筑波大学大学院博士課程

システム情報工学研究科修士論文

# 能動的インタラクションから生じる音響及び慣性データを用いた物体識別及び内容量認識

小口 雄斗

修士（工学）

（コンピュータサイエンス専攻）

指導教員　高橋 伸

2020年3月

## 概要

Internet of Things（IoT）と呼ばれる仕組みの登場により, 誰もが簡単にコンピュータを使用して生活で使う物の操作や管理ができるようになった. しかし日常生活で使用する様々な物に対して既存の IoT 製品やサービスでは適用することが難しい. そのためコンピュータと物のやりとりを拡張し, 生活を便利にするためには生活空間内の様々な物の情報を簡単な方法で取得できる手法が必要となる. そこで本研究では「物の識別」と「物の内容量の認識」の 2 点について, 人が物に触れる際の音や手の動きによりこれらを実現する手法を提案する. また提案手法を検証するために, 手周囲の音響信号や加速度信号, 角速度信号のセンシングデバイスと, センシングしたデータを用いて機械学習により特徴の学習と分類を行うプロトタイプシステムを構築した. 評価実験ではプロトタイプシステムを用いて物を叩く際のデータによる 13 種類の物体分類と, 容器を振る際のデータによる 5 種類の内容量分類の 2 つの実験を行い本手法を評価した. その結果, 物体分類について Leave-One-Group-Out-Cross-Validation による検証で 72% の正解率が示された. しかし今回の実験からは加速度信号と角速度信号の特徴量を用いることによる分類結果への寄与は示されなかった. また内容量分類についても音響信号のみを使用した検証で示された 58% の正解率が最も高い結果であり, 本手法にて加速度信号と角速度信号を用いることについて懸念点が残った. 一方で実験結果を踏まえて追加で行った調査では, 加速度信号と角速度信号の特徴量から置いてある場所やその使う状況などの情報の推定に有望であることが示された. したがって本手法は今後実装の改善により, 様々な生活用品や日用品, またその使用量などの識別に加えて, 使用している状況の判別などへの活用が期待できる.

# 目 次

# 図目次

# 表 目 次

# 第1章　序論

本章において, はじめに研究を行う上での背景について述べる. 次にそれを踏まえた本研究の目的とその実現のためのアプローチを示す. その後, 本研究の貢献と本論文の全体構成について述べる.

## 1.1　研究背景

インターネットの普及に伴い, 世の中の様々な物がインターネットを介して相互にやりとりする Internet of Things（IoT）と呼ばれる仕組みが登場した. そして近年ではスマートフォンやスマートウォッチ, スマートスピーカなどのスマートデバイスの普及とその性能向上により, コンピュータを日常生活に活用することが容易になった. 例えば Google Home や Amazon Alexa に代表されるスマートスピーカと, 連係可能なスマートホーム機器を用いることでエアコンやテレビなどの家電製品を声で操作することが可能となる. また Amazon Dash Replenishment[1] というサービスでは, 使用量や消耗状態を計測するセンサを内蔵したコーヒーメーカや電動歯ブラシなどを提供しており, ユーザはこれを用いることで日用品管理を自動化することができる.

このように市販の IoT 機器やスマートデバイスを用いることで誰もが簡単にコンピュータを生活環境に取り入れることができるようになった. 一方で例に示したようにその多くは電気的な配線が必要なリモコン装置や専用ハードウェアの導入が必要であり, また用途も電化製品への機能追加などに限られているため, 日常生活で使用する様々な物に対して適用することを想定したものではない. 例えばティッシュ箱などの家庭内で設置する個数や使用頻度, 取替頻度の多い日用品や, 飲み物や調味料などの食品, 衣料品や家具などに対しては, 既存の製品やサービスではユーザの生活の拡張に活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, 物の識別やその状態などの情報を簡易な手法によりコンピュータで認識できることが望まれる. 既存の物体情報の認識手法としてはカメラを用いた画像認識や対象物にセンサを取り付ける方法などがある. しかし画像認識を用いた手法では照明条件による認識精度への影響が課題として指摘されており [1], さらに生活空間内で使用する場合には撮影されるユーザのプライバシへの配慮が懸念される [2]. センサを取り付ける手法では生活用品に個別にセンサを取り付けることは実際の生活環境内において利便性に欠ける. また既存の製品や

[1] https://developer.amazon.com/ja/dash-replenishment-service

サービス同様にどちらの手法においても機器の設置や配線が必要でありセットアップコストが大きいことが指摘されている [3].

これらのことから生活で使用する様々な物がコンピュータとやりとりし, 物の持つ情報を用いてユーザの生活を便利にするためには, 既存の手法のような問題がなく, より簡単な方法で生活空間内の物の情報を取得する手法が必要である.

## 1.2 目的とアプローチ

物体の情報として本研究では「物の識別」と「物の内容量の認識」の 2 点に焦点を当てる. したがって本研究の目的は生活で使用する様々な物について, コンピュータを使用してこれら 2 点を実際の生活空間内に適用可能な手法で実現することである. この目的を達成するためのアプローチとして手で物に触れたり物を動かしたりする際に発せられる音と, その際の手の動きに着目する. 手にその動きや周囲の音をセンシングする装置を取り付けるだけであれば既存手法におけるセットアップコストの問題がなく, また適用できる対象物の範囲が広い. したがってユーザは物を叩いたり振ったりするような簡単なジェスチャや操作によって, より直感的に様々な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.

## 1.3 本論文の構成

本章以降の本論文の構成について示す. 第 2 章では本研究に関連する先行研究について示し, 本研究の差分を述べる. 第 3 章では手と物体の間で生じる音や運動情報を用いた物体認識や内容量識別を行う手法を提案し, その利用例について示す. 第 4 章では提案手法を元にその評価を行うに構築したプロトタイプシステムの実装について述べる. 第 5 章ではプロトタイプシステムを用いて物を叩くことによる物体分類と振ることによるや内容量分類の実験を行い, その精度について検証する. 第 6 章では実験結果を踏まえて本手法で使用したセンサデータの特徴量に関する議論と追加の調査を行い, 本手法の展望を示す. 第 7 章では本研究での取り組みについて結論を述べる.

# 第2章　関連研究

本研究で取り組む物体認識や内容量識別に関する先行研究を示すとともに, 本研究で提案する手法との差分について述べる. また本研究のプロトタイプシステム同様に手首に装着したデバイスを用いる研究の様々な先行事例を例示し, 本手法のアプローチの有望性を示す.

## 2.1　物体認識に関する研究

物体認識に関する先行研究について述べる. またそれらの特徴や利点を述べるとともに問題点について言及し, 本手法の利点を示す. Laput らの 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. この研究では対象物として機械や楽器などの使用時に特有の振動を発する物やモータを内蔵している物などを対象としており, 振動を発しない静的な物に関しては振動を発生させる装置の取り付けを必要とする. Xiao ら [5] や Maekawa ら [6] の研究では電磁気センサを取り付けたスマートフォンやグローブを使用することで触れた物体や把持した物体の認識を行っている. これらの研究では認識手法として物体の発する電磁気に基づくため電化製品への適用に限られている. また Bi ら [7] の研究では耳に取り付けたマイクと筋電センサにより咀嚼時の音と顎の筋電位の変化を用いて食べた物の分類を行っている. この研究においては対象物が食品のみに限定される. 以上の研究は対象物がその物の機能や特徴に依存している点で適用範囲が狭い. それに対して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, 認識対象の制限が少なく様々な物に適用できるという利点がある.

また本研究の手法同様に広範囲な物体を対象としたアプローチも先行事例で多く示されている. 代表的な手法としてはカメラによる映像や画像認識を用いた手法 [8, 9] が挙げられる. 例えば Kacorri ら [10] の研究では対象物を様々な角度から撮影し, 深層学習を用いてその画像を分類することで高精度な物体認識を実現している. また RFID タグや, QR コードのような二次元コードを用いる手法 [11, 12] がある. RFID タグや QR コードは製造コストが小さく製品の製造過程で付与することが容易であるため, 既に産業分野で実用化 [1][2] されている. 一方でこれらの手法では対象物の読み取りにスマートフォンのカメラや専用のリーダ機器を持つ必要があり, ユーザが生活空間内で使用するには利便性に欠ける. また屋内でカメラを用いる場合には撮影される人や風景についてプライバシの点で問題となりうる. 本研究の手法ではユーザの手周囲の限られた範囲のみをセンシングするため, それにより得られる情報は比較的具体性

[1] https://www.denso-wave.com/ja/adcd/fundamental/rfid/rfid/index.html
[2] https://www.denso-wave.com/ja/technology/vol1.html

が低い. またユーザの行動を妨げることがないため実生活への受容性が高いと考えられる.

本研究と同様のアプローチを用いた研究として Gong らの Knocker [13] がある. この研究はスマートフォンで叩いた物を認識する研究であり, 叩いた際の音とその加速度及び角速度を用いて物体の認識精度を評価している. また対象物を置く場所や騒音条件を複数パターン用意して検証しており, 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. 本研究は Knocker を参考としているが, Knocker がスマートフォンアプリケーションのトリガとして物体認識の利用を想定しているのに対し, 本研究ではより柔軟に生活空間に活用することを目標として, ユーザの手に直接適用する手法を構想する.

## 2.2 内容量認識に関する研究

本研究と同様に物の重量や内容量を計測する研究は過去にも取り組まれた事例がある. 例えば Fan ら [14] は周波数応答による生活用品の内容量計測を行う SoQr を構築した. SoQr 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直接生活用品に貼り付け, ヘッドホン端子を介してスマートフォンに接続し, スピーカから出力したチャープ信号に対するマイクから入力された周波数応答をスマートフォンのアプリケーション上で解析することにより, 19 種類の生活用品について高い精度での内容量推定を達成している. また VibeBin [15] 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, 加速度センサでその信号応答を取得し, これをコンピュータで解析することによりゴミ箱内の内容量を計測している. Wei ら [16] の研究では静電容量センサが内側に取り付けた専用のセンシング容器を使用して容器内部の液体の量を識別している.

これらの手法ではセンシング装置を直接取り付けたり専用のハードウェアを使用したりすることで高い精度での内容量の識別を実現している. 一方で実際の生活空間内に適用することを考慮した場合, 生活で使用する日用品や生活用品は使い捨てや買い換えの機会が多く, 先行手法の多くは機器の着脱や取り付けの面で手間である. また機器が物理的に生活用品に干渉することはユーザの使い心地の面でも利便性に欠ける. 本研究で提案する手法はユーザが物に触れる際の音と手の動きを用いるため取り付けコストを必要としない. また物体認識と組み合わせることで様々な物に対して同時適用が可能となると考えられる.

また著者らも過去に容器を振る際の音からの内容量の識別について取り組んでいる [17]. この研究では本研究のアプローチとは異なり, 空間内に設置したマイクにより 1 秒間容器を振る際の音を収集し, そのスペクトログラムを畳み込みニューラルネットワークにより学習及び分類することによる内容量の識別を試みた. 評価実験の結果では約 88% 程度の分類精度が示されていたが, 一方で 1 秒間容器を振理続けることは活用シナリオとして想定するのが難しい動作であることや, 深層学習のためのデータの収集が困難であることなどが指摘されていた. そのため本研究ではこのような問題を課題と捉え, 分類に使用する特徴量やその収集時間及び分析方法を見直しその評価を行う.

## 2.3 手首に取り付けるデバイスを用いた研究

本研究では手首に取り付けたデバイスにより手で物に触れる際の音や動きをセンシングし, 物体やその内容量の識別を試みる. 同様に手首に取り付けた装置を用いた研究はこれまでにも様々な形式で取り組まれてきた. 例えば手首に取り付けた装置により身体を入力インタフェースにする研究がある. Aoyama らの ThumbSlide [18] では時計のベルト内側に取り付けたフォトリフレクタを用いて手首の形状変化を認識することで, 親指位置に応じたスマートウォッチの片手操作を行っている. LumiWatch [19] ではプロジェクタと深度センサを用いて, 腕に画像情報の表示やそのコンテンツへのタッチ入力を実現する腕時計型のプロトタイプデバイスを開発している. Suzuki ら [20] は手首と前腕に取り付けた電極を用いて, 電気インピーダンス法により腕をタッチ入力インタフェースとしての利用する手法を提案している.

またスマートウォッチを他の機器への入力装置として用いる研究がある. Becker らの GestEar [21] ではスマートウォッチの内蔵センサで取得した音と加速度信号, 角速度信号から深層学習により簡易なジェスチャを認識し, スマートフォンやスマートホーム機器の操作に用いている. 同様に Shi ら [22] の研究ではスマートウォッチのマイクを用いて手で物を叩いた際の音から物体の認識を行い, 電子機器やコンピュータのショートカット操作への活用手法を提案している. これらの研究ではアプローチや用途が本研究の手法と類似しているが, 本研究では物体認識と内容量識別の 2 点について取り組み, 機器操作に限らず日常生活を支援するための様々な用途への活用を想定している.

このように手首に装置を取り付ける先行研究は様々存在している. また腕時計のように手首に機器を装着することは既に社会的に定着している. これらのことから本研究の手法のようにセンシングデバイスを手首に取り付けることは受容性や使い心地の面で有望であると考えられる.

# 第3章　能動的インタラクションによる物体情報の認識手法

本章では提案手法の概要とその動作原理, また提案手法を実際の生活に取り入れた際の利用イメージについて述べる.

## 3.1　概要

本研究における提案手法について述べる. 本研究では物を叩いたり振ったりした際の音や手の動きにより物体の識別や内容量の認識を行う手法を提案する（図 3.1）. 本手法ではユーザの手首にセンシングデバイスを装着することで, 手の周囲の音響信号と手の動きの慣性データ（加速度信号, 角速度信号）を収集する. そして収集したセンサデータからその特徴を機械学習で学習・分類することで物体識別や内容量認識を行う.

本手法はユーザの手と対象物のインタラクション時にのみ作用する. そのため対象物ごとに個別に装置を取り付ける必要がなく, ユーザは普段通りに物を使ったり, あるいは物に対して任意のジェスチャや操作を行ったりすることで本手法を利用可能となる.

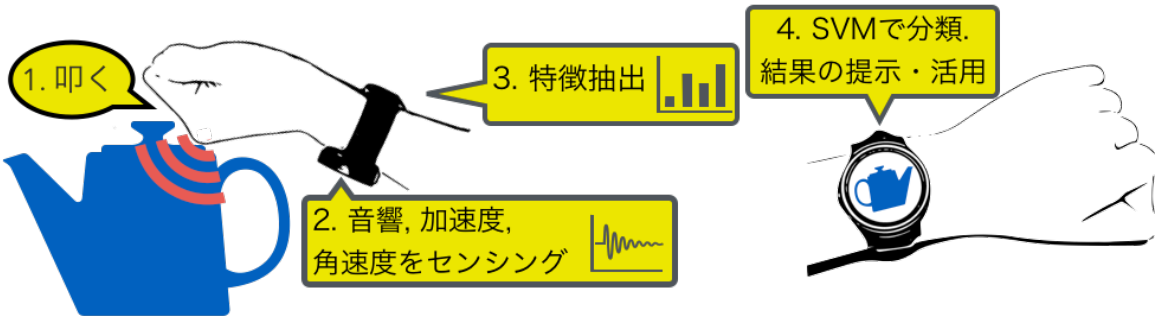



図 3.1: 提案手法の概要図

## 3.2 認識原理

物体識別や内容量認識を行う上での本手法の動作原理について述べる. 前提として物はその材質や形状により衝撃を加えた際の音が異なることが知られている [23]. 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, 一方で紙や布などの柔らかい素材で作られた物では叩いた際は衝撃が吸収されて低い音が響く. 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.

これらのことから物の素材や形状, 内容量などに応じて手で物を叩いたり振ったりすることで発生する音やその際に手に伝わる感触に違いが生じると窺える. したがって手の周囲の音と手の動きを用いることにより物体識別や内容量認識が可能となると考えられる. また加速度や角速度といった慣性情報は, 音響信号をベースとした物体認識のタスクにおいて類似物体間や騒音環境下での認識精度の向上に貢献することが示されている [13]. このことから本手法においても音響信号と加速度信号, 角速度信号の使用が物体識別や内容量認識に効果的であると見込まれる.

## 3.3 利用イメージ

### 3.3.1 調理支援システムとしての活用例

キッチンにて料理をする際の調理支援を行うシステムとしての活用例を示す. ユーザが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複数の調味料を混ぜ合わせる状況を想定する. 調理過程においてユーザがセンシングデバイスを着用した手で調味料のボトルを調理台に置く動作やその蓋を開ける際の動作などから, システムはユーザが使用している調味料を認識する. ユーザは調味料を注ぐ前後でボトルを振るようにする. この際にシステムはボトル内の調味料の減少量を認識する. システムは認識した調味料とその注がれた量をレシピと照らし合わせた上で, タブレット端末やセンシングに使用しているデバイス（スマートウォッチ等）を通してレシピの進行状況をユーザにフィードバックする. 以上によりユーザは計量容器を使わず, また頻繁にレシピ情報を確認することなく円滑に調理を進めることができる.

### 3.3.2 日用品管理の活用例

生活の中で使用する日用品の消費量を管理するシステムとしての活用例を示す. 例としてティッシュの在庫管理を想定する. システムの利用にあたり, ユーザはセンシングデバイスを着用した手で普段通りにティッシュを使用する. この際にシステムは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, ティッシュの使用量や箱内部の残量を認識し記録する. 以上により, ユーザはスマートフォンやタブレット端末などによってシステムにアクセスし可視化された残量を確認することで買い物計画などに活用ができる. またはシステムは

残量が一定量を下回った際にユーザのスマートフォンなどに通知しEC サービスによる購入を提案するといった使い方も考えられる.

### 3.3.3 読書を行う際の書籍管理や補助デバイスとしての活用例

日常生活における趣味や娯楽などの支援システムとしての本手法の活用例を示す. 例として読書を想定する. システムを利用するにあたり, ユーザは読書を行う際に本の表紙をセンシングデバイスを着用した手で叩くようにする. これでシステムは本のタイトルを識別し, ユーザがこれまでに読み勧めたページ数をスマートフォンやセンシングに使用するデバイス（スマートウォッチ等）を用いてフィードバックする. ユーザはこの情報を元に読書を再開する. この際にシステムは本のページを捲る動きや音から読んだページ数を認識し記録する. 以上によりユーザの読書状況を自動で管理することができる. またはじめに本を叩く際に, 読書をする場所や時間に応じて部屋の照明を点けたり, 本を叩く回数に応じてその明るさを変更したりするといった他のデバイスと協調させる使い方も考えられる.

# 第4章　プロトタイプシステム

本章では提案手法を検証するために構築したプロトタイプシステムの概要とその実装, また実装する上で行った予備実験について述べる.

## 4.1　システム概要

構築したシステムの全容についてその概要を述べる. 図 4.1 はシステムの全体構成である. 本システムは主にセンシングデバイスと分析用コンピュータの 2 点で構成されており, これらは同一のローカルエリアネットワーク（LAN）に接続されている. センシングデバイスはユーザの手首に装着され, 音響信号と加速度信号, 角速度信号をサンプリングし, 逐次各センサデータを UDP プロトコルにて分析用コンピュータに送信する. 分析用コンピュータでは周波数解析と機械学習による分類を行う分析プログラムが実行される. 分析用コンピュータは受信したセンサデータを分析プログラムに入力し, その分析結果を提示する.

本システムを実際に生活空間内で利用する場合には, センシングデバイス及び分析用コンピュータはスマートウォッチなどのウェアラブルデバイスに統合する. また識別結果はその本体やスマートフォン, スマートスピーカなどを通し, アプリケーションに適用した上でユーザにフィードバックすることを想定している. なお, 今回の実装ではセンシングデバイスとして M5StickC[24]（図 4.2）を使用し, 分析用コンピュータには MacBookPro を用いる. また分析用ソフトウェアとは別に機械学習に用いるデータセット構築のため, 起動中に受信したセンサデータをファイルに保存するデータ収集ソフトウェアがある.

## 4.2　使用手順

本システムを動作させる際の手順を以下に示す. なお, 分類器の学習に用いるデータセットは事前に構築済みであることを前提とする.

1. 分析用コンピュータで分析ソフトウェアを起動する
2. 分析ソフトウェアに分類器の訓練データを入力し学習させる, または学習済みの分類器のモデルを読み込ませる
3. センシングデバイスを起動する
4. センシングデバイスを装着した手で物を叩く, または振る
5. 分析用コンピュータの画面上に表示された分析ソフトウェアの処理結果を確認する

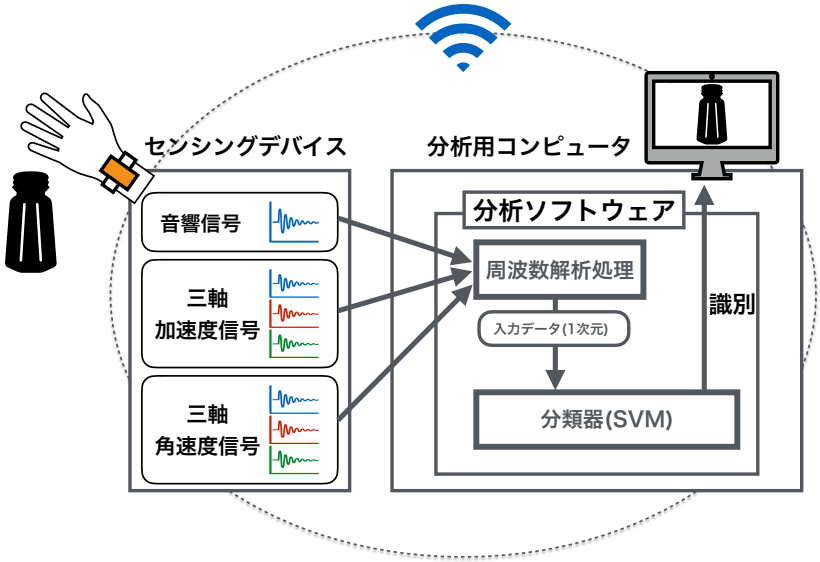



図 4.1: システム構成

図 4.2: センシングに用いるデバイス（M5StickC）

## 4.3 センシングデバイス実装

本システムのセンシングデバイスの概要とその実装について述べる. センシングデバイスとして使用する M5StickC は $240MHz$ で動作するデュアルコアマイクロコントローラ ESP32-PICO を搭載したオープンソースの開発基板である. 本体には 6 軸 IMU センサ MPU6886 やマイクロフォン SPM1423 が内蔵されており, ESP32-PICO とそれぞれ I2C と I2S にて接続されている. また Wi-FI や Bluetooth を利用可能な通信モジュールと本体駆動用のバッテリを内蔵しており, 専用のマウンタを用いることでユーザの手首に着用した状態で無線駆動することができる.

本システムの実装においてセンシングデバイスには, IMU センサによる加速度と角速度の信号データとマイクロフォンによる音響信号データをセンシングする処理, そしてセンシングしたデータを UDP プロトコルにて送信する処理を実装した. センシング処理に関して, 加速度と角速度の信号のサンプリングレートは $400Hz$ とし, 音響信号のサンプリングレートは $22050Hz$ と設定した. これについて, 複数のセンサについて異なる周期でセンシング処理を行う必要がある. また通信処理による処理の遅延も考慮するため, 実装が困難になる問題が発生する. そこで本実装ではマイクロコントローラの機能に着目することで効率の良い処理実装を図った. ESP32-PICO はデュアルコアプロセッサにより複数のタスクを並列に処理することができる. また ESP32-PICO は I2S インタフェースについて Direct Memory Access（DMA）をサポートしているためマイクロフォンでセンシングした音響信号を直接メインメモリ上の DMA バッファに書き込むことが可能である. したがってこれらを用いて本実装では加速度と角速度の信号をセンシングするタスク（タスク 1）と音響信号をセンシングして各データを UDP で送信するタスク（タスク 2）を並列処理するマルチタスク構成として実装を行った. マイクロフォンの制御に DMA コントローラを使用することで, タスク 1 ではサンプリングレートに従って $400Hz$ の周期で IMU センサから 3 軸加速度（図 4.3X,Y,Z）と 3 軸角速度（図 4.3Pitch,Yaw,Roll）の信号データを取得してキューに格納し, タスク 2 にて 10 ミリ秒ごとにタスク 1 のキューと DMA バッファに格納された各データを取り出し, それぞれ UDP プロトコルにて分析用コンピュータに送信することが可能となった.

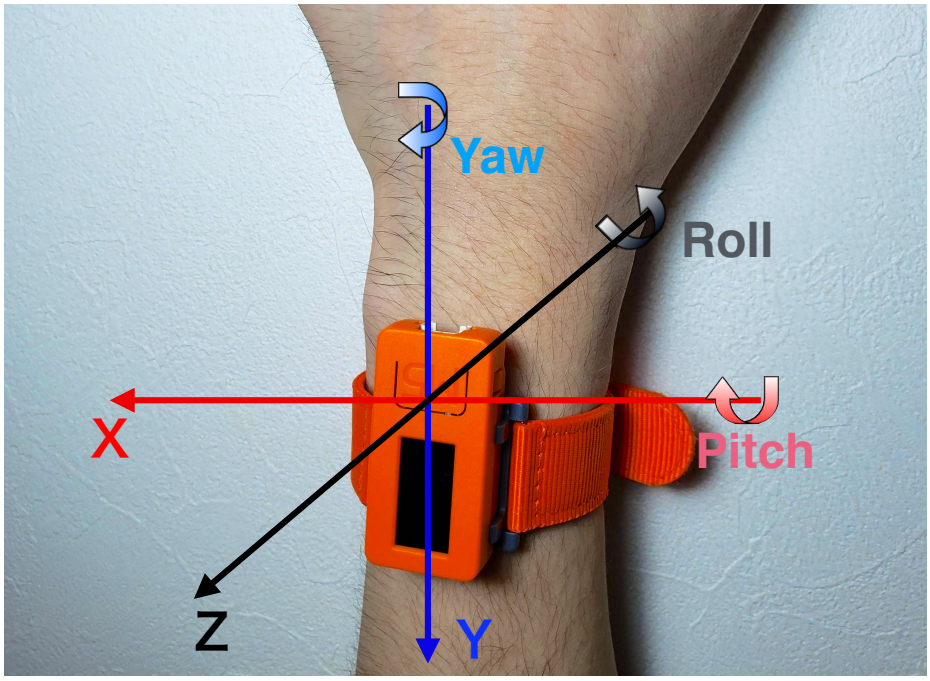



図 4.3: M5StickC 内蔵の 6 軸 IMU センサで取得する加速度と角速度のベクトル

## 4.4 分析ソフトウェア実装

本節では分析用コンピュータが受信したセンサデータを用いて物体識別や内容量認識を行う分析ソフトウェアの実装について述べる. 分析ソフトウェアは受信した信号データに周波数解析を行い特徴量抽出をするプログラムと, 抽出された特徴量のデータを学習・分類する分類器の 2 点から構成される. 後述する予備実験の調査により, 分析ソフトウェアが一度に処理する各センサデータの時間フレーム長 $W_l$ を叩く動作については 250 ミリ秒とし, 振る動作については 500 ミリ秒と設定する. これにより各センサデータの単位時間フレームあたりのサンプル $W$ はそのサンプリングレート $Sr$ より式 4.1 で表される. したがって分析ソフトウェアが一度に処理する音響信号及び加速度と角速度の信号の各データのサンプル数は, 叩く動作についてはそれぞれ 5512 と 100 であり, 振る動作についてはそれぞれ 11025 と 200 である.

$$
\begin{aligned}
W_l &= \begin{cases} 250 & (\text{叩く動作}) \\ 500 & (\text{振る動作}) \end{cases} \\
W &= Sr \times \frac{W_l}{1000}
\end{aligned}
\tag{4.1}
$$

これを踏まえて以下に周波数解析処理と分類器の実装, そして実装に伴って行った予備実験について示す.

### 4.4.1 周波数解析処理

分析ソフトウェアの周波数解析処理にて抽出する特徴量の概要とその抽出手法及び手順について述べる. 分析ソフトウェアでははじめに入力された各信号データに対して周波数解析を行い, 抽出した特徴量から分類器への入力データを作成する. 特徴量について, 加速度と角速度の信号に関しては 3 軸の各サンプル全体に対して高速フーリエ変換（FFT）を行った 3 次元周波数スペクトルを使用する. 音響信号に関しては同様の FFT による周波数スペクトル $F_{sound}$ に加えて, $F_{sound}$ にメルフィルタを適用したメルスペクトル $F_{mel}$ と, 短時間フーリエ変換（STFT) の加算平均 $F_{rms}$ の 3 種類を用いる. なお, いずれの処理においても窓関数はハミング窓とした.

特徴量抽出後, 分類器への入力データとして音響信号の各特徴量と加速度信号の特徴量 $F_{accel}$, 角速度信号の特徴量 $F_{gyro}$ を一次元に連結したデータが出力される. したがって入力データには以下の 3 種類の組み合わせが存在する.

**組み合わせ A** $F_{sound}, F_{accel}, F_{gyro}$

**組み合わせ B** $F_{mel}, F_{accel}, F_{gyro}$

**組み合わせ C** $F_{rms}, F_{accel}, F_{gyro}$

これらの組み合わせについて, 後述する評価実験では各特徴量の組み合わせごとの分類精度を比較し将来的なシステムの実装に役立てる. また各組み合わせごとに音響信号の特徴量 $F_{sound}, F_{mel}, F_{rms}$ のみを用いた場合についても検証し, 本手法にて加速度信号と角速度信号の特徴量を用いる有用性について評価する.

各特徴量の概要とその抽出処理について以下にその詳細を示す.

#### 高速フーリエ変換（FFT）による周波数スペクトル抽出

$F_{sound}$ 及び $F_{accel}, F_{gyro}$ の抽出処理について述べる. FFT により得られる周波数スペクトルの周波数分解能 $\Delta f$ はサンプル数を $N$ 点とした際にサンプリングレートを $f_s$ として, 式 4.2 の関係に表される.

$$\Delta f = \frac{f_s}{N} \tag{4.2}$$

そのため周波数分解能を向上させる目的で一般にゼロパディングと呼ばれる手法が用いられる. これは FFT するサンプルに任意の個数（$n$）の 0 を付与することでサンプル数を増やす手法である. また FFT の代表的なアルゴリズムである Cooley-Turkey 法はサンプル数が 2 の冪乗個であることを前提としている. したがってゼロパディングによりサンプル数（$N + n$）を 2 の冪乗とすることが FFT の前処理として適当である.

これらのことから, 本実装では前処理としてゼロパディングを施した音響信号と加速度信号, 角速度信号のサンプルについて FFT を行い $F_{sound}$ 及び $F_{accel}, F_{gyro}$ を抽出する. これにより叩く動作については, 前処理した音響信号のサンプル（8192 サンプル）と加速度信号と角速

度信号のそれぞれのサンプル（$3 \times 128$ サンプル）から, 4096 次元 の $F_{sound}$ と $3 \times 64$ 次元 の $F_{accel}$, $F_{gyro}$ が得られる. 振る動作については, 前処理した音響信号のサンプル（16384 サンプル）と加速度信号と角速度信号のそれぞれのサンプル（$3 \times 256$ サンプル）から, 8192 次元 の $F_{sound}$ と $3 \times 128$ 次元 の $F_{accel}$, $F_{gyro}$ が得られる.

またアナログ信号の AD 変換時におけるサンプリング可能な周波数の最大値をナイキスト周波数と呼ぶ. ナイキスト周波数は標本化定理によりサンプリングレート $f_s$ の $\frac{1}{2}$ と定義されている. したがって本実装で得られる周波数スペクトルの有効な最大周波数は, $F_{sound}$ については $11025Hz$ であり, $F_{accel}$ と $F_{gyro}$ については $200Hz$ である.

### 周波数スペクトルからメルスペクトルの抽出

$F_{sound}$ から $F_{mel}$ に変換する処理について述べる. メルスペクトルは人間の音響の知覚尺度であるメル尺度に基づいて周波数変換を行ったスペクトルである. 周波数 $f$ のメル周波数 $f_mel$ への変換は, $1000mel = 1000Hz$ を基準として, 式 4.3 で表される.

$$f_{mel} = 1127.010480 \ln \left( \frac{f}{700} + 1 \right) \tag{4.3}$$

本実装ではメル尺度への変換処理としてフィルタバンク分析 [25] を用いる. これはメル尺度上に等間隔に並べたメルフィルタと呼ばれる $L$ 個のバンドパスフィルタを周波数スペクトルに適用する手法である. 各フィルタ $W$ の領域ごとにスペクトルに平滑化処理を行い, フィルタごとの合計値の対数 $m$ を特徴量として用いることにより, スペクトルの特徴の明瞭化とその次元数を $L$ 次元まで圧縮する効果がある.

フィルタバンク分析における各フィルタのチャネル $l$ $(l = 1, \cdots, L)$ の周波数領域上の下限値, 中心値, 上限値をそれぞれ $k_lo(l), k_c(l), k_h(i)$ とした際に, 周波数スペクトル $F$ からメルスペクトルの算出は以下の関係式（4.4, 4.5, 4.6）により表される.

$$m(l) = \ln \sum_{k=k_{lo}}^{k_{hi}} W(k;l)|F(k)| \quad (l = 1, \cdots, L) \tag{4.4}$$

$$W(k;l) = \begin{cases} \frac{k - k_{lo}(l)}{k_c(l) - k_{lo}(l)} & \{k_{lo}(l) \leq k \leq k_c(l)\} \\ \frac{k_{hi}(l) - k}{k_{hi}(l) - k_c(l)} & \{k_c(l) \leq k \leq k_{hi}(l)\} \end{cases} \tag{4.5}$$

$$k_{hi}(l-1) = k_c(l) = k_{lo}(l+1) \tag{4.6}$$

本実装では $L = 20$ とし, 音響信号の特徴量としてその周波数スペクトル $F_{sound}$ から 20 次元のメルスペクトル $F_{mel}$ を抽出する.

### 短時間フーリエ変換（STFT）の加算平均

STFT の概要と $F_{rms}$ の抽出処理について述べる. STFT は短時間のサンプルに対するフーリエ変換を信号データの先頭から少しずつずらしながら適用する手法であり, 主にサウンドスペ

クトログラムなどの周波数と位相の時間変化の分析に用いられる. 本システムは現実の生活空間内での活用を理想としており, 効率の良い処理性能が求められる. したがって STFT により得られる各短時間サンプルの周波数スペクトルの加算平均を用いることで周波数スペクトルの次元数を削減し, これを音響信号の特徴量として用いることで $F_{sound}$ や $F_{mel}$ と比較して, 分類器の学習及び分類の高速化とその表現力の向上が見込めると考えた.

本実装では STFT の短時間サンプル数を 512 とし, フーリエ変換のアルゴリズムには FFT を用いる. これを音響信号データの先頭から 50% 重複してずらしながら適用し, 各区間の周波数スペクトルを抽出する. なお, データの終端付近では FFT するサンプル数が不足するため適宜ゼロパディングを用いる. 最後に抽出された各周波数スペクトル全体を加算平均することにより, 音響信号の特徴量として 256 次元の周波数スペクトル $F_{rms}$ が得られる.

### 4.4.2 SVM を用いた分類器

周波数解析により得られたデータを用いて学習及び分類を行う分類器の実装について述べる. 分析ソフトウェアにおける分類器の実装には教師あり機械学習手法の 1 つであるサポートベクタマシン（SVM）を用いる. そのため本システムを利用する際には事前にデータ収集ソフトウェアを使用して学習用データセットの構築及びそれを用いた分類器の学習を行う必要がある. SVM による分類器の実装にあたり, プログラミング言語 Python とその機械学習ライブラリである Scikit-learn を使用した. Scikit-learn には GridSearch と呼ばれる手法により, 分類器の学習時にデータセットに対して交差検証することでハイパパラメータを自動調整する機能がある. 本システムにおいてもこれを用いて分類器の学習ごとにハイパパラメータの調整を行う. この際パラメータの組み合わせについて, カーネル関数には'linear', 'rbf', 'poly', 'sigmoid' の 4 種類を使用し, 正則化パラメータに'1', '10', '100' の 3 種類, ガンマ値には'0.01' から'1.0' までの 50 段階を用いる.

## 4.5 予備実験: 分析ソフトウェアで処理するセンサデータの時間長調査

4.4 で示した分析ソフトウェアの実装において一度に処理するセンサデータの時間長について, その適切な長さを求めるために行った予備実験について述べる. 本システムではユーザの動作として物を叩く動作と振る動作を想定している. 叩く際の音は瞬間的に響き, 振る際の音は比較的長く響くことからそれぞれ信号データの変化に違いがあると考えられる. そこで叩く際と振る際のそれぞれについて, 分析ソフトウェアに入力する適切な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と振る際のセンサデータを収集して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 4.4A）, 本（図 4.4B）, スマートフォン（図 4.4C）, プラスチックボトル（図 4.4D）の 4 種類である. また振る対象は内容量の異なる食塩を封入した 3 つの円筒形プラスチック容器（図 4.5A: 全容量の 10%, 図 4.5B:全容量の 50%, 図 4.5C: 全容量の 90%）を用いる. 実験は各対象物についてデータ収集ソフトウェアを用いて 2 秒間の間に一度叩くまた振る際のセンサデータをそれ

ぞれ 5 回収集し, 得られたデータをグラフに描画した上で目視でその特徴を比較することにより行う. なお, 図 4.4 の机を除く 3 種類の対象物については机の上に置いた状態とする. また叩き方と振り方はそれぞれ左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（図 4.6）及び容器の側面を握って縦に振る動作（図 4.7）とした.

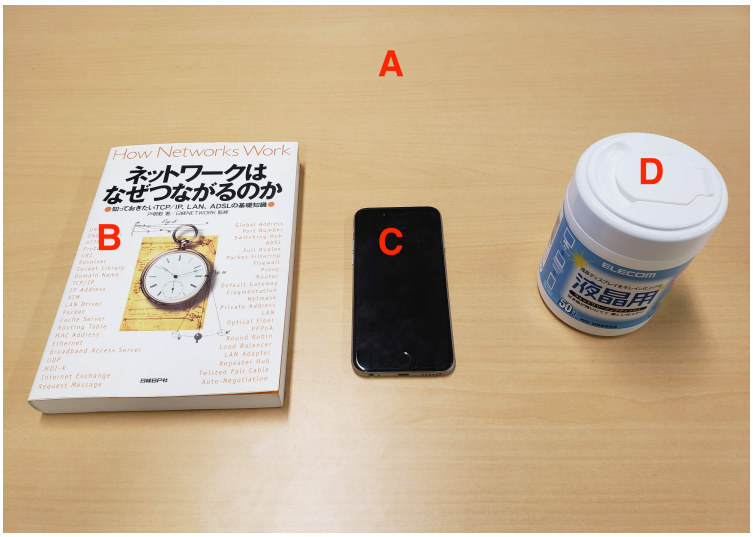



図 4.4: 4 種類の叩く対象物 (A: 机, B: 本, C: スマートフォン, D: プラスチックボトル)

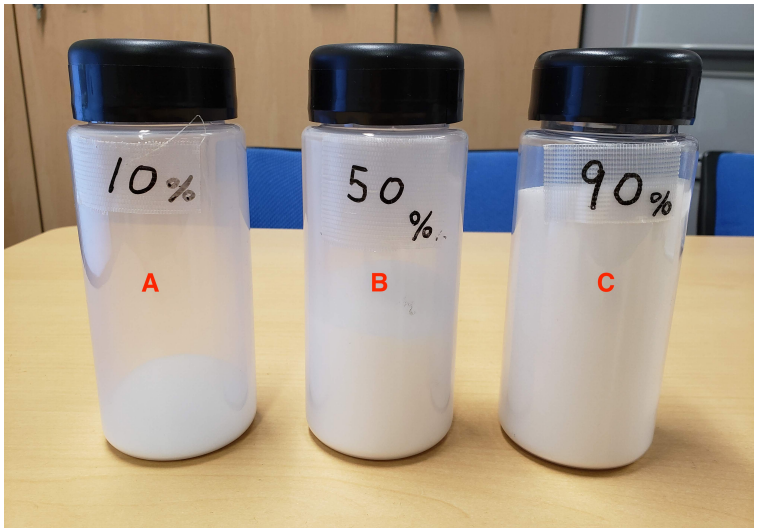



図 4.5: 内容量の異なる食塩を封入した 3 つの円筒形プラスチック容器 (A: 10%, B: 50%, C: 90%)

図 4.6: ノックするように叩く動作


図 4.7: 振る際の容器の握り方

収集した全センサデータの時系列変化について, グラフ描画した結果を図 4.8, 4.9, 4.10, 4.11 及び図 4.12, 図 4.13, 図 4.14 に示す. 叩いた際のセンサデータについてはいずれの対象物においても音響信号の最大ピークを基準としてその前方 100 ミリ秒, 後方 150 ミリ秒の 250 ミリ秒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. また同様に振る際のセンサデータについては音響信号の最大ピークを基準としてその前後 250 ミリ秒の 500 ミリ秒区間に変化が集中していると見られた. この結果から分析ソフトウェアが一度に処理する時間フレーム長について, 叩く動作については 250 ミリ秒とし振る動作については 500 ミリ秒と定めた.

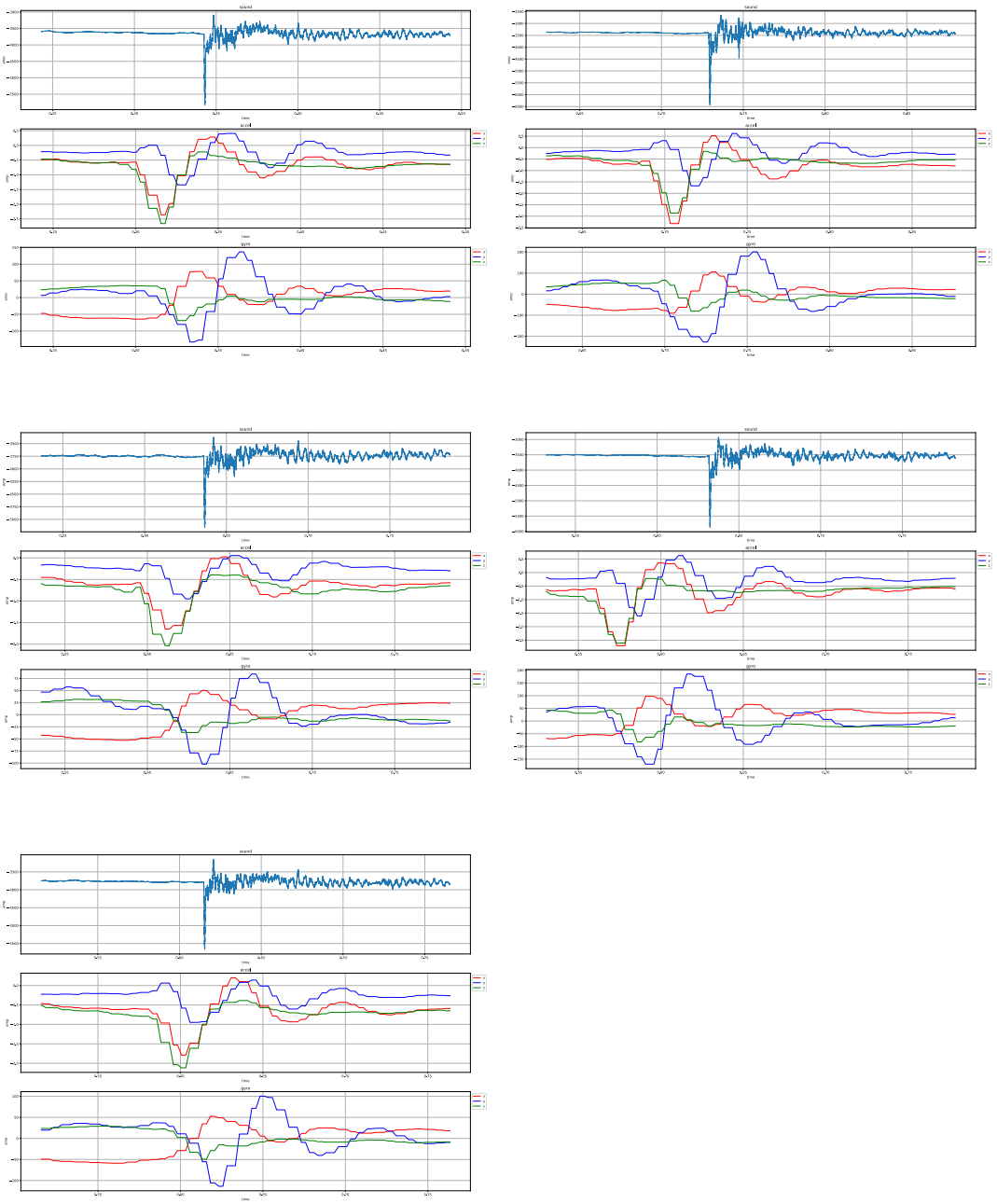


図 4.8: 机を叩いた際の 250 ミリ秒のセンサデータ

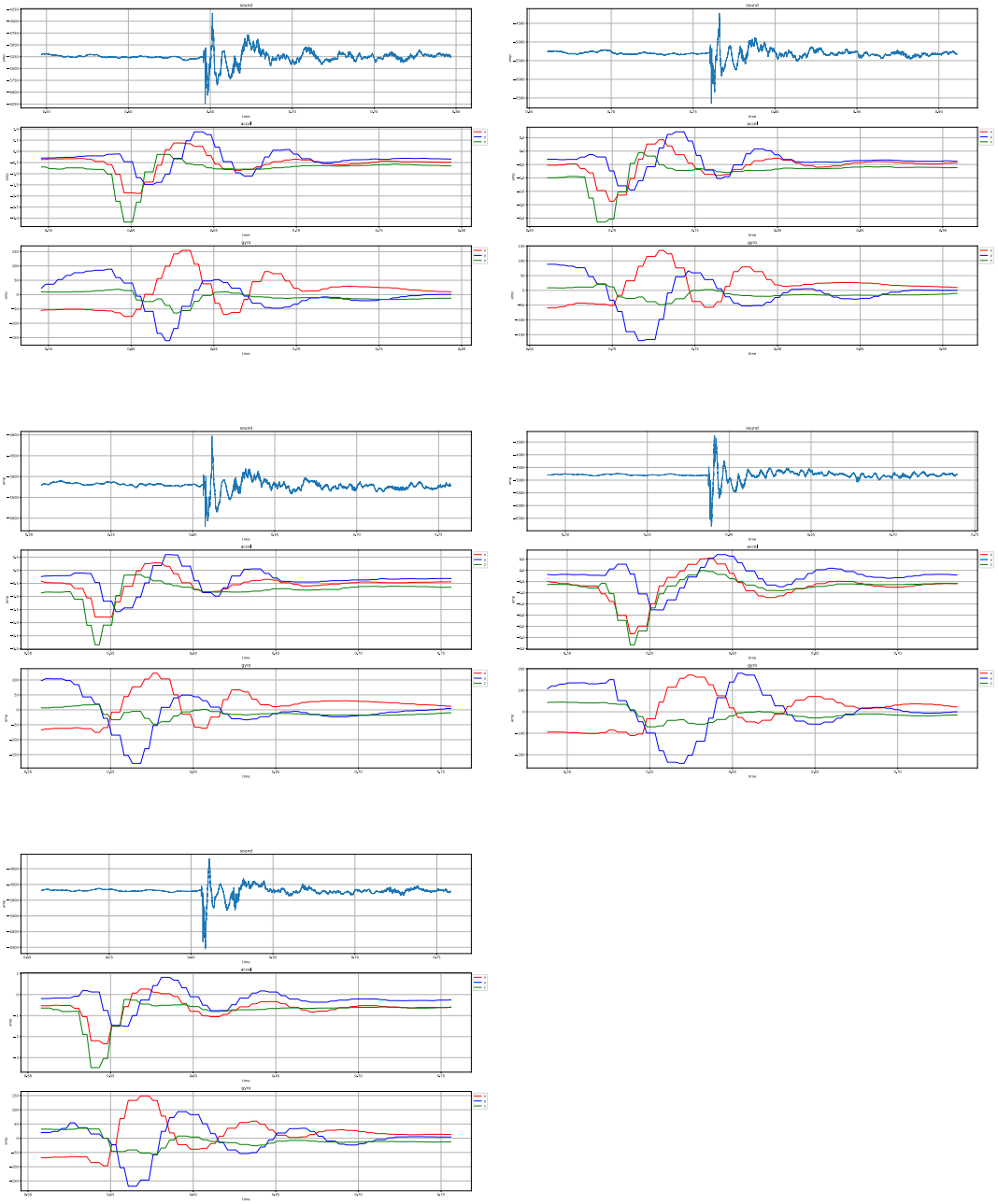


図 4.9: 本を叩いた際の 250 ミリ秒のセンサデータ

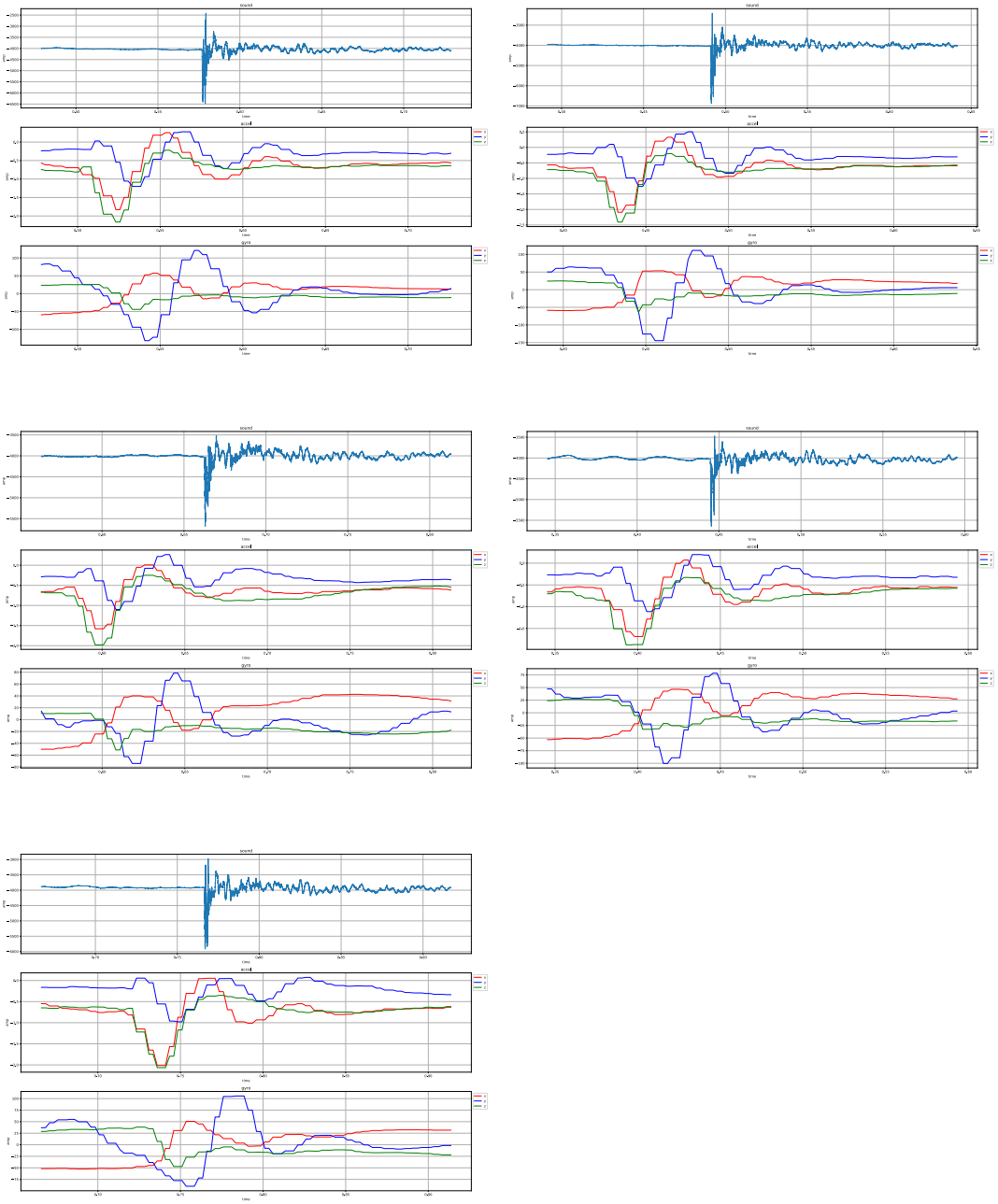


図 4.10: スマートフォンを叩いた際の 250 ミリ秒のセンサデータ

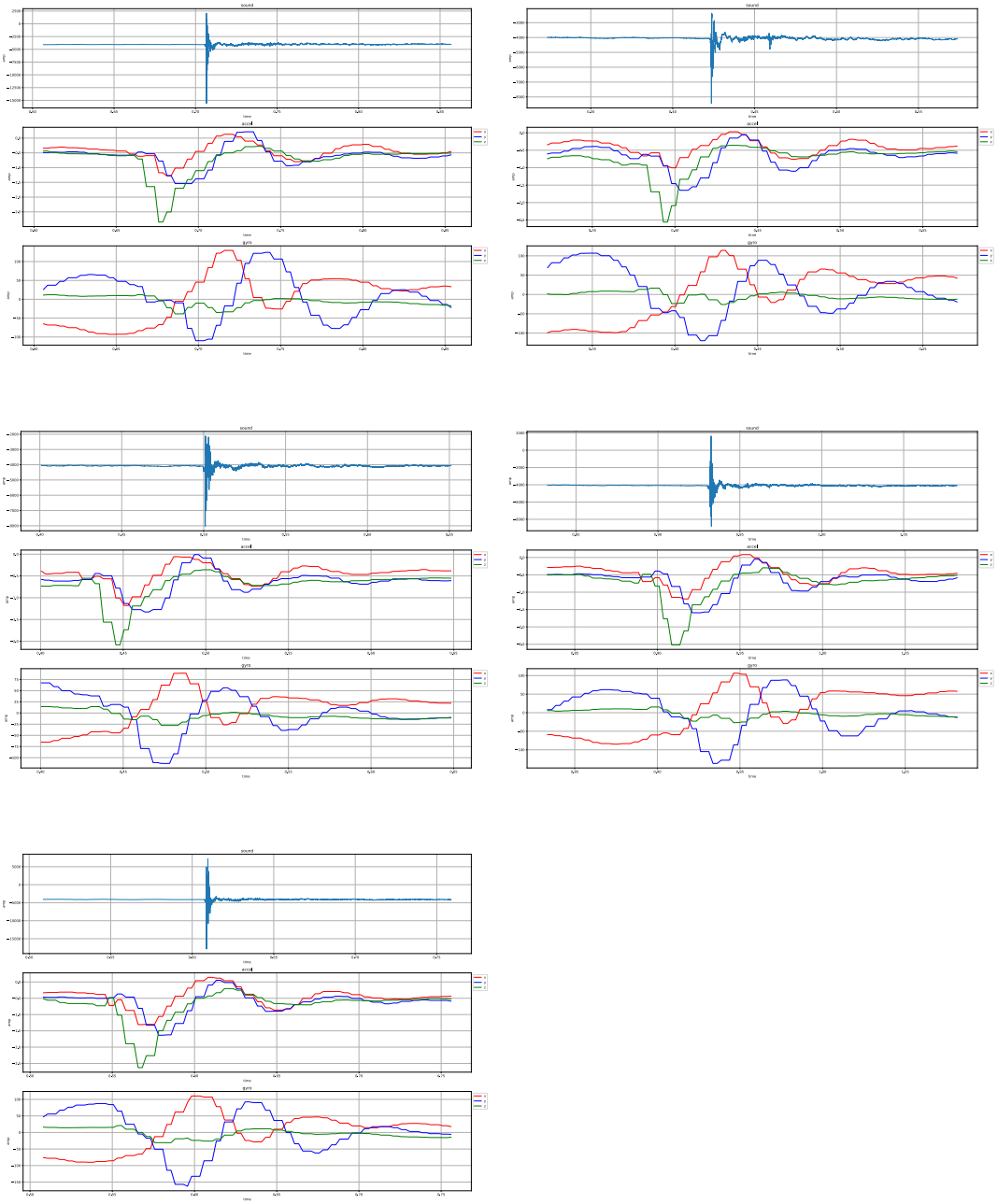


図 4.11: プラスチックボトルを叩いた際の 250 ミリ秒のセンサデータ

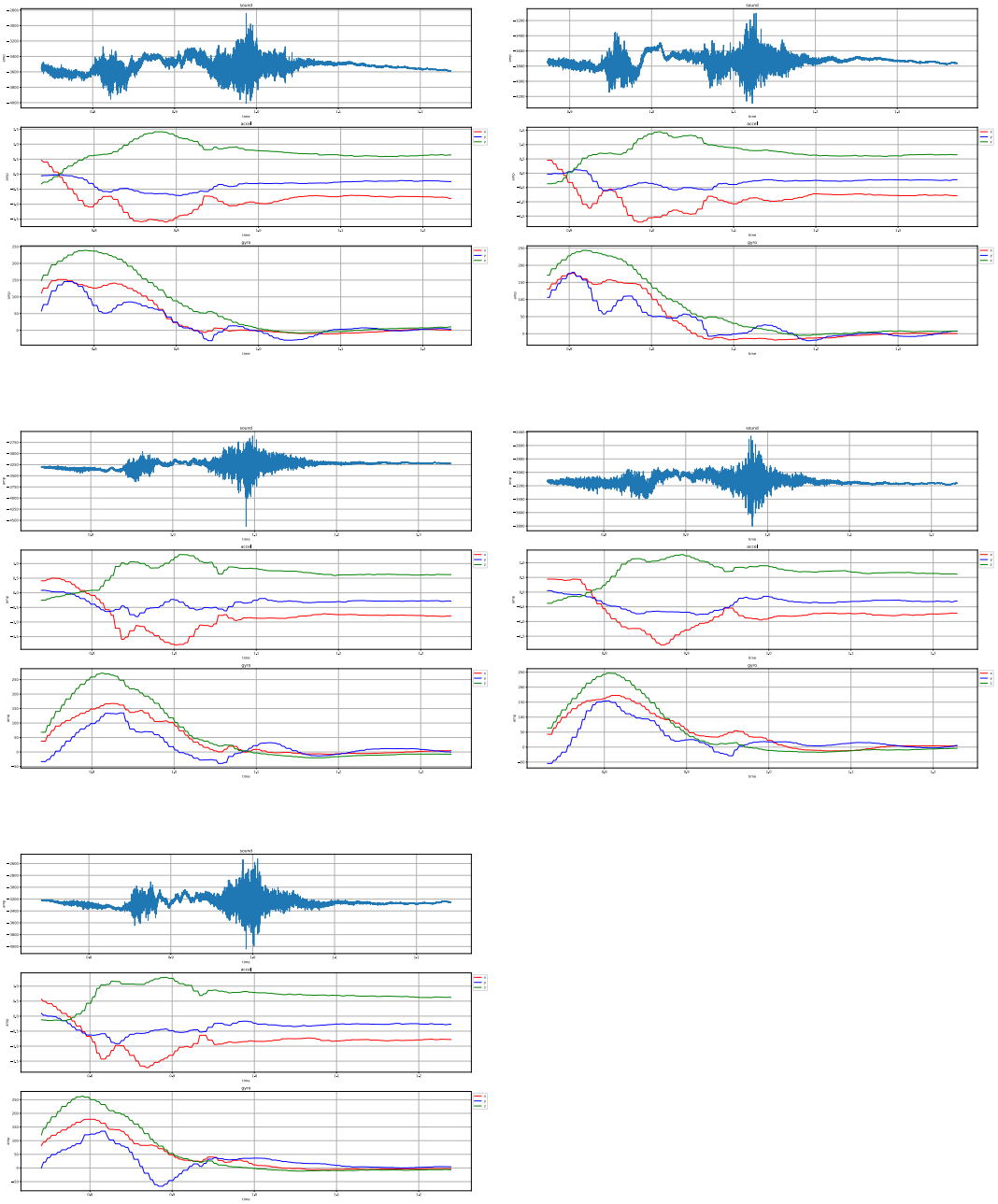


図 4.12: 全容量の 10%の食塩を封入した容器を振った際の 500 ミリ秒のセンサデータ

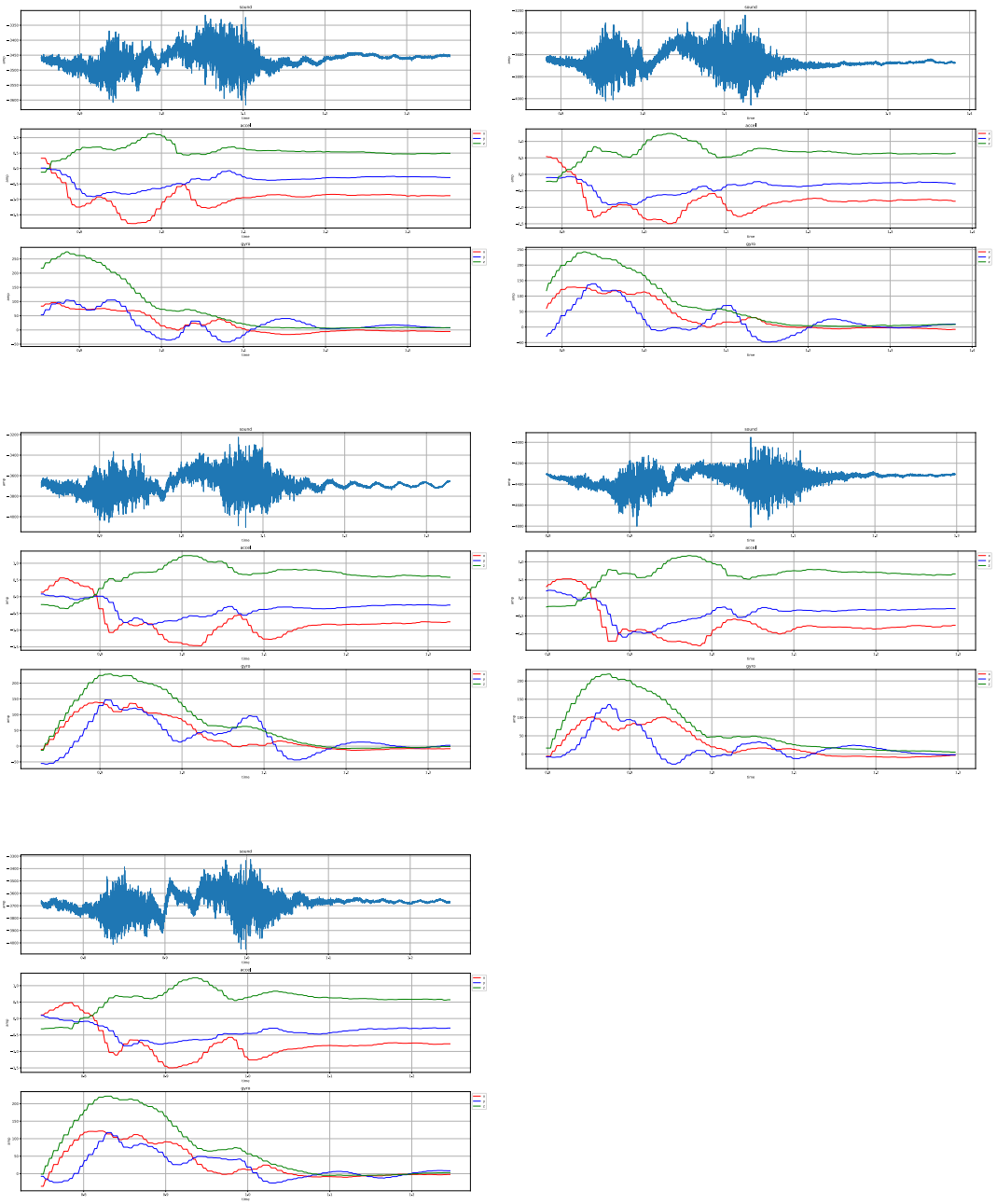


図 4.13: 全容量の 50%の食塩を封入した容器を振った際の 500 ミリ秒のセンサデータ

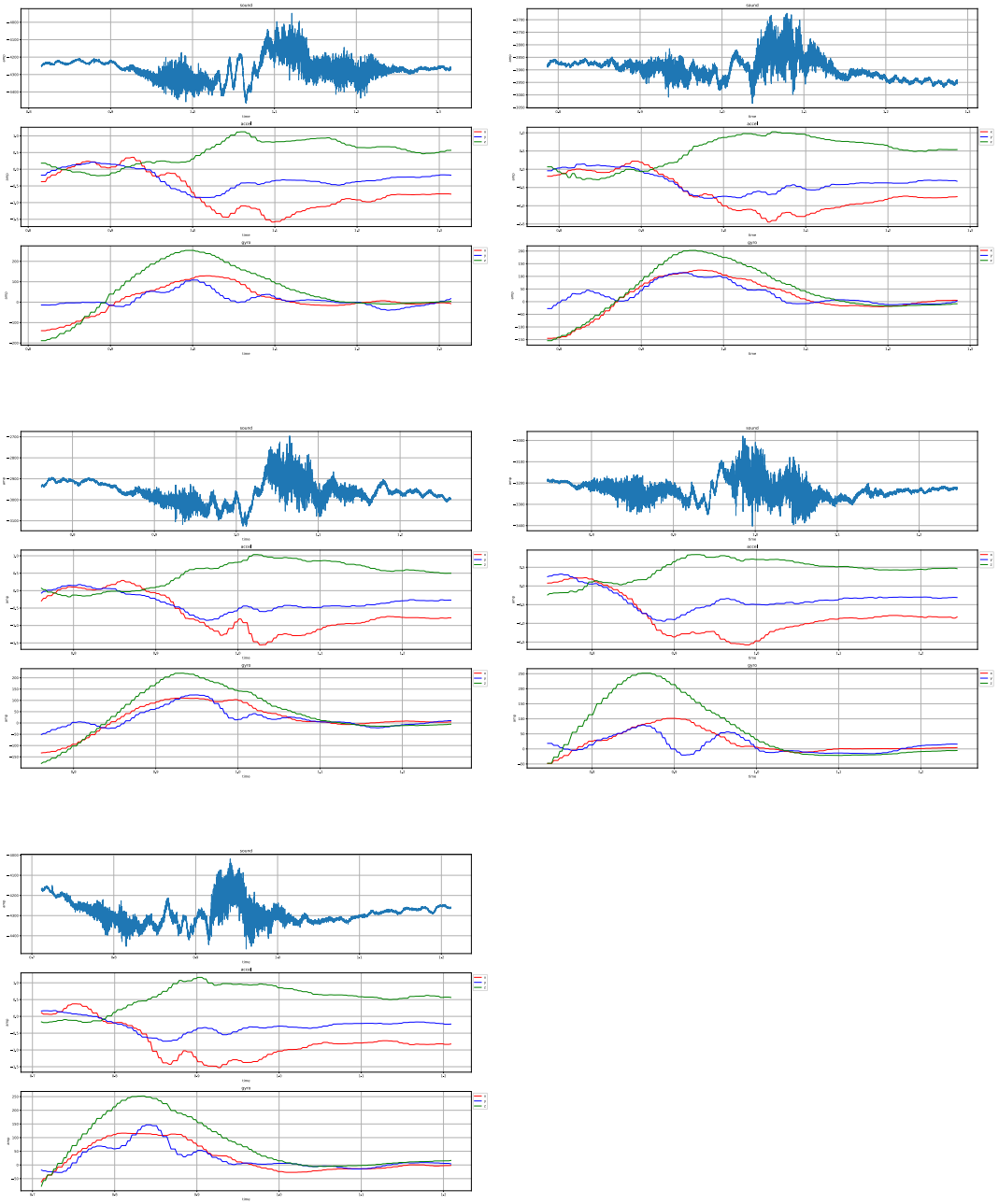


図 4.14: 全容量の 90%の食塩を封入した容器を振った際の 500 ミリ秒のセンサデータ

# 第5章　評価実験

実装したプロトタイプシステムを使用して, 物を叩いた際のデータによる物体分類と容器を振る際のデータによる内容量分類を行い, 分類精度を検証する. なお, 以下に示す実験はシステムの使用手順と異なり, 評価に用いるセンサデータを収集して特徴量抽出を行った上でデータセットとして保存し, それを用いて分類実験を行うことにより検証する. また分類器に入力するデータとして, 4 章で示した 3 種類の特徴量の組み合わせの場合とそれらの音響信号の特徴量のみを用いる場合でそれぞれ比較し, 本手法にて加速度信号と角速度信号を用いる有用性の評価を行う.

## 5.1　実験 1: 物を叩くことによる物体分類

この実験では本手法を生活空間内で使用することを想定して日用品や生活用品の分類に取り組む. 実験に用いる対象物は本（ラベル: book, 図 5.1A）, アルミ缶（ラベル: bottle-can, 図 5.1B）, プラスチックボトル（ラベル: bottle-pla, 図 5.1C）, ティッシュ箱（ラベル: box-paper, 図 5.1D）, ウェットティッシュ箱（ラベル: box-pla, 図 5.1E）, 机（ラベル: desk, 図 5.1F）, マウス（ラベル: mouse, 図 5.1G）, スマートフォン（ラベル: phone, 図 5.1H）, ガラスポット（ラベル: pot-pp, 図 5.1I）, ステンレスポット（ラベル: pot-sten, 図 5.1J）, プラスチックスプレー（ラベル: spray-pp, 図 5.1K）, タブレット端末（ラベル: tablet, 図 5.1L）の 12 種類である. 分類はこれら 12 種類に何も叩いていない状態 (ラベル: null) を加えた 13 クラスについて検証する.

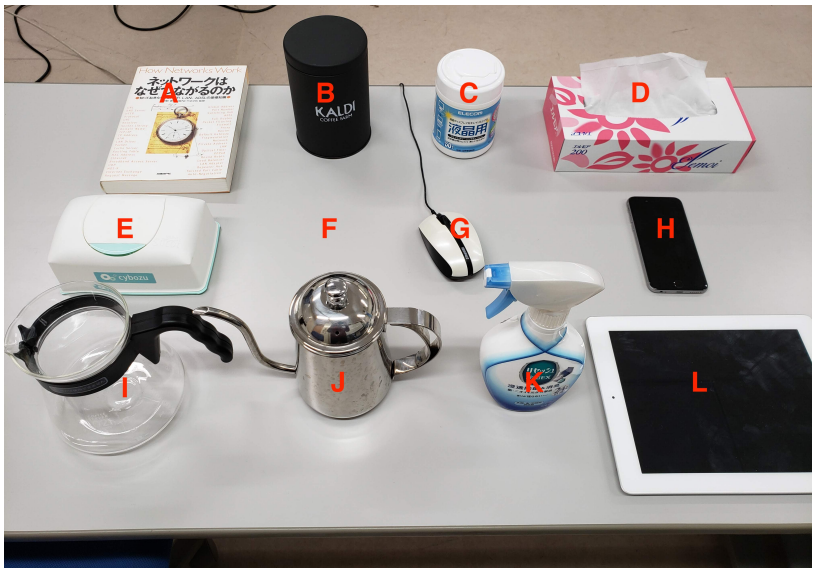



図 5.1: 実験 1 に使用する 12 種類の対象物

### 5.1.1 実験内容

実験は対象物を叩いた際のセンサデータの収集とデータセットの構築, そして分類器の学習とテストの 3 点について行う. 被験者は著者 1 名である. 以下に実験の手順とその内容について述べる

はじめにデータの収集を行う. 具体的な収集の手順は以下の通りである. なお, 説明の便宜上対象物に何も叩いていない状態も含め, その叩き方は空中で手をランダムに動かしたものとする.

**手順 1** 左手首にセンシングデバイスを装着する

**手順 2** 分析用コンピュータにてデータ収集ソフトウェアを 2 秒間起動する

**手順 3** 2 秒間の間に実験対象物を 1 回叩く

**手順 4** 同じ対象物について手順 2 から手順 3 を 10 回繰り返す

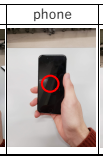
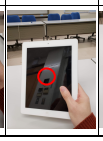
**手順 5** 30 秒程度休憩

**手順 6** 対象物を変更して手順 2 から手順 5 までを行う

**手順 7** 手順 2 から手順 6 までを全対象物について行った（1 セッション）後, これを 5 回繰り返す

収集は 13 クラスの分類対象についてそれぞれ 10 試行の収集を 1 セッションとして 5 セッション行う. 叩き方は図 4.6 と同様にノックするような叩き方とした. また各対象物は表 5.1 に示すように, 右手で持つ（hold）, 或いは机の上に置く（put）といった, それぞれ日常生活で使用する際に想定される状態とし, 赤丸で示した部分を叩くこととした. 以上の作業により合計で 650 試行分（13 種類 × 10 回 × 5 セッション）の各センサデータが収集される.

表 5.1: 叩く際の対象物の状態と叩く部分（赤丸で表示）の一覧

| 状態 | 右手で持つ（hold） | | | | | | （desk以外はdeskの上に）置く（put） | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| label | book | phone | pot-glass | pot-sten | spray | tablet | bottle-can | bottle-pla | box-paper | box-pla | desk | mouse |
| 図 |  |  |  |  |  |  |  |  |  |  |  |  |

データの収集後, 分析ソフトウェアの周波数解析処理により特徴量抽出を行う. この際各収集試行ごとに, 4 章の予備実験の調査結果に基づいて, 音響信号の最大ピークを基準とした 250 ミリ秒の区間のセンサデータから特徴量を抽出する. これにより得られた 650 組の $F_{sound}$, $F_{mel}$, $F_{rms}$, $F_{accel}$, $F_{gyro}$ をデータセットとして保存する.

最後に構築したデータセットを用いて 4 章で述べた特徴量の組み合わせごとに, 分析ソフトウェアの分類器を用いた分類精度を評価する. 分類についてはデータセットの 8 割にあたる

520 データ（各クラスごとに 40 データをランダムに選出）を訓練データとし残りの 130 データをテストデータとして分類を行う検証 1 と, データ収集時の各セッションを 1 グループとした Leave-One-Group-Out-Cross-Validation（LOGOCV）による検証 2 の 2 通りを行う. なお, 検証 2 で用いる分類器のハイパパラメータには検証 1 の Grid Search によって定められたものを設定する. また特徴量抽出から分類までの処理全体に要した時間についても評価を行う.

### 5.1.2 分類結果

4 章で示した音響信号の特徴量 $F_{sound}$, $F_{mel}$, $F_{rms}$ と加速度信号の特徴量 $F_{accel}$, 角速度信号の特徴量 $F_{gyro}$ の組み合わせ A, B, C について, 各組み合わせごとの検証 1 の結果を図 5.2, 図 5.3, 図 5.4 の混同行列に示す. 組み合わせ A について全体の正解率は 92% であったが, 'desk' と'phone' については F 値が 75% に落ち込む結果となった. 特に'phone' については材質が類似する'tablet' への誤分類が大きく表れた. 組み合わせ B については形状の影響か'book' と'tablet' 間でそれぞれ誤分類が目立つ. それ以外についても'book' は適合率が小さく, 全体的には再現率が 65% と低い結果となった. 組み合わせ C では平均 F 値と正解率がどちらも 93% と安定した結果となった.

検証 2 について, LOGOCV の平均正解率は組み合わせ A で 70%($SD = 20\%$), 組み合わせ B で 45%($SD = 11\%$), 組み合わせ C で 72%($SD = 19\%$) であった. 検証 1 の結果と比較していずれも正解率が 20% から 30% 程度低い結果であり, またテストデータのグループごとにも標準偏差が大きく表れていることから分類器の汎化性能の確認には至らなかった.

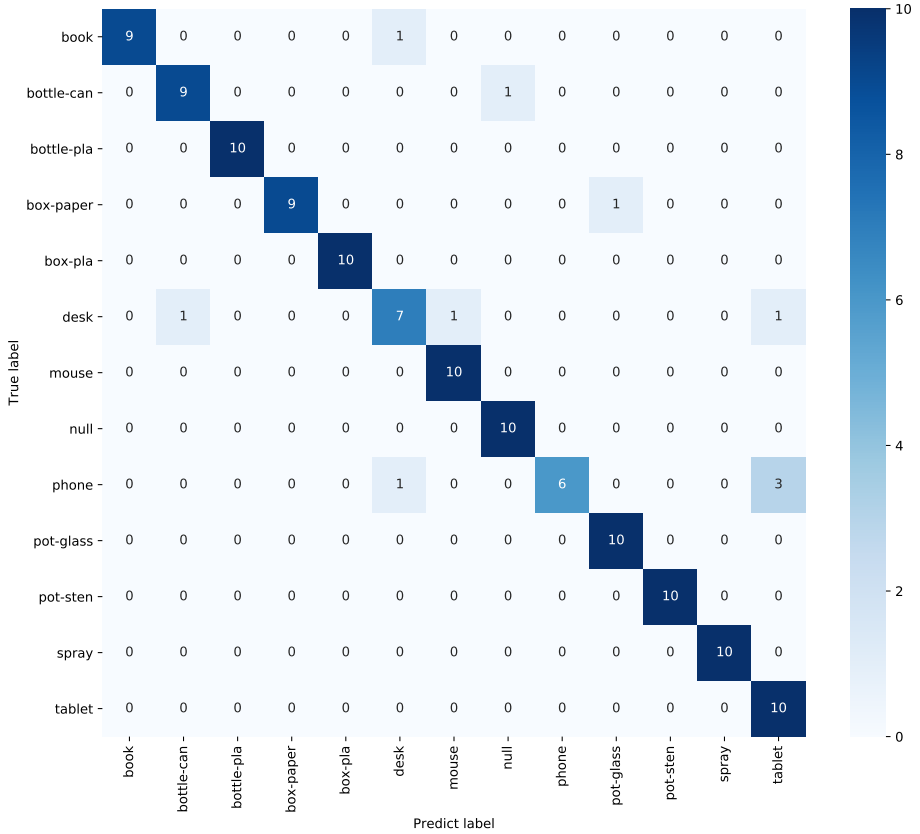



図 5.2: 組み合わせ A（$F_{sound}$, $F_{accel}$, $F_{gyro}$）を分類器への入力データとした際の物体分類結果

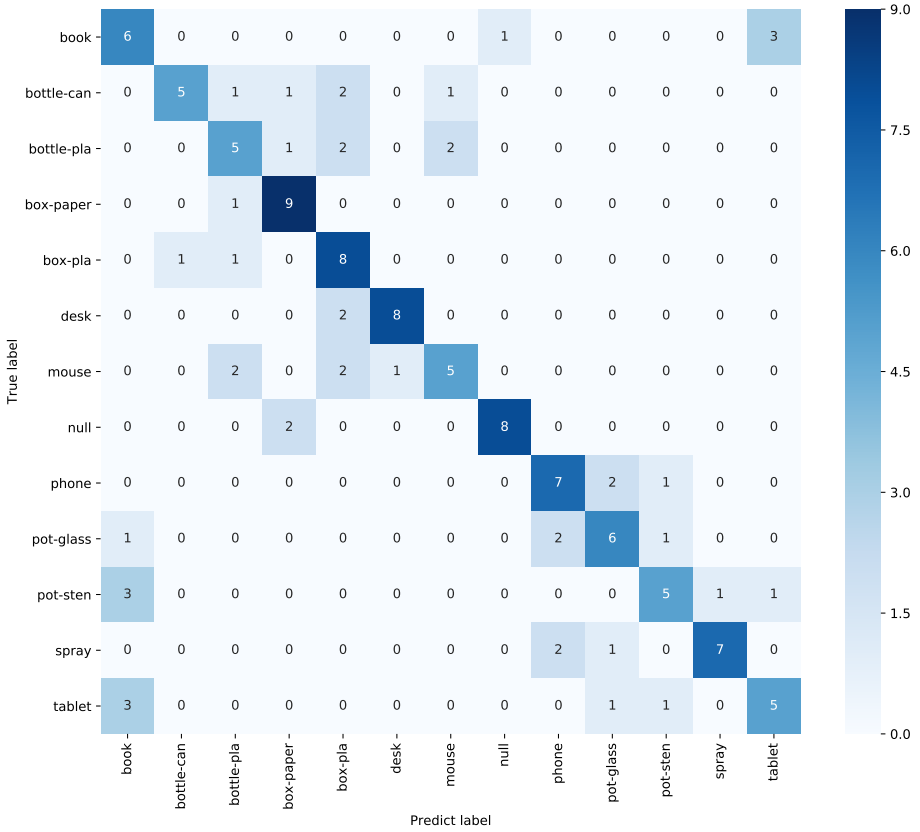



図 5.3: 組み合わせ B（$F_{mel}$, $F_{accel}$, $F_{gyro}$）を分類器への入力データとした際の物体分類結果

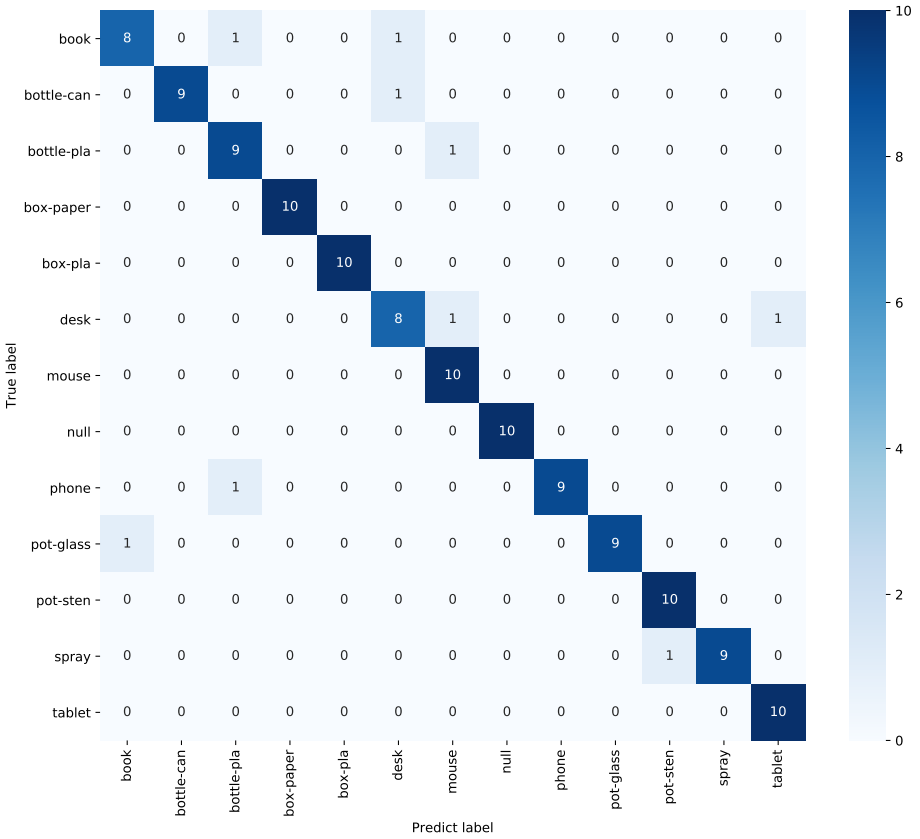



図 5.4: 組み合わせ C（$F_{rms}$, $F_{accel}$, $F_{gyro}$）を分類器への入力データとした際の物体分類結果

次に組み合わせ ABC からそれぞれ $F_{accel}$ と $F_{gyro}$ を除き, 分類器の学習とテストに音響信号の特徴量（$F_{sound}$, $F_{mel}$, $F_{rms}$）のみを用いた場合の検証 1 の結果を図 5.5, 図 5.6, 図 5.7 に示す. 組み合わせ ABC を用いた検証 1 の結果と比較していずれも局所的に再現率の低下が見られるが, 全体を通してはわずかに正解率が高い結果となった. 特に $F_{mel}$ のみを用いた場合については組み合わせ B と比べて 10% 以上の正解率向上が見られる.

検証 2 の LOGOCV の結果では, 分類器への入力データに $F_{rms}$ のみを用いた場合について平均正解率が 67%（$SD = 19\%$）となり, 組み合わせ C の結果を若干下回ったが, 各グループごとの分類結果を t 検定を用いて検定したところ有意差は認められなかった（$p = 0.72 > 0.05$）. したがって今回の実験において, 物体分類のデータセットとして音響信号の特徴量に加速度信号と角速度信号の特徴量を加えることによる分類精度への寄与率は低いと考えられる.

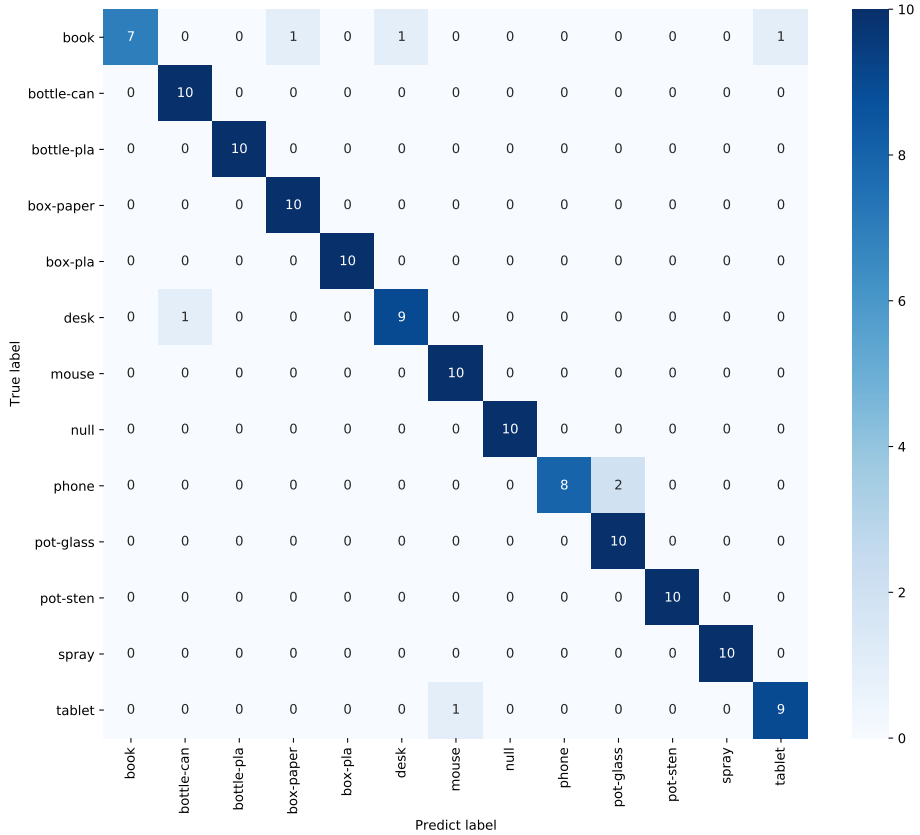



図 5.5: $F_{sound}$ のみを分類器への入力データとした物体分類結果

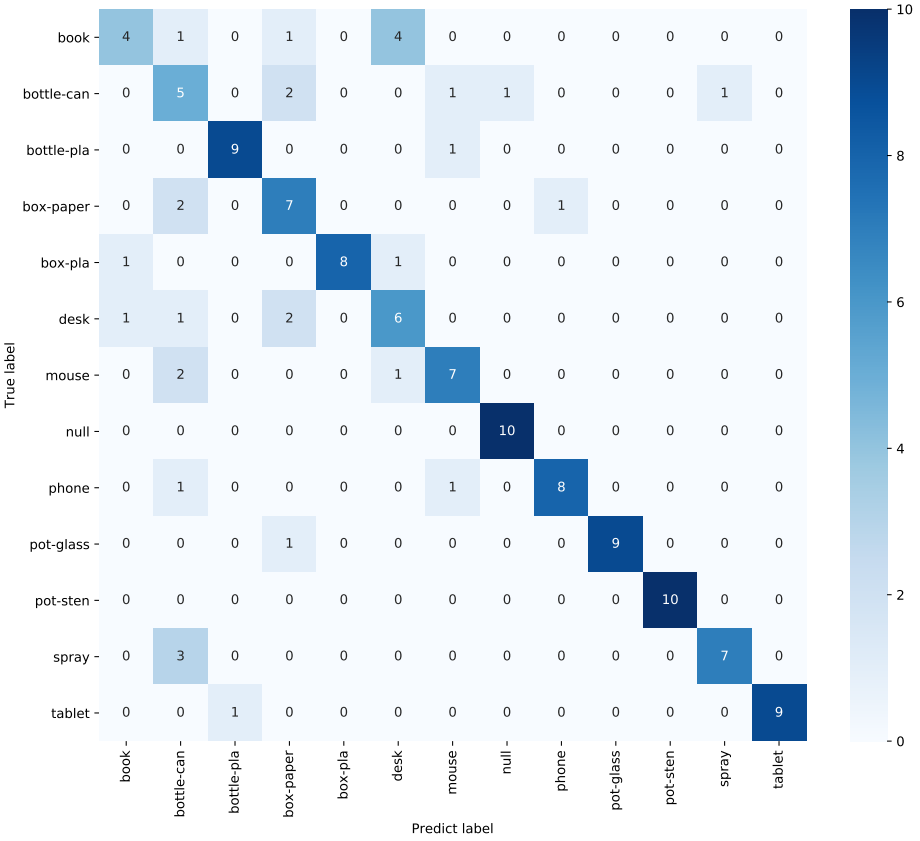



図 5.6: $F_{mel}$ のみを分類器への入力データとした物体分類結果

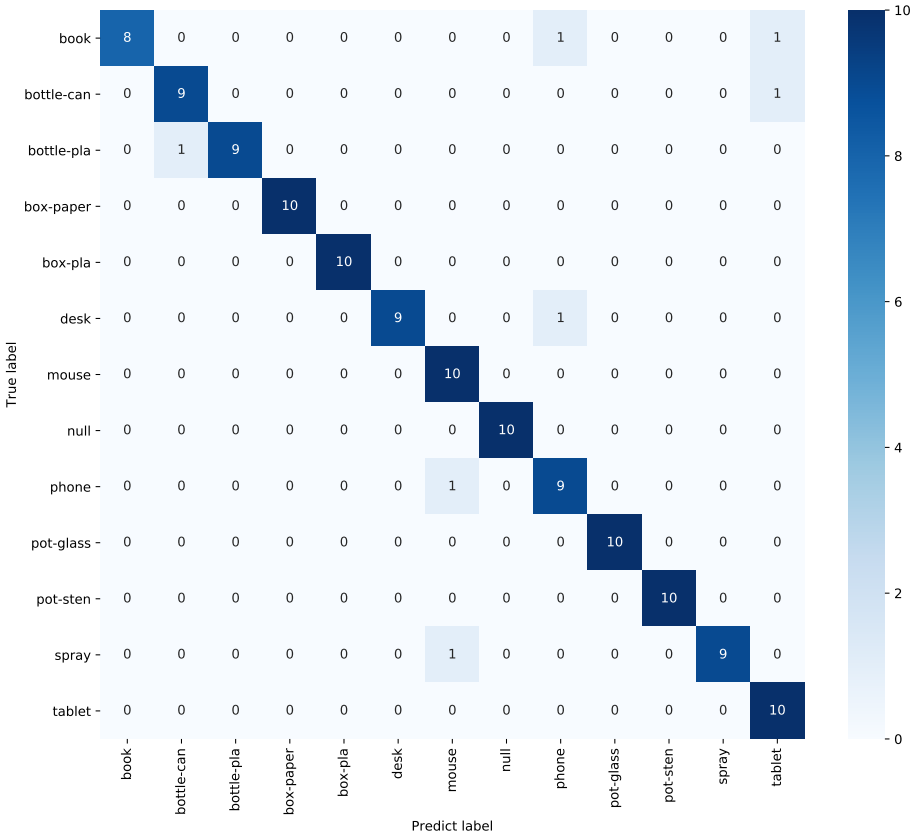



図 5.7: $F_{rms}$ のみを分類器への入力データとした物体分類結果

## 5.2　実験 2: 容器を振ることによる内容量分類

この実験では本手法により生活用品の内容量認識を目標として, その実現性を検証するために内容量の異なる容器について振る際のセンサデータを用いた内容量分類に取り組む. 実験にはそれぞれ異なる 5 段階の内容量（10%, 30%, 50%, 90%）の食塩を封入した 5 つの同一円筒形プラスチック容器（ラベル: salt_10, 図 5.8A）, （ラベル: salt_30, 図 5.8B）, （ラベル: salt_50, 図 5.8C）, （ラベル: salt_70, 図 5.8D）, （ラベル: salt_90, 図 5.8E）を使用する.

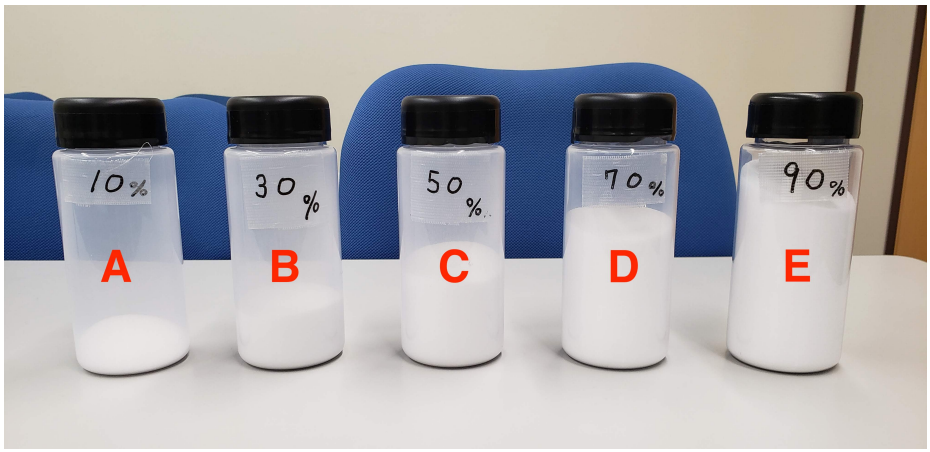



図 5.8: 実験 2 に使用する内容量の異なる 5 つの容器

### 5.2.1　実験内容

本実験は実験 1 と同様にデータ取得と特徴量抽出によるデータセット構築, そして分類器による分類結果の評価の 3 点について行う. 以下にこれらの手順と内容について述べるなお, 被験者についても同じく著者 1 名で行う.

はじめにデータの収集を行う. 収集手順は以下の通りである.

**手順 1** 左手首にセンシングデバイスを装着する

**手順 2** 分析用コンピュータにてデータ収集ソフトウェアを 2 秒間起動する

**手順 3** 2 秒間の間に容器を 1 回振る

**手順 4** 同じ内容量の容器について手順 2 から手順 3 を 10 回繰り返す

**手順 5** 30 秒程度休憩

**手順 6** 内容量の異なる容器に変更して手順 2 から手順 5 までを行う

**手順 7** 手順 2 から手順 6 までを全容器について行った（1 セッション）後, これを 5 回繰り返す

容器の振り方は図 4.7 同様に容器の側面を握り縦に振るような動作とした. 以上により 5 段階の内容量についてそれぞれ 10 試行の収集を 1 セッションとして, 5 セッションの収集を行うことで, 合計で 250 試行分（5 段階 $\times$ 10 回 $\times$ 5 セッション）の各センサデータが収集される.

データの収集後, 分析ソフトウェアの周波数解析処理を用いて, 各収集試行ごとに音響信号の最大ピークを基準とした 500 ミリ秒の区間のセンサデータについて特徴量の抽出を行う. これは 4 章の予備実験の調査結果に基づいた値である. 特徴量抽出処理により 250 組の $F_{sound}$, $F_{mel}$, $F_{rms}$, $F_{accel}$, $F_{gyro}$ を取得し, これをデータセットとして保存する.

最後に実験 1 と同様に, 4 章で述べた特徴量の組み合わせごとに分析ソフトウェアの分類器による分類精度を検証し, 実行時間の評価を行う. 分類検証はデータセットの 8 割にあたる 200 データ（各内容量ごとに 40 データをランダムに選出）を訓練データとし残りの 50 データをテストデータとして分類を行う検証 1 と, データ収集時の各セッションを 1 グループとした LOGOCV による検証 2 の 2 通り行う. 検証 2 で用いる分類器のハイパパラメータについては検証 1 の Grid Search によって定められたものを用いることとする.

### 5.2.2 分類結果

分類結果について述べる. 検証 1 では組み合わせ ABC をそれぞれ特徴量として用いた際の分類結果について, いずれの組み合わせにおいても同一の内容量（salt_90）に偏って分類されていた（図 5.9, 図 5.10, 図 5.11）. また検証 2 の LOGOCV の結果についても同様であった.

これらの結果を踏まえて音響信号の特徴量（$F_{sound}$, $F_{mel}$, $F_{rms}$）のみを用いて検証した際の分類結果を図 5.12, 図 5.13, 図 5.14 に示す. 検証 1 については, $F_{sound}$ 及び $F_{rms}$ をそれぞれ分類器への入力データとした際, これらに $F_{accel}, F_{gyro}$ を加えた組み合わせ A 及び組み合わせ C と同様に, salt_90 に偏った分類結果となった. 一方で $F_{mel}$ のみを用いた場合には最小容量（salt_10）と最大容量（salt_90）について 90% 以上の高い再現率が確認された. またその他の中間の内容量（salt_30, salt_50, salt_70）についても組み合わせ B と比較して分類結果に大きな偏りは見られず, 全体の正解率は 74% となった.

検証 2 の LOGOCV の結果について, $F_{mel}$ のみで検証した際に 58% の平均正解率が示された. その他の音響信号の特徴量（$F_{sound}, F_{rms}$）については検証 1 と同様の結果となった.

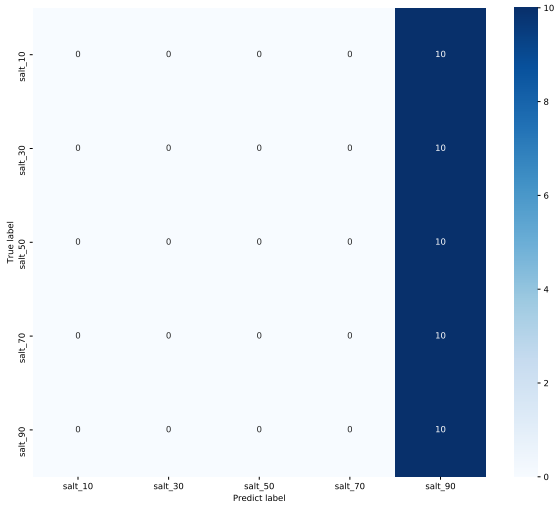



図 5.9: 組み合わせ A（$F_{sound}$, $F_{accel}$, $F_{gyro}$）を分類器への入力データとした際の内容量分類結果

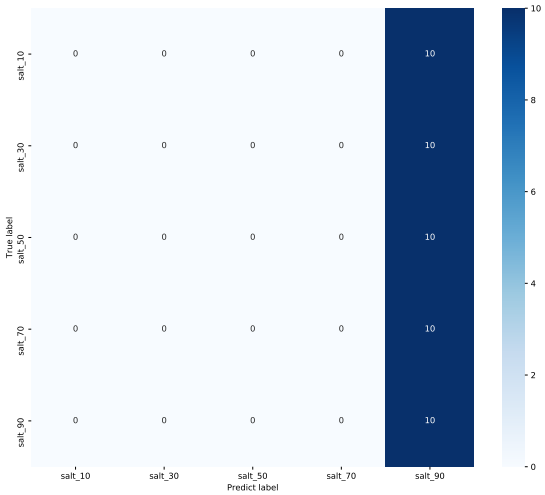



図 5.10: 組み合わせ B（$F_{mel}$, $F_{accel}$, $F_{gyro}$）を分類器への入力データとした際の内容量分類結果

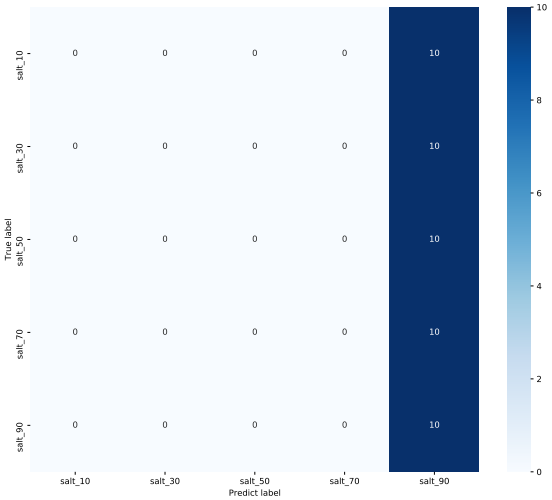



図 5.11: 組み合わせ C（$F_{rms}$, $F_{accel}$, $F_{gyro}$）を分類器への入力データとした際の内容量分類結果

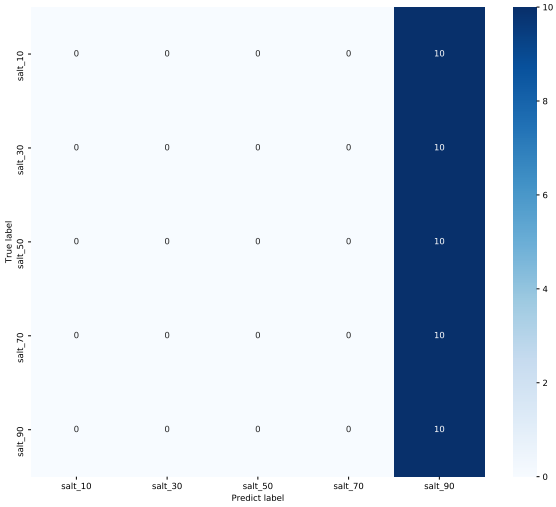



図 5.12: $F_{sound}$ のみを分類器への入力データとした内容量分類結果

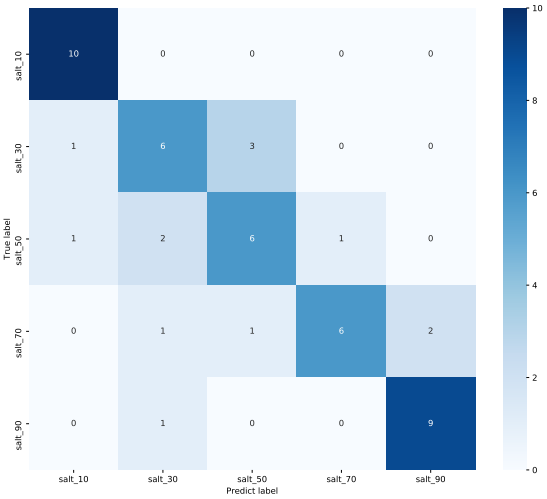



図 5.13: $F_{mel}$ のみを分類器への入力データとした内容量分類結果

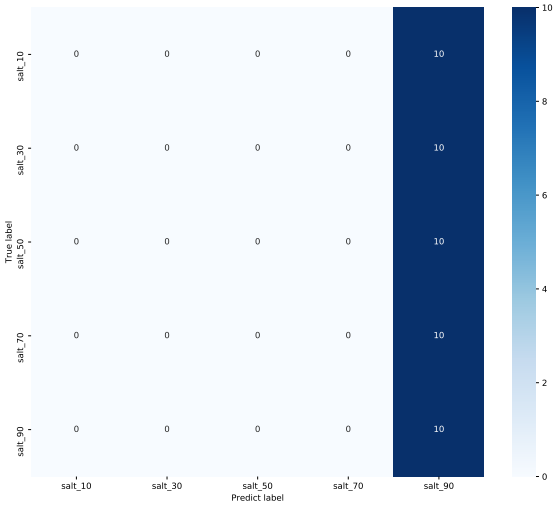



図 5.14: $F_{rms}$ のみを分類器への入力データとした内容量分類結果

# 第6章　議論と追加調査

5 章にて物体分類と内容量分類について検証した 2 つの実験について, これらの結果を元に議論する. またその内容を踏まえて追加の調査を行い, 本手法の展望について述べる.

## 6.1　音響信号の特徴量についての議論

本手法のプロトタイプシステムの評価に用いた 3 種類の音響信号の特徴量の実験結果への影響について議論する. 5 章にて行った実験では音響信号の特徴量として, FFT を分析ソフトウェアに入力する音響信号データの全サンプルに対して行った $F_{sound}$ と, $F_{sound}$ にメルフィルタによるメル周波数領域への変換と次元数の削減を行った $F_{mel}$, そしてデータに STFT を行った結果の加算平均 $F_{rms}$ を使用した. これらの特徴量について実験 1 の結果では, 加速度信号と角速度信号の特徴量（$F_{accel}, F_{gyro}$）の有無に関わらず, 分類器への入力データに $F_{mel}$ を用いた際の分類精度が $F_{sound}, F_{rms}$ を用いた場合と比較して非常に低い結果となった. $F_{mel}$ の次元数は 20 であり, $F_{sound}$（4096 次元）, $F_{rms}$（256 次元）と比べて非常に小さい. このことから特徴量ごとの次元数による情報量の差が分類精度に影響したと考えられる. 一方で実験 2 では $F_{mel}$ のみを用いた場合に分類精度の向上が見られた. このことについて, $F_{mel}$ の抽出過程で使用したフィルタバンク分析では周波数領域の低次成分の変化を微細に表現し, 高次成分の変化は緩やかに表現するように周波数スペクトルを変換する. したがって実験 2 の振る際の音については高い周波数帯において, 内容量ごとの値の差異が小さいことや振る際の環境ノイズが分布していたことなどが考えられる. また実験 2 にて組み合わせ B により検証した際に, $F_{mel}$ のみを用いた場合と比較して大きく偏って分類される結果となった. これについて実験 2 のデータ収集における手順 3（2 秒間の間に容器を 1 回振る）のような収集手法では, 内容量による振り方への影響が小さかったと考えられる. そのため $F_{accel}, F_{rms}$ は内容量ごとの値の差異が小さくなり, $F_{mel}$ と比較して次元数が大きい（各 $3 \times 64$ 次元）ために分類器の学習及び分類結果に偏りが見られたと予想される.

## 6.2　加速度信号と角速度信号を用いる意義についての議論

5 章の実験結果を踏まえて, 本手法において加速度信号と角速度信号を用いることの意義について議論する. 本稿で実装したプロトタイプシステムによる実験について, 実験 1 ではいずれのデータセットの組み合わせにおいても, 音響信号のみをデータセットに用いる場合と比較して, 実験結果への加速度信号と角速度信号の寄与は示されなかった. また実験 2 において

はデータセットの組み合わせ B の結果と $F_{mel}$ のみをデータセットに用いた場合の結果から, $F_{accel}$ と $F_{gyro}$ が分類結果に負の影響を与え得ることが確認された.

これらのことから本手法に加速度信号と角速度信号を用いることについて再検討する. 本研究の目標は生活空間内の多種多様な物について簡易な方法でその物体の認識や内容量の識別を行い生活に活用することである. その上で本研究の手法は設置する場所や使う場所が様々なものを対象としている. 例えばティッシュ箱は家庭内においてリビングや寝室など複数の部屋に常備されることが多く, また本やタブレット端末などはリビングのソファや食卓の椅子などに持ち歩いて使用される. したがって本研究の手法における物体認識については, 単一の物として生活用品や日用品などを認識するだけではなく, その設置された場所や使う状況などのコンテキストについても認識し適切に扱う必要がある. このことについて物に触れる際の手の位置や動かし方が効果的であり, 本手法でデータセットとして用いる加速度信号と角速度信号が役立つと考えられる.

## 6.3 加速度信号と角速度信号の有用性についての調査

6.2 の議論の内容を踏まえて加速度信号と角速度信号の有用性に関して追加の調査を行う. 調査には分析ソフトウェアの分類器を使用し, 5 章の実験 1 で構築したデータセットの $F_{accel}$ と $F_{gyro}$ を使用して収集時の対象物の状態（'hold', 'put'）についての簡易な分類検証により行う. 検証には対象物の 2 種類の状態についてデータセットの 8 割を訓練データとし残りの 2 割をテストデータとして用いた. その結果図 6.1 に示すように実験 1 で用いた対象物の 2 種類の状態については, 加速度信号と角速度信号から非常に高い精度で分類されることが示された. このことから本手法の物体認識において加速度信号と角速度信号は物体の置いてある場所や使う状況の推定への利用可能性が示唆される. また内容量識別においても振る以外の動作において加速度信号や角速度信号が有望であると考えられる. 例えば内容物として液体を用いる際に, かき混ぜやすさによってその量を識別する方法などが検討できる.

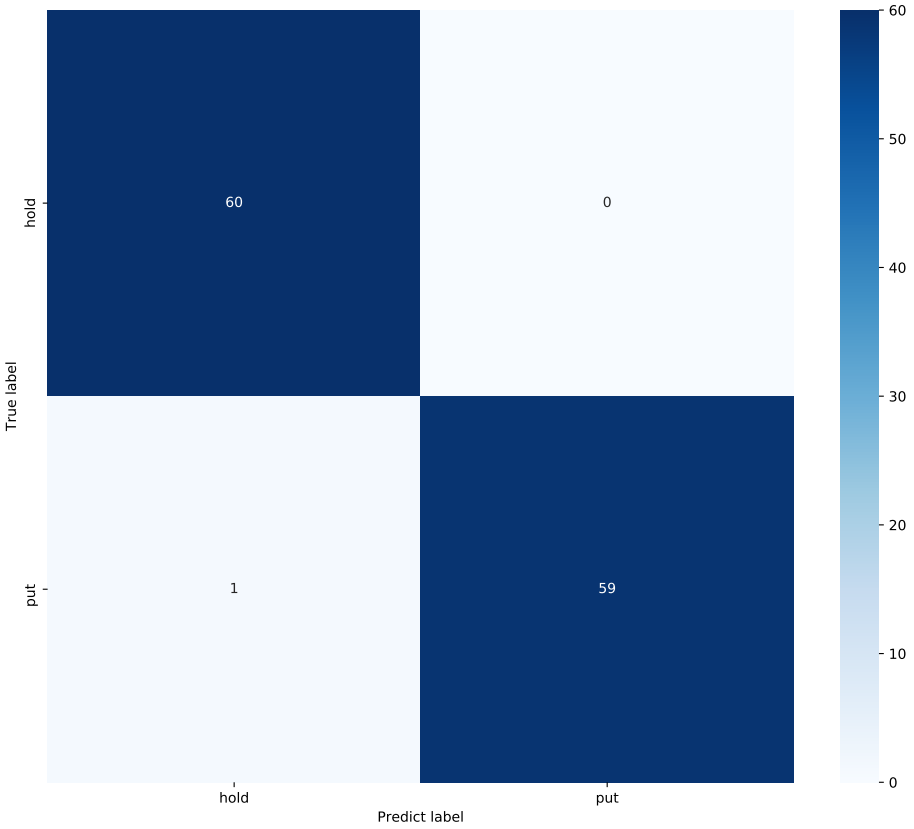



図 6.1: $F_{accel}$ と $F_{gyro}$ を用いた実験 1 の対象物の状態分類

# 第7章　まとめ

本研究では人が物に触れる際の音や手の動きを用いて物体認識や内容量識別を行う手法を提案した. また提案手法を検証するために, 手周囲の音響信号や加速度信号, 角速度信号をセンシングするデバイスと, それにより得られるセンシングデータから特徴量を抽出して, 機械学習により特徴の学習と分類を行うソフトウェアを実装したプロトタイプシステムを構築した.

評価実験ではプロトタイプシステムを用いて, 物を叩く際のデータによる 13 種類の物体分類と容器を振る際のデータによる 5 種類の内容量分類の 2 つの実験について取り組み, 分類器への入力データとして, 三種類の音響信号の特徴量（$F_{sound}, F_{mel}, F_{rms}$）ごとの分類精度の差と, 加速度信号と角速度信号の特徴量（$F_{accel}, F_{gyro}$）が実験結果に与える影響について評価した. その結果, 物体分類について音響信号の特徴量として $F_{rms}$ を用いた際の分類精度が最も高く, Leave-One-Group-Out-Cross-Validation（LOGOCV）による検証で 72% の正解率が示された. 一方で音響信号の特徴量のみを使用した場合について, 加速度信号と角速度信号の特徴量を用いることによる有意差は示されなかった. また内容量分類については $F_{mel}$ のみによる LOGOCV で 58% の正解率が示されたが, その他の特徴量をデータセットとした際についてはいずれも著しく低い分類精度が示される結果となった.

物体認識や内容量識別に向けた本論文の検証では, 加速度信号と角速度信号を用いる効果について懸念される. しかしその上で行った追加の調査では, 加速度信号と角速度信号の特徴量から対象物の置いてある場所やその使う状況といった情報を推定に有望であることが示された. したがって今後本手法の実装について改善を行った上で実際の生活空間内に用いることにより, 様々な生活用品や日用品, またその使用量などを簡単に認識し, その使用する状況に合わせて様々な用途に活用することが可能になると考えられる.

# 謝辞

本研究を進めるにあたり, 指導教員である高橋伸准教授には手法の考案及び実装, また論文の執筆について多くのご指導とご助言を賜りました. また志築文太郎准教授には情報特別演習の受講期間も含め, 長きに亘って大変お世話になりました. ここに深く御礼申し上げます.

インタラクティブプログラミング研究室の皆様には研究が難航した際にも相談に乗って様々な視点からのアドバイスをして頂き非常に感謝しております. また UBIQUITOUS チームの皆様とは合宿や課外のイベントへの参加, 日頃の様々な活動を通して交友が深まり, 3 年間楽しく過ごすことができました. 重ねて感謝いたします.

最後に大学及び大学院への進学を支援し研究生活を応援してくださった家族, そして学生生活をともに過ごした友人達, お世話になった方々に深く感謝申し上げます.

# 参考文献


[1] Kaifei Chen, Jonathan Fürst, John Kolb, Hyung-Sin Kim, Xin Jin, David E. Culler, and Randy H. Katz. 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, January 2018.

[2] Roberto Hoyle, Robert Templeman, Steven Armes, Denise Anthony, David Crandall, and Apu Kapadia. Privacy Behaviors of Lifeloggers Using Wearable Cameras. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp'14, pp. 571–582, New York, NY, USA, 2014. ACM.

[3] Yukitoshi Kashimoto, Kyoji Hata, Hirohiko Suwa, Manato Fujimoto, Yutaka Arakawa, Takeya Shigezumi, Kunihiro Komiya, Kenta Konishi, and Keiichi Yasumoto. Low-cost and Device-free Activity Recognition System with Energy Harvesting PIR and Door Sensors. In *Adjunct Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing Networking and Services*, MOBIQUITOUS'16, pp. 6–11, New York, NY, USA, 2016. ACM.

[4] Gierad Laput, Robert Xiao, and Chris Harrison. Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST'16, pp. 321–333, New York, NY, USA, 2016. ACM.

[5] Robert Xiao, Gierad Laput, Yang Zhang, and Chris Harrison. Deus em machina: On-touch contextual functionality for smart iot appliances. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 4000–4008, New York, NY, USA, 2017. Association for Computing Machinery.

[6] Takuya Maekawa, Yasue Kishino, Yasushi Sakurai, and Takayuki Suyama. Recognizing the use of portable electrical devices with hand-worn magnetic sensors. In *Proceedings of the 9th International Conference on Pervasive Computing*, Pervasive '11, pp. 276–293, Berlin, Heidelberg, 2011. Springer-Verlag.

[7] Shengjie Bi, Tao Wang, Ellen Davenport, Ronald Peterson, Ryan Halter, Jacob Sorber, and David Kotz. Toward a Wearable Sensor for Eating Detection. In *Proceedings of the 2017*

*Workshop on Wearable Systems and Applications*, WearSys'17, pp. 17–22, New York, NY, USA, 2017. ACM.

[8] Yu Zhong, Pierre J. Garrigues, and Jeffrey P. Bigham. Real time object scanning using a mobile phone and cloud-based visual search engine. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13, New York, NY, USA, 2013. Association for Computing Machinery.

[9] Kyungjun Lee and Hernisa Kacorri. Hands holding clues for object recognition in teachable machines. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, New York, NY, USA, 2019. Association for Computing Machinery.

[10] Hernisa Kacorri, Kris M. Kitani, Jeffrey P. Bigham, and Chieko Asakawa. People with visual impairment training personal object recognizers: Feasibility and challenges. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 5839–5849, New York, NY, USA, 2017. Association for Computing Machinery.

[11] Hanchuan Li, Can Ye, and Alanson P. Sample. Idsense: A human object interaction detection system based on passive uhf rfid. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 2555–2564, New York, NY, USA, 2015. Association for Computing Machinery.

[12] Jun Rekimoto and Yuji Ayatsuka. Cybercode: Designing augmented reality environments with visual tags. In *Proceedings of DARE 2000 on Designing Augmented Reality Environments*, DARE '00, pp. 1–10, New York, NY, USA, 2000. Association for Computing Machinery.

[13] Taesik Gong, Hyunsung Cho, Bowon Lee, and Sung-Ju Lee. Knocker: Vibroacoustic-based object recognition with smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 3, No. 3, pp. 82:1–82:21, September 2019.

[14] Mingming Fan and Khai N. Truong. Soqr: Sonically quantifying the content level inside containers. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp'15, pp. 3–14, New York, NY, USA, 2015. ACM.

[15] Yiran Zhao, Shuochao Yao, Shen Li, Shaohan Hu, Huajie Shao, and Tarek F. Abdelzaher. 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, September 2017.

[16] Qun Wei, Mi-Jung Kim, and Jong-Ha Lee. Development of capacitive sensor for automatically measuring tumbler water level with fea simulation. *Technol Health Care*, Vol. 26, No. S1, pp. 491–500, May 2018.

[17] 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, feb 2019.

[18] Shuhei Aoyama, Buntarou Shizuki, and Jiro Tanaka. Thumbslide: An interaction technique for smartwatches using a thumb slide movement. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pp. 2403–2409, New York, NY, USA, 2016. Association for Computing Machinery.

[19] Robert Xiao, Teng Cao, Ning Guo, Jun Zhuo, Yang Zhang, and Chris Harrison. Lumiwatch: On-arm projected graphics and touch input. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, New York, NY, USA, 2018. Association for Computing Machinery.

[20] Y. Suzuki, K. Sekimori, B. Shizuki, and S. Takahashi. Touch sensing on the forearm using the electrical impedance method. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 255–260, March 2019.

[21] Vincent Becker, Linus Fessler, and Gábor Sörös. Gestear: Combining audio and motion sensing for gesture recognition on smartwatches. In *Proceedings of the 23rd International Symposium on Wearable Computers*, ISWC '19, pp. 10–19, New York, NY, USA, 2019. Association for Computing Machinery.

[22] Lei Shi, Maryam Ashoori, Yunfeng Zhang, and Shiri Azenkot. Knock knock, what 's there: Converting passive objects into customizable smart controllers. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '18, New York, NY, USA, 2018. Association for Computing Machinery.

[23] Zhoutong Zhang, Qiujia Li, Zhengjia Huang, Jiajun Wu, Josh Tenenbaum, and Bill Freeman. Shape and material from sound. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pp. 1278–1288. Curran Associates, Inc., 2017.

[24] M5Stack A series of modular stackable development devices. M5StickC, 2019. `https://docs.m5stack.com/#/en/core/m5stickc`.

[25] 河原達也. 音声認識システム 改訂 2 版. オーム社, 2016.