

# 音響と振動のセンシングによる 充填率の異なる3Dプリンタ生成物の識別手法

八田 将志<sup>1</sup> 村尾 和哉<sup>1,2</sup>

**概要:** スマートフォンは多様なアプリケーションと機能で私たちの生活を豊かにしてきた。しかし、スマートフォンは手間のかかる作業や工程が多いという問題を抱えている。この問題は、ユーザが頻繁に使用するアプリケーションを短い時間で起動させることができれば手間のかかる作業や工程が短縮される。近年では音声認識技術やQRコード、バーコードなどの認識技術により、アプリケーションを短い時間で起動させることが可能となった。しかし、これらの技術は認識精度の低さや外的要因によりアプリケーション起動までにより時間がかかってしまう可能性がある。そこで本研究ではこれらの問題点を解決するために、音響と振動の特性が異なる3Dプリンタ生成物をスマートフォンで叩き、その特性の違いから3Dプリンタ生成物を認識し、その3Dプリンタ生成物に対応したアプリケーションを起動させるシステムを提案する。まず音響と振動の特性が異なる3Dプリンタ生成物を11種類用意し、スマートフォンで100回ずつ合計1,100回叩くことで、用意した3Dプリンタ生成物を分類した。結果として、全体では約81%、最大で96%の精度を得ることができた。

## 1. はじめに

スマートフォンは多様なアプリケーションと機能で我々の生活を豊かにしてきた。しかし、インストールされたアプリケーションが多く、目的のアプリケーションや機能を見つけることが困難であることや、目的の機能を実行するまでにタッチやスクロールなどの操作が多い、端末が大きく画面端に指が届かず操作しづらい、という問題点も挙げられる。具体的な例として、料理中にスマートフォンに搭載されているストップウォッチのアプリケーションを使用して時間を測りたい場合、まずスマートフォンの電源ボタンを押して画面を点灯させ、パスワードなどの認証を行ってホーム画面を表示し、ストップウォッチのアプリケーションを見つけてタッチして起動し、時間を入力して、計測を開始する必要がある。この数年で普及しつつあるPayPay[18]や楽天ペイ[19]などのスマートフォン決済においても、会計時にアプリを起動してバーコードを表示する煩雑さが問題点である。

音声認識技術の発展により、SiriやAlexaなどのAIアシスタントと呼ばれる発話解析・認識インタフェースを用いて、声によって迅速かつ容易にアプリケーションを起動できるようになった。先述の料理中のタイマ設定も、発話

解析・認識インタフェースを用いればハンズフリーで行うことができる。しかし、発話解析・認識インタフェースは、行いたい操作を言葉でどのように表現してよいのかわからないという自然言語による表現の困難さの問題がある。また、発話の認識精度の低さや、第三者の発話による誤操作、発話による情報漏洩も問題視されている。

QRコードやバーコードを用いて事前に定義されたアプリの起動や処理を行うアプローチもある。これらのコードはあらかじめ決められた情報を容易に、早く、正確に読み取るために開発されたデータ表現であり、現在では物流、製造、行政、医療、イベントなどのあらゆる分野で利用されている。しかし、コードと処理は一对一对応であり、タイマ設定のように時間の数値を設定するには数値ごとにコードが必要である。また、カメラでコードを撮影する必要があり煩雑さは残る。

キッチンであればタイマ、寝室であれば目覚まし時計、商店のレジであればスマホ決済アプリなど、場所ごとに利用するアプリケーションが決まっている場面があり、容易な操作でそれらを実行できることが望ましい。本研究では、スマートフォンで物体を叩いたときに発生する音響と振動に着目し、音響と振動の特性から物体を識別し、その物体に対応したアプリケーションをスマートフォンで起動させる手法の実現を目標とする。そのために、充填率の異なる3Dプリンタ生成物をスマートフォンで叩いたときに

<sup>1</sup> 立命館大学大学院情報理工学研究科

<sup>2</sup> 国立研究開発法人科学技術振興機構さきがけ

発生する音響と振動から、3D プリンタ生成物の充填率を識別する手法を提案する。本稿では3D プリンタ生成物は立方体とした。これにより、スマートフォンで3D プリンタ生成物を1回あるいは数回叩くだけで目的のアプリケーションを起動できる。タイム設定の時間なども3D プリンタ生成物を任意の回数叩けば設定できる。

以降、2節では関連研究を紹介する。3節では提案手法を説明し、4節では評価実験について述べ、最後に5節で本研究をまとめる。

## 2. 関連研究

本節では音を使用する物体認識に関する研究、音以外を用いた物体認識に関する研究および3D プリンタを用いた研究を紹介する。

### 2.1 音を使用した物体認識

物体認識に関する研究は多くされている。そのひとつとして、音響センシングを使用する物体認識の手法がある。岩瀬ら [1] はアクティブ音響センシングを用いてアクリルプレート上に置かれた物体の認識を行っている。アクティブ音響センシングとはスピーカから特定の音響信号を認識対象物に流し、同時にマイクで音響信号を受信し、両者の違いから物体の種類や状態を認識する手法である。アクリルプレートに対してスイープ信号と呼ばれる時間的に周波数が増える音響信号を与え、アクリルプレート上に置かれた物体によって変化する応答の周波数特性を解析する。これにより、配置物体の位置や種類を分類することができる。

中村ら [2] は超音波センサを用いて路上の障害物情報の検知を行っている。超音波センサ、GPS、地磁気センサから路上の障害物の位置を検知し、その情報を他者と共有することで危険検知の精度を向上させている。また、渡辺ら [3] は超音波とニューラルネットワークを用いて物体の認識を行っている。超音波を用いて非接触で物体を認識する方法では、スマートフォンで超音波を再生して物体に当てる必要があり、技術的には新しいがQRコードやバーコードと同様の煩雑さがある。

また、Gieradら [4] はスマートウォッチに搭載されている加速度データと生体音響データを使用して手のジェスチャを分類し、その情報から把持されている機械的に振動したオブジェクトを認識する手法を行っている。この研究では物体が振動していることが条件となっている。

Chrisらが提案したTapSense[13]は、スクリーン上でのタッチ入力に対して使用された指や爪などのオブジェクトの種類や位置を推定し、従来の画面を用いて指の入力方法を識別することができるタッチ操作の機能強化を行った。これは、指の衝撃から生じる音をセグメント化することで指の入力方法を分類している。Onoら [6] は、振動スピーカ

とマイクをセンサとして組み合わせて、既存の物体に対するさまざまなタッチジェスチャの認識を行っている。単純なタッチジェスチャの認識率は高かったが、手の姿勢の認識精度は低かった。Shanら [14] が提案したKnock-Knockは、テストセットのオブジェクトをマーカペンでノックすることで音データを取得し、ディープラーニングを使用して学習した。これにより、従来の分類器よりも認識精度が高くなることが確認できた。

Leiら [15] は、日常的なオブジェクトを使用して、そのオブジェクトをノックして発生する音を認識することで日常の環境を制御するBeatItを提案した。例えば、ドアをノックする音をライトを点灯するイベントと関連付けるなどを行うことで日常の環境を制御している。Gongら [10], [12] が提案したKnockerは、スマートフォンに搭載された加速度センサ、角速度センサ、マイクを用いてスマートフォンで叩いた物体を認識する手法を行っている。この研究では、市販のスマートフォンで物体認識を行うことができ、ペットボトルや箱など異なる物体の認識を行っている。本研究では形状が同一で充填率が異なる物体を生成し、それらを認識することを試みる。

### 2.2 音以外を使用した物体認識

音以外の情報を用いる物体認識に関する研究を紹介する。Maekawaら [5] は、磁気センサ搭載した手袋を用いて電子機器を持ったときに、電気機器から発生される磁場を感知して、電気機器の種類を認識する手法を提案している。磁力の異なる磁石や、モーターを用いて回転速度の異なる磁石などを埋め込むことで異なる磁界を発生させる物体を生成できれば、スマートフォンを近づけるだけで物体を識別できるが、実現は容易ではない。また、複数の物体を近接して使用すると、互いの磁場が影響することも考えられる。

唐津ら [8] のORGAは、物体に取り付けた加速度センサとユーザが身に着けたヘッドデバイスに搭載されたカメラで撮影した画像を取得し、物体を認識する。カメラで撮影した画像を解析し、移動物体を検出できた場合、同じ時間に加速度が大きく変動しているセンサを物体に取り付けられた加速度センサから検索する。これにより、あらかじめセンサに紐付けられた情報を取得することで物体の認識を行っている。Chrisら [7] は、バーコード状の溝が彫られた物体をコインなどでなぞることで溝のパターンに応じた音が発生し、その音を解析することで何の物体をなぞったか認識する手法を提案している。カメラでバーコードを読む必要はないが、正しく物体をなぞる必要がある。

### 2.3 3D プリンタを用いた研究

オウら [16] らは、Kinectを用いてキャプチャした3Dプリンタ生成物の深度マップを処理することで、その姿勢、位置、向きを認識した。従来のオブジェクト認識はほとん

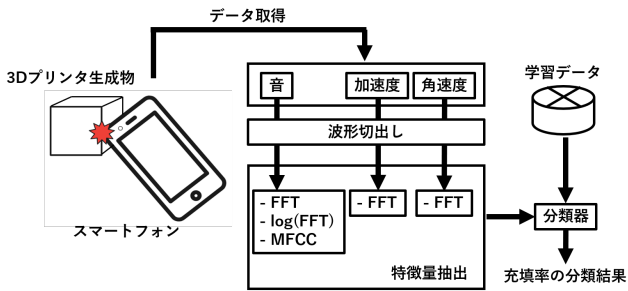


図 1 システム構成

どが画像認識に基づいており、3D プリンタ生成物とその姿勢などを認識する点では画像認識は役に立たないが、この手法であれば認識をすることができる。Katou ら [17] らは、3D プリンタ生成物の表面に導電性の点群を配置し、それを介することでタッチ入力が発生させる手法を提案している。タッチパネルなどの導電性素材を介したタッチ入力は、それらが接触している特定の箇所においてタッチ入力が発生させるか、スクロールのような直線動作のタッチ入力を発生させることしかできなかったが、この手法を用いることで手書き文字入力やポインティングなどの自由な 2 次元のタッチ操作が行える。Kubo ら [9] が提案した FabAuth は、アクティブ音響センシングを用いて共振特性の違いから 3D プリンタ生成物の充填率の識別を行っており、約 92%の精度で認識できている。このようなアクティブ音響センシングを用いる手法では物体に一定時間触れ続けられないといけない点や、専用のマイクとスピーカが必要である点が欠点として挙げられる。

### 3. 提案手法

本節では提案手法のシステム構成、センサデータの取得、センサデータの切出し、特徴量抽出、物体判定について説明する。

#### 3.1 システム構成

充填率の異なる 3D プリンタ生成物（以下、オブジェクト）は、叩いたときの音響と振動の特性が異なる。本研究では、充填率の異なる同一形状のオブジェクトを複数種類作成し、作成したオブジェクトをスマートフォンで叩くことによって、スマートフォンに内蔵されたセンサから得られた情報をもとに叩いたオブジェクトの分類を行う。

提案システムの構成を図 1 に示す。提案手法は、スマートフォンでオブジェクトを叩いた際の音データ、加速度データ、角速度データをスマートフォン内蔵のセンサで取得する。次に、オブジェクトを叩いた区間のセンサデータの切出しを行う。切り出した区間の音、加速度、角速度のそれぞれから特徴量を抽出する。最後に、事前に充填率の異なるオブジェクトごとの特徴量を学習した分類器を用いて、叩いたオブジェクトの分類を行う。

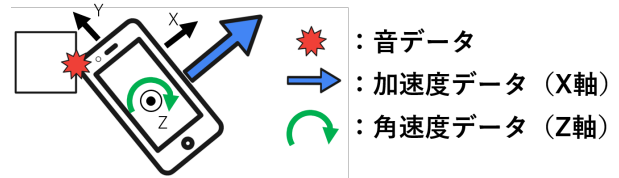


図 2 取得するセンサデータ

#### 3.2 センサデータの取得

はじめに、センサデータの取得を行う。取得するセンサデータは音データ、加速度データ、角速度データである。本研究では図 2 に示すように、スマートフォンを Z 軸を軸として回転する向きでオブジェクトを叩くと想定する。加速度センサは X 軸、Y 軸、Z 軸方向の 3 次元のデータを取得できるが、提案手法はオブジェクトを叩く際に最も影響の出る X 軸のデータのみを使用する。サンプリング周波数は 400Hz とする。角速度センサも同様に X 軸、Y 軸、Z 軸を回転軸とした 3 次元のデータを取得できるが、提案手法はオブジェクトを叩く際に最も影響の出る Z 軸のデータを使用する。サンプリング周波数は 400Hz とする。音データはスマートフォンのマイクから取得し、サンプリング周波数は 44,100Hz とする。これらのサンプリング周波数は一般的なスマートフォンで取得できる値であり、特別な機種や設定は必要としない。

#### 3.3 波形の切出し

波形の切出し方法について述べる。切り出す波形は、スマートフォンでオブジェクトを叩く動作によって発生する音、加速度、角速度の波形である。他人のスマートフォンやスマートフォン以外でオブジェクトを叩いたときに発生する音を自身のスマートフォンのマイクが拾って誤動作したり、端末をポケットに入れて走ったときの振動で誤動作することがないように、自身のスマートフォンでオブジェクトを叩いたときのみの波形を切り出す必要がある。

ここで、スマートフォンでオブジェクトを叩いたときの音データ、加速度データ、角速度データを同一時系列上にプロットしたものを図 3 に示す。図の区間ではオブジェクトが 2 回叩かれている。図中の加速度波形の赤い点は加速度のピークであり、この瞬間にオブジェクトが叩かれている。破線は加速度のピークの時刻を示しており、その時刻における角速度波形との交点も赤い点で示している。

一方、音データのピークは音データの波形の赤い点で示されている箇所であるが、スマートフォンの音声データの内部処理の仕様上、音がセンシングされてから記録されるまでには遅延があり、実際の時刻とタイムスタンプの間には 20~150[ms] のずれがあることが知られている [10]。図 3 の波形では、加速度のピークと音のピークにおよそ 100ms の遅延が生じている。提案手法は加速度データに閾値以上のピークが発生した瞬間から 200ms 以内に音データに閾値

以上のピークが発生した場合にスマートフォンでオブジェクトを叩いたと判定する。

ピークの検出には SciPy[20] の `signal.argrelmax` を用いた。 `argrelmax` は `order` というパラメータを設定することで、任意区間をスライドさせながら最大値をとるインデクスを取得でき、これによりピークを検出できる。本研究では閾値を  $1.0[G]$  に、 `order` を  $1000$  に設定して、加速度データのピーク検出を行った。 `order` が  $1000$  とは、ある時刻のサンプルの左右それぞれの側  $1000$  サンプルの区間、つまり  $2001$  サンプルの区間での最大値を探索することを意味する。加速度データのサンプリング周波数は  $400\text{Hz}$  であるため、  $5$  秒の区間に相当する。本研究の評価では連続してオブジェクトを叩くことは想定していないので `order` を  $1000$  としたが、 `order` を小さくすることで連続して叩く場合にも対応できる。検出されたピークに対して、オブジェクトを叩いた判定を行う。これにより、スマートフォンを手を持っている場合などの小さな加速度のピークが発生した場合を除去できる。

音データのピーク検出も同様に `argrelmax` を用いて、 `order` を  $10000$  に設定した。音データの閾値は  $0.1$  で、加速度のピークが発生した時刻から  $2$  秒間の間の波形を切り出し、ピーク検出を行った。これにより、スマートフォンをポケットに入れて走るなど加速度データに閾値以上のピークが発生した場合でも大きな音が発生していなければオブジェクトを叩いたと誤検出することを防ぐことができる。

なお、角速度データにはスマートフォンで物体を叩く瞬間以前からスマートフォンの回転による変化が表れている。叩く前の変化はオブジェクトの充填率とは無関係であり、加速度のピークの時刻以降に発生した角速度データの変化に物体固有の情報が含まれていると考えられるため、本研究では角速度データは切り出し開始点は加速度のピークと同じ時刻とした。加速度データの切り出し開始点は加速度のピークの時刻とした。

### 3.4 特徴量抽出

音、加速度、角速度データから特徴量を抽出する方法を説明する。

#### 3.4.1 音データ

音データから周波数パワースペクトル、対数周波数パワースペクトル、メル周波数ケプストラム (MFCC) の  $3$  種類の特徴量を抽出する。いずれの特徴量も検出した切り出し開始点から  $4096$  サンプルのウィンドウから抽出する。パワースペクトルはデータを高速フーリエ変換して、各周波数の成分に対して絶対値を取ったものである。高速フーリエ変換は離散フーリエ変換を計算機で高速に計算するためのアルゴリズムであり、離散フーリエ変換は波形の周波数特性の解析に使われるアルゴリズムである。高速フーリエ変換の式を次式に示す。  $x(n)$  は入力される音データで

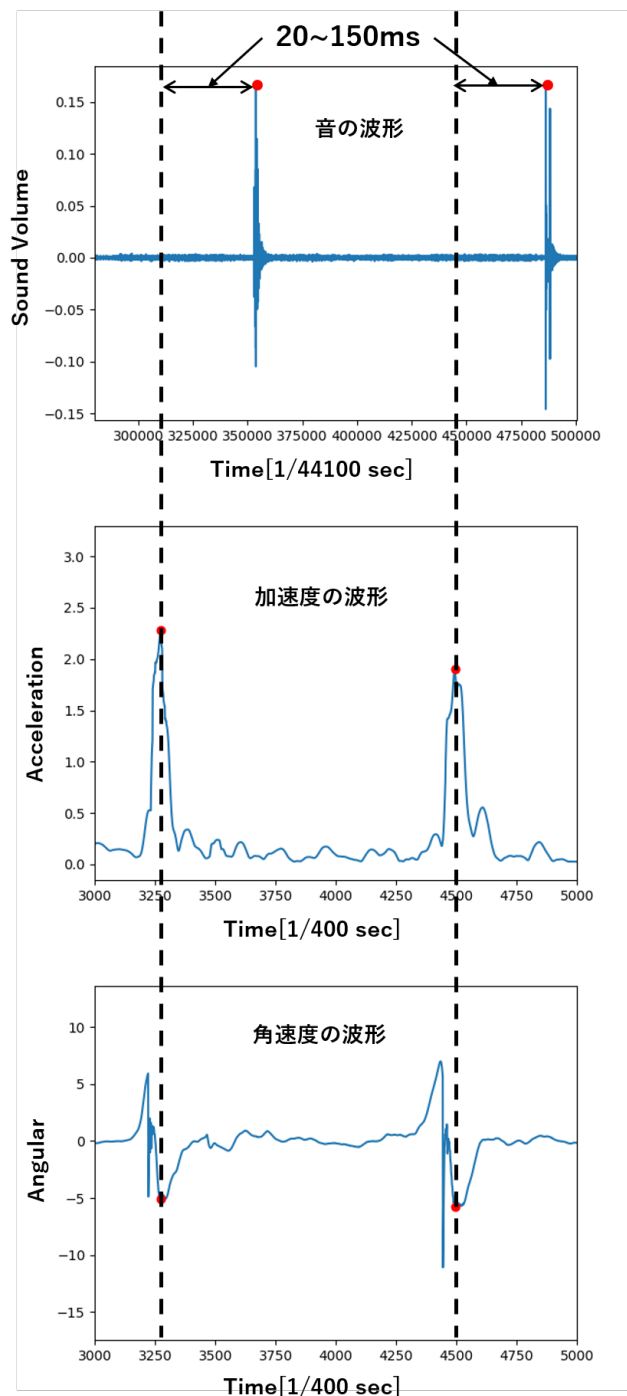


図 3 スマートフォンでオブジェクトを叩いたときの音データ、加速度データ、角速度データ

あり、  $N$  は音データの長さ (本研究では  $4096$ ) である。

$$X(k) = \sum_{n=0}^{k-1} w(n)x(n) \exp\left(-\frac{j2\pi kn}{N}\right) \quad (1)$$

通常、フーリエ変換は音声区間全体に対して行うのではなく、窓関数  $w(n)$  を用いることでその音声データを周期関数とみなして分析を行う。その中でも本研究では次式に示すハニング窓を使用した。

$$w(n) = 0.5 - 0.5 \cos \frac{\pi n}{N} \quad (2)$$

ハニング窓の最大値は1, 最小値は0であり, 取得した音データにこの関数を掛けることで切り出した区間を周期区間として扱う. 音データ  $x(n)$  に窓関数  $w(n)$  を掛けて離散フーリエ変換を行う. これにより, 取得した音データの周波数成分と振幅成分を抽出する. 高速フーリエ変換後の波形は中央を軸として左右対称となるため,  $k=1$  から 2048 までの 2048 次元を特徴量として使用する.

対数パワースペクトルは, 次式に従い, 高速フーリエ変換を行って得られた 4096 個のデータサンプルの対数を取ることで  $\log$  スケールに変換した 4096 個のデータサンプルを使用する.

$$X'(k) = \log(|X(k)|) \quad (3)$$

この値もパワースペクトルと同様に中央を軸として左右対称となるため,  $k=1$  から 2048 までの 2048 次元を特徴量として使用する.  $\log$  スケールは, 極端に大きいもしくは小さいデータを扱いやすくすることができる効果がある.

メル周波数ケプストラム (MFCC) は主に音声認識に用いられる特徴量である. ケプストラムとは対数パワースペクトルを離散コサイン変換することで得られる. ケプストラムは音の低周波数成分を引き延ばし, 高周波数成分を圧縮する特徴があり, ピッチ成分を除去し, 音の識別にとって重要な音響特性のみを抽出することができる.

はじめに, 次式に従い音データの周波数スケール  $f$  をメルスケール  $p$  へ変換する. メルスケールとは, 人間の感覚的な音の尺度であり, メルスケール上の値の間隔が同じであれば, 人間が感じる音の高低差も同じであると考えられている. 音データの周波数をメルスケールへ変換する変換式はピッチと周波数の関係として次式が Volkman と Stevens[11] によって一般的なものとして提示されているため, これを利用する.

$$p = 1127.010480 \log\left(\frac{f}{700} + 1\right) \quad (4)$$

式 4 を  $f$  について解くと, メルスケールの値  $p$  を周波数に逆変換する式が求められる.

$$f = 700(10^{\frac{p}{1127.010480}} - 1) \quad (5)$$

次に, メルフィルタバンクを作成する. 取得した音データのサンプリング周波数を  $f_s$  とすると, ウィンドウの長さが  $N$  であれば, 周波数のスケールは  $f(k) = \frac{k}{N}f_s$  となる. フィルタバンクは特定の範囲の値のみを有効なものとして扱う関数で, メルフィルタバンクとはメルスケールを等間隔に刻んだフィルタバンクである.  $M$  個のフィルタバンクの中心周波数  $f_c(m), m=1, 2, \dots, M$  を計算する. 分析対象の周波数帯について, その最小値  $f_{min}$  と最大値  $f_{max}$  をメルスケールに変換したものをそれぞれ  $p_{min}, p_{max}$  とする.  $\Delta p$  を次式で定義する

$$\Delta p = \frac{p_{max} - p_{min}}{M} \quad (6)$$

また,  $p_c$  を次式で定義する.

$$p_c(m) = m\Delta p \quad (7)$$

これにより,  $p_c(1), p_c(2), \dots, p_c(M)$  はメルスケール上で等間隔となる. これを式 5 に従って周波数に変換したものが中心周波数  $f_c$  となる. そして,  $f_c$  を用いてメタフィルタバンクの式を作成する.

$$H(k, m) = \begin{cases} 0 & (f(k) < (f_c(m-1))) \\ \frac{f(k)-f_c(m-1)}{f_c(m)-f_c(m-1)} & (f_c(m-1) \leq f(k) < f_c(m)) \\ \frac{f_c(m+1)-f(k)}{f_c(m+1)-f_c(m)} & (f_c(m) \leq f(k) < f_c(m+1)) \\ 0 & (f_c(m+1) < f(k)) \end{cases} \quad (8)$$

最後に, MFCC を計算する. 次式に従い,  $X(n)$  にメタフィルタバンクを掛けた値を対数に取る.

$$X'(m) = \log\left(\sum_{k=0}^{N-1} |X(k)|H(k, m)\right) \quad (9)$$

$X'(m)$  を次式に従い離散コサイン変換したものが MFCC となる.

$$c(i) = \sum_{m=1}^M X'(m) \cos \frac{i}{M}\left(m - \frac{1}{2}\right)\pi \quad (10)$$

ただし,  $i=1, 2, \dots, M$  である.

このようにして音データから得た MFCC の値のうち, データの特徴を最も顕著に表現できる先頭から 100 個 ( $i=1, 2, \dots, 100$ ) のデータの特徴量として使用する.

### 3.4.2 加速度データ・角速度データ

加速度データと角速度データから周波数パワースペクトルを特徴量として抽出する. 加速度のピークを開始点として 32 サンプルのデータを取り出し, 224 サンプルをゼロパディングして, 256 サンプルのデータを作成する. その後, 256 サンプルのデータに対して高速フーリエ変換を行い, 周波数パワースペクトルを得る. 角速度に対しても同様の手順で周波数パワースペクトルを得る. ゼロパディングを行うと, 高速フーリエ変換を行った際の周波数パワースペクトルの分解能を高くすることができ, オブジェクトと叩いたときに発生する高い周波数の振動を取得できる.

具体的には, 振幅の推定において, FFT ビンに対応しない周波数の正弦波の振幅を推定すると, 不正確な推定となる. FFT ビンとは, そのビンに収まる正弦波成分またはほかの十分に狭い帯域の信号を分離するためのフィルタである. ゼロパディングを行うことで FFT ビンに対応した周波数の振幅が正確になり, オブジェクト推定精度が向上する. ここで, 角速度データに対してゼロパディングを行ったときと行っていないときの周波数パワースペクトルを図 4 に示す. 図より, ゼロパディングを行わない場合

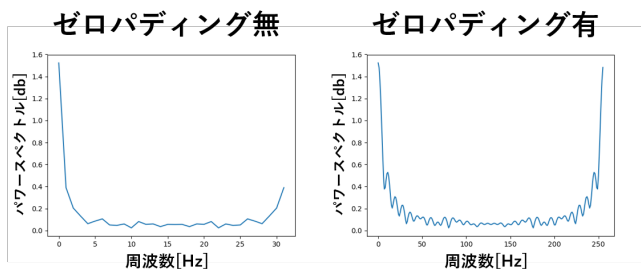


図 4 ゼロパディングの有無による周波数パワースペクトラム

は低周波数帯のパワースペクトルがほぼ一定で周波数による差異が無いように見えるが、ゼロパディングを行った場合は同じ周波数帯でもパワースペクトルの分解能が細かく表現されることで差異が明確になっていることがわかる。周波数スペクトルは中央を軸として左右対称であるため、音データと同様に加速度データと角速度データの周波数パワースペクトルの左半分の 128 個を使用する。このようにして抽出した周波数パワースペクトルを加速度データ、角速度データそれぞれの特徴量とした。

### 3.5 分類

得られた音の周波数パワースペクトル 2048 次元, 対数周波数パワースペクトル 2048 次元, MFCC100 次元, 加速度の周波数パワースペクトル 128 次元, 角速度の周波数パワースペクトル 128 次元の合計 4452 次元を入力とし, Support Vector Machine (SVM) で分類する。SVM は事前に充填率の異なる複数のオブジェクトを叩いて収集したデータを用いて学習モデルを構築しておき, 未知のオブジェクトを叩いたときのデータを入力して, 充填率の分類結果を出力する。

## 4. 評価実験

本節では, 提案手法の有効性を確認するために充填率の異なるオブジェクトを用いてオブジェクト分類精度を評価する実験を行った。

### 4.1 実験環境

3D プリンタで充填率の異なるオブジェクトを生成した。3D プリンタは Flashforge 社製 Creator Pro を用いて,  $60 \times 60 \times 60$  [mm] の立方体を充填率 0% から 100% まで 10% 刻みで変化させて 11 種類を各 1 個ずつ合計 11 個作成した。充填率 0%, 50%, 100% のオブジェクトの断面の写真を図 5 に示す。

作成したオブジェクトをスマートフォンで叩き, データを取得した。叩き方はユーザによって異なるため, さまざまな方向から実験者 1 名が各オブジェクトを 100 回ずつ叩き, 合計 1,100 サンプルのデータを取得した。スマートフォンでオブジェクトを叩く間隔は 300ms 以上空

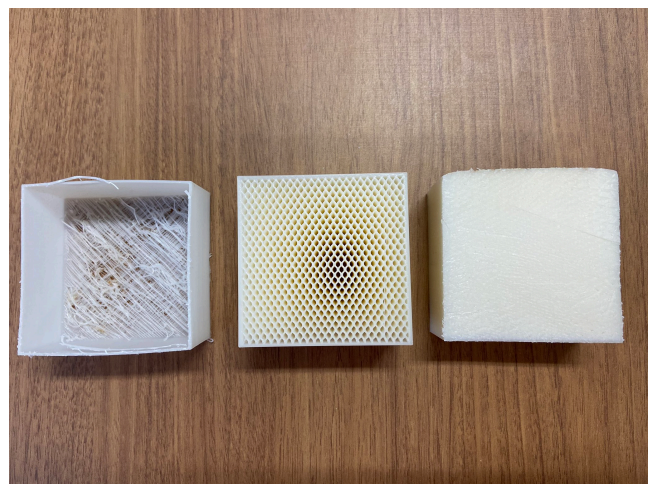


図 5 オブジェクトの断面図 (左から充填率 0%, 50%, 100%)

けて, 直前の叩く音や振動は次の叩くデータに含まれないようにした。使用したスマートフォンは HTC 社製 HTC U12+ で, センサデータを取得するアプリケーションとして Hasc Logger[21] を用いた。加速度データと角速度データは 400Hz, 音データは 44,100Hz でサンプリングした。本実験では, 分類可能な充填率の粒度を調査するために, 充填率を 11 種類 (すべて), 6 種類 (0%, 20%, 40%, 60%, 80%, 100%), 4 種類 (10%, 40%, 70%, 100%), 3 種類 (0%, 50%, 100%), 2 種類 (0%, 100%), の 5 つのグループを作成し, 各グループ内での分類を行った。

オブジェクトの分類にはデータマイニングツールである Weka[22] に搭載されている分類器 Support Vector Machine (SVM) を用いた。SVM は教師あり学習を用いるパターン認識モデルのひとつである。

各充填率 100 サンプルに対して 10 分割交差検証を行った。具体的には, 1 つのオブジェクトにつき 100 サンプルを 10 サンプルずつ 10 セットに分割し, 9 セット 90 サンプルで学習モデルを作成し, 残り 1 セット 10 サンプルでテストを行う。テストに使用するセットをローテーションしてすべてのセットでテストし, 10 回の結果の平均をとる。

### 4.2 結果と考察

11 種類, 6 種類, 4 種類, 3 種類, 2 種類のオブジェクト分類結果の混同行列, 再現率, 適合率, F 値をそれぞれ表 1, 表 2, 表 3, 表 4, 表 5 に示す。また, 各グループ分けにおける分類結果の平均 F 値を表 6 に示す。

表の見方について説明する。行が分類結果, 列が正解ラベルを示しており, 各セルには各正解ラベルに対する分類結果のサンプル数が示されている。行の値を合計すると 100 となる。左上から右下にかけての対角成分が正しく分類された結果を示している。適合率 (Precision), 再現率 (Recall), F 値 (F-measure) は以下の式で得られる。いずれも 1 に近いほど良い性能を示す。適合率は, ラベル  $L$  の

入力サンプルに対する、正しく  $L$  に分類されたサンプルの割合を示す。再現率は、ラベル  $L$  の出力サンプルに対する、正しく  $L$  に分類されたサンプルの割合を示す。F 値は再現率と適合率の調和平均であり、両者を合わせた性能を反映するため、一般的には F 値によって性能を評価する。

$$\text{適合率}(L) = \frac{\text{正しく } L \text{ に分類されたサンプル数}}{\text{ラベルの } L \text{ 入力サンプル数}} \quad (11)$$

$$\text{再現率}(L) = \frac{\text{正しく } L \text{ に分類されたサンプル数}}{\text{ラベル } L \text{ の出力サンプル数}} \quad (12)$$

$$\text{F 値}(L) = \frac{2 \times \text{適合率}(L) \times \text{再現率}(L)}{\text{適合率}(L) + \text{再現率}(L)} \quad (13)$$

結果から、11 種類、6 種類、4 種類、3 種類、2 種類のときの平均 F 値はそれぞれ 0.808, 0.866, 0.945, 0.947, 0.960 となり、分類する充填率の粒度が細くなるほど精度は低下するが、4 種類の F 値は 0.945 と高い値である。提案手法は音響と振動の特性からオブジェクトを分類しているため、充填率の差が小さいオブジェクトどうしてもは誤認識が生まれやすく、充填率の差が大きいオブジェクトどうしてもは誤認識が生まれにくいことがわかった。

次に、分類に影響を与えた特徴量について考察する。本実験で使用した特徴量は 5 種類であり、その中で音データを使用した特徴量は周波数パワースペクトル、周波数パワースペクトルの log スケール、MFCC の 3 種類、加速度データは周波数パワースペクトラムの 1 種類、角速度データは周波数パワースペクトラムの 1 種類である。11 種類のオブジェクトに対して、5 種類の特徴量を単独で用いた際の分類結果の混同行列を表 7, 表 8, 表 9, 表 10, 表 11 にそれぞれ示す。また、音の 3 種類の特徴量を併せて使用した際の分類結果の混同行列を表 12 に示し、加速度と角速度の 2 種類の特徴量を併せて使用した際の分類結果の混同行列を表 13 に示す。また、それぞれの特徴量の組合せにおける平均 F 値を表 14 に示す。

これらの結果から、音の特徴量 3 種類を使用した際の F 値は 0.789 であり、音が効果的に働いたといえるが、加速度、角速度の特徴量を追加すると F 値はわずかに上昇し 0.808 となった。音データに関しては、ひとつの特徴量のみで分類した際は、充填率の差が小さいオブジェクトどうしてもは誤認識が生まれやすく、充填率の差が大きいオブジェクトどうしてもは誤認識が生まれにくかった。また、本実験では叩き方をランダムにして実験を行ったため、加速度、角速度データでの分類結果の精度は低かった。この結果から、さまざまな叩き方による多くのデータを学習する必要があるといえる。

## 5. おわりに

本研究では、充填率の異なる 3D プリンタ生成物をスマートフォンで叩いてスマートフォン内蔵のマイク、加速度センサ、角速度センサを用いて 3D プリンタ生成物の充填率を分類する手法を提案した。実験によりスマートフォンでオブジェクト叩いたときの音データ、加速度データ、角速度データから充填率の異なる 11 種類のオブジェクトに対して、11 種類では F 値 0.808, 2 種類では 0.96 の精度で分類でき、提案手法の有効性を確認した。実験では各オブジェクトに対して 100 回叩いたデータを採取したが、今後はデータ数を増やして詳細に評価する。具体的に述べると、本実験では 1 種類のスマートフォン端末のみで実験を行ったため、異なるスマートフォン端末でも同様にオブジェクトを分類できるかどうかについての検証や、スマートフォンケースを装着した状態での検証、さまざまな叩き方での検証を行う。また、本提案システムをスマートフォンに実装し、実際にオブジェクトを叩いてスマートフォンアプリの起動し、インターフェースとしての使いやすさの評価を行う。さらに、日常生活環境での使用について、ノイズや人の会話、対象のオブジェクト以外にスマートフォンをぶつけることによる精度を評価する。

## 参考文献

- [1] 岩瀬大輝, 伊藤雄一, 秦 秀彦, 山下真由, 尾上孝雄, アクティブ音響センシングによる日常物体識別と位置推定, 情報処理学会インタラクショナル 2018, pp. 62–71 (2018).
- [2] 中村一輝, 中村嘉隆, 高橋 修, 超音波による路上障害物形状の認識手法の提案, マルチメディア, 分散, 協調とモバイルシンポジウム (DICOMO2015), pp. 149–156 (2015).
- [3] 渡辺澄夫, 米山正秀, 超音波映像法とニューラルネットワークを用いた 3 次元物体認識法, 日本音響学会誌, No. 47, Vol. 11, pp. 825–833 (1991).
- [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 (2016).
- [5] Takuya Maekawa, Yasue Kishino, Yasushi Sakurai, and Takayuki Suyama, Recognizing the Use of Portable Electrical Devices with Hand-Worn Magnetic Sensors, In Proceedings of International Conference on Pervasive Computing (PERVASIVE 2011), pp. 276–293 (2011).
- [6] Makoto Ono, Buntarou Shizuki, and Jiro Tanaka. 2013. Touch Activate: Adding Interactivity to Existing Objects using Active Acoustic Sensing. In Proc. of UIST '13. ACM, 31–40.
- [7] Chris Harrison, Robert Xiao, and Scott Hudson, Acoustic barcodes: passive, durable and inexpensive notched identification tags, In Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST 2012), pp. 563–568 (2012).
- [8] Yutaka Karatsu, Jin Nakazawa, Kazunori Takashio, Hideyuki Tokuda, ORGA : Visualize Sensor Data Information by Augmented Reality without Visual Marker, IEICE technical report, Vol. 110, No. 130, pp. 79–84

表 1 0%から100%まで10%刻み(11種類)の分類結果の混同行列

正解ラベル	分類結果											再現率	適合率	F 値
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%			
0%	94	0	0	1	2	1	1	0	0	1	0	0.855	0.940	0.895
10%	1	80	3	5	1	4	1	1	2	2	0	0.755	0.800	0.777
20%	0	4	93	0	0	2	0	0	1	0	0	0.969	0.930	0.949
30%	2	3	0	80	0	5	0	1	0	4	5	0.769	0.800	0.784
40%	6	0	0	2	83	2	3	1	0	2	1	0.838	0.830	0.834
50%	1	9	0	4	0	69	2	1	8	6	0	0.651	0.690	0.670
60%	0	3	0	1	2	4	81	1	1	0	7	0.862	0.810	0.835
70%	0	0	0	1	4	0	0	89	4	2	0	0.899	0.890	0.894
80%	2	3	0	3	2	11	0	2	70	4	3	0.737	0.700	0.718
90%	2	2	0	5	2	7	3	3	7	67	2	0.744	0.670	0.705
100%	2	2	0	2	3	1	3	0	2	2	83	0.822	0.830	0.826

表 2 0%, 20%, 40%, 60%, 80%, 100% (6種類)の分類結果の混同行列

正解ラベル	分類結果						再現率	適合率	F 値
	0%	20%	40%	60%	80%	100%			
0%	91	0	5	1	1	2	0.867	0.910	0.888
20%	0	95	1	0	4	0	0.941	0.950	0.945
40%	8	0	84	4	2	2	0.824	0.840	0.832
60%	0	1	5	83	6	5	0.847	0.830	0.838
80%	5	5	3	3	82	2	0.837	0.820	0.828
100%	1	0	4	7	3	85	0.885	0.850	0.867

表 3 10%, 40%, 70%, 100% (4種類)の分類結果の混同行列

正解ラベル	分類結果				再現率	適合率	F 値
	10%	40%	70%	100%			
10%	98	0	1	1	0.950	0.960	0.955
40%	2	89	3	6	0.921	0.930	0.925
70%	3	0	97	0	0.960	0.970	0.965
100%	2	6	0	92	0.948	0.920	0.934

表 4 0%, 50%, 100% (3種類)の分類結果の混同行列

正解ラベル	分類結果			再現率	適合率	F 値
	0%	50%	100%			
0%	95	4	1	0.960	0.950	0.955
50%	2	94	4	0.931	0.940	0.935
100%	2	3	95	0.950	0.950	0.950

表 5 0%, 100% (2種類)の分類結果の混同行列

正解ラベル	分類結果		再現率	適合率	F 値
	0%	100%			
0%	97	3	0.951	0.970	0.960
100%	5	95	0.969	0.950	0.960

表 6 11種類, 6種類, 4種類, 3種類, 2種類における分類結果の平均 F 値

充填率	11種類	6種類	4種類	3種類	2種類
F 値	0.808	0.866	0.945	0.947	0.960

(2010).

[9] Yuki Kubo, Kana Eguchi, Ryosuke Aoki, Shigekuni Kondo, Shuzo Azuma, and Takuya Indo, FabAuth: Printed Objects Identification Using Resonant Proper-

表 7 音データの周波数パワースペクトルのみを特徴量としたときの分類結果の混同行列

正解ラベル	分類結果										
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0%	84	11	0	0	0	1	2	0	0	1	1
10%	1	81	6	4	0	0	1	1	2	2	2
20%	0	18	81	0	0	0	0	0	0	0	1
30%	1	15	2	67	0	5	0	0	2	6	2
40%	1	11	0	2	69	3	5	2	2	4	1
50%	0	17	4	11	4	33	13	1	6	8	3
60%	0	3	0	3	4	9	66	1	6	3	5
70%	2	8	0	0	1	0	0	82	4	3	0
80%	2	14	3	5	1	3	2	8	51	5	6
90%	1	9	0	9	6	12	3	5	7	43	5
100%	0	7	3	2	1	3	2	6	9	4	63

表 8 音データの周波数パワースペクトルの log スケールのみを特徴量としたときの分類結果の混同行列

正解ラベル	分類結果										
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0%	84	3	0	4	2	3	2	2	0	0	0
10%	3	60	14	10	0	4	0	1	7	0	1
20%	0	17	75	3	0	2	1	0	1	0	1
30%	1	4	1	61	2	7	1	5	6	9	3
40%	3	2	1	1	73	3	6	7	1	2	1
50%	4	9	2	8	5	45	12	1	3	9	2
60%	0	3	1	3	8	6	69	0	3	3	4
70%	0	1	0	3	2	1	0	80	6	6	1
80%	3	10	4	7	3	5	3	6	45	7	7
90%	1	2	0	14	12	5	4	7	10	43	2
100%	1	6	6	0	4	2	1	9	4	3	64

ties of Their Inner Structures, Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19), pp. 1-6 (2019).

[10] Taesik Gong, Hyunsung Cho, Bowon Lee, and Sungju Lee, Knocker: Vibroacoustic-based Object Recognition with Smartphones, ACM IMWUT, Vol. 3, No. 3, Article 82, pp. 1-21 (2019).

[11] S.S. Stevens, J. Volkman, The relation of pitch to frequency: a revised scale, The American Journal of Psychology, Vol. 53, pp. 329-353 (1940).



表 9 音データの MFCC のみの特徴量としたときの分類結果の混同行列

正解ラベル	分類結果										
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0%	81	2	2	4	6	1	0	0	3	1	0
10%	4	68	6	1	2	10	1	0	2	5	1
20%	3	11	84	0	0	0	0	0	1	0	1
30%	3	5	1	70	0	6	3	2	0	7	3
40%	8	0	0	4	69	5	4	0	1	8	1
50%	2	13	0	8	4	46	2	2	9	11	3
60%	0	1	0	2	4	1	77	0	2	3	10
70%	0	1	0	2	3	2	0	81	7	2	2
80%	2	1	0	6	1	11	2	5	57	8	7
90%	0	8	0	9	8	13	4	6	1	48	3
100%	4	2	0	3	5	9	9	3	3	4	58

表 10 加速度データの周波数パワースペクトルのみの特徴量としたときの分類結果の混同行列

正解ラベル	分類結果										
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0%	0	0	2	0	2	0	0	92	1	1	2
10%	4	13	9	2	3	2	1	38	0	3	25
20%	0	0	24	0	0	4	0	37	0	0	35
30%	1	5	9	2	1	5	0	71	0	1	5
40%	1	6	14	1	9	7	4	30	1	3	24
50%	1	0	10	0	0	8	1	78	2	0	0
60%	1	4	31	0	5	7	1	20	0	2	29
70%	0	0	5	0	0	0	0	95	0	0	0
80%	1	0	3	0	0	5	0	87	0	0	4
90%	1	1	9	0	0	3	0	83	1	1	1
100%	2	2	22	0	2	0	1	6	1	2	62

表 11 角速度データの周波数パワースペクトルのみの特徴量としたときの分類結果の混同行列

正解ラベル	分類結果										
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0%	16	2	50	1	1	5	14	5	2	1	3
10%	7	25	33	1	9	1	17	3	0	1	3
20%	0	1	92	0	0	1	0	3	3	0	0
30%	2	3	81	1	2	1	7	2	1	0	0
40%	7	12	42	0	9	1	26	3	0	0	0
50%	2	0	92	0	0	0	1	1	4	0	0
60%	11	7	46	0	6	2	22	4	0	1	1
70%	6	2	89	0	0	0	0	1	2	0	0
80%	0	1	91	0	0	1	0	2	4	0	1
90%	0	0	93	0	2	0	3	2	0	0	0
100%	3	4	76	0	1	1	7	5	3	0	0

[12] Taesik Gong, Hyunsung Cho, BowonLee, and SungJu Lee. 2018. Identifying Everyday Objects with Smartphone Knock. In Proceedings of the 2018 CHI Conference Extended Abstracts on Human Factors in Computing Systems.ACM.

[13] ChrisHarrison,JuliaSchwarz,andScottE.Hudson.2011. TapSense:EnhancingFingerInteractiononTouchSurfaces.InProceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11).

表 12 音データの周波数パワースペクトル, 周波数パワースペクトルの log スケール, MFCC を特徴量としたときの分類結果の混同行列

正解ラベル	分類結果										
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0%	92	0	0	2	3	1	1	0	0	0	1
10%	0	83	5	5	0	4	0	1	1	1	0
20%	0	7	91	0	0	1	0	0	1	0	0
30%	2	3	0	79	2	5	0	0	1	4	4
40%	6	0	0	1	82	3	3	1	1	1	2
50%	1	10	0	5	3	67	2	1	6	4	1
60%	0	0	0	1	1	5	81	1	3	3	5
70%	0	1	0	1	3	0	0	89	4	2	0
80%	4	1	0	3	1	9	2	2	70	5	3
90%	0	2	0	5	6	6	6	3	5	62	5
100%	2	1	0	2	2	3	3	3	5	6	73

表 13 加速度と角速度の周波数パワースペクトルを特徴量としたときの分類結果の混同行列

正解ラベル	分類結果										
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0%	30	1	0	0	0	9	2	41	13	1	3
10%	9	26	4	1	8	3	9	29	2	1	8
20%	0	1	29	0	0	10	0	17	13	0	30
30%	1	2	8	4	2	15	4	50	11	1	2
40%	3	9	6	3	10	11	22	12	11	1	12
50%	2	0	6	0	0	18	1	54	16	0	3
60%	7	7	25	0	3	2	22	5	11	2	16
70%	4	2	0	0	0	11	2	70	11	0	0
80%	0	1	3	0	0	16	1	59	18	0	2
90%	0	2	8	0	1	22	2	48	14	2	1
100%	2	3	18	0	0	4	7	3	2	0	61

11).ACM,NewYork,NY,USA,627–636.

[14] ShanLuo,LeqiZhu,KasparAlthoefer,andHongbinLiu.2017. Knock-Knock:Acousticobjectrecognitionbyusingstackeddenoising autoencoders. Neurocomput.267,C(Dec.2017),18–24.

[15] Lei Shi, Maryam Ashoori, Yunfeng Zhang, and Shiri Azenkot. 2018. Knock Knock, What’s There: Converting Passive Objects into CustomizableSmartControllers.InProceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI ’18).ACM,NewYork,NY,USA,Article31,13pages.

[16] ギョクギョク オウ, 岡田 義広, 3D Printed Object Recognition Augmented Reality Framework based on the dealing with Point Cloud Data, 第 81 回全国大会講演論文集 2019(1), 109-110, 2019-02-28.

[17] Kunihiro Katou, Shouta Yamanaka, Yoshiaki Miyashita, 3D-printed Objects with 2D Touch Input, Transactions of Information Processing Society of Japan, 59(2), 670-680, 2018-02-15.

[18] PayPay, <https://paypay.ne.jp/>.

[19] 楽天ペイ, <https://pay.rakuten.co.jp/>.

[20] SciPy.org, <https://www.scipy.org/>.

[21] HASC Logger, <http://hasc.jp/tools/hasclogger.html>.

[22] Weka, <https://www.cs.waikato.ac.nz/ml/weka/>.

表 14 各特徴量の組合せにおける平均 F 値

特徴量	音 FFT	音 logFFT	音 MFCC	加速度 FFT	角速度 FFT	音 3 種	加速度, 角速度 2 種	5 種全部
F 値	0.654	0.632	0.672	0.130	0.109	0.789	0.241	0.808