Smart Health (2021)

Contents lists available at ScienceDirect

Smart Health

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

Wearable motion sensor-based chewing side detection

Shuangquan Wang[a,*], Gang Zhou[b], Amanda Watson[c], Lei Xie[d], Minglong Sun[b], 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
[d] State Key Laboratory for Novel Software Technology, Nanjing University, China

ARTICLE INFO

Communicated by S. Sarkar

2000 MSC:
68T05
68T10

Keywords:
chewing side preference
wearable motion sensors
muscle bulge
skull vibration
LSTM

ABSTRACT

Chewing side preference means a tendency to use one side to chew food more frequently than the other. Medical studies show that chewing side preference can result in lateral facial asymmetry, teeth abrasion, temporomandibular disorders, malocclusion, and stomach illness. To continuously detect chewing side preference and quantify its severity in daily life, several wearable sensor-based methods have been proposed in recent years. However, these methods are either intrusive or not fine-grained enough. In this paper, we propose a wearable motion sensor-based chewing side detection method. We observe that chewing activity generates mastication muscle bulge and skull vibration, which can be sensed by motion sensors worn on the mastication muscles. In addition, the muscle bulge and skull vibration of the chewing side are different from those of the non-chewing side. These observations motivate us to deploy motion sensors on the left and right temporalis muscles to detect chewing sides. We propose 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 are then calculated to characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side. A two-class classifier is trained using long short-term memory (LSTM), an artificial recurrent neural network, to model the data samples and classify chewing sides. 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%.

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 Tiwari et al. (2017). This tendency causes lateral asymmetry of chewing force and occlusal contact area Martinez-Gomis et al. (2008). Accordingly, it can lead to several diseases, such as lateral facial asymmetry Tiwari et al. (2017), teeth abrasion Lamontagne et al. (2013), temporomandibular disorders (TMD) Martinez-Gomis et al. (2008); Santana-Mora et al. (2013), malocclusion, and stomach illness AKJDental (2017). Clinical studies have demonstrated that the majority of people have a chewing side preference. Donnell et al. Donnell et al. (2004) 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 Donnell et al. (2004)". Tiwari et al. Tiwari et al. (2017) 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.

*Corresponding author
e-mail: spwang@salisbury.edu (Shuangquan Wang), gzhou@cs.wm.edu (Gang Zhou), aawatson@seas.upenn.edu (Amanda Watson), lxie@nju.edu.cn (Lei Xie), msun05@email.wm.edu (Minglong Sun), wjung01@email.wm.edu (Woosub Jung)

http://dx.doi.org/10.1016/j.smhl.xxxx.xx.xxx
Received 29 November 2020; Received in final form 1 January 2021; Accepted 1 January 2021; Available online 1 January 2021

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. Chung et al. (2017) 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. Lucena et al. (2018) 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 and thus not convenient to use in daily living.

One research problem is how to detect chewing sides accurately and less intrusively using wearable sensors. To solve this problem, 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 challenges. The first challenge is how to segment each chew from the continuous eating data. To address this challenge, we propose a heuristic-rules based method to exclude non-chewing data and segment each chew accurately. The second challenge is how to characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side. To address this challenge, 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. The third challenge is how to model and classify multi-dimensional data samples with unequal input lengths. To address this challenge, 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 types is collected from eight human subjects. The experimental results show that our proposed method is very promising.

The main contributions of this paper 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 differentiate muscle bulge and skull vibration between the chewing side and the non-chewing side.
2. To accurately identify each chew and recognize its chewing side, we propose a heuristic-rules based method to exclude non-chewing data and segment each chew from continuous eating data. To characterize the difference of muscle bulge and skull vibration between the chewing side and the non-chewing side, we propose to calculate the relative difference series of the left and right sensors.
3. 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 paper is organized as follows. Section 2 describes the system overview of our proposed method. Section 3 describes the motion sensor deployment. In Section 4, we describe the motion data collection and calibration. Section 5 describes the data segmentation and classification sample generation. In Section 6, we describe the chewing side classification with LSTM. Experiment and evaluation are presented in Section 7. Related work is introduced in Section 8, and Section 9 presents discussion. Finally, we draw our conclusion in Section 10.

2. System Overview

Fig. 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.

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 Wikipedia (2020d). Accordingly, the muscle bulges to some degree, which causes the translation and rotation movements of the sensors. The larger the chewing force, the greater the translation and rotation. The translation movements result in velocity changes that are directly sensed by an accelerometer Wikipedia (2020a). The rotation movements result in changes of orientation and angular velocity that are directly sensed by a gyroscope Wikipedia (2020c). 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.

Temporalis 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 Wikipedia (2020e). 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. As shown by the green colored muscle in Fig. 2, the temporalis

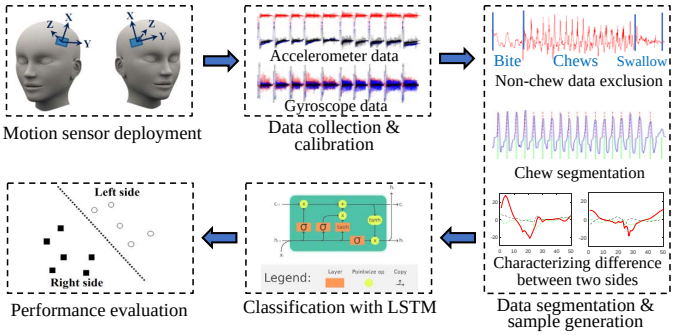

Fig. 1: Flowchart of our proposed chewing side detection method

is located at each side of the skull and in front of the ear TriggerPointSelfHelp.com (2020), where we often wear a headband, eye-glasses, or hat. This observation motivates us to embed motion sensors into a headband and deploy the sensors on the temporalis muscles.

Deploying sensors on the anterior temporalis is appropriate to detect chewing sides van der Bilt et al. (2008); MacDougall & Andrew (1953); Balkhi et al. (1993); da Silva Andrade et al. (2010). The temporalis is a fan-shaped muscle, which is divided into three parts: anterior fibers, mid fibers, and posterior fibers. The anterior fibers have a vertical orientation, mid fibers have an oblique orientation, and posterior fibers have more of a horizontal orientation Basit et al. (2019). Contraction of the anterior fibers results in elevation of the mandible (closing the mouth); contraction of the posterior fibers results in retrusion of the mandible; and the mid fibers are responsible for both elevation and retraction of the mandible Basit et al. (2019); Wikipedia (2021). As chewing activity mainly involves depression and elevation of the mandible, anterior temporalis muscles are suitable for sensing the muscle activity difference between the left side and right side and inferring chewing sides accordingly.

Considering the above observations, we deploy the left device at location A in Fig. 2, which is on the left anterior temporalis and near the boundary of the skull and the frontier of the left anterior temporalis. This location is similar to the EMG electrode location on the anterior temporalis in reference MacDougall & Andrew (1953). Similarly, the right device is deployed at the symmetrical location on the right temporalis muscle.

When the anterior temporalis contracts and bulges, the motion sensors mainly rotate around the X axis at location A. The blue arrows in Fig. 2 show the orientations of the X axis of the left accelerometer and gyroscope at locations A, B, and C, which are on the anterior, mid, and posterior fibers, respectively. Fig. 3 shows the gyroscope readings of X, Y, and Z axes at these three locations. We see that the amplitudes of gyroscope readings at location A concentrate on X axis. In addition, the amplitudes of gyroscope readings at location A are larger than those at locations B and C. This is because the anterior fibers are stronger than the mid and posterior fibers. The stronger the fibers, the larger the chewing force and muscle bulge, and the greater amplitudes of gyroscope readings.

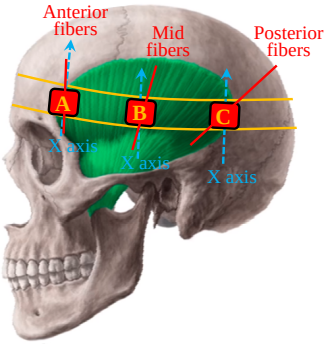

Fig. 2: Muscle fibers and sensor location

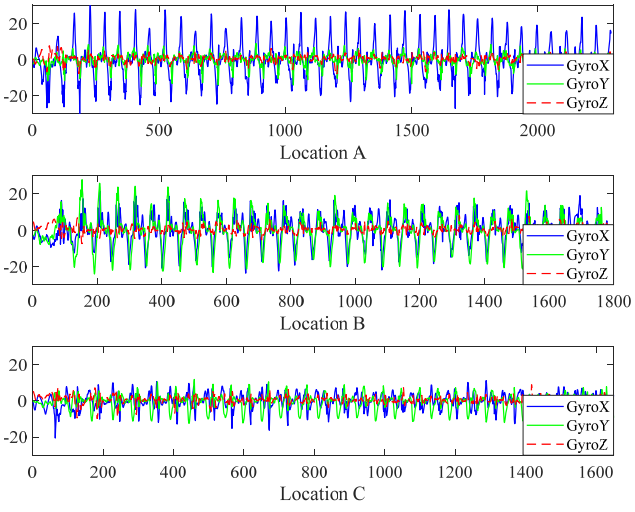

Fig. 3: Gyroscope readings at three locations

4. Motion Data Collection and Calibration

In this section, we first introduce motion data collection using wearable devices. Then, we describe how to calibrate data to eliminate the scaling and bias errors.

4.1. Motion Data Collection

Two small-size wearable devices Zhao et al. (2017) shown in Fig. 4 (a) are used to collect motion data. Each device contains a 3-axis accelerometer, a 3-axis gyroscope, and a 3-axis digital compass. We embed these two devices into a headband and deploy them on the left and right temporalis muscles, as shown in Fig. 4 (b). Fig. 4 (c) and (d) show the sensor orientations of the left and right devices from the subject's perspective. The X axes point upward; the Y axis of the left device points backward, and the Y axis of the right device points forward; the Z axes are perpendicular to their corresponding X-Y planes and point inside of the skull.

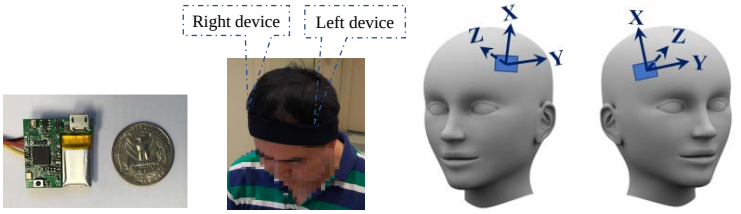

Fig. 4: Wearable device. (a) Sensing device Zhao et al. (2017); (b) Device deployment; (c) Sensor orientations of the left device; (d) Sensor orientations of the right device

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, folk, 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 Bluetooth Low Energy (BLE) in real-time. Then, the data are transferred to a PC for offline analysis.

4.2. 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.

Fig. 5 (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 Wang et al. (2005). 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 Shimmer (2017):

$$a_X = k_X \cdot m_X + b_X. \tag{1}$$

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

$$k_X = \frac{2 \cdot g}{m_X^+ - m_X^-}, \tag{2}$$

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

$$b_X = -\frac{k_X \cdot (m_X^+ + m_X^-)}{2}. \tag{3}$$

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. 5 (a) are shown in Fig. 5 (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.

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.

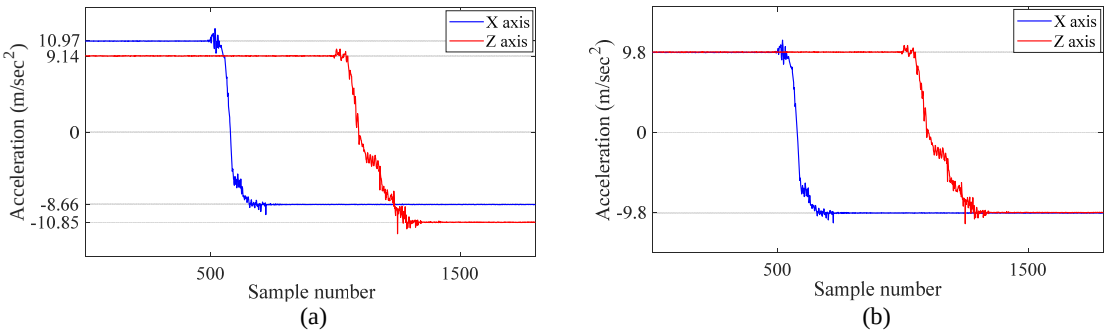

Fig. 5: Accelerometer data of the X axis and Z axis. (a) Before calibration; (b) After calibration

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 Fig. 4 (c) and (d), 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. Fig. 6 (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. Fig. 6 (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 Fig. 6 (a) - (c).

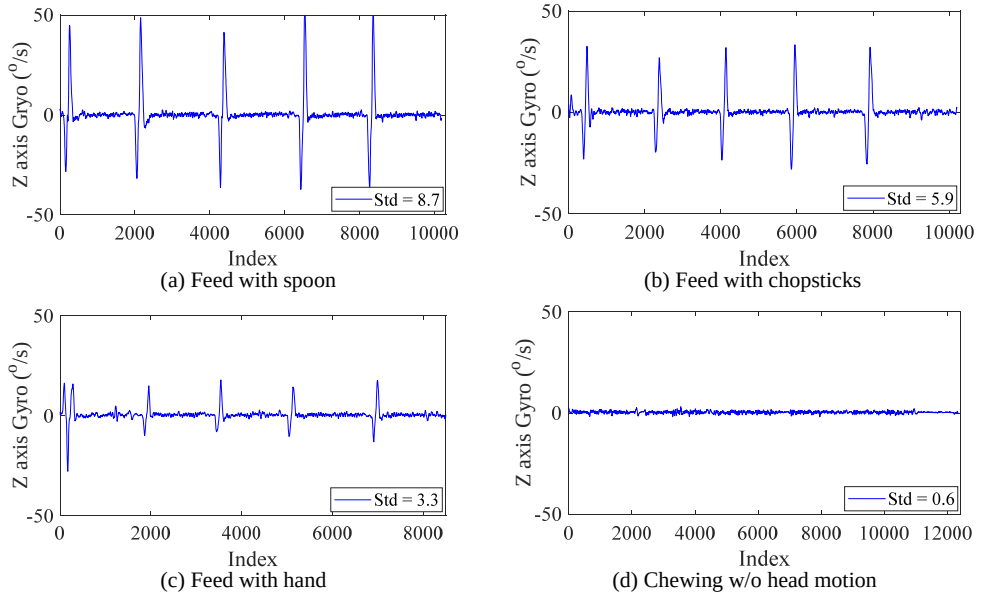

Fig. 6: Filtered Z axis gyroscope data

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. μ and σ, of the filtered Z axis gyroscope data. Then, we set two thresholds, $\mu - \sigma$ and $\mu + \sigma$, 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.

5.2. Swallowing Data Exclusion

The X axis of the left gyroscope is utilized to differentiate chewing data and swallowing data. As shown in Fig. 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. Fig. 7 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.

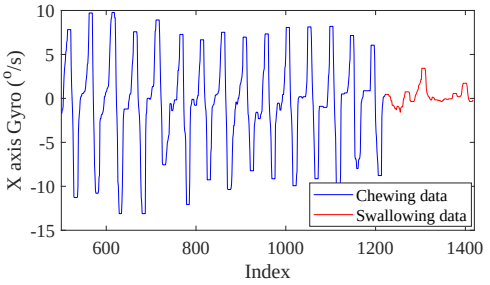

Fig. 7: Filtered X axis gyroscope data during chewing and swallowing

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 Fig. 8 (a), satisfies the following inequality for the chewing data:

$$\frac{3}{4} \cdot \frac{Rs}{f_{MFC}} \leq D_{P2P} \leq \frac{3}{2} \cdot \frac{Rs}{f_{MFC}} \tag{4}$$

where Rs is the sampling rate, and f_{MFC} means the chewing frequency. It is the frequency corresponding to the maximum frequency component (MFC) in the chewing frequency range. Here, the chewing frequency range is defined as $[0.5, 2.5]$ Hz Wang et al. (2018). $\frac{Rs}{f_{MFC}}$ represents the cycle length of one chew. If the distance between two neighboring peaks is less than $\frac{3}{4}$ cycle length or larger than $\frac{3}{2}$ cycle length, the data between these two peaks are considered as swallowing data and dropped.

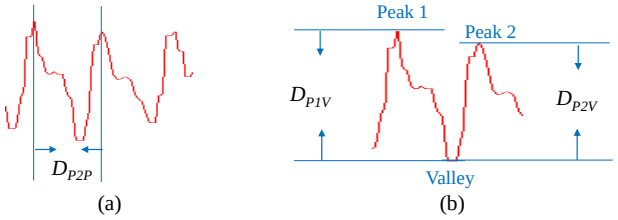

Fig. 8: Distance between peaks and amplitude difference

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 8 (b), satisfy the following two inequalities for the chewing data:

$$\begin{cases} D_{P1V} \geq T_{PV} \\ D_{P2V} \geq T_{PV} \end{cases} \tag{5}$$

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.

5.3. 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 15^{th}-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 Mathworks (2020b) (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 (4) and (5), 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.

5.4. 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 the following equation:

$$a_i = \sqrt{(a_X^i)^2 + (a_Y^i)^2 + (a_Z^i)^2}, \tag{6}$$

Here, $i = 1, 2, ..., n$. n is the data length of one chew segment. The resulting scalar acceleration series is $A = [a_1, a_2, ..., a_n]$. 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 Mathworks (2020g). The z-score algorithm normalizes each scalar acceleration series so that all the data in this series have mean 0 and standard deviation 1 Mathworks (2020g). 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.

6. Chewing Side Classification with LSTM

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

6.1. Biomechanical Evidence for Chewing Side Classification

Existing works provide strong evidence to support our proposed chewing side detection method. A clinical study van der Bilt et al. (2008) demonstrates that the left and right temporalis muscles have significant activity difference during unilateral chewing. This study selects 81 healthy subjects (13 men and 68 women) with a natural dentition to investigate whether an asymmetric task will introduce asymmetry in the muscle activities of masseter and anterior temporal muscles. The results demonstrate that "The masseter muscles showed no difference in activity between the ipsilateral side and the contralateral side during unilateral clenching. In contrast, the activity of the anterior temporal muscle on the ipsilateral side was significantly higher than on the contralateral side van der Bilt et al. (2008)." This conclusion indicates that our proposed chewing side detection method is valid.

6.2. Why Personalized Classification Model?

A subject's chewing patterns are consistent and can be accurately modeled. Clinical studies Woda et al. (2006); Lassauzay et al. (2000) 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 Woda et al. (2006)." 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 Woda et al. (2006); Lassauzay et al. (2000). For example, Lassauzay et al. Lassauzay et al. (2000) 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.

6.3. 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 Mathworks (2020d). In addition, it can handle arbitrary input lengths Tutorialspoint (2020). The classification samples obtained in Section 5 are temporal data with unequal lengths. Therefore, we utilize LSTM to model the data and classify chewing sides accordingly.

6.4. 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 Mathworks (2020d). 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 in this paper, 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 Fig. 9, we adopt a five-layer architecture introduced in the Matlab deep learning toolbox Mathworks (2020d,e): 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 Sindhu et al. (2019), 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 Mathworks (2020c). 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 Mathworks (2020f); 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 Mathworks (2020a) and then assigns a predication label (i.e. left side or right side) for each classification sample.

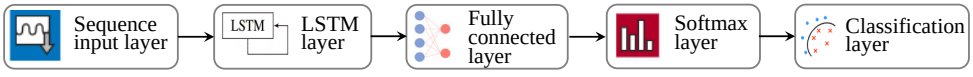

Fig. 9: Architecture of the five-layer LSTM model Mathworks (2020d,e)

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 Mathworks (2020h). The existing research results Du (2018); Li et al. (2014) 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 Mathworks (2020e) 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. Fig. 10 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 Fig. 10. 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 Mathworks (2020e) in ascending order before padding zeros. Fig. 11 shows the padding effect on the sorted classification samples. We see that the number of padded zeros are greatly reduced.

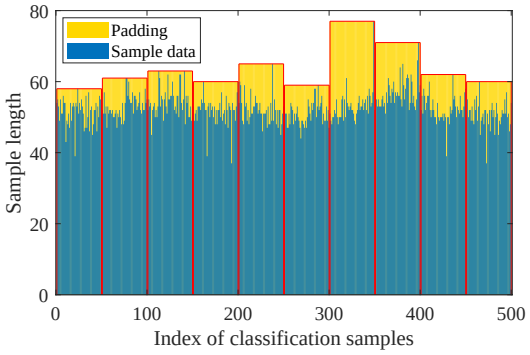

Fig. 10: Padding effect on unsorted samples

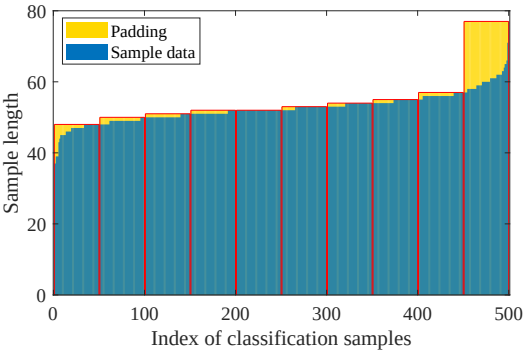

Fig. 11: Padding effect on sorted samples

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.

7.1. Experimental Setup

With the approval from the institutional review board (IRB), eight healthy human subjects (five male users and three female users) were recruited in our user study. Their demographic information is shown in Table 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 4. After initial deployment, the sensor locations may be slightly adjusted if the collected data are too noisy.

Table 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

Eight different food types are included in our user study. They are almond, cookie, 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 Fig. 12. The users eat one piece at a time. These eight foods are delivered and fed four different ways. The 1^{st} and 2^{nd} food types are fed with a spoon; the 3^{rd} and 4^{th} food types are fed with a fork; the 5^{th} and 6^{th} food types are fed with chopsticks, and the 7^{th} and 8^{th} food types are fed with a hand. Users 6 and 7 do not like eating gummi bear. They are allowed to skip it.

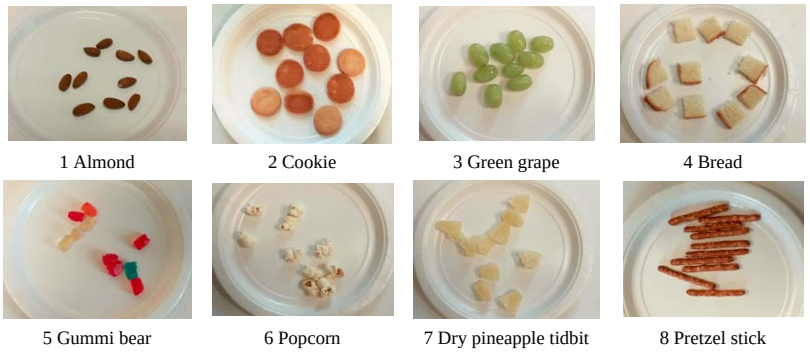

Fig. 12: The food served to each user

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 9.

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.

7.2. Performance of Chew Segmentation

The first question we would like to answer is how accurate the proposed chew segmentation method is. Table 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 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 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) and hence are dropped.

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 Wikipedia (2020b); Chen (2011) 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 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 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 3: Chewing side detection accuracy

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

Fig. 13 (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 Fig. 13 (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 WebmasterTeam (2020)".

Figure 13 (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 Fig. 13 (a). This indicates that our proposed method is valid for both male users and female users.

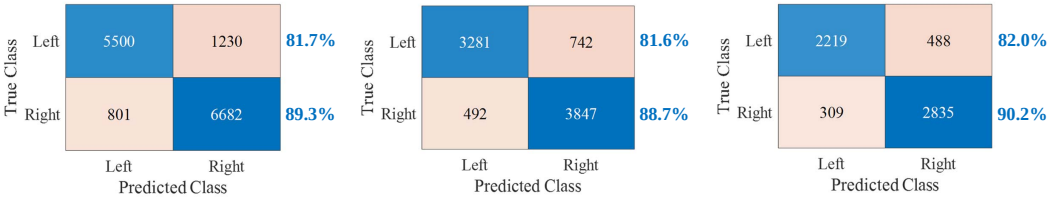

Fig. 13: Sum of the confusion matrices in the five-fold cross validation test

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 Fig. 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 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 $26mm \times 25mm$ Zhao et al. (2017). 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.

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 Fig. 14, 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.

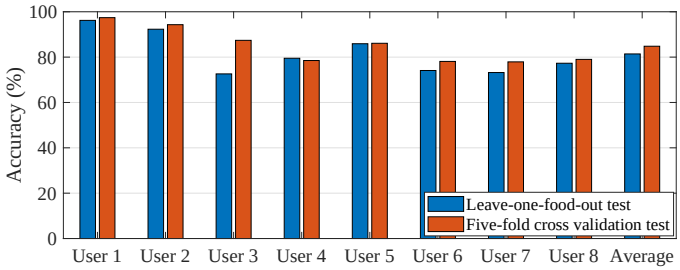

Fig. 14: Accuracy comparison between the leave-one-food-out test and five-fold cross validation test

Fig. 14 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 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.

Fig. 15 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.

Fig. 15: Sum of the confusion matrices in the leave-one-food-out test

7.6. Comparison with Existing Wearable Sensor-Based Methods

For the detection accuracy, it is infeasible to compare our proposed method with existing wearable sensor-based methods using the same dataset because the sensor types and sensor deployment are different. Here, we present a short performance comparison between our method and other existing methods, as shown in the Table 4.

Chung et al.'s work Chung et al. (2017) and our method have similar number of subjects and both accurately detect the chewing sides. However, their method is a coarse-grained method because each detection is based on a fixed-length window of 3 seconds instead of one chew. In addition, their method heavily relies on the sensitivity and deployment of the load cells, which are embedded in hinges on both sides of a pair of 3D-printed glasses. Moreover, the host devices (i.e. glasses) need to be customized to meet users' different head circumferences and wearing habits. Lucena et al. Lucena et al. (2018) did not directly report the final accuracy of chewing side detection. They indicated that "the results show no statistically significant difference between the proposed analysis and traditional clinical evaluation". Although their method and our proposed method both use motion sensors (i.e. accelerometers and gyroscopes), their method requires to deploy sensors on the user's jaw and forehead. This is obviously intrusive and not acceptable in daily living.

Table 4: Performance comparison of wearable sensor-based methods

	Sensor	**Host object**	**Subjects**	**Food types**	**Accuracy**
Chung et al. (2017)	Load cells	3D-printed glasses	10 subjects (7 males & 3 females)	3 food types	With the F1 score of 89%
Lucena et al. (2018)	2 devices (each has 1 accel. & 1 gyroscope)	None (fixed one device on jaw and the other to forehead)	60 subjects (30 males & 30 females)	3 food types	Not directly reported
Our method	2 devices (each has 1 accel. & 1 gyroscope)	Headband	8 subjects (5 males & 3 females)	8 food types	Average accuracy is 84.8%

8. Related Work

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 Flores-Orozco et al. (2015) take advantage of medical professionals' observations or special diagnosis devices. Nissan et al. Nissan et al. (2004) 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 Nissan et al. (2004)". Watarai et al. Watarai et al. (2013) 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. Mizumori et al. (2003) 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 Mizumori et al. (2003):

$$AI = \frac{Number\ of\ right\ side\ strokes - number\ of\ left\ side\ strokes}{Number\ of\ right\ side\ strokes + number\ of\ left\ side\ strokes} \times 100\% \tag{7}$$

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. Yamasaki et al. (2015) proposed to detect chewing sides by measuring the electromyographic (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. Chung et al. (2017) 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. Lucena et al. (2018) 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 Zhang & Amft (2017) 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.

9. 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 Gisel (1988) investigated the oral side preferences of 103 children. Because "placement of food on the preferred side might facilitate chewing Gisel (1988)", 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 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.

Our proposed chewing side detection method can be generalized to different users with different ages and eating habits (chewing time, chewing speed, chewing force, etc.). There are two main reasons. First, our method is based on an observation that the left and right temporalis muscles have different chewing force during unilateral chewing, which cause different muscle bulge and skull vibration accordingly. This observation proved to be valid for users of different ages and genders in an existing study van der Bilt et al. (2008). This study selects 81 healthy subjects, including 13 men (ages range from 22 to 62) and 68 women (ages range from 19 to 69). The results demonstrate that "the muscle activities of the right and left temporal muscles differed significantly during the unilateral clenching task: the ipsilateral side showed significantly more muscle activity van der Bilt et al. (2008)." Second, we build a personalized classification model for each user. Although the raw data of different users may vary a lot because of different eating habits, our proposed method is still valid as long as a user's left and right temporalis muscles have different chewing force during unilateral chewing.

Training a personalized classification model for each user may raise some concerns about the computation cost. The computation cost of the classification model mainly contains two parts: 1) computation cost for model training, which is caused by training an LSTM classification model based on a user's training samples; 2) computation cost for model testing, which is caused by classifying testing samples using the trained LSTM model. To directly compare the computation cost between model training and testing, here we do not consider the time cost for data segmentation and sample generation. We evaluate the computation cost of model training and model testing in the five-fold cross validation test in Section 7. This evaluation is done on a PC with an Intel Core i5 CPU (2.20 GHz) and 8 GB RAM. Experimental results show that the average time of eight users for training a personalized model is 18.3 seconds. Comparatively, the average time of eight users for classifying a testing sample is 3.5×10^{-4} seconds, which is much less than that of model training. A wearable device normally has a less powerful processor and smaller memory than a PC. For example, the wearable device used in our experiment utilizes a Cortex-M4 processor (64 MHz) and 64 kB RAM Nordic (2016). Thus, it may take a few dozens of minutes to train a classification model on the wearable device. However, model training is a one-time effort for a specific user in our proposed method. Thus, it can be done offline on a PC or workstation. Then the trained model is imported to the wearable device for online classification.

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. However, the size of food pieces and the food amount of each bite have little impact on the performance of our proposed chewing side detection method. This is because the detection is related to chewing instead of food intake.

The sensor deployment sites in our proposed method can be generalized to other subjects. We deploy motion sensors on the left and right anterior temporalis muscles to sense muscle bulge and skull vibration based on the observation in Fig. 3. This observation is supported by existing anatomical and clinical studies. Reference Korfage & van Eijden (1999) indicates that "during chewing and biting the anterior portions of the muscle are in general more intensively activated and they are capable of producing larger forces than the posterior portions." Reference Alarcon et al. (2000) selects 30 normocclusive subjects (16 girls and 14 boys; mean age, 12 years 5 months) to measure their EMG activities of anterior temporalis and posterior temporalis during chewing. Experimental results show that the anterior temporalis is more active than the posterior temporalis on both the left side (38.87 μV versus 22.70 μV) and the right side (40.90 μV versus 22.90 μV). Similarly, the EMG activity measurements in reference Blanksma & Eijden (1995) also show that "The anterior part of the temporalis muscle demonstrated a significantly higher peak activity than the posterior part during licorice chewing and 75% of the gum-chewing tasks." In addition, existing clinical studies MacDougall & Andrew (1953); Balkhi et al. (1993); da Silva Andrade et al. (2010) often deploy EMG electrodes at similar locations as our proposed method to sense unilateral chewing. For example, reference MacDougall & Andrew (1953) deploys three EMG electrodes on the anterior, mid, and posterior fibers of the temporalis, respectively, to investigate the bite force and muscle activity during maximum unilateral and bilateral clenching. Its EMG electrode location on the anterior fibers is the same as that in our proposed method. Based on the anatomical and clinical studies introduced above, we believe that the sensor deployment sites in our proposed method are common and can be generalized to other subjects.

Although the sensor deployment may vary a little bit for different users, it does not impact the validness of our proposed method. This is because we build a personalized classification model for each user. This classification model is valid as long as the sensor deployment locations of each specific user are consistent. We are also aware that our proposed method is a little sensitive to the sensor location bias for a specific user. A location matching algorithm is very helpful to guide the user to deploy sensors at the predefined locations. We will work on this algorithm in the future.

Our proposed method detects chewing sides using eating data. In a real-world application, an eating detection module is needed to distinguish eating activity from other daily activities, such as speaking, sitting, walking, and drinking. As the motion sensor-based eating detection Wang et al. (2018) has been investigated comprehensively, it is easy to develop and integrate it into our proposed method.

10. 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 fine-grained enough. In this paper, 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%.

Acknowledgments

Special thanks go to all the volunteers and all the anonymous reviewers. This work was supported by the U.S. National Science Foundation under grant CNS-1841129 (CSR EAGER).

References

AKJDental (2017). Harmful effects of unilateral chewing. http://www.akjdental.com/archives/3515.

Alarcon, J., Martin, C., & Palma, J. (2000). Effect of unilateral posterior crossbite on the electromyographic activity of human masticatory muscles. *American journal of orthodontics and dentofacial orthopedics*, *118*, 328–34. doi:10.1067/mod.2000.103252.

Balkhi, K. M., Tallents, R. H., Katzberg, R. W., Murphy, W., & Proskin, H. (1993). Activity of anterior temporalis and masseter muscles during deliberate unilateral mastication. *J Orofac Pain*, *7*, 89–97.

Basit, H., Eovaldi, B. J., & Siccardi, M. A. (2019). *Anatomy, Head and Neck, Mastication Muscles*. StatPearls Publishing.

van der Bilt, A., Tekamp, A., van der Glas, H., & Abbink, J. (2008). Bite force and electromyograpy during maximum unilateral and bilateral clenching. *Eur J Oral Sci*, *116*, 217–22.

Blanksma, N., & Eijden, T. (1995). Electromyographic heterogeneity in the human temporalis and masseter muscles during static biting, open close excursions, and chewing. *Journal of dental research*, *74*, 1318–27. doi:10.1177/00220345950740061201.

Chen, Y. (2011). *Ant colony optimization approach for stacking configurations*. Master's thesis, Lingnan University, Hong Kong.

Chung, J., Chung, J., Oh, W., Yoo, Y., Lee, W. G., & Bang, H. (2017). A glasses-type wearable device for monitoring the patterns of food intake and facial activity. *Scientific Reports*, *7*, 41690.

Donnell, S., Hector, M., & Hannigan, A. (2004). Chewing side preferences in children. *Journal of oral rehabilitation*, *31*, 855–60.

Du, H. (2018). Balance bias and variance: normalization techniques. In *ABCs2018*.

Flores-Orozco, E., Rovira, B., Peraire, M., Salsench, J., & Martinez-Gomis, J. (2015). Reliability of a visual analog scale for determining the preferred mastication side. *The Journal of prosthetic dentistry*, *115*.

Gisel, E. (1988). 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. doi:10.5014/ajot.42.6.378.

Korfage, J., & van Eijden, T. (1999). Regional differences in fibre type composition in the human temporalis muscle. *Journal of Anatomy*, *194*.

Lamontagne, P., Al-Tarakemah, Y., & Honkala, E. (2013). Relationship between the preferred chewing side and the angulation of anterior tooth guidance. *Medical Principles and Practice*, *22*, 545–549.

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. doi:10.1034/j.1600-0722.2000.00866.x.

Li, M., Zhang, T., Chen, Y., & Smola, A. J. (2014). Efficient mini-batch training for stochastic optimization. In *KDD '14*.

Lucena, C., Lacerda, M., Caldas, R., Lima Neto, F., & Rativa, D. (2018). Mastication evaluation with unsupervised learning: Using an inertial sensor-based system. *IEEE Journal of Translational Engineering in Health and Medicine*, *06*, 1–10.

MacDougall, J. D. B., & Andrew, B. L. (1953). An electromyographic study of the temporalis and masseter muscles. *J Anat.*, *87*, 37–45.

Martinez-Gomis, J., Lujan-Climent, M., Palau, S., Bizar, J., Salsench, J., & Peraire, M. (2008). Relationship between chewing side preference and handedness and lateral asymmetry of peripheral factors. *Archives of oral biology*, *54*, 101–7.

Mathworks (2020a). classificationlayer. URL: https://www.mathworks.com/help/deeplearning/ref/classificationlayer.html.

Mathworks (2020b). Fast fourier transform. URL: https://www.mathworks.com/help/matlab/ref/fft.html.

Mathworks (2020c). fullyconnectedlayer. URL: https://www.mathworks.com/help/deeplearning/ref/nnet.cnn.layer.fullyconnectedlayer.html.

Mathworks (2020d). Long short-term memory networks. URL: https://www.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html.

Mathworks (2020e). Sequence classification using deep learning. URL: https://www.mathworks.com/help/deeplearning/examples/classify-sequence-data-using-lstm-networks.html.

Mathworks (2020f). softmaxlayer. URL: https://www.mathworks.com/help/deeplearning/ref/nnet.cnn.layer.softmaxlayer.html.

Mathworks (2020g). Standardized z-scores. URL: https://www.mathworks.com/help/stats/zscore.html.

Mathworks (2020h). trainingoptions: Options for training deep learning neural network. URL: https://www.mathworks.com/help/deeplearning/ref/trainingoptions.html.

Mizumori, T., Tsubakimoto, T., Iwasaki, M., & Nakamura, T. (2003). Masticatory laterality - evaluation and influence of food texture. *Journal of oral rehabilitation*, *30*, 995–9. doi:10.1046/j.1365-2842.2003.01086.x.

Nissan, J., Gross, M. D., Shifman, A., Tzadok, L., & Assif, D. (2004). Chewing side preference as a type of hemispheric laterality. *J Oral Rehabil*, *31(5)*, 412–6. doi:10.1111/j.1365-2842.2004.01256.x.

Nordic (2016). nrf52832 product specification v1.1. URL: https://infocenter.nordicsemi.com/pdf/nRF52832_PS_v1.1.pdf.

Santana-Mora, U., LÃ§pez-CedrÃžn, J., Mora, M. J., Otero, X. L., & Santana-PenÃ■n, U. (2013). Temporomandibular disorders: The habitual chewing side syndrome. *PLoS One*, *8*, 1–7.

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.

da Silva Andrade, A., Gaviao, M. B. D., Gameiro, G. H., & Rossi, M. D. (2010). Characteristics of masticatory muscles in children with unilateral posterior crossbite. *Braz Oral Res*, *24*, 204–10.

Sindhu, I., Daudpota, S., Badar, K., Bakhtyar, M., Baber, J., & Nurunnabi, M. (2019). Aspect based opinion mining on studentâĂŹs feedback for faculty teaching performance evaluation. *IEEE Access*, *PP*, 1–1. doi:10.1109/ACCESS.2019.2928872.

Tiwari, S., Nambiar, S., & Unnikrishnan, B. (2017). Chewing side preference - impact on facial symmetry, dentition and temporomandibular joint and its correlation with handedness. *Journal of Orofacial Sciences*, *9*, 22–27.

TriggerPointSelfHelp.com (2020). Temporalis muscle: Location, action and trigger points. URL: http://www.gustrength.com/muscles:temporalis-location-action-and-trigger-points.

Tutorialspoint (2020). Tensorflow - cnn and rnn difference. URL: https://www.tutorialspoint.com/tensorflow/tensorflow_cnn_and_rnn_difference.htm.

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). Beijing, China: IEEE volume 2.

Wang, S., Zhou, G., Ma, Y., Hu, L., Chen, Z., Chen, Y., Zhao, H., & Jung, W. (2018). Eating detection and chews counting through sensing mastication muscle contraction. *Smart Health*, . doi:10.1016/j.smhl.2018.07.004.

Watarai, M., Ayano, R., Funatsu, T., Ooka, T., Takahashi, M., & Inoue, M. (2013). Three-dimensional motion analysis of lip and mandibular movements during mastication. *Dental Medicine Research*, *33*, 88–99. doi:10.7881/dentalmedres.33.88.

WebmasterTeam (2020). Classification with imbalanced datasets. URL: https://sci2s.ugr.es/imbalanced.

Wikipedia (2020a). Accelerometer. URL: https://en.wikipedia.org/wiki/Accelerometer.

Wikipedia (2020b). Cross−validation. URL: https://en.wikipedia.org/wiki/Cross$-$validation_(statistics).

Wikipedia (2020c). Gyroscope. URL: https://en.wikipedia.org/wiki/Gyroscope.

Wikipedia (2020d). Muscle. URL: https://simple.wikipedia.org/wiki/Muscle.

Wikipedia (2020e). Muscles of mastication. URL: http://en.wikipedia.org/wiki/Muscles_of_mastication.

Wikipedia (2021). Temporal muscle. URL: https://en.wikipedia.org/wiki/Temporal_muscle.

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.

Yamasaki, Y., Kuwatsuru, R., Tsukiyama, Y., Matsumoto, H., Oki, K., & Koyano, K. (2015). Objective assessment of actual chewing side by measurement of bilateral masseter muscle electromyography. *Archives of oral biology*, *60*, 1756–1762.

Zhang, R., & Amft, O. (2017). Monitoring chewing and eating in free-living using smart eyeglasses. *IEEE Journal of Biomedical and Health Informatics*, *PP*, 1–1.

Zhao, H., Wang, S., Zhou, G., & Zhang, D. (2017). 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 (pp. 392–401). Piscataway, NJ, USA: IEEE Press. doi:10.1109/CHASE.2017.123.