

アクティブ音響センシングを用いた野菜認識調理道具

西井 遥菜¹ 双見 京介¹ 村尾 和哉¹

概要：人は多くの場面で道具を使用して衣食住などの生産活動を行ってきた。こういった場面において、人がインタラク션을行った物体を認識できれば有用であるが、既存手法は複数の観点で課題を残していた。本研究では、ユーザが道具を介して物体とインタラク션을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、物体認識を効果的に行う手法を提案する。提案手法では、ユーザが使用する道具にスピーカとマイクを取り付け、スピーカから音響信号を流し、道具が物体に接触した際に、物体を伝搬した音響信号をマイクから取得し、その音響信号の周波数特性を解析することで、物体を認識する。このように提案手法では、認識したい物体にユーザの使用道具が接触すること、認識したい物体が固有の音響特性をもつことを利用している。本稿では、調理や食事といった道具を用いて食材とインタラク션을する場面を対象にして、食材を挟むトンガ型デバイスと、切る包丁型デバイスを実装した。そして、提案手法の有効性を評価する実験を7種類の野菜をもとに3つ行い、いずれの実験においても、88%以上の精度で野菜の種類を認識できることを確認した。

1. はじめに

人類はこれまでに、道具を使用して多様な物体とのインタラク션을行ってきた。人類最初の道具とされる石器は目的に応じて加工され、対象となる物体を挟む、切る、つぶす、耕す、すくう、掘る、書くなど多様な機能を持つようになり、食事や調理、農作業や狩猟、衣類作成や住居建築、絵画や学習用記録などの多様な目的のために利用されてきた。そして現代においても人は、衣食住や娯楽、学習などの活動のために、道具を用いている。

こういった、ユーザが道具を介して物体とインタラク션을する場面において、人がインタラク션을行った物体を認識できれば有用である。実空間内の物体を認識する技術は、近年のセンサやコンピュータの小型化や性能向上に伴って発展しており、様々なユーザ支援に利用されている。例えば、物体を認識してログを残すライフログシステムへの利用や、認識した物体からわかる状況に応じたサービスを提供するコンテキストウェアシステムへの利用が挙げられる。しかし、既存手法には課題がある。例えば、電子的なタグを認識すべき物体自体に仕込む手法 [1] は、認識したい物体にタグをつけるのが適切ではない場面には向かない。また、カメラを用いた物体認識手法 [6] は、認識すべき物体に人が被るせいでカメラ映像に物体が映らないオクルージョンが起りえる場面や、人に心理的負担を起したくない場面には向かない。

そこで、本研究では、ユーザが道具を介して物体とインタラク션을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、物体認識を効果的に行う手法を提案する。アクティブ音響センシング技術とは物体がもつ固有の音響特性を利用した物体認識手法であり、スピーカを用いて特定の音響信号を物体内部に伝搬させ、マイクで得たその応答の周波数特性を解析することで、物体の種類や状態を認識するものである。提案手法では、ユーザが使用する道具にスピーカとマイクを取り付け、スピーカから音響信号を流し、道具が物体に接触した際に、物体を伝搬した振動・応答をマイクから取得し、その応答の周波数特性を解析することで、物体を認識する。このように提案手法では、認識したい物体にユーザの使用道具が接触すること、認識したい物体が固有の音響特性をもつことを利用している。本稿では、調理や食事といった道具を用いて食材とインタラク션을する場面を対象にして、食材を挟むトンガ型デバイスと、切る包丁型デバイスを実装する。そして、それぞれのデバイスを用いて、提案手法の有効性を評価する3つの実験を7種類の野菜をもとに行う。

以下、2章では、本研究の関連研究について紹介する。3章では、提案手法の特徴とセンシング原理について述べる。4章では、提案手法のプロトタイプについて示したのち、5章でプロトタイプに対する評価実験並びにその評価や有用性について議論する。最後に、6章にて本研究をまとめる。

¹ 立命館大学情報理工学部

2. 関連研究

2.1 物体認識のためのセンシング技術

物体の種類や状態を認識する手法は数多く存在する。電子的なタグを認識すべき物体自体に仕込む手法は日常生活において広く普及している。Boukraa らの RF (Radio Frequency) タグは対象物に取り付けられたタグをタグ読み取り機で読み取り、タグを認識することで対象物の認識を行う [1]。Kim らは pantilt メカニズムを用いたステレオビジョンによるスマートタグシステムを提案した [2]。Kim らが開発した IRED デバイスを含むスマートタグを物体に取り付けることで、IC タグの情報を取得するロボットにオブジェクト認識を行わせることを実現した。また、Martin らの特別な視覚的マーカーを物体に取り付けることで、コンピュータビジョンアルゴリズムによって認識、追跡をリアルタイムビデオストリーム内で行うスタンドアロンアプリケーションを実現した [3]。Chris らは音響バーコードを用いて、材料を認識することに成功した。固有の ID をエンコードするサウンドを生成し、それを認識することで多種多様な材料を認識することができる [4]。これらタグを用いた手法は、認識できる物体の数が多く、認識精度が高いという利点があるが、認識したい物体すべてにタグをつける必要があるため、認識したい物体にタグをつけるのが適切ではない場面には向かない。例えば、調理場面においては食材にタグをつけることは適さない。これに対して提案手法は、ユーザの使用道具が物体に接触するだけで物体を認識できるものであり、本研究の対象場面に適している。

カメラの画像から空間内の物体を認識する手法は、多く提案されている。カメラを定点に設置する手法としては、Guo らの Zensors++がある [5]。これは、カメラを内蔵したスマートアプライアンスで、ここから得られた画像をもとに、画像内の物体の認識を一般化するために人にラベルづけを行ってもらったものである。また、カメラの視点移動をさせるために、安岡らは自律移動型ロボット（ドローン）を用いて認識システムの構築を行った [6]。これは、ドローンを飛行させ、その前方に取り付けられているカメラで室内の様子を撮影し、その画像を用いて室内にあるものの物体認識を行う。カメラの物体認識精度を向上させるための研究も盛んであり、例えば、塩濱らは、大画像から SIFT 特徴量を抽出することで物体の検出の高速化を実現させた [7]。これは、特定物体の認識を高速化するために、画像を分割し、その画像を順番に特徴点検出しマッチングすることで、高速化を図ったものである。また、野菜の認識においてもカメラを使用した手法は多くある。こういったカメラを用いた物体認識技術の利点としては、カメラの撮影範囲内の複数の物体を一括で認識でき、また、物体の見た目に特徴があれば多様な物体を高精度で認識できるこ

とが挙げられる。しかし、これらの技術は、認識すべき物体に人が被るせいでカメラ映像に物体が映らないオクルージョンの問題が起こる場面や、カメラの存在がプライバシー侵害などの観点から心理的負荷を生み得る場面には適さない。これに対して、提案手法は、対象場面においてユーザの使用道具が物体に必ず接触することを利用しているためオクルージョンの問題が起きず、また、心理的負担の生じるセンサも利用しない。

2.2 音響センシングを用いた認識手法

ある動作などを行うことによって発生した音を取得して分析、解析を行うパッシブ音響センシングを用いた技術はたくさんある。Brian らの研究では、指先の動作を行った際に生成される骨伝導音を用いることで、指先のジェスチャを認識する [8]。また、Chris らは質感ある素材の表面に指の爪をドラッグしたときに発生するサウンドを利用して、ジェスチャー入力表面を作成した [9]。Chris らはほかにも、身体の機械的振動を分析することにより、腕と手の指のタップ位置の推定を可能とした [10]。Roderick らは、触覚入力によって制御される手持ち式の対話型デバイスを用いて、これらの表面をなでたり引っ掻いたりすることによって発生する振動信号を分類することができる。これは、相互作用を制御するための機器として非常に安価に製造することができる [11]。このように音響センシングでは、ユーザと物体のインタラクションに多様な可能性を見出すことが可能となる。

スピーカとマイクを使用したアクティブ音響センシングを用いた研究は以前にも提案されている。大野らはこの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 [12][13]。これは、既存物体にコンタクトマイクとコンタクトスピーカを取り付け、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、この把持状態の認識を、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 [14]。また、Adiyan 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。これにより、身体上の接触位置やジェスチャを推測できる [15]。位置把握にもアクティブ音響センシングは用いられている。例えば、Kim らの実験では、携帯電話に内蔵されているスピーカの音と振動モーター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ことができる [16]。また、岩瀬らは、日常生活を支援するために、実空間内にある物体情報を取得するための物体情報認識手法を述べている [17]。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、間接的な情報の取得を可能とする。本研究では、このセンシング技法の新しい使い方を、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提案している。

3. 提案手法

提案手法は、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、物体認識を行うものである。提案手法では、ユーザが使用する道具にスピーカとマイクを取り付け、スピーカから音響信号を流し、道具が物体に接触した際に、物体を伝搬した振動・応答をマイクから取得し、その応答の周波数特性を解析することで、物体を認識する。このように提案手法では、認識したい物体にユーザの使用道具が接触すること、認識したい物体が固有の音響特性をもつことを利用している。

3.1 対象場面と提案手法の特徴

提案手法が有用な対象場面は、ユーザが道具を使い、ユーザと物体とのインタラクションがユーザの使用道具を介して行われる場面である。例えば、調理や食事の場面においては人と食材の接触は箸やトング、包丁を介する。また、農作業の収穫場面においては、人と収穫作物の接触は収穫ばさみを介する。この他にも狩猟や衣類作成、住居建築、絵画や学習用記録などの多様な活動において、人は道具を介して対象の物体とのインタラクションを行っている。これらの場面において人がインタラクションの対象となる物体を認識できれば有益なサービスに繋がると考えている。例えば、食材の使用ログを残すことができ、また、食材から調理段階を判断して、調理の手順や手本を示すなどの調理補助への応用も考えられる。

提案手法は次のような特徴から、この対象場面において有用と考えている。対象場面では、認識したい物体に認識のためのセンサなどをつけることが適さない場合がある。例えば、調理や食事の際の食材や、農作業時の収穫作物などに対して、電子的なタグなどを事前につけることは困難である。提案手法はユーザの使用道具だけにセンサを取り付けるだけで、物体認識を可能にすることを狙っている。ただし、認識したい物体に使用道具が接触する場面でのみ有用である。また、提案手法はユーザの使用道具が物体に接触するだけで物体を認識できる。そのため、ユーザは、物体の認識にあたって特別な動作などを行う必要がなく、通常通りに振る舞うだけで物体認識ができる。オクルージョンの問題が発生しないことも提案手法の利点である。例えばカメラであれば、認識すべき物体に人が被るせいでカメラ映像に物体が映らないオクルージョンの問題が起こる場面があるが、提案手法ではユーザの使用道具が物体に必ず接触することを利用しているためオクルージョンの問題が起きない。これに伴い、提案手法では、カメラのように心理的負荷を生み得るセンサは使用せず、また、人体にセンサを直接設置もしないため、心理的、身体的負担も低いと考えている。提案手法以外の方法は、これらの観点の

いずれかに欠点があると考えているため、ユーザが道具を使い、ユーザと物体とのインタラクションがユーザの使用道具を介して行われる場面では提案手法が有効であると考えている。

3.2 センシング原理

提案手法は物体のもつ音響特性を利用して、アクティブ音響センシング技術を用いて物体認識を行う。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、物体の持つ形状、材質、境界条件により決定される。本研究では、音響特性を決定する要素の一つである、物体の持つ材質の違いに着目する。今回の認識対象である食材において材質は栄養成分を指す。食材ごとに栄養成分の種類と含有量が異なることはわかっており、各食材の栄養成分の違いが、各食材の音響特性の違いを生むと考えられる。栄養成分としては、例えば、水分、炭水化物、タンパク質、脂肪、ビタミン、ミネラルなどがある。そして、この音響特性の違いが、食材を伝搬して得られる音の周波数特性に表れると考えている。

本研究では、この各食材から得られる音の周波数特性の認識のために、アクティブ音響センシングを用いる。具体的には次のようになる。まず、ユーザの使用道具に音を発するコンタクトスピーカと音を計測する圧電素子のコンタクトマイクを装着する。そして、様々な周波数の音を含むスイープ信号を振動スピーカから発し、使用道具が接触した物体を様々な周波数域において振動させ、物体を伝搬した音をピエゾマイクで取得する。そして、計測した音の周波数特性をFFT(高速フーリエ変換: Fast Fourier Transform)で解析し、各物体の音響特性を記録したデータベースをもとに、機械学習を用いて各物体の認識を行う。

4. 実装

本稿では、調理や食事といった道具を用いて食材とインタラクションをする場면을対象にして、プロトタイプシステムを実装した。また、ユーザの使用道具として、食材を挟むトング型デバイスと、食材を切る包丁型デバイスを実装した。

4.1 ハードウェア

プロトタイプシステムの構成を図1に示し、実装した全体図を図2に示す。プロトタイプシステムはコンタクトスピーカ、コンタクトマイクロフォン、信号処理と機械学習を行うコンピュータによって構成される。物体を振動させるコンタクトスピーカには、スライブ社のOMR20F10H-BP-310を使用した。その形状は、直径21[mm]、厚さ0.3[mm]以下の超薄型軽量の円盤形のものである。1[kHz]から40[kHz]以上の音を出力することができ、人には聞こえない周波数域

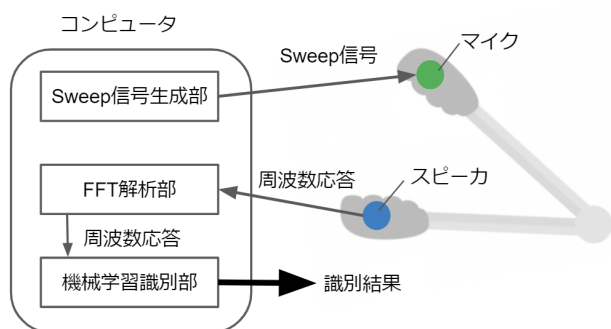


図 1 システム概要

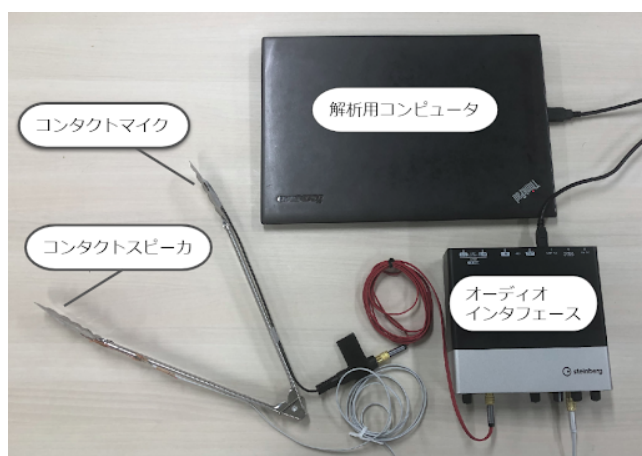


図 2 プロトタイプシステムの全体図

まで出力することが可能である。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ロフォンには、Shadow社のSH711 ピエゾタイプのピックアップを用いた。その形状は、トランデュサ径 20[mm]、ジャック 1/4 標準サイズ、厚さ 2[mm] の円盤型である。使用したコンタクトスピーカとコンタクトマイクロフォンを図 3 に示す。コンピュータとして、レノボ・ジャパン株式会社の ThinkPad X1Carbon (CPU : Intel Core i7-5600 2, 60GHz, RAM : 8.00GB) を使用した。コンピュータでは、コンタクトスピーカから発せられるアクティブ音響信号の作成、コンタクトマイクロフォンから得られた応答の入力、並びに FFT 解析による周波数特性の取得、機械学習を行う。音の入出力は USB 接続のオーディオインターフェース (Steinberg 社の UR 22mk) を介して行う。

食材を挟むトング型デバイスを図 4 に示す。トングには全長 30[cm] のすべてステンレスで作られた一般的なものをを用いた。マイクの装着位置については、物体を通過した応答信号を取得したいため、マイクとスピーカが食材を挟むことを考え、トングの 2 か所ある先端部分にそれぞれマイクとスピーカを装着した。食材を切る包丁型デバイスを図 5 に実装した。包丁にはハイカーボンステンレス製の切味工房を使用した。持ち手の部分は木製の約 290[mm] × 45[mm] × 14[mm] のサイズである。マイクの装着位置に



図 3 コンタクトマイクとコンタクトスピーカ

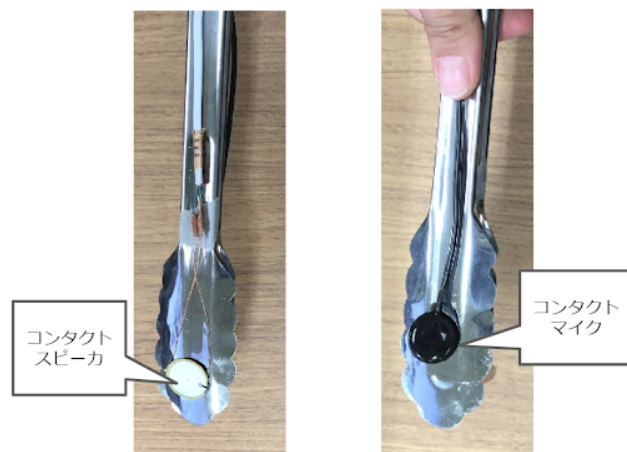


図 4 トングのセンサの取り付け位置



図 5 包丁のセンサの取り付け位置

については、物体を通過した応答信号を取得したいため、マイクとスピーカが食材を挟むことを考え、包丁の刃の前後に来るよう先端と付け根部分に装着した。

4.2 解析ソフトウェア

ソフトウェアはスイープ信号生成部分、FFT 解析部分、機械学習部分の 3 つのモジュールから成る。実装には、Audacity, python, weka (Waikato Environment for Knowledge Analysis) [18] を用いた。まず、スイープ信号生成部分では、200[ms] 間に 1[Hz] から 40000[Hz] まで直線的に音に変化する Sweep 信号を生成する。この周波数域は、先行研究 [12] の論文を参考に周波数域を決定した。人の可聴域は 1-20000[Hz] までであるため、それに加えて、人の非可聴域である 20000[Hz] から 40000[Hz] までの範囲を選択した。サンプリング周波数は 96.0[kHz] である。Sweep



図 6 実験 1 に用いた野菜

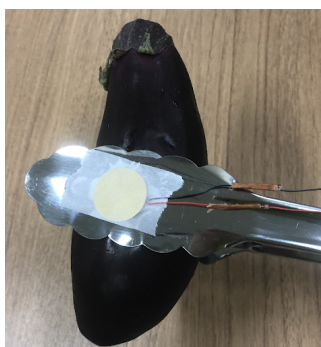


図 7 野菜をトング型デバイスで持っている様子

信号は Chirp 信号とも呼ばれる。音を繰り返して流すと周波数ギャップでインパルスノイズが発生するので、それ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。次に、FFT 解析部分では、ピエゾマイクから取得した振動応答を、FFT を用いて周波数応答に変換する。振動応答の取得の際のサンプリング周波数は 96.0[kHz] とし、32bit 浮動小数点でサンプリングした。コンピュータで FFT を行い、1-40000[Hz] の範囲の周波数スペクトルを得る。最後に機械学習部分では、Support Vector Machine (SVM) を用いて周波数特性から食材の識別を行う。

5. 評価実験

提案手法による野菜の認識率の検証を目的として評価実験を 3 つ行った。実験 1, 2 ではトング型デバイスを評価し、実験 3 では包丁型デバイスを評価した。

5.1 実験 1

この実験 1 では、挟んで持つタイプのトング型デバイスを用いて、カットなどがされていないそのままの野菜を挟むことで、挟んだ野菜の種類が認識できるかを評価した。

5.1.1 実験環境

図 6 に実験で認識する 7 種類の野菜を示す。今回使用したのは、しいたけ、ジャガイモ、トマト、ピーマン、ニンジン、なすび、玉ねぎの 7 種類を各 5 個体ずつである。これらは、一般的に利用される野菜として選択した。

表 1 実験 1 における各野菜の認識結果

野菜の種類	適合率	再現率	F 値
トマト	0.92	0.91	0.91
人参	0.86	0.89	0.87
ナス	0.85	0.85	0.85
ピーマン	0.93	0.88	0.90
玉ねぎ	0.87	0.89	0.88
ジャガイモ	0.90	0.86	0.88
しいたけ	0.86	0.90	0.88

表 2 実験 1 における認識結果のコンフュージョンマトリックス

トマト	265	3	9	1	0	5	7
人参	3	258	15	3	6	4	1
ナス	6	8	249	2	6	14	5
ピーマン	2	6	4	256	9	4	9
玉ねぎ	7	4	5	2	260	0	12
ジャガイモ	1	13	4	4	8	252	6
しいたけ	3	6	6	5	7	1	262

データ取得は次のように行った。実験者 1 名がトング型デバイスを使用して、野菜を約 60 秒間挟み続けることを、合計 35 個体の野菜すべてに対して行った。図 7 に野菜を持っている様子を示す。各野菜につき、約 60 秒間のデータを取得した。1 つのサンプルデータは長さ 2 秒をウィンドウとし、約 60 秒のデータに対してウィンドウを 50 % オーバラップのスライディングウィンドウで FFT を計算することで、全野菜で合計 2030 サンプルを用意した。機械学習の分類器には、SVM を使用した。得られた 2030 サンプルに対して、10 分割交差検証によって食材の分類精度を評価した。

5.1.2 評価結果

野菜の種類ごとの分類結果を表 1 に示し、コンフュージョンマトリックスを表 2 に示す。全サンプルデータに対する正答率は 88 % であった。表 1 から、どの種類の野菜の分類結果も F 値 0.85 以上となり、すべての種類を認識できていた。以上から、カットなどがされていないそのままの野菜を挟む条件において、提案手法の有効性を確認した。

5.2 実験 2

実験 2 では、挟んで持つタイプのトング型デバイスを用いて、カットされた野菜を挟むことで、挟んだ野菜の種類が認識できるかを評価した。

5.2.1 実験環境

使用した野菜はしいたけ、トマト、ピーマン、ナスの 4 種類を 1 個体ずつである。

データ取得は次のように行った。実験者 1 名がトング型デバイスを使用して、野菜を約 30 秒間挟み続けることを、合計 4 個体の野菜すべてに対して行った。図 8 にカットされた野菜の様子を示す。カットは個体全体を 1[cm] 単位でスライスした。そのため、1 つの個体から取得できるカッ



図 8 実験 2 でカットされた野菜の様子

表 3 実験 2 における各野菜の認識結果

野菜の種類	適合率	再現率	F 値
トマト	0.99	0.99	0.99
ナス	1.00	1.00	1.00
ピーマン	0.98	0.98	0.98
しいたけ	0.97	0.97	0.97

表 4 実験 2 における認識結果のコンフュージョンマトリックス

トマト	195	0	1	0
ナス	0	280	0	0
ピーマン	1	0	192	3
しいたけ	1	0	2	137

トされた被検体の数は異なる。各被検体につき、約 30 秒間のデータを取得した。1 つのサンプルデータは長さ 2 秒をウィンドウとし、約 30 秒のデータに対してウィンドウを 50 % オーバーラップのスライディングウィンドウで FFT を計算することで、1 つの被検体につき 28 サンプル、合計 812 サンプルを用意した。機械学習の分類器には、SVM を使用した。得られた 812 サンプルに対して、10 分割交差検証によって食材の分類精度を評価した。

5.2.2 評価結果

野菜の種類ごとの分類結果を表 3 に示す。コンフュージョンマトリックスを表 4 に示す。全サンプルデータに対する正答率は 99 % であった。表 3 から、どの種類の野菜の分類結果も F 値 0.97 以上となり、すべての種類を認識できていた。以上から、カットがされた野菜を挟む条件において、提案手法の有効性を確認した。

5.3 実験 3

実験 3 では、切るタイプの包丁型デバイスを用いて野菜をカットする条件で、カットした野菜の種類が認識できるかを評価した。

5.3.1 実験環境

図 9 に実験で認識する 7 種類の野菜を示す。今回使用したのは、しいたけ、ジャガイモ、トマト、ピーマン、ニンジン、なすび、玉ねぎの 7 種類を各 1 個体ずつである。

データ取得は次のように行った。実験者 1 名が包丁型デ



図 9 実験 3 に用いた野菜



図 10 実験 3 で野菜をカットしている様子

バイスを使用して、一般的な薄切りの要領で野菜を約 30 秒間刻み続けることを、合計 7 個体の野菜すべてに対して行った。図 10 に野菜をカットしている様子を示す。各野菜につき、約 30 秒間のデータを取得した。1 つのサンプルデータは長さ 2 秒をウィンドウとし、約 30 秒のデータに対してウィンドウを 50 % オーバーラップのスライディングウィンドウで FFT を計算することで、全野菜で合計 420 サンプルを用意した。機械学習の分類器には、SVM を使用した。得られた 420 サンプルに対して、10 分割交差検証によって食材の分類精度を評価した。

5.3.2 評価結果

野菜の種類ごとの分類結果を表 5 に示す。全サンプルデータに対する正答率は 92 % であった。表 5 から、どの種類の野菜の分類結果も F 値 0.80 以上となり、すべての種類を認識できていた。以上から、野菜をカットしていく条件において、提案手法の有効性を確認した。

5.4 考察

提案手法の結果から、提案手法が道具を介した物体認識を行うために有効であることを確認した。3 つの実験において、トングを用いて野菜をそのまま認識したもの、カットされた野菜をトングに挟んで認識したもの、野菜のカットする様子を包丁で認識したものの 3 点はどれも 90 % 程

表 5 実験 3 における各野菜の認識結果

野菜の種類	適合率	再現率	F 値
キャベツ	1.00	0.77	0.87
レタス	0.63	0.70	0.66
ナス	1.00	0.83	0.90
ジャガイモ	1.00	1.00	1.00
人参	0.90	0.93	0.92
大根	1.00	1.00	1.00
玉ねぎ	0.76	1.00	0.86

度の認識率を得ることができた。包丁型デバイスのような野菜との接触時間が短い道具においても野菜を認識できたことから、道具と認識対象との接触時間に関わらず、対象とする物体に固有の音響特性があれば、提案手法が有効に働くと考えられる。

また、今回選択した野菜以外にも、調理場面には様々な食材が存在する。よって、今後は、提案手法で対象となった食材以外でも提案手法が有効かを調査するために、今回選択した以外の野菜や、野菜以外の食材を利用して実験を行う。今回、物体の周波数特性に影響を与える要因として物体の形状や境界条件が主に考えられる。そのため、同じ個体であっても、切り方、持ち方、持つ部位を変化させたときに、認識できるのか、今後はこれらを考慮した実験を行う。

提案手法は、認識すべき物体に接触する道具をユーザが扱う場面で汎用的に利用できるため、今回の実験以外の目的にも応用できると考えている。例えば、菜箸やお玉などの調理器具類や、箸やフォークなどの食器類に、提案手法を適用することで、調理や食事の際の食材の認識に利用できる。また、農家の収穫場面において収穫鉢に提案手法を適用することや、運搬に使用されるロボットアームなどに仕込むことで収穫作物の認識に利用できると考えている。

6. おわりに

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。提案手法は、ユーザが使用する道具にスピーカとマイクを取り付け、道具が物体に接触した際に、物体を伝搬した音響信号を取得し、その音響信号の周波数特性を解析することで、物体を認識した。本稿では、調理や食事といった道具を用いて食材とインタラクションをする場면을対象にして、食材を挟むトング型デバイスと、切る包丁型デバイスを実装した。そして、提案手法の有効性を評価する実験を 7 種類の野菜をもとに 3 つ行い、いずれの実験においても、88 % 以上の精度で野菜の種類を認識できることを確認した。今後は、本稿とは異なる形状の道具に提案手法を適用させた評価や、野菜以外の対象の認識に提案手法が有効であるかを評価する予定である。

謝辞 本研究の一部は、JSPS 科研費若手研究 JP19K20330 によるものである。ここに記して謝意を表す。

参考文献

- [1] Boukraa Mustapha, and Shigeru Ando. Tag-based vision: assisting 3D scene analysis with radio-frequency tags. In Proceedings of the 5th International Conference on Information Fusion, ICIF' 02, Vol. 1, pp. 412–418, July 2002.
- [2] Jin-Young Kim, Chang-Jun Im, Sang-Won Lee and Ho-Gil Lee. Object Recognition using Smart Tag and Stereo Vision System on Pan-Tilt Mechanism. In Proceedings of the 12th International Conference on Control, Automation Systems, ICCAS' 05, pp.2379–2384, June 2005.
- [3] Martin Kaltenbrunner, and Ross Bencina. reactIVision: a computer-vision framework for table-based tangible interaction. In Proceedings of the 1st international conference on Tangible and embedded interaction, TEI ' 07, pp. 69–74, February 2007.
- [4] Chris Harrison, Robert Xiao, and Scott E. 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 '12, pp. 563–568, October 2012.
- [5] Anhong Guo, Anuraag Jain, Shomiron Ghose, Gierad Laput, and Jeffery p. Bigbam. Crowd-AI Camera Sensing in the Real World. In Proceedings on the 2018 ACM Interactive, Mobile, Wearable Ubiquitous Technologies, Vol. 2, No. 3, Article NO 111, September 2018.
- [6] 安岡里都, 蘭頭元春, 飯山将晃. ドローン搭載カメラ画像を用いた物体認識. ELCAS2018 Journal, Volume3, pp.85–87, 2018.
- [7] 塩濱教幸, 深谷健. SIFT を用いた特定物体認識の高速化と移動ロボットによる物体ハンドリングへの応用. 北海学園大学大学院工学研究科紀要 (12), pp. 25–30, 2012.
- [8] Brian Amento, Will Hill, Loren Terveen. The sound of one hand: a wrist-mounted bio-acoustic fingertip gesture interface. In Proceedings of the 2002 ACM SIGCHI International Conference on Human Factors in Computing Systems, CHI '02, pp. 724–725, April 2002
- [9] Chris Harrison, Scott E. Hudson. Scratch Input: Creating Large, Inexpensive, Unpowered and Mobile Finger Input Surfaces. In Proceedings of the 21th annual ACM symposium on User interface software and technology, UIST '08, pp. 205–208, October 2008
- [10] Chris Harrison, Desney Tan, Dan Morris. Skinput: appropriating the body as an input surface. In Proceedings of the 2010 ACM SIGCHI International Conference on Human Factors in Computing Systems, CHI '10, pp. 453–462, April 2010
- [11] Roderick Murray-Smith, John Williamson, Stephen Hughes, Torben Quaade. Stane: synthesized surfaces for tactile input. In Proceedings of the 2008 ACM SIGCHI International Conference on Human Factors in Computing Systems, CHI '08, pp. 1299–1302, April 2008
- [12] 大野誠, 志築文太郎, 田中二郎. アクティブ音響センシングを用いた把持状態認識. 一般社団法人情報処理学会, インタラクション' 13, pp. 5663, 一般社団法人情報処理学会, 2013.
- [13] Makoto Ono, Buntarou Shizuki, and Jiro Tanaka. Touch and Activate: Adding Interactivity to Existing Objects using Active Acoustic Sensing. In Proceedings of the 26th annual ACM symposium on User interface software and technology, UIST' 13, pp. 31–40, October 2013.

- [14] 河田大史, 寺田 努, 塚本昌彦. 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. 情報処理学会マルチメディア, 分散, 協調とモバイルシンポジウム, DICOMO' 15, 論文集, pp. 400–406, 2015.
- [15] Adiyana Mujibiyana, Xiang Cao, Desney S. Tan, Dan Morris, Shwetak N. Patel, and Jun Rekimoto. The sound of touch: on-body touch and gesture sensing based on transdermal ultrasound propagation. In Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces, ITS' 13, pp. 189–198, October 2013.
- [16] Kai Kunze and Paul Lukowicz. Symbolic Object Localization Through Active Sampling of Acceleration and Sound Signatures. In proceedings of the 9th International Conference on Ubiquitous Computing, UbiComp' 07, pp. 163–180, September 2007.
- [17] 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. 一般社団法人情報処理学会, インタラクション' 18, pp. 6271, 一般社団法人情報処理学会, 2018.
- [18] Weka Machine Learning Toolkit, <http://www.cs.waikato.ac.nz/ml/weka/>.