

アクティブ音響センシングを用いた手指の関節角度推定

小西 智樹^{1,a)} 角田 陸^{1,b)} 國分 晴利^{2,c)} 久保 勇貴^{3,4,d)} 志築 文太郎^{5,e)}

概要：リアルタイムの手指の関節角度の推定が行えると、ハンドジェスチャを用いたアプリケーションへの入力が可能になる。しかし、現在用いられている光学式のハンドトラッキングシステムには、トラッキングできる画角が固定されている、カメラの死角ではトラッキングができないなどの問題がある。そこで、本研究では手の位置に制限されず、周囲の環境にセンサを設置しないシステムにより手指の関節角度を推定することを目的とする。具体的には、アクティブ音響センシングを用いて手の周波数特性を取得し、この変化に基づき機械学習を用いて指の関節角度推定を試みた。本研究において、人差し指および中指の第3関節角度推定モデルを構築し、この精度調査を行った。結果、人差し指 $\pm 8.53^\circ$ および中指 $\pm 5.51^\circ$ の平均誤差の精度において各指の第3関節角度を推定可能であることを示した。

1. はじめに

現在用いられている OptiTrack [1] および HTC VIVE Pro 2 [2] などのハンドトラッキングシステムの多くは、周囲の環境に配置したセンサおよびカメラに依存するものが多い。これらには、トラッキングできる画角が固定されている問題およびカメラの死角ではトラッキングができないという問題がある。また、周囲の環境にセンサを設置する必要がある。それに対して、センサおよびカメラが配置されている環境以外において手指の関節角度推定ができると、様々な場面においてハンドジェスチャを用いた入力が可能となる。これらの問題に対して先行研究では、手袋型デバイス [3]、および手首に装着したカメラ [4] を採用した手指の関節角度推定が行われてきた。しかし、手袋を用いた場合、ユーザの手を覆う形態であるために、手指の動作を妨げる場合がある。また、カメラを用いた場合、プライバシーの問題が生じる。よって、本研究では、カメラを用いず、かつ手指の動作を妨げない形態において、手指の関節角度の推定を試みる。

我々は、カメラを用いず、かつ手指の動作を妨げずに手指の関節角度推定を実現するために、アクティブ音響セン

シングを活用する。アクティブ音響センシングとは、物体の形状、材質、および境界条件などに依存した固有の共振特性の変化を観察することにより、物体の状態の変化を取得する手法である [5]。アクティブ音響センシングには、マイクおよびスピーカの役割を持つ 1 対 1 のピエゾ素子を、センシング対象に貼り付ける必要がある。久保ら [6] は、手の甲に素子を貼り付け、アクティブ音響センシングを用いてハンドジェスチャの認識を行った。この研究では、手指の形状の推定が可能である一方、手指の関節角度の推定まではできていない。そこで本研究では、手指の各関節角度における手の共振特性を特徴量とした機械学習により、手指の関節角度推定を行う。手指の関節角度の実測値には、OptiTrack を通じて取得する手指の関節角度を用いる。手指の各関節角度における手の共振特性を各指の関節角度の実測値と紐づけ、これをラッソ回帰を用いて学習する。回帰モデルを用いることによって、共振特性から手指の関節角度を直接推測値として得られるモデルを構築する。

本稿において、アクティブ音響センシングによる手指の関節角度推定の基礎検討を行うために、人差し指および中指の第3関節角度推定の精度評価を行うための実験を実施した。1 名の実験参加者から収集したデータを用いて、回帰モデルによる各関節角度の推定値の平均誤差を求め、その精度を評価した。結果、人差し指および中指の第3関節角度をそれぞれ $\pm 8.53^\circ$ および $\pm 5.51^\circ$ の平均誤差において推定可能であることが示された。

2. 関連研究

本研究では、アクティブ音響センシングを用いて、手指

¹ 筑波大学 情報学群
² 筑波大学 情報理工学位プログラム
³ 筑波大学 システム情報工学研究科
⁴ NTT人間情報研究所
⁵ 筑波大学 システム情報系
a) konishi@iplab.cs.tsukuba.ac.jp
b) tsunoda@iplab.cs.tsukuba.ac.jp
c) kokubu@iplab.cs.tsukuba.ac.jp
d) kubo@iplab.cs.tsukuba.ac.jp
e) shizuki@cs.tsukuba.ac.jp



図 1 実験に用いたピエゾ素子およびケーブル。

の動作を妨げない形態において手指の関節角度を推定することを目的としている。そのため、本節においては、これまでに提案されたアクティブ音響センシングを用いた手指のジェスチャ識別手法、およびアクティブ音響センシングもしくはその他のセンサを用いた手指の関節角度の推定手法から本研究を位置づける。

2.1 アクティブ音響センシングを用いたジェスチャ識別手法および関節角度推定手法

アクティブ音響センシングを用いて複数種類のジェスチャを識別する手法が提案されている。

加藤ら [7] は、手首にマイクおよびスピーカを取り付けることによって、7 種類のハンドジェスチャを識別するシステムを提案した。横田ら [8] は指先にスピーカ、手首にマイクを取り付けることによって、8 種類のハンドジェスチャ、9 箇所の前腕部のタッチ、および 9 箇所の手の甲のタッチを識別するシステムを提案した。久保ら [6] は手の甲にマイクおよびスピーカを取り付けることによって、親指と他の指の指先、腹および付け根のタッチの識別を行った。また、親指を浮かせた状態および握った状態において親指が左、右および中央のどこにあるのかの識別および人差し指による手のひらの上、左、右、中央および下部分のタッチ箇所の識別を行った。これらの研究はいずれもあらかじめ決められたジェスチャの識別を行うものであり、指の関節角度の推定はできていない。これに対し、本研究ではアクティブ音響センシングを用いた関節角度の推定を目的としている。

Okawa ら [9] は人差し指に取り付けたスピーカおよび中指に取り付けたマイクを用いて人差し指の第 3 関節の角度を推定するシステムを提案した。この研究はアクティブ音響センシングを用いた関節角度の推定を行っているが、単一の周波数帯域しか利用していないため、複数の関節の角度推定が困難であると考えられる。そのため、本研究では

複数の周波数帯域を利用することにより、複数の関節の角度推定を行う。

2.2 その他センサを利用した関節角度の推定手法

Glauser ら [3] は手袋に静電容量センサを埋め込み、手形状を推定するシステムを提案した。また、Wu ら [4] は手首に装着したカメラを用いて指の動きに伴う手の甲表面の変化を捉え、各指の各関節角度を推定するシステムを提案した。

これらの研究において、手袋型デバイスは手指の動作を妨げる場合があり、カメラを用いる場合はプライバシーの問題があると考えられる。

3. 提案手法

OptiTrack を用いて取得した手指の 3 次元位置データから計算した手指の第 3 関節の角度を正解データとした機械学習を行うことにより、アクティブ音響センシングによって取得したデータのみを用いて手指の関節角度の推定を行う。

3.1 ハードウェア

ハードウェアは、久保ら [6] と同様にコンピュータ、オーディオインタフェース、およびピエゾ素子から構成されている。手に貼り付けるマイクおよびスピーカとしてピエゾ素子 (THRIVEK2512BP1, $25 \times 12 \times 0.23$ mm) を用いた。ピエゾ素子は 2 芯シールド (モガミ電線 超極細フレキシブル 2 芯シールド AWG36) を通してラインプラグ (Pro Audition AS-106) と接続され (図 1)、オーディオインタフェース (Steinberg UR24C) を通してコンピュータ (Windows11, CPU: Intel Core i9-12900K 3.20 GHz, RAM: 32 GB) に接続された。

3.2 ソフトウェア

ソフトウェアは Touch&Activate [5] を参考にしており、チャープ信号生成部分および音響解析部分から構成されている。

チャープ信号生成部分では 20–40 kHz の間のチャープ信号を連続的に流している。このチャープ信号は 20 ms で繰り返しされ、周波数は直線的に増加する。サンプリングレートは 96 kHz である。

音響解析部分ではマイクから得られた音響信号を周波数データに変換している。96 kHz でサンプリングされた音響信号に対し、8192 サンプル (約 85 ms) ごとに FFT を行い、0–48 Hz の 4097 点のデータを得る。これらのデータから、20–40 kHz 部分のデータのみを切り出し、1707 点のデータを得る。次に、ダウンサンプリングを行い、854 点のデータを得る。その後、周波数方向に 20 点の移動平均フィルタを用い、さらに 5 点の移動平均フィルタを用い

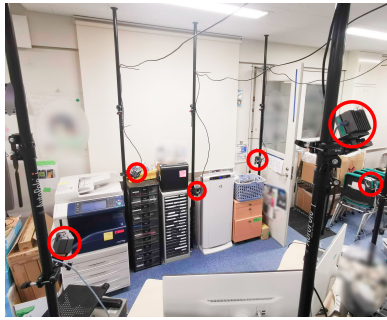


図 2 実験に用いた OptiTrack のカメラ。

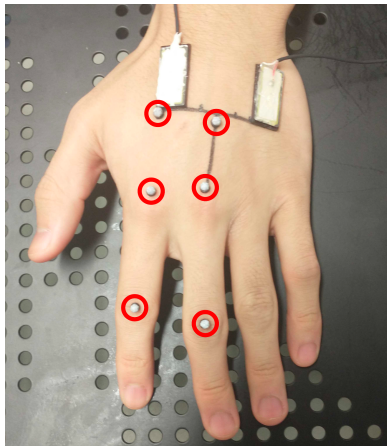


図 3 ピエゾ素子と OptiTrack マーカ（赤丸）を手についた様子。

て平滑化を行う。さらに、データ全体に対して最小値を 0, 最大値を 1 とする正規化を行う。これらの処理によって得られた 854 点のデータの特徴量とし、機械学習を用いて関節角度推定を行う。

3.3 データ同期

アクティブ音響センシングによる手の共振特性を計測すると同時に、手指の関節角度の正解データを得るために、OptiTrack による計測環境を構築した (図 2)。関節角度の正解データおよび周波数データからなる学習用データの作成方法を述べる。OptiTrack から得られた関節角度データおよび 854 点のデータからなる周波数データをそれぞれタイムスタンプとともに CSV ファイルに書き出す。各周波数データについて、周波数データの時刻に最も時間が近い関節角度データを結合し、機械学習に用いるデータとした。

4. 実験

提案手法の精度評価のために参加者 1 人に対し実験を行った。

4.1 実験参加者

実験参加者は著者 1 名 (男性, 22 歳, 右利き) であった。

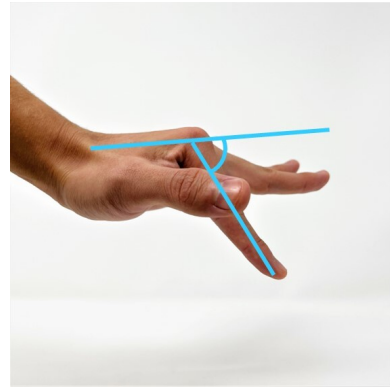


図 4 今回計測した手指の角度。第 3 関節の角度を計測している。

4.2 実験機器・環境

ピエゾ素子を貼付する場所は久保らの研究 [6] に基づき決定した。具体的には、中指の第 3 関節から 3.5 cm, ピエゾ素子の間隔は 2.8 cm とした。マーカはピエゾ素子に重ならないように人差し指, 中指の第 2 関節, 第 3 関節, および中手骨に沿って第 3 関節から 3 cm の位置に貼付した (図 3)。

手指の第 3 関節角度の実測値を収集するカメラとして OptiTrack PrimeX 13 を 6 台用い (図 2), トラッキングに必要なマーカは直径 3 mm の半球タイプの反射材を用いた。ソフトウェアは Motive 3.0.1 を用い, 120 Hz のサンプリングレートにおいてマーカの 3 次元座標データを取得した。得られた 3 次元座標データから, ベクトルの内積の計算結果を用いて関節角度を求めた。

4.3 実験手順

最初に, 皮膚貼付用の両面テープ (innovect SH 2477P-S) を用いて参加者の左手の甲に 3.1 節において述べたハードウェアのピエゾ素子を貼付した。実験の間, 参加者は椅子に座り台の上に腕を置き, 手指以外の腕および手首が動かないようにした。全ての指を伸ばした状態を 0° とし, 屈曲させる指以外の指を伸ばした状態を保ちながら, データ収集を行う指を 5° 刻みの角度において手のひら側に屈曲させ約 5 秒間固定するよう指示し, 65° までデータ収集を行った (図 4)。実験参加者には可能な限り屈曲させる指以外の指を曲げないように指示した。 $0-65^\circ$ までの 14 の角度に対して, 上記の動作を人差し指もしくは中指において 1 回ずつ行うことを 1 セッションとし, 各指について計 4 セッション行った。セッション間には 5 分以上の休憩を設けた。データ収集は 3.2 節のソフトウェアを用いた。本実験において収集した 7 秒間のデータの内, 最初 6.83 秒のデータ (80 データ) を学習に用いる。そのため合計 8960 (1 人 $\times$ 2 状態 $\times$ 4 セッション $\times$ 14 回 $\times$ 80 データ) のデータを, 各指の関節角度推定精度の評価に用いた。

機械学習部分のアルゴリズムには scikit-learn の LassoCV を用いた。正則化の強さを決める α は 0.1, 反復回数

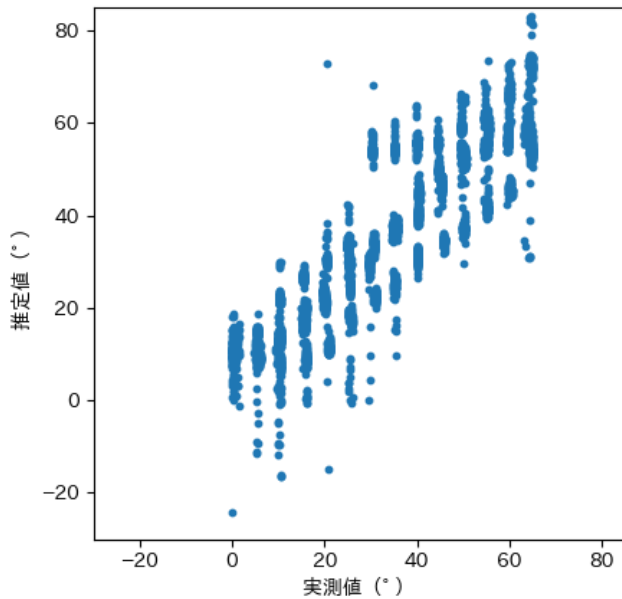


図 5 人差し指の角度の実測値（横軸）と推定値（縦軸）.

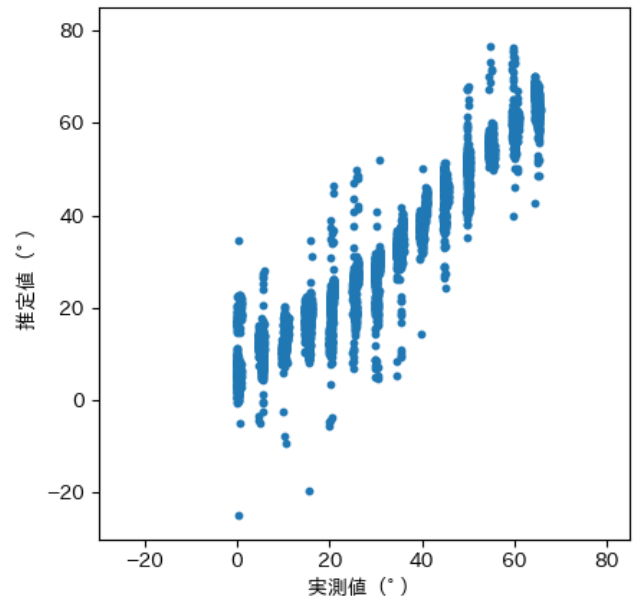


図 6 中指の角度の実測値（横軸）と推定値（縦軸）.

である max_iter は 100,000 とした.

4.4 実験結果および考察

得られたデータのうち、1セッションをテストデータ、残りの3セッションを訓練データとする1つ抜き交差検証を行い、各指について関節角度推定精度を求めた。まず、各指ごとに学習データを用いて、LassoCVにより関節角度推定の回帰モデルを作成した。このモデルに各テストデータを入力し得られた推定値と実測値との誤差を求めた。また、作成した回帰モデルの決定係数も求めた。その結果、人差し指において、平均誤差 (RMSE) は $\pm 8.53^\circ$ 、平均二乗誤差 (MSE) は 82.265、決定係数は 0.79 であった。中指において、RMSE は $\pm 5.51^\circ$ 、MSE は 32.329、決定係数は 0.92 であった。結果より、人差し指および中指ともに推定値のばらつきが見られる。これは各セッション間において統一できていない実験条件があること、および音響信号のノイズの影響があると考えられる。

5. 今後の展望

本節において、今後の展望を述べる。

5.1 実験参加者数の追加

人差し指および中指の第3関節の角度推定を行った結果、それぞれ $\pm 8.53^\circ$ および $\pm 5.51^\circ$ の平均誤差の精度において手指関節の推定が可能であることが示された一方で、本研究における実験の参加者数は1名であった。提案手法のさらなる性能調査として、より多くの実験参加者からデータ収集を行う必要がある。これにより、提案手法がどの程度の精度において手指関節の推定ができるかをより明確にすると同時に、ユーザに対する汎化性能の調査も行う。

5.2 複数指の関節角度の同時推定

本研究において、我々は人差し指および中指のみの第3関節を屈曲させ、関節角度の推定を行った。しかし、実際にハンドジェスチャを行う際には同時に複数指の第3関節の屈曲が行われることが想定されるため、複数指の第3関節を屈曲させた際の関節角度のデータ収集を行う必要がある。これにより、提案手法がどの程度の精度において同時に複数指の第3関節角度の推定ができるかの調査を行う。

6. おわりに

本研究は、手の位置に制限されず、かつ周囲の環境にセンサを設置することなく手指の関節角度を推定することを目的としている。このために、アクティブ音響センシングを用いて人差し指および中指の第3関節をそれぞれ屈曲させた際の音響データを取得し、機械学習を用いて第3関節の角度を推定した。今後、複数指の第3関節を同時に屈曲させた際の音響データを取得し、機械学習を用いて関節角度の推定を行う予定である。

参考文献

- [1] Acuity: OptiTrack モーションキャプチャ, <https://www.optitrack.jp> (最終閲覧日: 2022 年 12 月 13 日) .
- [2] HTC: HTC VIVE Pro 2, <https://www.vive.com/jp/product/vive-pro2-full-kit/overview/> (最終閲覧日: 2022 年 12 月 13 日) .
- [3] Glauser, O., Wu, S., Panozzo, D., Hilliges, O. and Sorkine-Hornung, O.: Interactive Hand Pose Estimation Using a Stretch-Sensing Soft Glove, *ACM Transactions on Graphics*, Vol. 38, No. 4 (online), DOI: 10.1145/3306346.3322957 (2019).
- [4] Wu, E., Yuan, Y., Yeo, H.-S., Quigley, A., Koike, H. and Kitani, K. M.: Back-Hand-Pose: 3D Hand Pose Estimation for a Wrist-Worn Camera via Dorsum Defor-

- mation Network, *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20, New York, NY, USA, Association for Computing Machinery, pp. 1147–1160 (online), DOI: 10.1145/3379337.3415897 (2020).
- [5] Ono, M., Shizuki, B. and Tanaka, J.: Touch & Activate: Adding Interactivity to Existing Objects Using Active Acoustic Sensing, *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, New York, NY, USA, Association for Computing Machinery, pp. 31–40 (online), DOI: 10.1145/2501988.2501989 (2013).
- [6] Kubo, Y., Koguchi, Y., Shizuki, B., Takahashi, S. and Hilliges, O.: AudioTouch: Minimally Invasive Sensing of Micro-Gestures via Active Bio-Acoustic Sensing, *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '19, New York, NY, USA, Association for Computing Machinery, pp. 1–13 (online), DOI: 10.1145/3338286.3340147 (2019).
- [7] Kato, H. and Takemura, K.: Hand Pose Estimation Based on Active Bone-Conducted Sound Sensing, *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, New York, NY, USA, Association for Computing Machinery, pp. 109–112 (online), DOI: 10.1145/2968219.2971403 (2016).
- [8] Yokota, T. and Hashida, T.: Hand Gesture and On-Body Touch Recognition by Active Acoustic Sensing throughout the Human Body, *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16 Adjunct, New York, NY, USA, Association for Computing Machinery, pp. 113–115 (online), DOI: 10.1145/2984751.2985721 (2016).
- [9] Okawa, Y. and Takemura, K.: Haptic-Enabled Active Bone-Conducted Sound Sensing, *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15 Adjunct, New York, NY, USA, Association for Computing Machinery, pp. 87–88 (online), DOI: 10.1145/2815585.2815732 (2015).