

画像によるドラムスティックの動作解析

上 遠 野 優^{†1} 子 安 大 士^{†1} 前 川 仁^{†1}

打楽器のリズムや演奏動作は音や曲想と密接な関わりがあり、その習得は非常に重要である。リズムの習得は様々な譜例を練習することで行えるが、演奏動作はその習得が難しく、学習の支援となるものが必要である。そこで本研究では、2台のカメラによる動画像から演奏動作を抽出し、どのようなリズムをどのような動作で演奏しているかを判定・提示するシステムの構築を目的とする。演奏動作として、基本の動作である4 basic strokes と、腕の動きに特徴のある Moeller stroke に着目する。本論文では動画像から演奏動作を抽出する手法と、そこからストロークを判定する特徴量を提案する。

Motion Analysis of Drumming by Image Processing

YU KATONO,^{†1} HIROSHI KOYASU^{†1} and HITOSHI MAEKAWA^{†1}

It is a hard work for players to acquire proper rhythm and adequate drumming motion while the both affect directly on sound and motif of the music.

To support its learning, we propose a method of analyzing the rhythm and drumming motion of the player by using image processing. In the method, several feature points including the stick tip, the grip, and the elbow of the player are imaged by two high-speed cameras and their movements are tracked by Kalman filter. The fourbasic stick movement and advanced Moeller stroke are both detected.

1. はじめに

打楽器の演奏において、曲に合ったリズムや音色、感情などを表現するために、演奏動作の習得は重要である。しかし、リズムや曲想に合った演奏動作を身につけるのは、熟練者による直接指導がなければ困難である。そこで演奏動作の学習支援のために、音に直接関係した演奏動作であるストロークに関する動作解析や、測定した学習者のストロークに対して、熟練者との違いを教示するシステムの開発が行われている¹⁾²⁾³⁾。これらは主に、一定の間隔で特定のストロークを行う場合の解析であり、打拍毎にストロークが異なるリズムについてはあまり触れていない。そのため、実際の演奏時に必要となるリズムと演奏動作の関連性の理解を得るのが難しい。

そこで本研究では、打楽器演奏を撮影した動画像から、どのようなリズムを演奏し、それに対して演奏者がどのようなストロークで演奏したかを判定・解析するシステムの構築を目指す。自分がどのようなストロークで演奏していたか、それが理想のストロークと

どう異なるかを知ることができるようなシステムがあれば、打楽器の学習支援に役立つと考える。本論文では演奏動作として4つの basic strokes と、特殊な奏法である Moeller stroke に着目する。

動作を解析する際、動作情報を得る手段としてモーションキャプチャシステムがよく使用されている。これは、マーキングした部分の3次元位置を比較的正確に得られる利点があるが、被験者の体にマーカーを装着する必要がある上に、そのシステムの規模は大きく、専用の部屋を要する場合が多い。しかし、学習支援として使用する場合を考えると、被験者が普段学習しているような環境で、容易にシステムを利用できるのが理想的である。

本研究では、カメラとPCのみで学習支援システムを構築することを考えている。カメラから動作情報を得ることで、マーカー装着等の被験者の負担を減らすことができ、さらにこの程度の機材であれば、被験者が普段学習している環境に設置できるため、容易にシステムを利用することができる。カメラは3次元情報を得るためには2台必要であるが、本論文の実験で示すように配置を工夫すれば1台のカメラでも十分に動作の解析は可能である。本論文では、動画像から画像処理や演奏動作の特徴を用いて演奏動作情報を得る手

^{†1} 埼玉大学大学院理工学研究科
Graduate School of Science and Engineering, Saitama University

法と、得られた情報から演奏動作の解析をする手法を提案する。

2. 打楽器演奏動作について

打楽器の演奏動作は音に直結するものであり、また曲想を表現をするものでもある⁴⁾⁵⁾。どのようなリズムをどのような動作で演奏するか、すなわちリズムに対する演奏動作の選びかた次第でその表現の仕方が変化する。

よく知られている動作として、4つの **basic strokes** がある⁶⁾。これは大きい音や強く打つ音 (アクセント) と小さい音を叩き分ける上でもっとも基本的なものであり、これを熟練することによってリズムを明確にしたり、曲のビート感を表現することができる。一つの打点につき一つの **basic stroke** が対応している。

また、アクセントが絡んだ複数打で構成される一つのグループを、大きな一つのストロークで演奏する **Moeller stroke** というのも存在する⁶⁾。また、本論文ではそのグループを **Moeller stroke** を用いないで演奏する場合を **Normal stroke** と呼ぶ。一つのグループにつき、どちらかの **stroke** が対応している。

2.1 単打のための演奏動作 – 4 basic strokes

演奏動作の基本となっている 4 **basic strokes** は、各打点の前後のスティックのポジションによって決まる。腕を振りあげた状態をアップポジション、打面付近に止めている状態をダウンポジションとすると、次のように区別される。

- **Full stroke**

アップポジションで動作開始し、打点後にリバウンドしてアップポジションに戻る。主に大きな音やアクセントを連続して演奏する場合に使用する。

- **Down stroke**

アップポジションで動作開始し、打点後にリバウンドを押えダウンポジションに止める。主に大きな音やアクセントで、その直後を小さく打つ場合に使用する。

- **Tap stroke**

ダウンポジションで動作開始し、打点後にダウンポジションに戻る。主に小さな音を連続して演奏す場言いに使用する。

- **Up stroke**

ダウンポジションで動作開始し、打点後にアップポジションまで戻す。主に小さな音で、その直後を大きく打つ場合に使用する。

2.2 複数打のための演奏動作 – Moeller stroke

アクセントが絡んだリズムを演奏するときやテンポ

が速いとき、滑らかな動きで演奏したい時に、**Moeller stroke** で演奏される場合がある。この奏法は肩胛骨から手の先までの関節とスティックのリバウンドをフル活用した鞭のような動きで、複数打で構成されるリズムパターンをひとつのストロークで演奏できるという奏法である。

本来は肩胛骨や指の細かな動きにも着目すべきであるが、本研究では画像からその動作を認識しやすいスティックの動きに着目する。直接音に関わる部分なのがスティックの動きであり、また **Moeller stroke** は肩や肘を動かさない場合も含むため、スティックの動きのみでこれを判定することを考える。前述のように、本論文では **Moeller stroke** を使用していない場合を **Normal stroke** と呼ぶ。

3. 動画像からの演奏動作の抽出

3.1 演奏動作の撮影

2台のカメラをどちらからも打面及び奏者が移るように配置し、同期をとって撮影を行う。今回想定しているカメラは Point Gray Research 社の Dragonfly Express である。フレームレートは 100fps で撮影を行う。また、動作の 3次元位置を取得するため、両カメラともキャリブレーションされているものとする。

本研究では **Moeller stroke** という高速に演奏する場合もカメラからの映像のみで解析しようとしているため、ある程度以上の高速カメラを必要とする。カメラと同期のとれたマイクを使用するなどして打点を正確に検出できれば、それを基に演奏動作を補完することで、より低速のカメラでも解析できる可能性はある。

3.2 打面座標系の提案

Intel の画像処理ライブラリである OpenCV を用いて簡易にキャリブレーションできるツールを作成し、カメラキャリブレーションに使用する。この簡易な実装によるキャリブレーションを行うと、校正用パターン上に原点を置き、世界座標系を定義する。しかし、打楽器の演奏動作を解析するにあたってこの座標系は使いづらい。そこで、打面を基準とした座標系である打面座標系を提案する。

打面座標系とは、原点を通る $x-z$ 平面が打面上にあり、打面に対して垂直に y 軸をとるような正規直交基底である。打面とスティックの位置関係が座標値からだけでも直感的に理解しやすく、動作解析に適していると言える。変換についての詳細は省略するが、カメラの外部パラメータを打面座標系との相対的な姿勢を表すように更新することで、この座標系を定義して

いる。

以降、本論文における3次元座標は、この打面座標系における座標値とする。

3.3 演奏動作の追跡

本論文では演奏動作としてスティックの先端であるチップ、持ち手の先端であるグリップ、手首、肘の4つの動作点の挙動に着目する。

演奏者に負荷をかけないように各動作点の追跡を容易にするため、撮影時に肘が見えるような服装で行うという状態を想定する。

初期フレームで各動作点の初期位置とテンプレートを作成し、その後の追跡には毎フレーム各動作点ごとにテンプレートマッチングを用いる。しかし、毎フレーム、画像の全範囲に対して探索を行うのは計算量が多く、演奏しているのと反対の手が認識されたりと誤認識も増えてしまう。そこで、次のような前提知識を基に探索の効率化を図る。

- (1) 背景差分による演奏者領域の抽出
- (2) スティックと手首の色情報に基づく探索領域の限定。
- (3) 動作点間の3次元空間内での距離の拘束。
- (4) Kalman Filterによる状態推定と予測。
- (5) エピポーラ拘束。

(1)について、あらかじめ演奏者が映っていないような画像を撮影し、それを基に背景差分を行う。(2)について、スティックの材質は木でありその多くは人の肌色に近い色をしている。よって画像内で肌色に近い部分だけを探索するようなマスクを生成する。以降、(3),(4),(5)について説明する。

3.3.1 動作点間の距離の拘束

演奏動作中において、腕と手、そしてスティックはほぼ剛体であるといえる。そしてそれらを連結すると、各動作点を間接としたリンク機構となる。すなわち各動作点間の3次元上での距離は常に一定である。よって、ある一点の位置が決定すれば、その前後の動作点の位置には距離の拘束ができる。動作点の中でもっとも追跡しやすいのは動きが少なく画像上の見かけの変化が少ない肘であるので、肘、手首、グリップ、チップの順で位置を特定すればこの拘束を利用できる。

まず肘の3次元位置を決める。そして、腕の姿勢が前時刻と同じであると仮定した場合の手首の推定位置を求める。その位置を基準に、1フレームのうちに肘を中心に腕は y 軸に平行な平面上を $\pm\Delta\theta$ 回転しようとして、手首が移動する可能性がある軌道を計算し、それを各カメラの画像に射影した部分を手首の存在範

囲とする。これを手首からグリップ、グリップからチップへと作業を行い、各動作点の画像上の存在範囲を限定する。

3.3.2 Kalman Filterによるスティック姿勢の推定と予測

高速カメラを用いているため、追跡する動作点の中で移動速度の速いチップとグリップの動きでも、フレーム間のごく短い時間では、線形に遷移するシステムとみなせる。

ここで、テンプレートマッチングの結果を観測値とし、Kalman Filterを用いてスティックの運動の時系列予測と推定を行う。線形に状態遷移するシステムの状態方程式は以下のように記述される⁷⁾。

$$\mathbf{x}_{t+1} = \mathbf{G}\mathbf{x}_t + \mathbf{w}_t \quad (1)$$

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{u}_t \quad (2)$$

ここで、 \mathbf{x}_t は時刻 t におけるシステムの内部状態を表すベクトル、 \mathbf{y}_t は時刻 t における観測値を表すベクトルである。また、 \mathbf{G} は内部状態の遷移を表す行列、 \mathbf{H} は内部状態を観測値に写像する行列である。 \mathbf{w}_t 、 \mathbf{u}_t はそれぞれプロセスノイズと観測ノイズである。ここで、チップの動きを平常時には等速運動モデル、運動方向が反転する打点時には完全弾性衝突モデルに近似する。位置 (x, y) 、速度 (\dot{x}, \dot{y}) を用いて \mathbf{x}_t 、 \mathbf{y}_t 、 \mathbf{G} 、 \mathbf{H} を以下のように定義する。

$$\mathbf{x}_t = \begin{pmatrix} x_t \\ y_t \\ z_t \\ \dot{x}_t \\ \dot{y}_t \\ \dot{z}_t \end{pmatrix}, \mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & \alpha & 0 & 0 \\ 0 & 1 & 0 & 0 & \alpha & 0 \\ 0 & 0 & 1 & 0 & 0 & \alpha \\ 0 & 0 & 0 & \alpha & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha \end{pmatrix}$$

$$\mathbf{y}_t = \begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix}, \mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

等速運動モデルでは $\alpha = 1$ となり、この時の \mathbf{G} を \mathbf{G}_1 、完全弾性衝突モデルでは $\alpha = -1$ となり、 \mathbf{G}_2 とする。各時刻においてどちらのモデルを選択するかは、それぞれのモデルを用いて予測した状態と、得られた観測値のマハラノビス距離を計算し、その値が小さいモデルを用いることで決定する。

これらにより得られた各時刻におけるチップとグリップの位置から、スティックの運動を決定する。

3.3.3 エピポーラ拘束

2台のカメラによって得られた画像の対応点を m_1 、 m_2 とおくと、次式が成り立つ。

$$\bar{m}_2^T F \bar{m}_1 = 0 \quad (3)$$

これをエピポラ拘束といい、 F は 2 台のカメラ間の相対的な位置や姿勢を表す行列である⁸⁾。片方の画像である 1 点が決まると、それに対応するもう片方の画像上の点は必ず上の拘束を満たすエピポラ線上に存在するため、探索範囲を限定することができる。

本研究では、各カメラのパラメータは既知であり、カメラ間の位置関係を表す F 行列を得ることができる。よって、第 1 カメラからの画像上で位置が分かれば、第 2 カメラの画像中の対応点はエピポラ拘束によりその存在範囲を限定できる。

これらの条件を利用することで、各動作点の位置探索の効率化を図る。

3.4 追跡実験

前述の手法で動作の追跡を行っている成功例を図 1 に示す。左側が第 1 カメラ、右側が第 2 カメラで得られた画像で、上から下に時系列になっている。白い丸で描かれた部分が追跡している各動作点の検出位置である。どちらのカメラからの画像でも各動作点が追跡できているのが確認できる。

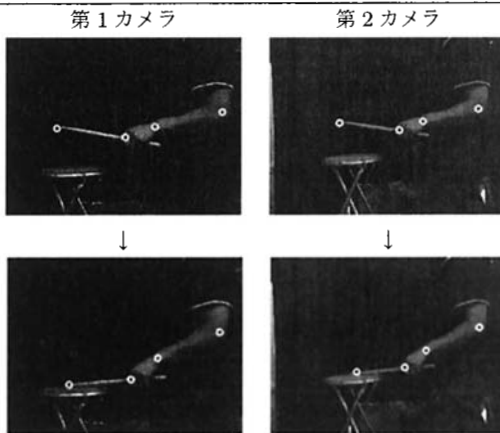


図 1 演奏動作の追跡成功例

しかし、図 2 に示すようにチップの追跡に失敗する場合もある。これは肘や手首、グリップの 3 次元位置の精度が悪く、チップの位置推定がうまくいっていないためである。検出される動作点の体に対する位置がフレーム毎にわずかに異なり、その誤差が 3 次元位置の精度に影響していると考えられる。テンプレートマッチングの動的更新や、画像上でのサブピクセル精度の位置検出が必要である。

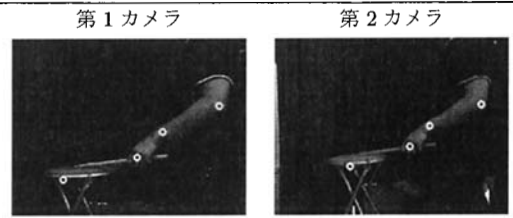


図 2 演奏動作の追跡失敗例

4. 演奏動作解析手法

追跡で得られた各動作点の軌道から、リズム推定及び動作解析を行う。ここで、解析するリズムの前に 4 拍分 4 分音符を演奏することを前提条件とする。解析をカメラの画像のみで行うため、リズムの基準となる 4 分音符の長さを把握するためである。

4.1 リズム推定

得られたチップの軌道から、 y 座標値がほぼ 0 になり、その前後で運動方向が変わる時を打点とする。この打点を基にリズムを推定する。

i 打目と $i+1$ 打目の間隔を h_i 、4 分音符の長さを H とすると $\frac{h_i}{H}$ は 4 分音符に対するその打点の長さである。各打点ごとにその長さを求めることによって演奏しているリズムを推定する。

4.2 特徴量の提案

動画からの演奏動作推定により、チップやグリップ等の位置と速度、スティックの角度、角速度の時系列データが得られる。これらのデータを基に、2 章で示した演奏動作を定量的に判定する以下の 5 つの特徴量を提案する。

- **ストローク比：Stroke Ratio (St_r)**
打点前後の打面に対するチップの最高点を各々 h_1 , h_2 とした時、その比である h_1/h_2 .
- **グリップ先行度：Grip Advance (Ga)**
チップ、グリップの動き出しから打点までの時間をそれぞれ t_1 , t_2 とした時、その比 t_2/t_1 .
- **打拍角度：Hit Angle (Ha)**
打点時における打面に対するスティックの角度.
- **打拍速度：Hit Speed (Hs)**
打点時におけるチップの速度を、演奏中のチップ速度の最大値で正規化した値.
- **腕動作分散：Arm variance (Av)**
打拍動作開始から終了までの手首、肘の位置の分散値。 x , y , z の 3 成分を持つ.

4.3 4 basic strokes の判定

2 章で説明したように、4 basic strokes は打点前後

のスティックポジションと、打点時のチップのスピードで区別されるので、ストローク比 Sr 、打拍速度 Hs を用いてこれを表現できる⁴⁾。各 basic strokes の Sr - Hs に対する分布は図3のようになっている。各ストロークを判定する閾値 s_1 , s_2 , v は実験的に値を求める。我々の行った予備実験から $s_1 = 0.7$, $s_2 = 1.45$, $v = 0.5$ と設定する。また、 $Hs > v$ となる打点はアクセントであるとする。

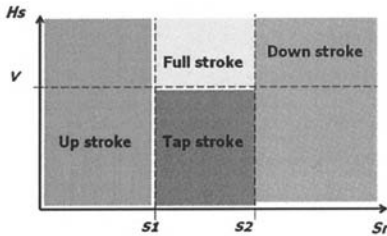


図3 4 basic strokes の Sr - Hs における分布

4.4 Moeller stroke の解析

2つのアクセントで囲まれた打点のグループをどのように演奏しているかで、Moeller, Normal stroke の判定を行う。各 stroke で演奏している時の、アクセント間の n 打における打拍角度 Ha の推移は図4のようになる。ここで、 i 打目から $i+n$ 打目までの Ha のグラフを、最小自乗法により直線 $Ha = Hg \times h + b$ に近似する。すると、Moeller stroke の時は Normal stroke に比べ、傾きを示す Hg が大きい値をとるのが分かる。

また、アクセント2打目を打つ直前(図中の $i+3$ 打目)ではグリップが高い位置にあるため、次打の際にチップに対してグリップが先行して動き出す。すなわちアクセント2打点目におけるグリップ先行度 Ga が大きい値をとる。これらを基に Moeller stroke の判定基準を提案したのが図5である。図中の閾値 g , a は実験的に値を求める。我々の行った予備実験から $g = 2.7$, $a = 1.3$ と設定する。

4.5 打点回りの腕の動作

打拍動作開始から終了までの手首や肘の動き、すなわち腕全体の動きもストロークの情報である。 Av の値が大きければその打点時に大きく腕を動かしていた、小さければ腕は動かさず手や指で打っていた、というようなストロークに対する付加情報を記述できる。

5. 動作解析シミュレーション実験

前章で述べた動作解析手法についてシミュレーション

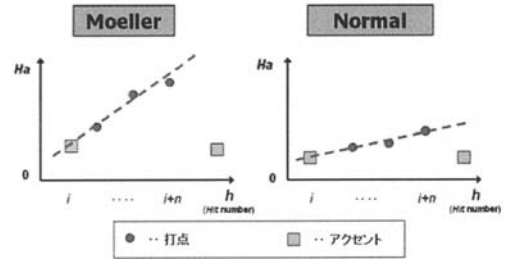


図4 アクセント間の打拍角度推移

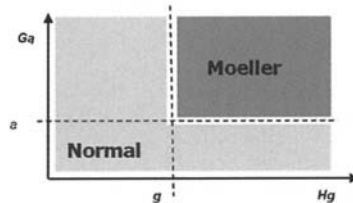


図5 Moeller, Normal stroke の Hg - Ga 分布

を行う。ここでいうシミュレーションとは、3章で述べた追跡手法を1台のカメラに適用し、得られる2次元データに付いて解析を行うことである。画像中の各動作点間の距離がほぼ一定であるように、撮影は演奏者の真横から行う。

5.1 実験内容

- 4 basic strokes を指定した順番で繰り返し演奏している動作を撮影し、指定通りにストローク判定が行われるかどうかを検証する。
- 同じ譜例を Moeller stroke, Normal stroke の両方で演奏し、その違いを正しく認識できるかを検証する。

5.2 結果

4 basic strokes を判定した結果を図6に示す。各打拍に対しどのストロークを指定したかを図中右上に記述されているような形状で表した。ほとんどの打拍について指定通りに判定が行われている。例えば、Full stroke と指定して演奏した打拍は、図3で Full stroke が分布する領域に集中している。

また、 Av の各ストロークにおける平均値を表1に示す。Tap stroke が他に比べ極端に小さい値を取って

表 1 各ストロークの平均 A_v

| A_v の平均値 | Full | Down | Tap | Up |
|------------|-------|-------|------|--------|
| x 成分 | 15.55 | 10.01 | 2.51 | 18.78 |
| y 成分 | 23.14 | 35.70 | 3.72 | 34.776 |

いる事が分かる。しかし、Tap stroke であっても他のストロークの値と同じような大きい値を取ることもある。すなわち、Tap stroke には腕を大きく動かしている打拍である、といったストローク以外の特徴を A_v の値により記述できることが分かる。

Moeller stroke と、Normal stroke を判定した結果を図 7 に示す。図中の 1 点は複数打拍で構成された 1 つのパターンを表している。こちらもほとんどのパターンについて、図 5 で想定した分布に従っている。

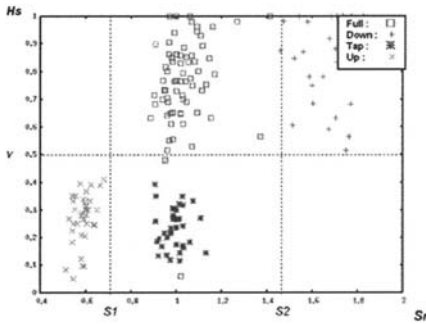


図 6 実験で得られた各ストロークの $Sr-Hs$ における分布

6. ま と め

本論文では、打楽器学習における演奏動作の習得の重要性と、いくつかの演奏動作の特徴について述べ、打楽器演奏シーンであることを利用した様々な条件をつけることで、動作の追跡を効率化し、得られるデータからストローク解析に適した特徴量と、それを用いた解析手法を提案した。

演奏動作の 3 次元追跡では、肘、手首、グリップの追跡精度は良好であった、しかし、チップの追跡精度が悪く、現時点では動作解析を行えるほどのデータが出ていないため、追跡精度の向上が必要である。また、動作解析シミュレーションでは、4 basic strokes, Moeller stroke 共に精度良く判定を行えることが確認

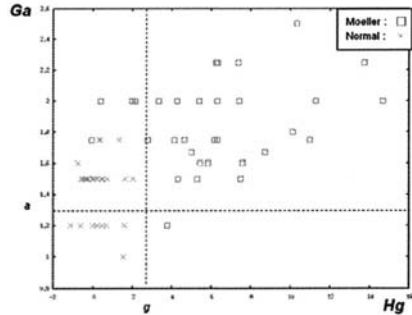


図 7 実験で得られた Moeller, Normal stroke の $Hg-Ga$ 分布

できた。これらをシステム化し、性能を評価していく必要がある。

さらに、より学習支援効果を得るため、3次元データを利用した演奏動作の任意視点映像も検討している。

参 考 文 献

- 1) S. Dahl and R. Huys, R, "Motor behaviour in drumming with different striking tools: Effects of tempo on control of rebound", The Neurosciences and Music - II. From perception to performance, May 5-8, 2005.
- 2) 辻, 西方, "リズムと打拍フォームの同時測定に基づく打楽器の演奏分析", 信学論, vol.J88-D-1, No.1, pp.99-107, 2005.
- 3) 辻, 西方, "リズムと打拍フォームに基づく打楽器学習支援システムの開発と評価", 信学論, vol.J88-D-1, No.2, pp.508-516, 2005.
- 4) 上遠野, 子安, 前川, "画像によるドラムスティックの動作解析", 第 5 回情報科学技術フォーラム (FIT2006), K-028, 2006.
- 5) S. Dahl and A. Friberg, "What can the body movements reveal about a musician's emotional intention?", In: Bresin, R. (Ed.), Proceedings of Stockholm Music Acoustic Conference (SMAC03), Vol II pp. 599-602, 6-9 August, 2003.
- 6) Jeff Queen, "THE NEXT LEVEL - rudimental snare drum techniques", Mark Wessels Publications, 2004.
- 7) 片山 徹, "新版応用カルマンフィルタ", 朝倉書店, 2000.
- 8) 佐藤 淳, "コンピュータビジョン - 視覚の幾何学 -", コロナ社, 1999.