物を叩いたり振ったりした際の音と手の動きを用いた物体識別及び内容量認識

小口 雄斗[1,a)]　志築 文太郎[2,b)]　高橋 伸[2,c)]

概要：コンピュータを生活に使用する物の管理に活用することは一般的なものとなった．一方で既存のサービスや認識手法では使用できる物やセットアップ環境に制限があり, 日常生活における様々な物やその内容量を認識して生活に活用することが難しい．そこで本研究では物に触れる際の手の動きとその周囲で発せられる音に着目し, 手で物を叩いたり振ったりした際の音響と慣性データ（加速度, 角速度）から物体やその内容量を機械学習により認識する手法を提案する．本稿では提案手法の検証を行うために M5StickC をセンシングデバイスとして用いた検証用システムを構築した．検証用システムのソフトウェア実装にあたり, 予備実験にて分類器の性能テストを行った結果 13 種類の分類について 94%の正解率を示した．

1. はじめに

IoT 機器の普及に伴いコンピュータを生活に使用する物の管理に活用することは一般的なものとなった．例えば Google Home や Amazon Alexa に代表されるスマートスピーカでは市販のスマートホーム機器と組み合わせることでエアコンやテレビなどを声で操作することが簡単に実現できる．また Amazon Dash Replenishment[*1] では使用状況を計測するセンサを内蔵した製品を使用することでその消費量や消耗状態を自動で認識し注文を行うサービスを提供している．しかしこれらは電力供給を必要とする電化製品や専用ハードウェアなどの制御を中心としており, 日用品や食品, 家具などの様々な物に対して適用することが難しい．

日常生活で使用される様々な物についてコンピュータとの連携やコンピュータを介した管理を行うために簡易に物体を識別する手法や内容量などの物体の状態を認識する手法が望まれる．既存の手法としてはカメラを用いた画像認識 [1] や対象物にセンサを取り付ける方法 [2] があるが, 前者は照明条件による認識精度への影響や撮影されるユーザのプライバシへの配慮が懸念され, また後者は機器の設置や配線, 取り付けの面で生活環境内で使用するには利便性に欠ける．

そこで本研究では物に触れたり物を動かしたりする際の手の動きとその周囲で発せられる音に着目する．音や動きをセンシングする装置を手に取り付けるだけであれば既存手法における制限や設置コストの問題がなく, ユーザはより簡単で直感的に物とコンピュータのやりとりを行い, 生活に役立てることができる．例えばインターネットに掲載された料理のレシピ情報に従って調理をする際に, 使う調味料のボトルを叩き, 注ぐ前後に振ることで「何の調味料」を「どのくらい注いだ」のかを自動で認識し, レシピと照らし合わせた上でスマートスピーカやスマートフォンを介してユーザにフィードバックすることにより調理の補助に活用できる．またティッシュ箱からティッシュを引き抜く音やその引き抜きやすさからどの程度残量が残っているのかを認識し, ユーザに購入時期を提案することで在庫管理に役立てることができる．さらに読書を行う際に「本を叩く」という動作で部屋の照明を明るくし, 叩き方に応じてその明るさを変更するといったスマートホーム機器の制御や操作への応用も考えられる．

本稿では物を叩いたり振ったりした際の音や手の動きによる, 物体認識や内容量認識を行う手法を提案するとともに, 提案手法の検証を行うためのシステム構築とその実装に関わる予備実験について述べる．

2. 関連研究

2.1 物体認識に関する研究

Knocker [3] はスマートフォンアプリケーションのトリガとしての利用を想定し, 本研究同様にスマートフォンで

[1] 筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻
[2] 筑波大学システム情報系
[a)] koguchi@iplab.cs.tsukuba.ac.jp
[b)] shizuki@cs.tsukuba.ac.jp
[c)] shin@cs.tsukuba.ac.jp
[*1] https://developer.amazon.com/ja/dash-replenishment-service

物を叩いた際の音とその加速度及び角速度を用いて叩いた物体の認識精度を評価した研究である．この研究では対象物を置く場所や騒音条件を複数パターン用意して評価しており, いずれのパターンにおいても音響のみを用いた場合と比較して, 加速度と角速度を特徴量に加えた際に識別精度が向上することが示されている．

ViBand [4] ではスマートウォッチに内蔵された加速度センサにて生体音響信号をセンシングし, 物に触れた際の信号の変化を用いて物体認識を行っている．この研究ではモータ駆動などの特有の振動を発する物を対象としており, 静的な物に関しては振動を発生する装置の取り付けを必要とする．

本研究の手法は物に触れる際の音や手の動きを用いるため, 認識対象の機能や特徴により制限されることがない．そのため日用品や生活用品などの様々な物への適用と日常生活への活用を想定している．

2.2 内容量認識に関する研究

日用品や生活用品の内容量を自動で計測する研究は過去にもいくつか先行例がある．SoQr [2] では一組のスピーカとマイクにより構成されたセンサデバイスを生活用品に貼り付け, スピーカから発した信号の周波数応答を用いて内容量の識別を高い精度で達成している．また VibeBin [5] ではゴミ箱に取り付けたモータによりゴミ箱を振動させ, 加速度センサで捉えた信号を解析することでゴミ箱内の内容量を計測している．

これらの研究では計測する対象物に個別に特殊な装置を直接取り付ける必要があり, 買い換えや使い捨てられることが多い日用品への適用が難しい．本研究の手法はユーザの手と物とのインタラクションにより生じる情報を使って物の認識やその内容量の計測を行うため, 様々な日用品や生活用品への適用が容易であり, 設置コストや運用コストの点で優れる．

3. 提案手法とシステム概要

本研究では物に対して手で叩くまたは振るといった動作により生じる音響と, その際の手の運動の慣性データ（3 軸加速度, 3 軸角速度）を収集し, これらの周波数特性を機械学習で分類することで物体の識別や内容量の認識を行う手法を検討する．物はその材質により叩いた際の音や感触が異なる．例えば木やガラスなどの固い素材で作られた物を叩いた際の音は高く, 叩いた衝撃が強く手に伝わる．一方で紙などの柔らかい素材で作られた物では叩いた際の衝撃が吸収され低い音が響く．また物を振る際においてはその中身が空に近い状態では容易に振ることが可能でありその際に発せられる音は高い．反対に満杯に近い状態では振るのに労力が必要であり発せられる音も低い．本手法はこれらの特性をコンピュータで分析することにより物体識別や内容量認識を試みる．

3.1 システム構成

本手法を検証するために構築したシステムの構成を示す．本システムはセンシングデバイスと分析用のコンピュータの 2 点で構成される（図 1）．これらは無線 LAN により同一ネットワーク上に接続されている．

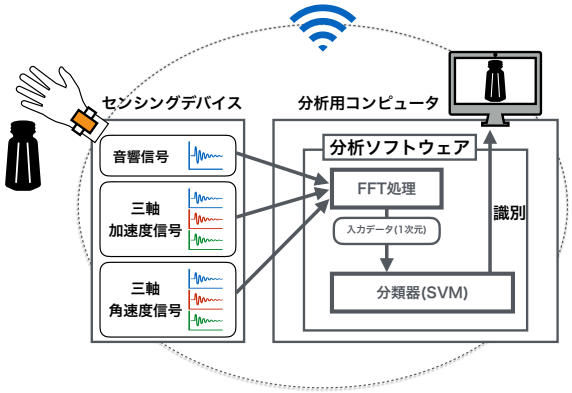

図 1 システム構成

センシングデバイスには図 2 に示す M5StickC[*2] を使用する．これは 6 軸 IMU センサやマイクロフォン, 無線通信モジュール及びマイクロコントローラと駆動用のバッテリを搭載したデバイスであり, 専用のマウンタを用いることで手首に装着することが可能である．センシングデバイスは内蔵センサにより音響及び慣性データをセンシングし, 逐次各センサデータを UDP プロトコルにてネットワーク上にストリーミング送信する．この際の音響のサンプリングレートは 22050Hz であり, 慣性データのサンプリングレートは 400Hz である．分析用コンピュータは受信したセンサデータを分析ソフトウェアに入力し認識された物体や内容量の提示を行う．尚, 実際の利用イメージとしてはセンシングデバイスと分析用コンピュータにはスマートウォッチなどのウェアラブルデバイスを想定しており, 認識結果はデバイス本体やスマートフォン, スマートスピーカなどからユーザにフィードバックされる．

3.2 分析ソフトウェア実装

センサデータの解析及び機械学習による分類を行うソフトウェア実装について示す．分析ソフトウェアでは入力されたセンサデータを時系列順に整列し, 250ms の時間フレーム単位で各センサデータに高速フーリエ変換（FFT）による周波数解析を行う．この時間フレーム長は後述する予備実験による調査で決定した．その後周波数解析により得られた各センサデータの周波数スペクトルを一次元の

[*2] https://m5stack.com/products/stick-c

図 2 センシングに用いるデバイス（M5StickC）

データとして結合し，これを分類器への入力データとする．分類器の実装には教師あり機械学習手法であるサポートベクタマシンを用いる．そのため本システムは事前に学習用データセットの構築及び分類器の学習が必要である．

今回実装した処理は叩いた際のセンサデータのみを対象としており，予備実験にてその信号の特徴により時間フレーム長を決定している．そのため今後振る際のデータも用いる際には我々の過去の研究 [6] の手法を参考に改めて最適な時間フレーム長の設定を行いシステムに組み込む．

4. 予備実験 1

分類器の学習用データセット及び入力データの生成に用いるセンサデータの時間フレーム長を決定するために，物を叩いた際のセンサデータを収集しそれらの時系列変化の特徴を調査した．被験者は著者 1 名であり，叩く対象は机（図 3A），本（図 3B），スマートフォン（図 3C），プラスチックボトル（図 3D）の 4 種類であり，机を除く 3 種類については机の上に置いた状態とした．実験は 2 秒間の間に一度叩く動作の収集を各 5 回行い，得られたデータをグラフに描画し目視で比較することで行う．叩く動作は手首にセンシングデバイスを装着した手でノックするような動作とした（図 4）．

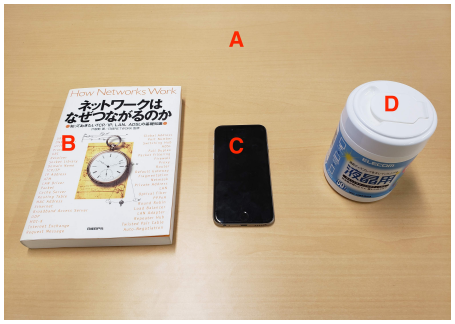

図 3 4 種類の叩く対象物 (A: 机, B: 本, C: スマートフォン, D: プラスチックボトル)

収集した全データを比較した結果を示す．実験結果から抜粋した図 5 は机を叩いた際のセンサデータの時系列変化のグラフである．図 5 同様にいずれの対象物においても

図 4 ノックするように叩く動作

音響のピークを基準としてその前方 100ms，後方 150ms の 250ms 区間に各センサデータの時系列変化の特徴が十分に収まることが確認された．この結果を元に分析ソフトウェアにおける時間フレーム長を 250ms と定めた．

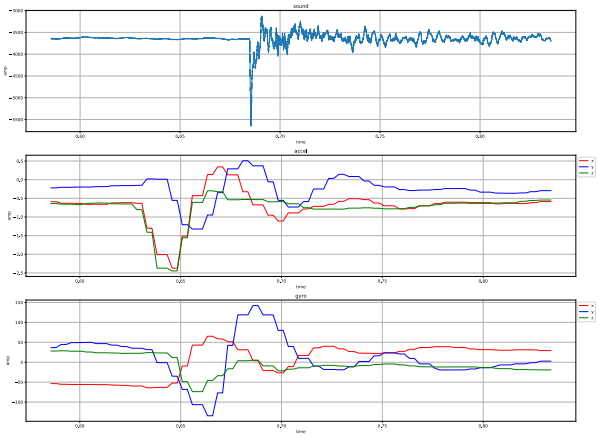

図 5 机を叩いた際の音響及び慣性データの時系列変化の例

5. 予備実験 2

本システムで実装した分類器の動作検証を行うために物を叩いた際の音響と慣性データを収集し，構築したデータセットを用いて物体分類の検証を行った．叩く対象は机（desk, 図 6A），本（book, 図 6B），スマートフォン（phone, 図 6C），タブレット（tablet, 図 6D），マウス（mouse, 図 6E），プラスチックボトル（bottle-pp, 図 6F），アルミ缶（bottle-can, 図 6G），ステンレスポット（pot-sten, 図 6H），電気ケトル（pot-pp, 図 6I），プラスチックスプレー（spray-pp, 図 6J），ティッシュ箱（box-tissue, 図 6K），ウェットティッシュ箱（box-pp, 図 6L）の 12 種類であり，机を除く 11 種類の物は机の上に置いた状態とした．分類対象はここに何も叩いていない状態（null）を加えた 13 種類である．

被験者は著者 1 名である．叩く動作は予備実験 1 と同様にノックするような動作とし，2 秒間に一度対象物を叩く動作の収集を各 30 回行った．その後得られた各対象物のデータに対して予備実験 1 で定義した基準に従って 250ms

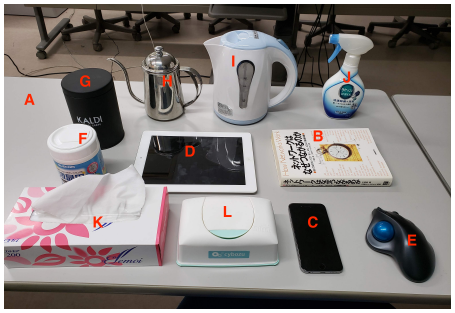

図 6　12 種類の叩く対象物

区間のデータのトリミング及び分析プログラムと同様の FFT 処理を行い, 分類器の学習用データセットとした.

構築したデータセットのうち各対象物についてそれぞれ 24 回収集分のデータを訓練データとして分類器の学習を行った. この際分類器には線形カーネルを使用した. そして残り各 6 回収集分のデータをテストデータとして学習した分類器に入力した.

分類結果を図 7 に示す. いずれの対象物においても F 値は 80%以上であり, 全体的な正解率 94%であった. したがって本システムで実装した分類器は今回用意した 1 人分のデータセットについては学習と分類テストの動作が確認された.

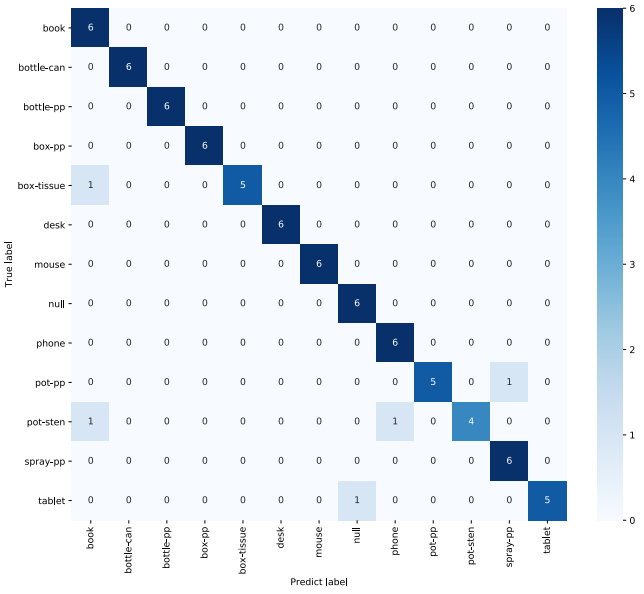

図 7　13 種類の分類結果の混同行列

6. まとめと今後の予定

本稿では簡易に物体や内容量の認識を行う方法として, 物を叩いたり振ったりした際の音響と手の慣性データ（3 軸加速度, 3 軸角速度）を用いて機械学習により分類, 識別する手法を提案した. そして提案手法を評価するためにセンシングデバイスと分析ソフトウェアを実装したシステムを構築し, 予備実験にて物を叩いた際のセンサデータの時間フレーム長を調査し, 用意したデータセットによる分類器の動作確認を行った.

今後はソフトウェアの調整により振る際のデータに対しても本システムを適用し, また実際に物体や内容量の認識精度の評価実験や複数被験者間における分類器の汎用性の確認を行う.

参考文献

[1] Chen, K., Fürst, J., Kolb, J., Kim, H.-S., Jin, X., Culler, D. E. and Katz, R. H.: SnapLink: Fast and Accurate Vision-Based Appliance Control in Large Commercial Buildings, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 1, No. 4, pp. 129:1–129:27 (2018).

[2] Fan, M. and Truong, K. N.: SoQr: Sonically Quantifying the Content Level Inside Containers, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, New York, NY, USA, ACM, pp. 3–14 (2015).

[3] Gong, T., Cho, H., Lee, B. and Lee, S.-J.: Knocker: Vibroacoustic-based Object Recognition with Smartphones, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 3, No. 3, pp. 82:1–82:21 (2019).

[4] Laput, G., Xiao, R. and Harrison, C.: ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers, *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, New York, NY, USA, ACM, pp. 321–333 (2016).

[5] Zhao, Y., Yao, S., Li, S., Hu, S., Shao, H. and Abdelzaher, T. F.: VibeBin: A Vibration-Based Waste Bin Level Detection System, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 1, No. 3, pp. 122:1–122:22 (2017).

[6] 小口雄斗，志築文太郎，高橋　伸：容器を振る際の音を用いた容量識別手法，情報処理学会第 81 回全国大会講演論文集，Vol. 2019, No. 1, pp. 363–364 (2019).