青山学院大学　理工学部　情報テクノロジー学科

2018（平成 30）年度卒業論文要旨

自然な食事環境に対応した骨伝導音を用いた咀嚼・嚥下・発話の分類手法の提案

近藤　匠海（15815041）
ロペズ研究室

1. はじめに

咀嚼とは，口に取り込んだ食べ物をかみ砕くことである．食事中の咀嚼回数が少ないと肥満につながることが明らかになった[1]．これに加え，食事中の会話と健康が関連していることより[2]，食事中に会話をすることが望ましい．また自由な食事環境での食事の詳細行動を認識することが可能になれば，食事行動と健康に関する研究に貢献すると考える．

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，自然な食事環境とは，日常生活における食事環境と定義する．また，自然な食事環境に対応した咀嚼・嚥下・発話の高精度分類を目標とする．本研究ではリアルタイムでの分類は行わないため，咀嚼，嚥下，発話に該当する音声区間を手動で抽出し，分類に用いる．

2. 関連研究

Amft らは，耳の内側にマイクロフォンを配置することによって質の高い咀嚼音を取得できると示した[4]．Bi らは，人がいつ，どのくらいの時間食事をしているか自動的に認識できるウェアラブルイヤピース「Auracle」を提案した[5]．しかし，咀嚼，嚥下，発話のような食事詳細行動は認識しなかった．

三井らは，食行動を改善するために骨伝導マイクロフォンを用い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[3]．しかし，どの食材にも対応できないことが問題点として挙げられる．

3. 自然な食事環境下での食事行動の分類手法

本研究では，自然な食事環境下での咀嚼・嚥下・発話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の収集を行った．収集した音声データを手動で「咀嚼・嚥下・発話」にラベリングした．ラベリングされたデータの詳細を表 1 に示す．

表 1．咀嚼・嚥下・発話のラベルデータ数

被験者（記録回数）	咀嚼	嚥下	発話
1 (5)	487	55	149
2 (1)	164	10	30
3 (1)	496	35	36
4 (1)	61	13	53
5 (1)	102	2	8
6 (1)	202	8	50
合計	1512	123	326

次に，特徴量抽出を行い，MFCC (Mel Frequency Cepstral Coefficient)，STE (Short Term Enenrgy)，パワースペクトルなどを含めた合計 75 個の特徴量を抽出した．表 1 が示すように，咀嚼以外のデータ数は咀嚼と比べて非常に少ない．よって，表 2 に示すように SMOTE (Synthetic Minority Oversampling TEchnique) により均衡なデータセットを構築した．

表 2．SMOTE を用い均衡したデータセット

ラベル名	ラベルデータの合計	訓練データ	テストデータ
咀嚼	1512	1209	303
嚥下	1512	1209	303
発話	1512	1210	302
合計	4536	3628	908

次に，咀嚼・嚥下・発話の分類を行うために，最適なモデルを選定した．Matlab の「分類学習器」による決定木，SVM (Support Vector Machine)，KNN (K-Nearest Neighbor algorithm)，アンサンブル分類器を元にした 15 モデルの交差検証結果を表 3 に示す．

表 3 が示すように，一番精度の高い中程度のガウス SVM（rbf カーネル）を用いることにした．

表 3. 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	83.9
	中程度の決定木	74.8
	粗い木	67.8
SVM	線形 SVM	84.2
	細かいガウス SVM	86.8
	中程度のガウス SVM	97.6
	粗いガウス SVM	82.4
KNN	細かい KNN	90.3
	中程度の KNN	82.2
	粗い KNN	64.5
	コサイン KNN	88.5
	3 次 KNN	81.0
	重み付き KNN	84.7
アンサンブル分類器	ブースティング決定木	81.7
	バギング決定木	95.5

4. モデルの最適化と分類性能評価

選択された分類モデルのパラメータを高精度な分類が実現するように調整した．また，特徴選択を行うことにより分類に関係の薄い特徴を削減し汎化性能を評価した．まず，特徴選択される前の特徴量を用いた SVM の正則化パラメータ C とガウシアンカーネルの幅の逆数を示す gamma を調整し，C=10，gamma=0.01 のときが最適であると結果が出た．このパラメータを用いて特徴選択を行った．

Sequential Forward Floating Selection を用いて特徴選択を行った．特徴量を 30 個まで限定し，特徴数 6 個のときに交差検証スコアが 80%を超え，16 個のときに 90%を超えた．先行研究で，咀嚼の検出精度が 80%を超えたリアルタイムでの咀嚼回数提示はユーザに違和感を与えなかったことより特徴数 6 個のときの汎化性能を評価した[3]．また，精度が 90%を超えると自動食事モニタリングの研究に貢献すると考え，特徴数 16 個のときの汎化性能も評価した．各個数の最適なパラメータを表 4 に示し，汎化性能評価に用いた．また，特徴数ごとの汎化性能評価結果を表 5 に示す．

表 4. 特徴数と交差検証スコア最大時パラメータ

特徴数	C	Gamma
6	100	1
16	10	0.1
30	100	0.1

表 5. 特徴数ごとの汎化性能評価結果

	F1 値			
特徴数	6	16	30	75
咀嚼	0.84	0.92	0.98	0.98
嚥下	0.88	0.95	0.99	0.98
発話	0.94	0.97	1.00	0.99

F1 値は，表 5 で示した結果になったが，特徴数 6 個のときは咀嚼の再現率が 80%を超えず 78%だった．また，16 個の特徴を用いたときは，咀嚼の再現率のみ 90%を超えず 88%だった．

5. まとめ

特徴選択された 16 個の特徴と 30 個の特徴を用いた場合， F1 値が 90%以上と非常に良い結果であった．また，6 個の場合も咀嚼と嚥下の F1 値が 84%，88%と 90%には届かなかったがリアルタイムでのフィードバックに利用するには問題ない結果であると考える．

本研究では自然な食事環境に対応した骨伝導を用いる咀嚼・嚥下・発話の分類を行うために，骨伝導マイクロフォンを用いて収集した音声データを咀嚼・嚥下・発話にラベリングし，rbf カーネルを用いた SVM により咀嚼・嚥下・発話の音声データを分類する手法を提案した．提案した手法では，90%を超える F1 値をもたらし，将来リアルタイムでの分類を行うために，特徴数が削減されても 80%を超える F1 値を得た．

今後の展望としては，本研究で提案した手法を用い，リアルタイムでの咀嚼・嚥下・発話の分類を目指す．これを実現するために，リアルタイムで咀嚼，嚥下，発話である音声区間を自動的に抽出するシステムを設計していきたい．

参考文献

[1] 岩崎正則，ほか：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.

[2] 岸田典子，ほか：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.

[3] 三井秀人，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. 562-566，2017.

[4] Oliver Amft, et al: Analysis of chewing sounds for dietary monitoring. In International Conference on Ubiquitous Computing, pp. 56-72. Springer, 2005.

[5] Shengjie Bi, et al: Auracle: Detecting eating episodes with an ear-mounted sensor. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 2, No. 3, p .92, 2018.