自然な食事環境下で収集した食事音声データによる食事詳細行動分類手法の提案

蒲地 遥[1] 近藤 匠海[1] 横窪 安奈[1] ロペズ ギヨーム[1]

概要：早食いの人ほど BMI が高いことと，食事中の会話が少ないと肥満の傾向があることが分かっている．そのため，食事中の咀嚼回数と会話を増やすことが望ましい．一方，食事行動の識別は実験環境下でしか行われていない．そこで本研究では，自然な食事環境下での食事行動の定量化を目的とし，自然な食事環境下で収集した食事音声データを利用して食事詳細行動の分類を行う．骨伝導マイクロフォンを用いた食事行動分類の研究は今までにも行われているが，リサンプリングのタイミングによる学習モデルの過学習の可能性がある．また，分類する行動の種類が十分でない．この研究では，日常的な食事環境での食事音声データを収集し，分類手法を評価する．

Classification Method of Eating Behavior by Dietary Sound Collected in Natural Meal Environment

HARUKA KAMACHI[1] TAKUMI KONDO[1] ANNA YOKOKUBO[1] GUILLAUME LOPEZ[1]

1. はじめに

肥満は生活習慣病を引き起こす恐れがある．厚生労働省は肥満予防のために対策を講じてきたが，肥満の患者数は 10 年前から減少していない [1]．また，早食いの人ほど肥満の基準となる BMI が高い傾向にあることが示された [2] ことから，肥満の防止にはゆっくりとよく噛んで食べることが重要であると考えられる．さらに，食事中に会話がある場合，生活習慣が規則正しく好き嫌いが少ないなど健康と関連があることから [3]，食事中の会話を増やすことが望ましい．

近年，市販されているウェアラブルデバイスにより，一日の消費カロリーの測定やこれに関連した人間の活動レベルのモニタリングが可能である．しかし，自然な食事環境下での食事行動を自動的に検出するデバイスはまだ市販されている状況にない．食事中の咀嚼や発話などの行動が検出可能となることで，咀嚼回数や食事中の会話時間をリアルタイムで食事者に提示可能となり，咀嚼回数の増加や発話意識の向上など食事行動への意識の改善が期待できる．

よって本研究では，自然な食事環境下での食事行動の定量化を実現することを目的とし，自然な食事環境に対応した食事詳細行動の高精度分類を目標とする．自然な食事環境とは，日常生活における食事環境とする．

2. 関連研究

音声を用いない食事行動解析・認識手法として特製デバイスを使用したものがある．Chun らは，食事検出のためのネックレスデバイスを提案した [4]．近接センサで顎骨の動きを捉え，食事と非食事行動を区別している．他に，メガネ型やヘッドバンド型デバイスも提案されているが，市販されていて誰でも容易に手に入るデバイスではないため，導入が難しい．また食事行動の認識は行われているが，食事の詳細な行動の分析はできていない．

Zhang らは，骨伝導マイクロフォンによる音情報を利用した食べ物のテクスチャ分析を行った [5]．高精度でのテクスチャ分類が達成されているが，実験環境下での実験のみであった．近藤らは，自然な食事環境下で骨伝導マイクロフォンを用いて食事音声データを収集し，咀嚼・嚥下・発話・その他の 4 種類の食事行動の分類を行った [6]．このとき，咀嚼のデータ数に対して他のデータが少ないことから Syn-

[1] 青山学院大学
Aoyama Gakuin University

thetic Minority Oversampling TEchnique (SMOTE) を利用し，データの均衡化を行っている．特徴選択を行った 7 個の特徴量で平均 92 %の精度，14 個の特徴量で 95 %以上の精度という結果になった．しかし，1 口の目安となる食べ物の嚥下は飲み物の嚥下と区別されていない．また，データの均衡化のために訓練データとテストデータへの分割前にオーバーサンプリングを行っているため，実際には存在しないデータをテストデータとして用いている可能性があり過学習が起きていることが考えられる．

以上から，本研究では自然な食事環境下での嚥下を区別したより詳細な食事行動 (咀嚼・発話・食べ物の嚥下・飲み物の嚥下) の分類を行う．そのため，耳の内側に配置が可能で容易に手に入る市販の骨伝導マイクロフォンを利用し，食事音声を収集する．

3. データセット作成と分類手法

本研究では自然な食事環境下での食事行動の分類を行うために，自然な食事環境下での食事音声データを収集した．リアルタイムの分類は行わないため，咀嚼・食べ物の嚥下・飲み物の嚥下・発話・その他に該当する音声区間を手動でラベリングし，分類に用いた．

3.1 自然な食事環境下での食事音声データ収集

食事音声データの収集にはスマートフォンと Bluetooth 通信を行う骨伝導マイクロフォンを使用した．スマートフォンは Google 社製の Google Pixel 3，骨伝導マイクロフォンは Motorola 社製の Finiti HZ800 Bluetooth Headset を用いた．Android アプリケーションソフトウェアにより食事音声データを収集した．

データ収集は 11 歳から 23 歳の男女合計 16 人を被験者とした．図 1 に示すように，被験者の片耳に骨伝導マイクロフォンを装着して行った．また，音声収集後に行うラベリング作業を補助するために，音声データと一緒に動画を撮影した．動画を撮影する際，被験者の口と喉が映るように撮影した．データ収集は食堂や一般家庭の食卓，研究室など自由な食事環境下で行い，被験者には日常生活と同じように食事することを求めた．

3.2 音声データのラベリング

収集したデータの整理は，機械学習による 5 種類の食事行動の分類モデル作成のための真値のデータセット作成を目的とした．真値のデータセットを作成するために，収集した音声データの咀嚼・食べ物の嚥下・飲み物の嚥下・発話・その他に該当する区間をラベリングした．

ラベリングする際に，音声のみでは各食事行動のラベリングを行うことが困難であるため，音声データと音声データ収集と同時に撮影した動画を同期させてラベリングを行った．撮影した動画は同期させることにより，映像音声

図 1 データ収集の様子

表 1 各ラベルの合計ラベルデータ数

ラベル名	咀嚼	食べ物の嚥下	飲み物の嚥下	発話	その他
合計	2001	119	83	555	201

から骨伝導マイクロフォンにより収集された音声データの音声に変換した．

音声データのラベリングは，音声データにラベルを付与できる音声分析用ソフトウェアの Praat[7] を用いて行った．5 つのラベルを設定した：咀嚼，食べ物の嚥下，飲み物の嚥下，発話，その他．それぞれのラベルを付与された音声区間のみを抽出したデータのまとめを表 1 に示す．

3.3 特徴量抽出

抽出した特徴量について表 2 に示す．咀嚼，嚥下と比べ発話の時間が長く振幅数が多いと考え，生データから生データ時間と振幅のピーク数，零交差数を抽出した．音声信号の特徴化に用いられる Amplitude Difference Accumulation (ADA) や Short term energy (STE) をかけ，それぞれ変換されたデータのデータ値の合計，最大値の 2 個ずつを特徴量として抽出した．また，生データを特徴化させ，咀嚼，嚥下，発話の振幅数に違いが出たことから，移動平均をかけたあとの零交差数を抽出した．

音声認識に頻繁に用いられているため，生データのパワースペクトル密度も特徴量として利用した．分類精度向上のために，咀嚼，嚥下の生データからそれぞれ選択されたデータとすべての生データ，パワースペクトル密度を相互相関にかけた．それぞれの相関結果の最大値を特徴量として利用した．本研究ではさらに，音声認識で頻繁に利用されるメル周波数ケプストラム係数 (MFCC) も特徴量として利用した．先行研究より本研究では，39 次の MFCC を抽出した [8]．以上より合計 75 個の特徴量を抽出した．

3.4 不均衡データセットの均衡化

ラベリング作業によって作成されたデータセットは表 1

表 2　抽出された 75 個の特徴量

特徴カテゴリ	説明	特徴数
生データ	データの長さ ピーク数 零交差数	3
パワー	最大値	1
Amplitude difference accumulation	全データの合計 最大値	2
Short term energy	全データの合計 最大値	2
移動中央値	零交差数	1
パワースペクトル密度 (PSD)	特定の周波数範囲の合計 中央周波数 特定の周波数範囲のバンドパワー 特定の周波数範囲の最大値になる周波数 一番と二番目に大きいピーク値の周波数	23
相互相関	生データ同士の相互相関の最大値 PSD 同士の相互相関の最大値	4
MFCCs	メル周波数ケプストラム係数	39

表 3　各ラベルの合計ラベルデータ数

ラベル名	咀嚼	食べ物の嚥下	飲み物の嚥下	発話	その他
合計	447	222	295	447	447

に示すように，咀嚼のデータ数は多いが，咀嚼以外のデータ数が咀嚼に比べて非常に少ない．よって不均衡なデータセットを均衡にするためにリサンプリングを行った．まず，咀嚼のデータをランダムに 500 個選択し，次に機械学習における分類モデル作成の際に訓練データのみに Support Vector Machine (SVM) による SMOTE を利用した．選択する個数を 500 個にした理由は次にデータ数の多かった発話のデータ数を参考にしたからである．リサンプリング後の訓練データ数の一例を表 3 に示す．

3.5　分類器毎の精度検証

本研究では，食事詳細行動の分類モデル作成のために機械学習を用いる．ここではすでにラベリングされたデータを用いるため，既知の入出力データを用いてモデルを訓練し出力を予測できる，教師あり学習の分類モデルを使用する．一般的な教師あり学習の分類器は，SVM や決定木，KNN，アンサンブル分類器などがあるが，最適な分類モデルを選定するために MATLAB の「分類学習器」アプリケーションを利用した．各モデルの交差検証結果を表 4 に示す．これにより，最も精度が高かった中程度のガウス SVM (rbf カーネル) を用いた．

4.　分類性能評価

自然な食事環境下で収集した食事音声データを用いた食事行動分類手法の分類性能評価にはテストデータセットを用いる．最適なパラメータは性能評価の前に調整した．

4.1　SVM のパラメータ調整

本研究で用いる学習モデルは，rbf カーネルを用いた SVM である．SVM を用いる際に，特徴量をスケール変換する必要がある．本研究では，平均が 0，分散が 1 になるよ

表 4　分類学習器によるモデルの交差検証結果

分類モデル		精度 [%]
決定木	複雑な木	68.6
	中程度の決定木	62.8
	粗い木	54.6
SVM	線形 SVM	73.1
	細かいガウス SVM	62.8
	中程度のガウス SVM	85.9
	粗いガウス SVM	68.2
最近傍分類器	細かい KNN	80.2
	中程度の KNN	67.3
	粗い KNN	54.0
	コサイン KNN	74.2
	3 次 KNN	65.1
	重み付き KNN	75.0
アンサンブル分類器	ブースティング決定木	68.5
	バギング決定木	82.4

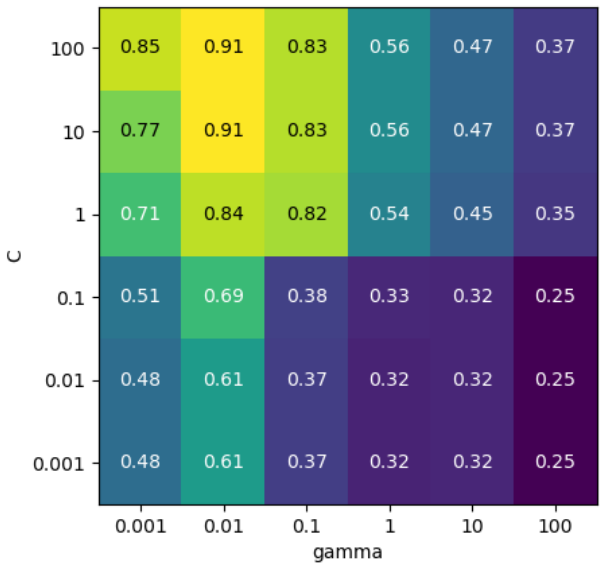

図 2　75 個の特徴量を用いたときの SVM の C と gamma の影響による交差検証精度

うスケール変換した．また，より高精度な分類を実現するためにパラメータ調整を行った．rbf カーネルのパラメータは，正則化パラメータ C とガウシアンカーネルの幅の逆数を表す gamma である．本研究ではパラメータの調整にグリッドサーチと呼ばれる手法を用い，合計 36 通りの C と gamma の組み合わせで交差検証を行った．各組合せの検証結果を図 2 に示す．グリッドサーチの出力結果では，75 個の特徴量を用いた SVM の最適なパラメータは C が 10，gamma が 0.01 となった．

4.2　汎化性能結果

テストデータを用いて，最適なパラメータに調整した SVM の汎化性能結果を出した．本研究では全 4 種類のデー

表 5　汎化性能評価結果（平均）

		F1 値			
特徴数		75	48	75	48
咀嚼		0.75	0.78	0.77	0.75
嚥下	食べ物	0.43	0.48	0.28	0.27
	飲み物			0.19	0.29
発話		0.90	0.89	0.91	0.90
その他		0.51	0.50	0.51	0.54

タセットに対する SVM の汎化性能結果を出した．まず，ラベリングした 5 種類の行動の分類と，食べ物の嚥下と飲み物の嚥下を 1 つの嚥下というラベルとして分類した場合の 4 種類の行動の分類である．また，ラベル数 4 つの場合とラベル数 5 つの場合それぞれに対して，75 個の特徴量と 48 個の特徴量を用いた SVM それぞれの汎化性能結果を出した．特徴量が 48 個に変更された理由は MFCC に関して 39 次ではなく，12 次を用いたからである．先行研究より 39 次の MFCC を用いていたが，音声認識では 12 次がよく用いられているため 12 次を用いた場合の結果も求めた．

今回は訓練データとテストデータの組み合わせを 10 個用意し，毎回訓練データのみにリサンプリングを行った．それぞれの条件の 10 回分の平均の結果を表 5 に示す．どの条件でも発話の F1 値は 89%以上，咀嚼の F1 値は 75%以上と良い結果となった．しかし，嚥下に関する F1 値は 50%以下と精度が低く，食べ物の嚥下と飲み物の嚥下に詳細化した場合はさらに悪い結果となった．嚥下が咀嚼と間違って予測されることが多かったことより，違いが明らかとなるような特徴量の追加・変更の必要がある．また，特徴量を MFCC に関して 39 次から 12 次に変更した場合でも結果に大きな差が出なかった．これより，MFCC に関して 12 次を用いても分類精度を維持できることが分かった．

5. まとめ

本研究では，自然な食事環境下で収集した食事音声データを用いた食事詳細行動の分類手法を提案した．特徴量を 75 個抽出し，rbf カーネルを用いた SVM により咀嚼・食べ物の嚥下・飲み物の嚥下・発話・その他の分類を行った．嚥下を分けた状態の場合，1 つにまとめた場合と特徴量を 75 個の場合と 48 個に減らした場合の結果より，発話・咀嚼の F1 値がどちらも 75%以上と良い結果となった．それに対し，嚥下の F1 値は 50%以下となり，食べ物の嚥下と飲み物の嚥下を区別する場合はさらに低い値となった．

嚥下が咀嚼と間違って予測されることが多かったことから，これらを差別化できるような特徴量の追加・変更が必要であると考えられる．特徴量を MFCC に関して 12 次を用いても精度が維持できることが分かった．

今後の展望としては，変更やラベリングの見直しなどにより高精度に分類できるような手法を検討し，リアルタイムでの食事行動の分類を目指す．

参考文献

[1] 厚生労働省．国民健康・栄養調査結果の概要　平成 29 年．https://www.mhlw.go.jp/content/10904750/000351576.pdf.

[2] 安藤雄一, 花田信弘, 柳澤繁孝．「ゆっくりとよく噛んで食べること」は肥満予防につながるか? ヘルスサイエンスヘルスケア, Vol. 8, No. 2, pp. 51–63, 2008.

[3] 岸田典子, 上村芳枝．学童の食事中における会話の有無と健康及び食生活との関連．栄養学雑誌, Vol. 51, No. 1, pp. 23–30, 1993.

[4] Keum San Chun, Sarnab Bhattacharya, and Edison Thomaz. Detecting eating episodes by tracking jawbone movements with a non-contact wearable sensor. *Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 2, No. 1, pp. 1–21, 2018.

[5] Hao Zhang, Guillaume Lopez, Ran Tao, Masaki Shuzo, Jean-Jacques Delaunay, and Ichiro Yamada. Food texture estimation from chewing sound analysis. In *in proc. of the 5th Int. Conf. on Health Informatics (HEALTHINF)*, pp. 213–218, 2012.

[6] Takumi Kondo, Haruka Kamachi, Shun Ishii, Anna Yokokubo, and Guillaume Lopez. Robust classification of eating sound collected in natural meal environment. In *in adj. proc. of ACM Int. Conf. on Pervasive and Ubiquitous Computing and ACM Int. Symp. on Wearable Computers*, pp. 105–108, 2019.

[7] Praat: doing phonetics by computer. http://www.fon.hum.uva.nl/praat/. (Accessed on 05/12/2020).

[8] 安藤純平, 斎藤隆仁, 川崎仁嗣, 片桐雅二, 池田大造, 峰野博史, 西村雅史ほか．咽喉音を利用した会話・摂食行動の認識．マルチメディア, 分散協調とモバイルシンポジウム 2017 論文集, Vol. 2017, pp. 116–123, 2017.