

特別研究報告書

調理加工に起因する振動音を用いた食材識別

指導教官 美濃 導彦 教授

京都大学工学部情報学科

三功 浩嗣

平成 18 年 2 月 10 日

調理加工に起因する振動音を用いた食材識別

三功 浩嗣

内容梗概

調理は日常的に行なわれる複雑な作業であるので、調理者の状況に応じて適切なアドバイスを行なうようなシステムがあれば便利である。このようなシステムを実現するためには、レシピ中のどの食材に対して調理を行なっているのかを識別することが必要となる。調理において、最も基本的な加工の一つは「切る」であるので、本研究では、調理者が「切る」という加工を施す際、その対象がレシピ中のどの食材であるのかを識別することを目標とする。

従来、調理者が扱っている食材を識別する際には、カメラにより調理を観測した映像から抽出した食材の視覚的特徴を用いてきた。しかし、調理では視覚的特徴のよく似た食材が同時に複数使われる場合も多く、識別がうまくいかない場合も少なくない。

これに対して、調理台を伝播する振動音は包丁が食材の表面を裂き、内部を通過する際に発生する振動を含むことから、食材の内部構造や大きさによって変化すると考えられる。そこで、食材を「切る」加工に起因する振動音波形から、食材固有の特徴を選出し、食材を識別する手法を提案する。

振動音を通常のマイクで取得したのでは、声や換気扇の音など、食材を切ることで生じる振動音とは無関係の音による影響が避けられない。そこで、まな板を置く作業台を振動が伝達しやすい硬化ガラスにし、その裏にコンタクトマイクを貼り付けることで、振動音を直接取得する。

食材を切る加工に起因する振動音は、まず包丁が食材表面に切れ込みを入れる際のピークが現れ、内部を進んでいく際の振動が続き、まな板に衝突した時点で最も大きなピークが生じるといった形を基本としている。そこで、切れ込み時の振幅、衝突時の振幅、食材内部を通過している間の振幅の3つが食材固有の特徴を反映しているのではないかと考えた。

まず、切れ込み時の振幅及び、衝突時の振幅が食材の堅さを反映していると考え、様々な堅さの食材についてこの値の分布の仕方を調べた。その結果、同じ食材でもこれら2つの値は広く分散しており、他の食材との重なりが大きく、この2つの特徴量だけでは食材の識別は困難であることがわかった。

次に、包丁が食材内部を通過している間の波形として、包丁がまな板に衝突

する直前の部分に注目した。包丁がまな板に衝突した時刻を、振動音波形に閾値を設けることにより検出し、その直前 0.2 秒間分を分析用セグメントとして抽出した。様々な種類の食材を切る振動音から得られた分析用セグメントに対してスペクトログラム分析を行ない比較したところ、食材の種類に依存した特徴が低周波部分に現れていることがわかった。できるだけ次元数が少なく、識別能力の高い特徴ベクトルを選定するため、様々な周波数解像度でスペクトログラムを計算し、低周波部分の特徴ベクトルとして抽出し、食材の種類の識別能力を調べた。その結果、周波数解像度 128 で取得したスペクトログラムの低周波部 8 次元の全時刻における平均ベクトル及び、分散ベクトルを結合し、16 次元の特徴ベクトルとして採用したものがこの条件を満たすことがわかった。

以上で述べた観測特徴による食材の識別性能を調べるため以下のような実験を行なった。まず、食材ごとに複数個の個体を用意し、各個体を切った際の振動音波形から、分析用セグメントの集合を抽出する。次に、各分析用セグメントを上で述べた 16 次元の特徴ベクトルに変換し、SVM(Support Vector Machine)を用いて特徴空間を食材の種類ごとに分割することで識別機を構成した。

まず、多くのレシピに登場する代表的な食材であるキャベツ、人参、キュウリ、玉葱、ピーマン、ジャガイモ、トマトについて本手法により識別を行なった。抽出した分析用セグメント全てにより学習した識別機を用いた場合は各食材とも 75.0%以上の識別率であった。さらに、未知の分析用セグメントの識別能力を調べるため個体ごとに交差検定を行なった。この結果、トマトの識別率は平均 17.8%と低かったが、他の食材については平均 52.9%以上で正しく識別できた。また、個体から抽出された一連の分析用セグメント集合の中で識別結果の多数決をとり、食材を決定すれば、トマト以外の食材は平均 88.9%の精度で識別できることがわかった。

次に、映像等により食材の候補が絞りこめた場合を想定し、同系色の食材の組み合わせであるジャガイモと玉葱の組、キュウリとピーマンの組を取り上げ、識別実験を行なった。この結果、各食材とも 81.4%以上の識別率を示した。これにより、本手法と、映像を併用することにより、識別率を向上させることが可能であることが示された。

より多くの個体について識別率の変化を観測することで、より頑強な識別機を構成していくことが今後の課題である。

Food Product Classification by Using a Vibration Sound made by Cutting Action

Hiroshi SANKOH

Abstract

Because cooking is complicated work on a daily basis, such system is expected that gives the advices to the cooker according to his/her situation. For realizing such system, we address the problem of classifying the name of the food product the cooker cooks from the list of food products in the recipe. Because "Cutting" is one of the most basic process in cooking, in this research, we aim classifying the food object the cooker is cutting.

In the previous research, food products are classified by their sight feature with camera. However, a cooker often cooks two or more food products that have similar sight feature at the same time. In such case, classification by sight feature does not perform.

We propose the method for classifying a food product using vibration sounds generated by the cooker cutting it. The vibration sounds contain the features of kitchen knife making a slit in the surface, passing over inside, and hitting on the cutting plate. It means that the vibration sounds change depend on an internal structure and the size of the food product. But, capturing the vibration sounds with a general microphone, irrelevant noises such as voice and air fan's sound are included in them. For solving this problem, we capture the vibration sounds directly, by the contact microphone put on the back of cooking counter.

The basic vibration sound generated by cutting process consisted of two peaks and the vibration between them. The first peak is occurred by a kitchen knife making a slit in the surface of a food product, and the second is occurred by the knife hitting on the cutting plate. The vibration between them is occurred during the knife is passing over inside the food product. According to this analysis, we picked up 3-features: amplitude of the first peak, amplitude of the second peak, and wave at passing, for classifying food products.

At the first, we extracted the amplitudes of first and second peaks about many kinds of food products and compared them. And we found that these two features widely spread even they came from the same kind of food product. It

means that the system cannot classify a food product by these 2 features.

Second, we evaluated the classification ability of "wave at passing". We extracted the vibration from 0.2 second just before the second peak to the second peak as an analytical segment from various kinds of food products. Then we calculated a spectrogram for each analytical segment and compare them. And, we discovered that the feature which depends on the kind of the food product appears in the low frequency part. We picked up 16 dimension feature vector consisted of average and variance of the spectrogram on the lowest 8 frequencies with 128 frequency resolution as classification features.

We conducted the following experiments to evaluate the classification ability of food products by the feature vector described above. We choosed "cabbage", "carrot", "cucumber", "onion", "green pepper", "potato", and "tomato" as subjects that were typical and frequently appeared in many recipes. And we prepared 4 to 6 substances for each food product, and extracted the sets of analytical segments from the vibration sounds by cutting each substance. Then, we converted each analytical segment into the 16 dimension feature vector described above. We trained a classifier for each food product by SVM (Support Vector Machine).

In the result, the classifier trained by all analytical segments succeeded to classify more than 75.0% of analytical segments for all food products. On one hold cross-validation, although classification rate of the tomato was low as 17.8% of analytical segments on average, the other food products were able to be classified more than 52.9% on average. When the system decided the classification result by a majority vote of sets of analytical segments extracted from one substance, it succeeded to classify 88.9% substance on the average except tomato.

We also evaluated the classification ability for the pair of the potato and the onion, and the pair of the cucumber and the green pepper as similar colored food products. The classification rates of both pairs showed more than 81.4%. It means that the system combined this structure with image recognition is able to improve the classification rate.

As a future work, we will construct such structure that integrates the classification result by image and the vibration sound.

調理加工に起因する振動音を用いた食材識別

目次

| | | |
|-------|------------------------|----|
| 1 | 緒論 | 1 |
| 2 | 振動音を用いた食材識別 | 3 |
| 2.1 | 従来手法との比較と振動音を用いた手法の提案 | 3 |
| 2.2 | 振動音の収録環境 | 3 |
| 3 | 食材加工に起因する振動音の性質 | 6 |
| 3.1 | 振動音の形状と「切る」動作との関係 | 6 |
| 3.2 | 識別に有効な振動音の特徴 | 7 |
| 3.2.1 | 振幅による特性 | 7 |
| 3.2.2 | 内部構造に起因する周波数特性 | 12 |
| 4 | 食材識別システム | 16 |
| 4.1 | システムのアルゴリズム | 16 |
| 4.2 | 識別機学習 | 18 |
| 5 | 実験・評価 | 19 |
| 5.1 | レシピ内の食材全てを対象とした場合の識別能力 | 19 |
| 5.2 | 映像等により絞り込めた場合の識別能力 | 21 |
| 5.3 | 考察 | 23 |
| 6 | まとめと今後の課題 | 24 |
| | 謝辞 | 24 |
| | 参考文献 | 24 |

1 緒論

調理者が新しいレシピに挑戦したり、うろ覚えのレシピを調理する際は、料理の本や料理番組を参照するのが一般的である。しかし、調理では扱う食材の種類や分量、加工の仕方など気をつけなければならない事柄が多く、その手順を調理前に完全に暗記しておくのは難しいため、調理中に何度も本やビデオを参照しなおさなければならない。このような場合、調理者は加工作業を中断しなければならないためわずらわしく、また火元から目を離さなければならなくなるため危険である。

そこで、調理者の調理行動を観測し、調理者が現在どの食材を加工しているかを把握して、適切な教示を行なうシステムの開発を目指している。このうち「切る」は、調理で用いられる食材の多くに施される加工であり、最も基本的な加工の一つである。また、調理過程の初期に行なわれる場合が多いため、この時点で食材を識別することができれば、その後、その食材に対するその他の加工に対し支援を行なう際にも有利であると考えることができる。そこで本研究では、調理者が食材を「切る」作業に注目して、その対象食材の識別を行なうことを目標とする。

従来、調理者が扱っている食材を識別する際には、カメラにより調理を観測した映像から抽出した食材の視覚的特徴を用いてきた [1]。しかし、調理台上には食材以外にも調理器具などが散在しているため、食材領域を抽出することが困難であった。また、調理では、視覚的特徴の類似した食材が同時に複数種類用いられる場合も多く、色だけでは判別が難しい。

一方、調理台を伝播する振動音は調理台上で行なわれる様々な動作を反映しているが、このうち包丁で食材を切ることによって生じる振動音は、物を置いたり引きずったりする際に生じる振動音とは性質が異なることから、その抽出が比較的容易である。また、この振動音は、包丁が食材の表面を裂き、内部を通過する際に発生する振動を含むことから、食材の内部構造や大きさによって変化すると考えられる。そこで、食材を「切る」加工に起因する振動音波形から、食材固有の特徴を見出し、食材を識別する手法を提案する。

まず、振動音を取得するための環境システムについて提案する。振動音を通常のマイクで取得したのでは、声や換気扇の音など、食材を切ることで生じる振動音とは無関係の音による影響が避けられない。そこで、まな板を置く作業

台の裏にコンタクトマイクを貼り付けることで、振動音を直接取得する。

次に、収録した波形を分析することで、「切る」加工に起因した振動音は食材の種類に依存した特徴を保持しており、これにより食材識別が可能であることを示す。このため、食材を1回切る際の波形を取り出し、その概形と、包丁と食材の関係について述べる。また、その形状が食材のどのような性質に関連しているかをデータ分析により明らかにする。

最後に、波形から見出された特徴量により、SVMを用いて識別する手法を提案する。食材ごとに学習用のサンプルを集め、各サンプルを特徴ベクトルに変換し、SVMを用いて特徴空間を食材領域に分割することで識別機を構成する。構成された識別機に、未知データを投入した場合に、どの程度の識別率を算出することができるのかを調べる。

以下、2章では、振動音を用いた食材識別の概要について述べる。3章では、振動音波形の形状と「切る」加工との関係、及び振動波形の各特徴が食材のどの特徴に関係しているのかについて述べる。4章では、識別システムの処理の流れと具体的な手法について述べる。5章では、識別能力を分析し、結果とその考察を述べる。6章では、本研究のまとめと今後の課題を述べる。

2 振動音を用いた食材識別

2.1 従来手法との比較と振動音を用いた手法の提案

本研究では、調理において調理者が切る加工を行なう際、その対象が、レシピの中のどの食材であるのかを識別することが目標である。システムには調理で用いられる食材及びその加工方法が記載されたレシピが与えられているものと仮定する。

食材の識別を行なうことを考えた場合、食材の特徴としては、色、形状、大きさ、堅さ、内部構造などが考えられる。従来は映像により食材の視覚的特徴を観測し、食材を識別する手法が用いられていた。色、形状、大きさ等の視覚的特徴の良く似た食材が同時に調理される場合は少なくない。また、調理では、前回に半分使った残りの食材を使うという状況も頻繁に発生するが、この場合、形状、大きさは既に変化しており、元の食材とは違うものと判断してしまう。これに対して、人参は表面から内部にかけて全体的に堅いが、トマトはやわらかい。また、ピーマンなどは内部が空洞になっている、と言った具合に食材の内部構造は異なるものが多い。また、食材の内部構造や堅さは変化しないため、この特性を識別に利用できれば、上記のような映像では識別が困難であるような食材を識別できる。

調理の作業場には、複数の食材、包丁、ボールといった様々なものが散在しており、調理者が切っている食材の領域を抽出することは困難である。調理者の腕や包丁などによるオクルージョンで食材特徴が十分抽出できないということも考えられる。これに対して「切る」という動作は通常まな板の上で行なわれるものであり、その振動音を抽出することは容易である。調理で扱う食材のうち、「切る」加工を施すものは多岐にわたる [2]。よって、その識別対象は決して狭いものではない。また、その多くが野菜であるということもわかっているので、本研究では識別対象を「切る」加工を施す野菜に限定する。

そこで、本研究では、調理者が切る対象がレシピ内のどの食材であるかを識別するのに、食材を切る際に生じる振動音を用いる手法を提案する。

2.2 振動音の収録環境

本節では、「切る」加工に起因する振動音の性質及び、振動音を収集するのに必要な機材、環境について述べる。

食材を切る際に生じる音は、食材の堅さや層の有無などの内部構造に関する情報を持っており、食材の識別に有効であると考えられる。しかし、この音を通常のマイクで取得する場合、人の話声、テレビの音、水を流す音など、切ることは無関係の音も同時に収録されてしまい、これらの除去が大きな問題となる。

一方、「切る」という動作は通常まな板の上で行なわれるものであり、食材を切ることで生じる振動音は、まな板を通じて作業台にも伝わる。この時、人の話声などの雑音も空気を伝わり作業台を振動させるが、その力は直接接触している食材を伝わる「切る」加工の振動音による力に比べれば微力である。そこで、まな板を置く作業台の裏に直接コンタクトマイクロフォンを貼り付け、食材を切ることで生じる振動音を取得することを考えた。これにより、作業台に直接接触していない場所で発生する雑音の影響は極力排除することができる。

本研究において「振動音」とは、まな板の上の食材を包丁で切る際に、食材を通じて作業台に伝播する振動音を、接触型音収録マイクであるコンタクトマイクロフォンで収録した音を指す。

キッチンの作業台は、振動音を伝播しやすい硬化ガラスでできており、この裏の4角に、AKG社製の、楽器用コンデンサーマイクとして広く用いられているMICROMIC 3シリーズのc411を設置した。本マイクは、通常は楽器の音収録に用いられる。振動を直接センサーで拾うので、まわりで音が鳴っていても、また比較的小さな音量でも、安定した正確な調律が可能である。各マイクが拾った振動音は、一台のUSB Audio Capture(EDIROL社 UA-1000)を通してPC内に取り込まれる。これを、Adobe社製のAudition1.0を用いて、各マイクごとに44.1kHzのサンプリング間隔・16bitのレンジで記録した。作業台付近の概観を図1に、キッチン全体の概観を図2示す。

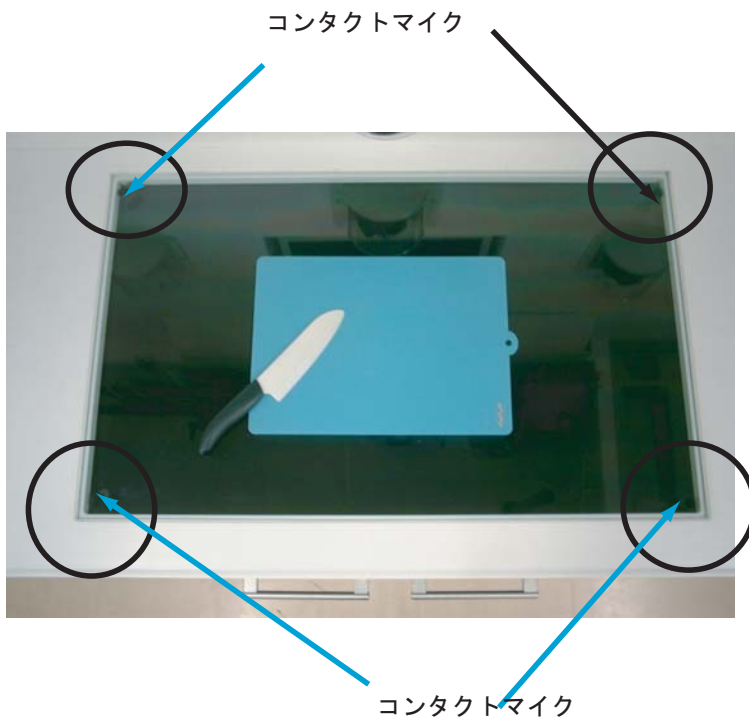


図 1: 作業台の概観



図 2: キッチン全体

3 食材加工に起因する振動音の性質

3.1 振動音の形状と「切る」動作との関係

食材を切る加工は、包丁が食材表面に切れ込みを入れ、さらに内部を進んでいき、刃がまな板に衝突してから持ち上げられるという動作を基本とする。本研究ではこの一連の動作に対応する波形部分を「基本セグメント」と呼ぶ。代表的な基本セグメントを図3に示す。横軸はサンプル数（サンプリング周波数44.1kHz）を示し、縦軸は各サンプル点における振幅を示す。

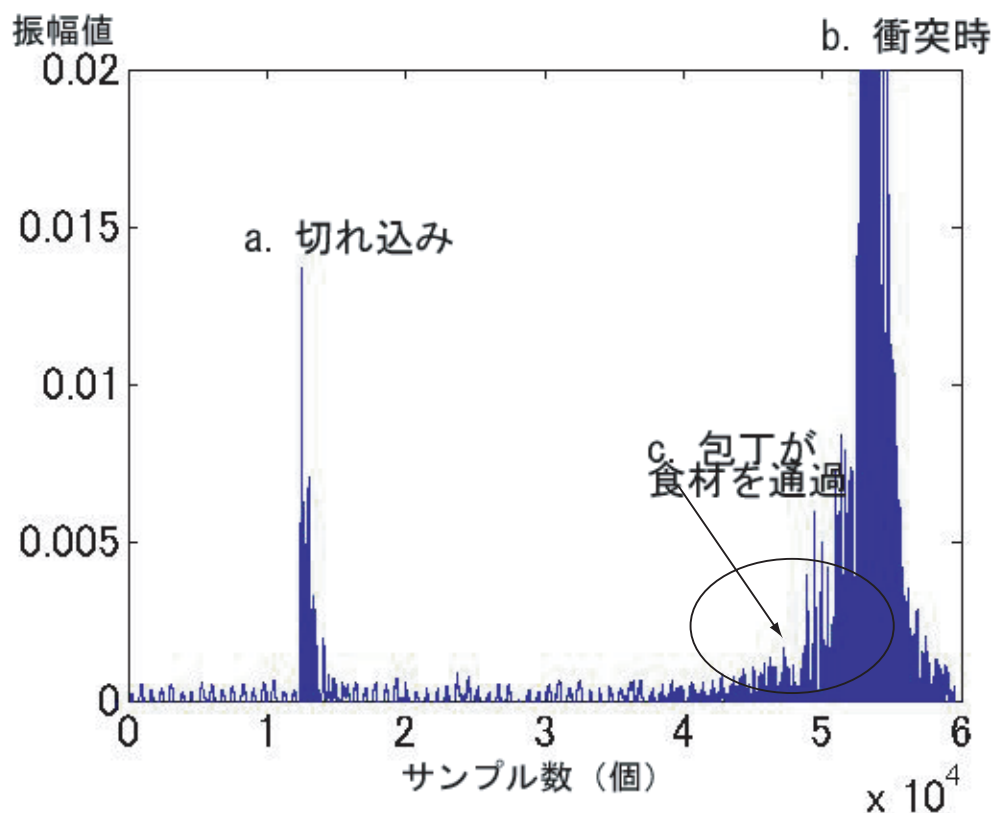


図3: 44.1kHz でサンプリングした基本セグメント

基本セグメントは以下で述べるような性質を持っている。

- 図3中aで示したピーク: 食材に包丁による切れ込みが入った際の振幅を表す。食材の表面の堅さが包丁による食材表面の破壊エネルギー、振動の伝播具合に影響し、この値が比例して大きくなると予測できる。キャベツのような葉ものや、皮を剥かれたものは、表面を切れ込む際の抵抗が小さいため、切れ込みによる振幅が小さく検出できないものも存在する。

- 図3中で示した部分 b: 包丁がまな板に衝突した際の衝突エネルギーを表す。食材内部が堅いものほど包丁をまな板に叩きつける力が強く、また切る速度が速いほど衝撃が大きいと考えられる。
- 図3中で示した部分 c: 包丁が食材の内部を通過している間、食材自身を伝わり作業台に伝播する振動による振幅を表す。食材の内部構造の違いを最も反映している部分であると考えられる。例えば、玉葱は内部が層構造になっているため、この部分に複数のピークが見られる。一方、ジャガイモは内部が均一であるため、この部分はほぼ一定値を取る。

作業台では、食材や調理器具を置いたりするため、これらの動作による振動も、作業台上に生じ、マイクにより収録される。これらの波形は、食材を「切る」加工に起因する波形である基本セグメントではない。これらの波形は、以下のような違いにより棄却することができると考えられる。物を置いたときの波形は振幅がほぼ0の状態をとり続けた後、物が置かれた時点で急に極大振幅をとるのに対して、基本セグメントは、包丁がまな板に衝突する際に極大振幅をとる前にも、包丁が切れ込みを入れたり、食材を通過することである程度大きな振幅が発生している。調理器具を置いた際に生じる振動音の一例を図4に示す。横軸はサンプル数(サンプリング周波数 44.1KHz)を示し、縦軸は各サンプル点における振幅を示す。

まな板や包丁の違いによる波形への影響が考えられる。まな板の厚みは、振動の伝播に影響するため、できるだけ減衰を抑える目的で、下敷き風の薄いものを用いた。包丁の違いは波形の特徴に影響すると考えられるので、できるだけ汎用性を高めるため、一般家庭において広く用いられているプラスチック製の物を用いた。これらまな板、包丁の制約は、システムの一般家庭での利用を前提にした場合、一般性を失わないと考えられる。

3.2 識別に有効な振動音の特徴

3.2.1 振幅による特性

本節では、図3の振動音波形の a(切れ込み時の振幅),b(まな板に衝突する際の極大振幅)の特徴が、どの程度の食材識別能力を持つかについて分析する。食材の堅さの違いは b の極大振幅に大きく影響すると予測される。食材が堅ければ、その食材を切る場合に大きな力が必要となり、勢いが付き、まな板に衝突する際に大きな振動を与えると考えられる。一方やわらかい食材であれば、大

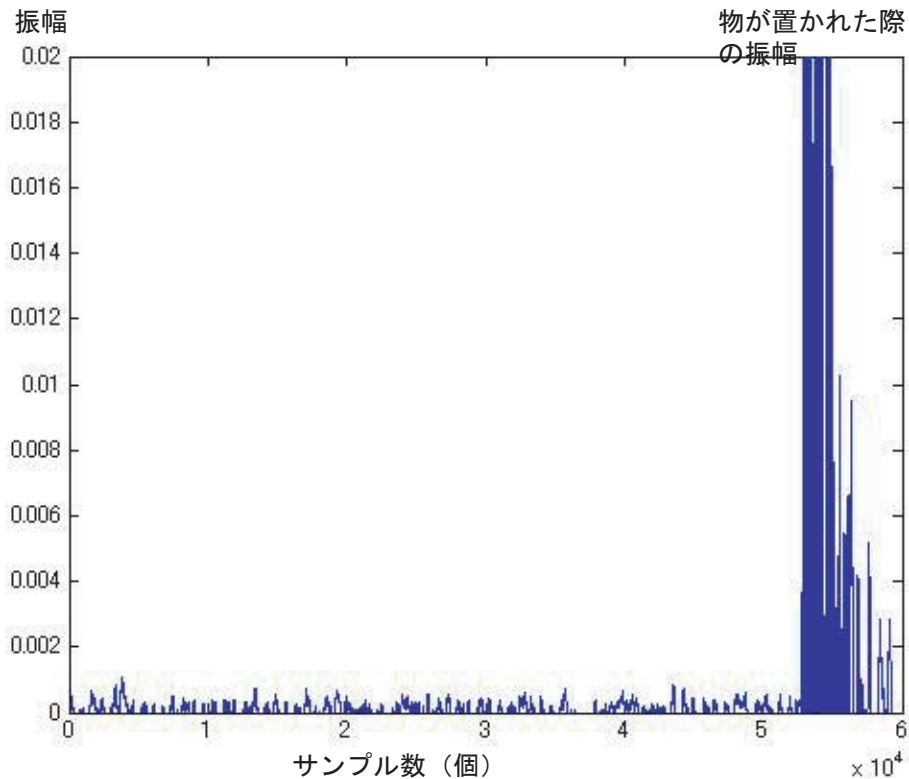


図 4: 調理器具を置いた際に生じる振動音の一例

きな力は必要なく、寧ろ調理者の側がその食材が潰れてしまわないように力を制御して切ることになり、まな板に衝突する際にはあまり大きな振動を与えないのではないかと予測できる。

また、同じ食材でも個体の大きさが違えば、極大振幅が変化することも考えられる。ここで言う大きさとは、食材の切れ込み部分の、まな板からの距離のことを指す。直感的には、食材が大きいほど包丁が進み始めてからまな板に衝突するまでの時間が長くなり、加速度が大きくなり、より大きな力でまな板に衝突するため、より大きな極大振幅を示すのではないかと考えられる。一方で、食材を通過する際の包丁にかかる抵抗力が働く時間も長くなるので、まな板に衝突する前の速度は逆に小さくなることも考えることもできる。

そこで、まず初めに極大振幅と食材の大きさの関係を調べる。次に極大振幅と食材の堅さの関係を調べ、最後に食材ごとの切れ込み振幅と極大振幅の分布をまとめる。以上の関係により極大振幅及び切れ込み振幅により、どの程度の食材を識別することができるのかを明らかにする。

[極大振幅と、食材の「大きさ」との関係]

食材として人参、玉葱、ジャガイモ、トマトの4種を調査対象とした。これらは、実際の調理環境において元の大きさのまま、或いは半分に切ってからと様々な大きさで切られることが多く、かつ頻繁に用いられる。各食材についてそれぞれ厚み1 cm、3 cmのサンプルを2つずつ用意した。このとき、食材の個体差によるばらつきも考慮し、2つのサンプルはそれぞれ異なる個体から採取した。3cmのものは食材の1個体の大きさを想定し、1cmのものは個体の半分の大きさを想定している。玉葱とジャガイモは皮を剥き、人参は皮を剥かずに、端から幅5mm程度の間隔で、一定の速度で包丁により輪切りにした音を収録した。図5(i),(ii)、図6(i)(ii)にそれぞれジャガイモ、玉葱、人参、トマトの1cm、3cmの個体を切った際の極大振幅のばらつきを示す。横軸は大きさ(cm)。縦軸は振幅を表す。これによると、大きさが異なっても同じような振幅の分布を示している。この結果から、食材の大きさは、振動音波形の極大振幅にはあまり影響を及ぼさないことが示唆される。

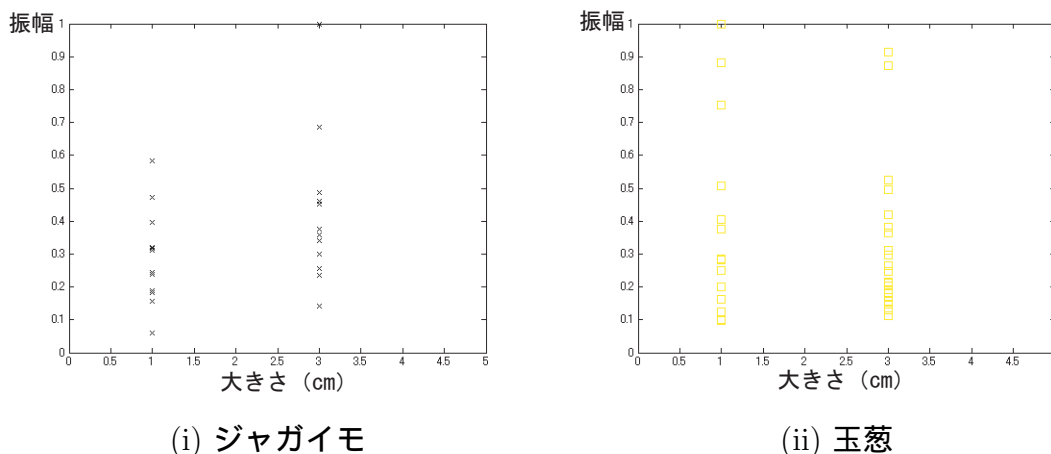


図5: 極大振幅と食材の大きさの関係

[極大振幅と、食材の堅さの関係]

極大振幅は、食材の堅さに比例して大きくなるのではないかと考えた。そこで、調理に頻繁に用いられ、かつ堅さが異なる食材として、人参、玉葱、ジャガイモ、トマト、キュウリ、レンコンの6種の野菜を調査対象とした。人参は皮を剥かずに立て方向に半分に割ったもの、玉葱は皮を剥き、二つのヘタを結ぶ

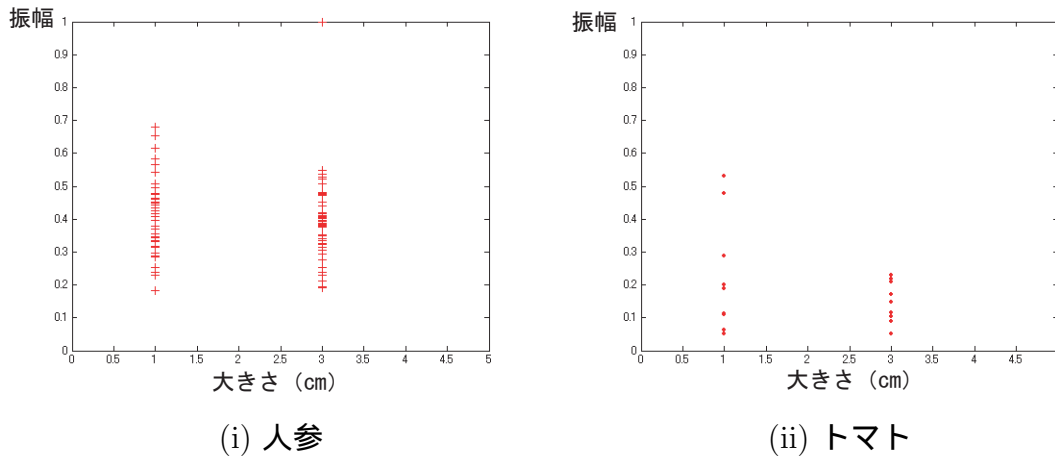


図 6: 128 段階の周波数領域でのスペクトログラム

直線で半分に切ったもの、ジャガイモは皮を剥いたもの、トマトはヘタに垂直な方向で半分に切ったもの、キュウリ、レンコンはそのままの大きさをそれぞれ1つのサンプルとして、各サンプルを端から幅 5 mm 間隔で一定の速度で輪切りにし、その際発生した振動音の極大振幅のばらつきを調べた。結果を図 7 に示す。横軸に、やわらかいと思われるものから順に、トマト、キュウリ、ジャガイモ、玉葱、人参、レンコンの順に並べている。縦軸は極大振幅を表す。この結果によると、トマトのようにやわらかいものでも、極大振幅は大きな値をとる場合があり、レンコンのように堅いものであっても、極大振幅が小さな値をとる場合がある。また、各食材を切った際の極大振幅は広く分布しており、食材ごとに排他的に分類することができないことがわかる。この結果、1つの基本セグメントの極大振幅だけでは、どの食材を切っているのかは判断できないということがわかった。

[食材の、切れ込み振幅と極大振幅の分布]

極大振幅だけでは食材識別は困難であるため、切れ込み振幅と組み合わせることで食材を識別できないかと考えた。先と同じ6種の野菜を調査対象とし、同様の条件で切った際発生した振動音波形を収集し、切れ込み振幅の分布の仕方を調べた。まず、振幅に閾値を設けることで包丁がまな板に衝突した際の時刻を検出し、その直前 T_1 秒間の波形を一つの基本セグメントとして抽出した。この抽出の際の時間間隔 T_1 は、抽出した基本セグメントに、切れ込み時点が含まれ、かつ一つ前の基本セグメントの減衰波形が含まれないよう食材ごとに設

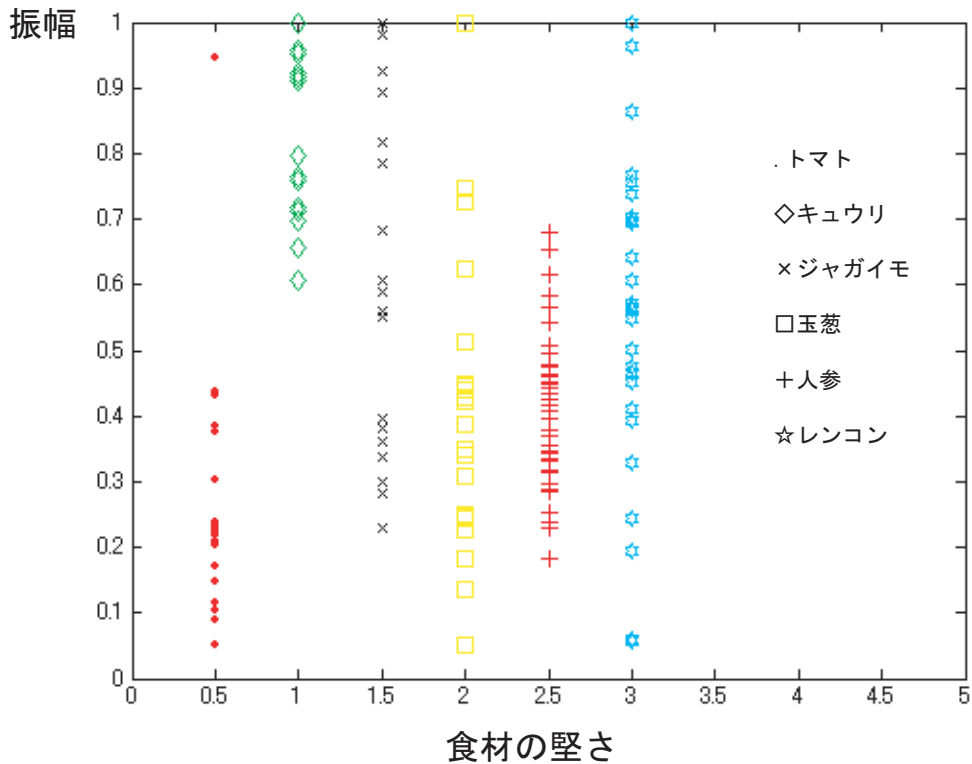


図 7: 各食材の極大振幅のばらつき

定した。各食材ごとの T_1 の値を表 1 にまとめる。

表 1: 各食材の T_1 の値

| 食材 | 玉葱 | ジャガイモ | 人参 | ピーマン | キュウリ | レンコン | トマト |
|--------------|-----|-------|-----|------|------|------|-----|
| T_1 の値 (秒) | 1.2 | 1.0 | 1.2 | 1.2 | 0.5 | 1.2 | 1.0 |

こうして抽出した基本セグメントに対して、切れ込み振幅、極大振幅の 2 つの特徴を分析した。切れ込み振幅は、手動で抽出した。各特徴量における食材の種類ごとの基本セグメントの平均、分散の値をを表 2 にまとめる。トマトに関しては切れ込み振幅の値があまりにも小さかったため、検出できなかった。

各特徴量とも平均値に対する分散値が大きく、食材間での重なりが大きく、トマトとそれ以外の判別が可能であるに過ぎないと考えられる。また、実験を進める中で、極大振幅や切れ込み振幅は、調理者がどのように切るかにより大きく変化する、すなわち、堅い食材であっても落ち着いて切れば小さな振幅を取

表 2: 各特長量の平均と分散

| 食材/特徴量 | 切れ込み振幅 | | 極大振幅 | |
|--------|------------------------|------------------------|--------|------------------------|
| | 平均 | 分散 | 平均 | 分散 |
| 玉葱 | 1.613×10^{-2} | 4.587×10^{-5} | 0.2544 | 2.837×10^{-2} |
| ジャガイモ | 6.689×10^{-3} | 1.358×10^{-5} | 0.3496 | 2.129×10^{-2} |
| 人参 | 1.080×10^{-2} | 2.629×10^{-5} | 0.3977 | 3.166×10^{-2} |
| ピーマン | 2.061×10^{-3} | 1.231×10^{-6} | 0.1816 | 2.322×10^{-2} |
| キュウリ | 5.539×10^{-3} | 1.340×10^{-5} | 0.4109 | 5.007×10^{-2} |
| レンコン | 1.418×10^{-2} | 3.378×10^{-5} | 0.4217 | 5.618×10^{-2} |
| トマト | *** | *** | 0.0413 | 3.570×10^{-4} |

るし、柔らかい食材であっても何らかのはずみで勢いがついてしまい、大きな振幅を取る場合が少なくないことがわかった。よって極大振幅や切れ込み振幅の値は必ずしも食材自身の特徴を反映しているわけではないので、これらの特徴は食材識別には有効では無いと言える。

次節では、基本セグメントにおいて、図 2 の c. 包丁が食材を通過している間の波形部分に注目して分析していくことにする。

3.2.2 内部構造に起因する周波数特性

先と同様に、包丁がまな板に衝突した際の時刻を検出し、その前 0.2 秒間分を分析用セグメントとして抽出する。この 0.2 秒という間隔は包丁が切れ込みを入れた後、一旦止まった後に進み始めてからまな板に衝突するまでの部分であり、図 3 の基本セグメントにおける c の部分に含まれる、食材を通過している間の波形部分であるといえる。得られた分析用セグメントに対して、内部構造の違いが各時刻における各周波数成分の強度（パワー）に影響を及ぼしているであろうと期待して、スペクトログラム分析を行った。スペクトログラムとは、各時刻における各周波数成分の強度を数値で表したものであり、強度 $X(t, \omega)$ は時刻 t と周波数 ω の関数となり、以下の数式により表される。

$$X(t, \omega) = \left| \int_{-\infty}^{\infty} x(t') \cdot h(t' - t) \cdot e^{-i\omega t'} dt' \right|^2$$

ここで、 $x(t)$ は振動音波形、 $h(t)$ は窓関数を表す。窓関数はハミング窓を用いた。分析用セグメントのスペクトログラムを求めるのには、MATLAB の Auditory Toolbox に用意されている spectrogram 関数 [3] を用いた。この関数では、切り

出した波形、窓幅、窓を動かす際の重なり幅を引数として入力することで、任意の周波数、時間解像度で、列に時刻、行に周波数が対応したスペクトログラム行列を出力する。

$$X(n, \omega) = \left| \sum_{m=-\infty}^{\infty} x(n+m) \cdot h(m) \cdot e^{-i\omega m} \right|^2$$

窓幅が大きい程、周波数解像度は高くなるが、時間領域における解像度は低くなり、これらはトレードオフの関係にある。また、窓の重なり幅が大きいほど時間領域における解像度が高くなる。玉葱、ジャガイモの分析用セグメントのスペクトログラム行列を可視化した例を図8に示す。

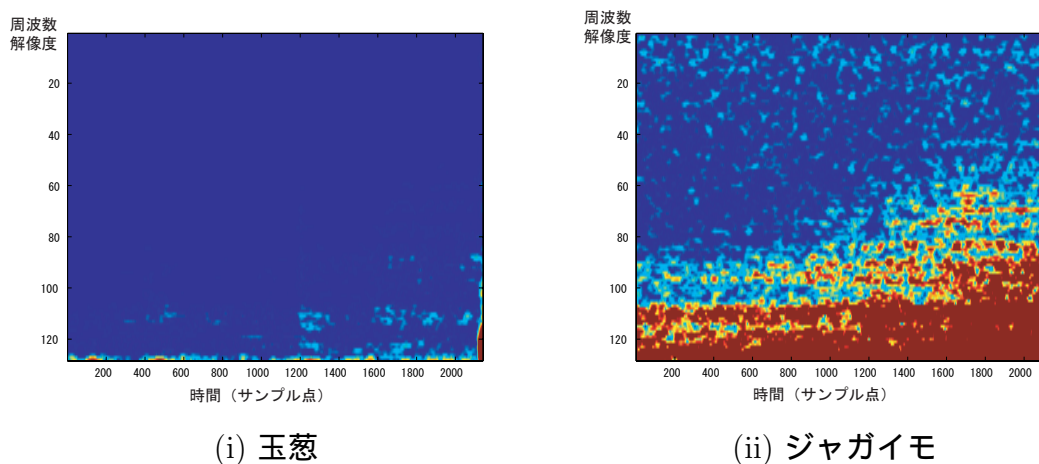


図8: 128段階の周波数領域でのスペクトログラム

各食材のスペクトログラム行列の可視化を行なったところ、低周波部分に食材の種類に依存した違いが現れることがわかった。よって、スペクトログラム行列の低周波部分を特徴ベクトルに用いることを考える。一般に、識別機を設計する場合、特徴の数を増やせば識別に用いることのできる情報量が増え、識別率の上昇が期待される。しかし、以下の3つの問題が生じる。

- 相関の高い特徴の組が混入し、次元を増やしても情報量が増えない
- 統計計算に要する計算量は少なくとも次元のべき乗のオーダーになる
- 有限個の学習パターンから識別機を設計する際、次元を高くしていくと誤識別率がかえって上昇してしまう（ヒューズの現象）

そこで、特徴ベクトルの次元はできるだけ小さく、かつ識別能力が高いことが望まれる。先に述べたように、窓幅、窓の重なり具合を変化させることで、任

意の周波数解像度でスペクトログラムを取得することができる。周波数解像度が高くなると、低周波部分のより細かな違いが明らかになり、識別率が上昇するのではないかと期待できるが、特徴ベクトルに採用する低周波部分の行数を増やす必要があるため、特徴ベクトルの次元が大きくなってしまい、上記の3つの問題が生じてしまう。そこで、周波数解像度を色々に変化させてスペクトログラムを取得し、低周波部分から抽出する次元を変化させることで、どの周波数解像度でスペクトログラムを取得するのが特徴ベクトルに適しているのかを調べる。

まずは周波数解像度 i ($i = 64, 128, 256, 512$) でそれぞれのスペクトログラム $X_i(t, \omega)$ ($t = 1, 2, \dots, t_i; \omega = 1, 2, \dots, i$) を算出する。ただし、 $t_i = 1071 \times \frac{i}{64}$ とする。各スペクトログラムの低周波成分からそれぞれ、

$$n_i = \frac{i}{16}$$

次元分を分析対象とする。特徴量としては、 $X_i(t, \omega)$ の低周波数部分

$$\omega = i - n_i + 1, \dots, i$$

の全ての時刻 t における平均ベクトル \mathbf{y}_i と、分散ベクトル \mathbf{z}_i の結合により、 $2n_i$ 次元の特徴ベクトル \mathbf{x}_i として算出される。 $\mathbf{y}_i, \mathbf{z}_i$ の第 j 成分はそれぞれ以下の式により得られる。

$$\mathbf{y}_i(j) = \frac{\sum_{t=1}^{t_i} X_i(t, i - (j - 1))}{t_i} \quad (j = 1, \dots, n_i)$$

$$\mathbf{z}_i(j) = \frac{\sum_{t=1}^{t_i} (X_i(t, i - (j - 1)) - \mathbf{y}_i(j))^2}{t_i} \quad (j = 1, \dots, n_i)$$

最後に、 $\mathbf{y}_i, \mathbf{z}_i$ を以下のように結合し、 \mathbf{x}_i を求める。

$$\mathbf{x}_i(j) = \mathbf{y}_i(j) \quad (j = 1, \dots, n_i)$$

$$\mathbf{x}_i(n_i + j) = \mathbf{z}_i(j) \quad (j = 1, \dots, n_i)$$

玉葱、ジャガイモ、人参の3種類についてそれぞれ4、4、2個体ずつ切り、抽出した分析用セグメントの数はそれぞれ63、45、37である。各分析用セグメントを上記の手順に沿って特徴ベクトル \mathbf{x}_i ($i = 64, 128, 256, 512$) に変換する。各周波数解像度ごとに、得られた特徴ベクトル全てを用いて食材ごとに学習し

た識別機により特徴空間を玉葱、ジャガイモ、人参の3つの食材領域に分割した。学習には、どの特徴成分が識別に寄与しているのかが一目でわかるという利点から、決定木を用いた。識別機により分割された特徴空間において、学習に用いた各特徴ベクトルが正しい領域に分類されている割合を調べた。結果は表3の通りである。各解像度ごとに、正しく識別できた特徴ベクトルの個数及び割合を示している。

表3: スペクトログラムの精度による認識率の違い

| 食材 (個数)/解像度 | 64 | 128 | 256 | 512 |
|-------------|------------|------------|------------|------------|
| 玉葱 (63) | 47 0.7460 | 57 0.9048 | 54 0.8571 | 57 0.9048 |
| ジャガイモ (45) | 37 0.8222 | 44 0.9778 | 40 0.8889 | 44 0.9778 |
| 人参 (37) | 27 0.7297 | 34 0.9189 | 30 0.8108 | 34 0.9189 |
| 平均精度 (145) | 111 0.7655 | 135 0.9310 | 124 0.8552 | 135 0.9310 |

この結果より、128段階の周波数解像度でスペクトログラムを取得し、低周波数部分8次元の平均ベクトル及び分散ベクトルを結合した16次元のベクトルが、次元数が小さく、識別能力の高いものであり、特徴ベクトルに適しているといえる。

図9に、16次元ベクトルにより上記サンプルを識別した際の決定木を示す。

まず第10成分の値により枝分かかれ、ある値以上のものはジャガイモと判断される。括弧内の数字は、先の実験で用いたサンプルのうち、21のものはジャガイモとして分類され、その全てが正しい分類であることを示す。以下第11成分の値で枝分かかれ、と言った具合に16次元の成分が満遍なく使われており、16次元全てが識別に有用であることがわかる。

Decision tree:

```
A10 > 1859.2 ⇒ (ジャガイモ) (21)
A10 ≤ 1859.2:
... A11 ≤ 747.56:
... A16 > 837.85:
: ... A3 ≤ 53.922 ⇒ (玉葱) (28/1)
: ... A3 > 53.922 ⇒ (人参) (4/1)
: ... A16 ≤ 837.85:
: ... A11 ≤ 280.33: ⇒ (玉葱) (18/2)
: ... A11 > 280.33:
: ... A8 ≤ 17.179 ⇒ (ジャガイモ) (9)
: ... A8 > 17.179:
: ... A10 ≤ 447.27 ⇒ (玉葱) (7)
: ... A10 > 447.27 ⇒ (ジャガイモ) (4)
A11 > 747.56:
... A5 > 90.842 ⇒ (ジャガイモ) (8/3)
... A5 ≤ 90.842:
... A8 > 73.746:
... A3 ≤ 55.103 ⇒ (玉葱) (7)
: ... A3 > 55.103 ⇒ (人参) (6/1)
... A8 ≤ 73.746:
... A15 > 2454.4 ⇒ (ジャガイモ) (3/1)
... A15 ≤ 2454.4:
... A7 ≤ 44.884 ⇒ (ジャガイモ) (4/1)
... A7 > 44.884 ⇒ (人参) (26)
```

図9: 識別で得られた決定木

4 食材識別システム

4.1 システムのアルゴリズム

本節では、本研究で提案する食材識別システムの処理の流れを説明する。

1. 識別機の学習

システムには調理者が調理しえる様々なレシピ情報が登録されているとする。各レシピには、その料理で用いられる材料のリストが記載されており、事前にシステムは以下の手順により、レシピごとに、そこで用いられる食材を識別するための識別機を学習しておく。識別機の学習の手順は以下の通りである。

レシピの材料名に相当する食材群を一つのクラス C_i 、食材の全体集合を C とすると、

$$C = \{C_1 = \text{"onion"}, C_2 = \text{"potato"}, \dots\}$$

で表される。システムは事前に各クラス C_i について、 N_i 個の個体 $D_{i,j}$ を切った際の振動音を収集している。

$$C_i \ni \{D_{i,1}, D_{i,2}, \dots, D_{i,j}, \dots, D_{i,N_i}\}$$

各個体 $D_{i,j}$ を切って得られた $n_{i,j}$ 個の分析用セグメント $F_{i,j}^{(k)}$ は次式の集合で表される。

$$D_{i,j} \ni \{F_{i,j}^{(1)}, F_{i,j}^{(2)}, \dots, F_{i,j}^{(k)}, \dots, F_{i,j}^{(n_{i,j})}\}$$

各食材 C_i ごとに $\sum_{j=1}^{N_i} n_{i,j}$ 個の分析用セグメント $F_{i,j}^{(k)}$ を抽出し、これらを 3.2.2 で述べた手順に従い、16 次元の特徴ベクトル $x_{i,j}^{(k)}$ に変換する。

各レシピ R_l では、 M_l 種類の食材集合が用いられているとすると、

$$R_l = \{C_{l_1}, C_{l_2}, \dots, C_{l_i}, \dots, C_{l_{M_l}}\}$$

で表される。 R_l に含まれる各食材 C_{l_i} には $\sum_{j=1}^{N_{l_i}} n_{l_i,j}$ 個の特徴ベクトル $x_{l_i,j}^k$ が用意されている。これら全ての特徴ベクトルを用いて、レシピ内に含まれる全ての食材で識別機 I_l を学習する。

また、映像等により食材の候補が絞られた場合や、上記の I_l では決定できない場合を想定し、 C の中から映像では識別が困難であるような食材の組み合わせ、 I_l においてよく似た識別率を算出するような食材の組み合わせ集合 S_m を用意しておく。 S_m は高々 2, 3 種の食材より構成されており、

$$S_m = \{C_{m_1}, C_{m_2}, C_{m_3}\}$$

と表される。ただし m の値により、 C_{m_3} は存在せず、その場合は 2 種の食材より構成されるものとする。 S_m に含まれる各食材で識別機 J_m を学習する。

2. データ観測

まず調理者がレシピ R_l を指定する。システムは調理の開始より振動音波形 $x(t)$ を観測し始める。振動音波形の振幅が閾値を超えた時点を、食材を切った際に包丁がまな板に衝突した時点の候補 p とする。 p から、包丁が食材を通過している部分として、その直前 0.2 秒間分の波形を分析用セグメントの候補として抽出する。各候補が、食材を切る加工に起因した波形であるかどうかは、包丁がまな板に衝突する前に食材を通過している部分が振幅として現れることにより判別できると考えられる。映像処理により食材の個体 D がまな板に入った時刻より出て行くまでの間に収集した分析用セグメントの候補集合のうち、上記の条件を満たすものを真の分析用セグメント集合 $F(D)$ として抽出する。

3. 識別

$F(D)$ に含まれる各分析用セグメントの一つ一つを3.2.2の手順に従い、16次元の特徴ベクトルに変換し、特徴ベクトル集合 $X(D)$ とする。 $X(D)$ に含まれる各特徴ベクトルが、レシピ R_l に対応する識別機 I_l において、どの食材領域に分類されるかを調べる。最後に、多数決を取ることで、レシピ内のどの食材であるかを決定する。

この際、識別機による分類で1番目と2番目の食材で多数決による差があまりみられないような場合は、この時点で食材を決定するのではなく、その2種類の食材より構成された S_m に対応する識別機 J_m において、 $X(D)$ に含まれる各特徴ベクトルが、どちらの食材領域に分類されるかを調べ、多い方をもってその食材であると決定する。

4.2 識別機学習

識別機による特徴空間の分割には、決定木、SVMなどが考えられる。決定木は、選んだ特徴量のうち、どの成分が識別に寄与しているのかが一目でわかるという利点がある。しかし、特徴空間を、特徴軸に平行な面でしか分割することができず、識別能力はあまり高いとは言えない。一方、SVMは、どの特徴成分が識別に寄与しているのかということは具体的には知ることができない。しかし、学習サンプルから「マージン最大化」という基準で線形しきい素子のパラメータを学習することで、未知データに対して高い識別率を達成することができるという利点を持つ [4]。そこで、実際の識別システムにおいてはSVMにより特徴空間をレシピに登場する各食材の領域に分割する。

5 実験・評価

本章では、本システムにより実際に食材を切って収録した振動音波形により、4章で採用した識別アルゴリズムで、どの程度食材を識別することができるのかを調べる実験を行なう。まず、レシピに登場する全ての食材により学習した識別機の精度について述べ、次に、映像等によりある程度食材候補が絞られた場合を想定し、それらの食材により学習した識別機の精度について述べる。

5.1 レシピ内の食材全てを対象とした場合の識別能力

[実験1: クローズドテスト]

1つのレシピで用いられる食材集合のプロトタイプとしてNESSLEのレシピ集[5]でよく用いられていたキャベツ、人参、キュウリ、玉葱、ピーマン、ジャガイモ、トマトの7種類の野菜を調査対象とした。これらの食材はレシピにおいて互いに共起しやすいという性質を持ち、かつ「切る」という加工を施すものである。食材ごとに2～4の個体を用意し、個体ごとに切った際の振動音波形を集めた。4章の方法により、7種類の食材よりそれぞれ42、77、66、63、46、44、16の分析用セグメントを抽出した。これら全てを学習サンプルとして、識別機を学習し、各食材を識別した際の識別率を表4に示す。

表4: クローズドテスト

| 食材 | キャベツ | 人参 | キュウリ | 玉葱 | ピーマン | ジャガイモ | トマト |
|-------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| キャベツ | <u>76.19</u> | 0.00 | 9.52 | 0.00 | 14.92 | 0.00 | 0.00 |
| 人参 | 0.00 | <u>97.40</u> | 0.00 | 0.00 | 0.00 | 2.60 | 0.00 |
| キュウリ | 0.00 | 0.00 | <u>96.97</u> | 3.03 | 0.00 | 0.00 | 0.00 |
| 玉葱 | 3.17 | 1.59 | 15.87 | <u>76.19</u> | 0.00 | 0.00 | 3.17 |
| ピーマン | 8.70 | 0.00 | 15.22 | 0.00 | <u>76.09</u> | 0.00 | 0.00 |
| ジャガイモ | 0.00 | 2.27 | 0.00 | 0.00 | 0.00 | <u>97.73</u> | 0.00 |
| トマト | 0.00 | 0.00 | 6.25 | 18.75 | 0.00 | 0.00 | <u>75.00</u> |

表中の各値は、行に対応する食材について得られた分析用セグメントが、識別機によりどの食材として識別されたかの割合を表している。対角線上の数値が、正しく識別された特徴ベクトルの割合を示す。各食材とも、複数個体から

なる分析用セグメントが75%以上の割合で正しい食材領域に分類されていることがわかる。

[実験2: オープンテスト]

実験1では、食材の個体差を考慮することなく、食材ごとに分析用セグメントを集め、それら全てを学習サンプルとして識別機を学習することで、食材間の大まかな違いを調べた。その結果、識別の可能性が十分にあることがわかった。本実験では、食材の個体差を考慮し、各食材の個体ごとに集められた分析用セグメントを1つの集合として、分けて収集する。収集された全ての集合から一つの集合を順に取り出し、それら以外の全ての集合の分析用セグメントにより学習された識別機により、どの程度の精度で識別されるのかを調べる交差検定を行なう。このためにはより多くの個体が必要である。そこで、実験1と同じ食材をそれぞれ4、4、4、6、5、4、4個体ずつ用意し、それぞれ合計で95、162、95、131、105、70、45の分析用セグメントを抽出した。各個体ごとの分析用セグメントの数を表5に示す。

表5: 個体ごとの分析用セグメント数

| 食材/個体番号 | 1 | 2 | 3 | 4 | 5 | 6 | 合計 |
|---------|----|----|----|----|-----|-----|-----|
| キャベツ | 21 | 24 | 24 | 26 | *** | *** | 95 |
| 人参 | 41 | 36 | 49 | 36 | *** | *** | 162 |
| キュウリ | 24 | 28 | 22 | 21 | *** | *** | 95 |
| 玉葱 | 15 | 17 | 26 | 24 | 23 | 26 | 131 |
| ピーマン | 18 | 21 | 21 | 21 | 24 | *** | 105 |
| ジャガイモ | 16 | 12 | 18 | 24 | *** | *** | 70 |
| トマト | 12 | 9 | 12 | 12 | *** | *** | 45 |

各食材の個体ごとに、正しく識別された分析用セグメントの割合を表6に示す。

キャベツ、人参、キュウリに関しては全ての個体において、それに含まれる特徴ベクトルの過半数を正しく分類した。また、2番目に確からしいと判断された食材との差も大きく、多数決を取った場合は正しい食材として識別された。

表 6: オープンテスト

| 食材/個体番号 | 1 | 2 | 3 | 4 | 5 | 6 | 平均 |
|---------|-------|-------|-------|-------|-------|-------|-------|
| キャベツ | 61.90 | 50.00 | 50.00 | 53.85 | *** | *** | 53.68 |
| 人参 | 90.24 | 72.22 | 85.71 | 72.22 | *** | *** | 80.25 |
| キュウリ | 87.50 | 85.71 | 72.73 | 52.38 | *** | *** | 75.79 |
| 玉葱 | 53.33 | 70.59 | 34.62 | 66.67 | 39.12 | 69.23 | 54.96 |
| ピーマン | 50.00 | 38.10 | 42.86 | 66.67 | 66.67 | *** | 53.33 |
| ジャガイモ | 75.00 | 41.67 | 50.00 | 45.83 | *** | *** | 52.86 |
| トマト | 0.000 | 22.22 | 33.33 | 16.67 | *** | *** | 17.78 |

ピーマンに関しては、過半数に達しない固体もあったが、いずれも多数決を取った場合は正しく分類された。玉葱とジャガイモに関しては、過半数に達しない個体がいづらか存在した。いずれの個体も、玉葱、ジャガイモと判断した特徴ベクトルの数が多く、この2種類で識別を誤る場合がよく見受けられた。ただし、この2種類のどちらかであるということがわかれば、改めて2種の食材による識別機に投入することで正しく識別することができる。トマトに関しては個体間での熟し方の違いにより、スペクトログラムにかなりばらつきがあったため、識別率は低かったのではないかと考えられる。

5.2 映像等により絞り込めた場合の識別能力

[実験3：玉葱とジャガイモの識別]

映像等により、調理者の切る食材がレシピ中である程度絞られた場合について、本手法が補完的役割を果たすかどうかを調べる。映像により色特徴から食材が絞られる場合が想定される。従来研究より、レシピに頻繁に登場し、映像ではなかなか識別が難しい対として玉葱とジャガイモの組が挙げられている。そこで、調理者の切っている食材が、玉葱またはジャガイモのどちらかであると仮定して、その上でどの程度識別できるのかを調べる実験を行なう。玉葱とジャガイモの2種類の食材のみで、実験2で集めた波形を用いた。個体数はそれぞれ6、4であり、各個体ごとの分析用セグメント数は表5の玉葱、ジャガイモの欄を参照されたい。1つの個体をテスト用データとして順に取り出し、残りの7個体の分析用セグメントにより学習した識別機で、どの程度の精度で識

別できるのか調べる交差検定を行なった。テスト用データの一連の分析用セグメントのうち、正しく識別されたものの割合 (%) を表 7 にしめす。

表 7: 玉葱とジャガイモの識別

| 食材/個体番号 | 1 | 2 | 3 | 4 | 5 | 6 | 平均精度 |
|---------|-------|-------|-------|-------|-------|-------|-------|
| 玉葱 | 80.00 | 76.47 | 84.62 | 79.17 | 82.61 | 88.46 | 82.44 |
| ジャガイモ | 87.50 | 75.00 | 83.33 | 79.17 | *** | *** | 81.43 |

どの個体に関しても、一連の分析用セグメントのうち、75%以上のものを正しく識別できた。よって、多数決をとることで、いずれの場合も正しく識別できたことになる。この結果より、玉葱とジャガイモという2種類の食材に絞られた場合は、本手法により識別できる可能性が高いと言える。視覚的にはうまく識別できないこの食材の組は、内部構造に結構な差がある。玉葱は内部が層構造になっており、分析用波形の複数の時点で振幅の極地を取るのに対して、ジャガイモは均一な構造であり、まな板に衝突するまでの間は一定に近い振幅を取る。本手法により、その内部構造の差をうまく利用できた結果ではないかと考えられる。

[実験 4 : キュウリとピーマンの識別]

映像等により、調理者の切る食材がキュウリかピーマンのどちらかであると絞られた場合について実験を行なった。調理者の切っている食材がキュウリまたはピーマンのどちらかであると仮定して、実験 2 で集めた振動音波形を用いて実験を行なった。結果は表 8 の通り。

表 8: キュウリとピーマンの識別

| 食材/個体番号 | 1 | 2 | 3 | 4 | 5 | 平均精度 |
|---------|-------|-------|-------|-------|-------|-------|
| キュウリ | 95.83 | 92.86 | 81.82 | 90.48 | *** | 91.58 |
| ピーマン | 88.89 | 76.19 | 90.48 | 85.71 | 75.00 | 81.90 |

どの個体についても、一連の特徴ベクトルのうち、75.00%以上のものを正しく識別できた。よって、多数決をとることで、いずれの場合も正しく識別でき

たことになる。しかし、本実験は2種類の食材のいずれかであることから、でたらめに識別したとしても50%の割合であたる。そこで、より高い精度の識別率を示すことが重要となる。この結果より、キュウリとピーマンの識別の可能性は高いと言える。単純な色特徴によりこの二つに食材が絞れた場合は、本手法を補完的に作用させることで、識別が可能となるだろう。ピーマンの内部は空洞であり、水分を多く含むキュウリとはその内部構造は大きく異なる。この違いが本手法により反映されていると考えられる。

5.3 考察

実験1によるクローズドテストの結果、いずれの食材も75%以上の精度で正しい領域に分類されることがわかった。これにより、本実験で選んだ7種の食材を同時に扱うようなレシピに関しては、全ての食材の識別の可能性があると言える。各食材とも、さらに多くの個体を集め、分析用セグメントを数多く抽出することで、より頑強な識別機の構築を目指すことが重要である。調理の際に同時に扱う様々な食材の組み合わせを想定し、特徴空間のクラス間分離面、識別率がどのように変化するかを調べていくことで、どの程度のレシピをカバーすることができるのかを調べる必要がある。

実験2により、未知データによるテストを行なった結果、トマト以外の食材については正しく識別できる可能性が高いことがわかった。しかし、個体間の識別率の差が大きく、識別機に用いた学習サンプル数がまだまだ足りないことがわかる。トマトに関しては個体差がかなり顕著であり、完熟していて、切ることで潰れてしまうものや、比較的堅いものまであり、抽出した分析用セグメントのスペクトログラムは大きな違いが見られた。こういった個体差は程度の差こそあれ全ての食材に見受けられるものである。より多くの個体を用意し、分析していくことが重要である。

5.2節の実験3, 4より、食材の種類が少ない場合はより高い精度の識別機を構成できることがわかった。レシピで扱う食材全体での識別において、過半数を獲得しなかった個体や、2番目に確からしいと判断された食材との差が小さな個体に関しては、その時点で食材を特定するのではなく、上位2、もしくは3種の食材で構成された識別機で食材を特定する方がより正確かもしれない。また、この結果により、本研究の振動音を用いた手法を、映像等による食材識別システムと補完的に使うことの可能性が示されたと言える。

6 まとめと今後の課題

本研究では、食材加工に起因する振動音を用いた食材識別について検討するために、まず振動音の定義を行い、その概要について説明した。それから、振動音波形の振幅のみで食材の識別がどの程度できるのかについて調べた。しかし、振幅だけでわかる情報はあまり無いという結論に至った。次に振動音波形から、包丁が食材内部を通過している部分に注目し、その部分を抽出し、スペクトログラム分析することで、得られたスペクトログラムの低周波部分に食材固有の特徴があるのではないかという結論に至った。特徴量を見出す課程で、具体的な識別システムのアルゴリズムの構想ができあがり、そのシステムでの識別能力を調べた。その結果、本研究で提案した識別システムでの食材識別の可能性が示された。ただし、特徴量についてはさらにサンプル数を増やし、より細かく分析することで、より識別に適したものが見つかる可能性は十分にある。スペクトログラムのみならず、他の振幅との組み合わせなども考慮する必要があるだろう。今後も有効な特徴量について分析していくことが重要である。また、より頑強な識別機を設計するためには、多くの食材サンプルが必要である。食材によっては個体差が大きく、同じ食材の中でも種類を分ける必要があるものも出てくるかもしれない。個体差をできるだけ吸収するような特徴量を選ぶことが重要であり、特徴量分析と、識別機の設計は同時並行で行なっていくべきであろう。

謝辞

本研究を行なうにあたり多大な御指導を頂いた美濃導彦教授、角所考助教授に深く感謝致します。また、日頃より熱心に指導して下さい、本報告書の作成においても御指導を頂きました美濃研究室の山肩洋子氏をはじめとする、認識グループ、並びに美濃研究室の皆様へ深く感謝しております。

参考文献

- [1] 山肩洋子, 尾原秀登, 角所考, 沢田篤史, 美濃導彦: 観測データに基づく生産加工活動認識のための物体・操作・目的情報の統合利用, 電子情報通信学会技術研究報告 PRMU2003, Vol. 300, pp. 95–100 (2004).

- [2] 野菜の種類: <http://www.h6.dion.ne.jp/~chusan55/shurui/1shurui.htm>.
- [3] Auditory Toolbox: <http://rvl4.ecn.purdue.edu/~malcolm/interval/1998-010/AuditoryToolboxTe%chReport.pdf>.
- [4] 栗田 多喜夫: サポートベクターマシン入門, <http://www.neurosci.aist.go.jp/~kurita/lecture/svm.pdf>.