複数センサ統合による調理中の状態変化認識

宮島　沙織、井川　　久

Cooking State Recognition Using Integrated Multiple Sensors

Saori MIYAJIMA, Hisashi IGAWA

抄　録

労働力不足の解消には、人間の複合的な判断力が求められる作業にも適用できる、高度な状態認識技術の開発が不可欠である。状態認識技術の高度化の方法の一つとして、複数の異なる種類の情報を組み合わせるマルチモーダル化が挙げられる。

そこで本研究では、人間の五感のように、複数の異なるセンサから得られた情報を統合した状態認識の実現を目指し、具体的なタスクとして加熱調理されている肉の焼き加減の認識に取り組んだ。加熱による肉の状態変化を計測可能なセンサを選定し、加熱調理中のデータを計測する実験を行い、計測したデータを用いて長期短期記憶ネットワークを学習することで、3段階の焼き加減の認識を行った。さらに、単一のセンサ情報のみを使用した場合と認識精度の比較を行うことで、マルチモーダル化の有用性を検証した。

キーワード：状態認識、マルチモーダル情報処理、センシング、時系列情報処理、AI

1. はじめに

労働力不足を解消するためには、より幅広い作業に適用可能な自動化技術の開発が必要である。近年の深層学習の活用により、判断基準の数値化や定量化が困難な認識タスクについても自動化が進みつつある。しかし、未だに人間にとっては簡単でも、自動化が困難な認識タスクも多く存在する。例えば製造品の不良検査において、表面のゆるやかな凹凸や質感の違いなど画像上で判別することが困難な差異も、作業者は目視と手で触れた感覚を組み合わせて判別できる。また農作物の選果作業でも、撮影が困難な内部の腐敗に対して、作業者は柔らかさや匂いをもとに判別することが可能である。このように、人間は意識することなく五感を組み合わせて判断することで、画像など単一の情報のみを判別に利用している既存装置では検出困難な異常を認識することが可能である。そこで本研究では、人間の五感のように複数の異なるセンサから得られた情報を統合（マルチモーダル化）することで、単一のセンサ情報のみでは困難なタスクの状態認識を目指す。複数種類の感覚にもとづいて判断する作業には、農作物の選果作業、楽器の調整作業など様々あるが、特に身近な作業として調理作業が挙げられる。調理作業は、加熱や泡立てなど、視覚情報のみでは正確な状態把握が困難な場合が多い。そのなかでも肉類の加熱調理においては、加熱が不十分の場合は食中毒を引き起こす恐れがあり、過度に加熱すると食感や風味を損なってしまうため、適切な加熱が求められる。先行研究においても、外観から内部が衛生上十分に加熱されたかを判断することは困難であることが示されている[1)]。一方で人間が肉の加熱調理を行う際は、外観だけではなく音や加熱面の温度、香り、蒸気や煙の発生など、複数の要素を統合して焼き加減を判断している。

以上により、単一の情報では認識が困難な肉内部の加熱状態の認識は、マルチモーダルセンシングを活用した状態認識のタスクとして適切であると考え、本研究では加熱調理中の肉の状態変化、すなわち焼き加減の認識に取り組んだ。

2. センサの選定と計測環境の構築

2.1　認識対象とする状態変化

肉の状態変化を認識するためには、まず変化がどのように現れるかを把握した上でその変化を計測可能なセンサを選定する必要がある。食肉とは家畜の筋肉であり、タンパク質、水分、脂肪その他で構成される。したがって、加熱により肉

事業名：職員研究奨励事業
課題名：複数種の感覚を利用した状況認識手法の検討

に生じる変化は、主に脱水とタンパク質の変性である。

食肉に含まれるタンパク質は大まかに、糸状の構造をもつ筋原線維タンパク質、結合組織を構成する筋基質タンパク質、液体である筋漿タンパク質に分けられる。筋原線維タンパク質の主成分であるミオシン・アクチンと、筋基質タンパク質の主成分であるコラーゲンは加熱によって収縮し保水性が下がる。この変化は肉の外観の変化のほか、水分の放出による音の変化および蒸気の発生として知覚可能である。

肉の色は筋漿タンパク質であるミオグロビンに含まれる鉄の状態に依存し、加熱により鮮やかな赤色になったのち、褐色へと変化する[2, 3]。熱源に接している面は内部よりも高温となり、約140℃でメイラード反応を生じ、約160℃からカラメル化して褐色となり、約200℃で炭化して黒くなる。色の変化は肉の外観から知覚できるうえ、メイラード反応やカラメル化では香気成分の発生を伴うため[4]、肉周辺の匂いから知覚できる。さらに、肉周辺の雰囲気温度は肉やフライパンなどの熱源の温度変化と関係していると考えられる。

2.2　センサの選定

先述のように、加熱により生じる変化は外観・湿度・温度・音・匂いの変化として計測可能と考えられる。そこで、これらの情報を計測するためのセンサ類を選定した。

外観の計測用として産業用カメラ（acA1300-200uc・Basler製）を、音の計測用としてマイク付きのWebカメラ（C922n・Logicool製）を使用した。匂いを計測する装置として、特定のガスを検出するセンサではなく、複数種類の匂い分子を検出可能な匂いセンサのうち、表1に示す3機種を候補として選定のための予備実験を行った。焼く前の肉・適度に焼いた肉・焦げた肉を常温の状態で密閉可能な保存袋に入れ、匂いセンサの検出器を同じ袋に入れて10分程度放置し、匂いを計測した。計測後、3状態の肉について計測値を比較し、区別が可能であるかを検討した。

さらに、本実験では加熱中の匂いの変化を計測するため、加熱時に発生する蒸気をセンサ素子に触れさせる必要がある。そこで、肉の加熱中に発生した湯気の温度を計測したところ、60℃から70℃程度であった。

表1　匂いセンサの比較

製品	検出素子	使用可能温度	区別
空気質センサ / Bosch製	金属酸化物	-45～85℃	○
MSSセンサ / Qception製	感応膜	0～55℃	△
nose@MEMS / I-PEX製	感応膜	0～40℃	○

予備実験の結果と各センサの使用可能温度を表1に示す。表の結果をもとに、各状態の肉の区別が可能かつ本実験で想定する70℃程度の環境下での使用が可能であることから、本研究では空気質センサ（BME688・Bosch製）を採用した。このセンサでは、匂いの変化を電気抵抗値［Ω］として計測する。また、空気質センサには温度センサ・湿度センサが内蔵されており、これらを利用して肉周辺の空間の温度と湿度を計測する。

以上により、外観・湿度・温度・音・匂いから加熱中の肉の状態変化を計測する機器を選定した。これらの機器を使用し、実験環境を構築した。

2.3　実験環境の構築

肉の加熱調理中は油を含む蒸気や煙が発生し高温になることから、安全確保のためにドラフトチャンバー内に実験環境を構築し、電気式のホットプレートを調理器具として使用した。ホットプレートの周辺に前節で選定した機器類を配置し、図1に示す実験環境を構築した。図中の肉の内部温度を計測する温度計（ステンレス保護管温度センサTR-0406・T&D製）は、焼き加減の正解値を得るためのリファレンスとしてのみ使用し、状態認識の入力としては使用しなかった。

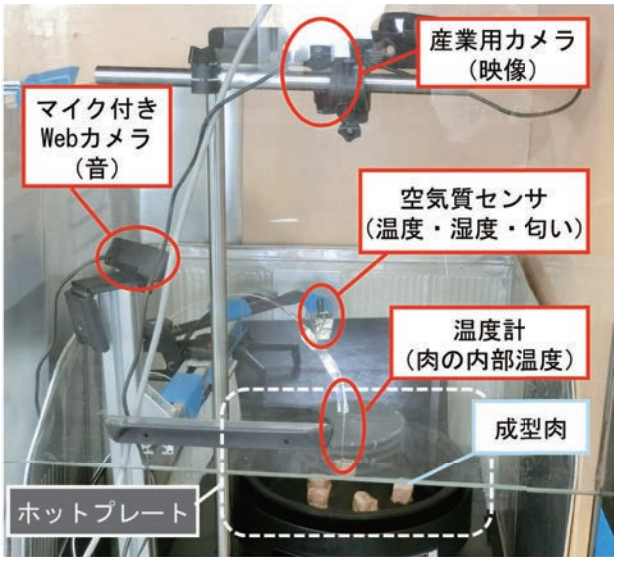

図1　構築した実験環境

3. 加熱中の肉の状態変化計測実験

構築した実験環境において、選定した装置により加熱調理中の外観・湿度・温度・音・匂いの変化を計測する実験を行った。加熱対象として、成分と大きさが均一な牛の成形肉（サイコロステーキ、１辺約20mmの立方体）を使用した。

実験では、次の手順に沿って成形肉を加熱した。

①ホットプレートを十分に加熱する。

②半解凍状態の成形肉を５個ホットプレート上に並べ、うち１個に温度計のプローブを刺す。

③内部温度を確認しながら肉を加熱する。途中、全面に焼き色がつくよう適宜焼き面を変更する。

④推奨加熱条件を満たした時刻を確認後、２個の肉を回収して加熱を続ける。

⑤肉の内部温度が80℃を超え、肉全体が焦げて黒色になったら加熱を終了する。

このような加熱実験を９回実施し、次章で実施する焼き加減の状態認識に使用する学習データを収集した。なお、空気質センサは１回の計測ごとに外気に30分当て、リフレッシュを実施した。また、空気質センサとリファレンス用温度センサのサンプリング周期は1sとした。

加熱前の肉、手順④で回収した肉、加熱終了後の肉を図２に示す。図より、加熱により肉が収縮し、色が変化していることがわかる。また各状態の肉の重量を計測（10個平均）すると、加熱前は１個あたりの重量が9.9gであったのに対し、手順④で回収した肉は１個あたり5.9g、加熱終了後の肉は１個あたり3.4gと、加熱によって重量が減少した。

図２　加熱前後の肉の外観比較

実験中の湿度と音の変化、匂いと内部温度の変化、さらに実験初期・中期・後期の画像の一例を図３に示す。図上側の音波形は、後述するハイパスフィルタで調理音以外のノイズを除去したものである。計測した音、湿度、映像を確認すると、肉の焼き面を変えた時刻付近で音が変化し、湿度が一時的に上昇したことから、肉が熱源に触れたことで収縮し水分が放出された変化を計測したものと考えられる。また、匂いセンサの変化を確認すると、全９回の実験に共通して、加熱開始後から値が減少し、肉全体が焦げた時刻付近で値が増加する傾向が見られた。

以上の結果より、選定したセンサ類によって加熱による肉の状態変化がデータとして計測された。このデータを使用し、深層学習による状態変化の認識を行う。

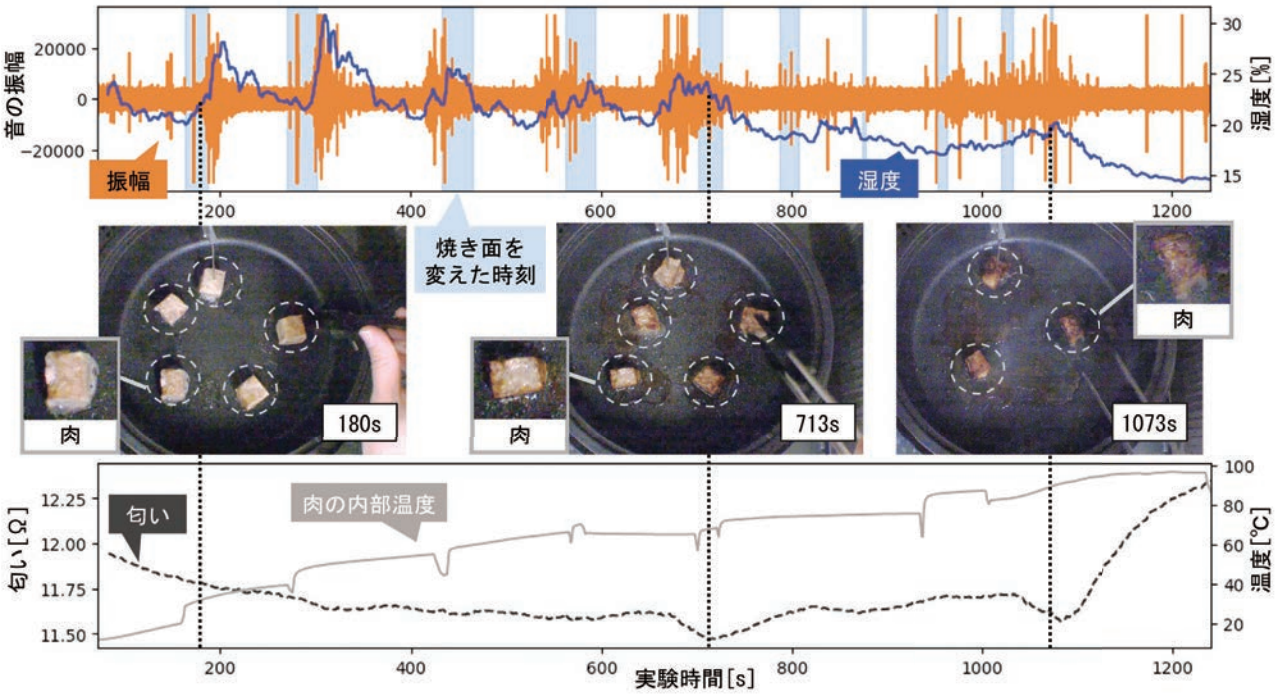

図３　音量と湿度の変化（上グラフ）、内部温度と匂いの変化（下グラフ）および撮影画像

4. LSTMによる状態変化認識

4.1 認識対象の状態定義

牛の成形肉には内臓などの部位も含まれているため、食品衛生の観点から内部温度70℃で3分またはそれに相当する温度・時間の組み合わせによる加熱が推奨されている。本研究ではこの推奨加熱条件にもとづき、認識対象とする状態を次の3状態と定義した。

A. 生焼け：肉が内部温度70℃で3分加熱される前の状態
B. 適度焼け：肉が内部温度70℃で3分加熱された後、内部温度80℃以下の状態
C. 焼きすぎ：肉の内部温度が80℃以上の状態

計測した内部温度をもとに各時刻の状態をA～Cに分類し、状態認識の真値とした。

4.2 学習データの作成

計測した実験データを用いて、学習データを作成した。計測したデータのうち温度・湿度・匂い・音は1次元のデータであるのに対し、映像は1フレームあたり横1280画素・縦1024画素の高次元情報を持つため、そのまま使用すると含まれる情報の大部分が映像の情報となり、他の1次元センサデータの特徴が反映されない可能性がある。また、音データも計測した波形データのままでは特徴の把握が困難である。そこで、音データと映像データの前処理を行い、それぞれ2次元の特徴量を抽出した。

a）音データの前処理

スペクトログラム表示により調理中の音の変化を確認したところ、ドラフトチャンバーの換気音の音量が大きく、調理中の音が隠れてしまっている状態が確認された。そこで、まずはドラフトチャンバーの換気音である低周波数成分をハイパスフィルタで除去した。ハイパスフィルタ適用前後の音の振幅（音量）の変化を図4に示す。

次に映像と音を合わせて確認したところ、肉から水分が放出された際に音量と音の高さ（周波数成分）が変化する傾向が観測されたため、音量に相当する振幅の二乗和（RMS）と、音の周波数成分の特徴に相当するスペクトル重心の2つを特徴量として抽出した。

b）映像データの前処理

2.2節で述べたように、加熱により肉は変色する。まず約40℃で色が鮮やかな赤色になり、高温になると褐色、灰褐色と変化する。また、肉の表面の焼き色は高温になると茶色から黒色に変化していく。したがって、加熱による色の変化は明度と彩度の変化に現れると考えられる。明度と彩度の変化を計測するため、画像から肉の領域を抽出し、肉の色の彩度と明度を特徴量とした。肉の領域抽出には物体検出用のAIモデルであるYOLOv11を学習し使用した。YOLOにより検出された肉の領域内について、各画素値をHSV表色系に変換し、その平均値を算出することで明度と彩度を求めた。明度と彩度を求める手順を図5に示す。

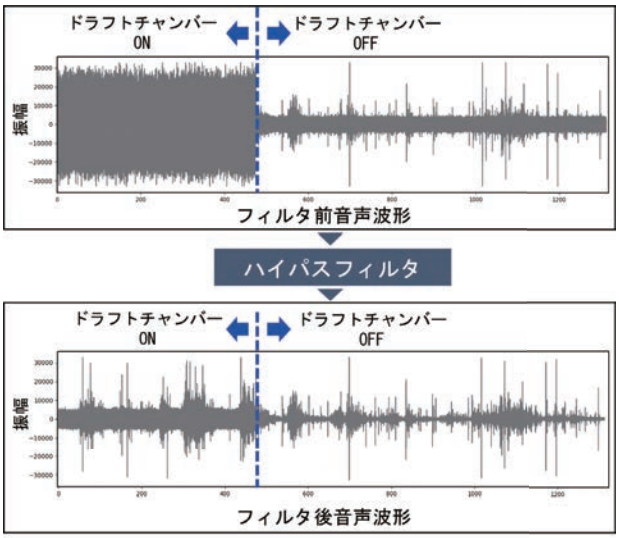

図4　ハイパスフィルタによる排気音低減効果

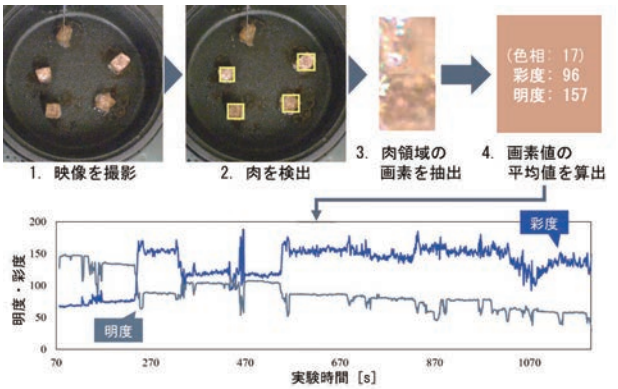

図5　映像から特徴量を抽出する手順

以上の処理によって抽出されたRMS・スペクトル重心・明度・彩度と、計測した温度・湿度・匂いを組み合わせた7次元のデータを使用して学習データを作成した。サンプリング周期は1sに統一した。

4.3 LSTMネットワークの学習

状態認識には時系列データ処理に適した長期短期記憶（LSTM: Long Short Term Memory）ネットワークというネットワークモデルを使用した。本研究ではTensorflowを用いて2つのLSTM層と1つの全結合層からなるネットワークを実装した。1秒の状態認識に使用する区間（Lookback）を直前の30秒分とし、9回分の実験データから約7000セットの訓練データを得た。このデータを使用して、バッチサイズ64、エポック数200としてLSTMネットワークの学習を行った。

実施した9回の実験のうち8回分を訓練データ、残り1回分はテストデータとし、訓練データ・テストデータの全組み合わせ9通りについて交差検証を行い、状態認識精度を評価した。

4.4 LSTMによる状態認識結果

認識精度を評価する指標として、機械学習分野において一般的に用いられる正解率（Accuracy）、適合率（Macro Precision）、再現率（Macro Recall）、F1値（Macro F1）を使用した。これらの評価指標はそれぞれ次の式(1)～(4)により算出される。

$$\text{Accuracy} = \frac{\sum_{l=1}^{L} \text{TP}_l}{N} \tag{1}$$

$$\text{Macro Precision} = \frac{1}{L}\sum_{l=1}^{L} p_l, \qquad p_l = \frac{\text{TP}_l}{\text{TP}_l + \text{FP}_l} \tag{2}$$

$$\text{Macro Recall} = \frac{1}{L}\sum_{l=1}^{L} r_l, \qquad r_l = \frac{\text{TP}_l}{\text{TP}_l + \text{FN}_l} \tag{3}$$

$$\text{Macro F1} = \frac{1}{L}\sum_{l=1}^{L} \frac{2 r_l p_l}{r_l + p_l} \tag{4}$$

式中のNは全データ数である。Lは認識する状態の個数で、今回は4.1節で述べた3状態A ～ Cが認識対象であるため、L＝3である。またTP_l・FN_l・FP_lは認識結果の分類で使用される指標であり、認識対象A ～ Cのうちのある状態 l について実際の状態と認識結果の組み合わせから次のように定義されている。

- TP_l：実際も認識結果も状態 l であるデータの数。
- FP_l：実際は状態 l ではないが、状態 l と誤認識されたデータの数。
- FN_l：実際は状態 l だが、状態 l と認識されなかったデータの数。

評価指標を計算した結果、全体の正解率は0.83、適合率は0.69、再現率は0.71、F1値は0.70であった。状態別（A・B・C）に適合率、再現率、F1値を求めた結果を表2に示す。表2より、状態別の評価指標を見ると、状態A（生焼け）の各評価指標は全て0.9以上だが、状態B（適度焼け）は0.5以下と低い値となっており、状態別の認識精度に差が生じていた。さらに、図6に示す混同行列より、データ全体に占める状態BとCの割合が低く、誤認識が起こりやすいのはAとB、BとCの隣り合った状態同士が多いことがわかる。

これらの結果より、提案手法によって生焼け（状態A）は精度よく認識されたが、適度焼けや焼きすぎを正しく認識するには改良が必要である。

4.5 マルチモーダル化の効果検証

次に、単一のセンサ情報のみを使用した場合と、提案手法である複数種類のセンサ情報を統合して使用した場合の認識精度を比較した。表3に、使用した入力情報別の正解率、適合率、再現率、F1値の算出結果を示す。表3より、正解率、適合率、再現率、F1値の全てにおいて、単一のセンサを使用した場合よりも提案手法の分類精度が高い結果となった。

一方で、交差検証の試行別に求めた評価指標を標本として

表2　状態別の認識精度

状態	適合率	再現率	F1値
全状態平均	0.697	0.713	0.703
A	**0.962**	**0.924**	**0.942**
B	0.371	0.476	0.417
C	0.759	0.738	0.749

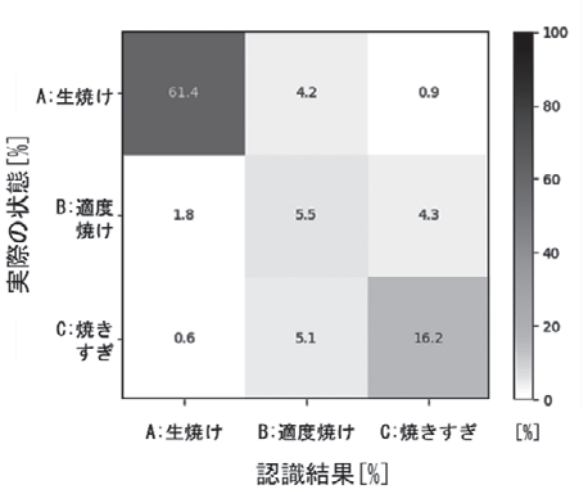

図6　全データ統合時の混同行列

表3　入力データ別の認識精度比較

入力	正解率	適合率	再現率	F1
全センサ	**0.831**	**0.697**	**0.713**	**0.703**
空気質	0.768	0.645	0.672	0.656
音	0.633	0.508	0.501	0.492
映像	0.801	0.667	0.653	0.660

優位水準を0.05とした有意差検定を行った結果、音データのみの場合と提案手法の比較では有意差が認められたが、映像のみ、空気質センサのみの場合と提案手法の比較では有意差が認められなかった。したがって、映像データと空気質データ（温度・湿度・匂い）については、他のセンサデータと組み合わせることで認識精度が有意に向上したとは言えない。

そこで、他の視点から提案手法であるセンサ統合の効果について考察する。有意差が認められない原因として、データのばらつきが大きいことが考えられる。試行別に求めた正解率の値を図7に示す。図7より、単一のセンサ情報のみの場合では試行により値のばらつきが大きいが、提案手法では大きな上下は見られない。表4に示す各評価指標の標準偏差を比較すると、数値上でもセンサ統合により認識精度のばらつきが低減したことが確認できる。

また、状態別（A・B・C）に適合率、再現率、F1値を求

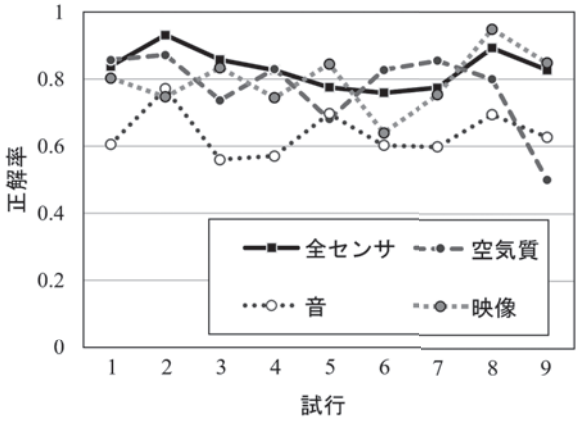

図７　試行別の正解率比較

表４　入力データ別の標準偏差比較

入力	正解率	適合率	再現率	F1
全センサ	**0.057**	**0.081**	0.121	0.095
空気質	0.120	0.119	0.122	0.123
音	0.070	0.082	**0.071**	**0.066**
映像	0.088	0.127	0.090	0.097

めた結果を図８に示す。図８より、映像のみの場合では状態B（適度焼け）の認識精度が低いが、他のセンサデータと統合することで認識精度が向上したことが確認できる。

以上により、複数種類のセンサを統合（マルチモーダル化）して状態認識を行うことで、各センサの不得意な領域を補い合い、精度よく安定した状態認識が可能となった。

本結果より、マルチモーダル認識を活用することで、ネットワークの構造を複雑にすることなく、認識精度の改善が可能であることが確認された。マルチモーダル化の効果を向上するには、データ統合時の重みの検討や、センサの選定が重要である。今回は各センサデータのみでも一定の認識精度が得られたため重みの検討は行わなかったが、極端に認識精度が低いセンサデータが含まれていた場合は、重みを変える、またはそのセンサデータを除外することを検討すべきである。

5. おわりに

本研究では、人間のように複数種類の感覚にもとづく状態認識の実現を目指し、加熱調理中の肉の焼き加減の認識をタスクとして設定した。調理中に計測した映像・湿度・温度・音・匂いのデータを用いてLSTMネットワークを学習することで、83%の正解率で成形肉の焼き加減認識に成功した。

今回は調理作業を対象としたが、本研究で取り組んだ複数種類のセンサデータを統合する状態認識手法は、農作物の選果、プラントの異常点検など様々な作業における状態認識タスクに適用可能である。今後は、これら他の作業への適用や、ロボットの強化学習への展開などに取り組む。

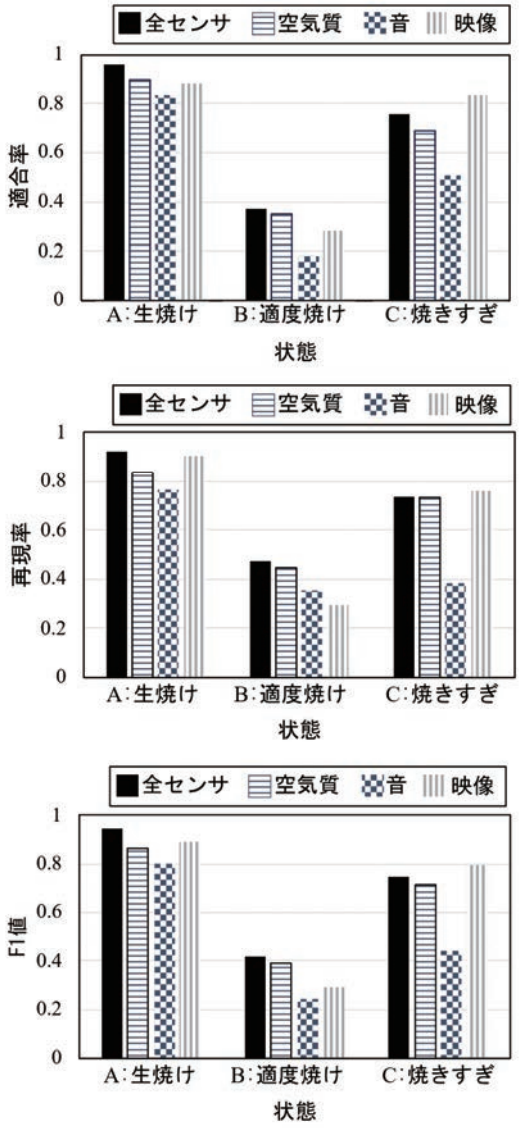

図８　状態別の認識精度比較

参考文献

1）内閣府食品安全委員会：令和２年度内閣府食品安全委員会調査事業報告書，cho20210040001,(2021)
2）泉本　勝利：岡山大学農学部学術報告，Vol.81, No.1 pp.81-100,(1993)
3）吉留 大雅，平井 経太，堀内 隆彦：日本色彩学会誌，Vol.41, No.6，p. 53-56,(2017)
4）臼井　照幸：日本食生活学会誌，Vol.26, No.1，pp.7-10,(2015)