

# 容器を振る際の音を用いた容量識別手法

小口 雄斗 <sup>†1</sup>

志築 文太郎 <sup>†2</sup>

高橋 伸 <sup>†2</sup>

<sup>†1</sup> 筑波大学 大学院システム情報工学研究科

<sup>†2</sup> 筑波大学 システム情報系

## 1 はじめに

音は空中や物体中を伝搬し、響き方によって複雑な情報を表現することができる。そのため産業分野においては超音波パルスや周波数スイープによる周波数応答を用いて、構造物の傷の検出などの非破壊検査 [1] で利用されている。また身近な例として、すいかを叩いた際の音からその水分量や甘みがわかることはよく知られている。

本研究では容器の容量に応じて容器を振った際の音が変わることに着目し、そのスペクトログラム画像を用いた畳み込みニューラルネットワークによる容量識別手法を提案する。本稿では内容物に食塩を用いた実験により 5 段階の容量の識別精度を検証し、本手法の有効性の評価を行った。

## 2 関連研究

容量識別に関する研究はこれまでも様々な手法により取り組まれてきた。例えば SoQr [2] 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対象の容器に直接貼り付けることで、スピーカから発した周波数スイープに対する周波数応答を用いて容量の識別を行う手法である。また VibeBin [3] 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容量を計測している。Wei ら [4] の研究では静電容量センサを内側に取り付けた容器を使用して容器内部の液体の量の識別に取り組んでいる。これらの手法では、特殊な装置を対象の容器に直接取り付けるコストを必要としている。また SoQr では非常に高い精度での容量識別を実現している一方で柔らかい素材の容器には適用が難しいことが指摘されている。本研究で提案する手法ではセンシングの際に屋内に設置したマイクを用いるため、容器の加工などの手間を必要とせず、また容器の形状や素材にかかわらず容量識別が可能となると考えられる。

## 3 提案手法

容器に触れる、あるいは容器を動かす際に発生する音はその容量によって異なる響き方を示す。例えば容器内が満杯に近い状態で容器を振った際の音は低く、空に近い状態では高い音になる。同様に容器を叩いたり置いたりした際の音も容量に応じて変化する。これらのことから容器から発せられる音はその容量を識別するのに有望であると考えられる。そこで容器を扱う際の音から以下の手順で特徴を抽出し容量の識別と提示を行う。

1. 容器を扱う際の音をマイクで録音し分析用のコンピュータに送信
2. 収集した音源に周波数分析を施しスペクトログラム画像を抽出
3. スペクトログラム画像を CNN のモデルに入力し容量を識別
4. 識別された容量を記録し、モバイルデバイス等を通してユーザに提示

本手法は容量の識別に音を用いるため生活用品を普段使う過程で発生する音からその容量を識別するというような活用ができる。例えば石鹸のボトルはポンプを押す際に音が発せられる。その音から石鹸を使用する度にボトル内の容量を記録することで、石鹸の補充する時期をユーザに通知したりや自動で購入を行ったりするような使い方ができる。

## 4 実装と実験

容器を振る際の音から本手法による容量の識別実験を行い、その有効性を評価した。容器にはプラスチック製の円筒形容器 (図 1) を使用し、容量は 5 段階 (20%, 40%, 60%, 80%, 100%) として、内容物には食塩を用いた。音を収集するマイクにはダイナミックマイク SHURE SM58 を使用し、コンピュータとの接続にはオーディオインタフェース Steinberg UR22 を用いた。

### 4.1 スペクトログラム画像の抽出

スペクトログラム画像には 1 秒間の音源から対数スケールで抽出したものをを用いる。出力及び保存には Python の画像描画ライブラリである matplotlib の specgram 関数を使用し、抽出過程における周波数分析には

Sonically Identifying Content Level by Shaking Container

<sup>†1</sup> Yuto Koguchi <sup>†2</sup> Buntarou Shizuki <sup>†2</sup> Shin Takahashi

<sup>†1</sup> Department of Computer Science, University of Tsukuba

<sup>†2</sup> Faculty of Engineering, Information and Systems, University of Tsukuba



図 1: 実験に用いたプラスチック製の円筒形容器  
サンプル数を 512, 窓関数をハミング窓とした高速フーリエ変換を使用する。

#### 4.2 CNN モデルの構築

CNN のモデルには既存のモデルである VGG16 [5] を参考に構築した 7 層の畳み込み層と 2 層の全結合層から成るモデルを用いる。実装にはオープンソースニューラルネットワークライブラリである Keras を使用した。構築したモデルの各畳み込み層のフィルタサイズは  $3 \times 3$  であり、活性化関数は中間層が ReLu, 出力層は SoftMax である。

#### 4.3 データセット生成

5 段階の各内容量の容器を振った際の音をマイクで録音し、各音源に対してスペクトログラム画像の抽出処理を行いデータセットを作成した。録音環境の騒音値は約 33 dB であり、これは聴覚的な目安では静かに分類 [6] される値である。録音を行う際のサンプリングレートは 44100 Hz とし量子化ビット数は 16 bit とした。また容器の握り方が一定となるように容器側面に目印を付けて握り方の変化を抑制し、容器を振る際のマイクからの距離は 10 cm とした。

振り続ける時間は 1 分間とし、音源の先頭から 1 秒間の音源の切り出しを 0.5 秒ずつスライドさせて行うことで一度の録音で図 2 に示すような 120 個のスペクトログラム画像が得られる。CNN のモデルが録音時の環境に依存しないようにするために、5 段階の各内容量についての振る際の音の録音を 6 日間実施し、合計で 3600 個のスペクトログラム画像を取得した。この内 5 日目までのスペクトログラム画像 3000 個を CNN のモデルの学習に使用する訓練用データとし、6 日目のスペクトログラム画像 600 個を学習させたモデルによる内容量識別の検証用データとした。

#### 4.4 実験結果

バッチサイズを 60, エポック数を 100 として訓練用データを用いて CNN のモデルの学習を行い、検証用データの識別を行った結果を以下に示す。

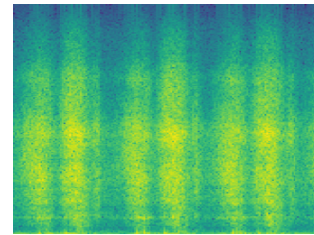


図 2: 抽出したスペクトログラム画像

表 1 は 600 個の検証用データについて、識別された内容量と実際の内容量の関係を表した混同行列である。全体の識別精度は約 88 % であった。このことから内容量を 5 段階とした場合についておよそその内容量を識別できる可能性が示された。

表 1: 内容量の識別結果の混同行列

		Predict Label				
		20%	40%	60%	80%	100%
True Label	20%	106	14	0	0	0
	40%	2	86	30	2	0
	60%	0	0	105	15	0
	80%	0	2	4	112	2
	100%	0	0	0	1	119

## 5 まとめと今後の展望

本研究では容器を振る際の音がその内容量によって変化することに着目し、容器を振る際の音のスペクトログラム画像と CNN を用いた内容量識別手法を提案した。評価実験では 5 段階の内容量の識別を検証し、その結果からは本手法の有用性が示唆された。

今後はデータ数を増強することにより識別精度の向上を試みる。また 5 段階以上のより細かい単位での内容量の識別を検証する。

## 参考文献

- [1] 飯塚幸理. 鋼管商品を支える非破壊検査技術. JFE 技報, No. 9, pp. 40–45, 2005.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, 2014.
- [6] World Health Organization. *Hearing loss due to recreational exposure to loud sounds: a review*. World Health Organization, 2015.