かな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 [MOTION] の 2 つのモードが存在する。この 2 つのモードの同時使用はできない。

5.2 TRIGGER

[TRIGGER] を押してボタンを点灯または画角内で素早く手を下から上に振ると、TRIGGER モードになる。このモードはハンドジェスチャーの種類で指定した MIDI ノートを送信し、カメラからの距離でピッチをコントロールが可能な、モノフォニック・シンセサイザーのように使用することができる。

5.2.1 基本操作

トリガーとなるハンドサインを選択し、MIDI 送信先の機器、鳴らしたい MIDI ノートナンバー、MIDI チャンネルを選択することで、そのハンドサインを取ったと判定した時に、指定した高さの音が鳴る。

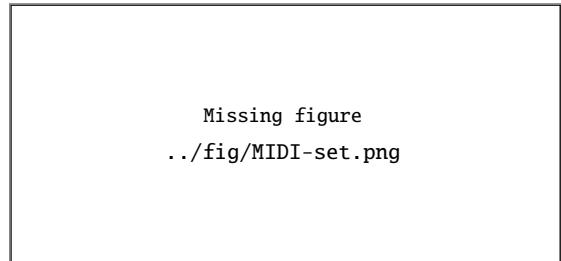


図 7: MIDI 送信設定画面

5.2.2 PITCH 機能

この機能を ON にすると、かざした手がカメラから遠いほど、スケールにしたがった高い音が出る。スケールとルート音、オクターブ範囲の変更が可能である。

5.2.3 DISTANCE 機能

この機能を ON にすると、かざした手のカメラからの距離で、指定した MIDI CC の値をコントロールすることができる。

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。

5.3 MOTION

[MOTION] を押してボタンを点灯または画角内で素早く手を上から下に振ると、MOTION モードになる。このモードは [カメラからの距離][手の開き具合][つまみを回す操作] のいずれかの操作で、MIDI CC をコントロールできる。

5.4 機能説明

5.4.1 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度々ある。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マーズに変化」とすることで、スムージングを可能にしている。

5.5 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 7 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。

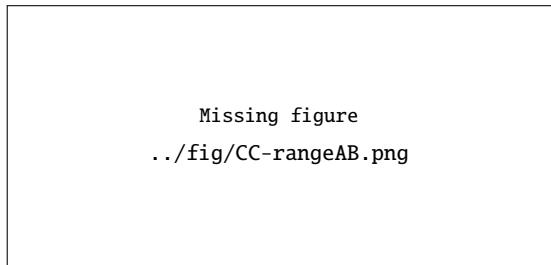


図 8: CC 範囲設定つまみ

5.5.1 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。

5.5.2 ノートナンバー・CC ナンバー

5.5.3 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。

5.5.4 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 8 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 9 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 8 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 8 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。

6. 考察

7. おわりに

参考文献

- [1] Cycling '74, <https://cycling74.com/>.

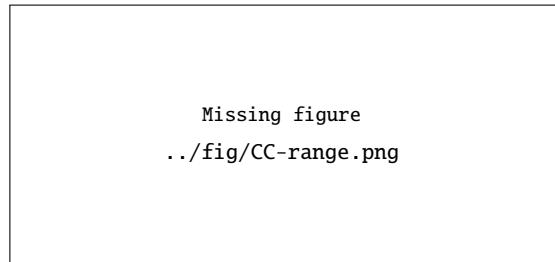


図 9: CC 範囲設定の詳細