Chewing Behavior Detection Based on Facial Dynamic Features

Cheng-Zhe Tsai[1], Chun-Chih Lo[1], Lan-Yuen Guo[2], Chin-Shiuh Shieh[1], and Mong-Fong Horng[1](✉)

[1] Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan
mfhorng@nkust.edu.tw

[2] Department of Sports Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

Abstract. Diet serves as the primary source of calorie intake for human beings, and maintaining a regular dietary intake is crucial for overall health. The pace or speed of chewing can significantly impact the body's response to food consumption. Traditionally, dietary monitoring has relied on manual assessment by clinicians, a process that is labor-intensive, time-consuming, and susceptible to inaccuracies. In this study, we introduce a novel image processing-based approach for quantitatively evaluating chewing and swallowing capabilities. In this method, facial recognition is employed to detect and calibrate facial features using the Dlib facial landmark model. This enables the precise identification of the mandible's position, facilitating the capture of the subject's chewing movements. Subsequently, signal processing techniques are applied to calculate the number of chewing instances. Experiments was conducted with five subjects of diverse genders and ages. The results indicated a mean absolute error of 6.48% in chewing count calculation. The proposed method offers the advantages of convenience and minimal error in comparison to similar studies.

Keywords: chewing instances · dietary intake · facial feature recognition

1 Motivation and Purpose

1.1 Motivation

The ingested food undergoes a pivotal chewing process within the mouth before proceeding to be swallowed and subsequently digested. The act of swallowing is immediately followed by chewing, creating a strong correlation between the two actions. Chewing plays a vital role in breaking down food into smaller fragments. Through coordinated efforts involving the tongue and oral muscles, saliva is absorbed to form cohesive food masses. Subsequent to ingestion, food progresses through the oral cavity, pharynx, esophagus, and eventually reaches the stomach. Figure 1 illustrates the sequential flow of swallowing, divided into five phases [1]: (a) Preparation phase, (b) Preparatory action phase, (c) Oral phase, (d) Pharyngeal phase, and (e) Esophageal phase. These phases

J.-S. Pan et al. (Eds.): ICGEC 2023, LNEE 1114, pp. 355–365, 2024.
https://doi.org/10.1007/978-981-99-9412-0_37

collectively form crucial components of the swallowing process. Any abnormalities in their function can be significant contributors to swallowing difficulties. In this study, our focus is on understanding chewing movements during the oral phase. We aim to characterize these movements comprehensively and propose a method for quantifying their frequency.

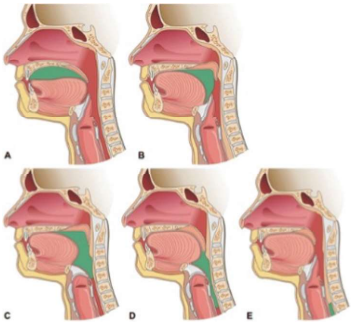

Fig. 1. The five phases of swallowing process [1]

1.2 Purpose

This study employs image recognition technology to precisely capture chewing events. Specifically, its objective is to devise a specialized chewing motion that enables image acquisition from the external perspective of the motion. To achieve this goal, the investigation utilizes camera equipment to systematically monitor subjects' eating behaviors. The camera equipment is strategically positioned to observe and document the dietary habits of the subjects. Captured images of the eating process are then transmitted back to a computer for subsequent analysis. This phase involves a thorough assessment of the distinctive characteristics exhibited in the subjects' chewing movements during their meals. Upon receiving the transmitted image messages, the study delves into comprehensive image analysis. This analysis serves as the foundation for the design of advanced algorithms adeptly calibrated to evaluate the subjects' eating behaviors. These algorithms encompass essential parameters, including the frequency of chewing occurrences and the pace of chewing. This approach provides an efficient methodology for accurately quantifying and comprehending the intricate dynamics of chewing during dietary intake.

2 Literature Review and Technical Analysis

2.1 Visual Characteristics of Chewing During Eating

During the act of eating, chewing is characterized by a consistent pattern of the mandible's cyclic opening and closing motions. This orchestrated movement, facilitated by the masticatory muscles, allows the teeth to grind and compress the food, transforming it into a cohesive mass suitable for the processes of swallowing, digestion, and

absorption. As visually depicted in Fig. 2, chewing involves the precise coordination of the masticatory muscles to regulate the movement of both the teeth and the mandible. This coordinated effort results in a rhythmic sequence of the jaw's opening and closing, effectively processing the ingested food within the confines of the oral cavity. Hence, the visual depiction of chewing is prominently tied to the motion of the mandible. The extent of the mandible's displacement emerges as a crucial quantitative measure, employed to assess the frequency of individual chewing sessions.

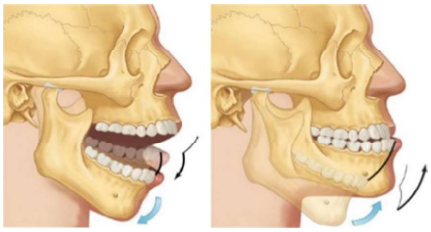

Fig. 2. Mandible motion during chewing [2]

2.2 Past Literature on Chewing Recognition

In 2008, Nishimura et al. introduced a novel approach involving a wireless wearable in-ear microphone for monitoring eating habits [3]. Their study featured the installation of a wireless microphone on the subject's ear to capture distinct chewing sound features. The sound signals generated while biting and chewing food were harnessed to track chewing motion. Through signal processing techniques, the number of chewing instances was computed. The final experimental outcomes revealed a 1.93% margin of error in estimating the frequency of chewing instances. In 2016, Farooq et al. presented an automated methodology for quantifying the number of chewing instances employing a wearable piezoelectric sensor [4]. The authors affixed a piezoelectric sensor element beneath the participant's ear, secured with a medical adhesive. This setup enabled the capture of vibrations generated during jaw movement during the act of chewing. The piezoelectric sensor translated these vibrations into voltage, producing a chewing signal that could be identified and quantified through peak detection methods. Their approach reported an average calculation error of 7% in determining the number of chews.

Commonly employed techniques for capturing chewing activity include utilizing microphones to record chewing sounds or employing piezoelectric sensors to detect and analyze chewing signals. Although these methods offer commendable accuracy, microphone recordings may be susceptible to environmental interference, potentially compromising their precision. Additionally, both microphone and piezoelectric sensor approaches necessitate physical attachment to the body, which can prove cumbersome and uncomfortable for certain subjects, thereby inconveniencing the testing environment. In contrast, the approach proposed in this study solely relies on image analysis to assess the number of chewing instances, offering a relatively more convenient alternative.

2.3 Dlib Facial Landmark Model

Dlib is a C + + based machine learning toolkit [5] designed for the development of machine learning, image processing, and natural language processing applications. This comprehensive toolkit offers an array of machine learning algorithms and tools tailored to image-related tasks, including image processing, feature extraction, face detection, object detection, and more. A noteworthy component within Dlib's repertoire is the Dlib facial landmark model, as depicted in Fig. 3. This model operates on the foundations of face detection and feature annotation within the facial structure using deep learning technology and trained using Convolutional Neural Networks (CNN) [6]. This model demonstrates an inherent capability to autonomously identify facial components within images, including eyes, nose, mouth, and other distinctive attributes. This model's efficacy is further substantiated by its training data, which comprises an extensive collection of real facial images captured in diverse scenarios. This comprehensive dataset empowers the model with the capability to recognize and annotate facial features across varying environmental scenarios.

In this study, the Dlib facial landmark model is leveraged to recognize the number of chewing instances performed by participants. The quantification of chewing sessions is predicated on the extent of mandibular bone displacement during the process of chewing and swallowing. The Dlib facial landmark model proficiently anticipates and accurately predicts mandibular features, which proves invaluable for the current study's objectives.

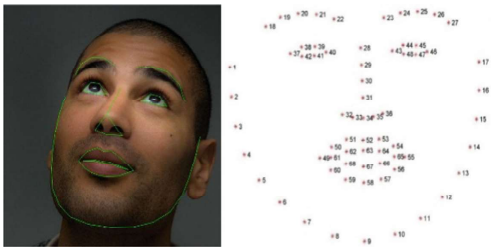

Fig. 3. Dlib Face Merker Model [7]

3 Methods and Procedures

3.1 Application Scenarios

The fundamental objective of this study is to employ an imaging approach that enables subjects to quantify their chewing instances while positioned in front of a camera, thereby eliminating any requirement to manipulate or wear sensory equipment. Figure 4 illustrates the envisioned application scenario. This scenario encompasses situating the subject within a well-illuminated environment, positioning them in a chair, and configuring a color camera along with a depth camera in front of the subject. This configuration serves to capture distinctive attributes of the subject's mouth and mandible, facilitating subsequent determination and analysis of chewing actions. During the testing phase, a

consistent quantity of food is provided to the participant. This ensures a uniform testing condition, allowing each participant to consume the same amount of food during the experiment. Once the participant is prepared, they are instructed to initiate eating. At this point, the camera initiates the recording of images depicting the participant's feeding process. These recorded images are subsequently transmitted to a computer for analysis. The collected data is then inputted into a pre-designed program crafted for the purpose of chewing instance quantification.

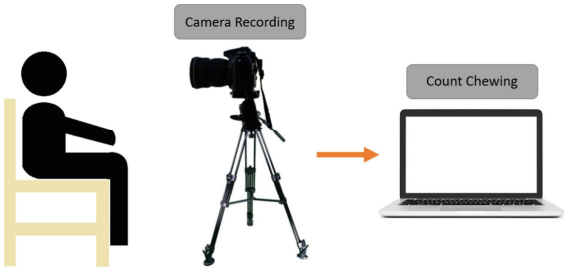

Fig. 4. Application Scenarios [7]

3.2 Chewing Calculation Process

The chewing signal is further processed by signal processing method, and then the chewing signal is peak detected and the number of chewing is calculated. The sequential procedure for calculating chewing instances is depicted in the flowchart illustrated in Fig. 5. This process is initiates by segmenting the captured images of the subject's eating activity into individual frames. Subsequent to segmentation, each frame undergoes recognition using the Dlib facial landmark model for face identification. This recognition procedure encompasses annotating facial features and generating corresponding coordinates for these features. Notably, the coordinates associated with the position of the subject's mandibular bone during the chewing motion within the video are extracted as the chewing signal. To facilitate comprehensive analysis, the chewing signal is subjected to signal processing techniques. This processed signal subsequently undergoes peak detection procedures, culminating in the computation of the number of chewing instances.

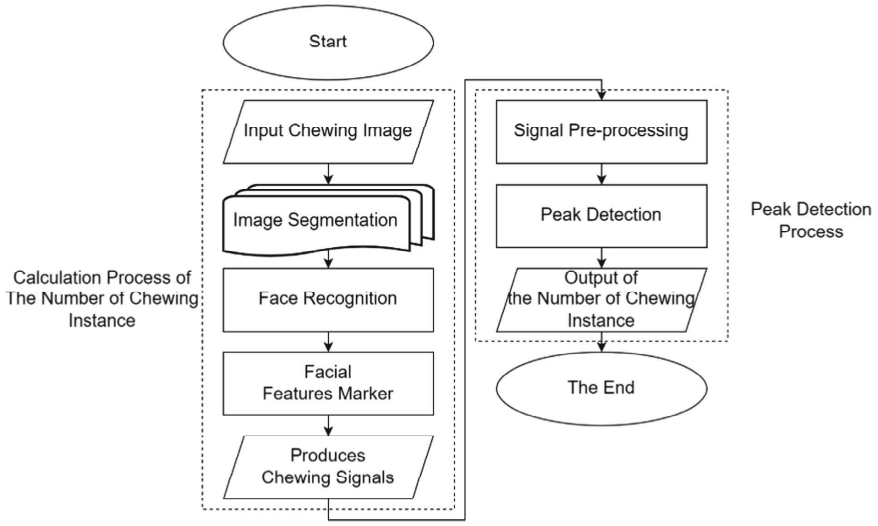

Fig. 5. Flowchart for Calculating the Number of Chewing Instances

3.3 Algorithm Design for Chewing Instances Calculation

The algorithm presented herein facilitates the precise calculation of chewing instances. Within the chewing signal, each peak manifest as a distinct chewing characteristic, reflecting the upward and downward displacement of the mandible during food consumption. These peaks within the chewing signal serve as the basis for computing the frequency of chewing events during eating. Figure 6 illustrates the calculation process of peak occurrences within the chewing signal.

The initial phase involves generating the chewing signal by converting the continuous sequence of the subject's eating activity into a singular image suitable for facial recognition. This image is then subjected to detailed analysis to identify the coordinates of the facial region. These coordinates are then employed in conjunction with a landmark model to characterize the entirety of facial features, thereby obtaining the corresponding features (i, j) coordinates. Due to the inherent variations in facial features and sizes across individuals, the algorithm calculates the distance $d_{(i,j)}(n)$ between the two eye points and the distance from the nose to the chin using Eq. 1. Furthermore, the ratios of distances $d_{(i1,j1)}(n)$ between the eyes and $d_{(i2,j2)}(n)$ between the nose and chin are determined using Eq. 2. These measurements collectively compose the primary masticatory signal $x(n)$ output. Initial analysis of the raw chewing signal reveals the presence of significant noise within its pattern. To address this, the initial step is to let the raw chewing signal undergoes processing via a shift-averaging filter, as depicted in Eq. 3. This filter effectively averages the raw signals $x(0), x(n-1)$, resulting in the generation of an improved chewing or swallowing signal $x^{'}(n)$.

Subsequently, the algorithm identifies peak counts within the signals, employing a moving average filter as a threshold T, this step plays a decisive role in identifying the signal peaks. When the chewing signal surpasses the defined threshold T, it triggers the generation of an amplified signal $y(n) = 1$. Conversely, when the signal is less than or

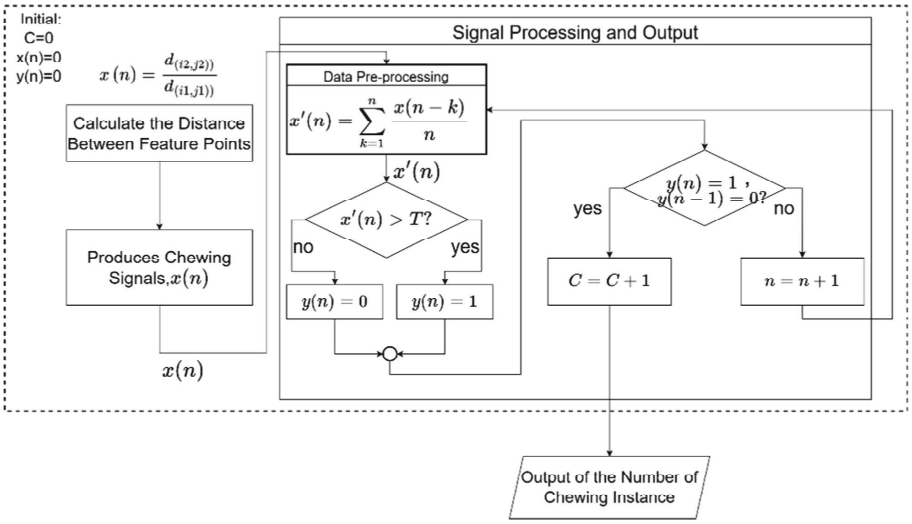

Fig. 6. Flowchart of Chewing Instance Calculation

equal to the threshold T, the signal $y(n) = 0$. Finally, peak counts denoted as C are accumulated by evaluating whether the condition $y(n) = 1$ and $(n-1) = 0$ is satisfied. Upon meeting these conditions, the count C is incremented. Otherwise, the assessment process continues.

$$d_{(i,j)}(n) = \sqrt{(x_i(n) - x_j(n))^2 + (y_i(n) - y_j(n))^2} \tag{1}$$

where $d_{(i,j)}(n)$ represents the Euclidean distance between the specified points, with $(x_i(n)$ and $y_i(n)$ denoting the coordinates of feature point i, and $x_j(n)$ and $y_j(n)$ representing the coordinates of feature point j.

$$x(n) = \frac{d_{(i2,j2)}(n)}{d_{(i1,j1)}(n)} \tag{2}$$

where $x(n)$ represents the ratio obtained from the distances between feature points $i_1 \backslash j_1$ and feature points $i_2 \backslash j_2$, and it serves as the output for the chewing raw signal.

$$x^{'}(n) = \frac{1}{n}[x(n-1), x(n-2), \ldots, x(0)] \tag{3}$$

where $x^{'}(n)$ is the average of the raw signals $x(n-1), x(n-2), \ldots, x(0)$, enabling the signal to undergo a low-pass filtering calculation process.

$$y(n) = \begin{cases} 1 \ if x'(n) > T \\ 0 \ \ otherwise \end{cases} \tag{4}$$

where T represents the threshold. If the value of $x\prime(n)$ is greater than T, then $y(n)$ is assigned the value 1; otherwise, it is assigned the value 0.

4 Experimentation and Performance Analysis

4.1 Purpose of the Experiment

The primary goals of the experiments conducted in this study are twofold: first, to establish the viability of capturing masticatory image features, and second, to demonstrate the efficacy of the newly developed masticatory capture technique based on image processing. In this section, a sequence of experiments will be conducted to explore the chewing characteristics exhibited by subjects during eating. This includes the formulation of a comprehensive set of measurement protocols for the participants, alongside the design, evaluation, and comparison of digital image-based feature capture methods employing various metrics. Additionally, the number of chewing instances of the subjects will be subjected to analysis, leading to the development of a novel approach for quantifying the number of chewing instances.

4.2 Characteristics of Chewing

In Fig. 7, the participant is seated in front of a USB camera while consuming food. Subsequently, the feeding images are fed into the Dlib facial landmark model. The resultant algorithm generates chewing signals, where the horizontal axis signifies the chewing signals measured across each image, and the vertical axis illustrates the amplitude of the chewing signals. Notably, a discernible pattern is observed in the form of changes in mandibular displacement during chewing. The signals depicted in Fig. 7 correspond to peaks resulting from the cyclical up and down movement of the mandible. The ensuing step involves the computation of these peaks in the masticatory signals, which are then presented as masticatory indicators.

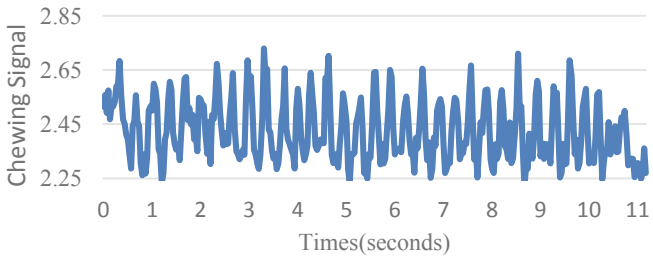

Fig. 7. Raw Signals of Chewing

4.3 Chewing Signal Data Pre-Processing

Due to the presence of substantial noise in the raw signal, it is imperative to eliminate this noise to enhance the accuracy of the assessment. To achieve this, a moving average filtering technique is employed to refine the raw chewing signal by retaining its fundamental attributes while effectively mitigating noise interference. The outcome of this process is showcased in Fig. 8, where the resultant low-pass filtering serves to amplify the significant chewing features, thereby preserving essential data characteristics.

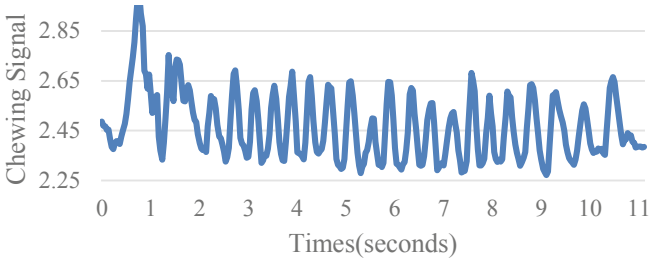

Fig. 8. Chewing Signal Low-Pass Filtering

4.4 Analysis of Experimental Results for Chewing Instance Calculation

The application of the dynamic threshold, achieved through the utilization of a moving average filter, plays a important role in this study by identifying the peaks within the chewing signal. As depicted in Fig. 9, the signal processing results for calculating the number of chewing instances are presented. The blue line denotes the outcome of the raw chewing signal following low-pass filtering, the red line signifies the threshold value generated through the application of the moving average filter, and the green line represents the signal utilized for chewing determination. Furthermore, a diverse group of five subjects, varying in age and gender, were engaged in consuming distinct foods in front of the camera. The manual recording of each subject's chewing instances was carried out as a benchmark for comparison against the program's calculations. The assessment of the accuracy was executed by leveraging Eq. 5, thereby computing the absolute percentage error in the calculated number of chews.

$$\delta = \left(\frac{|x - y|}{y} \right) * 100\% \tag{5}$$

where δ presents the absolute percentage error, x denotes the measured value, and y signifies the actual value.

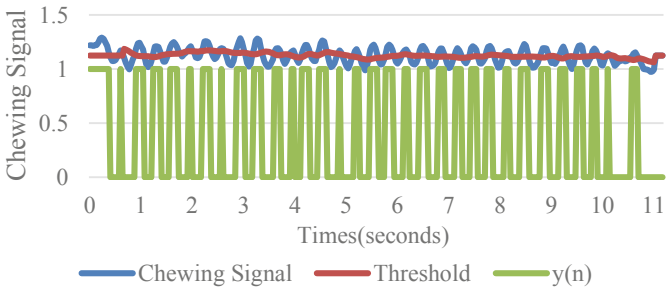

Fig. 9. Chewing Signal Processing

The experiment employed the developed chewing instance calculation system with participation from five healthy subjects. The experimental protocol involved placing a

uniform-sized cookie into the subjects' mouths, after which the program was initiated to detect and track chewing movements until the food was swallowed. The entire process, from initiation to completion, was timed and monitored by the program. The outcomes of the chewing tests are presented in Table 1, including the age of each of the five subjects, the corresponding count of chewing instances, and the duration of their chewing activities, among other relevant information.

Table 1. Experimental Results for the Number of Chewing Instances

Tested Person No	Age	Chewing Quantity Measurement	Actual number of chews	Time (sec)	Number of Chewing Instances (times/second)	Absolute error of calculation (%)
1	23	28	28	27.1	1.0	0
2	22	34	36	33	1.1	5.6
3	26	28	30	20	1.5	6.7
4	25	15	16	16	1.0	6.3
5	24	33	29	39	0.7	13.8

The experimental results demonstrated that the proposed method in this study yielded a mean absolute error of 6.48% in chewing calculation. By comparing the outcomes of this study to the chewing recognition results presented by Nishimura et al. and Farooq et al. [3, 4], our approach displayed a slightly higher error rate in calculating the number of chewing instances. However, the approach adopted here offers distinct advantages such as non-contact and convenience. These outcomes substantiate the effectiveness of the chewing instance calculation method employed in this study. The measurement technique and algorithm devised herein facilitated the derivation of chewing counts, thereby enabling an analysis of the subjects' eating behaviors.

5 Results and Discussion

In this study, we have developed a calculation process for determining the number of chewing instances using an image-based approach. This method effectively captures the chewing characteristics of subjects and incorporates an algorithm to accurately compute the number of chews. The experiment in this study involving five subjects from various age groups, the obtained results demonstrated an mean absolute error of 6.48% in the calculated number of chewing instances. This approach offers convenience and non-contact advantages when compared to previous chewing capture methods. Finally, the system developed through this study holds the potential to enhance participants' comprehension of their individual chewing patterns, offering valuable insights into their dietary habits.

Acknowledgement. The authors would like to thank the National Science Council in Taiwan R.O.C for supporting this research, which is part of the project numbered MOST 109–2221-E-992 -073 -MY3, NSTC 112–2622-8–992-009 -TD1 and NSTC 112–2221-E-992 -057 -MY3.

References

1. Carbo, A.I., Brown, M., Nakrour, N.: Fluoroscopic swallowing examination: radiologic findings and analysis of their causes and pathophysiologic mechanisms. Radiographics **41**(6), 1733–1749 (2021)
2. Mrzezo. Mechanics of Mandibular Movement. https://pocketdentistry.com/4-mechanics-of-mandibular-movement/
3. Nishimura, J., Kuroda, T.: Eating habits monitoring using wireless wearable in-ear microphone. In: 2008 3rd International Symposium on Wireless Pervasive Computing, pp. 130–132. IEEE (2008)
4. Farooq, M., Sazonov, E.: Automatic measurement of chew count and chewing rate during food intake. Electronics **5**(4), 62 (2016)
5. Dlib. http://dlib.net/
6. O'Shea, K., Nash, R.: An introduction to convolutional neural networks, arXiv preprint arXiv: 1511.08458, (2015)
7. Rosebrock, A.: Facial landmarks with dlib, OpenCV, and Python. https://pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/