mmJaw: Remote Jaw Gesture Recognition with COTS mmWave Radar

Awais Ahmad Siddiqi[1], Yuan He*[1], Yande Chen[1], Yimao Sun[1], Shufan Wang[1], and Yadong Xie[1]

[1]School of Software & BNRist, Tsinghua University, China
xidq19@mail.tsinghua.edu.cn, heyuan@tsinghua.edu.cn*,
{cyd22, sym21}@mails.tsinghua.edu.cn, wangshufan6622@outlook.com, ydxie@mail.tsinghua.edu.cn

***Abstract*—With the increasing prevalence of IoT devices and smart systems in daily life, there is a growing demand for new modalities in Human-Computer Interaction (HCI) to improve accessibility, particularly for users who require hands-free and eyes-free interaction in contexts like VR environments, as well as for individuals with special needs or limited mobility. In this paper, we propose teeth gestures as an input modality for HCI. We find that teeth gestures, such as tapping, clenching, and sliding, are generated by various facial muscle movements that are often imperceptible to the naked eye but can be effectively captured using mm-wave radar. By capturing and analyzing the distinct patterns of these muscle movements, we propose a hands-free and eyes-free HCI solution based on three different gestures. Key challenges addressed in this paper include user range identification amidst background noise and other irrelevant facial movements. Results from 16 volunteers demonstrate the robustness of our approach, achieving 93% accuracy for up to a 2.5m range.**

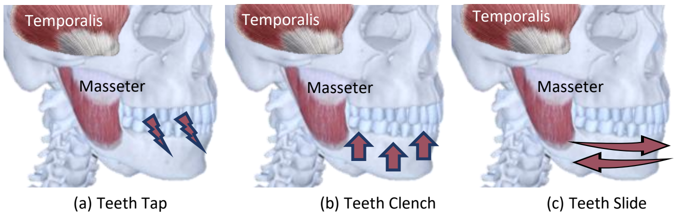
***Index Terms*—mmWave, Sensing, Human-Computer Interface, Teeth Gestures**

I. Introduction

In Human-Computer Interaction (HCI), advancements are continually sought to improve the interaction between humans and technology. As the demand grows for interfaces that require minimal effort on the part of users, innovative solutions are emerging to streamline this interaction process. Among these unique solutions, teeth gestures are an intriguing development. This paper introduces a novel method to explore and recognize different teeth gestures. Leveraging mmWave radar, we aim to offer a non-invasive method for remotely sensing teeth gestures, potentially alleviating discomfort associated with traditional proximal or invasive devices.

There have been substantial studies related to facial sensing in medical, HCI, and wearable domains. mmJaw distinguishes itself by emphasizing the following aspects: unlike past works that use dedicated hardware or invasive/intraoral devices requiring attachment to the teeth, causing discomfort to the user, we utilize mmWave radar for remote sensing of teeth gestures ensuring no discomfort to the user.

By harnessing these distinctive movements, we can develop an HCI system, named mmJaw, that offers versatility across various applications. This approach capitalizes on natural and intuitive motions, bypassing many limitations encountered with traditional HCI methods. Additionally, this solution holds promise for enhancing user experience and accessibility, particularly for individuals with disabilities, such as those who are paralyzed or partially paralyzed.



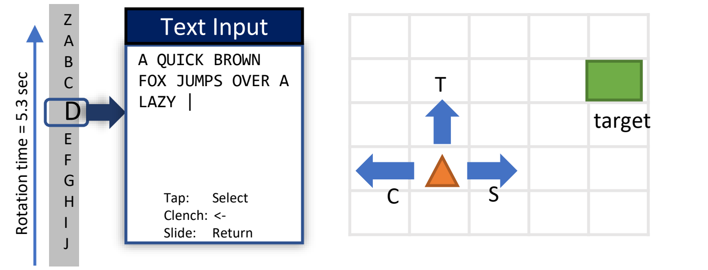
(a) Teeth Tap (b) Teeth Clench (c) Teeth Slide

Fig. 1. Stretching and relaxing of masseter muscle with different teeth activities

The teeth gestures like tapping, clenching, and sliding result from lower jaw actions, primarily involving the muscles of mastication: masseter, temporalis, medial pterygoid, and lateral pterygoid [1]. These muscles are tightly connected and contraction or relaxation of one muscle group can be sensed in another. While the medial and lateral pterygoid muscles are deep within the face, masseter and temporalis muscles are located near the skin as shown in Fig. 1. Their movements are strong enough to propagate to the edge of the face and can be detected by mm-wave radar [2]. By analyzing the phase variation pattern captured by the radar, we demonstrate the feasibility of localizing teeth gestures, thereby creating a human-to-machine interface. The challenges include weak signals, background noise, and localization of static targets. We address these issues with a series of sensing techniques, enabling the detection of three distinct gestures with high accuracy. Results from 16 volunteers show the system's robustness, despite mmJaw not requiring per-user training. This offers seamless interaction across different users and environments, ultimately paving the way for more intuitive and efficient human-computer interaction experiences. In this paper, we make the following contributions:

- We have proposed a novel HCI method that can be adapted by anyone with minimum training.
- We have leveraged the sub-mm level accuracy of mm-wave radar to detect minute motion patterns to develop different gestures.

*Yuan He is the corresponding author.



(a) mmJaw as text input modality (b) mmJaw as navigational control modality

Fig. 2. Potential applications of mmJaw

- To overcome the challenge of noise and distortion due to different activities, we did bio-metric verification of the target and trained a model that can distinguish between gesture and noise in different environments with high accuracy

The rest of the paper is arranged as follows: we build our case for the applicability of teeth gesture sensing in Sec. II. Then we define the gestures and describe the capability of mmwave radar to capture with preliminary experimentation in Sec. III. Sec. IV presents the overall design in response to the challenges during the implementation in real-life scenarios. Implementation and Evaluation are presented in Sec. V, followed by discussion and related work in Sec. VI and VII. We conclude our paper in Sec. VIII.

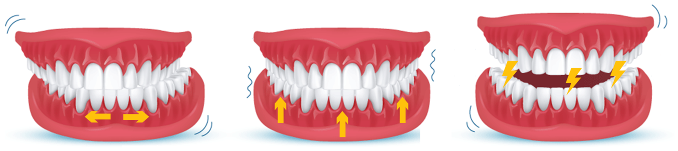
II. Potential Applications

Given the capabilities offered by mmJaw, various applications in multiple fields can be envisioned through teeth gesture sensing. Although the user experience and usability surveys are beyond the scope of this work, we explore the scenarios where mmJaw-based solutions are highly desirable.

1) mmJaw as keyboard: mmJaw can be used as a text input device (keyboard) where users can select scrolling text on the screen and input it. Different gestures may correspond to different text options, such as clicking to select text, clenching to backspace, and sliding to save as shown in Fig. 2(a). This functionality further expands the usability of mmJaw, making it a versatile tool for a wide range of applications, from basic computer interactions to more complex text input tasks.

2) mmJaw as mouse: For hands-free and eye-free interaction, mmJaw can be utilized as a mouse to perform tasks such as navigating, pointing, and clicking. This approach enables users to control their computer using teeth gestures, offering an alternative to traditional mouse interactions, all without the need for physical hand movements or visual focus. A visual representation of these tasks is given in Fig. 2(b), highlighting the potential of mmJaw to enhance accessibility and user experience in computing environments.

3) Medical applications: Symptoms preceding seizures, like teeth chattering and lip-smacking [3] [4], can be detected early by mmJaw, which doesn't require mouth devices and can monitor unattended patients. Dental disorders needing teeth movement monitoring, and artifacts in EEG caused by teeth clenching [5] can also be tagged using mmJaw.



(a) Sliding (b) Clenching (c) Tapping

Fig. 3. Different teeth activities.

4) Accessibility: IoT platforms, like Google's "Little Signals" [6], require physical button presses. IFTTT [7] based switches also rely on physical buttons, which can be integrated with mmJaw.

5) VR Sensing: Multiple studies [8] [1] use teeth gesture as an additional input method in VR systems. In a threat model, mmJaw can remotely eavesdrop on these subtle movements to detect potential vulnerability.

III. Background and Preliminaries

When humans move their mouth or jaw, it's mainly the lower jaw that moves, offering three degrees of freedom: up/down, left/right, and in/out. However, these movements are quite limited, especially the lateral and depth movements, restricting the range of possible gestures

Furthermore, unlike hand gestures or facial expressions, gestures relying on jaw and teeth tapping lack clear definitions. Hence, we initiate our study to identify the most common gestures that individuals can easily execute with minimal effort and learning. While there are numerous facial gestures performed by humans, we selected the gestures for this work with the following criteria: first gestures should be universal and secondly the gestures should not be prominently visible to the naked eye. The most common gestures identified were tapping, clenching, and sliding.

A. Gesture Definition

Tapping refers to brief, repetitive movements involving the teeth or jaw, akin to light, rapid touch. Clenching denotes a sustained, firm closure of the teeth, exerting pressure. While there is no relative movement between the teeth during clenching, the bulge of masseter muscle can be observed during the gesture even with the hand. Sliding involves a smooth, continuous movement of the lower and upper teeth along the axis of the teeth. The movement of the jaws/teeth during the execution of the gestures is shown in Fig. 3.

Normally in such systems, we expect to have two different types of movements. i.e. the muscle movement in response to the jaw movement and the bone-borne vibrations. These are the vibrations produced as a result of a jaw gesture such as a click and travel through the bone and manifest on the facial skin. In mmJaw, we only focus on the pattern of the first type of movement because firstly, among all the gestures, only the Clicking gesture produces sound, so it can be applied to all the gestures. Secondly, the bone-borne vibrations are too weak to be picked up by mm-wave radar.

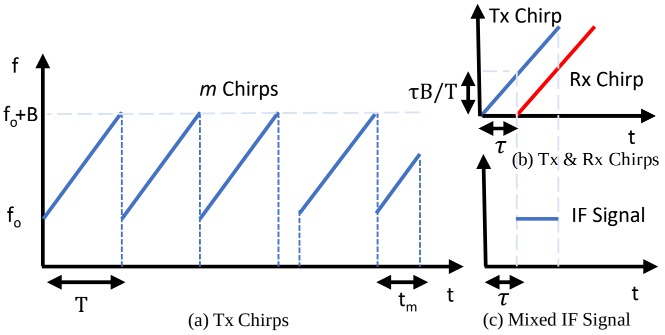


Fig. 4. Transmission and reception of chirps from mmwave radar

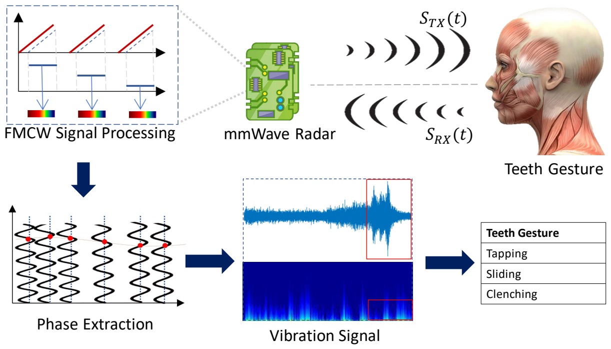


Fig. 5. Preliminary sensing model overview

While there are multiple approaches to sense the gesture information, such as microphone-based teeth activity sensing [9] [10] [4] [11] and IMU-based [12] [10], there is a need for the gadget to be attached to the face.

B. mmWave based Sensing

mmJaw utilizes the commodity mmwave radar to sense the facial motions caused by aforementioned gestures remotely. The short wavelength of mmWave radar enables higher resolution in both range and velocity measurements, making it a suitable candidate for teeth gesture detection [13]–[15].

The fundamental concept of detecting teeth gestures using mmWave Radar is to transmit mmWave signals toward the face and analyze the reflected signals received by the antenna. The transmitted signal $x_{Tx}(\tau)$ and the received signal $x_{Rx}(\tau)$ of a Frequency-Modulated Continuous Wave (FMCW) mmWave radar at time τ within a chirp period T are shown in Fig. 4 and can be expressed as follows:

$$x_T(\tau) = S_{Tx} \cdot e^{j(2\pi f_o \tau + \pi B \frac{\tau^2}{T})}, \tag{1}$$

$$x_{Rx}(\tau) = S_{Rx} \cdot e^{j(2\pi f_o (\tau - t_d) + \pi B \frac{(\tau - t_d)^2}{T})} \tag{2}$$

where S_{Tx} and S_{Rx} represent the signal strength (amplitude) of transmitted and received signals, f_o is the starting frequency, B is the bandwidth, and t_d is the round-trip time delay between transmitting and receiving as shown in Fig. 4(b).

The FMCW radar further mixes $x_T(\tau)$ and $x_{Rx}(\tau)$ and outputs the intermediate frequency signal $x(\tau)$ as follows:

$$x(\tau) \approx S_{Tx} S_{Rx} \cdot e^{j\left(4\pi B \frac{d}{cT} \tau + 4\pi f_o \frac{d}{c}\right)} \tag{3}$$

where d denotes the distance between reflecting objects and radar, and c denotes the speed of light.

In practice, $x(\tau)$ contains reflections from different objects at various distances. To separate these reflections based on their frequency components $4\pi B \frac{d}{cT}$, a range-FFT is used. When performing range-FFT, $x(\tau)$ is sampled at discrete time intervals $x[\tau_n]$, where $n = 0, 1, \ldots, N_\tau - 1$, and N_τ is the number of samples per chirp. The range-FFT (implemented as Discrete Fourier Transform F) is then performed on $x[\tau_n]$ to obtain frequency components $X[k]$, where $k = 0, 1, \ldots, N_\tau - 1$:

$$X[k] = F[x[\tau_n]] = \sum_{n=0}^{N_\tau - 1} x[\tau_n] \cdot e^{-j2\pi \frac{kn}{N_\tau}} \tag{4}$$

Each frequency component $X[k]$ corresponds to the reflection from the range within $\left[\frac{kc}{2B}, \frac{(k+1)c}{2B}\right]$.

The desired signal of the potential target X_{target} is selected from the range bin d_{target} where the gesture information lies:

$$X_{target} = X[d_{target}]. \tag{5}$$

Finally, the variations of X_{target} over slow time t are captured, denoted as $X_{target}(t)$:

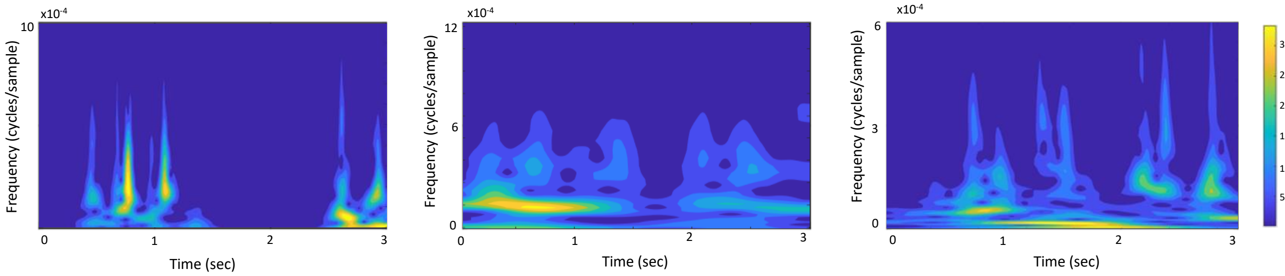
$$X_{target}(t) = S(t) \cdot e^{j\phi(t)}, \tag{6}$$

where $S(t)$ and $\phi(t)$ are the RSS and phase variation over time. The range bin of the jaw d_{target} can be approximated as constant when extracting $X_{target}(t)$.

C. Preliminaries

In our preliminary experiment, we established a controlled environment within a quiet and unoccupied room, with only one user present. Following the prompt of a beep sound, the user was instructed to execute the designated gesture, which was then automatically captured by the mmWave radar system. This experiment's primary objective was to assess the feasibility of our approach and ensure a noise-free environment, thus preventing signal contamination from nearby object movements. The results of the experiment revealed distinctly clear waveforms corresponding to different gestures, namely sliding, tapping, and clenching. This preliminary investigation served as a foundational step in validating the usability and intuitiveness of our approach, providing valuable insights into users' proficiency and comfort levels with the designated gestures. All the experiments are IRB-approved.

The phase information captured during the experiment was denoised as mentioned in Sec IV(A). The wavelet transform analysis of the denoised signals reveals distinct patterns for each gesture. In Fig. 6(a), the tap signal shows sharp, narrow peaks, indicative of brief, high-frequency transients characteristic of tapping. This is expected as taps produce short,



(a) Tapping

(b) Clenching

(c) Sliding

Fig. 6. The wavelet transform of de-noised phase of different gestures

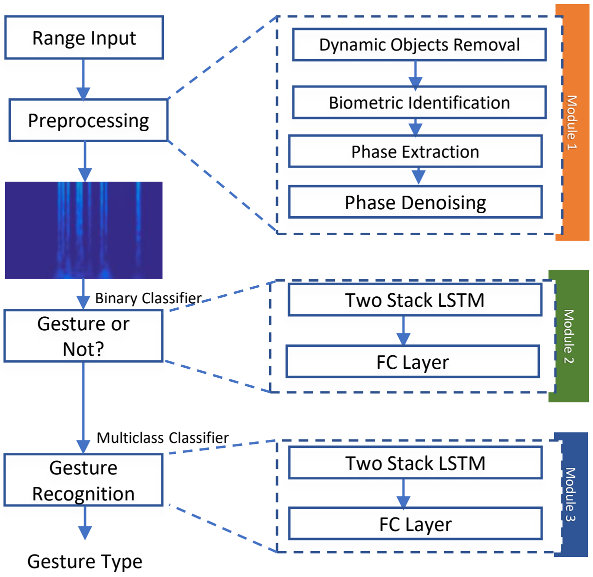


Fig. 7. Preprocessing and classification of detected gestures

impulsive signals with significant high-frequency components. In contrast, Fig. 6(b), representing the clench signal, displays broader, continuous regions of energy, primarily in the lower frequencies. This pattern aligns with the sustained muscle activation typical of clenching, resulting in a more prolonged and lower-frequency signal. Lastly, Fig. 6(c), which illustrates the sliding gesture, exhibits sustained, continuous energy over time with variable frequency content. This continuous band of energy reflects the nature of the sliding motion, where the signal varies in frequency depending on the speed and pressure of the gesture. These distinct wavelet transform patterns enable the differentiation between the gestures, highlighting the utility of wavelet analysis in gesture recognition and signal processing applications

IV. Design

Although our initial experiments showed promising results, real-world scenarios present additional challenges. One such challenge involves detecting subtle facial muscle movements induced by activities like tapping or clenching teeth, which occur at the millimeter level and are imperceptible to the naked eye. However, indoor environments introduce significant ambient noise to the raw mmWave signals. This noise stems from dynamic and stationary objects, including people moving around, as well as multi-path reflections from home appliances and walls.

To address this challenge, we comprehensively assess the impact of ambient noise, we conducted experiments in a living room measuring approximately 2.9m x 4.2m, using the same experimental setup as our preliminary study. Participants were instructed to move randomly within the room, allowing us to evaluate the effectiveness of our system under realistic conditions.

As illustrated in the Fig.7, our system design approach has 3 modules. Module 1 is the pre-processing step where we start with obtaining the phase signal and process it to eliminate the dynamic background noise. After this step, we are left with the static objects only. We apply bio-metric verification using heart-beat and breathing signals from the user. Upon the confirmation of the bin, we proceed with phase de-noising, followed by Gesture Detection (Module 2) and Gesture Classification (Module 3).

A. Dynamic Object Removal

Since the subjects performing the gestures are assumed to be static, we need to eliminate all the moving targets such as people walking in the background, door opening, moving fan etc. By visualizing the range information over time as shown in Fig. 8, the second FFT(Doppler FFT) enables the distinction between stationary and moving objects based on their velocity such as people walking [16].

B. Static Object Removal

For static objects, we've implemented a bio-metric verification approach, which involves testing for both heartbeat and respiration signals. The procedure of biometric verification is given as follows: after the elimination of the dynamic objects, we are only left with static range bins which are our potential range bins. In these range bins, two low-pass filters are applied in parallel for the detection of breath and heartbeat micromotion signals. since the breath rate of a normal human being is 12 to 18 breaths per minute and the heart rate is 60 to 100 beats per minute, the first filter passed the frequencies of 0.1 Hz to 0.6 Hz. The other filter passing frequencies were

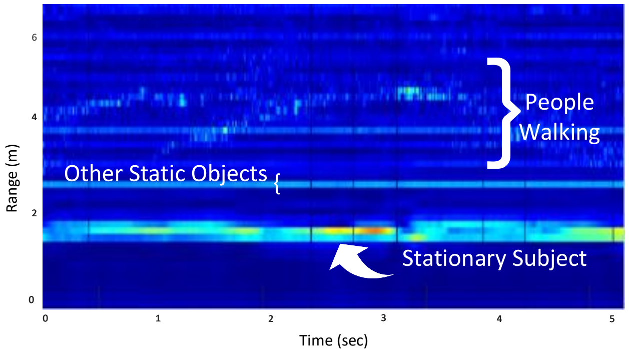


Fig. 8. Differentiation between static and dynamic objects

0.8 to 4.0 Hertz. Once the breathing rate or heart rate signal is detected in a specific bin, With the known range resolution of Rs, we can infer that the facial data will likely appear in the neighboring range bins of the heartbeat and respiration bins as shown in Fig. 9. All other range bins are ignored.

C. Noise removal from identified bin

The signal we obtain from the static and dynamic object removal is noisy. We use a wavelet-based denoising method to eliminate the undesired noise. It begins with the discrete wavelet transform (DWT), which decomposes the signal y into approximation coefficients $A_j(y)$ and detail coefficients $D_j(y)$ at each level j. The DWT can be represented as:

$$y = \sum_j A_j(y) + \sum_j D_j(y), \tag{7}$$

where $A_j(y)$ captures the coarse-scale information and $D_j(y)$ represents the detail or high-frequency components.

Next, the noise level σ in the signal is estimated using a method that assumes independence from the wavelet decomposition level. This estimate σ serves as a crucial parameter in thresholding the wavelet coefficients to suppress noise effectively.

For thresholding, a Bayesian approach is employed, specifically using the median of the wavelet coefficients. The threshold λ is calculated as:

$$\lambda = \sigma\sqrt{2\log(n)} \cdot Q^{-1}(p), \tag{8}$$

where n is the number of wavelet coefficients and $Q^{-1}(p)$ is the quantile function of a standard normal distribution corresponding to a chosen confidence level p. This threshold λ determines which wavelet coefficients are set to zero based on their magnitudes relative to λ.

After thresholding, the denoised signal $\hat{y}$ is reconstructed using the inverse discrete wavelet transform (IDWT), combining the modified approximation coefficients $A'_j(y)$ and detail coefficients $D'_j(y)$:

$$\hat{y} = \text{IDWT}(A'_j(y), D'_j(y)). \tag{9}$$

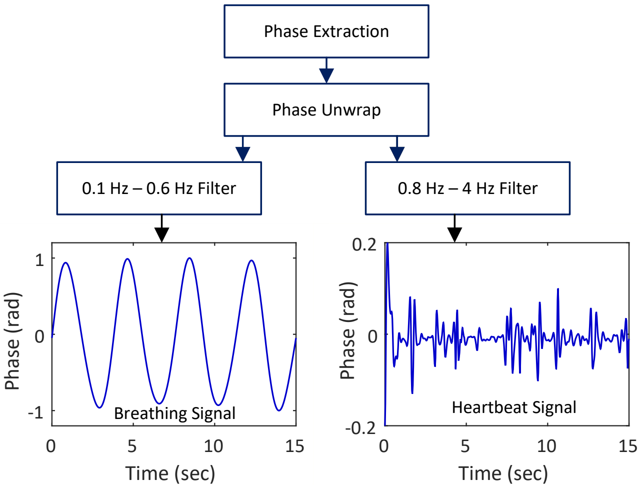


Fig. 9. For static candidate bins, the biometric features (breathing and heartbeat) extraction using low pass filters

The IDWT integrates the filtered approximation and detail coefficients back into the time domain, yielding a denoised signal $\hat{y}$ that retains essential features while reducing noise.

After the identification of the target bin by static object removal and biometric range bin selection, the wavelet transform of the target bin is carried out. The rationale for employing the wavelet transform in our analysis lies in its exceptional capability to provide both time and frequency localization of signals.

We again apply the wavelet transform to the denoised signal to better visualize both temporal and spatial features:

$$W_x(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-b}{a}\right) dt, \tag{10}$$

where $W_x(a, b)$ are the wavelet coefficients,a is the scale parameter,b is the translation (or shift) parameter and $\psi^*(t)$ is the complex conjugate of the mother wavelet.

Unlike traditional Fourier transform methods, which only offer frequency domain information, wavelet transform allows us to examine how signal characteristics evolve. This is particularly crucial for our application, as the gestures we are analyzing—taps, clenches, and sliding motions—have distinct temporal patterns and frequency components as evident in Fig 6.

D. Gesture Detection

Once the target is located after static and dynamic irrelevant object removal approaches mentioned in the aforementioned section, we find that the teeth signal is affected by various mouth movements, such as chewing, speaking, swallowing, smiling, yawning, coughing, sneezing, bruxism (teeth grinding), sucking, whistling, and different breathing patterns. These movements produce signals that can overlap or mimic the teeth-tapping signal, complicating the task of accurately

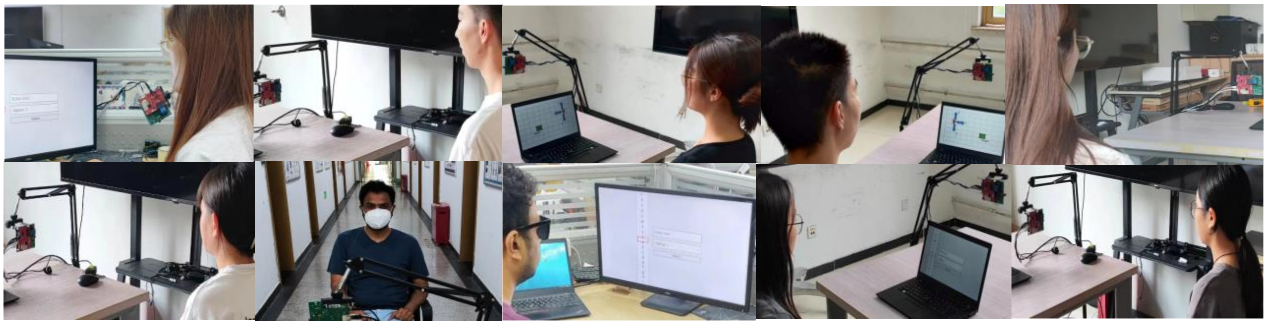


Fig. 10. Data collection and evaluation setup for different ranges, orientations, backgrounds, locations and with various accessories such as mask, glasses, etc.

distinguishing between them. The complexity arises from the diverse and dynamic nature of these activities, each generating unique yet sometimes similar signal characteristics. For this, we employed a binary classifier consisting of two LSTM layers followed by a dense output layer. Each LSTM layer, equipped with 128 units and tanh activation, sequentially processes input data, capturing intricate temporal dependencies. A fully connected layer with ReLU activation is added to further process the LSTM outputs before a final dense layer with sigmoid activation produces a probability score for binary classification The final dense layer employs a sigmoid activation function to produce a probability score, indicating the likelihood of the binary class.

E. Gesture Classification

In our study, an LSTM network is employed to classify high-dimensional sequence data with labels ranging from 1 to 3 corresponding to the 3 defined gestures. Similarly, the testing data consists of 65 samples with the same feature count and corresponding labels. The LSTM network architecture includes a sequence input layer for fixed-length sequences, followed by three LSTM layers: the first two with 200 hidden units each, outputting sequences, and the third with 200 hidden units, outputting the last hidden state. This is followed by two fully connected layers with 200 units each and ReLU activations, and an output layer consisting of a fully connected layer for 3 classes, a softmax layer, and a classification layer. The model is trained using the Adam optimizer, with a maximum of 50 epochs, a mini-batch size of 16, and an initial learning rate of 0.001, which gradually decreases. Validation checks are performed every 5 iterations to monitor performance on unseen data. Post-training, the model's classification accuracy is evaluated on test data, confirming its effectiveness in classifying high-dimensional sequences and demonstrating its potential for similar tasks.

V. Implementation and Evaluation

A. Implementation

We implement our system on a commercial mm-Wave radar Texas Instruments (TI) IWR6842 [17]. The ADC samples from the radar are captured by a TI DCA1000EVM data acquisition board [18]. For data processing, Intel Core i7-6500u was used with 32GB RAM. Python 3.7 and Matlab2022a are used for data processing. The configuration of mmWave radar is shown in Table 1.

B. Evaluation

The evaluation has been done in various settings for different mainly Detection and Classification modules shown in Fig. 10 with the following matrices.

C. Classification

For both classifiers (Gesture Detection and Gesture Classification) we use accuracy as a performance metric to evaluate the performance of mmJaw. It is calculated from True Positive(TP), True Negative(TN), False Positive(FP) and False Negative(FN) as follows: The accuracy A of a model is given by:

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

1) Impact of distance: In our experiments, we varied the distance between the subjects and the mm-wave radar, ranging from 1m to 2.5m. We noted a slight degradation in accuracy as the distance increased.

2) Impact of other activities: We now evaluate our system against different kinds of background noises. First, we check for the comparison of the system when there is a plane wall and when multiple people are walking behind the target as shown in Fig. 11 (e). mmJaw shows resilience against multiple background noises such as human movement or a moving fan and a door constantly opening and closing.

TABLE I
Radar Configuration Parameters

Parameter	Value	Parameter	Value
Frequency Slope	29.982 MHz/μs	Frame Periodicity	20.5 ms
ADC Samples/Second	5000K	Idle Time	20 μs
Chirp Cycle Time	80 μs	Start Frequency	77GHz
Frames	735	Sample/Chirp	256

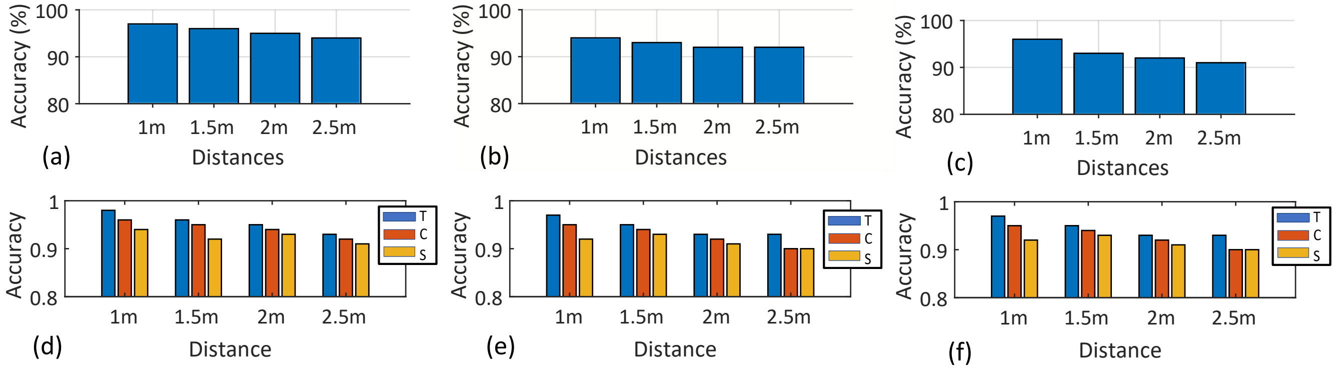


Fig. 11. Classification accuracy results at different ranges: (a) the accuracy of binary classifier in an indoor setting. (b) the binary classification's accuracy in a hallway setting. (c) the binary classification's accuracy with masks and glasses. (d) the gesture recognition classifier's accuracy in an indoor setting. (e) the classification accuracy in a hallway setting. (f) the classifier accuracy with other face accessories such as glasses and face masks

3) Impact of posture: It has been observed that as long as the user is facing the radar, irrespective of neck position, the gestures are detectable. This provides freedom to the user to move his/her neck from shoulder to shoulder. Fig. 10 shows the placement of the radar at multiple orientations relative to the users.

4) Impact of different facial accessories: Intuitively, facial accessories such as glasses and face masks should hinder the mmWave contact with the face, rendering it unusable under these circumstances. However, it has been observed that a mask touching the skin moves along, providing the same symmetric temporal pattern as the skin. Glasses, on the other hand, don't hinder the facial area of movement.

D. Results

The results presented in Fig. 11 illustrate the classification accuracies of Module 2 and Module 3 under varying conditions and distances. Fig. 11(a) and Fig. 11(b) depict the accuracy of a binary classifier in indoor and hallway settings, respectively. Both settings show a general trend of decreasing accuracy as the distance increases from 1 meter to 2.5 meters. This suggests that the classifier's performance is slightly sensitive to the proximity of the subject, with closer distances yielding higher accuracy.

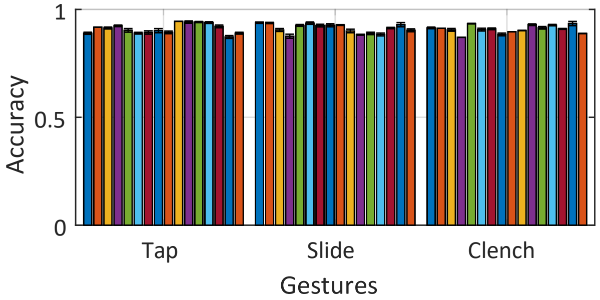


Fig. 12. Gesture accuracy for 16 users

In Fig. 11(c), the binary classifier's accuracy with masks and glasses is examined, showing a similar decreasing trend with increasing distance. This indicates that wearing masks and glasses does not significantly alter the overall trend of accuracy drop with distance, though it slightly affects the absolute accuracy levels.

Fig. 11(d), (e), and (f) focus on the gesture recognition classifier and its performance under various conditions. Fig.11(d) compares the accuracy across different distances in an indoor setting, using three types of gestures (T, C, and S). The accuracy remains relatively stable across different distances, though there are minor variations among the different gestures. Similarly, Fig. 11(e) shows the classifier's performance in a realistic setting (hallway), maintaining a consistent accuracy across distances with some fluctuations among gesture types.

Fig. 11(f) examines the classifier's accuracy when subjects wear additional face accessories such as glasses, masks, and a VR headset. The results indicate that these accessories do not dramatically affect the classifier's accuracy, which remains relatively consistent across different distances.

Fig. 12 shows the individual user accuracy for all users. Overall, the results highlight the robustness of the classifiers under various conditions. The presence of face accessories like masks and glasses appears to have a minimal impact on the overall accuracy, underscoring the classifiers' adaptability to real-world scenarios where such accessories are common.

VI. Discussion

The primary limitation of this work lies in the restricted number of gestures, stemming from the inherent difficulty humans face in performing concealed movements. Our study specifically focused on three basic teeth gestures: tapping (T), clenching (C), and sliding (S). However, by repeating these actions, additional gestures can be derived, such as double tap (TT), double clench (CC), and double slide (SS). Furthermore, combining these gestures in various sequences—like TS, TC, ST, SC, CT, and CS—expands the total number of recognizable gestures to 12. Extending this further to include gestures with up to three repetitions could significantly increase the number of possible gestures, though it might affect user comfort and ease of use.

This approach can also be extended to handle multiple users in different range bins. Moreover, the fixed Doppler threshold currently limits the system to recognizing the target only as a static object. Addressing this limitation remains a task for future research.

VII. Related Work

Multiple studies have focused on detecting and monitoring teeth activities. We categorize these studies into intraoral, involving sensors placed inside the mouth, and extraoral approaches, where sensors are positioned outside the mouth, such as over the face or in the ears.

A. Intra-Oral devices:

TongueBoard [19] is an oral interface that allows for subtle gestures and silent speech input through a palate retainer embedded with sensors to track tongue movements, enabling discreet and non-obtrusive interaction with digital devices. CanalSense [20] captures subtle changes in air pressure caused by movements such as smiling, frowning, or speaking, and translates these signals into recognizable patterns. TeethVib [12] introduces a novel method for monitoring teeth occlusion (the contact between teeth) by using vibration sensors embedded in dental retainers. It captures and analyzes vibrations during occlusion to reveal the functional dynamics of teeth interactions. Another study [21] directly embeds sensors into teeth to monitor oral activities. ChewIt [22] introduced an edible intraoral input interface that detects chewing movements, enabling hands-free, discreet inputs using the natural act of chewing for interaction with digital devices, eliminating the need for external sensors.

B. Extra-oral devices:

Bity [9] achieves high accuracy in distinguishing between different tooth clicks, enabling applications such as list navigation and keyboard input methods. This work highlights the potential of subtle, wearable devices for intuitive and accurate user input. Clench Interface [1] introduced a system that utilizes biting gestures for input. This technique leverages the act of clenching the teeth to interact with digital interfaces, providing a hands-free and subtle method of control. TeethTap [10] uses motion and acoustic signals to recognize discrete teeth gestures, improving gesture recognition with less obtrusive hardware. WiFace [23] recognizes facial expressions by detecting changes in Wi-Fi signals caused by facial muscle movements without the need for cameras or physical sensors placed directly on the face. mmFER [16] uses millimeter-wave radar for facial expression recognition in IoT. EarSense [4] monitors teeth activity via earphones. SonicFace [11] uses sound for facial expression detection.

VIII. Conclusion

In this paper, we demonstrate that jaw clenching tapping, and sliding teeth movements can be detected by mmwave radar with high accuracy. The pattern of these voluntary movements, though barely visible to the naked eye, can be analyzed to recognize muscle-related gestures for Human-Computer interaction. Despite challenges such as weak signals and interference from other movements, our sensing techniques achieve robust detection of three distinct gestures in different settings and distances up to 2.5m.

IX. Acknowledgement

This work is supported in part by the National Natural Science Foundation of China (No. 62425207 and No. U21B2007) and Postdoctoral Fellowship Program of CPSF (No. GZB20240356)

References

[1] X. Xu *et al.*, “Clench interface: Novel biting input techniques,” in *Proceedings of the ACM CHI*, 2019.

[2] C. Jiang *et al.*, “mmvib: micrometer-level vibration measurement with mmwave radar,” in *Proceedings of ACM MobiCom*, 2020.

[3] M. S. Duchowny *et al.*, “Video eeg diagnosis of repetitive behavior in early childhood and its relationship to seizures,” *Pediatric neurology*, 1988.

[4] J. Prakash *et al.*, “Earsense: earphones as a teeth activity sensor,” in *Proceedings of ACM MobiCom*, 2020.

[5] A. Hillebrand *et al.*, “Feasibility of clinical magnetoencephalography (meg) functional mapping in the presence of dental artefacts,” *Clinical Neurophysiology*, 2013.

[6] “Little signals,” 2024, accessed: 2024-07-03. [Online]. Available: https://littlesignals.withgoogle.com/

[7] “Ifttt,” 2024, accessed: 2024-07-03. [Online]. Available: https://ifttt.com/

[8] X. Shen *et al.*, “Clenchclick: Hands-free target selection method leveraging teeth-clench for augmented reality,” *Proceedings of ACM IMWUT*, 2022.

[9] D. Ashbrook *et al.*, “Bitey: An exploration of tooth click gestures for hands-free user interface control,” in *Proceedings of ACM MobileHCI*, 2016.

[10] W. Sun *et al.*, “Teethtap: Recognizing discrete teeth gestures using motion and acoustic sensing on an earpiece,” in *Proceedings of the ACM IUI*, 2021.

[11] Y. Gao *et al.*, “Sonicface: Tracking facial expressions using a commodity microphone array,” *Proceedings of ACM IMWUT*, 2021.

[12] S. Pan *et al.*, “Teethvib: Monitoring teeth functional occlusion through retainer vibration sensing,” in *Proceedings of IEEE/ACM CHASE*, 2021.

[13] J. Zhang *et al.*, “mmhawkeye: Passive uav detection with a cots mmwave radar,” in *Proceedings of IEEE SECON*, 2023.

[14] Y. He *et al.*, “Detection and identification of non-cooperative uav using a cots mmwave radar,” *ACM Transactions on Sensor Networks*, 2023.

[15] J. Zhang *et al.*, “A survey of mmwave-based human sensing: Technology, platforms and applications,” *IEEE Communications Surveys & Tutorials*, 2023.

[16] X. Zhang *et al.*, “mmfer: Millimetre-wave radar based facial expression recognition for multimedia iot applications,” in *Proceedings of ACM MobiCom*, 2023.

[17] *Texas Instruments Incorporated. 2020. IWR1642: Single-chip 76-GHz to 81- GHz mmWave sensor integrating DSP and MCU*. [Online]. Available: http://www.ti.com/product/IWR1642

[18] *Texas Instruments Incorporated. 2020. Real-time data-capture adapter for radar sensing evaluation module*. [Online]. Available: http://www.ti.com/tool/DCA1000EVM.

[19] R. Li, J. Wu, and T. Starner, “Tongueboard: An oral interface for subtle input,” in *Proceedings of ACM AH*, 2019.

[20] T. Ando *et al.*, “Canalsense: Face-related movement recognition system based on sensing air pressure in ear canals,” in *Proceedings of ACM UIST*, 2017.

[21] C.-Y. Li *et al.*, “Sensor-embedded teeth for oral activity recognition,” in *Proceedings of ACM ISWC*, 2013.

[22] P. Gallego Cascón *et al.*, “Chewit. an intraoral interface for discreet interactions,” in *Proceedings of ACM CHI*, 2019.

[23] Y. Chen *et al.*, “Wiface: Facial expression recognition using wi-fi signals,” *IEEE Transactions on Mobile Computing*, 2022.