

# GB-ABBX : ジェスチャーセンシングによるビートボックスパフォーマンスの拡張

大谷 健人<sup>1,a)</sup> 橋田 光代<sup>2</sup> 片寄 晴弘<sup>3</sup>

**概要:** ビートボックスは、人間の発声器官を使ってパーカッションサウンドを生成し、ビート、リズム、メロディなどの様々な音を作り出す音楽パフォーマンス技術である。最近ではサウンドエフェクトなどを用いた高度な技術や独自のスタイルが開発され、競技会も開催されるようになってきている。我々は IT 技術に応用したビートボックスパフォーマンスの拡張に取り組んでいる。シンポジウムでは、特にジェスチャーセンシングを用いたビートボックスパフォーマンスの拡張について技術的な構成を示すとともに実演を実施する。

## 1. はじめに

ヒューマンビートボックスは、人間の声道を多彩な打楽器に変える音楽技術であり、1970 年代の誕生以来、著しく進化してきた。初期のヒップホップ音楽を補完するために始まったビートボックスは、文化や地理的な障壁を超えて独立したパフォーマンスアートとしての地位を確立した。現在では、世界中にプロのビートボックスパーformerが存在し、さらには大勢の観客を集める世界的な競技会が毎年開催されるようになっている。

ヒューマンビートボックスの競技会にはいくつかの部門があり、中には、ボイスをトリガーとしたり、DAW やエフェクタとパラメータとして利用することが許される部門もある。この形態は、コンピュータ音楽の文脈で言えば、ライブエレクトロニクスと呼ばれる形態に相当する。ライブエレクトロニクスは 1960 年以降に成長した音楽形態の一つであり、これまでに様々な試みや技術的な蓄積がなされてきた [1] が、ビートボックスパフォーマンスとの接点は限定的である。

本研究では、ビートボックスパフォーマンスをライブエレクトロニクスの文脈で捉え、パフォーマンスジェスチャーによってパフォーマンスを拡張するシステム（環境）、『directaGesture Based Augmented Beatbox (GB-ABBX)』を提案する。

## 2. Human Beatbox

ヒューマンビートボックス [4] は、人間の発話器官、例えば口、舌、喉、唇などを使って様々な音を模倣する音楽の形式である。ヒューマンビートボックスは、音楽のリズムを作り出すための一部として、または単独のパフォーマンスとして使用される。この技術は、音楽の多様性を高め、新たな音楽的表現を可能にする。

ヒューマンビートボックスの起源は、1970 年代のアメリカ、特にニューヨークのヒップホップシーンに遡る。元来“ビートボックス”という呼称は、リズムパターンを自動で演奏するための装置の名称であった。ただ、その装置が高価なものであったため、当時それを購入するための十分なお金がない DJ やラッパーたちは、音楽のリズムを作り出すために口腔を使い始めた。そうして、ヒューマンビートボックスは音楽のリズムを作り出すための新たな方法として、または楽器が利用できない状況での即興的なパフォーマンスとして使用された。ヒューマンビートボックスは、ヒップホップ文化の一部として発展し、その後、世界中に広まった。

ヒューマンビートボックスは、その起源から大きく進化してきた。初期のヒューマンビートボックスは、主にシンプルなリズムとビートを模倣することのみに重点を置いていたが、時間とともに、より複雑なリズムと音色の模倣が可能になった。また、テクノロジーの進歩により、ビートボックスパフォーマンスはさらに進化し、ループステーションやエフェクトペダルを使用することで、より多様で豊かな音楽的表現を可能にした。これらの進歩は、ヒューマンビートボックスが音楽の新たな領域を探索するための

<sup>1</sup> 関西学院大学大学院理工学研究科

<sup>2</sup> 福知山公立大学情報学部

<sup>3</sup> 関西学院大学工学部

a) gwm00491@kwansei.ac.jp

プラットフォームとなり、その可能性を広げてきた。

### 3. GB-ABBX

ライブエレクトロニクスにおいては、パフォーマのジェスチャによって音楽パフォーマンスをリアルタイムで変容・展開する作業が行われる。ここでいうジェスチャとはいわゆるモーションジェスチャだけではなく音響的な特徴、音響ジェスチャも含まれ、さまざまな方式が検討されてきた。

先述の通り、ビートボックスパフォーマンスは様々な音楽機器やソフトウェアと組み合わせることで音楽的表現の幅が増えたが、音響ジェスチャとモーションジェスチャ双方の活用、多様なジェスチャからのトリガー利用やマッピング、音や映像処理を一体化することにより、表現の幅が飛躍的に向上する可能性がある。さらに、最近著しい技術的進展のある、カメラによる実時間での pose 推定技術を用いることで、ビートボックスパフォーマンスの可能性は大きく進むものとする。しかし、既存の音楽機器では、音の作り込みの段階、そして本番での実演においていくつか欠点が存在する。例えば、機器による音制御の場合、それらに内蔵されているエフェクトなどのパラメータを、機器を直接接触することでしか操作できないという点が挙げられる。そのため、必要な音制御の数に応じて必要な操作子の数が増えてしまい、マッピングの複雑化や、操作ミスなどによるパフォーマンス失敗率の増加が考えられる。また、機器によってボタンの数やペダルの数は限られているため、パラメータ制御の方法や個数が限定されてしまう。したがって、ビートボックスパフォーマンスにはさらなる IT 技術による支援が必要であるとする。

また、ビートボックスにおける競技会の規則の観点から考えた場合、その範疇を超えないような全く新しい IT 技術支援がいくつか可能であるとわかる。現在世界最大の規模を誇るビートボックス競技会である Grand Beatbox Battle(GBB)では、Producer Showcase という部門が存在する。この部門では、オーディオや MIDI に関するデバイス・ソフトウェアを無制限に使用することが許可されている。そして、Producer Showcase の Wildcard(動画予選) 規則 [2] の 5 番には "All sounds must be sourced or triggered by the mouth." と記載されている。つまり、"音のトリガーのためにマイクを使わないといけない" という規則ではないため、カメラに写る口の画像処理によって音を生成しても、現在の規則だけでは反則にすることが難しい。このように、現在の規則では規制できないような IT 技術支援はいくつか存在するため、それらを追求することで、現在の競技会において革新的なビートボックスパフォーマンスを行うことが可能であるとする。

GB-ABBX は、ビートボックスによって生成された音の制御を、カメラで捉えたポーズジェスチャーにより行うこ

とができる技術である。これにより、音楽機器を直接接触することなく機器内のパラメータなどを操作することができるようになるため、より柔軟で多様な音制御を行うことが可能となる。本研究では、手や顔のそれぞれのパーツの動きによって、ビートボックスに対する様々な音制御を割り当て・実行できるようにする。

### 4. 実装

GB-ABBX の実装には、MediaPipe と MAX を用いる。具体的には、MediaPipe によるリアルタイムジェスチャー検出を MAX 上で実装し、それによって得た情報を用いてビートボックスに対するオーディオ・MIDI 制御を行う。

MediaPipe[3] は、Google Research によって開発された、任意の感覚データに対して推論を実行するパイプラインを構築するためのフレームワークである。MediaPipe を用いることで、オブジェクト検出や顔の特徴点抽出、セグメンテーションなどが可能となる。

GB-ABBX では、大きく分けて 3 種類のジェスチャーセンシング手法を扱う。

- (1) ハンドトラッキング
- (2) フェイストラッキング
- (3) パターンマッチング

まず、ハンドトラッキングは、カメラにおける手の座標をリアルタイムで追跡することができる。具体的には、左手・右手それぞれに対して、手首に 1 つと、それぞれの指の先・第 1 関節・第 2 関節・付け根に 1 つずつ、合計 21 のポイントの座標を記録することができるが、本研究では手首の座標のみを制御に用いる。そして、フェイストラッキングも同様に、顔の座標を追跡することができる。これも顔にある様々なパーツの座標を追跡することができるが、本研究では口周りの座標のみを追跡し、それをもとに jawopen(顎の開き具合) の値を算出し、制御として用いる。最後に、パターンマッチングは、ハンドトラッキングやフェイストラッキングと組み合わせる。あらかじめジェスチャーのパターンを登録しておき、そのパターンに近いジェスチャーを検知することができる。

### 5. デモ

本研究では、以下の機器を用いてデモを行う。

- ダイナミックマイク "Shure BETA58A"
- オーディオインターフェース "Tascam US2 × 2"
- ノートパソコン "Macbook Pro"(MAX)
- ループステーション "BOSS RC-505 Mk2"

また、デモでは大きく分けて以下の 3 種類のジェスチャーセンシングによる制御を行う。

- (1) ハンドトラッキングによるボーカルのピッチ変換・エフェクト制御
- (2) フェイストラッキングによるシンセサイザーのトリガー

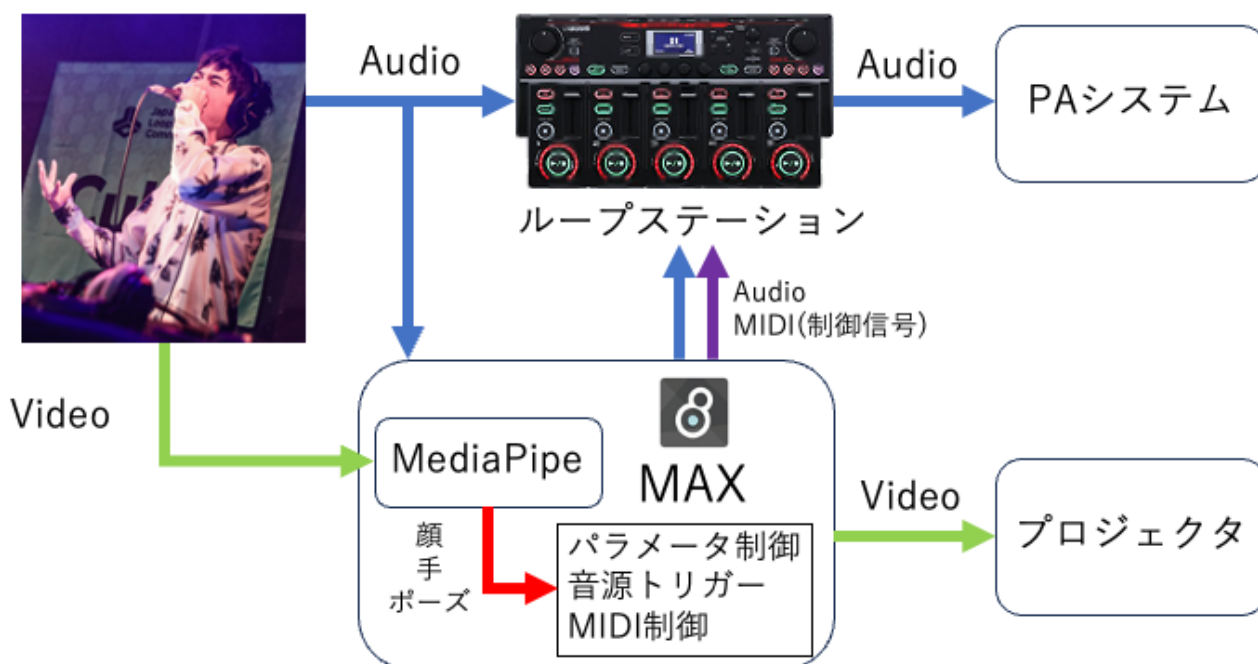


図 1 システムの全体図

### (3) ハンドトラッキングとパターンマッチングによる MIDI 送信

それぞれの制御を用いて、ループステーションによる曲を作成し、演奏する。システムの全体図は、図 1 の通りである。

ビートボックスパフォーマンスにおける入力音声と映像に分けられ、音声はオーディオインターフェースを通して、ジェスチャーによって制御する必要があるものは MAX、それ以外はループステーションに出力、映像は MAX 内の MediaPipe に出力される。MediaPipe によって、手・顔のジェスチャーと、パターンマッチングに使われるポーズが認識され、手のジェスチャーは、入力された音声に付加するエフェクトのパラメータ制御、顔のジェスチャーはシンセサイザーなどの音源のトリガー、パターンマッチングは決められた MIDI 情報の生成を行い、それぞれループステーションに出力される。最後に、直接入力された音声と、MAX を通して入力された音声が入力された音声がループステーションで合流し、それによってさらに加工されたオーディオが PA システムに出力される。

## 6. 展望とまとめ

今後の展望としては、GB-ABBX 自体の拡張と、ビートボックスパフォーマンスを拡張するためのさらなる手法の探索が挙げられる。本稿においては、特に手首と口の動きのみを制御として用いたが、今後はその他のパーツ等も制御に用いることができるように拡張していく予定である。

また、GB-ABBX というのは、あくまで ABBX のうちのひとつにすぎない。今後はまた別の手法を用いてビートボックスパフォーマンスを拡張していく予定である。

本研究は、様々なポーズジェスチャーによってビートボックスパフォーマンスを拡張するための手法、GB-ABBX を提案した。また、その技術を用いた曲を作成し、実演を行った。

## 参考文献

- [1] Cook, P.: Principles for Designing Computer Music Controllers, *Proceedings of the CHI'01 Workshop on New Interfaces for Musical Expression (NIME-01)*, Seattle, USA (2001).
- [2] Kilaa: GBB 2023 Wildcard Competition, Swissbeatbox (online), available from (<https://swissbeatbox.com/newsfeed/gbb-2023-wildcard-competition/>) (accessed 2023-07-28).
- [3] Lugaresi, C. et al.: MediaPipe: A Framework for Building Perception Pipelines (2019).
- [4] 河本洋一: 日本におけるヒューマンビートボックスの概念形成, *音楽表現学*, Vol. 17, pp. 33-52 (2019).