Smart Health (2021)

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

Inferring Food Types through Sensing and Characterizing Mastication Dynamics

Shuangquan Wang[a,*], Gang Zhou[b], Jiexiong Guan[b], Yongsen Ma[b], Zhenming Liu[b], Bin Ren[b], Hongyang Zhao[b], Amanda Watson[c], Woosub Jung[b]

[a]*Department of Mathematics & Computer Science, Salisbury University, USA*
[b]*Computer Science Department, William & Mary, USA*
[c]*PRECISE Center, University of Pennsylvania, USA*

ARTICLE INFO

Communicated by S. Sarkar

2000 MSC:
68T05
68T10

Keywords:
Food type recognition
Mastication dynamics
Food properties
Wearable device
Motion sensor

ABSTRACT

Unhealthy dietary structure leads to the prevalence of some chronic diseases, such as obesity, diabetes, and heart disease. Automatic food type recognition helps nutritionists and medical professionals understand patients' nutritional contents, provide accurate and personalized treatments, and evaluate therapeutic effects. Existing wearable sensor-based methods take advantage of microphone, electromyography (EMG), and piezoelectric sensors embedded in the wearable devices. However, these sensors are either easily impacted by ambient acoustic noise or intrusive and uncomfortable to wear. 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. In this paper, we present the first effort in using wearable motion sensors to sense mastication dynamics and infer food types accordingly. 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 and deploy the sensors on the temporalis muscles to sense mastication dynamics accurately and less intrusively. In addition, we extract 65 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 of these 15 human subjects is 82.3%. The accuracy of a single human subject is up to 93.3%.

1. Introduction

As one of the major causes of chronic diseases, unhealthy dietary structure leads to obesity, diabetes, and heart 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 Hales et al. (2017), more than 100 million Americans had diabetes or prediabetes in 2017 CDC (2017), and more than 600,000 Americans died of heart disease in 2009 CDC (2011). Deterioration of the situation forces people to actively monitor their dietary structures. Automatic food type recognition acts as a core function to monitor the dietary structure. For obesity disease, it provides nutritionists objective information to estimate carbohydrate intake amounts; for diabetes disease, it helps medical professionals understand patients' nutritional contents and provide timely feedback to the patients; for heart disease, it promotes patients to select heart-healthy diet and assists physicians to choose the best therapies.

*Corresponding author
e-mail: spwang@salisbury.edu (Shuangquan Wang), gzhou@cs.wm.edu (Gang Zhou), jguan@email.wm.edu (Jiexiong Guan), yma@cs.wm.edu (Yongsen Ma), zliu@cs.wm.edu (Zhenming Liu), bren@cs.wm.edu (Bin Ren), hyzhao@cs.wm.edu (Hongyang Zhao), aawatson@seas.upenn.edu (Amanda Watson), wjung01@email.wm.edu (Woosub Jung)

http://dx.doi.org/10.1016/j.smhl.xxxx.xx.xxx
Received 1 Jan 2021; Received in final form 1 Jan 2021; Accepted 1 Jan 2021; Available online 1 Jan 2021

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, electromyography (EMG), and piezoelectric sensors embedded in the wearable devices. The microphone-based method deploys a microphone sensor in the outer ear Amft et al. (2005) or at the throat area Yatani & Truong (2012). 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 Miyaoka et al. (2013), such as hardness, elasticity, fractorability, 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.

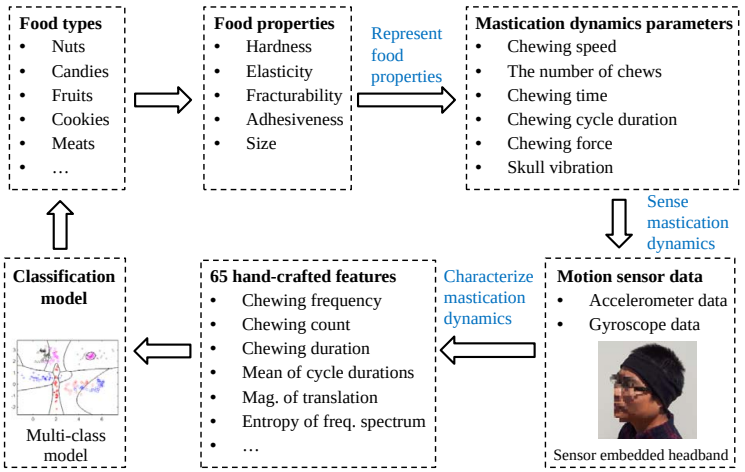

Fig. 1. Motivation of our proposed food type recognition method

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) 65 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. 1. To evaluate the performance of our proposed food type recognition method, we collect 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 of this paper are as follows: 1) 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; 2) We propose to embed motion sensors in a headband and deploy them on the temporalis muscles to sense mastication dynamics accurately and less intrusively; 3) We extract 65 hand-crafted features from each chewing sequence to explicitly characterize the mastication dynamics; 4) 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 82.3%. The recognition accuracy of a single human subject is up to 93.3%.

The remainder of this paper is organized as follows: Section 2 introduces how to represent food properties with mastication dynamics. Section 3 describes how to sense mastication dynamics with motion sensors. Section 4 introduces how to characterize mastication dynamics with hand-crafted features. Section 5 introduces food type classification model. Experiment and evaluation are presented in Section 6. Section 7 presents the comparison with existing wearable sensor-based methods. Related work is introduced in Section 8, and Section 9 presents discussion and future work. Finally, the conclusion is drawn in Section 10.

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.

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 Rahman & McCarthy (1999)." According to the aspects they describe, the food properties are classified into different categories, such as textural properties, tactile properties, appearance properties, rheology properties Rahman et al. (2016), 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 Anthimopoulos et al. (2014). The microphone-based method utilizes the acoustic properties during chewing or swallowing to recognize the food types Amft et al. (2005); Yatani & Truong (2012). Of all the food properties, some are related to the mastication Woda et al. (2006); Miyaoka et al. (2013), such as hardness, elasticity, fracturability, adhesiveness, and size. Hardness indicates the force required to break/chew the product Loret et al. (2011); elasticity describes the ability to deform and go back to its origin state Holm (2006); fracturability describes the ability to break food into pieces when it is bitten Paula & Conti-Silva (2014); adhesiveness indicates the ability of food to adhere to the teeth when chewed Paula & Conti-Silva (2014); size indicates the length, width and height of the food samples.

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 Zainudin (2019). 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 Woda et al. (2006). 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 Kohyama et al. (2008). 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 1 shows the food properties represented by each mastication dynamics parameter Woda et al. (2006); Miyaoka et al. (2013). 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 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

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.

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 Woda et al. (2006); Miyaoka et al. (2013). However, the EMG electrodes are required to be adhered on skin tightly, which is intrusive and uncomfortable. The 3D kinematics method Ferrario et al. (2005) 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 in 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.

3.2. Sensor Deployment

The temporalis is the best mastication muscle to deploy motion sensors. As shown in Fig. 2, there are four groups of mastication muscles: the masseter, the medial pterygoid, the lateral pterygoid, and the temporalis. The masseter is located on the side of the cheek; the medial pterygoid and the lateral pterygoid are located on the inner side of the mandible. Therefore, these three groups of muscles are not suitable for deploying motion sensors. The temporalis is a broad, fan-shaped muscle. It is located at each side of the skull and in front of the ear Strength (2010), 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.

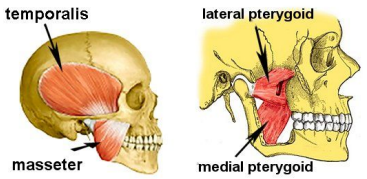

Fig. 2. Mastication muscles Hong (2014)

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 Zhao et al. (2017) shown in Fig. 3 (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.

A headband SleepPhones (2018) 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. 3 (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. 3 (c). Fig. 3 (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 upwards. The Y axis of the left device points backwards, and the Y axis of the right device points forwards. The Z axes of these two devices are vertical to their corresponding X-Y planes and point inside of the skull.

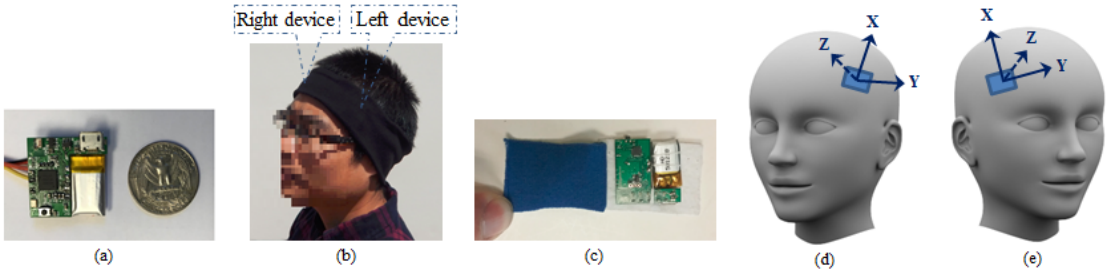

Fig. 3. Wearable device. (a) Hardware platform Zhao et al. (2017); (b) Device deployment; (c) Device covered by two polyster tapes; (d) Sensor orientations of the left device; (e) Sensor orientations of the right device

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. Characterizing Mastication Dynamics

In this section, we first describe motion data preprocessing. Then, we introduce the extraction of the 65 hand-crafted features.

4.1. Motion Data Preprocessing

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

4.1.1. Sensor Calibration

The motion sensor readings may not be accurate because of the scaling and bias errors. Additionally, these two errors vary from sensor to sensor. To eliminate them, we calibrate the accelerometer and gyroscope sensors separately for each device. For the accelerometer, the scaling factor and bias of each axis are calculated to linearly transform the raw sensor readings into the true acceleration outputs Wang et al. (2005); Shimmer (2017). For the gyroscope, the bias of each axis is measured and subtracted from the raw sensor readings. The scaling error of each axis is small. Thus, we ignore it for simplicity.

4.1.2. 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 2.2, 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. 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. 3 (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 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.

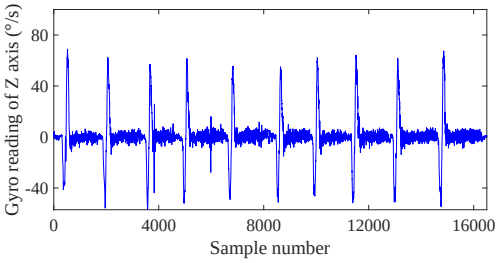

Fig. 4. Gyroscope data of the Z axis during eating

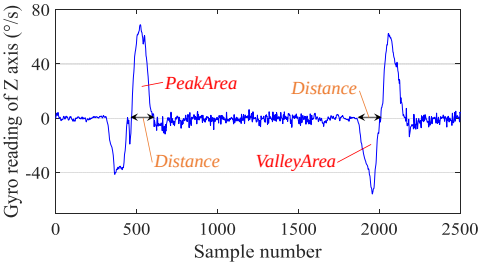

Fig. 5. Three metrics for head motion analysis

We propose three metrics (*PeakArea*, *ValleyArea*, and *Distance* between two neighboring zero-crossing points) to analyze head motions, as shown in Fig. 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.

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.

4.2. Feature Extraction

From each chewing sequence, we extract 65 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.

4.2.1. 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 of this sensor selection method is introduced in Appendix 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 MathWorks (2018a) 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 is taken as the chewing frequency feature. Here, we define the chewing frequency range as $[0.5, 2.5]$ Hz.

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 Wang et al. (2016). 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 MFC and the sampling rate, Rs. The distance threshold between two neighboring peaks, D_{P2P}, is defined as

$$D_{P2P} = \frac{3}{4} \cdot \frac{Rs}{MFC}. \tag{1}$$

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 $[p_1, \cdots, p_n]$. The cycle durations are expressed as $[t_{1,2}, \cdots, t_{n-1,n}]$, 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.

4.2.2. 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 9^{th}-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:

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

where $i = 1, 2, \ldots, L$. 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:

$$T_{Mag} = \frac{S_a^{Left} + S_a^{Right}}{L}, \tag{3}$$

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 and each gyroscope to characterize the skull vibration parameter. These features are calculated from the raw data between the first and last peaks. Let's take the accelerometer data as an example. First, using equation 2, 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 MathWorks (2018c) (without direct current component). From the single-sided amplitude spectrum, the MFC, entropy MathWorks (2018b), 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. Similarly, the same 28 features are extracted from the left and right gyroscopes.

In total, 65 hand-crafted features are extracted from each chewing sequence, as shown in Table 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 2. The 65 hand-crafted features for mastication dynamics characterization

No.	Feature name	Parameter	The data source
1	Chewing frequency	Chewing speed	Selected gyroscope data
2	Chewing count	# of chews	
3	Chewing duration	Chewing time	
4	Sequence length		
5	Maximum of cycle durations	Chewing cycle duration	
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	Left accelerometer (LA)
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	Right accelerometer (RA)
25	Entropy of frequency spectrum		
26	Energy of frequency spectrum		
27	Maximum frequency component		
28-37	Sum of spectrum comp. in 10 bins		
38	Number of mean-crossing	Skull vibration	Left gyroscope (LG)
39	Entropy of frequency spectrum		
40	Energy of frequency spectrum		
41	Maximum frequency component		
42-51	Sum of spectrum comp. in 10 bins		
52	Number of mean-crossing	Skull vibration	Right gyroscope (RG)
53	Entropy of frequency spectrum		
54	Energy of frequency spectrum		
55	Maximum frequency component		
56-65	Sum of spectrum comp. in 10 bins		

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 Woda et al. (2006); Lassauzay et al. (2000) 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 Woda et al. (2006)". This conclusion indicates that our proposed mastication dynamcis-based food type recognition method is valid.

Mastication dynamics-based food type recognition needs a personalized classification model for each subject. These existing works Woda et al. (2006); Lassauzay et al. (2000) also examined the inter-individual variation of mastication dynamics. All the masticatory parameters demonstrated a large variation between individuals Woda et al. (2006). For example, an experiment Lassauzay et al. (2000) 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 Lassauzay et al. (2000). Therefore, we train a personalized food type classification model for each subject.

We choose multilayer perceptron (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 Wikipedia (2018). 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.

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. Finally, we introduce the feature importance analysis.

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 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 3.

Table 3. The users' demographic information and data collection date

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

Table 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

As shown in Table 4, 11 food categories (20 types of food in total) are selected according to the following three criteria: 1) they 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. For some categories, we select multiple food types because their food properties have relatively obvious differences 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.

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 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.

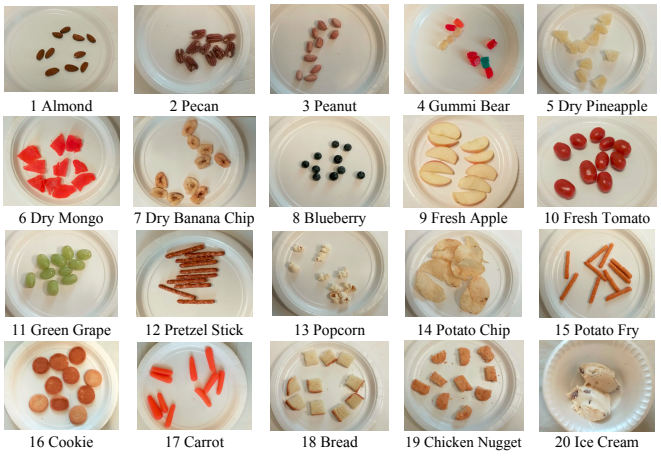

Fig. 6. The 20 types of food served

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. 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.

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 Pinapple Tidbit, seven pieces of Dry Mongo 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.

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 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 8^{th} (the Blueberry) and 20^{th} (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 5^{th}, 9^{th} and 11^{th} 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 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 MathWorks (2018d) 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 MathWorks (2018d). 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.

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 Witten et al. (2016) is used in our experiments. We adopt the default parameters for the MLP classifier in all the following experiments.

Table 6 and Table 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 82.3%. The accuracy of a single user is up to 93.3%; for the recognition of 20 food types, the average accuracy of 15 users is 71.0%. The accuracy of a single user is up to 87.6%. Compared with the average accuracy on the 20 food types, the average accuracy on the 11 food categories increases 11.3%. 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 nine times as accurate as the random classifier on the 11 food categories, and fourteen

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 5^{th}, 9^{th} and 11^{th} food types). Therefore, with a more accurate data segmentation algorithm, the performance of our proposed method could be further improved.

Table 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 (%)	80.6	82.5	93.3	92.3	74.2	85.0	82.2	87.2	86.6	90.6	75.6	65.0	78.5	74.2	86.8	82.3±7.8

Table 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 (%)	73.0	70.8	80.9	87.6	55.6	69.0	69.4	73.8	71.5	79.6	73.1	55.8	70.1	61.3	73.6	71.0±8.6

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 85.5%; the average accuracy of five female users (users 11 to 15) is 76.0%. For the recognition of 20 food types, the average accuracy of ten male users is 73.1%; the average accuracy of five female users is 66.8%. In these two experiments, the average accuracies of the male users are 12.4% and 9.2% higher than those 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 E. Youssef et al. (1997)". The stronger the chewing force, the larger the muscle bulge. Therefore, the motion data are more distinguishable for different food categories and food types.

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. 7. We see that the misclassified samples cluster in several areas. The first area is among the 1^{st}, 2^{nd}, and 3^{rd} food types. These three food types belong to the 1^{st} category (the Nuts). The second area is between the 5^{th} and 6^{th} food types, i.e. the Dry Pineapple Tidbit and Dry Mongo Slice. The third area is between the 10^{th} and 11^{th} food types, i.e. the Fresh Tomato and Green Grape. The fourth area is among the 13^{th}, 14^{th}, and 15^{th} food types. These three food types belong to the 6^{th} 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.

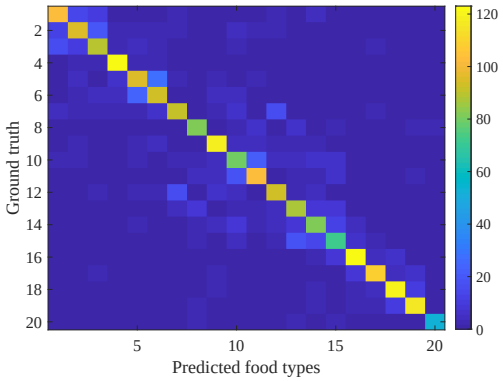

Fig. 7. Sum of the confusion matrices of the 15 users

6.4. Feature Importance Analysis

The 65 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 Witten et al. (2016) to rank all these 65 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 Witten et al. (2016)". Then, from the ranked features, the top ten 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.

Fig. 8 and Fig. 9 show the number of occurrences of the features on the datasets of 11 food categories and 20 food types, respectively. The features whose number of occurrences is greater than or equal to five are marked in these two figures. We see that the 1^{st} (chewing frequency), 3^{rd} (chewing duration), 4^{th} (sequence length), and 8^{th} (magnitude of translation) features have larger numbers of occurrences than all the others features except the 38^{th} feature. In the datasets of 11 food categories, they occurs 5, 6, 6, and 12 times, respectively. In the datasets of 20 food types,

they occurs 9, 7, 8, and 12 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.

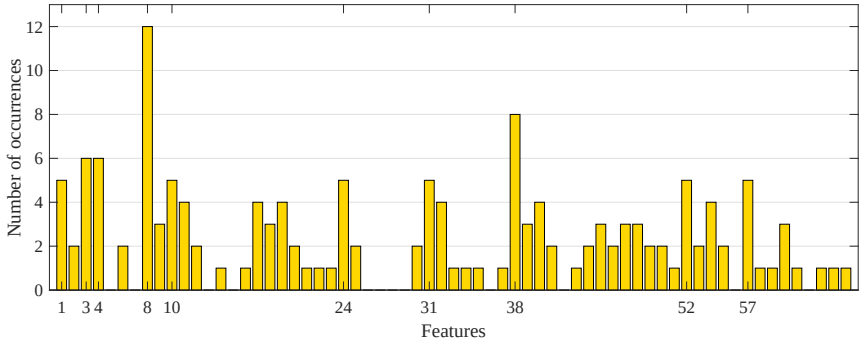

Fig. 8. The number of occurrences of the features in the datasets of 11 food categories. 1: chewing frequency, 3: chewing duration, 4: sequence length, 8: magnitude of translation, 10: number of mean-crossing (LA), 24: number of mean-crossing (RA), 31: sum of spectrum comp. in 4^{th} bin (RA), 38: number of mean-crossing (LG), 52: number of mean-crossing (RG), 57: sum of spectrum comp. in 2^{nd} bin (RG).

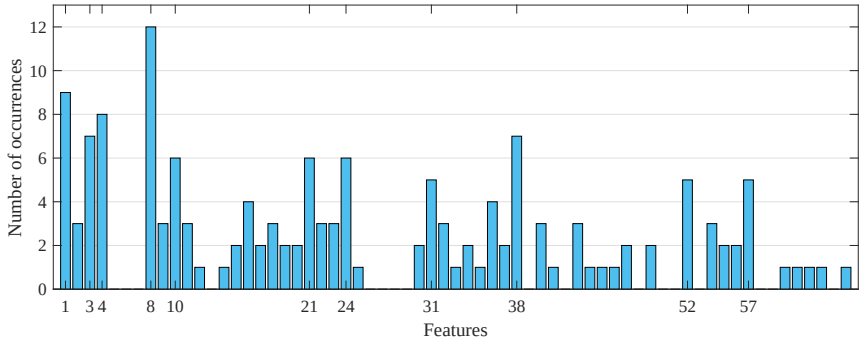

Fig. 9. The number of occurrences of the features in the datasets of 20 food types. 1: chewing frequency, 3: chewing duration, 4: sequence length, 8: magnitude of translation, 10: number of mean-crossing (LA), 21: sum of spectrum comp. in 8^{th} bin (LA), 24: number of mean-crossing (RA), 31: sum of spectrum comp. in 4^{th} bin (RA), 38: number of mean-crossing (LG), 52: number of mean-crossing (RG), 57: sum of spectrum comp. in 2^{nd} bin (RG).

The skull vibration is also important for food type recognition. The 10-65^{th} features characterize skull vibration. Here, six features and seven features each occurs at least five times in the datasets of 11 food categories and 20 food types, respectively. Of these features, four features are common to the datasets of 11 food categories and 20 food types. They are the 10^{th} (number of mean-crossing extracted from the left accelerometer), 24^{th} (number of mean-crossing extracted from the right accelerometer), 38^{th} (number of mean-crossing extracted from the left gyroscope), and 52^{th} (number of mean-crossing extracted from the right gyroscope) features. We see that these four features are the same feature (i.e. number of mean-crossing) extracted from two accelerometers and two gyroscopes separately. The number of mean-crossing feature represents the data fluctuation around the mean, which characterizes the skull vibration.

The mandible motions are less important for food type recognition in our proposed method. The 5-7^{th} features (maximum, mean, and std of cycle durations) characterize the mandible motions. However, these three features are not selected in the datasets of 20 food types. Only the 6^{th} feature occurs twice in the datasets of 11 food types. Though the mandible motions-related features may be helpful for the food type recognition, their contributions are relatively small.

7. Comparison with Existing Wearable Sensor-Based Methods

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 8. For Päßler et al.'s work Päßler et al. (2012) and Bi et al.'s work Bi et al. (2016), they only recognize seven food types. Our method recognizes 11 food categories (in total 20 types of food). Amft et al.'s work Amft & Tröster (2009) 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 Zhang et al. (2016) and Alshurafa et al.'s work Alshurafa et al. (2015), they only recognize 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.

Mirtchouk et al. (2016) present a pioneering work on sensor fusion based food type recognition. They accurately classify 40 food types. However, they require 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 report 62.3% accuracy when only Google Glass is used. However, for their Google Glass only solution, they still require 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. Comparatively, our method only uses motion sensors.

Table 8. Performance comparison of wearable sensor-based methods

	Sensor	**Host object**	**Subjects**	**Food types**	**Accuracy**
Päβler et al. Päßler et al. (2012)	2 microphones	Hearing aid package	51	7 food types & 1 drink	79% on 10% of all records
Bi et al. Bi et al. (2016)	1 microphone	Necklace-like device	12	7 food types	Average accuracy of 84.9%
Amft et al. Amft & Tröster (2009)	1 microphone	Ear-pad	3	19 food types	Accuracy of 80%
Zhang et al. Zhang et al. (2016)	1 EMG sensor	Eyeglass	8	5 food types	63% to 84% for sequences
Alshurafa et al. Alshurafa et al. (2015)	1 piezo. sensor	Necklace	10	2 categories	90% for liquid and solid
Mirtchouk et al. Mirtchouk et al. (2016)	Audio & motion sensors	Glass, earbud, smartwatch	6	40 food types	Accuracy of 82.7%
Our method	2 accel. & 2 gyroscopes	Headband	15	11 cat. (20 food types)	Average accuracy of 82.3%

8. Related Work

The existing food type recognition methods are divided into three categories: image-based methods, ambient sensor-based methods, and wearable sensor-based methods. The image-based methods Sharma et al. (2019); Anthimopoulos et al. (2014); Bossard et al. (2014); Yang et al. (2010); Kawano & Yanai (2013) take advantage of image processing algorithms to extract features from food pictures and build recognition models. This kind of methods can recognize more food items than other existing methods. But 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 Kadomura et al. (2014); Lester et al. (2010); Zhou et al. (2015) often embed sensors in specially designed tablewares to sense the physical or chemical properties of the food. However, these host objects are not convenient to carry and only available at particular locations.

Wearable sensor-based methods embed various sensors into wearable devices. The commonly used sensors include microphone Päßler et al. (2012); Amft et al. (2009); Bi et al. (2016); Amft & Tröster (2009); Amft (2010); Amft et al. (2005); Yatani & Truong (2012); Rahman et al. (2014), EMG Zhang et al. (2016); Zhang & Amft (2016), piezoelectric sensor Kalantarian et al. (2014); Alshurafa et al. (2015), and motion sensor Mirtchouk et al. (2016); Kim et al. (2012). Päßler et al. Päßler et al. (2012) integrate two microphones into a hearing aid package. One is placed in the ear canal to record the chewing sounds. The other is attached on the hearing aid to record the environmental sounds as the reference. A Viterbi algorithm based finite-state grammar (FSG) decoder is 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. Bi et al. (2016) design a microphone embedded hardware prototype. It is 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. Amft & Tröster (2009) embed a miniature microphone into an ear-pad case to recognize 19 standard food types. The reported accuracy is 80%. These above methods need to deploy a microphone in the outer ear Päßler et al. (2012); Amft & Tröster (2009); Amft et al. (2005) or at the throat area Bi et al. (2016); Yatani & Truong (2012). 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. Zhang et al. (2016) embed 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 Zhang & Amft (2016) combines EMG and an accelerometer to monitor chewing and identify a few food types. Kalantarian et al. Kalantarian et al. (2014) embed a piezoelectric sensor into a necklace to distinguish between solid and liquid. Alshurafa et al. Alshurafa et al. (2015) present a similar piezoelectric sensor embedded necklace. The sensor is 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 Ye et al. (2015); Mertes et al. (2015); Rahman et al. (2015); Biallas et al. (2015); Farooq & Sazonov (2018); Chun et al. (2018); Zhang et al. (2020, 2018) and count the number of chews Wang et al. (2015). However, there are only a few works on the motion sensor based food type recognition. Kim et al. Kim et al. (2012) utilize an accelerometer embedded wristband to detect 29 eating actions. The detection results are used to indirectly infer two different food types, the rice and noodle. Mirtchouk et al. Mirtchouk et al. (2016) combine 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 are 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.

9. Discussion and Future Work

There are a few practical guides for the deployment of our proposed system. 1) Choosing appropriate deployment locations for the left and right devices. The temporalis is a broad muscle. The muscle bulges at different locations may be different. According to our experience, the proposed system performs well at the locations where the muscle bulge is obvious. Therefore, we suggest identifying an appropriate location

for each device before deploying it. 2) Keeping the device locations and orientations consistent. An obvious location or orientation change may degrade the recognition performance. Thus, the length of the headband should be suitable for the user's head circumference. If the headband is too loose, the device locations may shift over time; if the headband is too tight, it is uncomfortable to wear.

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.

In our user study, some types of food are cut into equal pieces for the convenience of food intake. We are aware that a user can bite food by mouth in real life and the food amount of each bite may vary to some degree. Different food bite amount impacts several chewing duration-related features, such as chewing count, chewing duration, and sequence length. However, it has little impact on the majority of features, which are related to chewing speed, chewing force, and skull vibration. Thus, the classification accuracy may be reduced a little but not obviously. 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 swallowing and before next biting. These two noisy activities can be filtered using an eating/chewing detection module Chongguang Bi et al. (2017); Sen et al. (2018); Wang et al. (2020). We plan to integrate this module into our proposed method in the future.

Our proposed method can be generalized to users of different ages. Although senior persons or children are not included in our user study, we believe that the user age difference has little impact on the recognition performance of our proposed method. The differences in mastication dynamics are mainly from two impact factors: the inter-individual mastication variations and the differences in food properties (hardness, elasticity, fracturability, adhesiveness, and size). As described in section 5, our proposed method eliminates the first impact factor by training a personalized food type classification model for each subject. This personalized classification model specializes in distinguishing mastication dynamics of different food properties and only applies to the specific user himself/herself.

10. Conclusion

In this paper, 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 65 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 82.3%. The accuracy of a single human subject is up to 93.3%.

Appendix 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. We observe 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.

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. We first filter the data of each axis using a 9^{th}-order one-dimension median filter MathWorks (2018a) to reduce the noise. Then, we conduct Fourier transform on the filtered data of each axis to obtain its single-sided amplitude spectrum MathWorks (2018c) (without direct current component). 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,

$$R_{MFC} = \frac{MFC}{Sum(C_i)}, \tag{A.1}$$

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.

Acknowledgments

Special thanks go to all the volunteers and all the anonymous reviewers. This work was supported by NSF CNS-1841129.

References

Alshurafa, N., Kalantarian, H., Pourhomayoun, M., Liu, J. J., Sarin, S., Shahbazi, B., & Sarrafzadeh, M. (2015). Recognition of nutrition intake using time-frequency decomposition in a wearable necklace using a piezoelectric sensor. *IEEE Sensors Journal*, *15*, 3909–3916.

Amft, O. (2010). A wearable earpad sensor for chewing monitoring. In *2010 IEEE Sensors* (pp. 222–227).

Amft, O., Kusserow, M., & TrÃŰster, G. (2009). Bite weight prediction from acoustic recognition of chewing. *IEEE Transactions on Biomedical Engineering*, *56*, 1663–1672.

Amft, O., Stäger, M., Lukowicz, P., & Tröster, G. (2005). Analysis of chewing sounds for dietary monitoring. In *UbiComp'05* (pp. 56–72).

Amft, O., & Tröster, G. (2009). On-body sensing solutions for automatic dietary monitoring. *IEEE Pervasive Computing*, *8*, 62–70.

Anthimopoulos, M. M., Gianola, L., Scarnato, L., Diem, P., & Mougiakakou, S. G. (2014). A food recognition system for diabetic patients based on an optimized bag-of-features model. *IEEE JBHI*, *18*, 1261–1271.

Bi, Y., Lv, M., Song, C., Xu, W., Guan, N., & Yi, W. (2016). Autodietary: A wearable acoustic sensor system for food intake recognition in daily life. *IEEE Sensors Journal*, *16*, 806–816.

Biallas, M., Andrushevich, A., Kistler, R., Klapproth, A., Czuszynski, K., & Bujnowski, A. (2015). Feasibility study for food intake tasks recognition based on smart glasses. *Journal of Medical Imaging and Health Informatics*, *5*, 1688–1694.

Bossard, L., Guillaumin, M., & Van Gool, L. (2014). Food-101 – mining discriminative components with random forests. In *Computer Vision – ECCV 2014* (pp. 446–461). Cham: Springer International Publishing.

CDC (2011). Deaths: Final data for 2009. URL: `www.cdc.gov/nchs/data/nvsr/nvsr60/nvsr60_03.pdf`.

CDC (2017). National diabetes statistics report, 2017. URL: `https://www.cdc.gov/diabetes/data/statistics/statistics-report.html`.

Chongguang Bi, Xing, G., Hao, T., Jina Huh, Wei Peng, & Mengyan Ma (2017). Familylog: A mobile system for monitoring family mealtime activities. In *PerCom '2017* (pp. 21–30).

Chun, K. S., Bhattacharya, S., & Thomaz, E. (2018). Detecting eating episodes by tracking jawbone movements with a non-contact wearable sensor. *IMWUT*, *2*, 4:1–4:21.

E. Youssef, R., Throckmorton, G., Ellis, E., & Sinn, D. (1997). Comparison of habitual masticator patterns in men and women using a custom computer program. *The Journal of prosthetic dentistry*, *78*, 179–86. doi:`10.1016/S0022-3913(97)70123-9`.

Farooq, M., & Sazonov, E. (2018). Accelerometer-based detection of food intake in free-living individuals. *IEEE Sensors Journal*, *18*, 3752–3758.

Ferrario, V. F., Sforza, C., Lovecchio, N., & Mian, F. (2005). Quantification of translational and gliding components in human temporomandibular joint during mouth opening. *Archives of Oral Biology*, *50*, 507–515.

Hales, C. M., Carroll, M. D., Fryar, C. D., & Ogden, C. L. (2017). *Prevalence of Obesity Among Adults and Youth: United States, 2015-2016*. Technical Report Hyattsville, MD, USA.

Holm, K. (2006). The relations between food structure and sweetness. URL: `https://www.diva-portal.org/smash/get/diva2:943167/FULLTEXT01.pdf`.

Hong, K. A. (2014). The muscles of mastication. URL: `https://www.thousandoaksfamilydentistry.com/blog/2014/12/22/the-muscles-of-mastication`.

Kadomura, A., Li, C.-Y., Tsukada, K., Chu, H.-H., & Siio, I. (2014). Persuasive technology to improve eating behavior using a sensor-embedded fork. In *UbiComp '14* (pp. 319–329).

Kalantarian, H., Alshurafa, N., & Sarrafzadeh, M. (2014). A wearable nutrition monitoring system. In *BSN'2014* (pp. 75–80).

Kawano, Y., & Yanai, K. (2013). Real-time mobile food recognition system. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1–7).

Kim, H. J., Kim, M., Lee, S. J., & Choi, Y. S. (2012). 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* (pp. 1–5).

Kohyama, K., Sasaki, T., & Hayakawa, F. (2008). 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*, 1690–1695.

Lassauzay, C., Peyron, M. A., Albuisson, E., Dransfield, E., & Woda, A. (2000). Variability of the masticatory process during chewing of elastic model foods. *Eur J Oral Sci*, *108*, 484–492.

Lester, J., Tan, D., Patel, S., & Brush, A. J. B. (2010). Automatic classification of daily fluid intake. In *PervasiveHealth'2010* (pp. 1–8).

Loret, C., Walter, M., Pineau, N., Peyron, M. A., Hartmann, C., & Martin, N. (2011). Physical and related sensory properties of a swallowable bolus. *Physiology and Behavior*, *104*, 855–864.

MathWorks (2018a). 1-d median filtering. URL: `https://www.mathworks.com/help/signal/ref/medfilt1.html`.

MathWorks (2018b). Entropy of grayscale image. URL: `https://www.mathworks.com/help/images/ref/entropy.html`.

MathWorks (2018c). Fast fourier transform. URL: `https://www.mathworks.com/help/matlab/ref/fft.html`.

MathWorks (2018d). Standardized z-scores. URL: `https://www.mathworks.com/help/stats/zscore.html`.

Mertes, G., Hallez, H., Croonenborghs, T., & Vanrumste, B. (2015). 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* (pp. 1–5).

Mirtchouk, M., Merck, C., & Kleinberg, S. (2016). Automated estimation of food type and amount consumed from body-worn audio and motion sensors. In *UbiComp '16* (pp. 451–462).

Miyaoka, Y., Ashida, I., Tamaki, Y., Kawakami, S., Iwamori, H., Yamazaki, T., & Ito, N. (2013). Quantitative analysis of relationships between masseter activity during chewing and textural properties of foods. *Food and Nutrition Sciences*, *4*, 144–149.

Päßler, S., Wolff, M., & Fischer, W.-J. (2012). Food intake monitoring: an acoustical approach to automated food intake activity detection and classification of consumed food. *Physiol Meas.*, *33*, 1073–93.

Paula, A. M., & Conti-Silva, A. C. (2014). Texture profile and correlation between sensory and instrumental analyses on extruded snacks. *Journal of Food Engineering*, *121*, 9–14.

Rahman, M. S., & McCarthy, O. J. (1999). A classification of food properties. *International Journal of Food Properties*, *2*, 93–99.

Rahman, S. A., Merck, C., Huang, Y., & Kleinberg, S. (2015). Unintrusive eating recognition using google glass. In *PervasiveHealth '15* (pp. 108–111). Istanbul, Turkey.

Rahman, T., Adams, A. T., Schein, P., Jain, A., Erickson, D., & Choudhury, T. (2016). Nutrilyzer: A mobile system for characterizing liquid food with photoacoustic effect. In *SenSys âĂŹ16* (p. 123âĂŞ136).

Rahman, T., Adams, A. T., Zhang, M., Cherry, E., Zhou, B., Peng, H., & Choudhury, T. (2014). Bodybeat: A mobile system for sensing non-speech body sounds. In *MobiSys '14* (pp. 2–13).

Sen, S., Subbaraju, V., Misra, A., Balan, R., & Lee, Y. (2018). Annapurna: Building a real-world smartwatch-based automated food journal. In *2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)* (pp. 1–6).

Sharma, A., Misra, A., Subramaniam, V., & Lee, Y. (2019). Smrtfridge: Iot-based, user interaction-driven food item & quantity sensing. In *SenSys '19* (p. 245âĂŞ257). doi:`10.1145/3356250.3360028`.

Shimmer (2017). 9dof calibration application, user manual rev 2.10a. URL: `http://www.shimmersensing.com/images/uploads/docs/Shimmer_9DOF_Calibration_User_Manual_rev2.10a.pdf`.

SleepPhones (2018). URL: `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`.

Strength, G. U. (2010). Temporalis muscle: Location, action and trigger points. URL: `http://www.gustrength.com/muscles:temporalis-location-action-and-trigger-points`.

Wang, C., Lin, Z., Xie, Y., Guo, X., Ren, Y., & Chen, Y. (2020). Wieat: Fine-grained device-free eating monitoring leveraging wi-fi signals. *ArXiv*, *abs/2003.09096*.

Wang, E. J., Li, W., Hawkins, D., Gernsheimer, T., Norby-Slycord, C., & Patel, S. N. (2016). Hemaapp: Noninvasive blood screening of hemoglobin using smartphone cameras. In *UbiComp '16* (pp. 593–604).

Wang, S., Yang, J., Chen, N., Chen, X., & Zhang, Q. (2005). Human activity recognition with user-free accelerometers in the sensor networks. In *2005 International Conference on Neural Networks and Brain* (pp. 1212–1217). volume 2.

Wang, S., Zhou, G., Hu, L., Chen, Z., & Chen, Y. (2015). Care: Chewing activity recognition using noninvasive single axis accelerometer. In *UbiComp/ISWC'15 Adjunct* (pp. 109–112).

Wikipedia (2018). Multilayer perceptron. URL: `https://en.wikipedia.org/wiki/Multilayer_perceptron`.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Woda, A., Foster, K., Mishellany, A., & Peyron, M. A. (2006). Adaptation of healthy mastication to factors pertaining to the individual or to the food. *Physiology and behavior*, *89*, 28–35.

Yang, S., Chen, M., Pomerleau, D. A., & Sukthankar, R. (2010). Food recognition using statistics of pairwise local features. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (pp. 2249–2256).

Yatani, K., & Truong, K. N. (2012). Bodyscope: A wearable acoustic sensor for activity recognition. In *UbiComp '12* (pp. 341–350).

Ye, X., Chen, G., & Cao, Y. (2015). Automatic eating detection using head-mount and wrist-worn accelerometers. In *2015 17th International Conference on E-health Networking, Application Services (HealthCom)* (pp. 578–581).

Zainudin, M. H. (2019). Oral physio slides -5.mastication dynamics of occlusion. URL: `https://www.scribd.com/presentation/93171230/Oral-Physio-Slides-5-Mastication-Dynamics-of-Occlusion`.

Zhang, R., & Amft, O. (2016). Regular-look eyeglasses can monitor chewing. In *UbiComp '16* (pp. 389–392).

Zhang, R., Bernhart, S., & Amft, O. (2016). Diet eyeglasses: Recognising food chewing using emg and smart eyeglasses. In *BSN '2016* (pp. 7–12).

Zhang, S., Stogin, W., & Alshurafa, N. (2018). I sense overeating: Motif-based machine learning framework to detect overeating using wrist-worn sensing. *Information Fusion*, *41*, 37 – 47.

Zhang, S., Zhao, Y., Nguyen, D., Xu, R., Sen, S., Hester, J., & Alshurafa, N. (2020). Necksense: A multi-sensor necklace for detecting eating activities in free-living conditions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*, 1–26. doi:`10.1145/3397313`.

Zhao, H., Wang, S., Zhou, G., & Zhang, D. (2017). Gesture-enabled remote control for healthcare. In *CHASE '17* (pp. 392–401).

Zhou, B., Cheng, J., Lukowicz, P., Reiss, A., & Amft, O. (2015). Monitoring dietary behavior with a smart dining tray. *IEEE Pervasive Computing*, *14*, 46–56.