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

西井　遥菜 *　　双見　京介 *　　村尾　和哉 *†

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．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，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．食事場面を対象とし，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．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，書くなど多くの機能をもつようになり，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さまざま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々は，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．

このような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．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ともなって発展しており，例えば，IC タグを用いた物体認識には，RF タグ [1] や，IRED デバイスを用いたスマートタグ [2] がある．また，カメラを用いた物体認識技術には，Zensor++[3] や，光ストライプ投射法 [10] がある．ほかにも，アクティブ音響センシングを用いた物体認識手法 [9] などがある．こういった物体認識技術を利用することで，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，認識した物体に応じたサービスを提供するコンテクストアウェアシステムへの利用など，さまざまなユーザ支援に利用されている．

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．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，ユーザが使用する道具が認識したい物体に接触することと，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．

提案手法の有効性を検証するために，本研究では食事場面を対象として，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 [4][5]．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 [6]．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 [7]．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

Copyright is held by the author(s).
* 立命館大学大学院情報理工学研究科
† 科学技術振興機構さきがけ

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 [8]．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 [9]．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，食事場面を想定し，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に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．

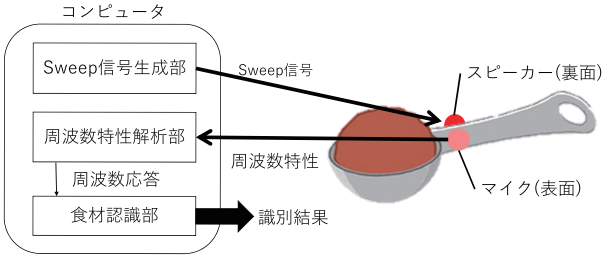

図 1: 提案システム概要

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 [4] 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材にも触れていないときの特徴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．まずはじめに，実験で使用するカトラリーの実装の説明およびスピーカとマイクの位置に関する予備実験を行う．その次に，予備実験の結果をもとに最適な位置にスピーカとマイクを取り付けたカトラリーを用いて 2 つのシナリオ（シナリオ 1 とシナリオ 2）で 14 種類の野菜に対して認識精度を評価する．シナリオ 1 は，各野菜につき複数の個体や複数種類のカットされた形を学習して，個体や形の違いに関わらず野菜の種類を認識できるかを評価した．シナリオ 2 は，野菜の種類だけでなく，カットされた形も正解ラベルに含めて，野菜の種類と切り方を同時に認識できるかを評価した．

[1] https://www.audacityteam.org/

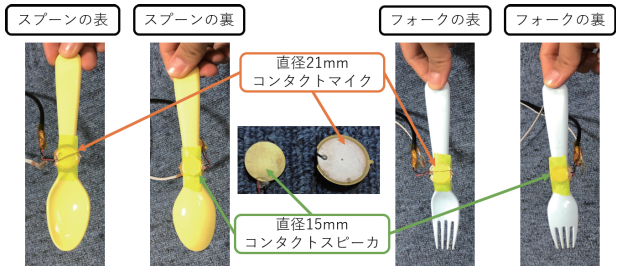

図 2: マイクとスピーカを取り付けたフォークとスプーン

4.1 スピーカとマイクの位置に関する予備実験

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ダ（FGT-15T-6.0A1W40）を使用した．直径 15mm，厚さ 0.3mm 以下の超薄型軽量な円盤形である．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ブ社の OMR20F10H-BP-310 を使用した．直径 21mm，厚さ 0.3mm 以下の超薄型軽量な円盤形である．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．

スピーカとマイクの位置の候補として，フォークやスプーンの先端，柄，柄尻の 3 か所が考えられるが，フォークやスプーンの先端にセンサを取り付けると，カトラリーとして機能しないため候補から除外した．したがって，スピーカとマイクの位置の組合せはフォークやスプーンの柄の表裏に取り付ける，柄と柄尻に取り付ける，柄尻の表裏に取り付ける 3 通りが考えられる．柄と柄尻に取り付ける場合に，マイクを柄に付けるか，柄尻に付けるかで 2 通りあるが，予備実験では柄にマイク，柄尻にスピーカを取り付けた．スピーカとマイクを 3 通りの方法で取り付けたフォークおよびスプーンを用いて，人参，キュウリ，サツマイモの 3 種類の野菜を認識した．機械学習モデルには，SVM(多項式カーネル）を使用した．フォーク型デバイスは，フォークの先端で野菜を刺した状態で振動データを取得した．スプーン型デバイスは，スプーンの先端に野菜をのせた状態で振動データを取得した．各野菜につき約 60 秒のデータを取得し，長さ 1 秒のウインドウをスライド幅 100%で切り出して FFT し，1 種類の野菜から 50 サンプル特徴量を抽出した．1 種類の野菜につき

表 1: スピーカとマイクの設置位置を変えたときのフォーク型デバイスとスプーン型デバイスの F 値

野菜の種類	mic=柄 spk=柄	mic=柄 spk=柄尻	mic=柄尻 spk=柄尻
人参	0.958	0.800	0.794
キュウリ	0.948	0.760	0.789
サツマイモ	0.906	0.775	0.746
人参	0.997	0.828	0.778
キュウリ	0.844	0.841	0.748
サツマイモ	0.783	0.669	0.829

3 個体分のデータを採取したため，全部で 1350 サンプルのデータ（50 ウィンドウ ×3 種類の野菜 ×3 個体 ×3 種類の取り付け位置）を得た．

フォーク型デバイスおよびスプーン型デバイスを用いたときの認識結果の F 値を表 1 に示す．結果より，マイクとスピーカをともに柄に取り付けたフォーク型デバイスにおいて平均 F 値 0.937，スプーン型デバイスにおいて平均 F 値 0.874 と最も高い精度を示した．よって以降の評価実験では，柄の部分の表と裏にスピーカとマイクを取り付けたフォークとスプーンを使用する．

4.2 評価実験で使用する野菜

評価実験では一般的に利用される野菜を選択するために，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[2] を参考にして，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 2 に示す 3 種類の形に切った．一例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，

[2] https://www.mhlw.go.jp/stf/houdou/0000096138.html

表 2: 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

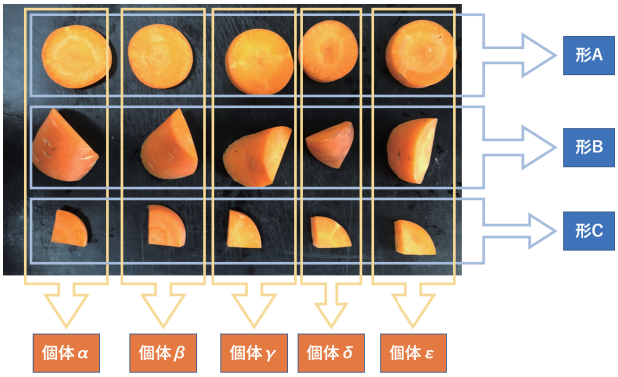

図 3: 個体と切り方の違いによる実験素材

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 秒のウインドサイズで FFT を計算し，1 素材あたり特徴量を 50 サンプル得た．その結果，10750 サンプルデータ（50 ウインドウ ×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．

4.4 シナリオ 1

シナリオ 1 では，野菜の個体やカットされた形に関わらず，野菜の種類を認識できるかを評価した．具体的には，個体の違いとカットされた形の違いに関わらず，同一の種類の野菜には同一の野菜ラベルを付けた．したがって，野菜に接触していない状態を含めてラベルは 15 クラスである．機械学習モデルは SVM，Random Forest，Naive Bayes で試したところ，最も性能の良かった SVM（多項式カーネル）を用いて 5-fold 個体別交差検証を行って比較した．つまり，ニンジンの個体 α，個体 β，個体 γ，

表 3: シナリオ 1 における各野菜の認識精度

	フォーク型デバイス			スプーン型デバイス		
野菜の種類	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

個体 δ を学習データとして学習モデルを構築してニンジンの個体 ϵ 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 3 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 シナリオ 2

シナリオ 2 では，提案手法が野菜の種類とカットされた形の両方を認識できるか評価した．具体的には，野菜の種類と形をラベルとした．例えば，ニンジンはニンジン A，ニンジン B，ニンジン C の 3 種類のラベルが存在し，ラベルは全部で 14 種類の野菜 ×3 種類の形 + デバイスと食材が接触していない 1 状態の合計 43 クラスである．機械学習モデルはシナリオ 1 と同様に SVM を用いた．

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 4 に示す．フォーク型デバイスでは F 値 0.532，スプーン型デバイスでは F 値 0.599 を示した．

4.6 考察

シナリオ 1 の 15 種類の野菜および状態の認識結果から，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考えられる．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

表 4: シナリオ 2 における各野菜の認識精度

	フォーク型デバイス			スプーン型デバイス		
野菜の種類と形	適合率	再現率	F 値	適合率	再現率	F 値
キャベツ_A	0.606	0.664	0.634	0.708	0.756	0.731
キャベツ_B	0.537	0.440	0.484	0.559	0.608	0.582
キャベツ_C	0.560	0.620	0.588	0.896	0.552	0.683
人参_A	0.292	0.276	0.284	0.745	0.840	0.789
人参_B	0.238	0.380	0.292	0.392	0.284	0.329
人参_C	0.246	0.268	0.257	0.397	0.544	0.459
白菜_A	0.702	0.916	0.795	0.718	0.988	0.832
白菜_B	0.138	0.052	0.076	0.870	0.616	0.721
白菜_C	0.337	0.352	0.344	0.806	0.896	0.848
キュウリ_A	0.927	0.808	0.863	0.533	0.384	0.447
キュウリ_B	0.865	0.868	0.866	0.488	0.644	0.555
キュウリ_C	0.963	0.940	0.951	0.698	0.812	0.750
ピーマン_A	0.593	0.768	0.669	0.427	0.376	0.400
ピーマン_B	0.652	0.584	0.616	0.352	0.376	0.364
ピーマン_C	0.471	0.388	0.425	0.293	0.316	0.304
レタス_A	0.312	0.388	0.346	0.248	0.208	0.226
レタス_B	0.361	0.316	0.337	0.451	0.668	0.539
レタス_C	0.468	0.520	0.492	0.190	0.224	0.206
玉葱_A	0.776	0.720	0.747	0.512	0.412	0.457
玉葱_B	0.773	0.600	0.676	0.446	0.496	0.470
玉葱_C	0.902	0.480	0.627	0.539	0.576	0.557
ジャガイモ_A	0.164	0.208	0.183	0.864	0.612	0.717
ジャガイモ_B	0.301	0.252	0.275	0.364	0.348	0.356
ジャガイモ_C	0.289	0.288	0.289	0.529	0.512	0.520
カボチャ_A	0.765	0.860	0.810	0.626	0.776	0.693
カボチャ_B	0.618	0.596	0.607	0.976	0.804	0.882
カボチャ_C	0.406	0.492	0.445	0.762	0.744	0.753
大根_A	0.495	0.596	0.541	0.522	0.484	0.502
大根_B	0.467	0.372	0.414	0.384	0.412	0.398
大根_C	0.434	0.672	0.527	0.725	0.716	0.720
ネギ_A	0.431	0.496	0.461	0.427	0.376	0.400
ネギ_B	0.679	0.668	0.673	0.461	0.308	0.369
ネギ_C	0.461	0.332	0.386	0.324	0.400	0.358
ほうれん草_A	0.703	0.548	0.616	0.639	0.652	0.646
ほうれん草_B	0.950	0.756	0.842	0.863	0.656	0.745
ほうれん草_C	0.694	0.800	0.743	0.748	0.688	0.717
サツマイモ_A	0.655	0.852	0.741	0.971	0.660	0.786
サツマイモ_B	0.330	0.232	0.272	0.809	0.900	0.852
サツマイモ_C	0.459	0.408	0.432	0.761	0.944	0.843
トマト_A	0.613	0.760	0.679	0.803	0.848	0.825
トマト_B	0.354	0.280	0.313	0.808	0.824	0.816
トマト_C	0.277	0.228	0.250	0.681	0.572	0.622
非接触状態	1.000	1.000	1.000	1.000	1.000	1.000
平均	0.541	0.536	0.532	0.612	0.600	0.599

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の認識ができると考えられる．

シナリオ 2 の 43 種類の野菜と切り方および状態の認識結果から，フォーク型デバイスは F 値 0.532，スプーン型デバイスは F 値 0.599 であった．この結果より，野菜を認識する際に形まで認識することは提案手法では困難であるといえる．シナリオ 1 とシナリオ 2 の結果より，野菜の種類の違いによる音響特性の違いは，野菜の形状の違いによる音響特性の違いよりも大きく，野菜を切ったときの個々の片の形状の違いは野菜の種類の認識に大きく影響しないといえる．

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．

提案手法は食事や調理以外の場面で活用できる．例えば，組立て作業において，ねじやナットはドライバやペンチなどの道具を用いて取り付けられる．写真で視覚的に指示をされても，似たねじを間違えることもあるが，提案手法を工具に適用することで，組立作業の工程において誤ったねじや部品を取り付けたことを検出できる．また，農作物の収穫において，収穫鋏や運搬に使用されるロボットアームなどに提案手法を適用することで，収穫物の種類や成熟度，サイズを判断して自動で仕分けして，収穫や出荷作業の負担を削減できると考える．

4.7 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．また，マイクやスピーカがカトラリーから外れて食事時に誤飲する可能性を避けるため，また洗浄に耐えるため，カトラリーの内部にデバイスを組み込むことが望ましい．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．今回の結果から，野菜の認識には野菜の形ではなく組成が効いていると考えられるが，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．食事場面を対象とし，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 クラスに対して個体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．また，14 種類の野菜を 3 種類の切り方で切り，42 種類の野菜の種類と形および食材が接触していない状態の計 43 クラスに対して個体別交差検証を行った結果，フォーク型デバイスにおいて F 値 0.532，スプーン型デバイスにおいて F 値 0.599 を得た．これにより，野菜の切り方の違いまでを認識することは困難であるが，切り方のばらつきを吸収して野菜の種類を高精度で認識できることがわかった．

参考文献

[1] B. Mustapha et al., Tag-based vision: assisting 3D scene analysis with radio-frequency tags, ICIF'02, Vol. 1, pp. 412–418 (2002).

[2] J. Kim et al., Object Recognition using Smart Tag and Stereo Vision System on Pan-Tilt Mechanism, ICCAS'05, pp. 2379–2384 (2005).

[3] A. Guo at al., Crowd-AI Camera Sensing in the Real World. IMWUT, Vol. 2, No. 3, Article. 111 (2018).

[4] 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).

[5] M. Ono at al., Touch and Activate: Adding Interactivity to Existing Objects using Active Acoustic Sensing. UIST'13, pp. 31–40 (2013).

[6] 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).

[7] A. Mujibiya et al., The sound of touch: on-body touch and gesture sensing based on transdermal ultrasound propagation. ITS'13, pp. 189–198 (2013).

[8] K.Kunze et al., Symbolic Object Localization Through Active Sampling of Acceleration and Sound Signatures. UbiComp'07, pp. 163–180 (2007).

[9] 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).

[10] O. Ozeki et al., Real-Time Range Measurement Device for Three Dimensional Object Recognition. IEEE TPAMI, Vol. 8, No. 4, pp. 550–554 (1986).