

2020

Dietary Monitoring Through Sensing Mastication Dynamics

Shuangquan Wang

William & Mary - Arts & Sciences, bering.sq.wang@gmail.com

Follow this and additional works at: <https://scholarworks.wm.edu/etd>



Part of the [Computer Sciences Commons](#)

Recommended Citation

Wang, Shuangquan, "Dietary Monitoring Through Sensing Mastication Dynamics" (2020). *Dissertations, Theses, and Masters Projects*. Paper 1593091934.

<http://dx.doi.org/10.21220/s2-zts3-1p33>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.

Dietary Monitoring through Sensing Mastication Dynamics

Shuangquan Wang

Williamsburg, VA, USA

Bachelor of Engineering, Wuhan Institute of Technology, China, 2000

Master of Engineering, Wuhan University of Technology, China, 2004

Doctor of Philosophy, Shanghai Jiao Tong University, China, 2008

A Dissertation presented to the Graduate Faculty
of The College of William & Mary in Candidacy for the Degree of
Doctor of Philosophy

Department of Computer Science

College of William & Mary
May 2020

APPROVAL PAGE

This Dissertation is submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy



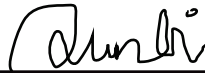
Shuangquan Wang

Approved by the Committee, May 2020



Committee Chair

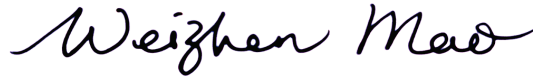
Gang Zhou, Professor, Computer Science
College of William & Mary



Qun Li, Professor, Computer Science
College of William & Mary



Robert Michael Lewis, Associate Professor, Computer Science
College of William & Mary



Weizhen Mao, Professor, Computer Science
College of William & Mary



Weisong Shi, Professor, Computer Science
Wayne State University

COMPLIANCE PAGE

Research approved by

College of William & Mary Protection of Human Subjects Committee

Protocol number(s):

PHSC-2018-02-26-12719-gzhou

PHSC-2019-02-28-13493-gzhou

PHSC-2020-02-06-14118-swang10

Date(s) of approval:

03/07/2018

03/13/2019

03/13/2020

ABSTRACT

Unhealthy dietary habits (such as eating disorders, eating too fast, excessive energy intake, and chewing side preference) are major causes of some chronic diseases, including obesity, heart disease, digestive system disease, and diabetes. Dietary monitoring is necessary and important for patients to change their unhealthy diet and eating habits. However, the existing monitoring methods are either intrusive or not accurate enough. In this dissertation, we present our efforts to use wearable motion sensors to sense mastication dynamics for continuous dietary monitoring.

First, we study how to detect a subject’s eating activity and count the number of chews. We observe that during eating the mastication muscles contract and hence bulge to some degree. This bulge of the mastication muscles has the same frequency as chewing. These observations motivate us to detect eating activity and count chews through attaching a triaxial accelerometer on the temporalis. The proposed method does not record any private personal information (audio, video, etc.). Because the accelerometer is embedded into a headband, this method is comparatively less intrusive for the user’s daily living than previously-used methods. Experiments are conducted and the results are promising. For eating activity detection, the average *accuracy* and *F-score* of five classifiers are 94.4% and 87.2%, respectively, in 10-fold cross-validation test using only 5 seconds of acceleration data. For chew counts, the average error rate of four users is 12.2%.

Second, we study how to recognize different food types. We observe that each type of food has its own intrinsic properties, such as hardness, elasticity, fracturability, adhesiveness, and size, which result in different mastication dynamics. Accordingly, we propose to use wearable motion sensors to sense mastication dynamics and infer food types. We specifically define six mastication dynamics parameters to represent these food properties. They are chewing speed, the number of chews, chewing time, chewing force, chewing cycle duration, and skull vibration. We embed motion sensors in a headband worn over the temporalis muscles to sense mastication dynamics accurately and less intrusively than other methods. In addition, we extract 37 hand-crafted features from each chewing sequence to explicitly characterize the mastication dynamics using motion sensor data. A real-world evaluation dataset of 11 food categories (20 types of food in total) is collected from 15 human subjects. The average recognition accuracy reaches 74.3%. The highest recognition accuracy for a single subject is up to 86.7%.

Third, we study how to detect chewing sides. We observe that the temporalis muscle bulge and skull vibration of the chewing side are different from those of the non-chewing side. This observation motivates us to deploy motion sensors on

the left and right temporalis muscles to detect chewing sides. We utilize a heuristic-rules based method to exclude non-chewing data and segment each chew accurately. Then, the relative difference series of the left and right sensors are calculated to characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side. To accurately detect chewing sides, we train a two-class classifier using long short-term memory (LSTM), an artificial recurrent neural network that is especially suitable for temporal data with unequal input lengths. A real-world evaluation dataset of eight food types is collected from eight human subjects. The average detection accuracy reaches 84.8%. The highest detection accuracy for a single subject is up to 97.4%.

TABLE OF CONTENTS

Acknowledgments	v
Dedication	vi
List of Tables	vii
List of Figures	viii
1 Introduction	2
1.1 Problem Statements	3
1.2 Contributions	5
1.3 Dissertation Organization	7
2 Related Work	9
2.1 Eating Detection and Chews Counting	9
2.2 Food Type Recognition	11
2.3 Chewing Side Detection	14
3 Eating Detection and Chews Counting through Sensing Mastication Muscle Contraction	17
3.1 Introduction	17
3.2 Background and Motivation	19
3.2.1 Four Mastication Muscles	19
3.2.2 Motivation	20
3.3 System Architecture and Implementation	24
3.3.1 Eating Activity Detection	25

3.3.2	Chews Counting	27
3.4	Experiment and Evaluation	29
3.4.1	Data Collection and Ground Truth	29
3.4.2	Evaluation of Eating Activity Detection	30
3.4.3	Evaluation of Chews Counting	34
3.5	Discussion and Future Work	36
3.6	Conclusion	37
4	Inferring Food Types through Sensing and Characterizing Mastication Dynamics	38
4.1	Introduction	38
4.2	Representing Food Properties with Mastication Dynamics	41
4.2.1	Food Properties	41
4.2.2	Food Property Representation	42
4.3	Sensing Mastication Dynamics	43
4.3.1	Why Motion Sensors?	43
4.3.2	Sensor Deployment	44
4.3.3	Motion Data Collection	45
4.4	Characterizing Mastication Dynamics	46
4.4.1	Motion Data Preprocessing	46
4.4.2	Feature Extraction	49
4.5	Food Type Classification	53
4.6	Experiment and Evaluation	54
4.6.1	Experiment Setup	54
4.6.2	Classification Dataset	57
4.6.3	Performance Evaluation with the 11 Food Categories and 20 Food Types	58
4.6.4	Feature Importance Analysis	61

4.6.5	Performance Evaluation of Single-Side Recognition	63
4.7	Comparison with Existing Wearable Sensor-Based Methods	65
4.7.1	Food Types	65
4.7.2	Performance Comparison	67
4.7.3	Intrusiveness Analysis	68
4.8	Discussion and Future Work	69
4.9	Conclusion	69
4.A	Sensor Selection for Extracting Chewing Cycles Dependent Features	70
5	Wearable Motion Sensor-Based Chewing Side Detection	72
5.1	Introduction	72
5.2	System Overview	75
5.3	Motion Sensor Deployment	75
5.4	Motion Data Collection and Calibration	77
5.5	Data Segmentation & Classification Sample Generation	78
5.5.1	Biting Data Exclusion	78
5.5.2	Swallowing Data Exclusion	80
5.5.3	Chew Segmentation	81
5.5.4	Classification Sample Generation	83
5.6	Chewing Side Classification with LSTM	84
5.6.1	Why Personalized Classification Model?	84
5.6.2	Classification Algorithm Selection	84
5.6.3	Architecture and implementation of LSTM model	85
5.7	Experiment and Evaluation	87
5.7.1	Experimental Setup	87
5.7.2	Performance of Chew Segmentation	89
5.7.3	Performance of Chewing Side Detection	90
5.7.4	Impact of Sensor Location Bias	91

5.7.5	Leave-One-Food-Out Test	92
5.8	Discussion	94
5.9	Conclusion	94
6	Future Work	96
7	Conclusion	99
	Bibliography	101
	Vita	115

ACKNOWLEDGMENTS

This dissertation is written with great support and help from many individuals. I would like to extend my heartfelt gratitude to all of them.

First and foremost, I would like to express my deepest appreciation to my advisor, Prof. Gang Zhou. With his encouragement and help, I started my new journey at the College of William & Mary. His expertise, consistent guidance, and continuous support helped me in all the time of my Ph.D. study and related research. Without his guidance and great efforts, this dissertation would not have been possible. I am very grateful to have such a good mentor and friend in my life.

I would also like to thank the rest of my committee members, Prof. Qun Li, Prof. Robert Michael Lewis, Prof. Weizhen Mao, and Prof. Weisong Shi, for serving on my dissertation committee as well as their insightful comments. In addition, I wish to show my gratitude to them for their caring and support in my Ph.D. study.

I want to thank Dr. Shuyin Jiao, Dr. Sarah Glosson, Prof. Zhenming Liu, Prof. Peter Kemper, Prof. Virginia Torczon, Prof. Pieter Peers, James Deverick, Prof. Bin Ren, Prof. Xu Liu, Prof. Chris Shenefiel, Prof. Andreas Stathopoulos, Vanessa Godwin, Dale Hayes, and Jacquelyn Johnson for their help in my study, research, and teaching work.

I wish to thank all the members of the LENS group past and present, Dr. Yongsen Ma, Dr. Hongyang Zhao, Dr. Qing Yang, Dr. Xin Qi, Dr. Ge Peng, Dr. David T. Nguyen, Dr. Daniel Graham, Dr. Yantao Li, Amanda Watson, Woosub Jung, and Minglong Sun for their constructive suggestions, generous assistance, stimulating discussions, and effective teamwork.

Last but not least, I would like to thank my family. Thanks go to my parents and my wife for their love and support. Thank my two kids for bringing me joy and happiness every day.

This dissertation is dedicated to my beloved wife, Yali, and our two kids, Samuel and Celine. Samuel started this journey with us, and Celine joined us in November 2018.

LIST OF TABLES

3.1	The <i>accuracy</i> of five classifiers in CVT	32
3.2	The <i>F-score</i> of five classifiers in CVT	32
3.3	The average <i>accuracy</i> and <i>F-score</i> in LOSOT	34
3.4	Ground truth of chewing of four users	35
3.5	Chews counting results of four users	35
4.1	Food properties represented by each mastication dynamics parameter	43
4.2	The 37 hand-crafted features for mastication dynamics characteriza- tion	53
4.3	The users' demographic information and data collection date . . .	55
4.4	The food types in each food category	55
4.5	Number of detected chewing sequences	57
4.6	Recognition accuracy on the 11 food categories	59
4.7	Recognition accuracy on the 20 food types	59
4.8	Total number of chews during eating all the food	60
4.9	Single-side recognition accuracy on the 11 food categories	64
4.10	Single-side recognition accuracy on the 20 food types	64
4.11	Food types tested by wearable sensor-based methods	66
4.12	Performance comparison of wearable sensor-based methods	67
5.1	The users' demographic information	88
5.2	Chew segmentation accuracy	89
5.3	Chewing side detection accuracy	90

LIST OF FIGURES

3.1	Mastication muscles [1]	19
3.2	(a) Data collection device and deployment; (b) Z axis direction . .	21
3.3	Acceleration data of eating and other six activities	22
3.4	Z axis acceleration and EMG data during eating	23
3.5	System architecture of the proposed method	24
3.6	Convex peak during food intake	28
3.7	Z axis data and reconstructed data	29
3.8	CVT results of five classifiers	31
3.9	ST results of four users	33
3.10	LOSOT results of four users	34
3.11	Less obvious EMG signal	37
4.1	Motivation of our proposed food type recognition method	40
4.2	Wearable device. (a) Hardware platform [2]; (b) Device deployment; (c) Device covered by two polyester tapes; (d) Sensor orientations of the left device; (e) Sensor orientations of the right device	45
4.3	Accelerometer data of the X axis and Z axis. (a) Before calibration; (b) After calibration	46
4.4	Gyroscope data of the Z axis during eating	48
4.5	Three metrics for head motion analysis	49
4.6	The 20 types of food served	56
4.7	Sum of the confusion matrices of the 15 users	61

4.8	The number of occurrences of each selected feature on the datasets of 11 food categories	62
4.9	The number of occurrences of each selected feature on the datasets of 20 food types	62
4.10	One chewing sequence. (a) Accelerometer data; (b) Gyroscope data	70
4.11	Single-sided amplitude spectrum of the gyroscope data in Fig. 4.10 (b)	71
5.1	Flowchart of our proposed chewing side detection method	75
5.2	Temporalis muscle fibers and sensor location	76
5.3	Filtered Z axis gyroscope data	79
5.4	Filtered X axis gyroscope data during chewing and swallowing . .	80
5.5	Distance between peaks and amplitude difference	81
5.6	Architecture of the five-layer LSTM model [3, 4]	86
5.7	Padding effect on unsorted samples	87
5.8	Padding effect on sorted samples	87
5.9	The food served to each user	88
5.10	Sum of the confusion matrices in the five-fold cross validation test	91
5.11	Accuracy comparison between the leave-one-food-out test and five- fold cross validation test	93
5.12	Sum of the confusion matrices in the leave-one-food-out test	93

Dietary Monitoring through Sensing Mastication Dynamics

Chapter 1

Introduction

Unhealthy dietary habits are very common in our daily life. Some representative ones are eating disorder, eating too fast, excessive energy intake, and chewing side preference. Eating disorder is a mental disorder represented by some abnormal eating habits, such as anorexia and bulimia [5]. The anorexia patients obsessively desire to lose weight by refusing to eat [6] or restricting the amount of calorie intake. The bulimia patients have recurrent episodes of binge eating. Then, they utilize some inappropriate compensatory methods to prevent weight gain [7]. Some people eat too fast and chew too little. According to the report of Daily Mail, “people who thoughtfully chew their food and don’t rush mealtimes not only avoid indigestion - they could be preventing diabetes as well [8].” Some people consume excessive calories than they need by eating a lot of high-calorie food, especially the fast food and snack food. Chewing side preference means a tendency to use one side to chew food more than the other side [9].

Unhealthy dietary habits are major causes of some chronic diseases, such as obesity, diabetes, heart disease, and digestive system disease. According to the statistics of the Centers for Disease Control and Prevention (CDC), more than one-third of adults in the United States were obese in 2015 - 2016 [10], more than 100 million Americans had diabetes or prediabetes in 2017 [11], more than 600,000 Americans died of heart disease in 2009 [12], and more than 22 million patients visited physician offices with digestive diseases as the primary diagnosis in 2016 [13]. Deterioration of the situation forces people

to actively monitor their dietary habits.

In recent years, some methods have been proposed to continuously monitor a subject’s dietary habits. Self-report diary based methods [14] are simple and straightforward but tedious and inaccurate. Audio based methods need to deploy sensors in the outer ear [15] or at the throat area [16], which is intrusive. Video based methods do not require a person to wear any sensor, but demand a camera to capture eating process [17] and hence bring privacy concerns. Motion sensor based methods are mainly used to infer eating behaviors through recognizing hand motions [18, 19] or head vibrations [20]. In addition, some other sensors are also utilized, such as electromyography (EMG) sensor and piezoelectric sensor [21]. However, these sensors need to be tightly adhered to skin, which is intrusive and discomforting.

1.1 Problem Statements

In this dissertation, we propose how to exploit wearable motion sensors to sense mastication dynamics for continuous dietary monitoring. Specifically, we work on the following three problems.

(1) Eating Detection and Chews Counting through Sensing Mastication Muscle Contraction. Existing eating detection and chews counting methods are either intrusive or not accurate enough. To solve this problem, we investigate the principle of eating activity and are inspired by following observations: 1) Eating activity is activated through a collaborative effort of four mastication muscles: the masseter, the medial pterygoid, the lateral pterygoid and the temporalis. The first three muscles are near mouth cavity and hence not convenient for sensor deployment. The temporalis is a broad, fan-shaped muscle located at the side of the skull and in front of the ear [22]. This is the area where people often wear a headband or hat. Therefore, the temporalis is most suitable for eating activity sensing; 2) During eating, the temporalis contracts to elevate the mandible, which results in the bulge of this muscle. We are hence motivated to recognize the eating activity through detecting the temporalis contractions and

bulges. This is done through embedding an accelerometer into a headband and attaching the accelerometer on the temporalis; 3) The bulge of the temporalis has the same frequency as chewing. Thus, the number of chews can also be counted through recognizing the frequency of periodic muscle bulges. Based on above observations, we propose to detect eating activity and count the number of chews simultaneously with a triaxial accelerometer. We embed the accelerometer into a headband to sense the temporalis contractions and bulges. However, there are two research questions. *First, how to accurately distinguish eating activity from other non-eating daily activities (reading/speaking, sitting, walking, drinking, coughing, standing, etc.)?* *Second, how to accurately count the number of chews from noisy accelerometer data?* To answer these two questions, we propose an eating activity detection method that extracts representative features to distinguish eating and non-eating activities. In addition, we propose to count the number of chews through identifying the primary periodicity of acceleration data.

(2) Inferring Food Types through Sensing and Characterizing Mastication Dynamics. To recognize food types accurately and less intrusively, we investigate the food properties and mastication dynamics. We are inspired by the following observations: 1) Food properties and mastication dynamics are highly correlated. Each type of food has its own intrinsic properties [23], such as hardness, elasticity, fracturability, adhesiveness, and size. Because the masticatory system is highly adapted to the food properties, the difference in food properties leads to the variance of corresponding mastication dynamics; 2) Mastication dynamics can be sensed by deploying a motion sensor on a mastication muscle. The contraction of a mastication muscle changes the shape of the muscle spindle to make it shorter and thicker. In addition, the muscle contractions are synchronized with the mandible movements. Therefore, the motion sensor can sense mastication dynamics through detecting the muscle contractions and deformations. Based on these observations, we are motivated to deploy motion sensors on the temporalis muscles to sense mastication dynamics and infer food types accordingly. However, this raises three research questions. *First, how to represent food properties using correspond-*

ing mastication dynamics? Second, how to deploy motion sensors on the mastication muscles to sense mastication dynamics accurately and less intrusively? Third, how to characterize the mastication dynamics using motion sensor data? To answer these three research questions, we define six parameters to represent the food properties and deploy motion sensors in a headband to sense the mastication dynamics. In addition, we extract 37 hand-crafted features to explicitly characterize the mastication dynamics.

(3) Wearable Motion Sensor-Based Chewing Side Detection. To detect the chewing sides during eating, we propose to deploy motion sensors on the left and right temporalis muscles to sense the muscle bulge and skull vibration of these two sides. However, there are three research questions. *First, how to segment each chew from the continuous motion sensor data? Second, how to characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side? Third, how to model and classify multi-dimensional data samples with unequal input lengths?* To answer these three questions, we utilize a heuristic-rules based method to exclude non-chewing data and segment each chew accurately. Then, we characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side through calculating the relative difference series of the left and right sensors. In addition, we utilize long short-term memory (LSTM), an artificial recurrent neural network, to model the data samples and classify chewing sides accordingly.

1.2 Contributions

This dissertation proposes three solutions towards wearable motion sensors-based continuous dietary monitoring. The overall contributions are as follows.

Eating Detection and Chews Counting through Sensing Mastication Muscle Contraction. We propose a novel eating activity detection and chews counting method. It is done through identifying the mastication muscle contractions using a triaxial accelerometer attached on the temporalis. Specifically, we make four contributions.

- We propose to detect eating activity and count chews through attaching a triaxial accelerometer on the temporalis. To our best knowledge, this is the first work on motion sensor-based sensing of mastication muscle contraction for simultaneous eating detection and chews counting.
- We design and develop an eating activity detection module. It extracts 23 representative time-domain and frequency-domain features to accurately differentiate eating activity from six other daily activities (reading/speaking, sitting, walking, drinking, coughing and standing) using only 5 seconds of acceleration data.
- We design and develop a chews counting module. It identifies the primary periodicity of highly noisy acceleration data and accurately count the number of chews.
- We evaluate the performance of the proposed method on a real-world dataset. Experimental results show that the average *accuracy* and *F-score* are 94.4% and 87.2%, respectively, for eating activity detection in 10-fold cross-validation test. The average error rate of chews counting for four users is 12.2%.

Inferring Food Types through Sensing and Characterizing Mastication Dynamics. We propose a motion sensor-based food type recognition method. It is done through sensing and characterizing the mastication dynamics using motion sensors deployed on the left and right temporalis muscles. Our main contributions are:

- We propose to infer food types through sensing mastication dynamics with wearable motion sensors. To our best knowledge, this is the first effort in using wearable motion sensors to sense mastication dynamics and recognize food types accordingly.
- We propose to embed motion sensors in a headband and deploy them on the temporalis muscles to sense mastication dynamics accurately and less intrusively.
- We define six mastication dynamics parameters to represent the food properties and extract 37 hand-crafted features from each chewing sequence to explicitly characterize the mastication dynamics.

- We evaluate the performance of our proposed method on a real-world dataset. Experimental results show that the average recognition accuracy of 15 human subjects is 74.3%. The recognition accuracy of a single human subject is up to 86.7%.

Wearable Motion Sensor-Based Chewing Side Detection. We propose a motion sensor-based chewing side detection method. It is done through sensing and characterizing the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side using motion sensors deployed on the left and right temporalis muscles. Our main contributions are:

- We propose to detect chewing sides through sensing muscle bulge and skull vibration with wearable motion sensors. To our best knowledge, this is the first effort in using wearable motion sensors to differentiating muscle bulge and skull vibration between the chewing side and the non-chewing side.
- We propose a heuristic-rules based method to accurately exclude non-chewing data and segment each chew from continuous motion sensor data.
- We propose to calculate the relative difference series of the left and right sensors to characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side.
- We evaluate the performance of our proposed method on a real-world dataset. Experimental results show that the average detection accuracy reaches 84.8%. The detection accuracy of a single human subject is up to 97.4%.

1.3 Dissertation Organization

The rest of this dissertation is structured as follows. In Chapter 2, we discuss related work. In Chapter 3, we present our eating detection and chews counting method. In Chapter 4, we present our food type recognition method. In Chapter 5, we present our

chewing side detection method. The future work is described in Chapter 6. Finally, we conclude in Chapter 7.

Chapter 2

Related Work

This chapter reviews related work in eating detection and chews counting, food type recognition, and chewing side detection, respectively.

2.1 Eating Detection and Chews Counting

A self-report diary [24] was often used to record the eating activities. However, people may record incorrectly [25] or forget to write a diary. Furthermore, few people are willing to estimate the eating speed or count the number of chews during eating. Recently, some progress has been made on automatic eating detection and/or chews counting, including audio based methods, video based methods and motion based methods.

Audio based methods. Reference [16] used a modified Bluetooth headset with embedded microphone to collect sounds in a user’s throat area. Time domain features, frequency domain features and cepstral features were extracted from the recorded sounds to train the classification model. The F-measure accuracy reaches 79.5% and 71.5% for laboratory study (12 activities) and small-scale in-the-wild study (4 activities), respectively. Reference [15] placed the microphone inside the ear canal to differentiate chewing from speech and silence. At the same time, this method can also differentiate several types of food. Even though the above related works demonstrate the validness of the audio based methods, a user may not be willing to wear a microphone near the throat

or in the ear during eating. Besides, audio recording raises potential privacy concerns.

Video based methods. An Active Appearance Model (AAM) was utilized in [17] to track a subject’s face and detect chewing activity from surveillance video. This is based on the observation that variations in AAM parameters have distinct periodicity during chewing. The experimental results demonstrated a cross-validated percentage agreement of 93.0%. The video based method needs no on-body sensor, and the video sequence can be acquired without any user intervention. However, this method brings many privacy concerns. Its accuracy is also affected by environmental lighting changes and face occlusion.

Motion based methods. In recent years, the popularity of wearable devices (smart glasses, smartwatch, smart wristband, etc.) has made motion based methods possible. Reference [26] combined accelerometers from a smartwatch and Google glass to recognize a user’s eating activity. Reference [27] only used a glasses mounted accelerometer to distinguish chewing from non-chewing activities. Similarly, reference [28] used motion sensors on Google glass to detect head movement and infer eating activity. Reference [20] integrated an EMG sensor and vibration sensor into 3D-printed eyeglasses for detecting chewing and identifying food categories. Reference [18] used a watch-like device, which is embedded with a micro-electro-mechanical gyroscope, to track wrist motions and detect food intake. Moreover, reference [29] used a smartwatch to detect eating activity. Reference [30] designed a sensor-embedded digital fork, Sensing Fork, to sense a child’s eating behavior. Furthermore, a mobile game named Hungry Panda was developed to encourage the child to eat diverse foods during mealtime. Although the above methods also utilize motion sensors, they mainly focus on detecting head motion and/or hand motion. Comparatively, our method aims to detect mastication muscle contractions and corresponding bulges. Moreover, these methods are mainly for eating or chewing activities detection. Chews counting is not included. Li et al [31] proposed a novel method that embeds a small accelerometer inside artificial teeth to capture unique motion characters during chewing, drinking, speaking and coughing. This method is too intrusive to

be widely accepted. In addition, safety issues and frequent battery changes also need to be addressed. In [32], we published a poster paper to present our preliminary idea, by which this technology had not been fully developed yet. In that poster publication, we used a single axis accelerometer to detect eating activity and count chews. However, this method requires much larger sliding windows. Besides, further experiments show that its eating detection accuracy is very low for the leave-one-subject-out test. We believe that it is because one axis acceleration data is not enough to distinguish motion characters between eating and other activities, and the extracted features are not representative enough.

Other methods. In addition to audio, video and motion based methods, there are some other ways for eating detection and chews counting. For example, references [21] and [33] deployed a piezoelectric sensor below the ear to capture the movement of the lower jaw and detect chewing rate. Reference [34] placed a piezoelectric sensor on the temporalis to detect the chewing bouts. However, the piezoelectric sensor needs to be attached on the skin tightly. This is intrusive and unfriendly for users.

2.2 Food Type Recognition

The existing food type recognition methods are divided into three categories: image based methods, ambient sensor based methods, and wearable sensor based methods. We introduce some representative works for each category as follows.

Image based methods. The image based methods take advantage of image processing algorithms to extract features from food pictures and build recognition models. This kind of methods mainly focuses on food appearances, such as shape, texture, color, and size. For example, Anthimopoulos et al. [35] proposed a bag-of-features (BoF) model-based method to distinguish 11 food categories. A set of characteristics was extracted and quantified to describe the visual content in the food images. Then this set of characteristics was used to train a classifier for food type recognition. The accuracy of five-fold

cross-validation test is 78%. Bossard et al. [36] introduced a large-scale benchmark dataset, Food-101. This dataset contains 101,000 images of 101 food categories. The authors used Random Forests to mine discriminative components in these images. The average accuracy of this method is 50.76%. Yang et al. [37] exploited the spatial relationship between the different ingredients of the food. The authors proposed to calculate pairwise statistics between local features to represent the food items and build a discriminative classifier. The recognition accuracies on 61 food types and 7 food categories are 28% and 78%, respectively. Kawano and Yanai [38] proposed a smartphone-based mobile food recognition system to recognize food items and estimate their calories. In general, the image-based methods can recognize more food items than other existing methods. However, it is difficult for them to distinguish food types with similar appearances. In addition, they are sensitive to environmental lighting conditions and view occlusion.

Ambient sensor based methods. Ambient sensor based methods often embed sensors in specially designed tablewares to sense the physical or chemical properties of the food. Kadamura et al. [30] introduced a sensor-embedded digital fork, the Sensing Fork. A single-pixel RGB color sensor was used to sample the food color and classify the food types. The F-measure accuracy of 10-fold cross-validation test on 17 food types is 87.5%. Lester et al. [39] proposed to embed spectrometer and pH/Conductivity probes into a cup to sense and classify liquids. The accuracy is up to 79% for 68 different drinks. Zhou et al. [40] embedded a fabric pressure sensor matrix on top of a dining tray and four force sensitive resistors (FSRs) under the four corners. These sensors were used to detect cutlery usage (cutting, poking, scooping, etc.) and infer the food types. The ambient sensor based methods are friendly to use. However, these host objects are not convenient to carry and only available at particular locations. This restricts the application scope of this kind of methods.

Wearable sensor based methods. Wearable sensor based methods embed various sensors into wearable devices. The commonly used sensors include microphone [41, 42, 43, 44, 45, 15, 16, 46], EMG [47, 20], piezoelectric sensor [48, 49], and motion sensor

[50, 51]. Päßler et al. [41] integrated two microphones into a hearing aid package. One was placed in the ear canal to record the chewing sounds. The other was attached on the hearing aid to record the environmental sounds as the reference. A Viterbi algorithm based finite-state grammar (FSG) decoder was used to recognize seven types of food and one drink. The food classification accuracy is 79% on a test set of 10% of all records. Bi et al. [43] designed a microphone embedded hardware prototype. It was worn on the subject’s neck to record acoustic signals during eating. The recognition accuracy on seven food types is 84.9%. Amft et al. [44] embedded a miniature microphone into an ear-pad case to recognize 19 standard food types. The reported accuracy is 80%. These above methods demonstrate the validness of audio based food recognition. However, they need to deploy a microphone in the outer ear [41, 44, 15] or at the throat area [43, 16]. Deploying a microphone in the outer ear may block the ear canal and affect daily communication; deploying a sensor at the throat area is intrusive and uncomfortable. Moreover, the audio recording may bring some privacy concerns.

Zhang et al. [47] embedded EMG electrodes into a 3D-printed eyeglass to detect chews and recognize five food types. The accuracy is between 43% and 71% for individual chewing cycles, and between 63% and 84% for intake sequences. Using a similar eyeglass, reference [20] combined EMG and an accelerometer to monitor chewing and identify a few food types. The accelerometer was placed on the temple of the eyeglass to sense the skull vibration. Kalantarian et al. [48] embedded a piezoelectric sensor into a necklace to distinguish between solid and liquid. Alshurafa et al. [49] presented a similar piezoelectric sensor embedded necklace. The sensor was contacted with the skin in the lower trachea. The F-measure accuracy on distinguishing liquids and solids is above 90%. However, the EMG and the piezoelectric sensor need to be adhered on skin tightly. It is intrusive and uncomfortable to wear.

Motion sensors are often used to detect eating [26, 27, 28, 52, 53, 54] and count the number of chews [32]. However, there are only a few works on the motion sensor based food type recognition. Kim et al. [51] utilized an accelerometer embedded wristband to

detect 29 eating actions. The detection results were used to indirectly infer two different food types, the rice and noodle. Mirtchouk et al. [50] combined a microphone embedded earbud, a Google Glass and two smartwatches to recognize 40 types of food and estimate the amount of food consumed. The motion sensors in the Google Glass and the smartwatches were used to catch the head and wrist motions, respectively. Comparatively, our proposed method utilizes motion sensors to directly sense the mastication dynamics and infer food types accordingly, instead of indirectly inferring food types from the head and wrist motions. In addition, their method requires using a camera to record the video during model training and testing. The video is used to manually annotate the exact time of food delivery, food intake, and chewing. This is obviously intrusive.

2.3 Chewing Side Detection

The existing methods of chewing side preference detection can be roughly divided into two categories: clinical methods and wearable sensor-based methods.

Clinical methods. The clinical methods [55] take advantage of medical professionals’ observations or special diagnosis devices. Nissan et al. [56] utilized direct visual observation to detect the chewing side preference. A piece of chewing gum was placed on the center of the dorsal aspect of the tongue. Then, the chewing side preference was determined by “observing the direction towards which the gum was moved by the tongue for the first cycle of mastication [56]”. Watarai et al. [57] attached six landmarks on a subject’s face, including the left and right corners of the mouth, the superior and inferior labial tubercles, the pogonion, and the tip of the nose. The movements of these landmarks were captured by two CCD cameras during chewing. Then, the three-dimensional coordinates and trajectories of these landmarks were extracted to differentiate the movements between the working-side (i.e. the chewing side) and the balancing-side (i.e. the non-chewing side). Mizumori et al. [58] used a Sirognathograph Analyzing System to record the movement traces of the incisal point during chewing. Then, these traces were compared with the reference traces to determine the side of each chewing stroke. Based

on the number of left side strokes and the number of right side strokes, an Asymmetry Index (AI) was calculated to quantitatively evaluate the masticatory laterality. The definition of AI is as follows [58]:

$$AI = \frac{\text{Number of right side strokes} - \text{number of left side strokes}}{\text{Number of right side strokes} + \text{number of left side strokes}} \times 100\% \quad (2.1)$$

Here, $AI = 0$ represents chewing equally on both sides; $AI = -100\%$ represents chewing only on the left side; $AI = 100\%$ represents chewing only on the right side. Yamasaki et al. [59] proposed to detect chewing sides by measuring the EMG activities of the bilateral masseter muscles. Based on the collected EMG data, a metric named “level of amplitude against the maximum voluntary contraction” was calculated for the left side and right side, respectively, for each chew. The side that has a larger metric value was considered as the actual chewing side. The above clinical methods require special diagnosis devices and/or the involvement of medical professionals, which are mainly available in clinical environments.

Wearable sensor-based methods. In order to detect chewing side preference continuously and conveniently in our daily living, several wearable sensor-based methods have been proposed in recent years. Wearable sensor-based methods embed sensors into various wearable devices to detect chewing sides. Chung et al. [60] embedded two load cells into hinges on both sides of a pair of glasses to recognize ingestive and facial behaviors, such as head movements, left chewing, right chewing, winking, and talking. The load cell measures the force amplified by the hinge. Similar to our proposed method, Chung et al.’s method takes advantage of the contraction and relaxation motions of the temporalis muscles. However, it senses the muscle motions indirectly using load cells embedded in the glasses. Our method senses the muscle bulge and skull vibration directly using motion sensors deployed on the temporalis muscles. Compared with our method, Chung et al.’s method heavily relies on the sensitivity and deployment of the load cells. In addition, it is a coarse-grained method because each motion detection is based on a fixed-length window of 3 seconds instead of one chew. Lucena et al. [61] attached

two motion sensors to a subject’s jaw and forehead to measure angle variations and detect chewing sides accordingly. However, attaching sensors on the jaw and forehead is obviously intrusive and not suitable for long-term wear. Zhang and Amft [62] embedded EMG electrodes into 3D-printed eyeglasses to detect eating and chewing activities. They indicated that “bilateral EMG measurement may serve to study chewing side variation”. However, they did not investigate this problem. The chewing side detection accuracy of this method is unknown. In addition, the EMG signal is easily affected by the hair.

Chapter 3

Eating Detection and Chews Counting through Sensing Mastication Muscle Contraction

3.1 Introduction

Nowadays chronic diseases have become one of the most serious threats to human health [63]. Eating detection and chews counting are helpful for monitoring eating time, estimating eating speed and food amount, changing unhealthy eating habits, and preventing related chronic diseases. In recent years, some methods have been proposed to recognize a subject's eating activity and count the number of chews. However, these methods are either intrusive or not accurate enough.

How to detect a subject's eating activity and count the number of chews in an accurate and noninvasive way? To answer this question, we investigate the principle of eating activity and are inspired by following observations: 1) Eating activity is activated through a collaborative effort of four mastication muscles: the masseter, the medial pterygoid, the lateral pterygoid and the temporalis. The first three muscles are near mouth cavity and hence not convenient for sensor deployment. The temporalis is a broad, fan-shaped muscle located at the side of the skull and in front of the ear [22].

This is the area where people often wear a headband or hat. Therefore, the temporalis is suitable for noninvasive sensing of eating activity; 2) During eating, the temporalis contracts to elevate the mandible, which results in the bulge of this muscle. We are hence motivated to recognize the eating activity through detecting the temporalis contractions and bulges. This is done through embedding an accelerometer into a headband and attaching the accelerometer on the temporalis; 3) The bulge of the temporalis has the same frequency as chewing. Thus, the number of chews can also be counted through recognizing the frequency of periodic muscle bulges.

In this chapter, we propose to detect eating activity and count chews simultaneously with a triaxial accelerometer. We embed the accelerometer into a headband to sense the temporalis contractions and bulges, which is convenient and less intrusive. Compared with existing audio or video based methods, our method only records the acceleration data. Therefore, our method has less privacy concerns.

We summarize our contributions as follows:

1. We propose to detect eating activity and count chews through attaching a triaxial accelerometer on the temporalis. To our best knowledge, this is the first work on motion sensor-based sensing of mastication muscle contraction for eating detection and chews counting simultaneously.
2. We design and develop an eating activity detection module. It accurately differentiates eating activity from six other daily activities (reading/speaking, sitting, walking, drinking, coughing and standing) using only 5 seconds of acceleration data.
3. We design and develop a chews counting module. It identifies the primary periodicity of highly noisy acceleration data and accurately count the number of chews.
4. We evaluate the performance of the proposed method on a real-world dataset. Experimental results show that the average *accuracy* and *F-score* are 94.4% and

87.2%, respectively, for eating activity detection in 10-fold cross validation test. The average error rate of chews counting for four users is 12.2%.

The rest of this chapter is organized as follows. Section 3.2 introduces background and motivation of the proposed method. Section 3.3 describes the system architecture and implementation of each module. Experiment and evaluation are presented in Section 3.4. Section 3.5 presents discussion and future work. Conclusion is drawn in Section 3.6.

3.2 Background and Motivation

In this section, we first introduce four mastication muscles. Then we present the motivation of acceleration data based eating activity detection and chews counting.

3.2.1 Four Mastication Muscles

From a physiological point of view, there are four mastication muscles: the masseter, the medial pterygoid, the lateral pterygoid and the temporalis [64], as shown in Fig. 3.1. During eating, these four muscles work together, enabling jaw open-close movements, to cut and grind the food.

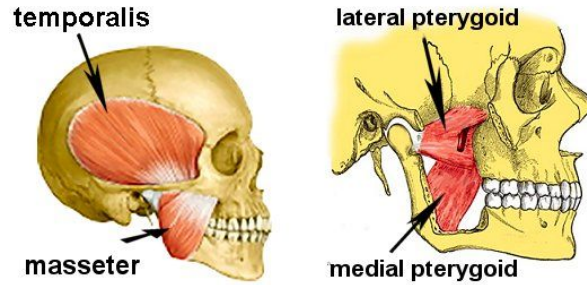


Figure 3.1: Mastication muscles [1]

The masseter is located on each side of a face. It connects the maxillae and the mandible, and primarily serves for elevating the mandible while the deep tissues help protrude the mandible forward [65]. The masseter is the most superficial muscle. It is also one of the strongest mastication muscles.

The medial pterygoid and the lateral pterygoid are located on the inner surface of the mandible. Contraction of the medial pterygoid helps elevate the mandible, and thus contributes to jaw-closing. However, the lateral pterygoid helps lower the mandible and open the jaw. It is the only mastication muscle for jaw-opening [66].

The temporalis is a broad, fan-shaped muscle located on the side of the skull and in front of the ear [22]. It is one of three muscles that close the jaw and clench the teeth.

From the above introduction, we see that the masseter, medial pterygoid and lateral pterygoid are in the face area and near mouth cavity. This area is inconvenient and invasive for sensor deployment. On the contrary, the temporalis covers a bigger area. This makes it convenient and less-invasive to attach a sensor, such as in a headband or hat.

3.2.2 Motivation

It is common knowledge that a muscle bulges when contracting. This motivates us to detect a subject’s eating activity using a triaxial accelerometer attached on the temporalis. Moreover, as the temporalis contracts and bulges once for each chew, the number of chews can also be counted by recognizing the frequency of periodic muscle bulges. In this subsection, we demonstrate the potential of utilizing acceleration data to detect eating activity and count the number of chews.

In the experiment, Shimmer2r wireless sensor platform [67] is used for data collection. Shimmer2r has an integrated accelerometer and can be connected to several types of external sensors, such as electrocardiogram (ECG), EMG, GPS, etc. We use triaxial accelerometer and EMG sensors to sample acceleration data and EMG data simultaneously. The acceleration data is utilized for eating activity detection and chews counting. The EMG data is used to obtain the ground truth of chews counting.

The sensor platform and its deployment are shown in Figure 3.2 (a). The EMG sensor has three electrodes: a positive electrode, a negative electrode and a neutral reference electrode. The sensing device is fixed in a headband using scotch tape. The user wears

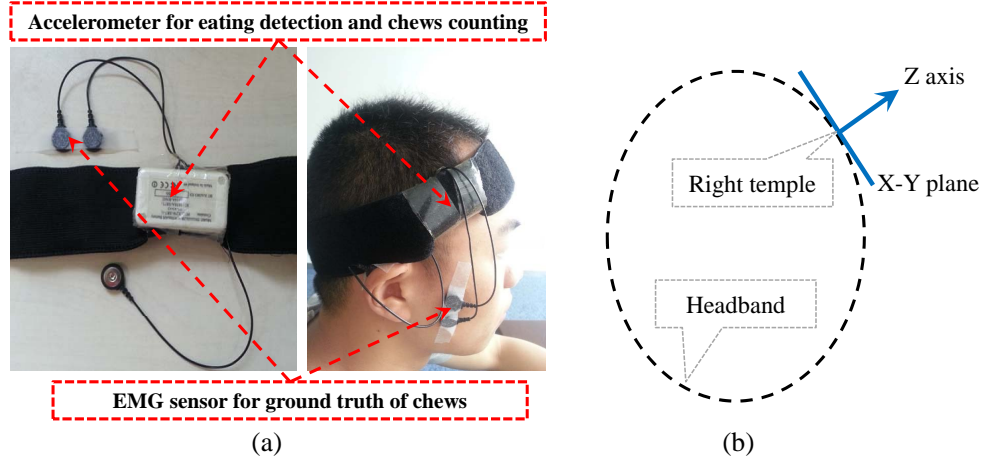


Figure 3.2: (a) Data collection device and deployment; (b) Z axis direction

the headband and places the device near the right temple. The headband is elastic and adjustable. The X axis and Y axis of the accelerometer are perpendicular. These two axes form the tangent plane of the skull at the contact position. The Z axis of the accelerometer directs outward and is vertical to the X-Y plane, as shown in Figure 3.2 (b). The accelerometer is calibrated as following: when one axis points downwards, its acceleration measurement is set to g (gravity); when it points upwards, the measurement is set to $-g$. For the EMG sensor, both positive and negative electrodes are attached on the right side of the face to detect contraction of the masseter. The neutral reference electrode is attached at the ear edge, where there is no muscle but just bone and skin. Hence, it is selected as the electrically neutral point of the body.

The sampling rates of the accelerometer and EMG sensors are the same and about 100Hz. All the collected data is wirelessly transmitted to the laptop through BlueTooth. Data of each continuous sampling process is stored in one file for post processing. The data of eating (while sitting) and six other non-eating daily activities (reading/speaking, sitting, walking, drinking, coughing and standing) are collected separately. The subject is served with multiple small pieces of watermelon. The reason for choosing watermelon will be discussed in Section 3.5.

Figure 3.3 shows the acceleration data of eating and six other daily activities. We

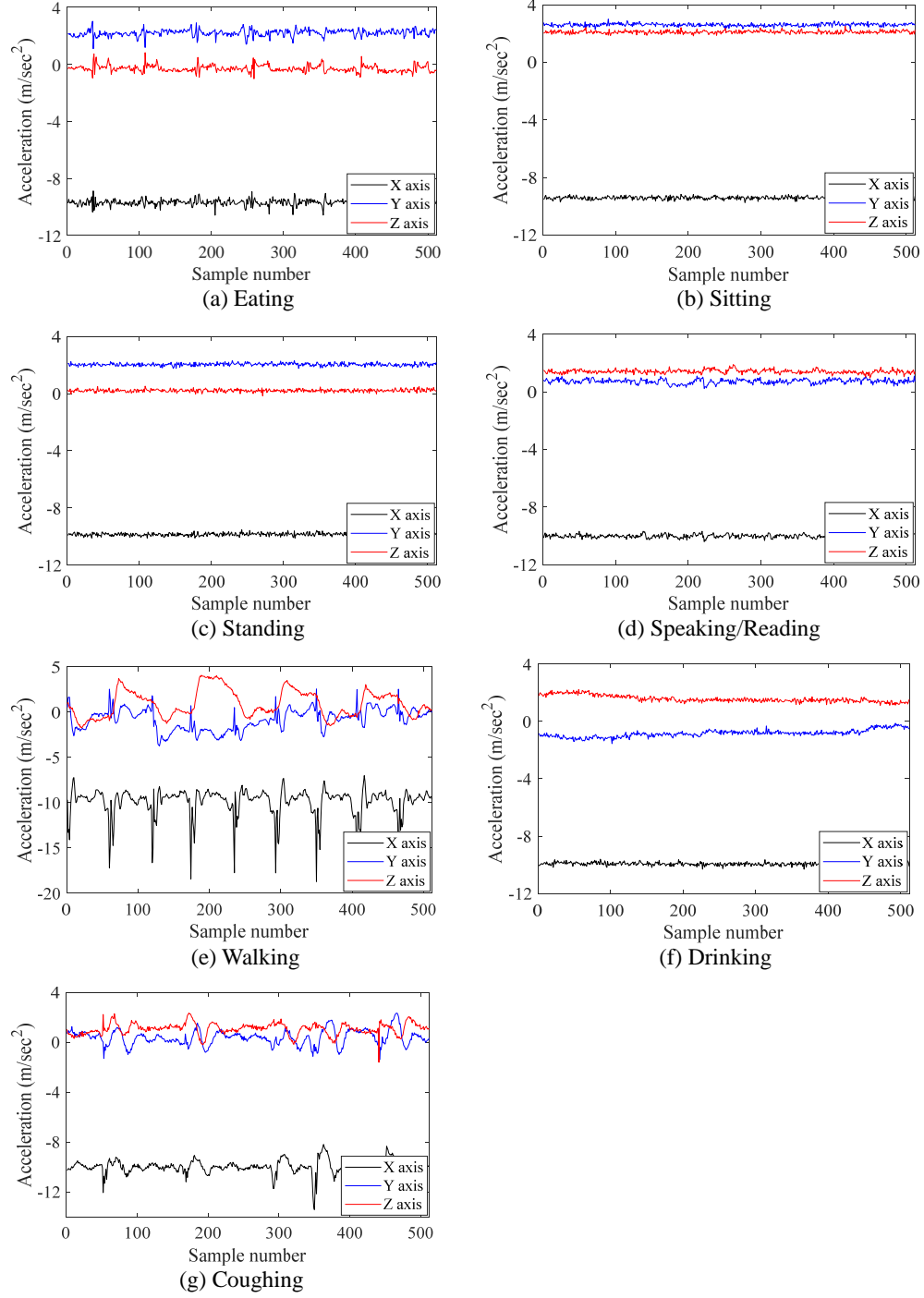


Figure 3.3: Acceleration data of eating and other six activities

observe that the acceleration data of eating activity shows clear periodic pattern, especially for the Z axis data. This is because the bulge of the temporalis is in the same

direction as the Z axis. In addition, the characters of the acceleration data indicate that it is possible to distinguish eating activity from other six activities. For walking and coughing activities, their fluctuation amplitude is much larger than that of eating activity. For sitting, standing, speaking/reading and drinking activities, even though they have similar fluctuation amplitude with that of eating activity, they have no clear periodicity. Therefore, we can extract fluctuation amplitude and periodicity related features to identify eating activity from other noisy activities.

Figure 3.4 shows an example window of raw Z axis acceleration data (m/sec^2) and EMG data (mVolts) during eating, indicated by a blue solid line and a red dashed line, respectively. We observe that: 1) both signals have the same periodic cycles. Each cycle corresponds to one chew; 2) the EMG signal has obvious spikes at the moments of masseter contraction. It almost equals to zero between two neighboring chews; 3) the Z axis acceleration data also has obvious increase during the muscle contraction, but there is some fluctuation between two neighboring chews.

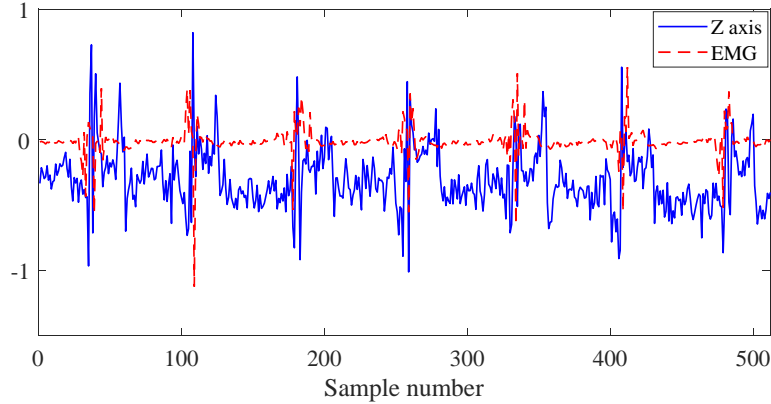


Figure 3.4: Z axis acceleration and EMG data during eating

According to the above observations, we believe that, despite many challenges, there is a high possibility of differentiating eating from other daily activities. Furthermore, it is feasible to count the number of chews only using Z axis acceleration data.

3.3 System Architecture and Implementation

Figure 3.5 shows the system architecture of the proposed method. It contains two main modules: an eating activity detection module and a chews counting module.

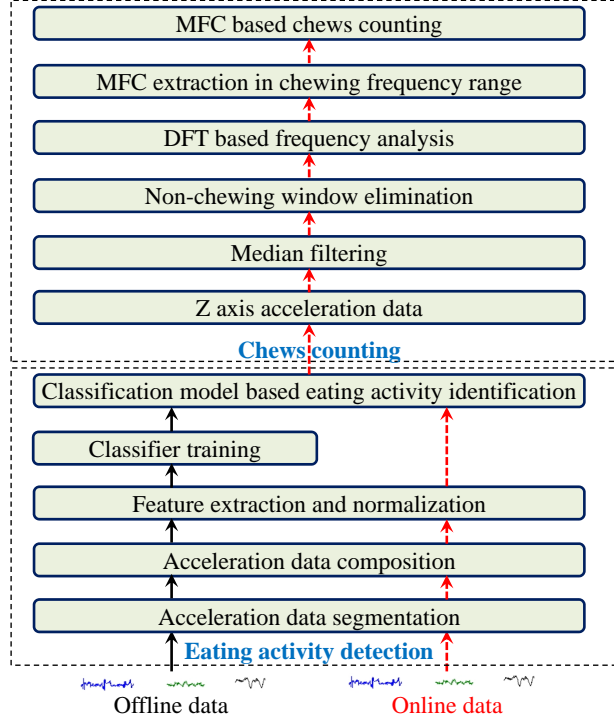


Figure 3.5: System architecture of the proposed method

The eating activity detection module includes two processes: offline training process and online testing process. They are marked with black arrows and red arrows, respectively. In the offline training process, with the offline data collected from the tri-axial accelerometer, a sliding window of length L without overlap is used to segment the sensor data. For each window, the eating activity detection module first composes the acceleration data of three axes, and then extracts representative features. By combining normalized feature vectors with corresponding class labels, the training dataset is built to train a two-class classifier for online recognition. The online testing process has the same operations of data segmentation, data composition, feature extraction and normalization with those of the offline training process. After that, unlabeled feature vectors

are fed to the trained classification model. Then, the eating activity detection results are obtained.

If one window is identified as eating activity, the raw Z axis acceleration data is used to count the number of chews in this window. The Z axis acceleration data is first filtered using median filter to remove some sampling noise. Then, the non-chewing window is eliminated through checking the difference of acceleration magnitude. For a chewing window, discrete Fourier transform (DFT) based frequency analysis is applied, and the maximum frequency component (MFC) in the chewing frequency range is extracted. Finally, the number of chews in this window is estimated based on the frequency corresponding to MFC and the time length of the sliding window.

We introduce the detailed implementation of the above two modules as following.

3.3.1 Eating Activity Detection

After segmenting the triaxial acceleration data with a sliding window of length L without overlap, we compose the i th sensor readings of three axes, a_X^i , a_Y^i and a_Z^i into one scalar acceleration a_i :

$$a_i = \sqrt{(a_X^i)^2 + (a_Y^i)^2 + (a_Z^i)^2} \quad (3.1)$$

where $i = 1, \dots, L$.

Based on the composed data, four groups of features are extracted to build the feature vector for this window of data. The first group consists of six time domain features: the maximum, the minimum, the 1st quartile, the 2nd quartile, the 3rd quartile, and the number of mean cross.

The second group consists of four amplitude statistics features extracted from the composed window data. They are defined [68] as:

$$Amplitude : \mu_{amp} = \frac{1}{L} \sum_{i=1}^L a_i \quad (3.2)$$

$$\text{Standard deviation : } \sigma_{amp} = \sqrt{\frac{1}{L} \sum_{i=1}^L (a_i - \mu_{amp})^2} \quad (3.3)$$

$$\text{Skewness : } \gamma_{amp} = \frac{1}{L} \sum_{i=1}^L \left(\frac{a_i - \mu_{amp}}{\sigma_{amp}} \right)^3 \quad (3.4)$$

$$\text{Kurtosis : } \beta_{amp} = \frac{1}{L} \sum_{i=1}^L \left(\frac{a_i - \mu_{amp}}{\sigma_{amp}} \right)^4 - 3 \quad (3.5)$$

The third group consists of four amplitude statistics features extracted from single-sided amplitude spectrum (without direct current component) after Fourier transform [69]. These features can be computed using the above four formulas after replacing L and a_i with $\frac{L}{2}$ and s_i , respectively. Here, s_i means the i^{th} component of single-sided amplitude spectrum.

According to [70], chewing activity mainly occurs in the range between 0.94 Hz (5th percentile) and 2.17 Hz (95th percentile). Here, we define the chewing frequency range as 0.5 Hz to 2.5 Hz. Then, the single-sided amplitude spectrum (without direct current component) can be partitioned into three bands: (0, 0.5) Hz, [0.5, 2.5] Hz and (2.5, $SF/2$] Hz. SF means the sampling frequency of accelerometer. Three features are extracted from each band to form the fourth group of features. They are the MFC, the location (i.e. the index) of the MFC, and the spectral energy. The spectral energy is defined as the sum of squared spectrum components in each band.

In total, 23 features are extracted. To eliminate the scaling effects among different features, all the features are normalized using the z-score normalization algorithm [71].

The eating activity detection is formulated as a two-class classification problem. The positive class corresponds to eating activity, while the negative class corresponds to other daily activities, such as speaking/reading, sitting, standing, walking, drinking, coughing, etc. Five commonly used classification algorithms are compared: decision tree (DT), nearest neighbor (NN), multi-layer perceptron (MLP), support vector machine (SVM) and weighted support vector machine (WSVM).

DT algorithm builds a pattern classifier from a labeled training dataset using a divide-and-conquer approach. It recursively selects the attribute that is used to partition the training dataset into subsets until each leaf node in the tree has a uniform class membership [72]. NN algorithm is an instance-based learning method. It only stores the training samples but does not generate a specific classification model. During classification, the distances between the test sample and all training samples are calculated. The test sample is assigned the same class label as its nearest neighbor. MLP algorithm is a feedforward artificial neural network model. It maps a set of inputs onto a set of appropriate outputs. It uses a supervised technique called backpropagation to train the network and obtain the parameters [73]. SVM algorithm is based on the foundation of statistical learning theory. It gains promising empirical performance in the fields of nonlinear and high dimensional pattern recognition [72, 74].

The above four algorithms have different rationales and model structures. Comparison of their recognition results should demonstrate performance of the proposed method on eating activity detection in a comprehensive and unbiased way. WSVM can deal with the uneven class size problem of SVM by assigning larger weights to classes with fewer samples [75]. Therefore, it is also included for comparison.

3.3.2 Chews Counting

The Z axis acceleration data is used to count the number of chews in one sliding window. The chews counting module contains the following five steps:

Step 1: Median filtering. We first use a 7^{th} -order one-dimensional median filter [76] to remove the sampling noise in the acceleration data. The median filter runs through the sliding window sample by sample, and replaces each sample with the median of neighboring samples [77].

Step 2: Non-chewing window elimination. In the eating activity detection module, one window is identified as eating or non-eating activity. An eating window may contain not only chewing, but also food intake and swallowing. From Figure 3.2 (b) we can see

that, when the user bows down his head and feeds food into his mouth, the Z axis directs downwards in some degree. Accordingly, there is a positive decomposition of gravity on the Z axis. The gravity decomposition generates a large convex peak of the Z axis acceleration data, as shown in Figure 3.6. Normally, the magnitude of the convex peak is much larger than the acceleration variation during chewing. For simplicity, we calculate the difference between the maximum acceleration and the minimum acceleration in one window. If the difference is larger than a predefined threshold, *MagDiff*, this window is considered as a non-chewing window and eliminated.

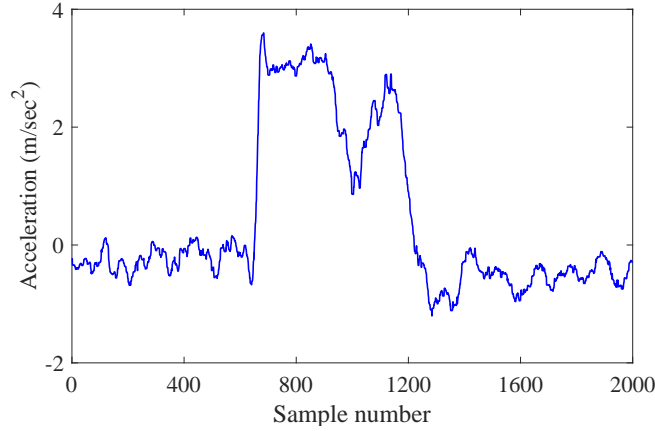


Figure 3.6: Convex peak during food intake

Step 3: DFT based frequency analysis. For one chewing window, we need to count the number of temporalis bulges. The most straightforward method is to count the peaks of Z axis acceleration data. However, as shown by the red line in Figure 3.7, the Z axis acceleration data is very noisy even after median filtering. There are lots of false peaks caused by the vibration of the skull during chewing. One observation is that the chewing frequency is consistent with the primary periodicity of the acceleration data. We use the MFC to reconstruct the acceleration data only with primary periodicity. The reconstructed data is shown as the blue line in Figure 3.7. Therefore, for chews counting, we propose to utilize DFT to transform the acceleration data from time domain into frequency domain.

Step 4: MFC extraction in chewing frequency range. After DFT based frequency

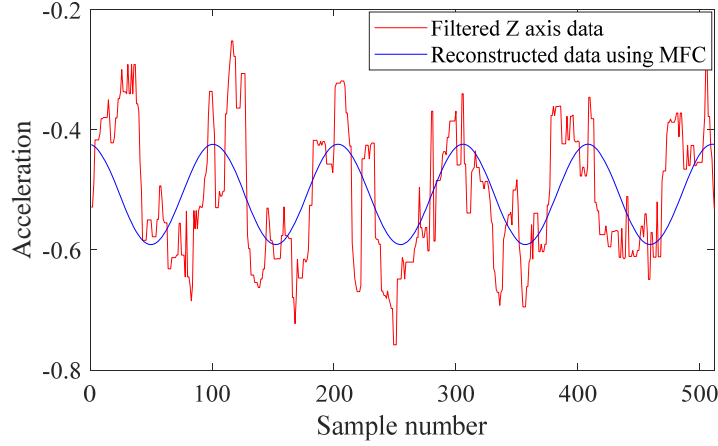


Figure 3.7: Z axis data and reconstructed data

analysis, we extract the MFC in the chewing frequency range, i.e. $[0.5, 2.5]$ Hz.

Step 5: MFC based chews counting. We take the frequency corresponding to MFC, f_{MFC} , as the approximate chewing frequency. Then we estimate the number of chews in one window by multiplying f_{MFC} with the time length of the window. The time length can be obtained through dividing the window length by the sampling rate.

3.4 Experiment and Evaluation

In this section, we introduce the experimental evaluation on eating activity detection and chews counting.

3.4.1 Data Collection and Ground Truth

We recruit four volunteers to collect the experimental dataset [32] of seven daily activities, including eating, reading/speaking, standing, sitting, walking, drinking and coughing. For eating, reading/speaking, standing, sitting and walking activities, one volunteer performs each of them for 6 to 9 minutes. For drinking and coughing activities, one volunteer performs each of them about 30 seconds. For reading/speaking and walking, the volunteers are asked to perform in three different speeds (slow, moderate and fast), and each speed for 2 to 3 minutes. The acceleration data of three axes and the EMG

data are sampled simultaneously. In total, about 150 minutes of data is collected.

These activities are manually labeled during data collection. To serve as ground truth of chews counting, the EMG data is manually identified and counted to obtain the total number of chews for each volunteer.

3.4.2 Evaluation of Eating Activity Detection

Three tests are conducted to evaluate the eating detection performance of the proposed method. 1) Cross validation test (CVT). CVT combines all subjects' samples to form the dataset. It uses the cross validation method to evaluate the general eating detection accuracy on multiple subjects; 2) Self test (ST). ST only uses the samples of the subject himself/herself to form the dataset. For the ST evaluation of each subject, the same cross validation method as above is used; 3) Leave-one-subject-out test (LOSOT). LOSOT uses the samples of all subjects except one to form the dataset and train the classification model accordingly. Then the model is tested using the samples of the excluded subject. LOSOT shows how generic the detection model is for unknown subjects.

Weka toolkit [72] is used for classifier training and testing. For DT, the J48 algorithm is used. For SVM and WSVM, the LibSVM wrapper for Weka [78] is used. We adopt the default parameters for all classifiers in the following experiments. For cross validation, the fold number is set to 10. Because the samples of negative class are about three times that of positive class, the weights of WSVM are set to 3 for positive class and 1 for negative class.

Four evaluation metrics are used to quantify the classification performance. They are *accuracy*, *precision*, *recall* and *F-score*, which are defined as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.6)$$

$$precision = \frac{TP}{TP + FP} \quad (3.7)$$

$$recall = \frac{TP}{TP + FN} \quad (3.8)$$

$$F\text{-score} = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (3.9)$$

where TP denotes true positive, TN denotes true negative, FP denotes false positive, and FN denotes false negative.

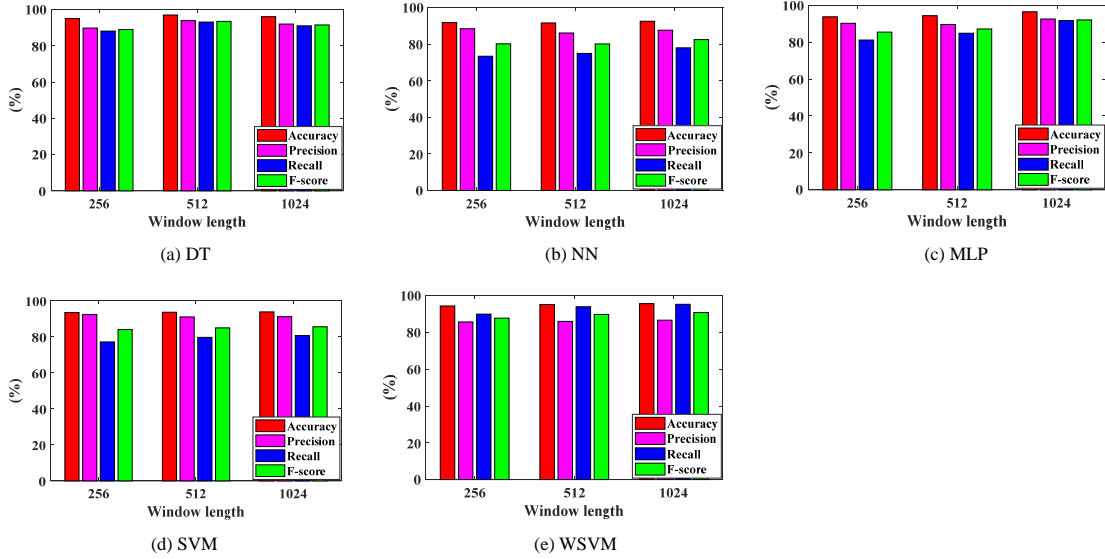


Figure 3.8: CVT results of five classifiers

Cross validation test (CVT). Figure 3.8 shows the *accuracy*, *precision*, *recall* and *F-score* of CVT for different window lengths: 256, 512 and 1024. From Figure 3.8 we see that: (1) For all these five classification models, their *accuracy* is larger than 90%, and their *F-score* is larger than 80%. Table 3.1 and Table 3.2 give the *accuracy* and *F-score* of the five classifiers. The average *accuracy* of the five classifiers are 93.8% (window length of 256), 94.4% (window length of 512) and 95.0% (window length of 1024), respectively. The average *F-score* of the five classifiers are 85.4% (window length of 256), 87.2% (window length of 512) and 88.5% (window length of 1024), respectively. This indicates that the proposed eating activity detection method is able to accurately distinguish eating activity from the six other daily activities; (2) The *accuracy* outperforms *precision*,

recall and *F-score* in all these cases. Through comparing equations 3.6, 3.7 and 3.8, we know that the *accuracy* metric not only takes *TP* and *FP* into consideration, but also *TN* and *FN*. High *accuracy* indicates that *TN* is much larger than *FN* and *FP*. That is to say, all the models can identify most negative samples; (3) For DT, NN, MLP and SVM, their *precision* is better than *recall*. This is because that, according to equations 3.7 and 3.8, *FP* is smaller than *FN*. This implies that all these four models misclassify more positive samples as negative class than negative samples as positive class. In other words, these models are biased to negative class. Through assigning a larger weight to the positive class, WSVM reverses the bias and obtains higher *recall* than *precision*; (4) As to the five classifiers, DT, MLP and WSVM outperform NN and SVM. Specifically, DT performs best for window length of 256 and 512. MLP performs best for window length of 1024; (5) The classification performance improves with the increase of window length, but the improvement is only two to three percent. As longer window length causes larger time delay, in all the following experiments, the window length is set to 512 to balance the accuracy and delay.

Table 3.1: The *accuracy* of five classifiers in CVT

Win.	DT	NN	MLP	SVM	WSVM	Average
256	95.2%	91.9%	93.8%	93.5%	94.4%	93.8%
512	97.1%	91.7%	94.4%	93.7%	95.2%	94.4%
1024	96.2%	92.6%	96.5%	93.9%	95.7%	95.0%

Table 3.2: The *F-score* of five classifiers in CVT

Win.	DT	NN	MLP	SVM	WSVM	Average
256	89.1%	80.3%	85.5%	84.1%	87.8%	85.4%
512	93.6%	80.2%	87.2%	85.0%	89.8%	87.2%
1024	91.6%	82.6%	92.1%	85.6%	90.8%	88.5%

Self test (ST). Figure 3.9 depicts the ST results of these four users. For user 1 and user 2, the ST performances of the five classifiers on almost all these four metrics are highly accurate. Both the *accuracy* and *F-score* are higher than those of CVT. For user 3, the *F-score* of SVM is a little low. For user 4, while the *accuracy* of all these classifiers is

above 90%, the F -score of NN and SVM is lower than 80%. However, DT and MLP still perform quite good for user 4. The $accuracy$ of DT and MLP is larger than 95%, and the F -score of DT and MLP is above 90%. We believe that the performance difference between different classifiers is mainly because that we adopt the default parameters for all classifiers. The performance of these classifiers could be improved after parameter optimization.

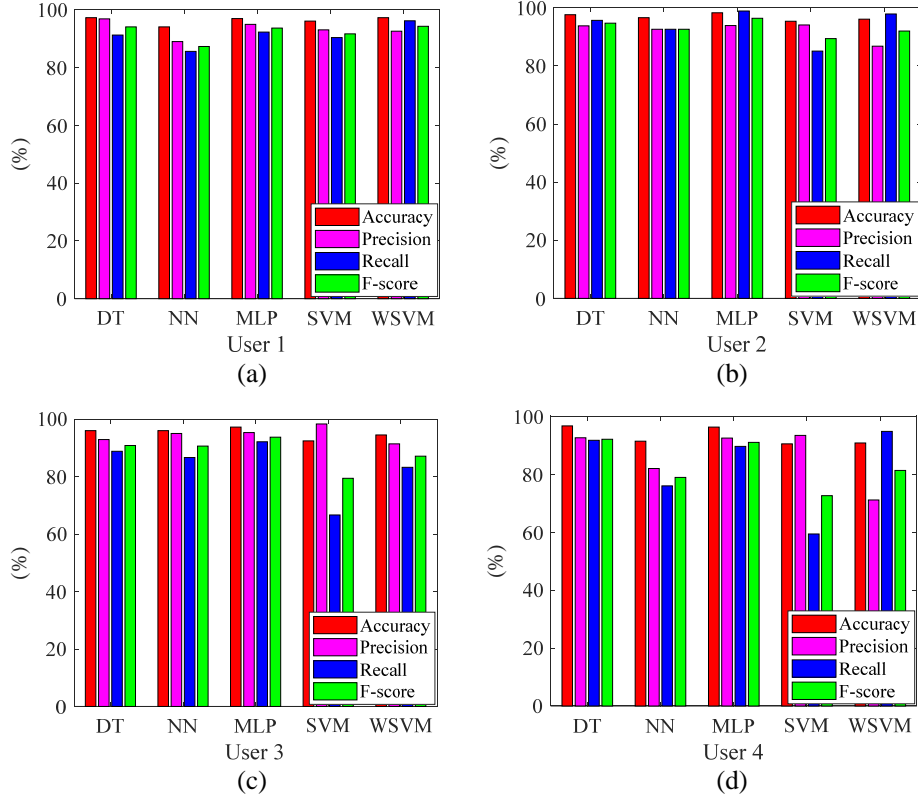


Figure 3.9: ST results of four users

Leave-one-subject-out test (LOSOT). Figure 3.10 presents the LOSOT results of four users. Comparatively, the performance of LOSOT falls below that of CVT and ST. This is reasonable as the data of testing user is not included in the training dataset. Table 3.3 shows the average $accuracy$ and F -score of the five classifiers for these four users. For user 1, user 2 and user 3, the average $accuracy$ is between 89.8% and 93.4%, and the average F -score is between 76.6% and 85.1%. These are still good. For user

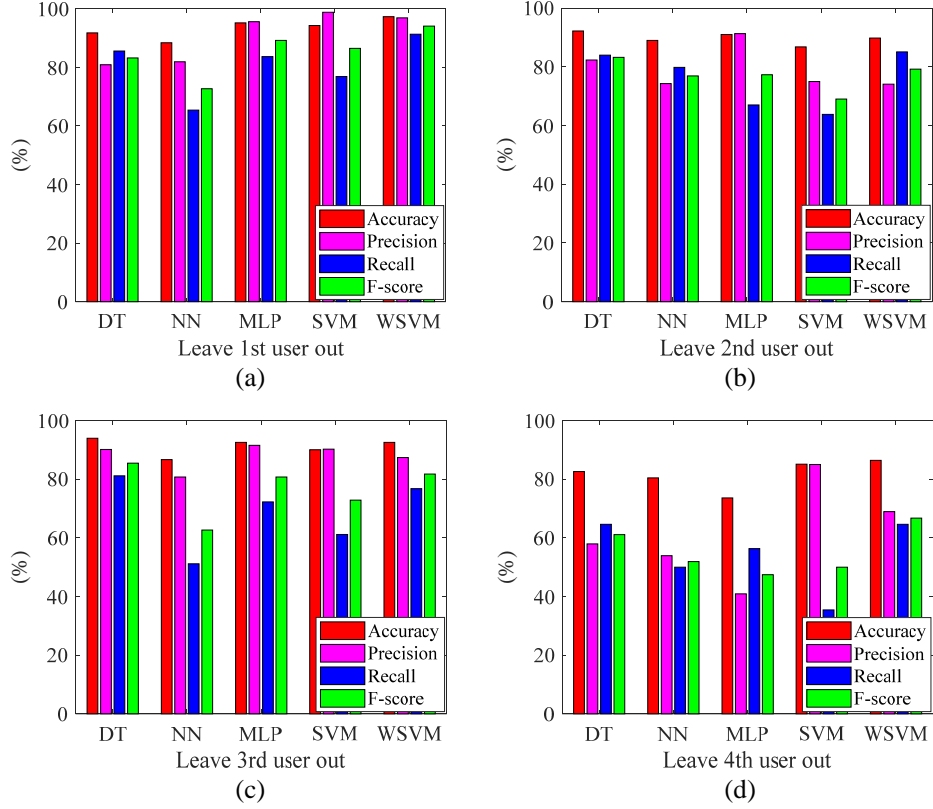


Figure 3.10: LOSOT results of four users

4, the average *accuracy* is 81.6%, but the average *F-score* is only 55.4%. Why is the detection performance of user 4 lower than that of the other three users? We believe that it is because of sensor position bias, which causes larger variance of the sampled acceleration data for user 4.

Table 3.3: The average *accuracy* and *F-score* in LOSOT

	User 1	User 2	User 3	User 4
Average <i>accuracy</i>	93.4%	89.8%	91.1%	81.6%
Average <i>F-score</i>	85.1%	77.1%	76.6%	55.4%

3.4.3 Evaluation of Chews Counting

We evaluate the chews counting accuracy for each user in the following experiments. For non-chewing window elimination, the difference threshold, *MagDiff*, is set to 3. To describe the chews counting accuracy, the following detection error rate is used:

$$Error\ rate = \frac{|Detection - Ground\ truth|}{Ground\ truth} \times 100\% \quad (3.10)$$

Table 3.4 describes the ground truth in terms of the number of chews, chewing time and chewing frequency of these four users. Compared with the chewing frequency range reported in reference [70], the chewing frequencies in Table 3.4 are a little low. This is because we do not exclude the time spent in biting and swallowing.

Table 3.5 depicts the chews counting results for four users. The error rates are 9.9%, 21.8%, 4.0% and 13.2%. The average error rate of four users is about 12.2%. Comparatively, the chews counting performance on user 2 is worse. After examining the Z axis data of these four users, we think the reason may be that the Z axis data of user 2 is a little more noisy than the Z axis data of other three users.

Table 3.4: Ground truth of chewing of four users

	User 1	User 2	User 3	User 4
Number of chews	473	596	323	380
Chewing time (Sec.)	532	481	492	461
Chewing frequency	0.9 Hz	1.2 Hz	0.7 Hz	0.8 Hz

Table 3.5: Chews counting results of four users

	User 1	User 2	User 3	User 4
Chews counting	520	466	310	330
Error rate	9.9%	21.8%	4.0%	13.2%

In Step 2 of chews counting module, we drop the whole window if the acceleration magnitude difference is larger than the predefined threshold. The dropped window may contain a few chews. Therefore, in most cases, the chews counting results are underestimated, as we can see from the results of user 2, 3 and 4. The best solution is to design a segmentation algorithm to extract whole chewing segments for chews counting.

For the threshold of acceleration magnitude difference, we use a fixed value for all the users. Considering the user difference and sensor location variance, a user-dependent and online adaptive threshold should be better.

3.5 Discussion and Future Work

We chose multiple pieces of watermelon as the food in all evaluation experiments. Watermelon is one of the softest foods. Thus, the user chews it with little effort. According to commonsense, the more strength the mastication muscles use, the greater the muscle bulge is. Therefore, if eating watermelon can be accurately recognized by the proposed method, it is reasonable to expect that eating harder food could also be accurately recognized. However, for some foods such as soups and drinks, there is no need to chew. Eating these foods is more like drinking activity. Thus, the proposed method is mainly for detecting eating solid foods.

In our experiment, all collected data is wirelessly transmitted to a laptop for offline analysis. This experimental setup is energy-consuming and only for performance evaluation and validation of the proposed method. For real-world application, we plan to embed the eating detection model and chews counting model into the headband for online analysis.

The experimental data for different activities is collected separately. However, in real life scenarios, a person may conduct two or more activities simultaneously, such as eating while walking, eating while talking, etc. Then, the collected data may be dominated or contaminated by these noisy activities. In these scenarios, the performance of the proposed method may be degraded. One possible solution is to separate the signal from the data for each individual activity before eating activity detection and chews counting. We leave this for future work.

For the ground truth in terms of the number of chews, the EMG data is used as the reference to identify and count each chew. The EMG signal during food intake is ignored. At the same time, as shown in the green boxes in Figure 3.11, occasionally the EMG signal may be less obvious and hard to identify. Thus, the ground truth may not be perfectly obtained, but the error is very small according to our observation.

For the feature extraction, when the window length is 256, there is only one frequency component in the band of (0, 0.5) Hz. Thus, the MFC location feature in this band always

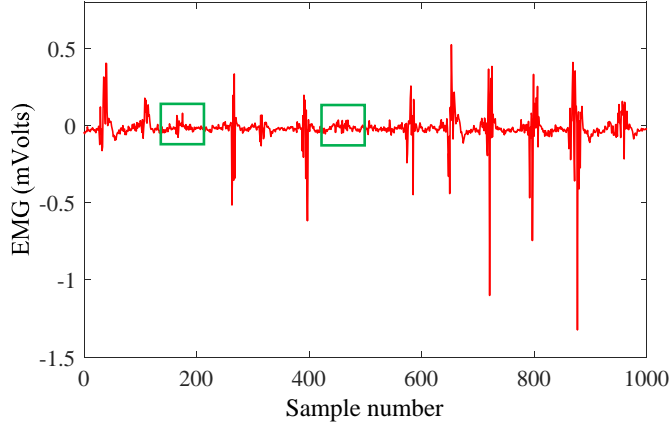


Figure 3.11: Less obvious EMG signal

equals 1, which is useless for classification. We delete this feature in the CVT experiment for window length of 256.

The proposed eating detection method is complementary to other methods for recognizing more complex dietary activities. We also leave this for future work.

3.6 Conclusion

In this chapter, we propose a novel eating activity detection and chews counting method. It is done through identifying the mastication muscle contractions using a triaxial accelerometer attached on the temporalis. The accelerometer is embedded in a headband, and only the acceleration data is recorded. Therefore, the proposed method is less invasive and privacy-preserving. Experiments are conducted with multiple human subjects. The results demonstrate that the proposed method accurately distinguishes eating activity from other daily activities using only 5 seconds of acceleration data. Moreover, the average error rate of chews counting for four users is 12.2%.

Chapter 4

Inferring Food Types through Sensing and Characterizing Mastication Dynamics

4.1 Introduction

As one of the major causes of chronic diseases, unhealthy dietary structure leads to obesity, diabetes, and heart disease. Automatic food type recognition acts as a core function to monitor the dietary structure. It provides medical professionals objective information to understand patients' nutritional contents, estimate carbohydrate intake amounts, choose the best therapies, and provide timely feedback to the patients.

To recognize food types continuously and conveniently in daily living, some wearable sensor-based methods have been proposed in recent years. These methods take advantage of microphone, EMG, and piezoelectric sensors embedded in the wearable devices. The microphone-based method deploys a microphone sensor in the outer ear [15] or at the throat area [16]. This method is easily impacted by ambient acoustic noise. In addition, the earphone or headphone may block the ear canal and affect daily communication. Plus, deploying a sensor at the throat area is intrusive and uncomfortable. The EMG and piezoelectric sensors need to be tightly adhered to the skin. They are obviously

intrusive and not suitable for long-time wear.

How to recognize food types accurately and less intrusively using wearable sensors? To answer this question, we investigate the food properties and mastication dynamics. We are inspired by the following observations: 1) Food properties and mastication dynamics are highly correlated. Each type of food has its own intrinsic properties [23], such as hardness, elasticity, fracturability, adhesiveness, and size. Because the masticatory system is highly adapted to the food properties, the difference in food properties leads to the variance of corresponding mastication dynamics; 2) Mastication dynamics can be sensed by deploying a motion sensor on a mastication muscle. The contraction of a mastication muscle changes the shape of the muscle spindle to make it shorter and thicker. In addition, the muscle contractions are synchronized with the mandible movements. Therefore, the motion sensor can sense mastication dynamics through detecting the muscle contractions and deformations.

Based on these observations, we are motivated to deploy motion sensors on the mastication muscles to sense mastication dynamics and infer food types accordingly. However, this raises three research questions. *First, how to represent food properties using corresponding mastication dynamics? Second, how to deploy motion sensors on the mastication muscles to sense mastication dynamics accurately and less intrusively? Third, how to characterize the mastication dynamics using motion sensor data?* To answer these three research questions, 1) we define six mastication dynamics parameters to represent the food properties. They are chewing speed, the number of chews, chewing time, chewing force, chewing cycle duration, and skull vibration; 2) we embed motion sensors in a headband and deploy the sensors on the temporalis muscles to sense mastication dynamics accurately and less intrusively; 3) 37 hand-crafted features are extracted from each chewing sequence to explicitly characterize the mastication dynamics. Based on the extracted features, we train a multi-class classification model to recognize the food types. The motivation of our proposed food type recognition method is shown in Fig. 4.1. To evaluate the performance of our proposed food type recognition method, we collect

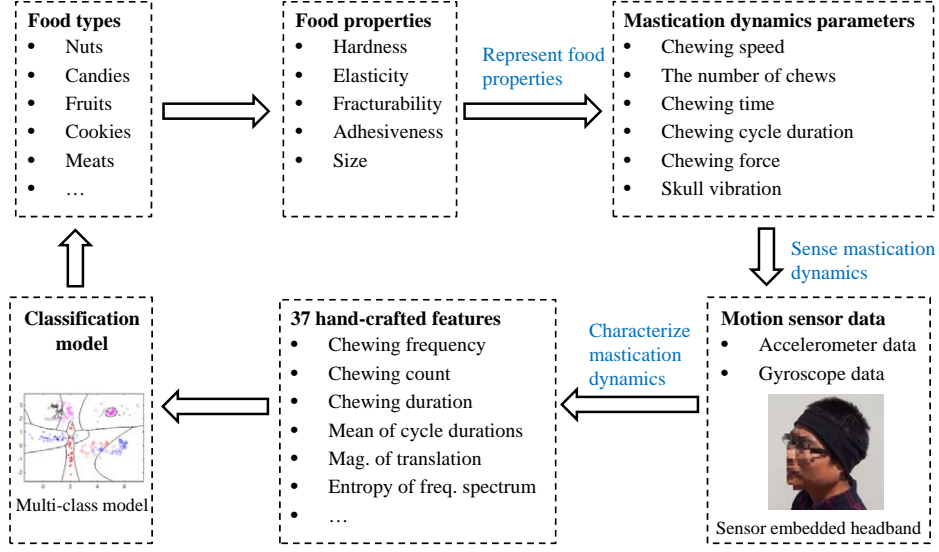


Figure 4.1: Motivation of our proposed food type recognition method

a real-world dataset of 15 human subjects for 11 food categories (20 types of food in total). The experimental results are very promising.

The main contributions are as follows:

- We propose to infer food types through sensing mastication dynamics with wearable motion sensors. To our best knowledge, this is the first effort in using wearable motion sensors to sense mastication dynamics and recognize food types accordingly.
- We propose to embed motion sensors in a headband and deploy them on the temporalis muscles to sense mastication dynamics accurately and less intrusively.
- We extract 37 hand-crafted features from each chewing sequence to explicitly characterize the mastication dynamics.
- We evaluate the performance of our proposed method on a real-world dataset. Experimental results show that the average recognition accuracy of 15 human subjects is 74.3%. The recognition accuracy of a single human subject is up to 86.7%.

The remainder of this chapter is organized as follows: Section 4.2 introduces how to represent food properties with mastication dynamics. Section 4.3 describes how to sense

mastication dynamics with motion sensors. Section 4.4 introduces how to characterize mastication dynamics with hand-crafted features. Section 4.5 introduces food type classification model. Experiment and evaluation are presented in Section 4.6. Section 4.7 presents the comparison with existing wearable sensor-based methods. Section 4.8 presents discussion and future work. Finally, the conclusion is drawn in Section 4.9.

4.2 Representing Food Properties with Mastication Dynamics

In this section, we first introduce the food properties. Then, we describe how to represent the food properties with mastication dynamics.

4.2.1 Food Properties

A food property is defined as “a particular measure of the food’s behavior as a matter, or its behavior with respect to energy, or its interaction with the human senses, or its efficacy in promoting human health and well-being [79].” According to the aspects they describe, the food properties are classified into different categories, such as textural properties, tactile properties, appearance properties, rheology properties, acoustic properties, and flavour properties.

Different types of food have different food properties. Thus, the food properties are capable of identifying the food types. The image-based food type recognition method takes advantage of the appearance properties (color, texture, shape, etc.) to distinguish the food types [35]. The microphone-based method utilizes the acoustic properties during chewing or swallowing to recognize the food types [15, 16]. Of all the food properties, some are related to the mastication [80, 23], such as hardness, elasticity, fracturability, adhesiveness, and size. Hardness indicates the force required to break/chew the product [81]; elasticity describes the ability to deform and go back to its origin state [82]; fracturability describes the ability to break food into pieces when it is bitten [83]; ad-

hesiveness indicates the ability of food to adhere to the teeth when chewed [83]; size indicates the length, width and height of the food samples.

4.2.2 Food Property Representation

These mastication related food properties are highly correlated to the corresponding mastication dynamics. Different food properties provide different stimulus to the masticatory system and lead to the variance of corresponding mastication dynamics. For example, the chewing speed of the soft food is higher than that of the hard food [84]. Therefore, the mastication dynamics are able to represent the food properties and infer the food types accordingly.

The mastication dynamics mainly include two aspects, the mastication muscle activities and the mandible motions [80]. To represent the muscle activities and mandible motions, some parameters are extracted from a single chew, a specific stage (e.g. the early, middle, or late chewing stage), or a whole chewing sequence [85]. The muscle activities related parameters include chewing speed, the number of chews, chewing time, and chewing force. The mandible motions related parameters include time, amplitude, and velocity of opening or closing the mouth.

We specifically define six mastication dynamics parameters to represent the mastication related food properties. We observe that normally the food type between two neighboring bites does not change. Thus, we extract mastication dynamics parameters from a whole chewing sequence to represent these food properties. Compared with the parameters extracted from a single chew or a specific stage, the parameters extracted from a whole chewing sequence are more robust and complete. These six mastication dynamics parameters are chewing speed, the number of chews, chewing time, chewing force, chewing cycle duration, and skull vibration. The first four parameters are used to represent the muscle activities. The chewing cycle duration indirectly indicates the amplitude and velocity of mandibular movement. The skull vibration is not included in the existing study of mastication dynamics. However, it is very useful to characterize

the vibrations of skull bone during mastication.

Table 4.1 shows the food properties represented by each mastication dynamics parameter [80, 23]. We see that each mastication dynamics parameter represents several food properties. Through combining all these parameters, it is highly possible to distinguish different food types.

Table 4.1: Food properties represented by each mastication dynamics parameter

Parameter	Food properties
Chewing speed	Hardness, elasticity, adhesiveness
The number of chews	Hardness, fracturability, adhesiveness, size
Chewing time	Hardness, fracturability, adhesiveness, size
Chewing force	Hardness, fracturability, adhesiveness
Chewing cycle duration	Hardness, elasticity, adhesiveness, size
Skull vibration	Hardness, fracturability

4.3 Sensing Mastication Dynamics

In this section, we first describe why we choose motion sensors to sense mastication dynamics. Then, we introduce the sensor deployment on the mastication muscles. Finally, we introduce the motion data collection.

4.3.1 Why Motion Sensors?

Existing works often use EMG sensor and 3D kinematics method to sense the mastication muscle activities and the mandible motions, respectively. The EMG sensor is utilized to record the electrical signals generated by the mastication muscles during contraction [80, 23]. However, the EMG electrodes are required to be adhered on skin tightly, which is intrusive and uncomfortable. The 3D kinematics method [86] deploys several markers on the head, mandible and reference plane. The infrared video cameras are used to record the markers' movements. Then, the 3D coordinates of the mandible are extracted to calculate the parameters of mandible motions. This method is obviously intrusive and only used in clinical study.

Mastication dynamics can be sensed by deploying a motion sensor on a mastication

muscle. One observation is that the mastication muscle contraction changes the shape of the muscle spindle to make it shorter and thicker. Accordingly, the mastication muscle bulges to some degree; the greater the chewing force, the larger the muscle bulge. In addition, the muscle contractions are synchronized with the mandible movements. Therefore, through deploying a motion sensor on a mastication muscle, the sensor is able to directly sense the mastication muscle activities. Plus, it is also capable of inferring the mandible motions. For example, the chewing cycle duration indirectly indicates the amplitude and velocity of the mandible movement. Therefore, the mastication dynamics are sensed through detecting the muscle contractions and deformations. Moreover, the motion sensor can also catch the skull vibration during mastication. Different from the EMG sensor or 3D kinematics method, the motion sensor needs no skin contact and is easily embedded into a headband or hat. Thus, it is less intrusive and more comfortable to wear.

4.3.2 Sensor Deployment

The temporalis is the best mastication muscle to deploy motion sensors. As discussed in subsection 3.2.1, the masseter, medial pterygoid and lateral pterygoid are in the face area and near mouth cavity, where is inconvenient and intrusive for sensor deployment. The temporalis is located at each side of the skull and in front of the ear [22], where people often wear a headband or hat. This motivates us to embed motion sensors into a headband and deploy the sensors on the temporalis muscles.

A subject may prefer to chew the food mainly on the left side or the right side. Therefore, the mastication muscles and corresponding mastication dynamics of these two sides are not symmetric. To accurately sense the mastication dynamics, two small-size hardware platforms [2] shown in Fig. 4.2 (a) are deployed on the left and right temporalis muscles, respectively. Each device contains a 3-axis accelerometer, a 3-axis gyroscope, and a 3-axis digital compass. Only the accelerometer and gyroscope are used in our proposed method.

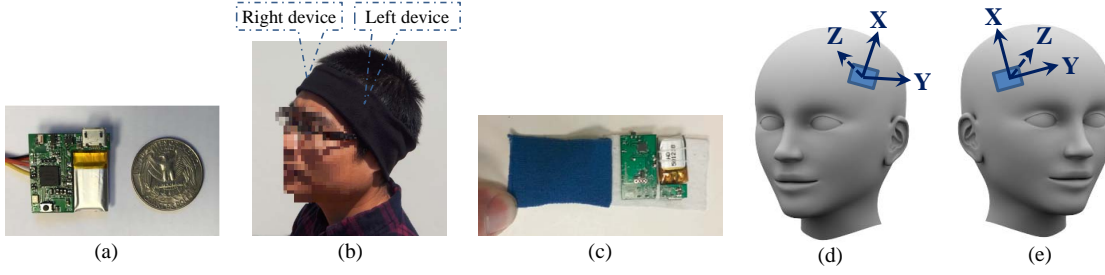


Figure 4.2: Wearable device. (a) Hardware platform [2]; (b) Device deployment; (c) Device covered by two polyester tapes; (d) Sensor orientations of the left device; (e) Sensor orientations of the right device

A headband [87] is utilized as the host object. It is made by polyester and spandex materials and is comfortable to wear. We open the headband at the upper side and insert two wearable devices in it. Fig. 4.2 (b) shows the deployment of these two devices on the left and right temporalis muscles, respectively. Each device is covered by two polyester tapes to protect its components and prevent the location change after the deployment, as shown in Fig. 4.2 (c). Fig. 4.2 (d) and (e) indicate the sensor orientations of the left and right devices (from the subject’s perspective), respectively. The X axes of these two devices point upward. The Y axis of the left device points backward, and the Y axis of the right device points forward. The Z axes of these two devices are vertical to their corresponding X - Y planes and point inside of the skull.

4.3.3 Motion Data Collection

Fifteen human subjects were recruited for data collection. Each of them sits in front of a table and is served with different types of food. The food is cut into pieces (if necessary) and put on a paper plate. They eat the food one piece at a time using a spoon. For some types of food that are inconvenient to scoop, they feed themselves with their hands but act like using spoons. While a subject eats the served food, the accelerometer and gyroscope on each device sample simultaneously. The sampling rate is 100 Hz. The sampled data on these two devices are transmitted to a mobile phone through Bluetooth Low Energy (BLE) in real time. On the mobile phone, a software is developed to receive

all the data and store them on the local storage. The data of each food type are stored in a separate file. Then, all the data files are transferred to a PC for offline analysis.

4.4 Characterizing Mastication Dynamics

In this section, we first describe motion data preprocessing. Then, we introduce the extraction of the 37 hand-crafted features.

4.4.1 Motion Data Preprocessing

The motion data preprocessing includes two parts, the sensor data calibration and data segmentation. We introduce them separately as follows.

Sensor Data Calibration. The motion sensor data may not be accurate because of the scaling and bias errors. To eliminate them, we calibrate the accelerometer data and the gyroscope data separately for each device.

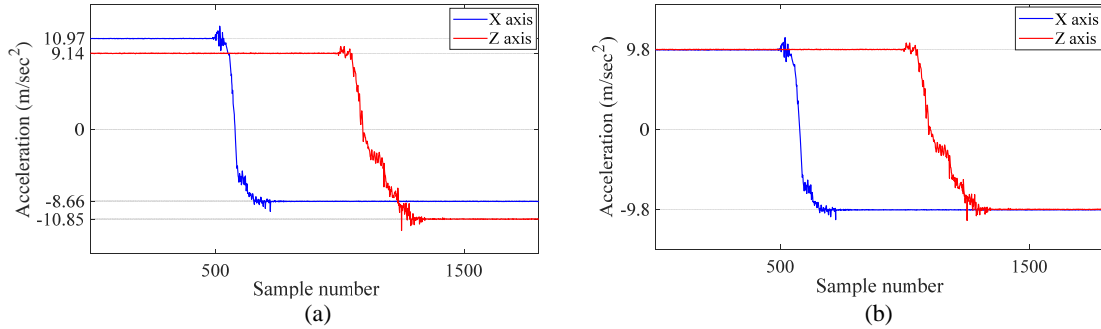


Figure 4.3: Accelerometer data of the X axis and Z axis. (a) Before calibration; (b) After calibration

Fig. 4.3 (a) shows an example of the raw accelerometer data of the X axis and Z axis. For the X axis, the raw sensor data are 10.97 and -8.66 when it points to the positive and negative directions, respectively. The acceleration difference is 19.63, which is about $2g$. The g is the gravitational acceleration, which is set to 9.8 m/sec^2 . Although their scaling error is not obvious, the raw sensor data are clearly biased toward the positive direction. For the Z axis, the raw sensor data are 9.14 and -10.85 when it points to the

positive and negative directions, respectively. The acceleration difference is 19.99, which is larger than $2g$. The raw sensor data of the Z axis have not only the scaling error but also the bias error.

We calibrate accelerometer data of each axis in the following way [88]. First, we assume the raw sensor data of the X axis is m_X and the true acceleration output of the X axis is a_X . They have the following linear relationship [89]:

$$a_X = k_X \cdot m_X + b_X. \quad (4.1)$$

where k_X is the scaling factor of the X axis. It scales the acceleration difference between the positive and negative directions to $2g$. k_X is defined as

$$k_X = \frac{2 \cdot g}{m_X^+ - m_X^-}, \quad (4.2)$$

where m_X^+ and m_X^- are the raw sensor data of the X axis when it points to the positive and negative directions, respectively. b_X is the bias of the X axis, and is defined as

$$b_X = -\frac{k_X \cdot (m_X^+ + m_X^-)}{2}. \quad (4.3)$$

In the same way, the raw data of the Y and Z axes are calibrated. The calibrated accelerometer data of the X and Z axes in Fig. 4.3 (a) are shown in Fig. 4.3 (b). We see that these two errors are eliminated.

For the gyroscope data, the bias of each axis is measured and subtracted from the raw sensor data. The scaling error of each axis is very small. Thus, we ignore it for simplicity.

Data Segmentation. We segment the continuous sensor data and extract the chewing sequences through analyzing the subjects' head motions during biting. As we introduced in Section 4.2, the mastication dynamics parameters are extracted from a whole chewing sequence. As chewing normally happens between two neighboring bites, the chewing sequences can be segmented by detecting biting or chewing actions [90, 32]. Here, a general

biting or chewing detection method is not our focus. Our main goal is to characterize the mastication dynamics through manually extracting features from a segmented chewing sequence. As our dataset is collected mainly by using a spoon, the subjects bow their heads before biting and raise their heads after biting. Therefore, we simply segment the sensor data by analyzing the subjects' head motions during biting.

The gyroscope data of the Z axis are appropriate for head motion analysis. According to the sensor orientations in Fig. 4.2 (d) and (e), when a subject bows or raises his/her head, the two wearable devices rotate around their own Z axes in opposite directions. We randomly choose the Z axis on the left device for head motion analysis. Fig. 4.4 shows the gyroscope data of eating 10 pieces of food. When the subject bows his/her head, the device rotates clockwise around the Z axis. The gyroscope data are negative and form a valley. When the subject raises his/her head, the device rotates counter-clockwise. The gyroscope data are positive and form a peak.

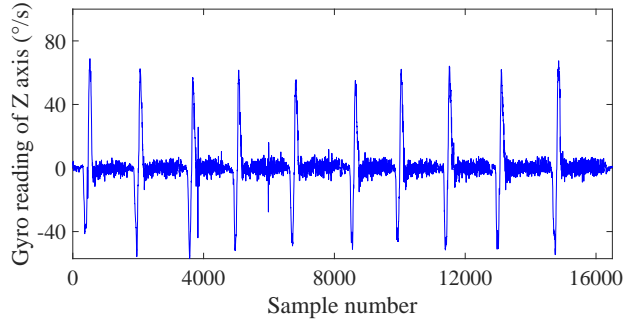


Figure 4.4: Gyroscope data of the Z axis during eating

We propose three metrics ($PeakArea$, $ValleyArea$, and $Distance$ between two neighboring zero-crossing points) to analyze head motions, as shown in Fig. 4.5. $PeakArea$ is the accumulation of the gyroscope data in a positive peak. It represents the degree of raising the head. Similarly, $ValleyArea$ is the accumulation of the gyroscope data in a negative valley. It represents the degree of bowing the head. The third metric is based on one observation, i.e. the distance between two neighboring zero-crossing points before and after one peak or valley is relatively larger than the distance between two neighboring zero-crossing points during chewing.

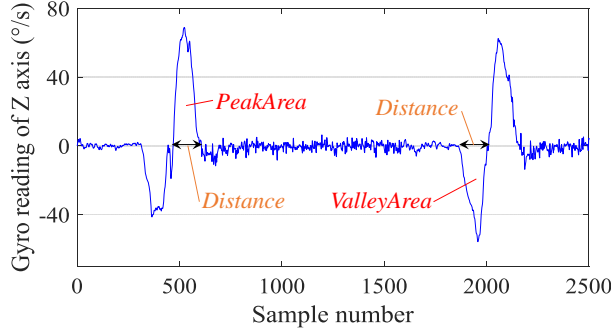


Figure 4.5: Three metrics for head motion analysis

With above three metrics, we segment the sensor data as following. We first filter the gyroscope data using a moving average filter of span s . Then, all the zero-crossing points are detected. If the *Distance* between two neighboring zero-crossing points is larger than $DisThres$, and the *PeakArea* between them is larger than $PeakAreaThres$, the data segment between these two zero-crossing points is identified as raising head. Similarly, if the *Distance* between two neighboring zero-crossing points is larger than $DisThres$, and the *ValleyArea* between them is smaller than $ValleyAreaThres$, this data segment is identified as bowing head. The data segment between the end point of raising head and the start point of following bowing head is taken as a whole chewing sequence. If a chewing sequence is incomplete and shorter than len , it is dropped.

4.4.2 Feature Extraction

From each chewing sequence, we extract 37 hand-crafted features to characterize the mastication dynamics. These features are divided into two sets: 1) chewing cycles dependent features. These features characterize the chewing speed, the number of chews, chewing time, and chewing cycle duration parameters; 2) chewing cycles independent features. These features characterize the chewing force and skull vibration parameters.

Chewing Cycles Dependent Features. We propose a metric, R_{MFC} , to select one sensor whose data is most regular and obvious to extract the chewing cycles dependent features. We observe that if the data of one sensor is more regular and obvious than the

data of other sensors, its energy should be more concentrated on a small range of frequencies. Accordingly, we define R_{MFC} as the ratio of the maximum frequency component (MFC) in the chewing frequency range to the sum of all the frequency components. The sensor whose data has the largest R_{MFC} is selected. The details and justification of this sensor selection method is introduced in Appendix 4.A. From the selected sensor data, we extract the chewing cycles dependent features as follows:

Chewing speed. One feature, chewing frequency, is extracted to characterize the chewing speed parameter. To compute this feature, we first filter the selected sensor data using a 9^{th} -order one-dimension median filter [76] to reduce the noise. Then, we conduct Fourier transform on the filtered sensor data. The frequency corresponding to the MFC in the chewing frequency range, f_{MFC} , is taken as the chewing frequency feature. Here, we define the chewing frequency range as $[0.5, 2.5]$ Hz [90].

The number of chews. One feature, chewing count, is extracted to characterize the number of chews. Because each peak in the data of a chewing sequence corresponds to one muscle contraction, we estimate the chewing count through counting the number of peaks in the selected sensor data. One problem is that some noise spikes may be falsely counted as true peaks and exaggerate the counting results. To solve this problem, we first filter the selected sensor data using the 9^{th} -order median filter. Then, we propose a distance threshold and an amplitude threshold to further eliminate the noise spikes. For the distance threshold, we refer to the method in [91]. It requires that the distance between two neighboring peaks is no less than $\frac{3}{4}$ cycle length. The cycle length is calculated based on the f_{MFC} and the sampling rate, Rs . The distance threshold between two neighboring peaks, D_{P2P} , is defined as

$$D_{P2P} = \frac{3}{4} \cdot \frac{Rs}{f_{MFC}}. \quad (4.4)$$

The second threshold, P_{amp} , is for the peak amplitude. If the amplitude of a peak is less than this threshold, this peak is ignored. The details of the proposed peak detection method are described as follows:

- Step 1: Zero-crossing point detection. We scan the filtered sensor data sequentially and find out all zero-crossing points.
- Step 2: Peak detection. One peak with the maximum amplitude is detected between any two neighboring zero-crossing points.
- Step 3: False peak elimination. We scan all the detected peaks from beginning to end. If the amplitude of one peak is less than P_{amp} or the distance between it and its prior peak is less than D_{P2P} , this peak is dropped.

The number of remaining peaks is taken as the chewing count.

Chewing time. Two features are extracted to characterize the chewing time parameter: chewing duration and sequence length. We define the chewing duration feature as the distance between the first peak and the last peak, L . However, the chewing duration does not include the time spent on swallowing, which is also useful to distinguish the food types. Therefore, we extract segment length feature to include the swallowing time. It is defined as the distance between the first point and the last point of the selected sensor data.

Chewing cycle duration. Three features are extracted to characterize the parameter of chewing cycle duration. Suppose n peaks are detected in the selected sensor data. Their indices are $[p_1, \dots, p_n]$. The cycle durations are expressed as $[t_{1,2}, \dots, t_{n-1,n}]$, where $t_{i,i+1} = p_{i+1} - p_i$ is the duration between the i^{th} chew and the $i+1^{th}$ chew. Then, the maximum, mean, and standard deviation of the cycle durations are extracted as features.

Chewing Cycles Independent Features. In the following, we introduce how to extract features to characterize the chewing force and skull vibration parameters.

Chewing force. Two features are extracted to characterize the chewing force parameter: the magnitude of translation and the magnitude of rotation. Chewing force indicates the contraction intensity of the mastication muscles. The contraction of the temporalis generates both translation and rotation movements of the wearable devices. The greater

the mastication force, the larger the translation and rotation. Accordingly, we propose to use the accelerometer and gyroscope data to quantify the translation and rotation, respectively. To compute the magnitude of translation, we first filter the accelerometer data of each axis using the 9th-order median filter. According to the peak detection results, we extract filtered 3-axis accelerometer data from the first peak to the last peak. Then, the i^{th} filtered data of these three axes, a_X^i , a_Y^i and a_Z^i , are composed into one scalar acceleration a_i using Equation 3.1. Here, $i = 1, 2, \dots, L$.

The accumulative acceleration, S_a , is defined as $S_a = \sum_{i=1}^L a_i$. Because a subject may chew the food on either the left side or the right side, we calculate the accumulative accelerations for the left accelerometer and right accelerometer separately, which are expressed as S_a^{Left} and S_a^{Right} . Then, the magnitude of translation, T_{Mag} , is formulated as:

$$T_{Mag} = \frac{S_a^{Left} + S_a^{Right}}{L}, \quad (4.5)$$

Similarly, from the 3-axis gyroscope data, we calculate the magnitude of rotation, R_{Mag} , using the same method above.

Skull vibration. We extract 14 features from each accelerometer to characterize the skull vibration parameter. These features are calculated from the raw accelerometer data between the first and last peaks. First, using Eq. 3.1, the sensor readings of the three axes are composed into scalar accelerations, which are not sensitive to the sensor orientations. From the composed acceleration data, we calculate the number of mean-crossing, i.e. the times of the data goes across its mean. We take it as the first skull vibration feature. Then, the Fourier transform is conducted on the composed acceleration data to compute its single-sided amplitude spectrum [69] (without direct current component). From the single-sided amplitude spectrum, the MFC, entropy [92], and energy are calculated as the second, third, and fourth skull vibration features. The energy is defined as the sum of squared spectrum components. Finally, we partition all the spectrum components into 10 bins, i.e. (0, 5] Hz, (5, 10] Hz, ..., (45, 50] Hz, according

to their corresponding frequencies. The spectrum components in each bin are summed together as one feature. These 10 features are used to represent the energy distribution of skull vibration at different frequency intervals. Altogether 28 features are extracted from the left and right accelerometers.

In total, 37 hand-crafted features are extracted from each chewing sequence, as shown in Table 4.2. The first column is the feature number. The second column shows the feature name. The third column shows the mastication dynamics parameters characterized by these features. The last column indicates the data source where each feature is extracted.

Table 4.2: The 37 hand-crafted features for mastication dynamics characterization

No.	Feature name	Characterized parameter	The data source
1	Chewing frequency	Chewing speed	Selected gyroscope data
2	Chewing count	The number of chews	
3	Chewing duration	Chewing time	
4	Sequence length	Chewing cycle duration	
5	Maximum of cycle durations		
6	Mean of cycle durations		
7	Std of cycle durations		
8	Magnitude of translation	Chewing force	Two accelerometers
9	Magnitude of rotation	Chewing force	Two gyroscopes
10	Number of mean-crossing	Skull vibration	The left accelerometer
11	Entropy of frequency spectrum		
12	Energy of frequency spectrum		
13	Maximum frequency component		
14-23	Sum of spectrum comp. in 10 bins		
24	Number of mean-crossing	Skull vibration	The right accelerometer
25	Entropy of frequency spectrum		
26	Energy of frequency spectrum		
27	Maximum frequency component		
28-37	Sum of spectrum comp. in 10 bins		

4.5 Food Type Classification

Existing works provide strong evidence to support our proposed food type recognition method. Using well controlled food stimuli and strict criteria, existing research on mastication [80, 93] demonstrated the stability of intra-individual mastication dynamics. The experimental results clearly showed that “there are no significant differences between the values of the masticatory parameters for a given individual who is asked to chew the same food several times [80]”. This conclusion indicates that our proposed mastication

dynamics-based food type recognition method is valid.

Mastication dynamics-based food type recognition needs a personalized classification model for each subject. These existing works [80, 93] also examined the inter-individual variation of mastication dynamics. All the masticatory parameters demonstrated a large variation between individuals [80]. For example, an experiment [93] selected 15 young male subjects to chew four food products. The results showed that the parameters of mandible motions and muscle activities varied up to 3-fold among these subjects [93]. Therefore, we train a personalized food type classification model for each subject.

We choose MLP as the multi-class classifier. This is because the mapping function from the hand-crafted features to the food types is implicit and highly likely nonlinear. MLP is a feedforward artificial neural network model. It specializes in modeling nonlinear mapping from the input neurons to the output neurons [73]. It learns the connection weights between neurons using backpropagation technique. The learned network is a very good approximation of the mapping function from these features to the food types.

4.6 Experiment and Evaluation

In this section, we first introduce the experiment setup, followed by the classification dataset. Then, we evaluate the performance of our proposed method with 11 food categories and 20 food types. Next, we introduce the feature importance analysis. Finally, we present the recognition performance on either the left side or the right side.

4.6.1 Experiment Setup

With the approval from the institutional review board (IRB), ten male users and five female users were recruited to collect the experimental data. Their demographic information is shown in Table 4.3, including age, gender, weight, head circumference, dominant feeding hand, and whether they wear glasses or not. The sensor deployment and data collection method are the same to those in Section 4.3.

Twenty types of food are selected according to the following three criteria: 1) they

Table 4.3: The users' demographic information and data collection date

User	Age	Gender	Weight (lbs)	Head cir. (cm)	Feeding hand	With glasses	Data collection date (1 st part; 2 nd part)
User 1	39	Male	200	58	Right	No	03/30/2018; 04/13/2018
User 2	34	Male	134	56	Right	No	03/30/2018; 10/01/2018
User 3	31	Male	145	58	Right	No	03/30/2018; 04/13/2018
User 4	29	Male	132	56	Right	Yes	04/02/2018; 04/13/2018
User 5	28	Male	138	58	Right	Yes	04/02/2018; 04/13/2018
User 6	32	Male	120	54	Right	Yes	04/03/2018; 04/13/2018
User 7	29	Male	170	58	Right	Yes	10/03/2018; 10/23/2018
User 8	28	Male	150	58	Right	Yes	10/19/2018; 10/20/2018
User 9	41	Male	159	58	Right	No	10/19/2018; 10/20/2018
User 10	24	Male	165	58	Right	Yes	10/22/2018; 10/23/2018
User 11	24	Female	100	58	Right	Yes	04/14/2019; 04/12/2019
User 12	41	Female	128	56	Right	Yes	04/13/2019; 04/12/2019
User 13	27	Female	126	59	Right	Yes	04/14/2019; 04/13/2019
User 14	37	Female	130	56	Right	Yes	04/15/2019; 04/16/2019
User 15	23	Female	110	57	Right	Yes	04/16/2019; 04/17/2019

Table 4.4: The food types in each food category

Food category	The food types included
1 Nuts	1 Almond; 2 Pecan; 3 Peanut
2 Gum Candy	4 Gummi Bear
3 Dry Fruit Slices	5 Dry Pineapple Tidbit; 6 Dry Mongo Slice; 7 Dry Banana Chip
4 Fresh Fruits	8 Blueberry; 9 Fresh Apple Slice; 10 Fresh Tomato; 11 Green Grape
5 Pretzel	12 Pretzel Stick
6 Corn and Fry	13 Popcorn; 14 Potato Chip; 15 Potato Fry
7 Cookie	16 Cookie
8 Vegetable	17 Carrot
9 Bread	18 Bread
10 Meat	19 Chicken Breast Nugget
11 Frozen Cream	20 Ice Cream

have different food properties (hardness, elasticity, fracturability, adhesiveness, and size); 2) they are commonly eaten food. All the food is bought from the Food Lion Grocery Store; 3) each type of food contains only one composition. We do not include food types that contain multiple compositions, such as the sandwich or hamburger. These food types are divided into 11 categories, as shown in Table 4.4.

According to the users' feedback in the preliminary experiment, it is very difficult for them to eat all the food at a time. Therefore, the data collection for each user is done on two different days, as shown in Table 4.3. For each male user, he eats the first 15 types of food on one day and the remaining five types of food on another day. Because the female users prefer a more balanced division of the food between two days, we adjust the data collection process for the female users accordingly. For each female user, she eats

the first 12 types of food on one day and the remaining eight types of food on another day.

For each of the first 19 types of food, we serve 10 pieces to each user. For the Ice Cream, we serve it in a bowl and ask users to take 10 bites (except that user 1 takes 12 bites). The pictures of the served food are shown in Fig. 4.6. For each type of food, the users chew half on the left side and the other half on the right side. The users eat one type of food at a time. After finishing one type of food, they may drink some water or have a rest until they feel comfortable to eat another type of food.

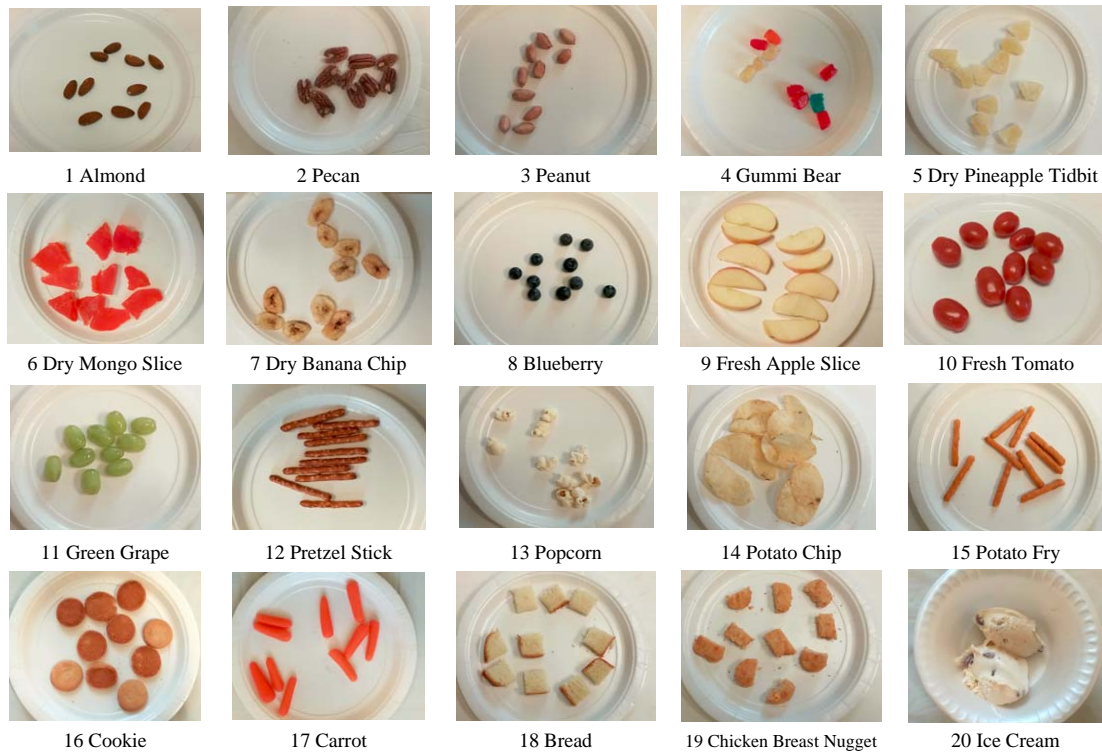


Figure 4.6: The 20 types of food served

A few users dislike some types of food. They are allowed to skip them. User 10 does not eat the Carrot. User 11 does not eat the Almond, Peanut and Gummi Bear. In addition, user 13 eats only five pieces of Dry Pineapple Tidbit, seven pieces of Dry Mango Slice, and eight pieces of Dry Banana Chip.

The data recording starts just before one user eats one type of food and stops just

after the user finishes it. Its corresponding food type is manually labeled.

4.6.2 Classification Dataset

The sensor data are segmented to extract the chewing sequences. For the data segmentation, the span of the moving average filter, s , is set to 31. The thresholds of three segmentation metrics ($DisThres$, $PeakAreaThres$, and $ValleyAreaThres$) are set to 75, 2000, -2000, respectively. The length threshold of the chewing sequence, len , is set to 300, which represents 3 seconds.

The last chewing sequence is dropped in the real application because there is no following biting action. However, in our lab experiment, the data collection stops just after the user finishes one type of food. Thus, the remaining data after the last bite is also a complete chewing sequence, and we include it in the following experiments.

Table 4.5 shows the number of detected chewing sequences. Of all the 2952 chewing sequences, 2742 are detected. The detection rate reaches 92.9%. Most missing sequences belong to the 8th (the Blueberry) and 20th (the Ice cream) food types. These two types of food need little chewing efforts. Thus, some chewing sequences are shorter than the configured length threshold, len , and hence are dropped. User 12 has 11 detected chewing sequences for the 5th, 9th and 11th food types. The reason may be the user occasionally bows her head to have a look at her smart phone. This head motion is falsely identified as biting. Thus, one chewing sequence is segmented into two incomplete sequences.

Table 4.5: Number of detected chewing sequences

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Subtotal
User 1	10	10	10	10	10	10	10	8	10	10	10	10	10	10	10	10	10	10	10	8	196/202
User 2	10	10	10	10	10	9	10	0	10	10	10	10	5	9	7	10	10	10	10	6	176/200
User 3	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	6	196/200
User 4	10	10	10	10	10	10	10	9	10	10	10	10	9	10	9	10	10	10	9	8	194/200
User 5	10	9	10	10	9	10	10	10	9	10	9	6	9	8	7	10	9	10	9	4	178/200
User 6	10	10	10	10	10	10	10	10	10	10	10	10	9	9	9	9	9	8	10	7	190/200
User 7	10	10	10	10	10	10	10	1	10	9	10	10	10	10	9	10	9	10	10	2	180/200
User 8	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	200/200
User 9	10	9	10	10	10	10	10	6	10	10	10	9	10	10	9	10	10	10	10	4	187/200
User 10	9	10	10	10	10	10	10	7	10	9	10	10	10	10	10	10	—	10	10	6	181/190
User 11	—	10	—	—	9	9	9	10	10	10	9	10	10	10	8	10	10	10	10	5	159/170
User 12	10	9	9	8	11	8	9	9	11	9	11	9	9	8	7	10	10	10	1	2	170/200
User 13	10	10	10	6	7	5	6	1	8	5	9	9	2	7	9	10	9	10	10	3	146/190
User 14	10	10	9	10	10	10	10	10	9	9	9	10	10	9	9	10	10	9	10	7	190/200
User 15	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	9	10	10	199/200

We extract features from each chewing sequence to form a feature vector. For the

peak detection, the peak amplitude threshold, P_{amp} , is set to 4. Of all the 2742 detected chewing sequences, 34 chewing sequences contain no more than one peak. The first reason is that some users chew only a few times or do not chew at all when eating the Ice Cream. The second reason is that the number of peaks is under-estimated for some chewing sequences, because the amplitude of some real peaks is smaller than the peak amplitude threshold. As the chewing duration feature is defined as the distance between the first and last peaks, it equals zero for these chewing sequences, which is unreasonable. Thus, these chewing sequences are dropped. The resulting sample numbers of these 15 users are 196, 171, 194, 194, 178, 187, 180, 195, 186, 181, 156, 163, 144, 186, and 197, respectively.

The feature vectors of each user are normalized using the z-score algorithm [94] to eliminate the scaling effects among the features. The z-score algorithm normalizes each feature so that all the samples of this feature have mean 0 and standard deviation 1 [94]. The normalized feature vectors are combined with the corresponding labels of 11 food categories or 20 food types. Then, the classification datasets are obtained for model training and testing.

4.6.3 Performance Evaluation with the 11 Food Categories and 20 Food Types

The evaluation experiments are conducted for each user separately. The 10-fold cross-validation test is utilized to evaluate the recognition performance. The MLP classifier in the Weka toolkit [72] is used in our experiments. We adopt the default parameters for the MLP classifier in all the following experiments.

Table 4.6 and Table 4.7 show the recognition accuracies on the 11 food categories and 20 food types, respectively. We see that our proposed method accurately recognizes these 11 food categories and 20 food types. For the recognition of 11 food categories, the average accuracy of 15 users reaches 74.3%. The accuracy of a single user is up to 86.7%; for the recognition of 20 food types, the average accuracy of 15 users is 64.8%.

The accuracy of a single user is up to 77.0%. Compared with the average accuracy on the 20 food types, the average accuracy on the 11 food categories increases 9.5%. If we use a random classifier, its recognition accuracies on the 11 food categories and 20 food types are 9.1% and 5.0%, respectively. Comparatively, our proposed method is eight times as accurate as the random classifier on the 11 food categories, and thirteen times as accurate as the random classifier on the 20 food types. These results do approve the concept of our design. In addition, the performance evaluation is based on the data segmentation results. As we introduced in Section 6.2, the data segmentation results are not entirely correct (e.g. user 12 has 11 detected chewing sequences for the 5th, 9th and 11th food types). Therefore, with a more accurate data segmentation algorithm, the performance of our proposed method could be further improved.

Table 4.6: Recognition accuracy on the 11 food categories

User ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Avg.±Std
Accy (%)	79.6	79.5	85.6	85.1	66.3	78.6	73.3	86.7	69.9	79.0	74.4	53.4	68.8	54.8	79.7	74.3±10.2

Table 4.7: Recognition accuracy on the 20 food types

User ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Avg.±Std
Accy (%)	77.0	64.9	75.8	75.8	52.2	70.6	58.3	74.9	65.6	72.4	67.9	49.1	55.6	42.5	69.0	64.8±10.8

The recognition performance of the male users is better than that of the female users. For the recognition of 11 food categories, the average accuracy of ten male users (users 1 to 10) is 78.4%; the average accuracy of five female users (users 11 to 15) is 66.2%. For the recognition of 20 food types, the average accuracy of ten male users is 68.8%; the average accuracy of five female users is 56.8%. In these two experiments, the average accuracies of the male users are about 12% higher than that of the female users. One possible reason is that the male users have stronger chewing force than the female users. According to a clinical study on the habitual mastication patterns of 20 male users and 17 female users, “men used significantly greater chewing force than women [95]”. The stronger the chewing force, the larger the muscle bulge. Therefore, the motion data are more distinguishable for different food categories and food types.

The recognition accuracies are possibly related to users' eating speed. From Table 4.6 and Table 4.7 we also see that the recognition accuracies vary a lot for different users. For the recognition of 11 food categories, the standard deviation of accuracies for 15 users is 10.2%. For the recognition of 20 food types, the standard deviation of accuracies for 15 users is 10.8%. To investigate the accuracy variation for different users, we examine users' eating speed. Table 4.8 shows the total number of chews during eating all these 20 types of food. We see that users 4, 6, and 8 chew more than 5000 times in total. They eat slower than other users. Their recognition accuracies on the 11 food categories are 85.1%, 78.6%, and 86.7%, respectively. These accuracies are obviously higher than the average accuracy of 74.3%. The same is true for their recognition accuracies on the 20 food types. On the contrary, users 5, 7, 9, and 13 chew less than 3000 times in total. They eat faster than other users. Their recognition accuracies on the 11 food categories are 66.3%, 73.3%, 69.9%, and 68.8%, respectively. These accuracies are lower than the average accuracy of 74.3%. The same is true for users 5, 7, and 13 on the recognition accuracies of 20 food types. User 9's accuracy on 20 food types is 0.8% higher than the average accuracy. According to the above observations, we believe that the recognition accuracies are related to users' eating speed. When a user eats very fast, the corresponding chewing sequences are short. Thus, the extracted features might be less distinguishable among different food categories and food types.

Table 4.8: Total number of chews during eating all the food

User ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
# of chews	4554	3131	3840	6177	2863	5357	2869	5587	2554	3098	3698	3073	2333	4856	4884

To examine the misclassified samples among 20 food types, we sum the classification confusion matrices of these 15 users together and show it in Fig. 4.7. We see that the misclassified samples cluster in several areas. The first area is among the 1st, 2nd, and 3rd food types. These three food types belong to the 1st category (the Nuts). The second area is between the 5th and 6th food types, i.e. the Dry Pineapple Tidbit and Dry Mongo Slice. The third area is between the 10th and 11th food types, i.e. the

Fresh Tomato and Green Grape. The fourth area is among the 13th, 14th, and 15th food types. These three food types belong to the 6th category (the Corn and Fry). Clearly, misclassification often happens among food types with similar food properties and accordingly similar mastication dynamics. This indicates that our proposed method specializes in recognizing the food types with different food properties. This conclusion is consistent with the motivation of our proposed method.

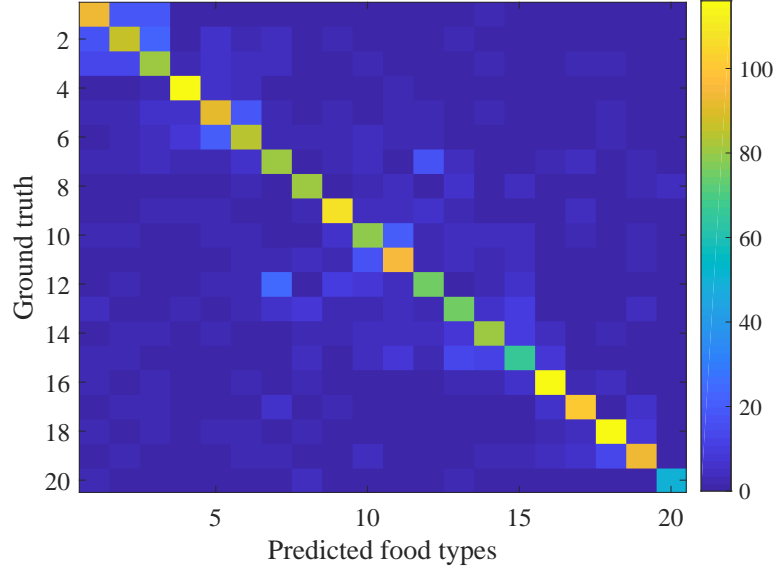


Figure 4.7: Sum of the confusion matrices of the 15 users

4.6.4 Feature Importance Analysis

The 37 hand-crafted features have different importances for the classification models. Identifying the most important features are very helpful for the nutritionists and medical professionals to understand our proposed food type recognition method.

The importances of these features are evaluated as follows. We utilize the *InfoGainAttributeEval* algorithm in Weka toolkit [72] to rank all these 37 features for each user’s dataset. The *InfoGainAttributeEval* algorithm “evaluates the worth of an attribute (i.e. feature) by measuring the information gain with respect to the class [72]”. Then, from the ranked features, the top five features are selected. Based on all the selected features

from 15 users, we count the number of occurrences of each feature. This number of occurrences is used to roughly represent the importance of a feature. The larger the number, the more important this feature is.

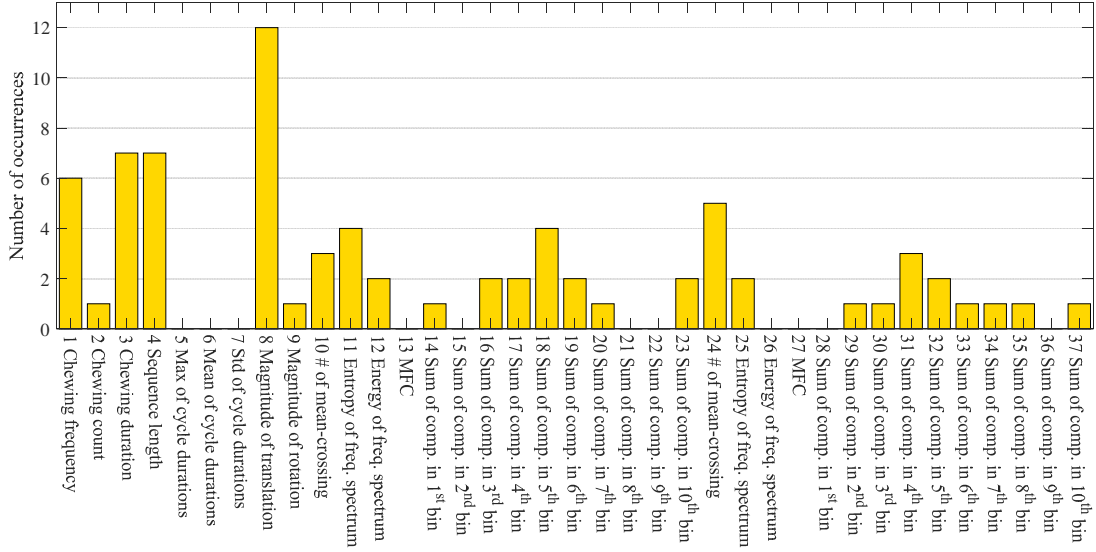


Figure 4.8: The number of occurrences of each selected feature on the datasets of 11 food categories

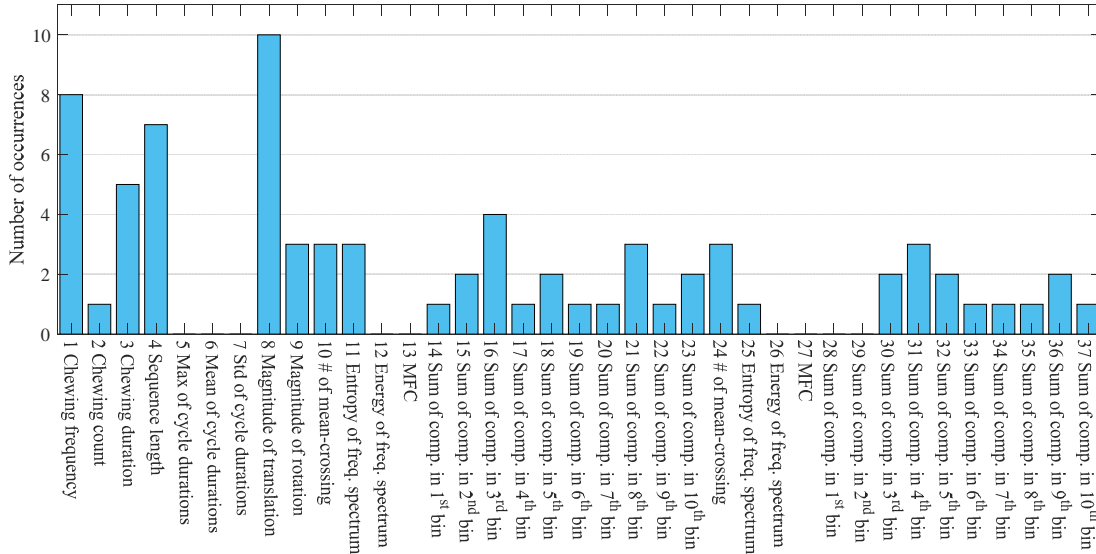


Figure 4.9: The number of occurrences of each selected feature on the datasets of 20 food types

Fig. 4.8 and Fig. 4.9 show the number of occurrences of each selected feature on

the datasets of 11 food categories and 20 food types, respectively. We see that the 1st (chewing frequency), 3rd (chewing duration), 4th (sequence length), and 8th (magnitude of translation) features have larger numbers of occurrences than other features. In the datasets of 11 food categories, they occur 6, 7, 7, and 12 times, respectively. In the datasets of 20 food types, they occur 8, 5, 7, and 10 times, respectively. These four features characterize the chewing speed, chewing time, and chewing force parameters, which represent the temporalis muscle activities. The above observation indicates that the mastication muscle activities are most important for food type recognition.

The skull vibration is also important for food type recognition. The 10-37th features characterize skull vibration. We see that most of these features occur a few times. Of these features, the 11th feature (entropy of frequency spectrum on the left device) and the 24th feature (number of mean-crossing on the right device) are two important ones. The 11th feature occurs 4 and 3 times in the datasets of 11 food categories and 20 food types, respectively. The 24th feature occurs 5 and 3 times in the datasets of 11 food categories and 20 food types, respectively. This indicates that the skull vibration has the characteristic of high frequency irregularity and data fluctuation.

The mandible motions are less important for food type recognition in our proposed method. The 5-7th features (maximum, mean, and std of cycle durations) characterize the mandible motions. However, these three features are not selected. Though they may be helpful for the food type recognition, their contributions are relatively small.

4.6.5 Performance Evaluation of Single-Side Recognition

In our proposed method, two devices are used to sense the mastication dynamics simultaneously. If only one device is utilized to sense the mastication dynamics of either the left side or the right side, what is the recognition performance?

To answer this question, we evaluate the performance of our proposed method using the left device and the right device, respectively. First, we segment the data of each device using its own gyroscope data of the Z axis. Then, we extract features from each

chewing sequence. Compared with the two-side recognition (i.e. using both the left and right devices), the feature extraction process of the single-side recognition (i.e. using either the left device or the right device) has three differences. First, to calculate the chewing cycles dependent features, one sensor whose data is most regular and obvious is selected from the sensors on the host device. Second, the 8th feature (magnitude of translation) and the 9th feature (magnitude of rotation) are calculated from the data of the accelerometer and gyroscope, respectively, on the host device. Third, for the left-side recognition, 24-37th features (skull vibration related features calculated from the right accelerometer) do not exist. Similarly, for the right-side recognition, 10-23th features (skull vibration related features calculated from the left accelerometer) do not exist. Therefore, in total there are 23 features for the single-side recognition.

Table 4.9: Single-side recognition accuracy on the 11 food categories

User ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Avg.±Std
Left (%)	74.0	73.3	85.1	75.3	65.0	65.4	62.2	72.7	65.8	63.5	59.1	57.3	59.7	50.5	70.9	66.7±8.7
Right (%)	77.3	72.2	77.9	84.8	53.6	74.8	72.8	88.8	70.3	78.5	73.7	59.9	76.6	66.5	78.4	73.7±8.9

Table 4.10: Single-side recognition accuracy on the 20 food types

User ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Avg.±Std
Left (%)	60.2	59.3	75.4	69.1	45.8	54.8	54.4	58.1	57.8	50.3	55.8	39.6	44.4	42.0	57.3	55.0±9.7
Right (%)	70.2	52.5	66.2	74.9	46.9	63.2	54.3	73.5	57.8	66.1	67.9	42.7	60.7	49.2	67.8	60.9±10.0

The single-side recognition is able to distinguish the food types accurately. Table 4.9 and Table 4.10 show the single-side recognition accuracies on the 11 food categories and 20 food types, respectively. We see that the single-side recognition is also capable of distinguishing the 11 food categories and 20 food types accurately, although the average accuracies of the single-side recognition are slightly lower than the average accuracies of the two-side recognition in Table 4.6 and Table 4.7. The average accuracies of the left-side recognition on the 11 food categories and 20 food types are 66.7% and 55.0%, respectively. The right-side recognition outperforms the left-side recognition and gets close performance as the two-side recognition. The average accuracies of the right-side recognition on the 11 food categories and 20 food types are 73.7% and 60.9%, respectively. The above experimental results indicate that, in spite of slight accuracy degradation, our

proposed method works for the single-side recognition.

The performances of the single-side recognition vary for different users. For most users, the performance of the right-side recognition is better than that of the left-side recognition. However, for users 2 and 3, the left-side recognition outperforms the right-side recognition. The reason may be each user has his/her own preference to chew the food on the left or right side. Thus, the mastication muscles of the left and right sides are not exactly symmetric. Accordingly, the sensor data of these two sides have different capability to characterize the mastication dynamics.

4.7 Comparison with Existing Wearable Sensor-Based Methods

In this section, we present a comparison between our proposed method and the existing works. We first introduce the food types tested by these methods. Then, we compare their recognition performances. Finally, we analyze the intrusiveness of these methods.

4.7.1 Food Types

The wearable sensor-based methods utilize different sets of food types. Table 4.11 shows the food types tested by the existing methods and our proposed method. The food types of each method are sorted in alphabetical order. We see that these methods utilize different sets of food types. The reason may be that these methods use different sensors. For example, Päßler et al. [41], Bi et al. [43], and Amft et al. [44] used microphones to recognize the food types. Accordingly, they prefer the food types with different acoustic properties. Similarly, Zhang et al. [47] utilized EMG. Alshurafa et al. [49] used a piezoelectric sensor. Mirtchouk et al. [50] took advantage of both microphones in an earbud and motion sensors in the Google Glass and smartwatches. Thus, they selected the food types that have different acoustic properties and hand/head motions during food delivery and food intake. Compared with these existing methods, our proposed

method utilizes only motion sensors in a headband to sense the mastication dynamics. Therefore, we select 11 food categories with different mastication related food properties. They are the Nuts, Gum Candy, Dry Fruit Slices, Fresh Fruits, Pretzel, Corn and Fry, Cookie, Vegetable, Bead, Meat, and Frozen Cream. These food categories have different hardness, elasticity, fracturability, adhesiveness, and size. For some categories, we select multiple food types because their food properties have obvious difference even in the same category. For example, in the category of Corn and Fry, three food types are included. They are the popcorn, potato chip, and potato fry.

Table 4.11: Food types tested by wearable sensor-based methods

	Food types
Päßler et al. [41]	Apple, carrot, chocolate, peanut, potato chips, pudding, walnut, drink (juice/water)
Bi et al. [43]	Apple, carrot, cookie, peanut, potato chips, walnuts, water
Amft et al. [44]	Apple, biscuit, bread, carrot, chocolate, cooked chicken, cooked potato, corn-bar, french fries, lettuce, maize, muffin, orange, pasta, peanut, pepper, potato chips, toast, waffles
Zhang et al. [47]	Banana, biscuit, carrot, jelly baby, toast
Alshurafa et al. [49]	Liquid (hot tea, water) and solid (chocolate, nuts, patty)
Mirtchouk etc. [50]	Almonds, apple, bagel, banana, bread, broccoli, burger, burrito, celery, chicken, chipotle, chips, chocolate, citrus, danish, drink, dumpling, eggs, fries, ice cream, peanuts, peppers, pie, pizza, popcorn, potato, pretzel, rice, salad, sandwich, shake, shrimp, soup, spaghetti, steak, strawberries, sushi, tacos, wings, yogurt
Our method	Almond, apple slice, blueberry, bread, carrot, chicken breast nugget, cookie, dry banana chip, dry mongo slice, dry pineapple tidbit, grape, gummi bear, ice cream, peanut, pecan, popcorn, potato chip, potato fry, pretzel stick, tomato

Our proposed method has more food types than most of these existing methods. In total, 20 food types are included in our user study, as shown in Table 4.12. Comparatively, Päßler et al.’s method [41] included seven food types and one drink; Bi et al.’s method [43] tested seven food types; Amft et al.’s method [44] had 19 food types; Zhang et al.’s method [47] contained five food types; and Alshurafa et al.’s method [49] only included two food categories. Our proposed method contains more food types than the above methods. We also notice that Mirtchouk etc.’s method [50] included more food types than our proposed method. However, this method required a user to wear an earbud, a Google Glass and two smartwatches simultaneously. In addition, it required using a camera to record the video for manually annotation. This method is obviously

intrusive.

Table 4.12: Performance comparison of wearable sensor-based methods

	Sensor	Host object	Intrusive	# of subj.	# of food types	Accuracy
Päßler et al. [41]	2 microphones	Hearing aid package	Median	51	7 food types & 1 drink	79% on 10% of all records
Bi et al. [43]	1 microphone	Necklace-like device	High	12	7 food types	Average accuracy of 84.9%
Amft et al. [44]	1 microphone	Ear-pad	Low	3	19 food types	Accuracy of 80%
Zhang et al. [47]	1 EMG sensor	Eyeglass	Median	8	5 food types	63% to 84% for sequences
Alshurafa et al. [49]	1 piezo. sensor	Necklace	High	10	2 categories	90% for liquid and solid
Mirtchouk etc. [50]	Audio and motion sensors	Glass, earbud, smart-watch	Median	6	40 food types	Accuracy of 82.7%
Our method	2 accel. & 2 gyroscopes	Headband	Low	15	11 cat. (20 food types)	Average accuracy of 74.3%

4.7.2 Performance Comparison

For the recognition accuracy, it is infeasible to compare with existing wearable sensor-based methods using the same dataset because the sensors used are different. Here, we present a short performance comparison between our proposed method and other existing methods, as shown in Table 4.12. For Päßler et al.’s work [41] and Bi et al.’s work [43], they only recognized seven food types. Our method recognizes 11 food categories (in total 20 types of food). Amft et al.’s work [44] and our method have a similar number of food types, and both accurately recognize these food types. However, they use a microphone, which is easily impacted by ambient acoustic noise. For Zhang et al.’s work [47] and Alshurafa et al.’s work [49], they only recognized five food types and two food categories, respectively. Our method recognizes 20 food types and 11 food categories, which are much more than their methods. Mirtchouk et al. [50] presented a pioneering work on sensor fusion based food type recognition. They accurately classified 40 food types. However, they required a user to wear an earbud, a Google Glass and two smartwatches simultaneously during eating. Our method needs only a headband. We are also aware that they reported 62.3% accuracy when only Google Glass is used. However, for their Google Glass only solution, they still required using a camera to record the video during model training and testing. The video was used to manually annotate the exact time of food delivery, food intake, and chewing. Comparatively, our method only

uses motion sensors.

4.7.3 Intrusiveness Analysis

In the following, we analyze the intrusiveness of these wearable sensor-based methods.

Päßler et al. [41] deployed two microphones in a hearing aid package. The in-ear microphone was attached on the earmold that was placed in the ear canal. The reference microphone was attached on the hearing aid that hooked over the top of the ear. The earmold blocks the ear canal and affects daily communication. In addition, wearing a hearing aid package may restrict some daily activities (e.g. running). Bi et al. [43] deployed a microphone on a user’s neck to record acoustic signals during eating. This sensor deployment is intrusive. Amft et al. [44] embedded a microphone in an ear-pad case. Compared with the headphone, the ear-pad reduces ear occlusion and increases user comfort. However, it still blocks the ear canal and affects daily communication to some degree.

Zhang et al. [47] embedded EMG electrodes into 3D-printed eyeglasses. Although the eyeglasses is commonly used in daily living, the skin contact of electrodes at the nose and ear areas is intrusive. Alshurafa et al. [49] embedded a piezoelectric sensor against the throat to detect skin motion in the lower trachea during ingestion. This method is obviously intrusive and uncomfortable for long-term wear.

Mirtchouk et al. [50] utilized an earbud, a Google Glass and two smartwatches to collect sensor data simultaneously. Though these three kinds of devices are commonly used in daily living, wearing all these devices at the same time certainly interferes the user’s daily living.

Our proposed method embeds motion sensors into a headband. The motion sensors need no skin contact. In addition, the headband is often used by some people for fashion and practical purposes (e.g. hold hair away from the face or eyes). We are also aware that some people are not used to wearing headband. For those who do not prefer to wear a headband, we can use a hat as the host object and embed the sensors at the inside

edge of the hat.

4.8 Discussion and Future Work

In this section, we discuss some implementation issues of our proposed method and introduce some future work.

In our experiment, we only include food types that contain one composition. For the food types that consist of multiple compositions (such as the sandwich, pizza, and hamburger), the sensed mastication dynamics represent a mixture of food properties of all the compositions. Accordingly, the performance of the recognition model may be reduced, especially when these compositions are included as independent food types. Moreover, the variety of the composition proportion makes this problem more difficult. We will investigate this problem in the future.

The evaluation of our proposed method is done in a lab environment, in which some confounding factors in real life scenarios are not considered. For example, people eat while talking and drinking. The talking and drinking activities normally happen after a swallow and before the next bite. These two noisy activities can be filtered using an eating/chewing detection module [90]. We plan to integrate this module into our proposed method in the future.

4.9 Conclusion

In this chapter, a motion sensor-based food type recognition method is proposed. We observe that each type of food has its own intrinsic properties, such as hardness, elasticity, fracturability, adhesiveness, and size. Different food properties result in different mastication dynamics. Through embedding motion sensors in a headband and deploying the sensors on the temporalis muscles, the mastication dynamics can be captured accurately. Based on these observations, we propose the first effort in using motion sensors to sense mastication dynamics and infer food types accordingly. We define six mastication

dynamics parameters to represent these food properties. From each chewing sequence, we extract 37 hand-crafted features to explicitly characterize the mastication dynamics using motion sensor data. Experiments are conducted with 15 human subjects on 11 food categories (in total 20 types of food). The average recognition accuracy of these 15 human subjects is 74.3%. The accuracy of a single human subject is up to 86.7%.

4.A Sensor Selection for Extracting Chewing Cycles Dependent Features

Each chewing sequence contains data from 12 sensors. They are the X , Y and Z axes of the accelerometer and gyroscope of the left and right devices. Fig. 4.10 shows the 3-axis accelerometer data and the 3-axis gyroscope data in one chewing sequence. We clearly see that the gyroscope data is more regular and obvious than the accelerometer data, especially for the X axis and the Y axis. Therefore, we propose a metric, R_{MFC} , to select one sensor whose data is most regular and obvious from the X and Y axes of the left and right gyroscopes.

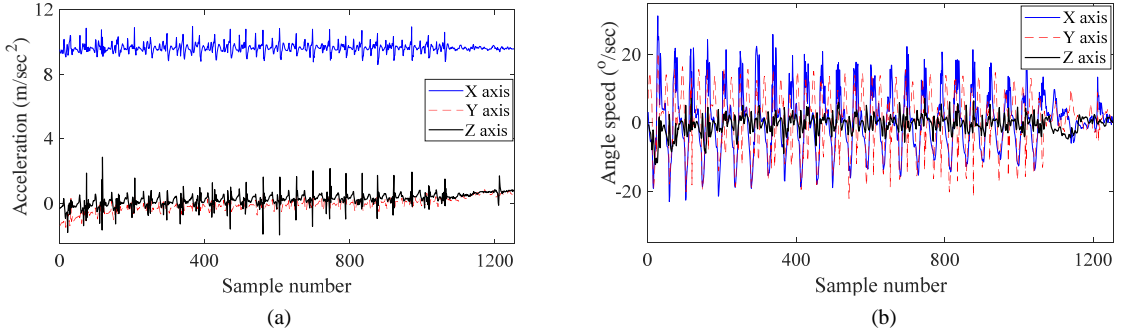


Figure 4.10: One chewing sequence. (a) Accelerometer data; (b) Gyroscope data

We observe that if the data of one sensor is more regular and obvious than the data of other sensors, its energy should be more concentrated on a small range of frequencies. Taking the 3-axis gyroscope data of one chewing sequence in Fig. 4.10(b) as an example, we first filter the data of each axis using a 9th-order one-dimension median filter [76] to reduce the noise. Then, we conduct Fourier transform on the filtered data of each

axis. The single-sided amplitude spectrum [69] (without direct current component) of the 3-axis gyroscope data is shown in Fig. 4.11. We see that the X and Y axes have obvious spikes near 2.25 Hz and 4.5 Hz. The spikes near 2.25 Hz indicate the frequency amplitude of single chewing cycles, and the spikes near 4.5 Hz indicate the frequency amplitude of double chewing cycles. Based on the above observation, we define R_{MFC} as the ratio of the maximum frequency component (MFC) in the chewing frequency range to the sum of all the frequency components. That is,

$$R_{MFC} = \frac{MFC}{\text{Sum}(C_i)}, \quad (4.6)$$

where C_i is the amplitude of the i^{th} frequency component. A large R_{MFC} indicates that the energy of the corresponding sensor data is concentrated on a small range of frequencies. Accordingly, the sensor whose data has the largest R_{MFC} is selected for extracting the chewing cycles dependent features.

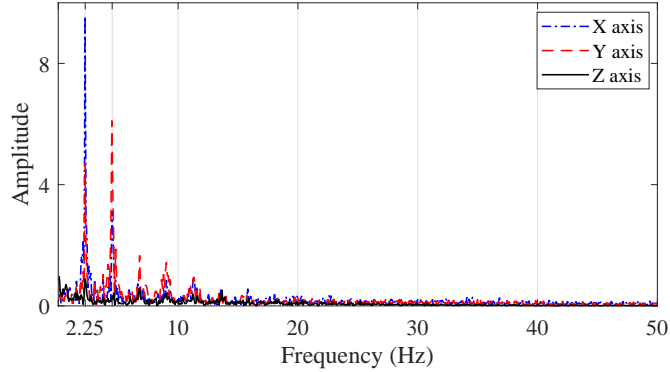


Figure 4.11: Single-sided amplitude spectrum of the gyroscope data in Fig. 4.10 (b)

Chapter 5

Wearable Motion Sensor-Based Chewing Side Detection

5.1 Introduction

Chewing side preference is an unhealthy dietary habit that is very common but often overlooked. A subject with chewing side preference demonstrates a tendency to chew food mainly on one side [9]. This tendency causes lateral asymmetry of chewing force and occlusal contact area [96]. Accordingly, it can lead to several diseases, such as lateral facial asymmetry [9], teeth abrasion [97], temporomandibular disorders (TMD) [96, 98], malocclusion, and stomach illness [99]. Clinical studies have demonstrated that the majority of people have a chewing side preference. Donnell et al. [100] examined 57 children aged 6-8 years old. The percentage with a chewing side preference varies “from 70% of the caries free group to 92% of the group with caries, pathology and pain [100]”. Tiwari et al. [9] examined 76 healthy adults with a mean age of 20.8. They observed that 75 of these adults had a chewing side preference. However, more than half were not aware their habit of chewing side preference.

In order to detect chewing side preference and quantify its severity in a continuous and convenient way, several wearable sensor-based methods have been proposed in recent years. Chung et al. [60] embedded two load cells into hinges on both sides of a pair of

glasses to recognize ingestive and facial behaviors, such as head movement, left chewing, right chewing, winking and talking. This method heavily relies on the sensitivity and deployment of the load cells. In addition, it is a coarse-grained method because each motion detection is based on a fixed-length window of 3 seconds instead of one chew. Lucena et al. [61] attached two motion sensors to a subject’s jaw and forehead to sense the jaw movements and detect chewing sides. This method is obviously intrusive.

One research challenge is how to detect chewing sides accurately and less intrusively using wearable sensors. To address this challenge, we investigate mastication muscle bulge and skull vibration during chewing. We observe that: 1) chewing activity generates mastication muscle bulge and skull vibration. During chewing, the mastication muscles contract and relax rhythmically. When the mastication muscles contract, they bulge to some degree; when these muscles relax, they return to the original shape. In addition, food grinding and tooth friction during chewing cause skull vibration. 2) muscle bulge and skull vibration can be sensed by wearable motion sensors. When we deploy motion sensors on the mastication muscles, the muscle bulge and skull vibration cause translation and rotation movements of these sensors. 3) the muscle bulge and skull vibration of the chewing side are different from those of the non-chewing side. A subject normally exerts larger chewing force on the chewing side than that on the non-chewing side. Thus, the muscle bulge and skull vibration of the chewing side are relatively larger than those of the non-chewing side.

Based on the above observations, we propose to deploy motion sensors on the mastication muscles to sense muscle bulges and skull vibrations and differentiate chewing sides accordingly. There are three research questions. *First, how can we segment each chew from the continuous motion sensor data? Second, how can we characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side? Third, how can we model and classify multi-dimensional data samples with unequal input lengths?* To answer these questions, 1) we propose a heuristic-rules based method to exclude non-chewing data and segment each chew accurately; 2) we propose to calculate

the relative difference series of the left and right sensors to characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side; 3) we utilize long short-term memory (LSTM), an artificial recurrent neural network, to model unequal-length data samples and classify chewing sides accordingly. To evaluate the performance of our proposed method, a real-world dataset of eight food categories is collected from eight human subjects. The experimental results show that our proposed method is very promising.

The main contributions are as follows:

1. We propose to detect chewing sides through sensing muscle bulge and skull vibration with wearable motion sensors. To our best knowledge, this is the first effort in using wearable motion sensors to differentiating muscle bulge and skull vibration between the chewing side and the non-chewing side.
2. We propose a heuristic-rules based method to accurately exclude non-chewing data and segment each chew from continuous motion sensor data.
3. We propose to calculate the relative difference series of the left and right sensors to characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side.
4. We evaluate the performance of our proposed method on a real-world dataset. Experimental results show that the average detection accuracy reaches 84.8%. The detection accuracy of a single human subject is up to 97.4%.

The remainder of this chapter is organized as follows. Section 5.2 describes the system overview of our proposed method. Section 5.3 describes the motion sensor deployment. In Section 5.4, we describe the motion data collection and calibration. Section 5.5 describes the data segmentation and classification sample generation. In Section 5.6, we describe the chewing side classification with LSTM. Experiment and evaluation are presented in Section 5.7. Section 5.8 presents discussion. Finally, we draw our conclusion in Section 5.9.

5.2 System Overview

Figure 5.1 shows the flowchart of our proposed chewing side detection method. Two motion-sensing devices are deployed on the left and right temporalis muscles. From each device, we collect both accelerometer data and gyroscope data simultaneously. The collected sensor data are first calibrated to eliminate the scaling and bias errors. Then, non-chewing data are excluded using a heuristic-rules based method. After segmenting the sensor data for each chew, we calculate the relative difference series of the left and right sensors to characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side. By combining the relative difference series with corresponding labels (i.e. left side and right side), we obtain the dataset that is utilized to train and test an LSTM classification model.

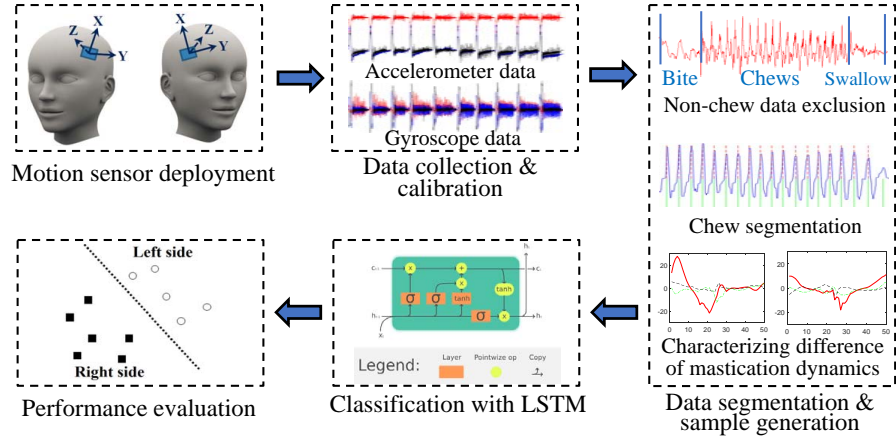


Figure 5.1: Flowchart of our proposed chewing side detection method

5.3 Motion Sensor Deployment

Accelerometer and gyroscope are the most suitable sensors for sensing muscle bulge and skull vibration. When a mastication muscle contracts, its muscle spindle becomes shorter and thicker. Accordingly, the muscle bulges to some degree, which causes the translation and rotation movements of the sensors. The translation movements result in velocity changes that are directly sensed by an accelerometer [101]. The rotation

movements result in changes of orientation and angular velocity that are directly sensed by a gyroscope [102]. During chewing, the upper and lower teeth rub against each other to grind the food, which causes skull vibration. Skull vibration is a mechanical oscillation that is directly sensed by an accelerometer.

In Figure 5.2, the green colored muscle is the temporalis. It is the most suitable muscle to deploy motion sensors. We humans have four mastication muscles: the masseter, the medial pterygoid, the lateral pterygoid, and the temporalis [64]. The masseter, medial pterygoid, and lateral pterygoid are located in the face area and near the mouth cavity. This area is not suitable for deploying a sensor. The temporalis is located at each side of the skull and in front of the ear [22], where we often wear a headband, eyeglasses, or hat. This observation motivates us to embed motion sensors into a headband and deploy the sensors on the temporalis muscles.

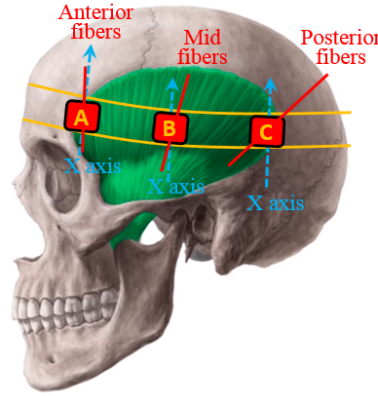


Figure 5.2: Temporalis muscle fibers and sensor location

We deploy motion sensors on the anterior fibers of the temporalis muscles. The fibers of a temporalis muscle are divided into three parts: anterior fibers, mid fibers, and posterior fibers. As indicated in reference [103], “the anterior fibers have a vertical orientation, mid fibers have an oblique orientation, and posterior fibers have a more of a horizontal orientation [103]”, as shown by the red lines in Figure 5.2. The anterior and mid fibers are responsible for elevating the mandible; the posterior fibers are responsible for retracting the mandible [103]. Therefore, the anterior and mid fibers are stronger

than the posterior fibers. Accordingly, they have a larger bulge than the posterior fibers during contraction. The blue arrows in Figure 5.2 show the orientations of the X axis of the left accelerometer and gyroscope at three different locations. Compared with locations B and C, location A has the smallest angle between the X axis and the muscle fibers. The smaller the angle is, the more regular the sensor readings are. Considering the above two observations, we deploy the left device at location A, which is on the anterior fibers and near the edge of the temporalis muscle. Similarly, the right device is deployed at the symmetrical location on the right temporalis muscle.

5.4 Motion Data Collection and Calibration

Two small-size wearable devices [2] shown in Figure 4.2 (a) are used to collect motion data. Each device contains a 3-axis accelerometer, a 3-axis gyroscope, and a 3-axis digital compass. The device deployment and sensor orientations are the same as those in Figure 4.2.

Eight human subjects were recruited to collect motion sensor data. The subjects sit in front of a table and eat eight different types of food. The food is cut into pieces (if necessary) and served on a paper plate. The subjects eat the food one piece at a time using a spoon, fork, chopsticks, or hand. When a subject eats the served food, the sensors sample simultaneously at 100 Hz. Only the accelerometer and gyroscope data are used in our proposed detection method. The sampled data are transmitted to a mobile phone through BLE in real-time. Then, the data are transferred to a PC for offline analysis.

To eliminate the scaling and bias errors, we calibrate the collected accelerometer and gyroscope data separately using the same method introduced in Section 4.4.

5.5 Data Segmentation & Classification Sample Generation

In this section, we first describe how to exclude biting data and swallowing data. Then, we present how to segment each chew. Finally, we describe how to generate one classification sample for each chew.

5.5.1 Biting Data Exclusion

Biting actions are normally accompanied by head motions. An eating cycle involves three actions: biting, chewing, and swallowing. Before a subject bites the food, he/she normally bows down his/her head to some degree to approach the food. After biting, he/she starts to chew the food during or after raising his/her head. The chewing process contains multiple continuous chews and is followed by swallowing. The above eating cycles repeat until the subject finishes all the food.

The Z axis of the left gyroscope is used to detect head motions and infer biting actions. Based on the sensor orientations shown in Figure 4.2, when a subject bows down his/her head, the left device rotates around its Z axis in a clockwise direction, and the right device rotates around its Z axis in a counter-clockwise direction; when a subject raises his/her head, the left device rotates around its Z axis in a counter-clockwise direction, and the right device rotates around its Z axis in a clockwise direction. Therefore, the Z axis of either the left gyroscope or the right gyroscope can be utilized to detect head motions and infer biting actions accordingly. In our proposed method, we randomly select the Z axis of the left gyroscope.

We observe that the Z axis gyroscope data during biting have larger deviations than those during chewing. Biting actions are low-frequency activities. To eliminate the high-frequency noise, the Z axis gyroscope data are filtered using a moving average filter of span s . Here, s is set to 31. Figure 5.3 (a) - (c) show the filtered Z axis gyroscope data when a subject eats five pieces of food using a spoon, chopsticks, and hand, respectively.

We see that when the subject bows down his/her head, the Z axis gyroscope data are negative and form a valley; when the subject raises his/her head, the Z axis gyroscope data are positive and form a peak. The standard deviations of these three data series are 8.7, 5.9, and 3.3, respectively. Figure 5.3 (d) shows the filtered Z axis gyroscope data when the subject chews food without head motion. The standard deviation is 0.6, which is much smaller than those in Figure 5.3 (a) - (c).

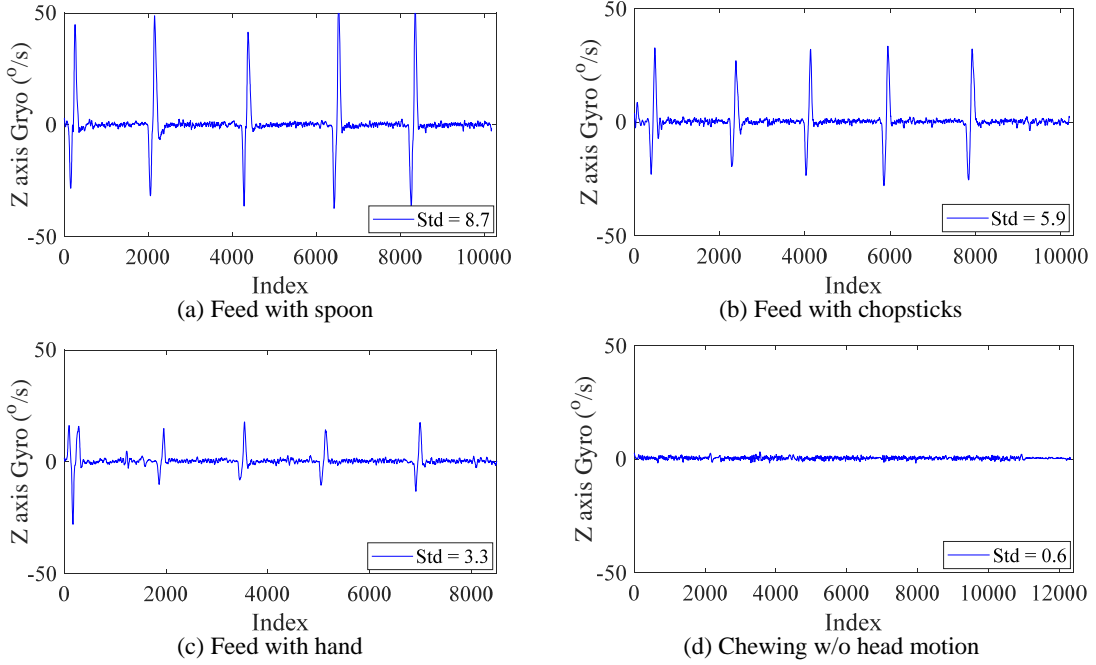


Figure 5.3: Filtered Z axis gyroscope data

We exclude biting data through detecting the deviation of filtered Z axis gyroscope data. To exclude biting data, we first calculate the mean and standard deviation, i.e. $Mean$ and Std , of the filtered Z axis gyroscope data. Then, we set two thresholds, $Mean - Std$ and $Mean + Std$, as the lower bound and upper bound of the chewing data. The sensor readings larger than the upper bound or smaller than the lower bound are considered as biting data and hence dropped. Excluding biting data may partition the data series into multiple segments. Each segment contains chewing data and swallowing data. A few segments may be incomplete because of noise or intense head motions. We set a length threshold, len , to drop the segments that are shorter than this threshold.

Here, len is set to three seconds.

5.5.2 Swallowing Data Exclusion

The X axis of the left gyroscope is utilized to differentiate chewing data and swallowing data. As shown in Figure 5.2, we deploy the left device at location A. When the temporalis muscle contracts and bulges, the device rotates around the X axis to some degree. The X axis gyroscope data are first filtered using a n^{th} -order median filter, which eliminates outliers while preserving details of the data. Here, n is the length of the sliding window that the median is calculated from, and we set n to 15. Figure 5.4 shows an example of the filtered X axis gyroscope data during chewing and swallowing. We see that: 1) the distances between two neighboring peaks of the chewing data are very regular, which show the periodic and rhythmic muscle contractions. Comparatively, the distances between two neighboring peaks of the swallowing data are irregular; 2) the amplitude of the chewing data is much larger than that of the swallowing data. The reason is that the temporalis muscles are not responsible for swallowing. Accordingly, the muscle bulge is not obvious during swallowing.

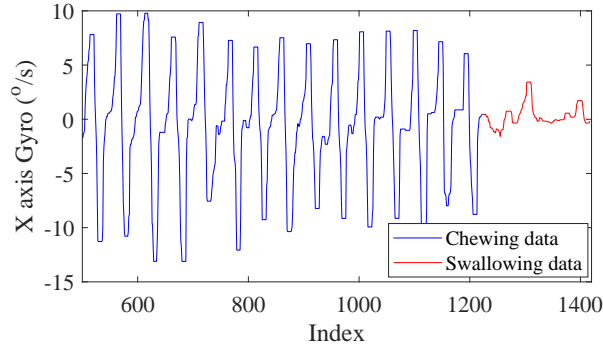


Figure 5.4: Filtered X axis gyroscope data during chewing and swallowing

We propose two heuristic rules to exclude swallowing data. The first heuristic rule is the distance between two neighboring peaks, i.e. D_{P2P} shown in Figure 5.5 (a), satisfies the following inequality for the chewing data:

$$\frac{3}{4} \cdot \frac{Rs}{f_{MFC}} \leq D_{P2P} \leq \frac{3}{2} \cdot \frac{Rs}{f_{MFC}} \quad (5.1)$$

where Rs is the sampling rate, and f_{MFC} means the chewing frequency. It is the frequency corresponding to the MFC in the chewing frequency range. Here, the chewing frequency range is defined as $[0.5, 2.5]$ Hz [90]. $\frac{Rs}{f_{MFC}}$ represents the cycle length of one chew. If the distance between two neighboring peaks is less than $\frac{3}{4}$ cycle length [91] or larger than $\frac{3}{2}$ cycle length, the data between these two peaks are considered as swallowing data and dropped.

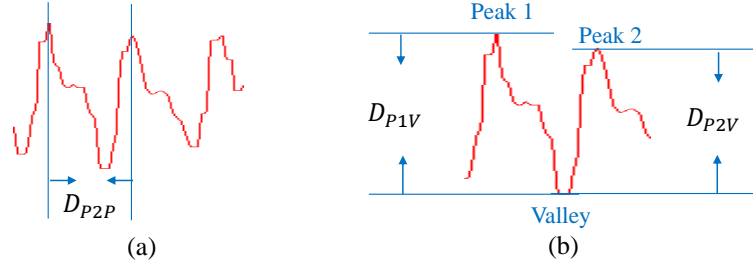


Figure 5.5: Distance between peaks and amplitude difference

The second heuristic rule is the amplitude differences between two neighboring peaks and the valley between them, i.e. D_{P1V} and D_{P2V} shown in Figure 5.5 (b), satisfy the following two inequalities for the chewing data:

$$\begin{cases} D_{P1V} \geq T_{PV} \\ D_{P2V} \geq T_{PV} \end{cases} \quad (5.2)$$

where T_{PV} is the predefined minimum amplitude difference for the chewing data. If either one of these two inequalities is not satisfied, the data between these two neighboring peaks are considered as swallowing data and dropped.

5.5.3 Chew Segmentation

The X axis data of the left gyroscope is used to segment each chew. After excluding biting data, we conduct chew segmentation on the X axis gyroscope data using the above

two heuristic rules. The steps of the proposed chew segmentation method are described as follows:

- Step 1: Median filtering. We use a 15th-order median filter to eliminate outliers in the X axis gyroscope data.
- Step 2: Chewing frequency determination. The Fourier transform is conducted on the filtered X axis gyroscope data to compute the single-sided amplitude spectrum [104] (without direct current component). Then, the chewing frequency, f_{MFC} , is obtained through locating the frequency corresponding to the MFC in the chewing frequency range.
- Step 3: Zero-crossing point detection. We scan the filtered X axis gyroscope data sequentially and find out all zero-crossing points.
- Step 4: Peak/valley detection. The maximum and minimum data points are detected between any two neighboring zero-crossing points. If the absolute value of the maximum data point is larger than that of the minimum data point, the maximum data point is selected as a peak; otherwise, the minimum data point is selected as a valley. If there are two neighboring peaks or valleys, the one with the larger absolute value is kept and the other is dropped.
- Step 5: Data segmentation for each chew. We scan all the peaks and valleys from beginning to end. If any two neighboring peaks and the valley between them satisfy the aforementioned two heuristic rules in inequalities 5.1 and 5.2, the data segment between these two peaks is considered as a chew.
- Step 6: Segmentation of calibrated data traces. Based on the segmentation locations of the X axis gyroscope data, all the calibrated data traces are segmented accordingly. We extract the calibrated data of all the left and right sensors for each chew, which are used to generate one classification sample.

5.5.4 Classification Sample Generation

Based on the extracted sensor data of each chew, we calculate the relative difference series of the left and right sensors to characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side. There are two relative difference series: accelerometer difference series and gyroscope difference series.

To calculate the accelerometer difference series, we first compose the i^{th} data points of three axes, a_X^i , a_Y^i and a_Z^i , into one scalar acceleration a_i using Equation 3.1. Here, $i = 1, 2, \dots, n$. n is the data length of one chew segment. The resulting scalar acceleration series is $A = [a_1, a_2, \dots, a_n]$. We calculate the scalar acceleration series for the left accelerometer and right accelerometer separately, which are expressed as A_{left} and A_{right} . Because the left and right temporalis muscles are not identical, they may have different sizes, thicknesses, and strengths. Thus, their bulge degrees may be different, which causes different acceleration amplitudes. To eliminate the scaling effects between A_{left} and A_{right} , we normalize them separately using the z-score algorithm [94]. The z-score algorithm normalizes each scalar acceleration series so that all the data in this series have mean 0 and standard deviation 1 [94]. The normalized scalar acceleration series are \hat{A}_{left} and \hat{A}_{right} . Finally, the accelerometer difference series is calculated as $D_{Accel} = \hat{A}_{left} - \hat{A}_{right}$. Similarly, from the gyroscope data of each chew, we calculate the gyroscope difference series, $D_{Gyro} = \hat{G}_{left} - \hat{G}_{right}$, using the same method above.

One classification sample is generated for each chew. We combine D_{Accel} and D_{Gyro} to form a two-dimensional data series. Through combining this data series with corresponding label (i.e. left side or right side), one classification sample is generated. All classification samples constitute a classification dataset, which is utilized to train and test a classification model.

5.6 Chewing Side Classification with LSTM

In this section, we first explain why we need a personalized classification model for chewing side detection. Then, we describe why we use LSTM to train the classification model. Finally, we introduce the architecture and implementation of the LSTM model.

5.6.1 Why Personalized Classification Model?

A subject’s chewing patterns are consistent and can be accurately modeled. Clinical studies [80, 93] demonstrated the stability of intra-individual masticatory parameters, such as EMG activity, chewing frequency, and the number of chews. Woda et al. concluded that “the lack of a significant difference for the same individual has been clearly shown by using well controlled food stimuli and strict inclusion criteria [80].” The stability of intra-individual masticatory parameters indicates that a subject’s chewing patterns can be accurately modeled.

Chewing side detection requires a personalized classification model. Clinical studies also showed that the masticatory parameters vary a lot between individuals [80, 93]. For example, Lassauzay et al. [93] recruited 15 male students between 21 and 25 years old to eat four food products and then measured their masticatory parameters. The results showed that the mean muscular work, which is calculated from the EMG data and characterizes the magnitude of the chewing force, varies from 0.57 mV.s to 1.58 mV.s, the chewing frequency varies from 1.19 Hz to 1.70 Hz, and the number of chews varies from 19 to 57. Therefore, we propose to train a personalized chewing side detection model for each subject.

5.6.2 Classification Algorithm Selection

LSTM is the most suitable classification model for our proposed method. To classify multi-dimensional time series, traditional classification models (decision tree, support vector machine, etc.) need to explicitly extract features from the time series data. However, the best features for chewing side detection are implicit and unknown. The

feature extraction process may lose some useful information and impact the detection accuracy. Another option is using deep learning models, such as convolutional neural network (CNN) and LSTM, to automatically learn these implicit features from the time series data. CNN is suitable for spatial data with fixed-size inputs. Comparatively, LSTM is suitable for temporal data. An LSTM is a special type of recurrent neural network. It is capable of learning long-term dependencies between time steps of sequence data [3]. In addition, it can handle arbitrary input lengths [105]. The classification samples obtained in Section 5.5 are temporal data with unequal lengths. Therefore, we utilize LSTM to model the data and classify chewing sides accordingly.

5.6.3 Architecture and implementation of LSTM model

A simple LSTM model is utilized in our proposed method. The LSTM models can be roughly divided into two categories: simple LSTM models and deeper LSTM models [3]. A simple LSTM model only contains one LSTM layer; a deeper LSTM model contains two or more LSTM layers. A deeper LSTM model has a higher learning and representation capability than a simple LSTM model. However, the more layers it has, the higher the computation and storage cost it requires. In addition, a deeper LSTM model is easier to overfit than a simple LSTM model. We choose a simple LSTM model in our proposed method. There are two reasons: 1) although our method is implemented as an offline system, its future implementation in a real-world application should be online and on-device. A simple LSTM model is suitable for the limited resources of wearable devices; 2) compared with a deeper LSTM model, a simple LSTM model has a higher generalization capability and thus is more robust in a real-world application.

Chewing side detection is formulated as a sequence classification problem. As shown in Figure 5.6, we adopt a five-layer architecture introduced in the Matlab deep learning toolbox [3, 4]: 1) Sequence input layer: The sequence input layer inputs the classification samples to the LSTM layer; 2) LSTM layer: The LSTM layer learns long-term dependencies between time steps of the input data. It then outputs the last hidden state

h_t [106], a D -dimensional vector, to the next layer. Here, D is the number of hidden units, which is set to 20; 3) Fully connected layer: A fully connected layer follows the LSTM layer. It maps each output of the LSTM layer to a c -dimensional vector. This is done through multiplying h_t by a weight matrix and then adding a bias vector [107]. Here, c is the output size of the fully connected layer, which is equal to the number of all possible classes. In our proposed method, $c = 2$. For a sequence classification problem, the fully connected layer must be followed by a softmax layer and then a classification layer [108]; 4) Softmax layer: The softmax layer applies a softmax activation function to the output of the fully connected layer and generates a probability distribution over all possible classes; 5) Classification layer: Based on the probabilities obtained in the softmax layer, the classification layer computes the cross-entropy loss [109] and then assigns a predication label (i.e. left side or right side) for each classification sample.



Figure 5.6: Architecture of the five-layer LSTM model [3, 4]

Mini-batch training is often used to speed up and optimize the process of stochastic gradient descent optimization during training a LSTM model. Mini-batch training splits the training dataset into multiple mini-batches. Each mini-batch is used to separately evaluate the gradient of the loss function and update the weights [110]. The existing research results [111, 112] have demonstrated that mini-batch training can not only improves the computational efficiency but also avoids dropping into local minima.

Mini-batch training requires the classification samples in each mini-batch to have the same length. To accommodate this requirement, a method named zero padding [4] is often used. Zero padding method adds zeros to the samples so that all the samples in a mini-batch have the same length as the longest one. One question is that adding zeros introduces noise into the samples. The more zeros added, the more noise introduced. Figure 5.7 shows the padding effect on 500 unsorted chewing side classification samples. The size of each mini-batch is set to 50, as indicated by the red lines in Figure 5.7. We see

that quite a few zeros are added to the samples in each mini-batch. To reduce the number of padded zeros, we sort the classification samples by length [4] in ascending order before padding zeros. Figure 5.8 shows the padding effect on the sorted classification samples. We see that the number of padded zeros are greatly reduced.

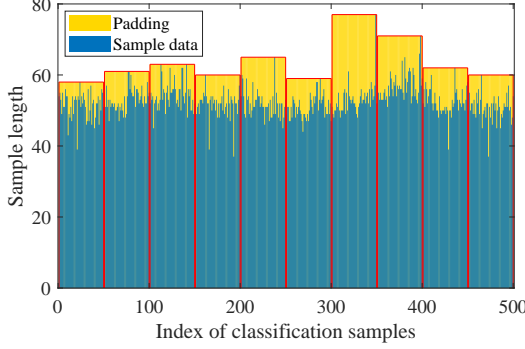


Figure 5.7: Padding effect on unsorted samples

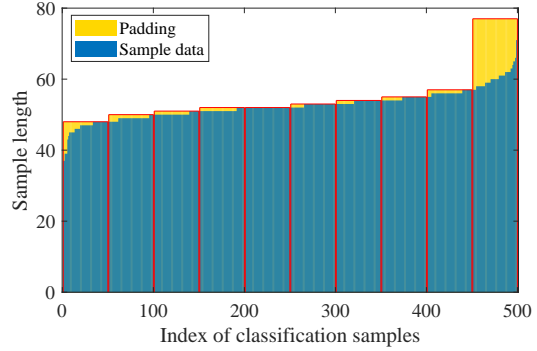


Figure 5.8: Padding effect on sorted samples

5.7 Experiment and Evaluation

In this section, we first describe the experimental setup. Then, we present the performances of chew segmentation and chewing side detection. Next, we analyze the impact of sensor location bias. Finally, we evaluate the accuracy of the proposed method for unknown food types.

5.7.1 Experimental Setup

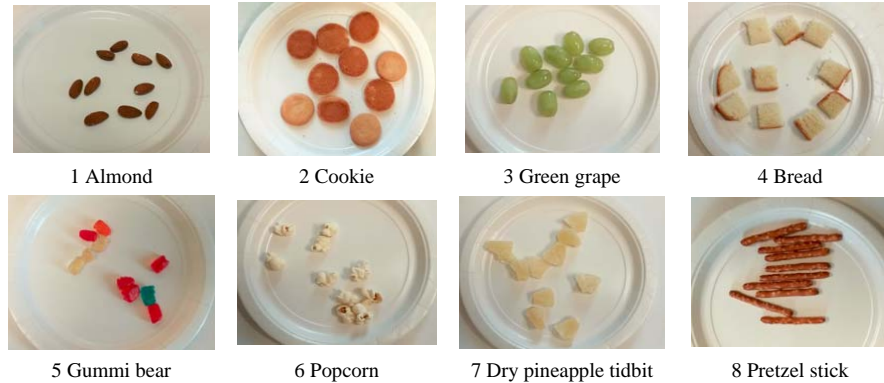
With the approval from the institutional review board (IRB), five male users and three female users were recruited in our user study. Their demographic information is shown in Table 5.1, including age, gender, weight, head circumference, dominant feeding hand, and whether they wear glasses or not during eating. The sensor deployment and data collection method are the same as those in Section 5.4. After initial deployment, the sensor locations may be slightly adjusted if the collected data are too noisy.

Eight different food types are included in our user study. They are almond, cookie,

Table 5.1: The users’ demographic information

User	Age	Gender	Weight (lbs)	Head cir. (cm)	Feeding hand	With glasses
User 1	41	Male	200	58	Right	No
User 2	26	Female	116	53	Right	Yes
User 3	41	Female	128	56	Right	No
User 4	29	Male	150	58	Right	Yes
User 5	29	Male	150	60	Right	Yes
User 6	24	Female	100	58	Right	Yes
User 7	31	Male	174	58	Right	Yes
User 8	31	Male	160	58	Right	Yes

green grape, bread, gummi bear, popcorn, dry pineapple tidbit, and pretzel stick. These food types have different food properties and are bought from the Food Lion Grocery Store. For each food type, we serve 10 pieces to each user, as shown in Figure 5.9. The users eat one piece at a time. These eight foods are delivered and fed four different ways. The 1st and 2nd food types are fed with a spoon; the 3rd and 4th food types are fed with a fork; the 5th and 6th food types are fed with chopsticks, and the 7th and 8th food types are fed with a hand. Users 6 and 7 do not like eating gummi bear. They are allowed to skip it.

**Figure 5.9:** The food served to each user

The classification samples are labeled based on the corresponding chewing sides. For each food type, we ask the users to chew the first five pieces on the left side and the second five pieces on the right side. Accordingly, all the classification samples corresponding to the first five pieces are labeled as “left side”. All the classification samples corresponding to the second five pieces are labeled as “right side”. We do not consider the situation of

chewing on both sides in our experiment, which will be discussed in Section 5.8.

To obtain the ground truth of the number of chews, we video-recorded the whole eating process using a PC-embedded camera. Then, we replay the videos slowly and manually count the number of chews by observing users’ jaw movements.

5.7.2 Performance of Chew Segmentation

The first question we would like to answer is how accurate the proposed chew segmentation method is. Table 5.2 shows the ground truth, the number of detected chews, and the chew segmentation accuracy for each user. Here, the chew segmentation accuracy is defined as the ratio of the number of detected chews to the ground truth. We see that the proposed method detects and segments chews accurately. The average accuracy of these eight users is 85.9%. The accuracy of user 2 is up to 94.4%.

Table 5.2: Chew segmentation accuracy

User	Ground truth	# of detected chews	Accuracy
User 1	2897	2538	87.6%
User 2	2158	2037	94.4%
User 3	1967	1674	85.1%
User 4	2903	2519	86.8%
User 5	1430	1130	79.0%
User 6	2347	2140	91.2%
User 7	1293	1091	84.4%
User 8	1378	1084	78.7%

From Table 5.2 we also observe that the numbers of detected chews are smaller than the ground truths. This indicates that some chews are not detected. The reason may be that the temporalis muscle bulge in the late stage of chewing (i.e. right before swallowing) is much smaller than that in the early and middle stages. In the late stage of chewing, the food has already been chewed into very small particles. Thus, the chewing force in the late stage is smaller than that in the early and middle stages. A small chewing force results in a small temporalis muscle bulge. Accordingly, some chews do not satisfy the second heuristic rule in inequality 5.2 and hence are dropped.

5.7.3 Performance of Chewing Side Detection

The second question we would like to answer is how accurate the proposed chewing side detection method is. To answer this question, we conduct a five-fold cross validation test [113, 114] on the classification dataset of each user. The classification dataset is split into five mutually exclusive and exhaustive folds. Each time, one fold is selected as the testing dataset; the remaining four folds are combined together to form the training dataset. A classification model is trained on the training dataset and then tested with the testing dataset. The testing accuracies of these five folds are averaged. This average accuracy is taken as the accuracy of the five-fold cross validation test.

Our proposed method detects chewing sides accurately. Table 5.3 shows the accuracies of the five-fold cross validation test for these eight users. We see that the chewing sides are accurately classified. The average detection accuracy of these eight users reaches 84.8%. The detection accuracy of user 1 is up to 97.4%.

From Table 5.3 we also see that the detection accuracies of different users vary to some degree. The standard deviation of the detection accuracies is 7.8%. The difference between the highest accuracy and the lowest accuracy is 19.5%. According to our experience, this may be because the sensors are not deployed at the optimal locations for some users. We will investigate this problem in the following subsection.

Table 5.3: Chewing side detection accuracy

	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	Avg.±Std
Accy (%)	97.4	94.3	87.4	78.5	86.1	78.1	77.9	79.0	84.8±7.8

Figure 5.10 (a) shows the sum of classification confusion matrices of these eight users in the five-fold cross validation test. We see that both left side chews and right side chews are accurately classified. Of all the left side chews, 81.7% are correctly classified; of all the right side chews, 89.3% are correctly classified. From Figure 5.10 (a) we also observe that more left side chews are misclassified than right side chews. More specifically, 1230 left side chews are misclassified as right side, and 801 right side chews are misclassified as left side. This indicates that the classification models are biased to right side a little

bit. The reason may be the classification datasets are not completely balanced. For the same type and amount of food, users chew on the right side slightly more than the left side. The right side has 7483 samples, and the left side has 6730 samples. Normally, “classification learning algorithms are biased towards the majority class and therefore there is a higher misclassification rate for the minority class instances [115]”.

Figure 5.10 (b) and (c) show the sum of the confusion matrices of male users and female users, respectively, in the five-fold cross validation test. We have the same observations as those from Figure 5.10 (a). This indicates that our proposed method is valid for both male users and female users.

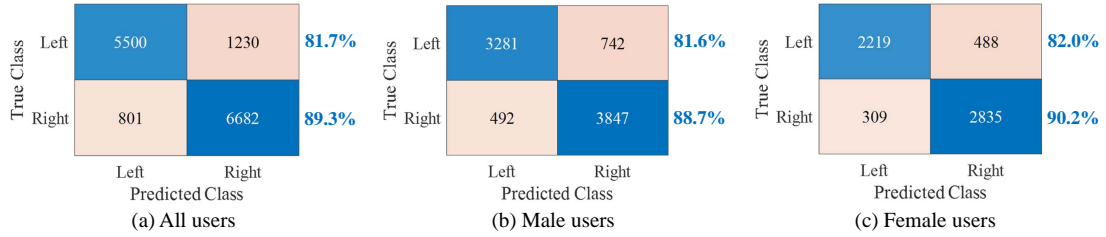


Figure 5.10: Sum of the confusion matrices in the five-fold cross validation test

5.7.4 Impact of Sensor Location Bias

The third question we would like to answer is how much impact a sensor location bias has on the detection accuracy. To answer this question, we ask user 1 to collect data once again. The left device is deployed one centimeter backward from the location A in Figure 5.2. Similarly, the right device is moved one centimeter backward.

The detection model is sensitive to sensor location bias. We train a classification model using the dataset collected at the previous location and test this model using the dataset collected at the biased location. The detection accuracy is 45.6%. Conversely, we train a classification model using the dataset collected at the biased location and test this model using the dataset collected at the previous location. The detection accuracy is 29.0%. The average detection accuracy of these two tests is 37.3%, which is 47.5% lower than the average detection accuracy of the five-fold cross validation test.

There are two potential reasons for the model sensitivity to the sensor location bias. The first reason is that the muscle bulges of different locations are very irregular. As we described in Section 5.3, the direction and strength of the muscle fibers are very irregular. Accordingly, muscle bulges of different locations vary a lot. When the sensor location is changed, the muscle bulge and the corresponding sensor data change accordingly. The second reason is that the sensor data collected at the biased location is too noisy. This is because the circuit board is a rigid body and covers an area of $26mm \times 25mm$ [2]. The muscle bulge of any point in this area may introduce noise and affect the sensed data. This problem can be solved through detaching the small sensor module from the circuit board and only deploying sensors at the specified locations.

5.7.5 Leave-One-Food-Out Test

The fourth question we would like to answer is how generic a classification model is for unknown food types. To answer this question, we conduct a leave-one-food-out test on the classification dataset of each user. The leave-one-food-out test uses classification samples of all food types except one to train a classification model. Then, this model is tested using the classification samples of the excluded food type. This process repeats for all food types. The average of the testing accuracies is taken as the accuracy of the leave-one-food-out test.

Our proposed method is capable of detecting chewing sides accurately for unknown food types. In Figure 5.11, we compare the accuracies of the leave-one-food-out test and the accuracies of the five-fold cross validation test. The average detection accuracy of the leave-one-food-out test is 81.4%, which is only 3.4% lower than the average detection accuracy of the five-fold cross validation test. This indicates that our proposed classification models can be generalized to unknown food types.

Figure 5.11 also shows that, for all the users except user 4, the accuracies of the leave-one-food-out test are a little lower than those of the five-fold cross validation test. This is reasonable because the data of the testing food type are not included in the

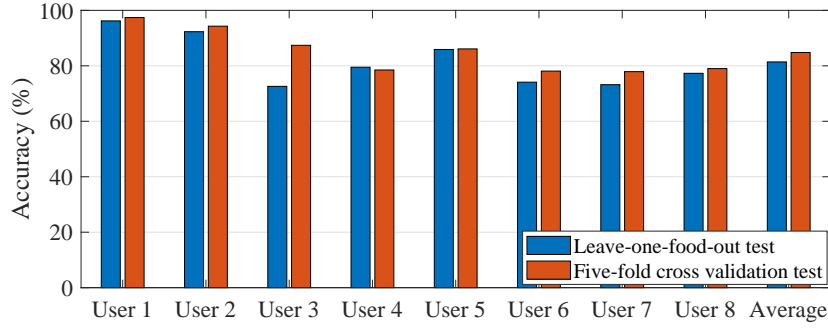


Figure 5.11: Accuracy comparison between the leave-one-food-out test and five-fold cross validation test

training dataset. For user 4, the accuracy of the leave-one-food-out test is only 1% higher than that of the five-fold cross validation test. This may be because we adopt the same parameters for all the users when training the classification models. These parameters are not optimal for the five-fold cross validation test of user 4.

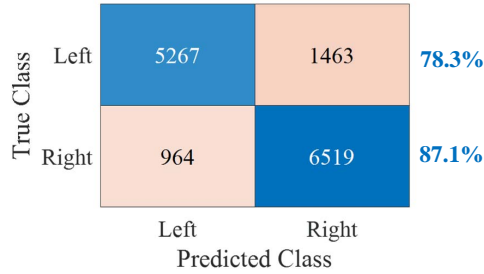


Figure 5.12: Sum of the confusion matrices in the leave-one-food-out test

Figure 5.12 shows the sum of the confusion matrices of these eight users in the leave-one-food-out test. We observe that both left side chews and right side chews are accurately classified for unknown food types. 78.3% of the left side chews are correctly classified, and 87.1% of the right side chews are correctly classified. Similar to the five-fold cross validation test, there are more right side chews than the left side chews in the leave-one-food-out test. Accordingly, the classification models are slightly biased to the right side.

5.8 Discussion

In our experiment, we only consider two cases: chewing on the left side and chewing on the right side. We are aware that sometimes a subject may chew on both sides. However, we do not consider this case in our user study. There are two reasons. First, according to the clinical studies, chewing on both sides happens much less frequently than chewing on the left or right side. For example, Gisel [116] investigated the oral side preferences of 103 children. Because “placement of food on the preferred side might facilitate chewing [116]”, Gisel conducted this investigation through observing the initial placement of food by the tongue. The experimental results show that only 2.5% observations are placing food on both sides. Second, chewing on both sides normally happens in the late stage of chewing. At this stage, the food has been chewed into particles, which are distributed on both sides. As we indicated in Section 5.7, the chewing force in the late stage is much smaller than that in the early and middle stages. Accordingly, the chews in the late stage may not satisfy the second heuristic rule and hence are excluded.

Eating activity may be accompanied by other activities, such as talking and drinking, which normally happen after a swallow and before the next bite. These noisy activities can be easily filtered through integrating an eating detection module [90] into our proposed method.

5.9 Conclusion

Chewing side preference can result in some diseases, such as lateral facial asymmetry, teeth abrasion, temporomandibular disorders, malocclusion, and stomach illness. Existing wearable sensor-based methods are either intrusive or not accurate enough. In this chapter, we propose a wearable motion sensor-based method to detect chewing sides accurately and less-intrusively. This is done through embedding motion sensors in a headband to sense temporalis muscle bulge and skull vibration. We utilize a heuristic-rules based method to exclude non-chewing data and segment each chew accurately.

The relative difference series of the left and right sensors is then calculated to characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side. To model the data samples with unequal input lengths, LSTM is utilized to train a classifier and recognize chewing sides accordingly. Experiments are conducted with eight human subjects on eight food types. The results demonstrate that the proposed method accurately detects chewing sides. The average detection accuracy of these eight subjects reaches 84.8%. The detection accuracy of a single human subject is up to 97.4%.

Chapter 6

Future Work

In Chapters 3 to 5, we have introduced three research projects. In each project, we have listed some corresponding future work. Here, we list more future research topics in terms of the whole dissertation.

- **Food amount estimation.** Carbohydrate intake estimation provides medical professionals with important information to diagnose and monitor some chronic diseases, such as obesity and diabetes. Our food type recognition method has demonstrated that it can accurately recognize different food types. In order to calculate the amount of carbohydrate intake, we also need to estimate the amount of consumed foods. According to the clinical study [80], the food amount of each bite is correlated with some masticatory parameters, including the number of chews, chewing time, EMG activity, vertical/lateral amplitude, and closing velocity [80]. These masticatory parameters can be directly calculated or indirectly estimated from the motion sensor data, which are collected from motion sensors deployed on the temporalis muscles. Therefore, it is highly possible to estimate the amount of consumed foods. There are two research questions. *First, how to accurately estimate the masticatory parameters that can not be directly calculated from the motion sensor data (e.g. EMG activity)? Second, how to model the mapping from these masticatory parameters to the amount of consumed foods?*

- **Energy-efficient sensing.** A wearable device may have a very limited power supply because of the constraints of its size and weight. How can we reduce the energy consumption of data sensing to extend the lifetime of a wearable device? There is a potential solution, which is a combination of the following two strategies: 1) event-based high-power sensor triggering. We noticed that the energy consumption of the accelerometer is about 1.5% of the energy consumption of the gyroscope [117]. Thus, we can use one accelerometer in either the left or the right device to recognize eating and non-eating activities, as we described in Chapter 3. All the other sensors are turned on only when a user is eating. This strategy can greatly reduce energy consumption. However, the sensor triggering always occurs after detecting an eating activity and hence is delayed for a while; 2) eating detection with a dynamic sampling rate. For our eating detection method in Chapter 3, the accelerometer is sampled at 100Hz. This sampling rate is appropriate for distinguishing eating from some similar daily activities (e.g. speaking/reading) but may be too high for distinguishing eating from other non-similar activities (e.g. sitting). Thus, we can dynamically adjust the sampling rate of the accelerometer for different activities. This strategy can further reduce energy consumption. There are three research questions. *First, how can we minimize the delay of sensor triggering? Second, how can the eating detection model adapt to accelerometer data with a dynamic sampling rate? Third, how can we minimize the sampling rate for each specific activity?*
- **Bluetooth radio-based dietary monitoring.** Although the motion sensor-based dietary monitoring methods introduced in this dissertation are very promising, a few users in our user study expressed their unwillingness to wear motion sensor during eating. Without the motion sensor attached on the mastication muscle, the proposed methods do not work anymore. This motivates us to think about how to monitor people’s diet and eating habits using other alternative non-intrusive sensors. We are considering replacing the motion sensors with BLE transceivers.

The BLE transceivers have already been widely used in commercial products like Apple Bluetooth earbuds. We plan to attach two BLE transceivers at the end of two earpieces of eyeglasses. One transceiver acts as the transmitter, and the other transceiver acts as the receiver. The BLE signals sent from the transmitter go across a user's head and are received by the receiver. During eating, the mandible moves up and down. This movement changes the shape of the oral cavity and also the transmission channel of the BLE signals. Accordingly, the strength of received BLE signals changes synchronously with the mandible movement and shows a periodic pattern. Therefore, dietary monitoring can be done by analyzing the strength series of the received BLE signals. There are three research questions. *First, how can we characterize the eating activity and mastication dynamics only using the BLE strength series? Second, how can we personalize the recognition models so that they can adapt to the head size of a specific user? Third, how can we minimize the energy consumption of the BLE transceivers while preserving the recognition performance?*

Chapter 7

Conclusion

In this dissertation, we present our efforts in using wearable motion sensors to sense mastication dynamics for continuous dietary monitoring. Specifically, we work on the following three topics:

First, we propose a motion sensor-based eating activity detection and chews counting method. Eating activity detection is formulated as a two-class classification problem. Twenty-three representative time-domain and frequency-domain features are extracted from a few seconds of acceleration data to accurately distinguish eating activity from other non-eating activities. In addition, the number of chews is accurately counted by identifying the primary periodicity of acceleration data. Experimental results show that the average *accuracy* and *F-score* are 94.4% and 87.2%, respectively, for eating activity detection in 10-fold cross-validation test. The average error rate of chews counting for four users is 12.2%.

Second, we propose a motion sensor-based food type recognition method. We define six mastication dynamics parameters to represent the food properties. They are chewing speed, the number of chews, chewing time, chewing force, chewing cycle duration, and skull vibration. To sense mastication dynamics accurately and less intrusively, we embed motion sensors in a headband and deploy the sensors on the left and right temporalis muscles. We specifically define 37 hand-crafted features, which are extracted from each chewing sequence to explicitly characterize the mastication dynamics. Based on the

extracted features, a neural network model is trained to recognize 11 food categories (20 types of food in total). Performance evaluation on a real-world dataset demonstrates that the average recognition accuracy of 15 human subjects is 74.3%. The recognition accuracy of a single human subject is up to 86.7%.

Third, we propose a motion sensor-based chewing side detection method. We present a heuristic-rules based method to exclude non-chewing data and segment each chew accurately. We calculate the relative difference series of the left and right sensors to characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side. To model multi-dimensional data samples with unequal input lengths, we utilize LSTM to train a two-class classifier for chewing side classification. A real-world evaluation dataset of eight food types is collected from eight human subjects. The average detection accuracy reaches 84.8%. The highest detection accuracy for a single subject is up to 97.4%.

Bibliography

- [1] Kari Ann Hong. The muscles of mastication. <https://www.thousandoaksfamilydentistry.com/blog/2014/12/22/the-muscles-of-mastication>, 2014.
- [2] Hongyang Zhao, Shuangquan Wang, Gang Zhou, and Daqing Zhang. Gesture-enabled remote control for healthcare. In *Proceedings of the Second IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies*, CHASE '17, pages 392–401, Piscataway, NJ, USA, 2017. IEEE Press.
- [3] Long short-term memory networks. <https://www.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html>.
- [4] Sequence classification using deep learning. <https://www.mathworks.com/help/deeplearning/examples/classify-sequence-data-using-lstm-networks.html>.
- [5] Wikipedia. Eating disorder. https://en.wikipedia.org/wiki/Eating_disorder, 2020.
- [6] Lexico US dictionary. Anorexia. <https://www.lexico.com/en/definition/anorexia>, 2020.
- [7] National Eating Disorders Association (NEDA). Bulimia nervosa. <https://www.nationaleatingdisorders.org/learn/by-eating-disorder/bulimia>, 2018.

- [8] Something to chew on: Eating meals slowly could reduce your risk of diabetes. <http://www.dailymail.co.uk/health/article-2141077>.
- [9] Shreyasi Tiwari, Supriya Nambiar, and Bhaskaran Unnikrishnan. Chewing side preference - impact on facial symmetry, dentition and temporomandibular joint and its correlation with handedness. *Journal of Orofacial Sciences*, 9:22–27, 01 2017.
- [10] Craig M. Hales, Margaret D. Carroll, Cheryl D. Fryar, and Cynthia L. Ogden. Prevalence of obesity among adults and youth: United states, 2015-2016. Technical report, Hyattsville, MD, USA, 2017.
- [11] CDC. National diabetes statistics report, 2017. <https://www.cdc.gov/diabetes/data/statistics/statistics-report.html>, 2017.
- [12] CDC. Deaths: Final data for 2009. www.cdc.gov/nchs/data/nvsr/nvsr60/nvsr60_03.pdf, 2011.
- [13] National Center for Health Statistics. National ambulatory medical care survey: 2016 national summary tables. https://www.cdc.gov/nchs/data/ahcd/namcs_summary/2016_namcs_web_tables.pdf, 2016.
- [14] Kristin L. Schneider, Emily Panza, Bradley M. Appelhans, Matthew C. Whited, Jessica L. Oleski, and Sherry L. Pagoto. The emotional eating scale. can a self-report measure predict observed emotional eating? *Appetite*, 58(2):563–566, 2012.
- [15] Oliver Amft, Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. Analysis of chewing sounds for dietary monitoring. In *Proceedings of the 7th International Conference on Ubiquitous Computing*, UbiComp ’05, pages 56–72. Springer-Verlag, 2005.
- [16] Koji Yatani and Khai N. Truong. Bodyscope: A wearable acoustic sensor for activity recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp ’12, pages 341–350, New York, NY, USA, 2012. ACM.

- [17] Steven Cadavid, Mohamed Abdel-Mottaleb, and Abdelsalam Helal. Exploiting visual quasi-periodicity for real-time chewing event detection using active appearance models and support vector machines. *Personal and Ubiquitous Computing*, 16(6):729–739, 2012.
- [18] Yujie Dong, Adam Hoover, Jenna Scisco, and Eric Muth. A new method for measuring meal intake in humans via automated wrist motion tracking. *Applied psychophysiology and biofeedback*, 37:205–15, 04 2012.
- [19] Sen Zhang, Marcelo H Ang, Wendong Xiao, and Chen Khong Tham. Detection of activities by wireless sensors for daily life surveillance: Eating and drinking. *Sensors (Basel, Switzerland)*, 9(3):1499–1517, 2009.
- [20] Rui Zhang and Oliver Amft. Regular-look eyeglasses can monitor chewing. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp ’16, pages 389–392, New York, NY, USA, 2016. ACM.
- [21] E. S. Sazonov and J. M. Fontana. A sensor system for automatic detection of food intake through non-invasive monitoring of chewing. *IEEE Sensors Journal*, 12(5):1340–1348, 2012.
- [22] Temporalis muscle: Location, action and trigger points. <http://www.gustrength.com/muscles:temporalis-location-action-and-trigger-points>.
- [23] Y. Miyaoka, I. Ashida, Y. Tamaki, S. Kawakami, H. Iwamori, T. Yamazaki, and N. Ito. Quantitative analysis of relationships between masseter activity during chewing and textural properties of foods. *Food and Nutrition Sciences*, 4(2):144–149, 2013.
- [24] V. A. Vance, S. J. Woodruff, L. J. McCargar, J. Husted, and R. M. Hanning. Self-reported dietary energy intake of normal weight, overweight and obese adolescents. *Public Health Nutr.*, 12(2):222–7, 2009.

- [25] Dale A. Schoeller. How accurate is self-reported dietary energy intake? *Nutrition Reviews*, 48(10):373–379, 1990.
- [26] Xu Ye, Guanling Chen, and Yu Cao. Automatic eating detection using head-mount and wrist-worn accelerometers. In *2015 17th International Conference on E-health Networking, Application Services (HealthCom)*, pages 578–581, Oct 2015.
- [27] Gert Mertes, Hans Hallez, Tom Croonenborghs, and Bart Vanrumste. Detection of chewing motion using a glasses mounted accelerometer towards monitoring of food intake events in the elderly. In *International Conference on Biomedical and Health Informatics*, pages 1–5, 10 2015.
- [28] Shah Atiqur Rahman, Christopher Merck, Yuxiao Huang, and Samantha Kleinberg. Unintrusive eating recognition using google glass. In *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth '15*, pages 108–111, 2015.
- [29] Edison Thomaz, Irfan Essa, and Gregory D. Abowd. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '15*, pages 1029–1040. ACM, 2015.
- [30] Azusa Kadomura, Cheng-Yuan Li, Koji Tsukada, Hao-Hua Chu, and Itiro Siio. Persuasive technology to improve eating behavior using a sensor-embedded fork. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '14*, pages 319–329, New York, NY, USA, 2014. ACM.
- [31] Cheng-Yuan Li, Yen-Chang Chen, Wei-Ju Chen, Polly Huang, and Hao-hua Chu. Sensor-embedded teeth for oral activity recognition. In *Proceedings of the 2013 International Symposium on Wearable Computers, ISWC '13*, pages 41–44. ACM, 2013.

- [32] Shuangquan Wang, Gang Zhou, Lisha Hu, Zhenyu Chen, and Yiqiang Chen. Care: Chewing activity recognition using noninvasive single axis accelerometer. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, UbiComp/ISWC'15 Adjunct, pages 109–112, New York, NY, USA, 2015. ACM.
- [33] Muhammad Farooq and Edward Sazonov. Automatic measurement of chew count and chewing rate during food intake. *Electronics*, 5(4), 2016.
- [34] Muhammad Farooq and Edward Sazonov. Segmentation and characterization of chewing bouts by monitoring temporalis muscle using smart glasses with piezoelectric sensor. *IEEE Journal of Biomedical and Health Informatics*, 21(6):1495–1503, 2017.
- [35] M. M. Anthimopoulos, L. Gianola, L. Scarnato, P. Diem, and S. G. Mougiakakou. A food recognition system for diabetic patients based on an optimized bag-of-features model. *IEEE Journal of Biomedical and Health Informatics*, 18(4):1261–1271, July 2014.
- [36] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101 – mining discriminative components with random forests. In *Computer Vision – ECCV 2014*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, pages 446–461, Cham, 2014. Springer International Publishing.
- [37] Shulin Yang, Mei Chen, Dean A. Pomerleau, and Rahul Sukthankar. Food recognition using statistics of pairwise local features. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2249–2256, 2010.
- [38] Y. Kawano and K. Yanai. Real-time mobile food recognition system. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–7, Portland, OR, USA, June 2013. IEEE.

- [39] J. Lester, D. Tan, S. Patel, and A. J. B. Brush. Automatic classification of daily fluid intake. In *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*, pages 1–8, Munich, Germany, March 2010. IEEE.
- [40] B. Zhou, J. Cheng, P. Lukowicz, A. Reiss, and O. Amft. Monitoring dietary behavior with a smart dining tray. *IEEE Pervasive Computing*, 14(4):46–56, Oct 2015.
- [41] Sebastian Päßler, Matthias Wolff, and Wolf-Joachim Fischer. Food intake monitoring: an acoustical approach to automated food intake activity detection and classification of consumed food. *Physiol Meas.*, 33(6):1073–93, Jun 2012.
- [42] O. Amft, M. Kusserow, and G. Trster. Bite weight prediction from acoustic recognition of chewing. *IEEE Transactions on Biomedical Engineering*, 56(6):1663–1672, June 2009.
- [43] Y. Bi, M. Lv, C. Song, W. Xu, N. Guan, and W. Yi. Autodietary: A wearable acoustic sensor system for food intake recognition in daily life. *IEEE Sensors Journal*, 16(3):806–816, Feb 2016.
- [44] Oliver Amft and Gerhard Tröster. On-body sensing solutions for automatic dietary monitoring. *IEEE Pervasive Computing*, 8(2):62–70, April 2009.
- [45] O. Amft. A wearable earpad sensor for chewing monitoring. In *2010 IEEE Sensors*, pages 222–227, Kona, HI, USA, Nov 2010. IEEE.
- [46] Tauhidur Rahman, Alexander T. Adams, Mi Zhang, Erin Cherry, Bobby Zhou, Huaishu Peng, and Tanzeem Choudhury. Bodybeat: A mobile system for sensing non-speech body sounds. In *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys ’14, pages 2–13, New York, NY, USA, 2014. ACM.
- [47] R. Zhang, S. Bernhart, and O. Amft. Diet eyeglasses: Recognising food chewing using emg and smart eyeglasses. In *2016 IEEE 13th International Conference on*

- Wearable and Implantable Body Sensor Networks (BSN)*, pages 7–12, San Francisco, CA, USA, June 2016. IEEE.
- [48] H. Kalantarian, N. Alshurafa, and M. Sarrafzadeh. A wearable nutrition monitoring system. In *2014 11th International Conference on Wearable and Implantable Body Sensor Networks*, pages 75–80, Zurich, Switzerland, June 2014. IEEE.
 - [49] N. Alshurafa, H. Kalantarian, M. Pourhomayoun, J. J. Liu, S. Sarin, B. Shahbazi, and M. Sarrafzadeh. Recognition of nutrition intake using time-frequency decomposition in a wearable necklace using a piezoelectric sensor. *IEEE Sensors Journal*, 15(7):3909–3916, July 2015.
 - [50] Mark Mirtchouk, Christopher Merck, and Samantha Kleinberg. Automated estimation of food type and amount consumed from body-worn audio and motion sensors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp ’16, pages 451–462, New York, NY, USA, 2016. ACM.
 - [51] H. J. Kim, M. Kim, S. J. Lee, and Y. S. Choi. An analysis of eating activities for automatic food type recognition. In *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–5, Hollywood, CA, USA, Dec 2012. IEEE.
 - [52] Martin Biallas, Aliaksei Andrushevich, R Kistler, Alexander Klapproth, Krzysztof Czuszyński, and Adam Bujnowski. Feasibility study for food intake tasks recognition based on smart glasses. *Journal of Medical Imaging and Health Informatics*, 5:1688–1694, 12 2015.
 - [53] M. Farooq and E. Sazonov. Accelerometer-based detection of food intake in free-living individuals. *IEEE Sensors Journal*, 18(9):3752–3758, May 2018.
 - [54] Keum San Chun, Sarnab Bhattacharya, and Edison Thomaz. Detecting eating

- episodes by tracking jawbone movements with a non-contact wearable sensor. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1):4:1–4:21, March 2018.
- [55] Elan Flores-Orozco, Bernat Rovira, Maria Peraire, Juan Salsench, and Jordi Martnez-Gomis. Reliability of a visual analog scale for determining the preferred mastication side. *The Journal of prosthetic dentistry*, 115, 10 2015.
- [56] J. Nissan, M. D. Gross, A. Shifman, L. Tzadok, and D. Assif. Chewing side preference as a type of hemispheric laterality. *J Oral Rehabil*, 31(5):412–6, 05 2004.
- [57] Mao Watarai, Rika Ayano, Takahiro Funatsu, Takafumi Ooka, Mari Takahashi, and Mitsuko Inoue. Three-dimensional motion analysis of lip and mandibular movements during mastication. *Dental Medicine Research*, 33:88–99, 01 2013.
- [58] T Mizumori, T Tsubakimoto, M Iwasaki, and T Nakamura. Masticatory laterality - evaluation and influence of food texture. *Journal of oral rehabilitation*, 30:995–9, 11 2003.
- [59] Yo Yamasaki, Rika Kuwatsuru, Yoshihiro Tsukiyama, Hiroshi Matsumoto, Kyosuke Oki, and Kiyoshi Koyano. Objective assessment of actual chewing side by measurement of bilateral masseter muscle electromyography. *Archives of oral biology*, 60:1756–1762, 10 2015.
- [60] Jungman Chung, Jungmin Chung, Wonjun Oh, Yongkyu Yoo, Won Gu Lee, and Hyunwoo Bang. A glasses-type wearable device for monitoring the patterns of food intake and facial activity. *Scientific Reports*, 7:41690, 01 2017.
- [61] Caroline Lucena, Marcelo Lacerda, Rafael Caldas, Fernando Lima Neto, and Diego Rativa. Mastication evaluation with unsupervised learning: Using an inertial sensor-based system. *IEEE Journal of Translational Engineering in Health and Medicine*, 06:1–10, 04 2018.

- [62] Rui Zhang and Oliver Amft. Monitoring chewing and eating in free-living using smart eyeglasses. *IEEE Journal of Biomedical and Health Informatics*, PP:1–1, 04 2017.
- [63] The top 10 causes of death. <http://www.who.int/mediacentre/factsheets/fs310/en/>.
- [64] Wikipedia. Muscles of mastication. https://en.wikipedia.org/wiki/Muscles_of_mastication, 2018.
- [65] Masseter muscle: Location, actions, and trigger points. <http://www.gustrength.com/muscles:masseter-muscle-actions-and-trigger-points>.
- [66] Lateral pterygoid muscle. https://en.wikipedia.org/wiki/Lateral_pterygoid_muscle.
- [67] Shimmer sensor platform. www.shimmersensing.com.
- [68] Xiaoling Wang. *High Accuracy Distributed Target Detection and Classification in Sensor Networks Based on Mobile Agent Framework*. PhD thesis, Knoxville, TN, USA, 2004.
- [69] MathWorks. Fast fourier transform. <https://www.mathworks.com/help/matlab/ref/fft.html>, 2018.
- [70] J.M.C. Po, Jules Kieser, Luigi Gallo, A.J. Tesenyi, P Herbison, and Mauro Farella. Time-frequency analysis of chewing activity in the natural environment. *Journal of dental research*, 90(10):1206–10, 08 2011.
- [71] Jiawei Han and Micheline Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2000.
- [72] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal. *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2016.

- [73] Wikipedia. Multilayer perceptron. https://en.wikipedia.org/wiki/Multilayer_perceptron, 2018.
- [74] J. Yang, Shuangquan Wang, Ningjiang Chen, Xin Chen, and P. Shi. Wearable accelerometer based extendable activity recognition system. In *2010 IEEE International Conference on Robotics and Automation*, pages 3641–3647, May 2010.
- [75] XULEI YANG, QING SONG, and YUE WANG. A weighted support vector machine for data classification. *International Journal of Pattern Recognition and Artificial Intelligence*, 21(05):961–976, 2007.
- [76] MathWorks. 1-d median filtering. <https://www.mathworks.com/help/signal/ref/medfilt1.html>, 2018.
- [77] Median filter. https://en.wikipedia.org/wiki/Median_filter.
- [78] Y. EL-Manzalawy and V. Honavar. Wlsvm: Integrating libsvm into weka environment. 2005.
- [79] M. S. Rahman and O. J. McCarthy. A classification of food properties. *International Journal of Food Properties*, 2(2):93–99, 1999.
- [80] Alain Woda, K. Foster, Anne Mishellany, and M A Peyron. Adaptation of healthy mastication to factors pertaining to the individual or to the food. *Physiology and behavior*, 89:28–35, 2006.
- [81] C. Loret, M. Walter, N. Pineau, M. A. Peyron, C. Hartmann, and N. Martin. Physical and related sensory properties of a swallowable bolus. *Physiology and Behavior*, 104(5):855–864, June 2011.
- [82] K. Holm. The relations between food structure and sweetness. <https://www.diva-portal.org/smash/get/diva2:943167/FULLTEXT01.pdf>, 2006.

- [83] A. M. Paula and A. C. Conti-Silva. Texture profile and correlation between sensory and instrumental analyses on extruded snacks. *Journal of Food Engineering*, 121:9–14, 2014.
- [84] M. H. Zainudin. Oral physio slides -5.mastication dynamics of occlusion. <https://www.scribd.com/presentation/93171230/Oral-Physio-Slides-5-Mastication-Dynamics-of-Occlusion>, 2019.
- [85] K. Kohyama, T. Sasaki, and F. Hayakawa. Characterization of food physical properties by the mastication parameters measured by electromyography of the jaw-closing muscles and mandibular kinematics in young adults. *Biosci. Biotechnol. Biochem.*, 72(7):1690–1695, 2008.
- [86] V. F. Ferrario, C. Sforza, N. Lovecchio, and F. Mian. Quantification of translational and gliding components in human temporomandibular joint during mouth opening. *Archives of Oral Biology*, 50:507–515, 2005.
- [87] SleepPhones. https://www.amazon.com/AcousticSheep-SleepPhones-Classic-Headphones-Medium/dp/B0046H8ZHS/ref=sr_1_2?s=wireless&ie=UTF8&qid=1540844243&sr=1-2&keywords=sleepphones, 2018.
- [88] Shuangquan Wang, Jie Yang, Ningjiang Chen, Xin Chen, and Qinfeng Zhang. Human activity recognition with user-free accelerometers in the sensor networks. In *2005 International Conference on Neural Networks and Brain*, volume 2, pages 1212–1217, Beijing, China, Oct 2005. IEEE.
- [89] Shimmer. 9dof calibration application, user manual rev 2.10a. http://www.shimmersensing.com/images/uploads/docs/Shimmer_9DOF_Calibration_User_Manual_rev2.10a.pdf, 2017.
- [90] Shuangquan Wang, Gang Zhou, Yongsan Ma, Lisha Hu, Zhenyu Chen, Yiqiang Chen, Hongyang Zhao, and Woosub Jung. Eating detection and chews counting through sensing mastication muscle contraction. *Smart Health*, 07 2018.

- [91] Edward Jay Wang, William Li, Doug Hawkins, Terry Gernsheimer, Colette Norby-Slycord, and Shwetak N. Patel. Hemaapp: Noninvasive blood screening of hemoglobin using smartphone cameras. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, pages 593–604, New York, NY, USA, 2016. ACM.
- [92] MathWorks. Entropy of grayscale image. <https://www.mathworks.com/help/images/ref/entropy.html>, 2018.
- [93] C. Lassauzay, M. A. Peyron, E. Albuisson, E. Dransfield, and A. Woda. Variability of the masticatory process during chewing of elastic model foods. *Eur J Oral Sci*, 108(6):484–492, 2000.
- [94] MathWorks. Standardized z-scores. <https://www.mathworks.com/help/stats/zscore.html>, 2018.
- [95] Riad E. Youssef, Gaylord Throckmorton, Edward Ellis, and Douglas Sinn. Comparison of habitual masticator patterns in men and women using a custom computer program. *The Journal of prosthetic dentistry*, 78:179–86, 09 1997.
- [96] Jordi Martnez-Gomis, Mar Lujan-Climent, Sara Palau, Josep Bizar, Juan Salsench, and Maria Peraire. Relationship between chewing side preference and handedness and lateral asymmetry of peripheral factors. *Archives of oral biology*, 54:101–7, 10 2008.
- [97] P. Lamontagne, Y. Al-Tarakemah, and E. Honkala. Relationship between the preferred chewing side and the angulation of anterior tooth guidance. *Medical Principles and Practice*, 22:545–549, 08 2013.
- [98] Urbano Santana-Mora, Jos Lpez-Cedrn, Mara J. Mora, Xos L. Otero, and Urbano Santana-Penn. Temporomandibular disorders: The habitual chewing side syndrome. *PLoS One*, 8:1–7, 04 2013.

- [99] A.K.J Stomatological Hospital. Harmful effects of unilateral chewing. <http://www.akjidental.com/archives/3515>, 2017.
- [100] S Donnell, Mark Hector, and Ailish Hannigan. Chewing side preferences in children. *Journal of oral rehabilitation*, 31:855–60, 09 2004.
- [101] Accelerometer. <https://en.wikipedia.org/wiki/Accelerometer>.
- [102] Gyroscope. <https://en.wikipedia.org/wiki/Gyroscope>.
- [103] Hajira Basit, Benjamin J. Eovaldi, and Marco A. Siccardi. *Anatomy, Head and Neck, Mastication Muscles*. StatPearls Publishing, 2019.
- [104] Fast fourier transform. <https://www.mathworks.com/help/matlab/ref/fft.html>.
- [105] Tensorflow - cnn and rnn difference. https://www.tutorialspoint.com/tensorflow/tensorflow_cnn_and_rnn_difference.htm.
- [106] Irum Sindhu, Sher Daudpota, Kamal Badar, Maheen Bakhtyar, Junaid Baber, and Mohammad Nurunnabi. Aspect based opinion mining on students feedback for faculty teaching performance evaluation. *IEEE Access*, PP:1–1, 07 2019.
- [107] fullyconnectedlayer. <https://www.mathworks.com/help/deeplearning/ref/nnet.cnn.layer.fullyconnectedlayer.html>.
- [108] softmaxlayer. <https://www.mathworks.com/help/deeplearning/ref/nnet.cnn.layer.softmaxlayer.html>.
- [109] classificationlayer. <https://www.mathworks.com/help/deeplearning/ref/classificationlayer.html>.
- [110] trainingoptions: Options for training deep learning neural network. <https://www.mathworks.com/help/deeplearning/ref/trainingoptions.html>.

- [111] Heming Du. Balance bias and variance: normalization techniques. In *ABCs2018*, 2018.
- [112] Mu Li, Tong Zhang, Yuqiang Chen, and Alexander J. Smola. Efficient mini-batch training for stochastic optimization. In *KDD '14*, 2014.
- [113] Cross-validation. [https://en.wikipedia.org/wiki/Cross-validation_\(statistics\)](https://en.wikipedia.org/wiki/Cross-validation_(statistics)).
- [114] Yijun Chen. *Ant colony optimization approach for stacking configurations*. Master's thesis, Lingnan University, Hong Kong, 2011.
- [115] Classification with imbalanced datasets. <https://sci2s.ugr.es/imbalanced>.
- [116] Erika Gisel. Development of oral side preference during chewing and its relation to hand preference in normal 2- to 8-year-old children. *The American journal of occupational therapy : official publication of the American Occupational Therapy Association*, 42:378–83, 07 1988.
- [117] InvenSense Inc. Mpu6500 product specification (revision 1.1). www.invensense.com, 2014.

VITA

Shuangquan Wang

Shuangquan Wang has been working on his second Ph.D. degree in the Department of Computer Science, College of William & Mary since Fall 2016. He is working with Dr. Gang Zhou in the fields of wearable and mobile computing, smart health, and human activity recognition.

Shuangquan Wang earned a bachelors degree in Computer Science and Technology in 2000 from Wuhan Institute of Technology. Then, he pursued a master's degree in Traffic Information Engineering & Control in 2004 from Wuhan University of Technology. In 2008, he obtained his first Ph.D. degree in Pattern Recognition and Intelligent Systems from Shanghai Jiao Tong University. After that, he worked at the Nokia Research Center (Beijing) as a postdoc researcher for two years. Before joining the College of William & Mary, he worked at the Institute of Computing Technology, Chinese Academy of Science for six years.