

骨伝導音を用いた自動セグメンテーションによる 食事行動検出手法の提案

近藤 匠海¹ 蒲地 遥¹ 横窪 安奈¹ ロペズ ギヨーム¹

概要：食事時の咀嚼回数が少ないと肥満に繋がることが明らかになった。さらに、リアルタイムで咀嚼回数を食事者に提示することで、早食いを防ぎ、食事に対する意識を改善することが期待できる。また、食事行動の識別は実験環境下でしか行われていない。本研究では、自然な食事環境下での食事行動定量化による食事行動の意識改善を促すシステムを提供することを目的として、骨伝導音を利用した食事詳細行動を自動で正確にセグメントする手法の開発を行った。骨伝導マイクロフォンによって収集した食事音声データの食事詳細行動に該当する音声を抽出し、さらに本手法によって抽出された食事詳細行動の音声を分類するモデルの提案と咀嚼回数推定も行った。セグメンテーション評価では適合率は88.1%、各クラスの再現率の平均は70.5%となった。

Eating Activity Detection Method by Automatic Segmentation Using Bone Conduction Sound

TAKUMI KONDO¹ HARUKA KAMACHI¹ ANNA YOKOKUBO¹ GUILLAUME LOPEZ¹

1. はじめに

咀嚼とは、口に取り込んだ食べ物を噛み砕くことである [1]。Zhu らは、ゆっくり食べることがメタボリックシンドローム発症の予防になると示唆している [2]。さらに、食べる速さが早いほど BMI が高い傾向にあることから [3]、早食いと肥満は強い関係にあることが明らかとなった。また、咀嚼回数が少ないと肥満に繋がることが [4] や、嚥下前の咀嚼回数が多いと主観的な食欲が減ること [5] も明らかとなっている。食べる速さに加え、食事時の会話も健康と関連があることが明らかとなっている [6], [7]。

近年、市販されているウェアラブルデバイスはカロリー出力に関連した人間の活動レベルをモニタリングすることが可能である。さらに実験環境において食事時にリアルタイムでの咀嚼回数をフィードバックすることにより、咀嚼回数が増加するという結果が示された [8], [9]。しかし、自由な環境下での食事行動を自動的に検出するデバイスはいまだに存在しない [10]。

食事の際に1回の咀嚼、つまり口の中にある食物を1回噛む動作を認識することが可能になれば、リアルタイムで食事者に咀嚼回数の少なさや早食いである事を判定しフィードバックすることが可能である。さらに、食事時の会話も検出することが可能になれば、フィードバックによって食事時の会話の増加が期待できる。咀嚼や嚥下・発話のような食事の詳細行動を認識することが可能になれば、食事者の肥満の防止や健康促進につなげることができると考える。

以上から本研究では、自然な食事環境下での食事行動の定量化実現によるフィードバックを用いた食事者の食事意識改善を目的とし、目的達成のために以下を目標とする。まず自然な食事環境に対応した咀嚼・嚥下・発話・その他(ノイズなど)といった食事詳細行動に該当する骨伝導音の自動セグメンテーションの実現、さらに食事詳細行動に該当する骨伝導音を80%以上で分類するモデル提案と誤差率20%以下で咀嚼回数を推定することである。

2. 関連研究

音声をを用いない食事行動解析・認識手法として特製デバ

¹ 青山学院大学
Aoyama Gakuin University

イスを使用したものがある。Zhang らは自動的に摂食活動を監視できるようにするためのマルチセンサネックレス「NeckSense」を開発した [11]。近接センサでデバイスと顎の距離から咀嚼行動、環境光センサで食物を口に運ぶ摂食行動、慣性計測ユニット (IMU) センサで食事をする際の前傾姿勢の動きを捉えており、これらの特徴から咀嚼と食事エピソードの検出を行っている。他にメガネ型デバイスや近接センサを用いたネックレスデバイスも提案されているが、市販されていて誰でも手に入れられるデバイスではないため導入が難しい。また食事の検出は行われているが、食事の詳細な行動の分析はできていない。

Amft らの咀嚼音解析 [12] や Papapanagiotou らの耳装着型デバイスによる食事イベント検出 [13] により、耳の内側にマイクロフォンを設置することで質の高い咀嚼音が取得できることが示唆された。Zhang らは、骨伝導マイクロフォンによる音情報を用いて食べ物のテキスト分析を行った [14]。高精度でのテキスト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。三井らは、骨伝導マイクロフォンを利用して咀嚼回数と発話状態をリアルタイムで認識しユーザにフィードバックするシステムを提案した [9]。咀嚼判定と発話時間の測定は高精度で達成しているが、食材が限定的であるなど自然な食事環境下での実験ではなかった。我々は、骨伝導マイクロフォンを用いた自由な食事環境下での食事詳細行動分類手法の提案を行った [15], [16], [17]。被験者に日常生活と同じ環境で自由に食事を行うように指示し食事データを収集、1 回の咀嚼・嚥下・発話・その他 (ノイズなど) を含めた 4 つのクラス分類 [15], [16] と、さらに嚥下を食物の飲み込みと飲み物の飲み込みで分けた 5 クラスの分類を行った [17]。咀嚼と発話の F1 値は 70% 以上だったが、それ以外のクラスは低い分類精度となった。先行研究では骨伝導マイクロフォンを用いた自然な食事環境下での食事詳細行動の自動セグメンテーションおよび、高精度認識は行われていない。

以上から本研究では、食品を限定しない自然な食事環境下での咀嚼・嚥下・発話・その他 (ノイズなど) の自動セグメンテーションと分類を行いたいため、耳の内側に配置でき、容易に手に入る市販の骨伝導マイクロフォンを利用する。

3. データの収集

本研究では骨伝導音を用いた自動セグメンテーションによる食事行動検出を行うために、自然な食事環境下での食事音声データの収集を行った。

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 Pixel3, Bluetooth 通信を



図 1 データ収集の様子

行う骨伝導マイクロフォンは Motorola 社製の Motorola Finiti HZ800 Bluetooth Headset を用いた。

データ収集では、11 歳から 32 歳の男性・女性 14 名を被験者とした。図 1 に示すように、被験者の片耳に Bluetooth 通信を行う骨伝導マイクロフォンを装着して行った。また、音声データ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。データ収集は、食堂や一般家庭の食卓、レストラン、研究室など自由な食事環境で行い、被験者には日常生活と同じように食事するよう求めた。

また、嚥下クラスのデータが他のクラスと比べて非常に少なかったことから、嚥下データのみのデータ収集も行った。被験者に飲み物を数杯飲むように指示し、上記と同じように骨伝導マイクロフォンとスマートフォンを用いて、22 歳から 42 歳の男性・女性 8 名を被験者としてデータ収集を行った。

3.2 収集データの整理

収集した食事音声データと嚥下データの整理を行った。このデータの整理は、自動セグメンテーション手法の評価と食事詳細行動の分類を行う際に用いる真値のデータセットを作成することを目的とした。真値のデータセットを作成するために、骨伝導マイクロフォンを用いて収集した音声データの咀嚼・嚥下・発話・その他 (ノイズなど) に該当する区間の音声をラベリングした。また、ラベリングを行う際に、食事行動をしているがどの食事行動に該当するか判断できない音声区間を「食事」ラベルとして定義した。このラベルを作成することによって、セグメンテーション評価を行う際に、食事行動を行っている音声区間をセグメントしたのに誤ってセグメントを行ったと評価することを避ける。

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時に同時に撮影した動画を同期させてラベリングを行った。本研究では堅牢なデータセット作成のために、2人または3人でラベリングを行った。複数人でラベリングを行い全員が一致した箇所のラベルを採用し、一致しなかった箇所は「食事」ラベルとして採用した。ラベリングの際にはPraat[18]とアノテーションツールELAN[19]を補助ツールとしてラベリングに使用した。ラベルの種類は「咀嚼・嚥下・発話・その他・食事」の5種類に設定し、ラベルを付与された音声区間のみを抽出しデータセットを作成した。4名の被験者は骨伝導マイクロフォンの装着不備により音声データの質が悪かった。装着不備に陥った原因として、データ収集時に被験者が骨伝導マイクロフォンの完全な装着状態を理解していなかったことを挙げる。本研究では、装着不備により質の悪い被験者のデータは除外したデータセットを使用した。実際に使用したデータの合計を表1に示す。

表1 実際に使用したデータセットの合計

	Chew	Swallow	Talk	Other	Eating
Total	2274	332	448	536	219

4. 自動セグメンテーション手法

食事詳細行動に該当する骨伝導マイクロフォンで得た音声波形を抽出するため、自動セグメンテーションアルゴリズムの設計と評価を行う。本研究の目的で述べた1回の咀嚼を認識するために、咀嚼1回分の音声区間を抽出可能なセグメンテーションアルゴリズムを設計する。目標は、咀嚼1回分、嚥下、発話、その他に該当する音声区間を80%以上の精度でセグメンテーションするアルゴリズムを設計することである。

4.1 咀嚼のモデル化

咀嚼1回分の音声区間を自動抽出するセグメンテーションアルゴリズムを設計するにあたって、咀嚼のモデル化を行った。Schiboniらは咀嚼1回を四段階のシーケンスとしてモデル化し[20]、Zhangらはこの四段階のシーケンスをそれぞれP1, P2, P3, P4と定義した[21]。各段階の定義を以下に示し、図2に咀嚼プロセスの特徴を示す。咀嚼モデルより、Zhangらは咀嚼の特徴からP1の音声信号の振幅はP3より大きく、P4の期間はP2よりも長い事、1秒間に検出される咀嚼回数は3回以下であると仮定することによって、ルールを作成し咀嚼回数推定を行った。

- P1: 下顎を閉じ、食物を押しつぶす。
- P2: 下顎を動かさずに小休止し、噛み合わせている。
- P3: 下顎を開き、上歯と下歯がかみ合っていない。

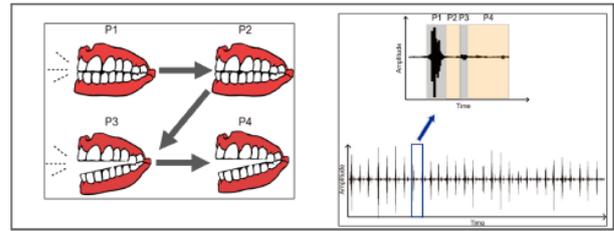


図2 咀嚼プロセスの特徴 ([20], [21] より)

- P4: 休止し、下顎は開いて動かない。

また、我々は咀嚼の開始時間と直後の連続した咀嚼の終了時間差の分布を出した。収集した音声データをラベリングした咀嚼データを使用した(3.2章)。この結果より、咀嚼の開始時間と直後の咀嚼の終了時間差は0.4秒以上であることが示された。さらに、咀嚼と嚥下の生データを比較した。両方の生データは一度音声信号が大きく振動したあとに無振動の間を経て、もう一度振動している。しかし、振幅の大きさに注目すると、咀嚼は後の振動の振幅が最初の振動と比べて小さく、嚥下は2つの振動の振幅の大きさが変わらないことが分かる。以上の咀嚼モデル、咀嚼と嚥下の振幅の大きさを考慮してセグメンテーションアルゴリズムを設計した。

4.2 音声信号の特徴化

食事詳細行動に該当する音声の抽出のため、Short Term Energy (STE) を用いて音声信号を特徴化させた。処理したい信号を s 、サンプル数のシフトを n 、窓関数を $w(n)$ として式1に示す。また、窓関数は矩形窓を採用した。本研究ではSTEの窓サイズは0.04s、シフトサイズは0.01sとした。高解像度でセグメンテーションを行いたかったため、シフトサイズを0.01sに設定した。また、先行研究である三井らの研究がオーバーラップ75%で窓をシフトしていたため[9]、本研究もオーバーラップ75%を採用した。そのため、窓サイズはオーバーラップ75%を満たす0.04sとした。STEによって特徴化された咀嚼1回分の信号(図3)より、咀嚼フェーズのP1とP3はSTEで特徴化できることが分かった。よって、閾値を設定することでP1とP3に該当する音声区間が抽出可能であることが分かる。

$$STE(n) = \sum_{m=-\infty}^{\infty} (s(m) * w(n-m))^2 \quad (1)$$

4.3 セグメンテーションアルゴリズム

食事詳細行動に該当する音声区間を抽出するためにセグメンテーションアルゴリズムを設計した。咀嚼1回分の特徴を考慮して以下のようなセグメンテーションルールを作成した。

- (1) STEの値が設定した閾値を超えた音声区間を抽出する。

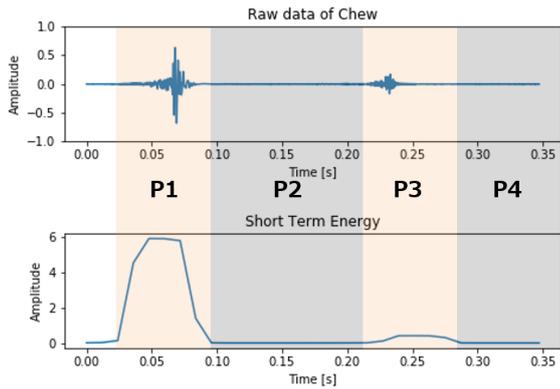


図 3 咀嚼 1 回分の音声生データと STE によるデータ

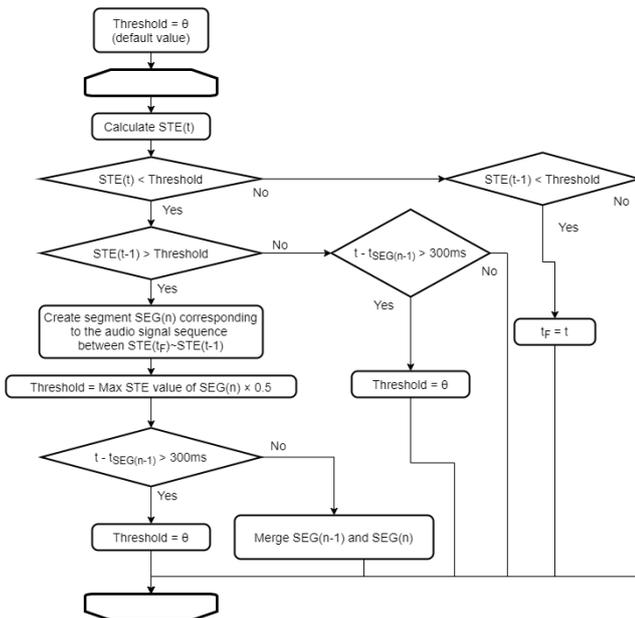


図 4 提案セグメンテーションアルゴリズムのフローチャート

- (2) セグメントされた音声区間の終了時間が直前のセグメント開始時間から 400ms 以内であればマージする。
- (3) セグメントされた音声区間の開始時間から 400ms 以内は、閾値を直前セグメントの STE の最大値の半分を設定する。

上記のルールに則ったセグメンテーションアルゴリズムのフローチャートを図 4 に示す。SEG は抽出されたセグメントを指し、SEG(n) は n 番目に抽出されたセグメントを指す。

また、提案したセグメンテーション手法と比較を行うために、2つのセグメンテーション手法を設計し利用した。1つ目のセグメンテーション手法は、三井らが用いた咀嚼回数推定アルゴリズムを元にしたセグメンテーション手法である [9]。三井らは音声信号を STE によって特徴化し、設定した固定の閾値を超え次に閾値を下回ったときに咀嚼 1 回と推定していたため、本研究では設定した閾値を超えてから下回るまでの音声区間をセグメントした。STE の窓

サイズは 0.08s で、シフトサイズは 0.02s であった。2つ目のセグメンテーション手法は、Zhang らが提案した [21] 手法に基づいたものであり、提案したセグメンテーション手法の動的な閾値を固定された閾値に変更したものを採用した。これは、咀嚼 1 回を複数回としてセグメントすることを避けるためにセグメントをマージするアルゴリズムとなっている。

4.4 セグメンテーション評価指標の定義

本研究で提案するセグメンテーション手法を評価するために、真陽性、偽陽性、真陰性、偽陰性を定義した。定義を以下に示す。

- 真陽性 (TP)
 - 咀嚼 1 回分のラベルまたは他のラベルをセグメントとして抽出する。
 - * 咀嚼以外のラベルは同じラベルであれば 2 個以上セグメント内に含まれていても TP とみなす。
 - * その他ラベルはノイズのため、その他以外のラベルと混ざっていても TP とみなす。
- 偽陽性 (FP)
 - 既にセグメントとして抽出されている咀嚼ラベルを新しくセグメントとして抽出する。
 - セグメント内に 2 種類以上のラベルを含む、または咀嚼ラベル 2 個以上を含む。
 - セグメント内にラベルが存在しない。
- 真陰性 (TN)
 - セグメントとして抽出されなかった区間にラベルが存在しない。
- 偽陰性 (FN)
 - セグメントとして抽出されなかった区間にセグメントされなかったラベルが存在する。

定義された TP, FP, TN, FN を用いて、セグメンテーション評価指標の定義を式 2, 式 3, 式 4 に示す。適合率 (*Precision*) は嚥下、発話、その他に該当する音声区間または咀嚼 1 回分の音声区間としてセグメンテーションによって抽出されたセグメントが正しいセグメントである割合を評価する。また、各クラスごとの再現率 ($Recall_{class}$) は、ラベリングされた音声区間が正確にセグメントされた割合を評価する。さらに、再現率 (*Recall*) はマクロ平均により、全クラスに対する再現率を求める。

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall_{class} = \frac{TP_{class}}{N_{class}} \quad (3)$$

$$Recall = \frac{Recall_{chew} + Recall_{swallow} + Recall_{talk} + Recall_{other} + Recall_{eating}}{5} \quad (4)$$

4.5 セグメンテーション評価結果

設計したセグメンテーションアルゴリズムの評価を行った。評価を行う際にオーバーラップの割合を設定し、セグメントされた音声区間にラベリングされた音声区間の重なっている割合がオーバーラップで設定された割合を超えていればセグメント内にラベルが存在するとみなした。オーバーラップの割合は10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%の9通りで評価を行った。4.4章で定義した指標に従って、各オーバーラップごとの適合率、クラスごとの再現率、全クラスの再現率を結果として出した。また、提案したセグメンテーション手法以外の手法と比較するために4.3章で述べた手法も評価した。各手法のオーバーラップごとのセグメンテーション評価結果を表2, 表3, 表4に示す。

表2 提案セグメンテーション手法の評価結果

Overlap [%]	Precision [%]	Recall[%]					Average
		Chew	Swallow	Talk	Other	Eating	
10	88.1	75.9	67.3	87.2	63.4	58.4	70.5
20	87.3	78.0	56.0	87.6	63.6	59.8	69.0
30	79.5	68.3	43.5	88.8	63.6	62.1	65.3
40	67.2	50.0	35.1	89.8	63.6	63.5	60.4
50	58.8	41.1	30.6	89.5	63.6	63.5	57.7
60	54.8	31.0	24.6	89.3	63.8	64.4	54.6
70	51.5	25.9	21.8	90.0	63.8	64.8	53.3
80	47.7	20.0	16.9	90.5	63.8	64.8	51.2
90	42.0	11.3	10.9	91.0	64.0	65.3	48.5

表3 三井らの手法に基づくセグメンテーション手法の評価結果

Overlap [%]	Precision [%]	Recall[%]					Average
		Chew	Swallow	Talk	Other	Eating	
10	80.7	60.6	75.0	86.2	70.2	68.0	72.0
20	80.3	61.7	74.6	86.2	70.2	68.0	72.2
30	78.8	68.2	63.3	86.7	70.4	68.9	71.5
40	73.6	75.1	48.8	87.9	70.6	68.9	70.3
50	61.9	60.5	37.9	88.8	66.7	69.9	64.8
60	54.2	47.5	28.6	88.8	70.8	69.9	61.1
70	50.8	41.7	24.6	88.8	70.8	70.3	59.3
80	49.0	38.7	21.4	89.1	71.0	70.3	58.1
90	46.9	35.0	19.4	89.1	71.0	70.3	57.0

表4 Zhangらの手法に基づくセグメンテーション手法の評価結果

Overlap [%]	Precision [%]	Recall[%]					Average
		Chew	Swallow	Talk	Other	Eating	
10	86.6	74.0	67.7	84.1	65.2	60.3	70.3
20	87.1	77.1	58.9	84.8	66.0	61.2	69.6
30	85.1	75.3	49.2	86.2	67.1	63.5	68.3
40	80.6	69.3	43.1	88.1	67.3	64.8	66.6
50	75.8	62.5	38.7	88.1	64.2	64.8	63.7
60	71.9	57.1	33.5	88.4	67.5	65.3	62.3
70	68.1	51.4	30.6	88.4	67.5	65.8	60.7
80	62.0	42.3	25.0	89.5	67.5	66.2	58.1
90	51.8	27.3	17.3	90.7	67.5	66.2	53.8

4.6 提案セグメンテーション手法の考察

セグメンテーション評価結果より、各セグメンテーション手法はオーバーラップの割合が低いほど適合率、再現率が高かった。また、オーバーラップの割合が高くなるにつれて、評価指標の結果が低くなる傾向が各手法の評価指標よりわかる。

各手法の適合率を比較して、提案したセグメンテーション手法のオーバーラップ10%が88.1%と一番高かった。しかし、オーバーラップ90%で42.0%と一番悪い結果を出しているのも提案したセグメンテーション手法であった。一方で、提案したセグメンテーション手法に近いアルゴリズムを用いたZhangらのセグメンテーション手法は、オーバーラップ10%で86.6%で2番目に高い結果を出しており、90%のときに51.8%と他の手法のオーバーラップ90%の適合率と比較して良い結果を出している。また、オーバーラップ10%から80%まで60%以上の適合率を出しているのはZhangらのセグメンテーション手法である。よって、3つのセグメンテーション手法の中で、複数の食事行動、咀嚼複数回、食事詳細行動に該当しない音声抽出しない手法はZhangらのセグメンテーション手法であるといえる。

次に、各手法の再現率の結果を比較する。各手法の咀嚼クラスの再現率の結果より、一番高い結果は提案したセグメンテーション手法のオーバーラップ20%における78.0%だった。しかし、オーバーラップ90%で11.3%となった。これは3つの手法の中で一番低い結果だった。オーバーラップの割合が低いときに再現率が他の手法と比べて高く、割合が高いときは非常に低い再現率となった原因として、動的な閾値により咀嚼のP1部分のみをセグメントし、P3部分はセグメントされなかったためだと考える。なので、咀嚼クラスのセグメンテーションに関しては、再現率の評価だけではなく実際に抽出されたセグメントが咀嚼として分類されるのかを調べる必要がある。また、嚥下に関しては三井らのセグメンテーション手法が一番良い結果を出した。他の手法が三井らの結果と比較して低い結果だったのは、セグメントをマージするアルゴリズムであるため、嚥下の他に別の食事詳細行動に該当とする音声区間をマージしてしまったことが原因だと考えられる。発話に関しては、どの手法も80%を超える再現率を出していた。比較している全手法はエネルギーベースのセグメンテーションを採用しており、発話の音声信号は振幅が大きく、長く振動することによりセグメンテーションしやすいからだと考える。

5. 食事詳細行動の分類手法提案

提案する自動セグメンテーション手法によって抽出された食事詳細行動に該当する音声区間を高精度に分類するモデルを作成し、分類性能と作成モデルを利用した咀嚼回数推定の精度を評価する。

5.1 食事詳細行動の分類手法

分類に機械学習を用いるため、ラベリングされたデータセットから特徴量を抽出する。分類精度を高くするため合計 31 個の特徴量を抽出した (表 5)。

表 5 抽出された 31 個の特徴

Feature category	Description	Number of features
Raw data	Sum	1
	Average	1
	Variance	1
	Kurtosis	1
	Skewness	1
	Zero Crossing count	1
Amplitude difference accumulation	Sum	1
	Average	1
Zero crossing rate	Sum	1
	Average	1
Frequency domain features	Peak frequency	1
	Band total amplitude	4
	Band peak frequency	4
	Mel-Frequency cepstral Coefficients (MFCC)	12

ラベリング作業により作成されたデータセットは表 1 で示すように、咀嚼のデータ数は多いが、咀嚼以外のデータは咀嚼と比べ非常に少ない。そこで、不均衡なデータセットを均衡にするため、データセットの分割後に訓練データのみ Support Vector Machine (SVM) に基づく Synthetic Minority Oversampling TEchnique (SMOTE) を用いてリサンプリングを行った。分類モデルにはランダムフォレストを選択し、パラメータ調整を行わずに学習を行った。用いた学習データは各クラスのデータ数が均衡化されたあとの訓練データである。

5.2 分類モデル汎化性能結果

作成した分類モデルが未知データに対して正確に予測する能力を評価するため、シャッフル分割交差検証を採用し汎化性能結果を出した。訓練データとテストデータを 4:1 に分割し、10 回分割と検証を繰り返し平均結果を求めた。結果を表 6 に示す。

表 6 シャッフル分割交差検証 10 回による平均結果

Class	Precision[%]	Recall[%]	F1[%]
Chewing	90.3	89.7	89.7
Swallowing	45.6	37.8	41.1
Speaking	92.3	90.0	91.1
Other (noise etc.)	65.5	77.7	71.0

咀嚼クラスと発話クラスの F1 値は約 90% と高い結果を得ているが、その他クラスと嚥下クラス、特に嚥下クラスは低い結果となった。

5.3 咀嚼回数推定

作成したデータセットに提案したセグメンテーション手法と分類手法を用いて、自然な食事環境下での骨伝導音データによる咀嚼回数推定評価を行った。先行研究より評

価指標は誤差率を使用した。定義を式 5 に示す。

$$Error = \left| \frac{A_{CNT} - E_{CNT}}{A_{CNT}} \right| * 100 \quad (5)$$

ここで、 A_{CNT} は真の咀嚼ラベルの個数、 E_{CNT} は推定された咀嚼クラスの個数とする。

咀嚼回数推定の評価は被験者ごとに行うため、分類モデルの学習には評価を行う被験者以外の食事データと嚥下収集実験で得たデータを学習データとして扱う。そして、評価を行う被験者の食事データをセグメンテーション手法によりセグメントし、セグメントされた音声データの特徴量を抽出し分類を行う。評価を行う被験者の食事データ内に存在する咀嚼ラベルの個数を A_{CNT} とし、セグメントされた音声データから咀嚼と分類された個数を E_{CNT} とする。結果を表 7 に示す。

表 7 提案セグメンテーション手法による咀嚼回数推定結果

Number of chews	Number of chewing detections[%]	Error
773	690	10.7
205	124	39.5
485	458	5.6
124	134	8.1
96	99	3.1
101	114	12.9
144	155	7.6
158	104	34.2
52	65	14.8
127	103	18.9

全被験者の平均は 15.5% の誤差率となっており、80% 以上の精度で推定を行う目標を達成した。

6. 結論と今後の展望

本研究では、骨伝導音を用いた食事行動検出のための自動セグメンテーション手法を提案した。咀嚼モデルを参考にしたセグメンテーション手法を提案し、先行研究で用いられた咀嚼回数推定手法をベースにしたセグメンテーション手法と比較した。提案手法はセグメンテーションの適合率が 88.1% と一番高い結果となった。各クラスの再現率の平均は 70.5% であった。

また、セグメンテーションによって抽出された音声を食事詳細行動に分類するための分類モデルの作成と汎化性能評価を行った。咀嚼・発話では F1 値約 90% と良い結果を得たが、嚥下は 50% 以下だった。さらに、セグメンテーション手法と分類モデルを使用し、咀嚼回数推定を行った。これらを組み合わせることで誤差率 15.5% と目標を達成した。

今後は、目標である咀嚼クラス再現率 80% を超えるような、より正確に咀嚼一回分を抽出する新たなセグメンテーション手法を検討したい。さらに提案したセグメンテーション手法と分類モデルを用いることで、リアルタイムでの咀嚼回数推定を行い、食事者にどのようなフィードバックを与えれば早食いを防げるかを調査したい。

謝辞

この研究は、ロツテ財団の支援により研究が行われました。この場を借りて深く御礼申し上げます。

参考文献

- [1] 日本咀嚼学会. 日本咀嚼学会からの発信. <http://sosyaku.umin.jp/info/file/info01.pdf> (accessed:2021/1/29).
- [2] Bing Zhu, Yasuo Haruyama, Takashi Muto, and Takako Yamazaki. Association between eating speed and metabolic syndrome in a three-year population-based cohort study. *Journal of Epidemiology*, Vol. 25, No. 4, pp. 332–336, 2015.
- [3] 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.
- [4] 岩崎正則, 葭原明弘, 宮崎秀夫. 成人期および高齢期における咀嚼回数と体格の関連. 口腔衛生学会雑誌, Vol. 61, No. 5, pp. 563–572, 2011.
- [5] Yong Zhu, Walter H. Hsu, and James H. Hollis. Increasing the number of masticatory cycles is associated with reduced appetite and altered postprandial plasma concentrations of gut hormones, insulin and glucose. *British Journal of Nutrition*, Vol. 110, No. 2, p. 384–390, 2013.
- [6] 森脇弘子, 岸田典子, 上村芳枝, 竹田範子, 佐久間章子, 寺岡千恵子, 梯正之. 女子学生の健康状況・生活習慣・食生活と小学生時の食事の楽しい会話との関連. 日本家政学会誌, Vol. 58, No. 6, pp. 327–336, 2007.
- [7] 岸田典子, 上村芳枝. 学童の食事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.
- [8] 三井秀人, 中野和哉, 磯山直也ほか. 咀嚼回数向上支援システム. 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. 1733–1737, 2017.
- [9] 三井秀人, 小原丈, 横窪安奈ほか. 骨伝導音を用いたリアルタイム咀嚼・発話判定精度向上手法の提案. マルチメディア, 分散協調とモバイルシンポジウム 2018 論文集, Vol. 2018, pp. 562–566, 2018.
- [10] N. A. Selamat and S. H. M. Ali. Automatic food intake monitoring based on chewing activity: A survey. *IEEE Access*, Vol. 8, pp. 48846–48869, 2020.
- [11] Shibo Zhang, Yuqi Zhao, Dzung Tri Nguyen, Runsheng Xu, Sougata Sen, Josiah Hester, and Nabil Alshurafa. Necksense: A multi-sensor necklace for detecting eating activities in free-living conditions. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 4, No. 2, June 2020.
- [12] Oliver Amft, Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. Analysis of chewing sounds for dietary monitoring. In *International Conference on Ubiquitous Computing*, pp. 56–72. Springer, 2005.
- [13] Vasileios Papapanagiotou, Christos Diou, Lingchuan Zhou, Janet Boer, Monica Mars, and Anastasios Delopoulos. A novel chewing detection system based on ppg, audio, and accelerometry. *IEEE Journal of Biomedical and Health Informatics*, Vol. PP, pp. 1–1, 11 2016.
- [14] Hao Zhang., Guillaume Lopez., Ran Tao., Masaki Shuzo., Jean-Jacques Delaunay., and Ichiro Yamada. Food texture estimation from chewing sound analysis. In *the International Conference on Health Informatics (HEALTHING 2012)*, pp. 213–218, 2012.
- [15] T. Kondo, H. Shiro, A. Yokokubo, and G. Lopez. Optimized classification model for efficient recognition of meal-related activities in daily life meal environment. In *2019 Joint 8th International Conference on Informatics, Electronics Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision Pattern Recognition (icIVPR)*, pp. 146–151, 2019.
- [16] Takumi Kondo, Haruka Kamachi, Shun Ishii, Anna Yokokubo, and Guillaume Lopez. Robust classification of eating sound collected in natural meal environment. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, UbiComp/ISWC '19 Adjunct*, p. 105–108. Association for Computing Machinery, 2019.
- [17] Haruka Kamachi, Takumi Kondo, Anna Yokokubo, and Guillaume Lopez. Classification method of eating behavior by dietary sound collected in natural meal environment. In *Activity and Behavior Computing*, pp. 135–152. Springer, 2020.
- [18] Praat. <http://www.fon.hum.uva.nl/praat/> (accessed: 2021/1/29).
- [19] Elan. <https://archive.mpi.nl/tla/elan> (accessed: 2021/1/29).
- [20] Giovanni Schiboni and Oliver Amft. Automatic dietary monitoring using wearable accessories. In *Seamless healthcare monitoring*, pp. 369–412. Springer, 2018.
- [21] Hao Zhang, Guillaume Lopez, Masaki Shuzo, Jean-Jacques Delaunay, and Ichiro Yamada. Mastication counting method robust to food type and individual. In *HEALTHING*, 2012.