

# アクティブ音響センシングを用いた食材認識カトラリーの設計と評価

西井 遥菜<sup>1</sup> 双見 京介<sup>1</sup> 村尾 和哉<sup>1,2</sup>

**概要:** 本研究では、ユーザが道具を介して物体とインタラクションをする場面において、ユーザが使用する道具にアクティブ音響センシング技術を適用させることで、ユーザが道具を介して接触した物体を認識する手法を提案する。食事場면을対象とし、ユーザに食事と関係無い動作を行わせることなくユーザが食べた食材の認識を行うカトラリーとして、食材を刺すフォーク型デバイスと、食材をのせるスプーン型デバイスを実装した。提案手法の有効性を評価する実験を 14 種類の野菜およびデバイスと食材が接触していない状態の計 15 種類に対して行った結果、フォーク型デバイスで F 値 0.916、スプーン型デバイスで F 値 0.949 が得られたことを確認した。

## 1. はじめに

人類はこれまでに、道具を使用して多種多様な物体とのインタラクションを行ってきた。そこで、ユーザが道具を介して物体とインタラクションをする場面において、人がインタラクションを行った物体を認識できれば有用である。

本研究では、ユーザが道具を介して物体とインタラクションをする場面において、ユーザが使用する道具にアクティブ音響センシング技術を適用させることで、ユーザが道具を介して接触した物体を認識する手法を提案する。

提案手法の有効性を検証するために、本研究では食事場면을対象として、ユーザが食べた食材の認識を、ユーザに特別なふるまいをさせることなく行うためのカトラリーを開発する。具体的には、食材を刺すフォーク型デバイスと、食材をのせるスプーン型デバイスを実装した。評価実験では、それぞれのデバイスを用いて提案手法の食材の認識精度の評価を 14 種類の野菜と食材が接触していない状態の計 15 種類に対して行う。

## 2. 関連研究

本研究と同じくアクティブ音響センシングを用いた認識手法には次のものがある。大野らはアクティブ音響センシングを用いたタッチ認識技術を提案している [1][2]。これは、物体にコンタクトスピーカとコンタクトマイクを取り付け、物体の触り方の違いを取得することで把持状態を認識する手法である。河田らは、把持状態の認識を物体では

なくユーザにセンサを取り付けることにより行った [3]。また、Mujibiya らは低周波の超音波で皮膚表面を共振させ、身体上の別地点で受け取ることで身体上の接触位置やジェスチャを推測している [4]。このほか、測位にもアクティブ音響センシングは用いられている。Kunze らは、携帯電話に内蔵されているスピーカの音と振動モータからの振動を利用して、加速度センサとマイクを用いて応答を分析することで、携帯電話の位置を把握する手法を提案している [5]。岩瀬らは、アクティブ音響センシングを仕込んだアクリル板の上の物体の種類と位置の認識を行った。また、複数物体の同時識別も可能とした [6]。このようにアクティブ音響センシングを用いた認識では、人が自発的に動作して音や振動を発生させなくても、物体に触れているだけで物体の情報を取得できる。本研究では、ユーザと物体とのインタラクションが道具を介して行われる場面において、アクティブ音響センシングの新しい応用を提案している。

## 3. 提案手法

本節では、フォークやスプーンといったカトラリーに対してアクティブ音響センシングを適用し、カトラリーが接触した食材を認識する手法について述べる。

### 3.1 想定環境

本研究では、食事場면을想定し、フォークやスプーンなどのカトラリーにアクティブ音響センシングを適用し、カトラリーで刺したり、すくったりした食材を識別する手法を提案する。アクティブ音響センシングをカトラリーに適用するアプローチを採用した理由を述べる。人の生活にお

<sup>1</sup> 立命館大学大学院情報理工学研究科

<sup>2</sup> 科学技術振興機構さきがけ

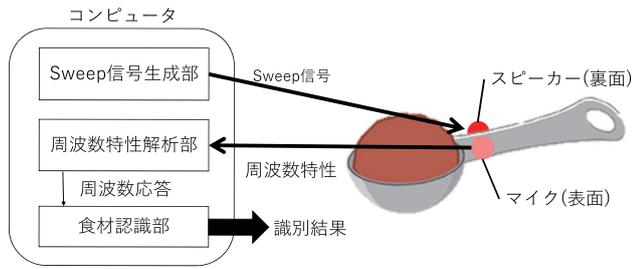


図 1 提案システム概要

いて食事は必須の活動であり、健康的な生活を送るために食事を記録することは有用である。既存の食事の記録手法として、食事をする前にメニュー全体の写真を撮る方法や、食事の前後で人が記録する方法がある。また、調理するレシピの記録や食事の後に食べた物を思い出しながら記録する方法などがある。しかし、写真を撮って記録する方法は、料理の中に含まれるすべての食材を記録することは難しい。また、食べた順序を記録するためにはビデオで食事中記録し続けなければならないため面倒であるし、プライバシーの問題を含む。人が記録する方法では、記録し続けるための記録者の高いモチベーションが必要であり、記入漏れや誤記録といった記録ミスの回避も難しい。また、食材に直接センサやタグを取り付ける方法は衛生面や手間の観点から現実的ではない。

これらの既存手法に対して、提案手法はユーザが使用するカトラリーのみでセンサを取り付けるため、カトラリーで食材に接触するだけで物体を認識できる。カトラリーで食材に接触する動作は、人が食事をするときに必ず生じる動作であるため、ユーザは食材を認識させるための特別な動作を行う必要がなく、通常通りに食事をするだけで実際に摂取した食材を認識できる。

### 3.2 提案手法の流れとシステム構成

提案システムの概要を図 1 に示す。提案システムはコンタクトスピーカとコンタクトマイクが搭載されたカトラリー（フォークおよびスプーン）と、Sweep 信号生成部、周波数特性解析部、食材認識部から成るコンピュータによって構成される。

提案手法は食材のもつ音響特性をもとに、食材の認識を行う。食材を含むすべての物体には固有の音響特性があり、共振モード、共振周波数、モード減衰比によって表せられる。これは、物体の形、材質、境界条件により決定される。これらの音響特性の差を利用することで、食材の種類の違いを識別することができる。

提案手法では、各食材から得られる音の周波数特性を区別するために、アクティブ音響センシングを用いる。具体的にはまず、ユーザが使用するカトラリーに Sweep 音を発するコンタクトスピーカと反響音を取得するコンタクト

マイクを取り付ける。そして、周波数が時間とともに変化する Sweep 音をコンピュータの Sweep 信号生成部で生成して振動スピーカから発してカトラリーを振動させる。その振動はカトラリーを経由してコンタクトマイクから取得されるが、カトラリーに接触する食材の種類や形状、カトラリーの状態によって変化する。そして、コンタクトマイクから取得した音（振動）の周波数特性を得るために、固定長のウィンドウで切り出した音を FFT（高速フーリエ変換：Fast Fourier Transform）してパワースペクトルを得る。そして周波数特性では、事前に収集した各食材のパワースペクトルを学習した機械学習モデルを用いて、カトラリーに接触している未知の食材の認識を行う。

### 3.3 Sweep 信号生成部

Sweep 信号生成部では、200ms のあいだに周波数が 20kHz から 40kHz まで直線的に変化する Sweep 信号（Chirp 信号とも呼ばれる）を繰り返し生成する。この周波数帯域は、先行研究 [1] の論文を参考にして決定した。具体的には、人の可聴域は年齢にもよるが 20Hz から 20kHz 程度であり、人の非可聴域である 20,001Hz から 40,000Hz までの範囲と合わせて記録することにした。サンプリング周波数は 96kHz である。Sweep 信号を繰り返し流すと、繰り返しの切れ目の部分で周波数が大きく変化することによるインパルスノイズが発生するため、ノイズを軽減するために音の繰り返しの瞬間にクロスフェード処理を行う。この Sweep 信号はフリーのオープンソースオーディオソフトウェアである Audacity\*1 を用いて作成した。

### 3.4 周波数特性解析部

周波数特性解析部では、カトラリーのマイクから取得した振動応答を、FFT を用いてパワースペクトルに変換する。振動応答は 96kHz、16bit でサンプリングした。振動応答の時系列データから 96000 サンプルをごとに FFT を行いパワースペクトルを得る。サンプリング定理により、48000 点の 0Hz から 48kHz までの周波数のパワースペクトルを得られるが、この中から人の非可聴域かつ Sweep 信号の周波数である 20,001Hz から 40,000Hz の領域を抽出する。このことから、20000 点のパワースペクトルが得られる。

### 3.5 食材認識部

食材認識部では、事前に取得した特徴量ベクトルと正解の食材ラベルからなる学習データを用いて認識モデル（評価では SVM を利用）を構築しておき、入力された未知の食材の特徴量ベクトルを認識する。学習データには認識したい食材の特徴量ベクトルに加えて、カトラリーが何の食

\*1 <https://www.audacityteam.org/>

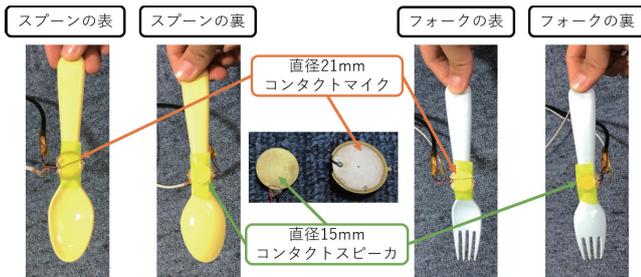


図 2 マイクとスピーカを取り付けたフォークとスプーン

材にも触れていないときの特微量ベクトルを含めている。これによりカトラリーで食材に触れているときは食材名が出力され、食材に触れていないときは「触れていない」と出力される。

#### 4. 評価実験

提案手法の有効性を示すために、野菜を認識対象として実験を行う。

##### 4.1 提案システム

本研究で使用する食材を刺すフォーク型デバイスと食材を乗せるスプーン型デバイスを図 2 に示す。図ではマイクとスピーカをともに柄に付けている。フォークとスプーンはプラスチック製で、ともに全長 16cm である。物体を振動させるコンタクトスピーカには圧電サウンド (FGT-15T-6.0A1W40) を使用した。直径 15mm、厚さ 0.3mm 以下の超薄型軽量の円盤形である。コンタクトスピーカの発した音を取得するコンタクトマイクはスライプ社の OMR20F10H-BP-310 を使用した。直径 21mm、厚さ 0.3mm 以下の超薄型軽量の円盤形である。音の入出力は USB 接続のオーディオインタフェース (Steinberg 社の UR 22mkII) を介して行う。また、取得した時系列の振動応答を FFT する周波数特性解析部、機械学習を行う食材認識部のコンピュータとして、レノボ・ジャパン社の ThinkPad X1 Carbon (CPU: Intel Core i7-5600 2.60GHz, RAM: 8.00GB) を使用した。

##### 4.2 評価実験で使用する野菜

評価実験では、厚生労働省が発表している日本人における野菜の摂取量ランキング<sup>\*2</sup> を参考に、一般的に利用される野菜である、大根、玉葱、キャベツ、白菜、人参、ほうれん草、トマト、キュウリ、カボチャ、レタス、ネギ、ピーマン、ジャガイモ、サツマイモの 14 種類の野菜を採用した。

野菜の個体ごとの差異に関わらず提案手法が野菜を認識できるかを検証するために、1 種類の野菜につき 5 個体を用意した。また、切った形による認識精度の変化を調査するために、各個体を表 1 に示す 3 種類の形に切った。一例

表 1 野菜の切り方

野菜の種類	形 A	形 B	形 C
キャベツ	ざく切り	千切り	みじん切り
人参	いちょう切り	乱切り	輪切り
白菜	ざく切り	千切り	みじん切り
キュウリ	斜め切り	乱切り	輪切り
ピーマン	乱切り	輪切り	短冊切り
レタス	ざく切り_芯	ざく切り_複数	ざく切り_葉
玉葱	みじん切り	短冊切り	串切り
ジャガイモ	半切り	4 分の 1 切り	輪切り
カボチャ	乱切り_大	薄切り	乱切り_小
大根	輪切り	いちょう切り	短冊切り
ネギ	斜め切り	輪切り	筒切り
ほうれん草	葉_半分	葉_複数	葉_1 枚
サツマイモ	輪切り	斜め切り	乱切り
トマト	賽の目切り	半月切り	串切り

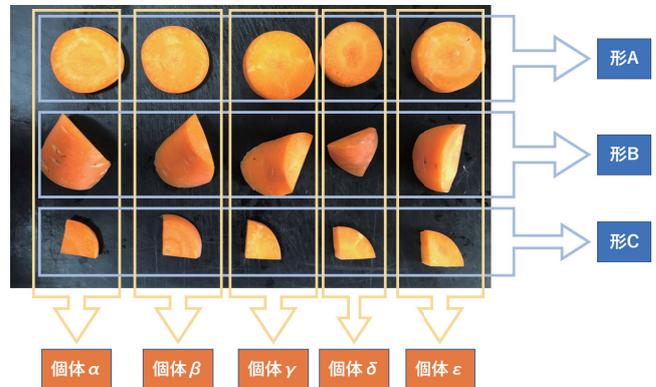


図 3 個体と切り方の違いによる実験素材

としてニンジンの実験素材を図 3 に示す。1 本のニンジン (個体) から形 A (いちょう切り)、形 B (乱切り)、形 C (輪切り) の 3 種類の形に 1 個ずつ切り出した。つまり、実験の素材として 14 種類の野菜 × 5 個体 × 3 種類の形 = 210 素材を用意した。なお、野菜の切り方に関しては、野菜ごとに一般的な切り方を選択した。そのため、野菜ごとに 3 種類の切り方は異なる。

##### 4.3 実験環境とデータセット

フォーク型デバイスについて、前述の 210 素材すべてに対してフォークの先端で素材を刺した状態で 1 素材につき 60 秒間、振動データを取得した。スプーン型デバイスについてもスプーンの皿に素材をのせた状態で同様に振動データを取得した。また、食事をしない状態を認識するため、デバイスと食材が接触していない状態を 60 秒間、5 回振動データを取得した。つまり、野菜と合わせて全部で 215 素材ある。取得したデータから、長さ 1 秒のウィンドサイズで FFT を計算し、1 素材あたり特微量を 50 サンプル得た。その結果、10750 サンプルデータ (50 ウィンド × 3 種類の形 × 5 個体 × 14 種類の野菜 + デバイスと食材が接触していない 1 状態 × 5 回分) を得た。

\*2 <https://www.mhlw.go.jp/stf/houdou/000096138.html>

表 2 各野菜の認識精度

野菜の種類	フォーク型デバイス			スプーン型デバイス		
	適合率	再現率	F 値	適合率	再現率	F 値
キャベツ	0.952	0.956	0.954	0.986	0.967	0.976
人参	0.801	0.909	0.852	0.923	0.899	0.911
白菜	0.853	0.819	0.835	0.919	0.999	0.957
キュウリ	0.871	0.848	0.859	0.966	0.991	0.978
ピーマン	0.987	0.987	0.987	0.981	0.983	0.982
レタス	0.922	0.971	0.945	0.821	0.912	0.864
玉葱	0.908	0.813	0.858	0.934	0.976	0.954
ジャガイモ	0.912	0.921	0.916	0.945	0.919	0.932
カボチャ	0.888	0.931	0.909	0.953	0.937	0.945
大根	0.868	0.844	0.856	0.940	0.979	0.959
ネギ	0.982	0.923	0.951	0.980	0.904	0.940
ほうれん草	0.989	0.989	0.989	0.956	0.808	0.876
サツマイモ	0.856	0.912	0.883	0.976	0.984	0.980
トマト	0.970	0.915	0.942	0.975	0.979	0.977
非接触状態	1.000	1.000	1.000	1.000	1.000	1.000
平均	0.917	0.916	0.916	0.950	0.949	0.949

#### 4.4 結果

提案手法による食材の認識精度を評価するために、14種類の野菜と食材非接触状態の計15種類の分類を行った。機械学習モデルはSVM（多項式カーネル）を用いて5-fold個別交差検証を行って比較した。つまり、ニンジン $\alpha$ 、個体 $\beta$ 、個体 $\gamma$ 、個体 $\delta$ を学習データとして学習モデルを構築してニンジン $\epsilon$ をテストデータとして認識し、テストデータをローテーションするようにした。

フォーク型デバイスとスプーン型デバイスの認識精度の適合率、再現率、F値を表2に示す。フォーク型デバイスではF値0.916、スプーン型デバイスではF値0.949を示した。

#### 4.5 考察

評価結果から、フォーク型デバイスはF値0.916、スプーン型デバイスはF値0.949であった。フォーク型デバイスとスプーン型デバイスの両方において高い認識精度が得られたことから、食材の刺し方や乗せ方にばらつきがあり、カトラリーと野菜の接触方法の異なっても提案手法が有効であるといえる。この結果から、箸やナイフなど食材との接触方法が異なるほかのカトラリーに対しても提案手法が汎用的に利用できると思われる。また、異なる個体を学習しても野菜を認識できることがわかったため、認識時とは別の野菜の個体を事前に学習したモデルを利用でき汎用性は高いといえる。さらに、同じ野菜で同じ切り方でも形は多少異なるため、同じ切り方のなかでの形のブレの影響は小さいといえる。実際の食事場面では、出現頻度の高い複数の切り方で切った野菜を事前に学習しておけば、高い精度で食材の種類が認識できると考えられる。

提案手法の応用として、食事のログが考えられる。ユー

ザが何の食材をどういった順番でどのタイミングで食べたかを記録できる。ほかには、提案手法を調理器具に適用した調理支援がある。包丁やお玉などの調理器具に提案手法を適用することで、調理しながら食材を認識できるため、調理過程の記録（レシピ自動作成）や調理段階に応じた調理方法の提示など調理インタラクションに利用できる。

#### 4.6 リミテーション

提案手法について本論文で検証できていない点として以下が考えられる。一点目は、検証における食材の種類である。今回選択した野菜以外にも、肉や魚、キノコ類など、さまざまな食材が存在するため、今後は野菜以外の食材でも提案手法の有効性を調査する。二点目は、デバイスの実用性である。実装したプロトタイプデバイスはカトラリーとPCは有線接続されているが、食事時に使用するため無線通信であることが望ましい。三点目は、食材の組成が変化した場合の認識精度の変化の調査である。調理段階で茹でたり焼いたりすることで野菜の組成が変化し、認識精度が低下する可能性がある。

#### 5. おわりに

本論文ではユーザが道具を介して物体とインタラクションをする場面において、ユーザの使用道具にアクティブ音響センシング技術を適用させることで、物体認識を行う手法を提案した。食事場면을対象とし、フォーク型デバイスとスプーン型デバイスを実装した。提案手法の有効性を評価する実験を14種類の野菜およびデバイスと食材が接触していない状態の計15クラスに対して個別交差検証で実施し、フォーク型デバイスにおいてF値0.916、スプーン型デバイスにおいてF値0.949が得られた。今後は、フォークとスプーン以外のカトラリーに提案手法を適用させた評価や、食事や調理以外の場面においても提案手法の有効性を検証する。

#### 参考文献

- [1] 大野 誠ら, アクティブ音響センシングを用いた把持状態認識, インタラクション'13, pp. 56–63 (2013).
- [2] M. Ono et al., Touch and Activate: Adding Interactivity to Existing Objects using Active Acoustic Sensing. UIST'13, pp. 31–40 (2013).
- [3] 河田大史ら, 振動のアクティブセンシングによるウェアラブル所持物体認識システム, DICOMO'15, pp. 400–406 (2015).
- [4] A. Mujibiyah et al., The sound of touch: on-body touch and gesture sensing based on transdermal ultrasound propagation. ITS'13, pp. 189–198 (2013).
- [5] K. Kunze et al., Symbolic Object Localization Through Active Sampling of Acceleration and Sound Signatures. UbiComp'07, pp. 163–180 (2007).
- [6] 岩瀬大輝ら, 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, インタラクション'18, pp. 62–71 (2018).