SmarTeeth: Augmenting Manual Toothbrushing with In-ear Microphones

Qiang Yang
Department of Computer Science and Technology
University of Cambridge
Cambridge, United Kingdom
qiang.yang@cl.cam.ac.uk

Yang Liu
Department of Computer Science and Technology
University of Cambridge
Cambridge, United Kingdom
yl868@cam.ac.uk

Jake Stuchbury-Wass
Department of Computer Science and Technology
University of Cambridge
Cambridge, United Kingdom
js2372@cam.ac.uk

Kayla-Jade Butkow
Department of Computer Science and Technology
University of Cambridge
Cambridge, United Kingdom
kjb85@cam.ac.uk

Emeli Panariti
Faculty of Dentistry, Oral and Craniofacial Sciences
King's College London
London, United Kingdom
K21163606@kcl.ac.uk

Dong Ma
School of Computing and Information Systems
Singapore Management University
Singapore, Singapore
dongma@smu.edu.sg

Cecilia Mascolo
Department of Computer Science and Technology
University of Cambridge
Cambridge, United Kingdom
cecilia.mascolo@cl.cam.ac.uk

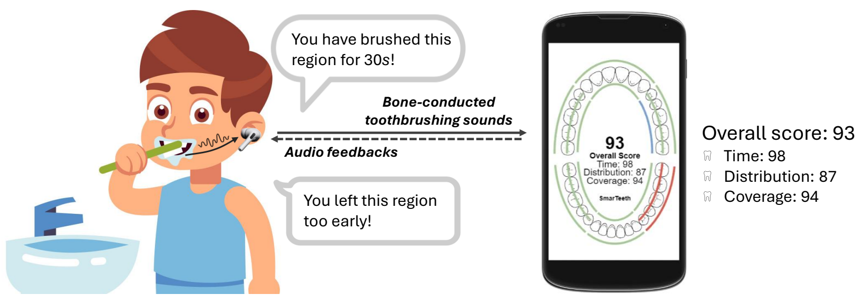


Figure 1: An illustration of SmarTeeth, an earphone-based toothbrushing monitoring system using in-ear microphones that augments manual toothbrushing by integrating brushing surface tracking, a feature typically found in high-end electric toothbrushes. SmarTeeth is based on a key observation–toothbrushing sounds that travel along bones from the oral cavity to ear canals can be captured by in-ear microphones for toothbrushing monitoring. The distinct propagation paths of brushing sounds from various dental locations to each ear canal provide the foundation for our methods to accurately identify different brushing surfaces. As long as the user wears a pair of earphones while brushing their teeth, SmarTeeth can alert users through earphones if they brush their teeth for too short or too long and can also evaluate the overall brushing performance, even with manual toothbrushes.

Abstract

Improper toothbrushing practices persist as a primary cause of oral health issues such as tooth decay and gum disease. Despite the availability of high-end electric toothbrushes that offer some guidance, manual toothbrushes remain widely used due to their simplicity and convenience. We present SmarTeeth, an earable-based toothbrushing monitoring system designed to augment manual toothbrushing with functionalities typically offered only by high-end electric toothbrushes, such as brushing surface tracking. The underlying idea of SmarTeeth is to leverage in-ear microphones on earphones to capture toothbrushing sounds transmitted through the oral cavity to ear canals through facial bones and tissues. The distinct propagation paths of brushing sounds from various dental locations to each ear canal provide the foundational basis for our methods to accurately identify different brushing locations. By extracting customized features from these sounds, we can detect brushing locations using a deep-learning model. With only one registration session (~ 2 *mins*) for a new user, the average accuracy is 92.7% for detecting six regions and 75.6% for sixteen tooth surfaces. With three registration sessions (~ 6 *mins*), the performance can be boosted to 98.8% and 90.3% for six-region and sixteen-surface tracking, respectively. A key advantage of using earphones for monitoring is that they provide natural auditory feedback to alert users when they are overbrushing or underbrushing. Comprehensive evaluation validates the effectiveness of SmarTeeth under various

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713893

conditions (different users, brushes, orders, noise, *etc.*), and the feedback from the user study (N=13) indicates that users found the system highly useful (6.0/7.0) and reported a low workload (2.5/7.0) while using it. Our findings suggest that SmarTeeth could offer a scalable and effective solution to improve oral health globally by providing manual toothbrush users with advanced brushing monitoring capabilities.

CCS Concepts

• **Human-centered computing** → *Ubiquitous and mobile computing design and evaluation methods.*

Keywords

Toothbrushing monitoring, Acoustic sensing, Earable devices

ACM Reference Format:
Qiang Yang, Yang Liu, Jake Stuchbury-Wass, Kayla-Jade Butkow, Emeli Panariti, Dong Ma, and Cecilia Mascolo. 2025. SmarTeeth: Augmenting Manual Toothbrushing with In-ear Microphones. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3706598.3713893

1 Introduction

Oral diseases pose a significant public health challenge for countries and populations worldwide, with tooth decay and severe periodontal disease being the most common conditions affecting millions of people [61]. According to WHO statistics [59], more than one-third of the world's population lives with untreated tooth decay. Toothbrushing is a simple and effective strategy to prevent these conditions. The established Bass toothbrushing technique emphasizes that all dental surfaces should be brushed for an adequate amount of time, however overbrushing for long periods can lead to enamel erosion and gum recession [1]. Therefore, some high-end smart electric toothbrushes offer features such as brushing timers and surface detection [5] to guide users' brushing. However, a substantial portion of the world's population still prefers manual toothbrushes because of their simplicity and convenience. According to the US Census in 2020 [66], more than 250 million Americans still resort to manual toothbrushes. As a result, improving oral hygiene practices for manual toothbrush users remains a pressing and significant issue.

Although several approaches have been explored to empower manual toothbrushes with various toothbrushing monitoring capabilities, existing work has various limitations. Camera-based systems [12, 23], which place cameras in front of the user's mouth, although capable of capturing coarse-grained brushing activity, suffer from occlusions caused by hands and the mouth, limiting detection granularity and raising privacy concerns. IMU-based solutions [13, 22, 38], which integrate IMUs into manual toothbrushes or leverage IMUs in wearable devices like smartwatches, exhibit sensitivity to hand and head movements, imposing constraints on users' hand movements and postures while brushing [25, 48]. Some alternatives utilize toothbrushing sounds collected from a nearby smartphone [45] or external earpiece microphones [60] to monitor the toothbrushing process. However, these methods are vulnerable to interference from ambient noise sources such as running water, compromising their reliability [52].

In recent years, the market for Active Noise Cancellation earbuds has experienced significant growth, with an estimated market size of USD 17.88 billion in 2024, projected to reach USD 34.42 billion by 2029 [2]. This rapid proliferation has led to the widespread adoption of ANC earbuds, which are equipped with various sensors (e.g., in-ear and out-ear microphones), for various applications, including health monitoring [63]. A recent work, ToothFairy [71], demonstrates the feasibility of using these smart earbuds (*Earables*) to explore the intensity levels of in-ear sounds caused by *electric toothbrush* vibration to detect brushing locations. However, it cannot work with *manual toothbrushes.* Therefore, in this paper, we introduce SmarTeeth, an earable-based tooth-brushing monitoring system using in-ear microphones that augments manual toothbrushing by integrating functionalities typically found in high-end electric toothbrushes, such as brushing surface tracking. Using commercially available earphones to augment manual toothbrushing eliminates the need for users to procure additional purpose-made electric toothbrushes. The earphones, positioned at the upper extremity of the body, are free from interference from limb motions (unlike sensors on the arm/wrist/hand) [80]. Also, the in-ear microphones exhibit high resilience to environmental noise due to effective noise occlusion [53]. Moreover, a key advantage of using earphones for monitoring is that they provide a natural auditory interface to promptly alert users when they are overbrushing or underbrushing. As illustrated in Fig. 1, the high-level idea of SmarTeeth is to utilize the on-board in-ear microphones of earphones to capture brushing sounds. These sounds originate from friction between the brush bristles and the tooth surface and then transmit through bones and facial tissues to the ear canal. Since brushing different dental locations involves distinct bone-conduction pathways, the captured in-ear audio reveals specific characteristics for each teeth. This provides the foundational basis from which we can extract relevant features from the audio signals and employ deep learning techniques to accurately predict the brushing locations. With continuous tracking the brushing areas, it can provide timely audio feedback to inform the user if underbrushing or overbrushing is detected.

Turning the intuitive concept of SmarTeeth into reality needs to address the following challenges. First, variations in brushing locations, combined with differences in brushing force, speed, and toothbrush types, result in significant discrepancies in brushing sounds [13]. Achieving fine-grained toothbrushing tracking using sound features must account for these area-specific variations while remaining unaffected by these external factors [45, 60]. As a result, there is a need for a method to customize features that inherently characterize different brushing locations. Secondly, the sounds captured in the two ear canals vary significantly when brushing at different times due to changes in the wearing seal state [53], which can lead to performance degradation. To address this, we need to develop a method to unify the feature distribution across varying wearing conditions to eliminate this effect. Lastly, because of the similar brushing sounds of adjacent teeth surfaces, it is extremely challenging to distinguish them, since they only have slight differences in their sound properties and transmission paths.

To overcome these challenges, we propose the following technical approaches. First, we establish a signal propagation model to characterize how toothbrushing sounds travel from various teeth positions to both ear canals. Based on that, we customize features related to the propagation channel, which inherently reflects different brushing locations. Secondly, we extract the coherence level and phase from the propagation channels to unify the feature distribution, thereby minimizing the impact of variations caused by different wearing states. Lastly, given the temporal continuity inherent in the toothbrushing process, we exploit modified audio features and smoothing techniques to enhance the recognition results to differentiate surfaces in close proximity, achieving fine-grained toothbrushing tracking.

We collaborate with a dentist on this study design to ensure that it addresses clinically-relevant aspects of toothbrushing. We designed a custom pair of earphones and introduced a feedback mechanism that alerts users when overbrushing or underbrushing of a certain surface is detected. Additionally, we developed an app that visualizes the brushing score and duration for each dental surface to provide users with comprehensive feedback. We conducted a comprehensive evaluation and user study of SmarTeeth (N=13). The participants and the dentist also provided suggestions on form factor optimization and functionality enhancement to ensure that SmarTeeth integrates seamlessly into users' daily routines with better use experience (Sec. 6.6). To summarize, this paper makes the following contributions:

- To the best of our knowledge, SmarTeeth is the first toothbrushing monitoring system using in-ear microphones, which augments manual toothbrushing to provide fine-grained toothbrushing tracking, a feature originally owned by high-end electric toothbrushes. Our experiment shows the techniques of SmarTeeth can also be applied to low-end electric toothbrushes that do not have tracking functionality.
- The comprehensive evaluation (N=13) shows that SmarTeeth achieves an average accuracy of 92.7% for detecting six regions and 75.6% for sixteen tooth surfaces with only one registration session (2 *mins*) for a new user. With three registration sessions, the performance can be boosted to 98.8% and 90.3% for six-region and sixteen-surface tracking, respectively.
- We propose a deep learning approach using customized channel-related features for fine-grained toothbrushing tracking, which is robust to the variability in brushing habits and the wearing states of earphones. We also design three feedback mechanisms to improve users' brushing habits: underbrushing/overbrushing alerting, brushing duration visualization, and brushing score evaluation. The subjective user study (N=13) demonstrated that our system is highly useful (6.0/7.0) and has a low workload (2.46/7.0).

2 Related Work

2.1 Toothbrushing Monitoring

2.1.1 Vision/light-based Approaches. Vision-based solutions employ cameras to track brushing processes. For instance, Playful Toothbrush [23] utilized a web camera placed in front of the user's mouth to track an LED-coded toothbrush extension, aiding users in learning proper brushing techniques. Akifusa *et al.* [12] attached a tiny camera within the head of a UV-LED toothbrush to visualize the plaque removal efficacy of electric toothbrushes. LiT [25] adopts two photosensors in commercial LED toothbrushes to monitor the toothbrushing process. However, camera-based systems are susceptible to occlusions caused by the hand and mouth, and privacy concerns may arise due to the invasive nature of video recording. Additionally, some LEB-based work also requires modifications to attach dedicated sensors on toothbrushes, limiting their usability.

2.1.2 IMU-based Approaches. Another subset of the literature leverages IMUs to track brushing motions. IMUs can be used on the modified toothbrush handle to estimate brushing motions [24, 40, 44, 47]. Social Brush [22] utilized an IMU attached to the brush handle to detect different brushing regions. Li *et al.* [48] attached an IMU sensor and five pressure sensors on the brush handle to estimate brushing regions and forces with Random Forest models. However, integrating IMUs may require modifying the toothbrushes. Another approach is to utilize the IMU of the smartwatch on the user's wrist to monitor the toothbrushing process. Huang *et al.* [38] employed a Naive Bayes classifier to recognize brushing surfaces using accelerometer, gyroscope, and magnetometer data from a wristwatch. MET [39] tracks brushing coverage for 15 surfaces of teeth with a magnetic sensor array. Hygiea [52] exploits wrist-worn IMUs to achieve fine-grained toothbrushing activity recognition with an LSTM model. Similarly, mORAL [14] can detect oral health behaviors such as brushing and flossing passively from wrist-worn IMUs. mTeeth [13] detects teeth surfaces being brushed with a manual toothbrush in the natural free-living environment using wrist-worn inertial sensors. BrushBuds [78] uses IMU sensors on earphones to track six toothbrushing regions but struggles with tracking fine-grained surfaces. While using wrist/earable-IMU does not need toothbrush modification, these systems are sensitive to hand or head movements, posing a constraint on the natural brushing posture. Additionally, the accumulation drift and vibrations of electric toothbrushes negatively affect detection accuracy [25, 48].

2.1.3 Audio-based Approaches. Compared to the IMU-based approaches, audio-based methods are resilient to the effects of user motion. Korpela *et al.* [45, 46] applied hidden Markov models (HMM) to recognize brushing surfaces based on audio collected from smartphones placed nearby. Ouyang *et al.* [60] utilized two throat microphones and the external microphones of an earphone for toothbrushing monitoring. However, the external microphone can be easily disturbed by ambient noise [52]. ToMoBrush [83] embeds a microphone in the brush head to record toothbrushing sounds and detect dental diseases. EarSense [62] utilizes in-ear audio to recognize tooth activities and validate the feasibility of toothbrushing monitoring, but it can only distinguish coarse-grained horizontal areas. Inspired by EarSense, ToothFairy [71] explores the intensity levels of in-ear sounds caused by electric toothbrush vibration to detect brushing locations. However, it relies on the vibration of electric toothbrushes and cannot work on manual toothbrushes. In comparison, SmarTeeth augments *manual toothbrushing* with a new modality, in-ear audios, to achieve 16-surface (including both inner and outer teeth surfaces) fine-grained toothbrushing monitoring also works properly on electric toothbrushes.

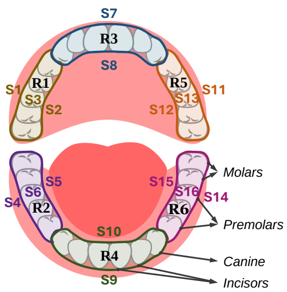
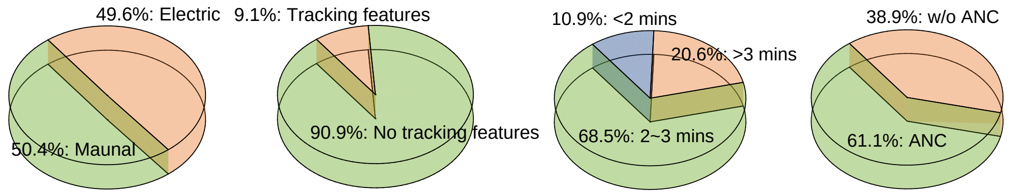


Figure 2: Sixteen surfaces.



(a) Toothbrush type **(b) Tracking features** **(c) Brushing duration.** **(d) ANC function.**

Figure 3: Feedback from 359 individuals in the motivational survey.

2.2 Earable Applications with In-ear Microphones

Various sensing modalities have been widely explored for numerous IoT applications, such as RF [26, 41, 77], camera [19], LoRa [81, 82], and IMU [50, 51]. Among them, acoustic sensing has gained significant attention [33, 37, 75, 79, 85], with in-ear microphones in earphones being utilized as an effective tool for human sensing [21, 29]. OESense [53] leverages the occlusion effect of the sounds inside the human ear to recognize human gestures and activities. Authentication systems such as HeartPrint [20], EarEcho [34], EarGate [32], and EarDynamic [73] authenticate users based on unique acoustic signatures in the ear canals. Additionally, researchers have also explored the use of in-ear microphones to measure physiological parameters, including respiratory [49, 54], heartrate [18, 67], dietary [17], and lung function [76]. Using an in-ear speaker and microphone, Amesaka *et al.* [16] propose a system that detects facial gestures based on ear canal deformations. ToothSonic [72] utilizes in-ear tooth-tapping sounds as a fingerprint to perform earable authentication. Aligned with the advancements in earphone-based sensing and computing platforms, our proposed work harnesses widely-used earphones to augment manual toothbrushing for fine-grained brushing surface detection, bringing intelligent toothbrushing monitoring technologies to a substantial portion of manual toothbrush users.

3 Preliminaries

3.1 Oral Anatomy and Structure

As shown in Fig. 2, the human oral cavity typically contains 32 teeth in adults, including incisors, canines, premolars, and molars [55]. Incisors are located at the front and are used for cutting, while canines are pointed teeth adjacent to the incisors, serving to tear food. The premolars and molars are located toward the back of the mouth and are used to grind and chew food. Oral medicine typically divides the dentition into six regions (sextants, R1 to R6) [28]: left, middle, and right, for both the upper and lower dental arches. The left or right side contains premolars and molars, and the middle region includes four incisors and two canines. As a result, most toothbrushing monitoring systems [13, 25, 38, 60] detecting brushing activity across these six regions. In addition, each region consists of multiple surfaces, totaling 16 surfaces overall (S1 to S16 in Fig. 2). For instance, the left and right sides include three surfaces: inner, outer, and chewing surfaces, while the middle region only has inner and outer surfaces. This fine-grained division enables a comprehensive assessment of brushing coverage.

3.2 Toothbrushing Duration and Quality

Incomplete coverage and insufficient brushing time are the primary causes of dental diseases. Failure to adequately clean all tooth surfaces can lead to the accumulation of plaque and tartar, increasing the risk of tooth cavities, bleeding gum, and other oral health issues [8]. Conversely, excessive brushing, or overbrushing, can lead to gum recession and enamel erosion, leading to dental sensitivity [1]. The Bass technique, recommended by the American Dental Association (ADA) [8], involves placing the toothbrush at a 45-degree angle to the gumline and using short back-and-forth or circular motions to clean both the teeth and gums effectively. This brushing technique should touch all surfaces—inner, outer, and chewing-ensuring the removal of plaque from all the tooth surfaces and along the gumline for optimal oral hygiene. The consensus recommendation is that people should brush their teeth twice a day each for two minutes, and each surface should be brushed for an even time [8].

As highlighted in related works [13, 25, 38], monitoring brushing regions/surfaces and ensuring sufficient brushing time are foundational steps in improving oral hygiene. Many commercial high-end electric toothbrushes, such as OralB io10 [58], also use surface tracking to guide users toward better brushing cleanness and ensure all areas are brushed adequately, thereby enhancing overall brushing quality [9]. The dentist collaborator highlighted that brushing time and coverage directly link to the brushing quality and detecting brushing regions/surfaces and duration is the most important aspect for addressing poor brushing habits, which are particularly critical for manual toothbrush users, who often lack guidance on coverage and timing.

4 Survey: Understanding Toothbrushing Practice

To better understand user habits, preferences, the challenges associated with current toothbrushing practices, and the potential for the adoption of (ANC) earphones, we conducted a survey (N=359) through the university online forum. This survey aimed to assess several key aspects: participants' oral health status, brushing habits, preferences between manual and electric toothbrushes, and their openness to adopting new technologies for monitoring toothbrushing. We report key findings from the following perspectives.

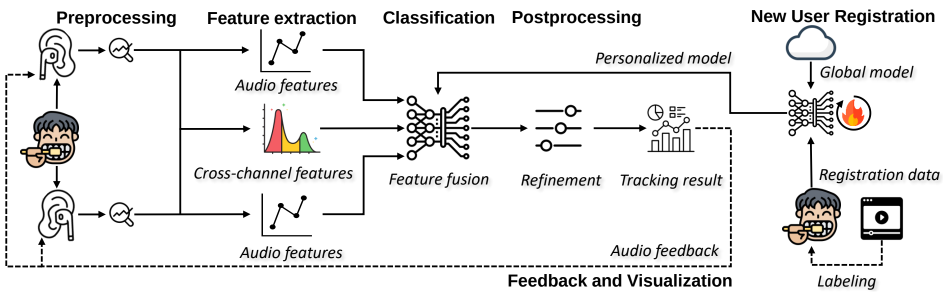


Figure 4: The system overview of SmarTeeth.

- **Demographics.** The survey included 359 participants aged between 18 and 65, with a balanced gender distribution of 47% men, 49.6% women, and the remainder identifying as non-binary or preferring not to say. The largest age group was 25–34 years (43.1%), followed by 18–24 years (20.2%), and 35–44 years (13.2%). Additionally, 11.7% of participants were aged 55 or older. The responses provide good gender and age diversity.
- **Oral Health Status and Brushing Habits**. The majority of respondents (76.5%) brush their teeth twice a day, while a smaller percentage brush once (17.7%) or three times (5.9%) a day. 59.4% of respondents have had cavities, and 40.7% have had sensitive teeth, reflecting a significant prevalence of dental issues within the surveyed group (in total 72.7%). However, regular dental visits are not a common practice, with only 43.1% visiting a dentist yearly and 33.9% going only when they have a dental problem. This indicates that although the majority of the participants presented with a history of dental needs, preventative dental care does not appear to be a common practice. Therefore, there is a need for a method that can help them monitor their dental practice on a daily basis.
- **Toothbrush Type and Limitations**. As shown in Fig. 3(a), A slight majority of participants use manual toothbrushes (50.4%) due to reasons such as simplicity (43%) and/or feeling that electric toothbrushes are unnecessary (17%). Other reasons mentioned include "I have implants", "do not want to charge", "more eco-friendly", "I find the vibration of an electric toothbrush extremely annoying", and "electric toothbrush makes me dizzy". In addition, as shown in Fig. 3(b), among all participants, only 9.1% of them used toothbrushes with tracking features, highlighting a gap in brushing monitoring capabilities in both manual and most electric toothbrushes.
- **Awareness of Brushing Practices**. Figure 3(c) shows while 47.7% of manual toothbrush users think they are aware of their brushing duration, 20.6% of them tend to overbrush their teeth (3∼5 *min*) and 10.9% of them tend to underbrush their teeth (< 2*min*), and only 68.5% of them brush teeth with a proper duration (2∼3 *min*). Furthermore, among these answers, most (71.8%) of participants are uncertain if they spend equal time on different areas of their mouth. However, interestingly, 90.7% of them believe the time spent on brushing is related to oral health issues. These results show many participants do not brush their teeth with a proper duration but recognize the importance of proper brushing habits.
- **Adoption of Earphones**. In recent years, the market for Active Noise Cancellation earbuds has experienced significant growth, with estimated market size of USD 17.88 billion in 2024, projected to reach USD 34.42 billion by 2029 [2]. As shown in Fig. 3(d), a substantial 89.9% of participants of our survey own earphones, with 61.1% featuring ANC functionality. Additionally, 85.7% use their earphones daily or several times a week, indicating high potential for integrating earphone-based solutions into daily routines. 67.4% of the respondents said that they would be willing to wear earphones while brushing, and a substantial 89.9% think that they will benefit from our system if their earphones can help improve their oral health. This result demonstrates that participants are open to adopting earphones for daily toothbrushing monitoring.

In summary, our motivational study clearly indicates a significant gap in current oral hygiene practices. Daily monitoring of dental practices is needed to improve oral health status. Despite the availability of high-end electric toothbrushes with sophisticated features, the majority of individuals still rely on manual brushes or low-end electric toothbrushes without tracking functions due to their simplicity and cost-effectiveness. However, this choice leaves people vulnerable to improper brushing, which can lead to potential oral health problems. Fortunately, from the survey result, people convey a high openness to adopt a system that can enhance their brushing habits, especially one that integrates seamlessly with devices they already use daily, such as earphones, which motivates our work. SmarTeeth addresses this need by offering an accessible and innovative solution that empowers users to improve their oral hygiene practices with a pair of earphones, bridging the gap between manual and advanced smart toothbrushing technologies.

4.1 System Overview

The primary objective of our work is to *detect which tooth region/surface the user is brushing to provide brushing feedback and evaluation to the user.* Most existing smart electric toothbrushes typically achieve brushing monitoring at a four (left/right side plus upper/lower jaw) or six-region level. Only a few high-end models like the Oral-B io10 (550 USD) [15] can detect all 16 surfaces [58].

Our aim is to leverage the in-ear microphone of widely available ANC earphones to augment manual toothbrushing for achieving effective 6-region and even 16-surface toothbrushing monitoring,

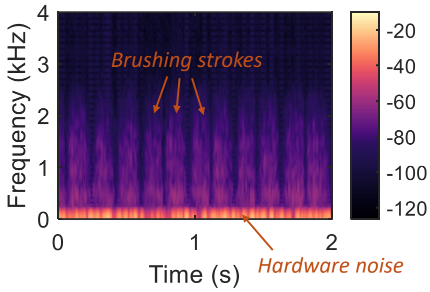


Figure 5: The spectrogram of in-ear toothbrushing sounds. Most sounds are distributed below 2.5 kHz, and the hardware noise is below 100 Hz.

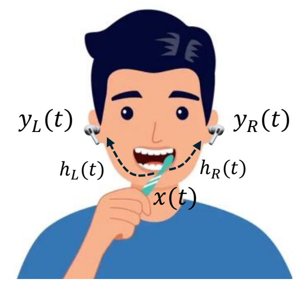


Figure 6: The signal propagation model of SmarTeeth. The in-ear microphones can capture binaural toothbrushing sounds in two ear canals.

thereby promoting oral hygiene and preventing dental diseases. Since no commercial earphones allow open access to the raw audio data [76], we designed and fabricated a custom pair of earphones for our evaluation (detailed in Sec. 6).

5 SmarTeeth System Design

Figure 4 shows the system overview of SmarTeeth. The toothbrushing sounds in ear canals will be recorded by the in-ear microphones of earphones and then forwarded to the preprocessing module. The preprocessing module (Sec. 5.1) is responsible for filtering the noise, detecting the toothbrushing events, and segmenting audio with sliding windows. After that, the feature extraction module (Sec. 5.2) will extract the audio features from the left and right channels and extract the channel-related features between both channels. Following this, the extracted features are fused in the classification module (Sec. 5.3) and used to predict the brushing regions/surfaces. Then, the prediction results are fed into the postprocessing module (Sec. 5.4) to refine the output with temporal constraints. For new users, they need to follow the video instructions for several toothbrushing sessions to personalize the model (Sec.5.5). Finally, the system visualizes the brushing habit and provides feedback for users (Sec. 5.6).

5.1 Signal Preprocessing

After collecting the toothbrushing sounds in the ear canals, we first perform preprocessing to filter the noise and detect the brushing activities. Figure 5 illustrates the spectrogram of a toothbrushing audio clip. We can see that most sounds are distributed below 2.5 kHz, and hardware causes strong noise interference below 100 Hz. Therefore, we design a 20-order bandpass Butterworth filter [57] with the cutoff frequencies of [100, 2500] Hz. SmarTeeth uses the short-time energy-based approach [84] to detect toothbrushing events. Specifically, we segment audio signals into $300ms$ sliding windows (the time spent on a typical back-and-forth brushing stroke) with a 50% overlap. The sounds travel through bones and soft tissues at speeds up to 1000 m/s, causing an extremely short delay between two ears [62], which will not affect the sliding windows. Then, the short-time energy is calculated for each window. We use the silence period upon starting as the reference to calculate the energy mean μ and standard deviation σ of each window. If the signal energy E exceeds $\mu + 8\sigma$ and lasts for one second in both left and right channels, we detect the last window of that second as the start point. Similarly, if E is lower than this threshold for one second in both channels, we regard the last window of that second as the endpoint. The list of windows between the start point and the endpoint will be fed into the feature extraction module for surface detection.

5.2 Feature Extraction

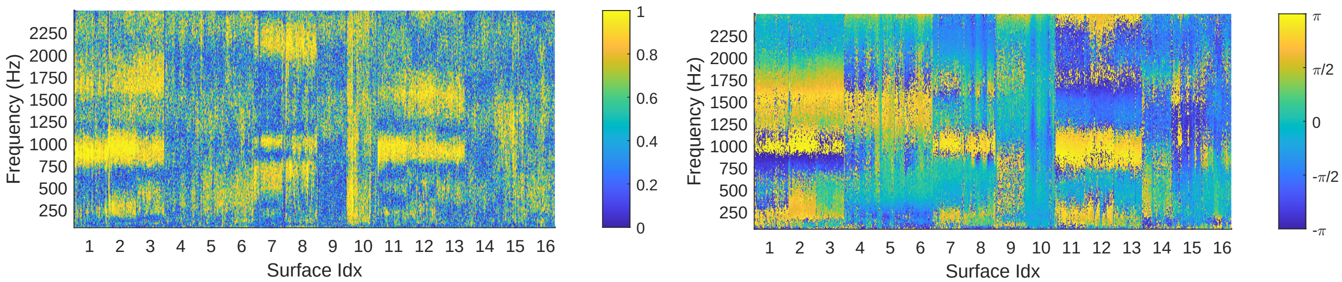
Upon detecting the toothbrushing sound, we can extract audio features from each sliding window and perform classification. Previous audio-based works [45, 60] typically directly utilize traditional audio features like Mel-frequency cepstral coefficients (MFCC) and time/frequency-domain statistic features (mean, deviation, *etc.*), which are vulnerable to many external factors, such as brushing force levels, bristle firmness, and wearing states. In this section, we make a key observation of the signal propagation model and extract the channel-related features, inherently characterizing different propagation paths from the teeth to both ear canals, which are more robust to the external variance.

5.2.1 Signal Propagation Model. As shown in Fig. 6, the fraction sound $x(t)$ between the tooth and brush bristle during toothbrushing will propagate through the bone and face tissues to the ear canals. Subsequently, the in-ear microphones can capture the toothbrushing sounds $y_L(t)$ and $y_R(t)$ for the left and right channels, respectively. We can model the propagation process as a Linear Time-Invariant (LTI) system [56] as follows:

$$\begin{bmatrix} y_L(t) \\ y_R(t) \end{bmatrix} = x(t) * \begin{bmatrix} h_L(t) \\ h_R(t) \end{bmatrix} + \begin{bmatrix} n_L(t) \\ n_R(t) \end{bmatrix} \quad (1)$$

where $n_L(t)$ and $n_R(t)$ are noise. $*$ is the convolution operation. $h_L(t)$ and $h_R(t)$ are the impulse response of the propagation channels between the tooth locations and both ear canals. Given different propagation paths, the channel information $h_L(t)$ and $h_R(t)$ fundamentally characterize the difference among the toothbrushing sounds of different teeth.

Ideally, we can use the Least Squares (LS) algorithm [43] to calculate $h(t)$ with $x(t)$ and $y(t)$. However, the problem is that we cannot capture the original fraction sounds $x(t)$ occuring on the tooth surface during toothbrushing, which hinders us from using $h(t)$ as a feature to predict toothbrushing surfaces. In this paper, instead of directly calculating the two separated channel responses, we utilize the redundancy of binaural audio outputs and propose to calculate the cross-channel responses, which implicitly characterize



(a) Cross-channel coherence **(b) Cross-channel phase**

Figure 7: An illustration of cross-channel features.

the properties of both channels and, more importantly, without the need for original fraction sounds $x(t)$.

Considering that most current earbuds can fit the user's ear canal very well (some earbuds can even detect the wearing seal state [3, 27]), the inward-facing in-ear microphone mainly captures the sound in the ear canal and is less susceptible to external noise due to the occlusion effect [53]. Therefore, we can omit the noise component in Eq. 1 and transform this equation to the frequency domain:

$$Y_L(f) = X(f)H_L(f) \tag{2}$$

$$Y_R(f) = X(f)H_R(f) \tag{3}$$

where $Y(f)$, $X(f)$, and $H(f)$ are the frequency representation of $y(t)$, $x(t)$, and $h(t)$ after Fast Fourier Transform (FFT), respectively. If we divide the Eq. 2 by the Eq. 3, then the original fraction sounds $X(f)$ can be canceled out as the common terms:

$$\frac{Y_L(f)}{Y_R(f)} = \frac{\cancel{X(f)}H_L(f)}{\cancel{X(f)}H_R(f)} = \frac{H_L(f)}{H_R(f)} = \alpha H_L(f)\overline{H_R(f)} = H_{LR}(f) \tag{4}$$

where α is a constant, $\overline{H_R(f)}$ is the conjugate operation, and $H_{LR}(f)$ is the obtained cross-channel frequency response.

5.2.2 Cross-channel Feature Extraction. As the cross channel $H_{LR}(f)$ is a complex matrix, we can use its magnitude and phase as the features. The cross-channel magnitude indicates the correlation level between the left and right channels at different frequencies, in other words, how the left channel aligns with the right channel in different frequency components. The cross-channel phase reflects the time lag between the sound propagating through two channels [42]. Figure 7(a) and Figure 7(b) illustrate the cross-channel magnitude and phase of 16 different surfaces in a complete toothbrushing cycle. We can observe that the frequency responses when brushing different surfaces are different, especially for the surfaces belonging to separated regions. For example, the cross-channel phases of S1, S2, and S3 in Fig. 7(b) have a majority of positive values since they are located at the left side of the mouth. Accordingly, most negative values are observed in the cross-channel phases of S11, S12, and S13, which are on the right side. We note that the cross-channel response is the superposition of all multipath between left and right channels, which is also the reason why it can be used as the feature to perform brushing surface classification.

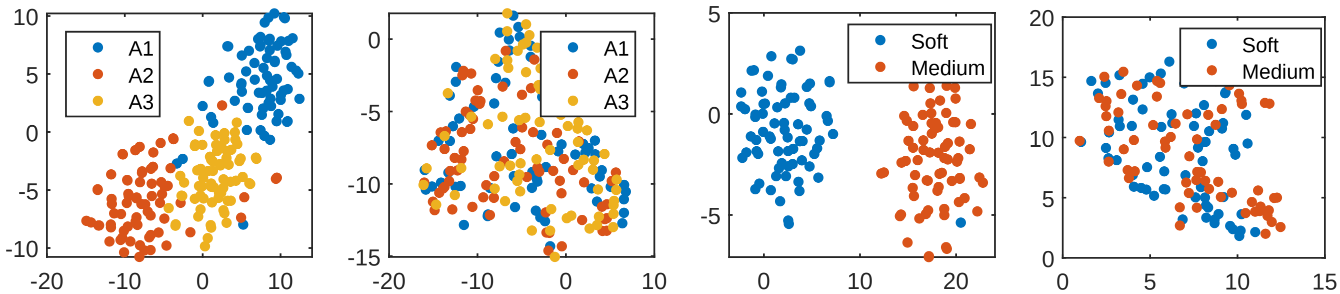
However, we cannot directly use the cross-channel features in Eq. 4. This is because the cross-channel magnitude is sensitive to the signal amplitude variation between left and right channels. Considering the user may brush teeth while wearing the earphones with different occlusion levels, the amplitude of in-ear sounds of two channels may vary up to $40dB$ [53]. Thus, the cross-channel magnitude will also differ across different wearing attempts. To deal with this problem, instead of directly using the absolute cross-channel amplitude, we calculate the cross-channel coherence to normalize it into [0, 1]:

$$C_{LR}(f) = \frac{|H_{LR}(f)|^2}{H_{LL}(f)H_{RR}(f)}, P_{LR}(f) = \text{angle}(H_{LR}(f)) \tag{5}$$

where $H_{LL}(f)$ ($H_{RR}(f)$) is the cross-channel feature but with two identical left (right) channels, *i.e.*, the amplitude of the left (right) channel. The angle$(\cdot)$ is the function to get the phase value of complex numbers. In this way, we can unify the cross-channel amplitude to [0, 1] and the phase to $[-\pi, \pi]$. Moreover, we use Welch's overlapped averaged periodogram method [74] to calculate cross-channel features to reduce the non-stationary variance.

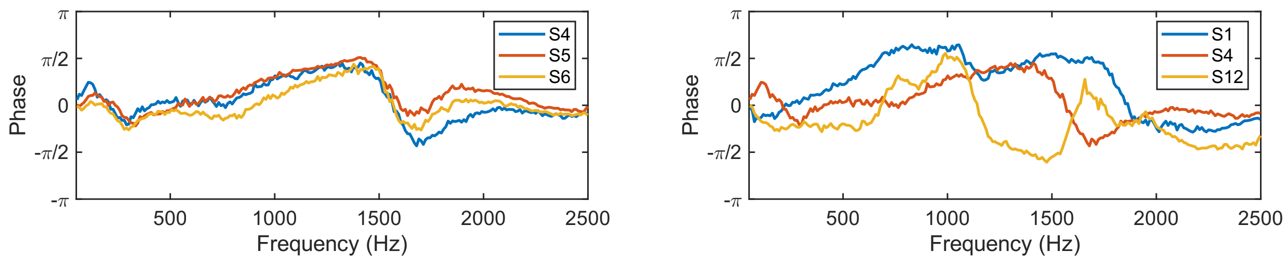
5.2.3 Advantages of Cross-channel Features. Compared to traditional statistic audio features, cross-channel features offer several advantages: *(1) Independence from the original toothbrushing sound.* Our approach does not require knowledge of the original toothbrushing sound. Instead, it extracts features from its propagation path, thus alleviating the impact of variations in brushing force, toothbrush material, and other factors on the toothbrushing sound. *(2) Normalization for consistency.* By normalizing cross-channel features to a uniform distribution, we eliminate the influence of different wearing conditions on sound intensity, enhancing the robustness of the features across varied scenarios. To validate it, we conducted a study involving a user brushing teeth with the toothbrushes of different firmness levels (*i.e.*, soft and medium). This study was repeated three times, with the user removing and re-wearing the earphones between each attempt (*i.e.*, A1, A2, and A3). Figure 8 provides a visualization of the MFCC feature and cross-channel features reduced to 2D using t-SNE [69]. We can see that the MFCC features exhibit clustering into three different groups across the three different attempts, while the cross-channel features have a consistent distribution. The same observation goes to the different bristle firmness, since the cross-channel features characterize the channel properties, robust to the variation of original brushing sounds.

5.2.4 Feature Formulation. After extracting the cross-channel features, we can use them to predict the brushing regions/surfaces.



(a) MFCC (b) Cross-channel features (c) MFCC (d) Cross-channel features

Figure 8: Feature visualization (MFCC *v.s.* cross-channel features) for different attempts (a, b) and bristle firmness (c, d).



(a) Intra-region cross-channel features (b) Inter-region cross-channel features

Figure 9: Intra-region and inter-region cross-channel features.

However, we observe that the intra-region difference of the cross-channel features is relatively less than the inter-region difference. As shown in Fig. 9(a), while the cross-channel phase features of different surfaces (S4, S5, and S6) in the same region R2 have some extent of difference, it is much smaller compared to the surfaces (S1, S4, and S12) in the different regions (Fig. 9(b)). This aligns with our expectations because cross-channel features characterize the propagation path between the teeth and the ear canal. Even though the different surfaces of the same tooth have slightly different propagation paths, they are still relatively similar compared to the different-region case. Considering that different surfaces have different shapes and textures, we also include MFCC and Gammatone Cepstral Coefficients (GTCC) features [68] from both channels to incorporate cross-channel features to recognize the surfaces close to each other. GTCC is derived from the gammatone filterbank, which is a set of auditory filters that mimic the frequency selectivity of the human auditory system, which have been found to be robust to audio noise and interference [64]. Given the significant variance in the sound intensity caused by different wearing conditions, we remove the first dimensions of MFCC and GTCC, which are directly related to sound energy. Therefore, we name them MMFCC (Modified MFCC) and MGTCC, which are concatenated together as a 24-dimensional feature for each channel. We conducted an ablation study to evaluate the performance contributions of channel-related features and modified MFCC/GTCC features in Sec. 6.5.

5.3 Classification

With the feature extracted in the previous step, we can perform deep learning to classify them into different regions/surfaces. Figure 10 shows the structure of the deep learning model. For each 300ms sliding window, we first concatenate the extracted MMFCC and MGTCC features of both channels together to a 48-dimension feature vector. For cross-channel features, since we need a high frequency resolution to capture the detailed response of the propagation channel, the feature dimension is 228 with a 10 Hz frequency resolution. Given the high feature dimension, we first use a dense layer to encode the cross-channel coherence and phase into two 48-dimensional embeddings. Recall that cross-channel coherence indicates the correlation level between both channels at different frequencies, so we can naturally regard the coherence as a weight to the cross-channel phase at different frequencies. Thus, we use the *softmax* and *relu* as the activation functions for coherence and phase embeddings, respectively, and multiply them as the final cross-channel features. Next, the cross-channel features extracted from both channels as well as the MMFCC and MGTCC features extracted from separated channels are fused by concatenation to generate a 96-dimension feature vector for classification. We use three dense layers as the classification backbone. To prevent overfitting, each dense layer is followed by a batch normalization layer and a dropout layer with a dropout rate of 0.2.

5.4 Postprocessing

To further enhance tracking performance, we propose a postprocessing method for calibrating sudden intra-region shifts and outlier classification results. Figure 11(a) shows the classification results from the learning model for a complete toothbrushing cycle. We

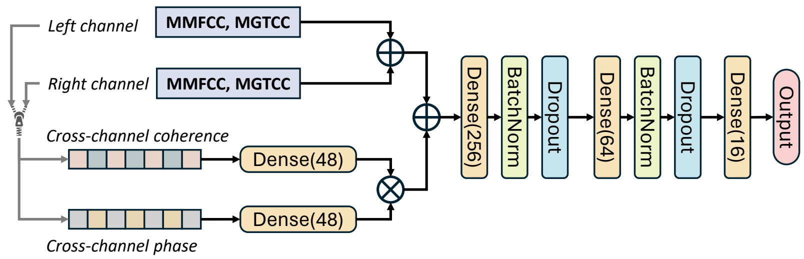
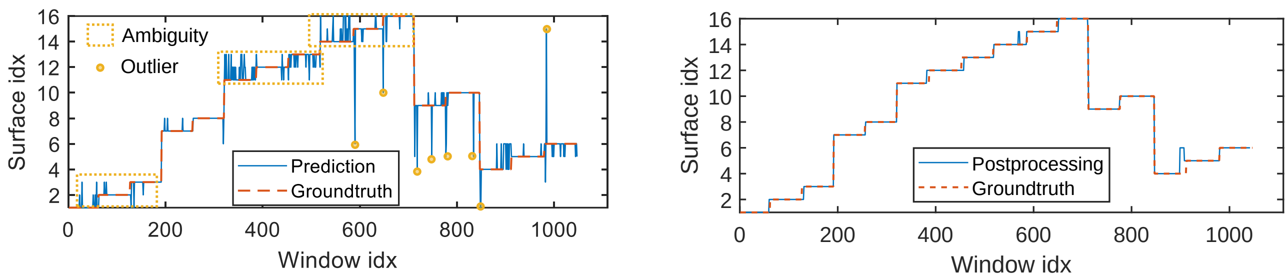


Figure 10: The deep learning model structure of SmarTeeth.



(a) Before postprocessing

(b) After postprocessing

Figure 11: Postprocessing with temporal constraints.

can see that the classification results exhibit some sudden shifts due to the inherent channel ambiguity caused by the adjacent tooth surfaces and occasional prediction outliers. Given that the prediction results represent a continuous time series of brushing surfaces, we can impose temporal continuity constraints to address these sudden shifts and outlier points. Specifically, it is unlikely for the brushing surface to shift dramatically between different regions within a very short time frame (*e.g.*, 150*ms*). By smoothing out these abrupt changes and anomalies, we can enhance the overall coherence and accuracy of the toothbrushing monitoring system.

Specifically, to enforce temporal continuity in the predicted result, we employ a modified 1D Symmetric Nearest Neighbor (SNN) filter [36]. This filter originally aimed to smooth out sudden variations in an image by replacing each pixel with the average of its nearest symmetric neighbors. Given the discrete time series of surface prediction, we introduce a 1-D modification to the SNN filter: if the values of the nearest neighbors of a data point on both sides are equal, we replace the data point with this common value; otherwise, we retain the original value. The advantage of this filter is that it maintains continuity within a surface while preserving the boundaries between different surfaces. Mathematically, we can formulate the filter as follows:

$$s'_i = \begin{cases} s_{i-\delta}, & \text{if } \exists \delta < k \text{ and } s_{i-\delta} = s_{i+\delta} \\ s_i, & \text{otherwise} \end{cases} \tag{6}$$

where s'_i is the filtered value of s_i at position i, k is the filter window size, and δ is the nearest neighbors of i. We apply the SNN filter with varying window sizes of 1, 3, and 5 sequentially to filter the classification results at different temporal scales.

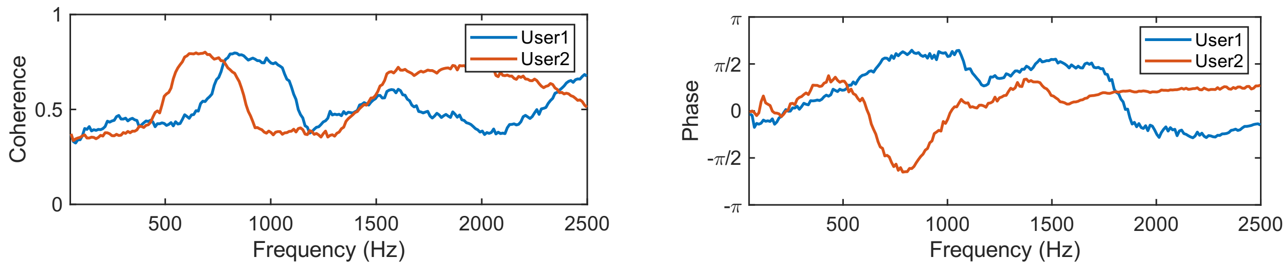
Figure 11(b) shows the filtered result after the temporal constraints compared to ground truth. We can see that most abrupt variances and outliers are filtered out. We note that this continuity-based postprocessing method is also applicable to other toothbrushing orders since it only utilizes the temporal continuity within a short time. By leveraging temporal constraints, we achieve significant improvements in the accuracy and consistency of the classification results, thereby enhancing the utility and reliability of our proposed approach for real-world applications.

5.5 New-user Registration

Since we primarily utilize propagation channels as features, which depend on the physiological structure (*e.g.*, head size, head bone, and facial muscles) of a person, these features vary among different users. As shown in Fig. 12, user 1 and user 2 have different cross-channel coherence and phase when brushing the same teeth surface S1. Consequently, we cannot directly apply a model trained on one user's data to a completely new user. To address this, we employ a transfer learning strategy to personalize the trained model for new users. As shown in Fig. 4, a new user can follow these steps to complete the registration process:

(1) Before brushing, the user downloads the global model and SmarTeeth instructional video.
(2) During brushing, the user follows the video's animated instructions (*i.e.*, surface and duration) to label their toothbrushing data.
(3) After brushing, the few-shot labeled data is used to fine-tune the pretrained model, personalizing it for the new user.
(4) For higher tracking accuracy, the user can repeat the registration process with additional sessions as needed.

Figure 15 shows the guide video. There is a blue line flashing to indicate the surface to brush and a countdown clock to remind the user of the time left for this surface. The animation video guides



(a) Cross-channel coherence **(b) Cross-channel phase**

Figure 12: Cross-channel features of different users.

users to brush all 16 dental surfaces, with an 8-second brushing time for each surface. The total time for one brushing session is about 2 minutes, which aligns with the recommendations of ADA [8]. We provide a preset brushing sequence, but users can also customize the order to suit their preferences. The new user is asked to brush their teeth following the video instructions strictly: this "script" serves as the labeled ground truth. During this registration, the guide video also serves as an opportunity for users to learn how to achieve comprehensive toothbrushing.

Theoretically, incorporating multiple registration sessions leads to higher performance, as it allows the model to capture a more comprehensive characterization of the user's physiological structure and brushing habits. In real-life scenarios, users have the flexibility to decide the number of sessions they want to contribute based on the desired performance. These sessions do not require users to brush consecutively in one day; instead, users can simply enable the app and follow the guide video during their regular brushing routines. The SmarTeeth system will automatically incorporate the data for fine-tuning. This approach delivers substantial improvements while imposing minimal additional effort, achieving a balance between accuracy and user convenience for effective new-user registration. In Sec. 6.4, we evaluated the system performance with different numbers of registration sessions.

5.6 Toothbrushing Feedback

Based on the brushing surface detection results, we design feedback strategies to improve users' brushing habits by providing timely alerts and visualization. This section elaborates on the three primary feedback mechanisms: over-brushing feedback and under-brushing reminder, brushing duration visualization, and brushing score.

5.6.1 Overbrushing Alert and Underbrushing Reminder. Overbrushing can lead to enamel erosion and gum recession. Hence, timely alerts help prevent these adverse effects. Our system utilizes in-ear microphones to monitor the brushing locations and identify if any specific area is being overbrushed. Typically, the recommended brushing time is 2 minutes (*e.g.*, 8 seconds for each surface or 20 seconds for each region [8]), but sometimes dentists advise patients to brush a specific area for a slightly longer period to clear stubborn plaque or stains caused by smoking. Therefore, the users can customize this parameter based on their dentist's advice. The earphone used by SmarTeeth provides a natural feedback interface to users. When overbrushing is detected or a region has brushed for the customised duration, a gentle alert through the earphone speaker prompts the user to move to another area. Conversely, brushing teeth for insufficient time will accumulate dental plaque and could lead to tooth decay and periodontal diseases. SmarTeeth also identifies areas that have been underbrushed. The system analyzes the tracking results to pinpoint regions that were brushed for insufficient duration (*e.g.*, 5 seconds for each surface [39]). The system reminder through the earphone speaker encourages them to rebrush these areas and ensure a mimimum brushing duration. This audio feedback mechanism helps maintain proper brushing practices, preventing underbrushing.

5.6.2 Brushing Score. Collaborating with a dentist, we introduce a brushing score to quantify the quality of each brushing session. The overall score will be displayed on the app and is composed of three sub-scores: time score, distribution score, and coverage score. The time score evaluates the overall duration of the brushing session. A total time within the recommended range results in a high score, encouraging users to brush for an adequate amount of time. Conversely, if the user's brushing time exceeds or does not reach the recommended time, the score decreases. This score is calculated as 1 minus the absolute difference between the actual brushing time and the recommended time, divided by the recommended time; The distribution score evaluates the distribution of brushing time across different areas (surfaces). It ensures that no particular area is neglected or overly focused on. This score takes the average brushing time score for all separate dental areas (surfaces); The coverage score indicates how completely the user has brushed different areas of the mouth. It ensures that all areas, including hard-to-reach spots, receive sufficient attention. The coverage score is defined as the ratio between the number of brushed areas (surfaces) and total areas (surfaces). Only if the brushing time exceeds half of the required duration will it count as an effective brushed area (surface).

The overall toothbrushing score integrates all three aforementioned scores into a weighted sum. We have set the weights to 0.35, 0.4, and 0.25 to emphasise the importance of brushing all surfaces for sufficient time [30]. Users can adjust these weights based on their dentist's recommendations. Fig. 1 showcases an overall brushing score and its sub-scores for one brushing cycle. This score helps users evaluate their brushing quality and identify areas for improvement. Over time, users can track their scores and observe improvements in their brushing habits, fostering better oral practice.

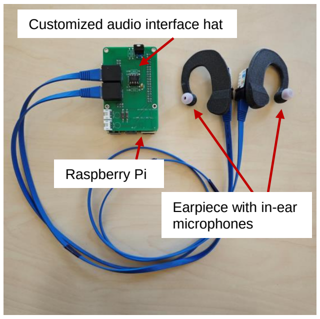


Figure 13: SmarTeeth hardware.

Figure 14: Experiment scenario.

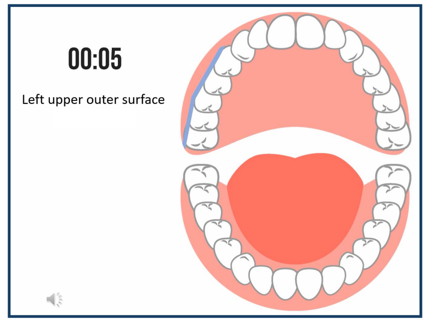


Figure 15: Guide video for toothbrushing.

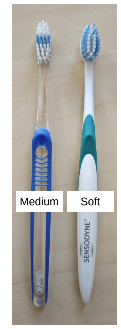


Figure 16: Toothbrushes.

5.6.3 Brushing Duration Visualization. Like many commercial smart toothbrushing apps [9], SmarTeeth generates a visualization to indicate the brushing score and the brushing duration for each area of the mouth to enhance user awareness and improve brushing habits. As shown in Fig. 1, this visualization is accessible via a companion app on the web or smartphones, providing a clear visual representation of which areas received adequate attention and which did not. The heatmap uses color coding to indicate different brushing durations – for instance, blue for underbrushed areas, green for adequately brushed, and red for overbrushed zones. This visual feedback of the brushing score and duration enables users to quickly assess their brushing patterns and adjust their brushing habits in subsequent sessions. By consistently using the toothbrushing visualization, users can develop a more balanced and effective brushing routine, ensuring that all areas of the mouth receive proper care. We will further analysis the brushing habit conveyed by this figure in Sec. 6.5.

6 Implementation and Evaluation

6.1 Implementation

Hardware. It is becoming increasingly common for ANC earphones to contain in-ear microphones, but no commercial earphones allow open access to the raw audio data [76]. Therefore, we designed and fabricated a prototype earphone for our evaluation. As shown in Fig. 13, we encapsulate a Knowles SPU1410LR5HQB microphone [7] within a 3D-printed ear-mounted housing case. We designed a custom PCB to connect the microphones in each ear to a HiFiBerry DAC+ADC pro audio HAT [4] installed on a Raspberry Pi 4B. The in-ear sound is sampled at a rate of 44100 Hz. The whole system is powered by a 5V power bank to ensure portability. Figure 14 shows the experiment scenario. During toothbrushing sessions, users wear custom ear-mounted earphones and fit them into the ear canal, so that the in-ear microphone can capture the inner-ear sound signals during toothbrushing.

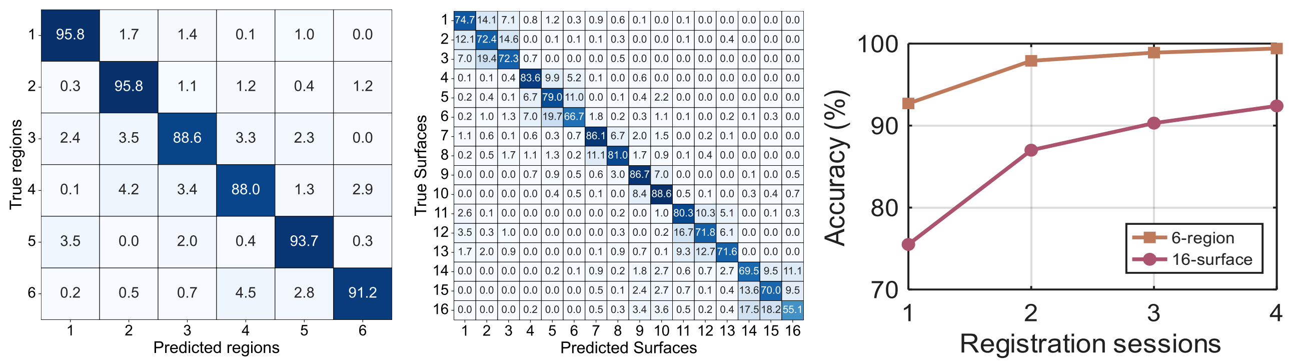
Software. We deployed SmarTeeth on a Raspberry Pi 4B, where the signal processing functions are implemented in Python, taking approximately 40.8 *ms* per 300 *ms* audio window. The deep learning model, implemented using TensorFlow, was trained on a PC with the Adam optimizer, a learning rate of 0.001, and a batch size of 200. The model converged quickly within 20 epochs, with a final model size of 59K. Inference takes 47 *ms* per window. The total processing time remains under 100 *ms*, ensuring that SmarTeeth delivers real-time feedback, providing users with immediate brushing surface detection and guidance without noticeable delays. This prototype serves as proof of concept. Modern smartphones, however, far surpass the computational power and hardware inference capabilities of the Raspberry Pi used here. If we can access in-ear microphone data from commercial earphones and deploy the system on a smartphone, the processing speed could be further improved.

6.2 Study Design

Ground Truth. Given that camera-based acquired ground truth is often compromised by hand and mouth obstruction [13, 25], many prior studies have required participants to brush their teeth following guide instructions to ensure accurate ground truth collection [31, 40, 48, 62]. In line with this, we also use video-guided brushing instructions to collect ground truth data, as described in Sec. 5.5. This instructional video is integrated into the companion app, allowing new users to easily follow the brushing steps for quick data registration and seamless use of our system.

Data Collection. We recruited 13 volunteers (7 female and 6 male) to participate in our experiment, which was approved by the Ethics Committee of our institution. Before data collection, participants underwent a brief training session to familiarize themselves with the video instructions. They wore the earphone prototype and brushed their teeth five times on five separate workdays (one session per day). Each session involved two rounds of brushing—once with a soft-bristle toothbrush and one with a medium-bristle toothbrush (Fig. 16). During data collection, we also instructed participants to vary their brushing order, introduced different levels of ambient noise in the room, and conducted a case study using low-end electric toothbrushes. This dataset is collected in the laboratory setting to design and benchmark the model (Sec. 6.3~6.5)

Open Settings. To evaluate SmarTeeth's feedback and performance in open settings, after system development, we re-invited all participants to use the SmarTeeth prototype while brushing freely according to their own habits for two sessions. In this evaluation, users did not follow video instructions, so another method of obtaining ground truth was necessary. We observed that line-of-sight obstructions from the user's hands and mouth make visually distinguishing all 16 individual surfaces challenging [13, 25], whereas six-region identification remains reliable. Consequently, an observer



(a) 6-region tracking

(b) 16-surface tracking

(c) Registration with more sessions

Figure 17: Overall performance of SmarTeeth. The values in cells are percentages (%).

manually recorded ground truth data for six regions and made every effort to annotate the specific 16 fine-grained surfaces. Compared to regions, we note that these surface annotations serve only as a "silver" standard rather than a definitive ground truth, providing an approximate reference for comparing SmarTeeth's 16-surface recognition results under realistic conditions (Sec. 6.5).

Evaluation Procedure. Many electric toothbrushes on the market offer tracking capabilities which are limited to six regions of the mouth, such as Oral-B iO Series 5, 6, 7 & 8 [6]. Only a few high-end models like the Oral-B io10 (550 USD) [15] can detect all 16 surfaces [58]. Therefore, we first evaluate SmarTeeth's tracking performance across the six regions and then assess its performance in tracking all 16 surfaces. Since SmarTeeth aims to provide comprehensive tracking of all 16 surfaces for manual toothbrush users, we subsequently analyze the impact of various factors based on its 16-surface performance in the following sessions. We also invited users to brush their teeth with and without activating the audio feedback function to visualize and analyze their brushing habits. After the experiment, all users were invited to rate the usability and satisfaction of our system's monitoring performance and feedback mechanisms. We report the findings of this user study in Sec. 6.6.

6.3 Overall Performance

In this evaluation, we trained the model on data from 12 users and then fine-tuned it using different numbers of random sessions from a new user. The model was then tested on the remaining sessions from that user. This protocol assesses the model's adaptability to new users with minimal registration data, reflecting a realistic deployment scenario where the system needs to adapt to new users after brief calibration. We first illustrate the performance with one-session registration and then gradually increase the fine-tune sessions to find a balance between the registration overhead and the performance.

Six-region Detection. Figure 17(a) shows the confusion matrices for the 6-region classification. In the cross-user case, SmarTeeth achieves an overall accuracy of 92.7%, demonstrating its effectiveness in detecting 6 distinct dental regions. We observe that performance is slightly lower for the middle regions (R3, middle upper, and R4, middle lower) compared to the left and right regions (*e.g.*, R2 and R5). This is likely because, when brushing the middle region, the brush head may inadvertently contact parts of the adjacent left and right regions due to limited space in the mouth and cause ambiguity. In addition, we observed zero values along the matrix edges, suggesting that our postprocessing techniques help to reduce occasional misclassifications and outliers.

Sixteen-surface Detection. Figure 17(b) shows the confusion matrix for the 16-surface classification task. The observations from this evaluation closely align with those from the six-region detection experiments. Specifically, the system achieved an accuracy of 75.6% under the cross-user evaluation. We can observe that ambiguity occurs between neighboring surfaces and regions, where adjacent brushing locations with similar propagation channels can lead to misclassification. For example, the precision for S2 (upper left inner) is 72.4%, with 12.1% and 14.6% of samples misclassified to S1 (upper left outer) and S3 (upper left chewing), respectively. Compared to the six-region evaluation, SmarTeeth performs less effectively on 16-surface detection. This difference is expected due to the greater complexity of this task and the heightened ambiguity among adjacent surfaces within the same region, which makes accurate tracking more challenging.

Registration with More Sessions. In the previous evaluation, we used only a single registration session (2 mins) to fine-tune the pretrained model for new users to evaluate minimum calibration effectiveness. Figure 17(c) further shows the accuracy with one to four registration sessions. We observe a steady improvement in performance as the number of registration sessions increases. After applying postprocessing, the accuracy of six-region tracking increases from 92.7% using one session to 99.4% using four sessions, highlighting the benefit of additional registration data. For 16-surface tracking, the accuracy increases from 75.6% with one session to 92.4% with four sessions. With three registration sessions (~ 6 *mins*), the performance can be boosted to 98.8% and 90.3% for six-region and sixteen-surface tracking, respectively.

Remarks. Our results demonstrate that just three sessions (a total of six minutes) are sufficient to achieve superior performance. These sessions do not require users to brush consecutively in one day; instead, users can simply enable the app and follow the guide video during their regular brushing routines. The SmarTeeth system will automatically incorporate the data for fine-tuning. This

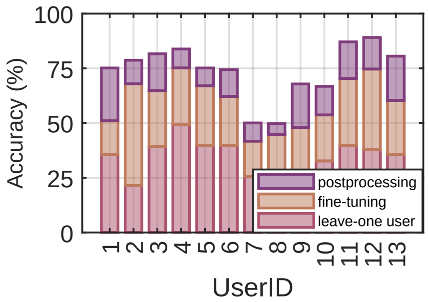


Figure 18: Impact of users.

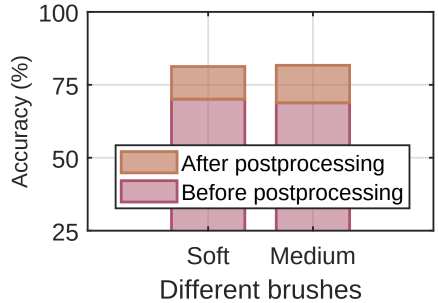


Figure 19: Impact of different brushes.

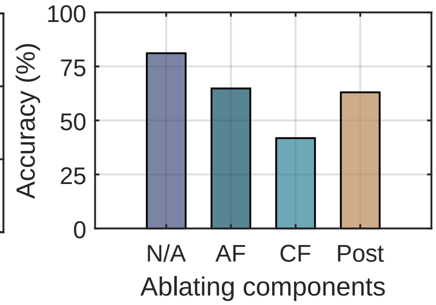


Figure 20: Ablation study.

approach delivers substantial improvements while imposing minimal additional effort, making three sessions an optimal balance between accuracy and user convenience for effective new-user registration.

6.4 Impact of Practical Factors

Impact of Users. Figure 18 shows the performance across different users. When we directly applied for the leave-one-user-out analysis, the average accuracy is only 34.2%. This result was expected, given the channel-related features used in our system, which are inherently related to individual physiological characteristics. Since each user has a unique oral anatomy and head structure, directly testing the model on new users leads to significant bias. After model fine-tuning with one registration session, we observed a significant improvement in accuracy, with performance increasing to 60.1%. This fine-tuning method helps calibrate the feature distribution of the new user. Then, by employing postprocessing techniques (*i.e.*, temporal constraints), we were able to further refine the model's predictions and enhance its accuracy. As a result, the accuracy was boosted to 73.9%. This evaluation demonstrates the efficacy of the fine-tuning approach in enhancing the model's adaptability to different users. Note that this result is averaged over 13 users rather than across all samples. It is somewhat inconsistent with the cross-user performance in Sec. 6.3, as the number of samples is different between users.

Interestingly, we observed that the performance of users 7 and 8 was notably lower compared to other users. This is because these users had relatively smaller ear pinnae, making it challenging to securely fit the in-ear microphone in the ear canal. In such cases, the in-ear microphone is susceptible to interference, and the assumption of omitting noise term in Eq. 1 cannot hold, leading to inaccuracies in the extracted channel-related features. Upon excluding users 7 and 8 from the analysis, we observed a substantial improvement in performance, with average accuracy reaching 80.1%. In the future, we plan to print custom earphone casings of varying sizes. By offering users earphone casings tailored to their ear shape and size, the secure fit could mitigate interference and enhance the accuracy of these users. Moreover, current earphones are lightweight and come with ear tips of different sizes. Some of them can even detect the seal state of the earphone [3], which can effectively address this challenge.

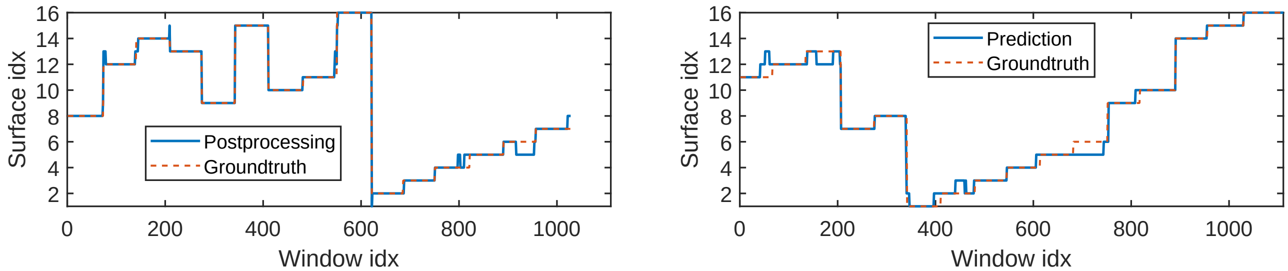
Impact of Brushes. To evaluate the impact of different toothbrushes, we divided all the data into two groups according to the toothbrush type (*i.e.*, soft and medium). The model was then trained on one toothbrush (*e.g.*, soft) and tested on another (*e.g.*, medium), and vice versa. As shown in Fig. 19, the accuracies for soft and medium brushes are 70.1% and 68.8%, respectively. Yet, after applying postprocessing techniques, the performance is improved to 81.3% and 81.7%, respectively, which shows the robustness of our system to different brush bristles due to cross-channel features.

Impact of Brushing Orders. Different individuals may have their own brushing orders and habits. Despite the diversity across users, the methodology of SmarTeeth remains applicable. To validate it, we collected the data using the instruction videos with different brushing orders as shown in Fig. 21. We directly use the model trained with the data following the order in Fig. 11(b) to evaluate these two new brushing orders. Figure 21 illustrates the comparison between ground truth and predicted outcomes of two brushing cycles. The respective performances are 82.8% and 79.4%, showing no significant deviation from our previous results. This is because the SNN-based smoothing method only constrains the temporal continuity within a short time frame while preserving the boundaries between different surfaces. As a result, the detection accuracy remains consistent regardless of the sequence in which the surfaces are brushed. We also further validate this through an in-the-wild case study in Sec. 6.5.

Impact of Environmental Noise. To evaluate the impact of environmental noise, we conducted experiments where three users with good occlusion states brushed their teeth while a nearby loudspeaker played music at a typical daily noise level ranging from 50 to 60 dB. We use the model trained with data in quiet scenarios to test these brushing sessions. The accuracy is 73.4%, which improved to 80.8% after postprocessing. This is because of the sealing effect of the eartips, which effectively isolates external sounds. Consequently, any residual noise within the ear canal became negligible. Additionally, the toothbrushing sound propagates through bone conduction, which inherently offers higher fidelity compared to subtle external interference transmitted through the air and the attenuation caused by the eartips.

6.5 Method Evaluation and Extension

Ablation Study. To understand the individual contributions of different components, we conducted an ablation study. We replicated the 16-surface evaluation in Sec. 6.3 for the users with good seal states while excluding various modules: modified audio features (AF), cross-channel features (CF), and postprocessing (Post). The performance of the system without ablation (N/A) was 80.1%.



(a) Order 1 (b) Order 2

Figure 21: Impact of different brushing orders.

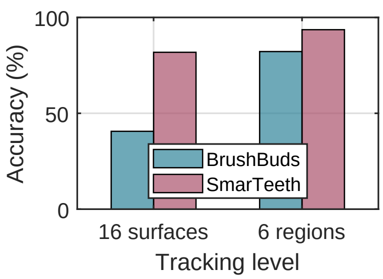


Figure 22: Baseline comparison.

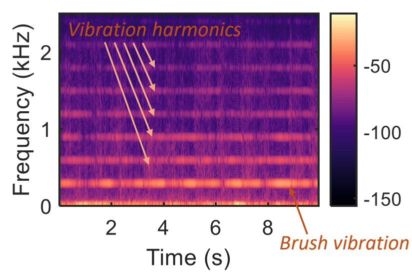


Figure 23: Electric toothbrushing sounds.

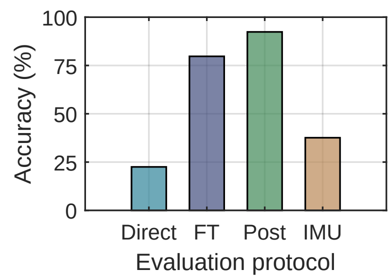


Figure 24: Testing on electric brushes.

Removing modified audio features resulted in the performance decreasing to 64.8%. This reduction can be attributed to the sound timbre information captured by these features, which aids in distinguishing between adjacent surfaces in the same region. Without them, the system encounters more ambiguity in surface classification. When removing cross-channel features, the performance significantly drops to 41.8%. This decline highlights the sensitivity of sole audio features to variation between sessions, which can introduce bias as we discussed in Sec. 5.2.2. Excluding postprocessing, the system has a substantial performance decrease to 63%. This significant drop underscores the importance of postprocessing in mitigating ambiguity and handling the outlier effect effectively.

Baseline Comparison. BrushBuds [78] leverages IMU sensors on earphones to track toothbrushing locations. To establish a baseline for performance comparison, we attached two IMU sensors to our earphone prototypes and implemented BrushBuds. We collected data from five users, with five sessions per user, to ensure a fair comparison. Figure 22 compares the performance of BrushBuds and SmarTeeth at both 16-surface and 6-region tracking levels. BrushBuds achieves an accuracy of 40.6% for 16-surface classification and 82.2% for six-region classification, which is 41.2% and 11.4% lower than SmarTeeth's respective accuracies. The ineffectiveness of the earable IMU-based approach in tracking 16 surfaces can be attributed to several factors. First, the differences in IMU signals caused by brushing different regions are minimal because IMUs on earphones primarily capture coarse head motions incurred by brushing rather than fine-grained vibrations, and they become even smaller when distinguishing between individual surfaces. Second, we observed that users often unintentionally move their heads while brushing, introducing noise that further degrades IMU signal quality. In contrast, SmarTeeth utilizes in-ear toothbrushing sounds, which can effectively capture the distinct friction sounds through bone conduction from different surfaces and is inherently more robust to head movements, enabling superior tracking performance.

Extending to Electric Toothbrushes. Considering that many low-end electric toothbrushes also do not have the brushing monitoring function, we also conducted a case study to evaluate whether SmarTeeth could also augment low-end electric toothbrushes. We note that our experiments with electric toothbrushes are primarily to validate the feasibility of the method; however, the main focus of SmarTeeth remains on manual toothbrushes. Participants were asked to use the Mijia T300 electric toothbrush (20 USD) [10] for three sessions. As shown in Fig. 23, the brushing sound using electric toothbrushes exhibits distinct patterns compared to that of manual toothbrushes. We can observe the fundamental frequency and multiple harmonics of toothbrush vibrations in the frequency domain. As electric toothbrushes do not require manual brushing strokes and users only need to position the brush head at different teeth, no stroke-related signals were observed.

As shown in Fig. 24, we first tested the model trained on manual toothbrush audio directly on the electric toothbrush data (Direct), and the accuracy significantly reduces to 22.5%. This result is expected since we use the propagation-channel features as well as the brushing sound features (*i.e.*, MFCC/GTCC) to achieve fine-grained toothbrushing monitoring. Even though these features work well across different manual brushes, the powerful excitation vibration of electric toothbrushes, several orders of magnitude higher than manual toothbrushing sounds, has significantly disturbed the MFCC/GTCC features, leading to a performance drop without any finetuning. Subsequently, we conducted model fine-tuning with one electric toothbrushing registration session, and the performance increased to 79.7% (FT). After postprocessing (Post), the accuracy improved to 92.4%. These results indicate that in addition to manual toothbrushes, the methodology of SmarTeeth is also applicable to

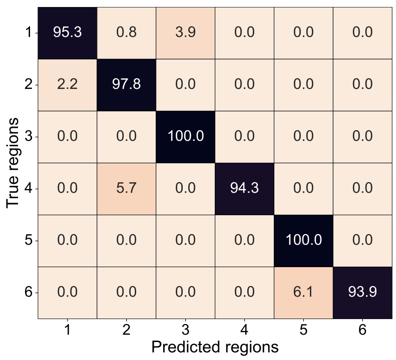


Figure 25: Confusion matrix of six-region performance in-the-wild.

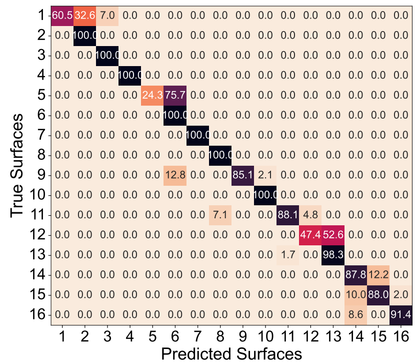


Figure 26: Confusion matrix of 16-surface performance in-the-wild.

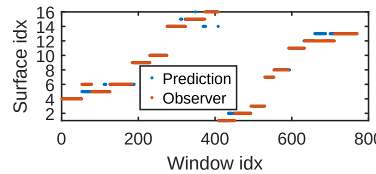


Figure 27: A brushing cycle in the wild.

Figure 28: Plaque test. The tablet stained residual plaque red.

electric toothbrushes, which yields even better performance compared to manual toothbrushes. We speculate that the performance improvement may be attributed to the stronger excitation vibration of electric toothbrushes, resulting in more stable and accurate channel-related features.

We also compared our system's performance with BrushBuds [78] on electric toothbrushes. BrushBuds achieved an accuracy of 37.6% (IMU in Fig. 24), which is significantly lower than our system. This inferior performance can be attributed to two factors: first, unlike manual toothbrushes, electric toothbrushes do not require back-and-forth brushing. Users generally simply hold the brush against the teeth. Thus, the differences in IMU signals caused by brushing different surfaces are minimal, making fine-grained surface tracking challenging. Second, IMU signals are highly susceptible to interference from the strong vibrations generated by electric toothbrushes, further degrading their reliability [25].

Another work specifically designed for electric toothbrushes, ToothFairy [71], reports an accuracy of 92.4% for identifying dental quadrants (four regions). For a direct comparison, we evaluated SmarTeeth by dividing the teeth into the same four regions as well. The accuracy of SmarTeeth is 98.8%, which outperforms ToothFairy [71] by 6.4%. We attribute this improvement to fundamental differences in methodology. ToothFairy relies on fitting parameters based on the vibration energy generated by electric toothbrushes, which can be influenced by various factors such as brushing style (*e.g.*, the force applied, brushing speed, and motion patterns) and toothbrush condition (*e.g.*, wear and tear, and battery life). In contrast, SmarTeeth leverages propagation-channel features, where the distinctions between left, right, upper, and lower channels are inherently more pronounced, enabling more robust region classification.

Toothbrushing In-the-wild. To evaluate the performance of SmarTeeth in more realistic settings, we conducted an in-the-wild case study as described in Sec. 6.2. Figure 25 shows the confusion matrix for the six-region detection performance in the wild. The overall accuracy is 97.1%, which is slightly higher than our previous results. The detection accuracy for some regions like R3 and R5 even reaches 100%. This is because the participants become more accustomed to using our system and achieve better ear canal occlusion. We also compared the 16-surface detection results in the wild with observer-annotated reference. Fig. 27 shows the surface tracking result of a participant. Figure 26 presents the confusion matrix for 16-surface tracking in this setting, with an overall accuracy of 81.6%. Overall, SmarTeeth's detection results generally align well with the observer's annotations, which serve as a silver reference rather than a definitive ground truth. We can observe some discrepancies between the SmarTeeth predictions and the annotations. For example, samples identified as S6 by SmarTeeth were sometimes annotated as S5, and similar mismatches occurred between S1 and S2, as well as S12 and S13. These cases are likely caused by the observer's line-of-sight obstructions while recording the data. Despite with silver reference, the results demonstrate that SmarTeeth performs effectively in a real-world and open setting.

Plaque Test. The plaque test is a common method used to help identify areas of teeth that have been missed after toothbrushing. To intuitively visualize brushing effectiveness, we conducted a plaque test where users chewed a disclosing tablet after brushing freely. The tablet stained residual plaque red, providing a clear visual of areas that were insufficiently cleaned. Fig. 28 and Fig. 29 show both a (mirrored) photo taken after the plaque test and the toothbrushing report generated by SmarTeeth. According to the SmarTeeth tracking results, the user did not spend sufficient time brushing the left side, particularly on the outer surfaces. This observation is supported by the plaque test results, which reveal residual plaque in the left outer areas (highlighted with boxes in Fig. 28). This indicates that insufficient brushing time in certain areas leads to plaque buildup. Interestingly, we found that this user is left-handed and tends to focus more on the right side, which feels more natural to brush. Conversely, the left side, particularly the outer and inner surfaces, is less accessible and often brushed only briefly. The user admitted to not being aware of this habit before and noted that

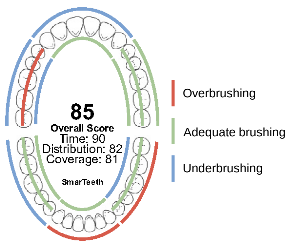
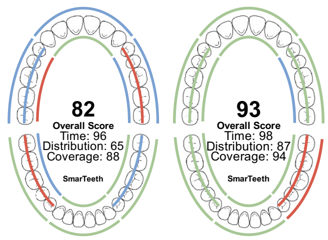


Figure 29: Toothbrushing report.



(a) Disable feedback **(b) Enable feedback**

Figure 30: Audio feedback evaluation.

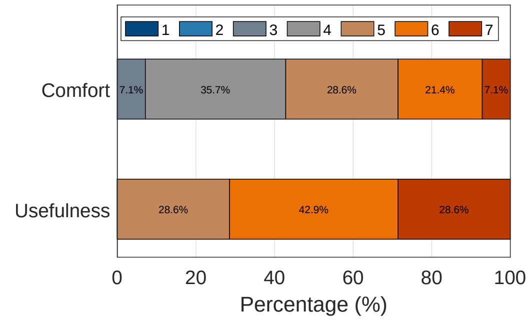


Figure 31: Usefulness and comfort evaluation.

the feedback provided by SmarTeeth was highly useful in helping improve his brushing habits.

6.6 Evaluation of User Experience

After system development, we re-invited all participants to use the SmarTeeth prototype during toothbrushing and rate the usability and satisfaction of our system's monitoring performance and feedback mechanisms. The results and users' comments provide valuable insights for further development and optimization of the system.

Audio feedback. Figure 30(a) shows the brushing visualization in our app of a user for a brushing cycle with audio feedback disabled. The overall brushing score is 82. Specifically, the total brushing duration was about 1.92 minutes, which is very close to the recommended 2 minutes, resulting in a high time score of 96. However, the distribution score is relatively low at 65, indicating that the user did not brush each tooth surface evenly and properly. Specifically, we can observe that this user brushed the lower regions more thoroughly than the upper regions. Additionally, the outer surfaces of the lower teeth were sufficiently brushed, while certain inner surfaces (*e.g.*, S5) and chewing surfaces (*e.g.*, S16) were underbrushed likely due to their harder-to-reach locations. In addition, the user overbrushed the left lower chewing surface (*i.e.*, S6). In contrast, most of the upper surfaces were not brushed for a sufficient duration, possibly because brushing the upper teeth is more inconvenient than brushing the lower ones. Similar to the lower teeth, the user also overbrushed two upper chewing surfaces. Out of the 16 surfaces, 14 were brushed for more than half of the required time, leading to a coverage score of 88. This brushing time visualization allows the user to intuitively understand their brushing habits and make necessary adjustments, such as shortening the brushing time for some surfaces or dedicating additional time to others.

We then invited this user to activate the audio feedback function and brush their teeth again. Figure 30(b) shows the toothbrushing visualization with the feedback enabled. The overall score significantly improved to 93, indicating that the feedback helped the user adopt better brushing habits. From the figure, we can observe that the user performed notably better on the upper arch compared to the previous session. As a result, the distribution score increased from 65 to 87, and the coverage score improved from 88 to 94. However, we noticed that the user still underbrushed the right-upper-inner surface. Interestingly, when the system reminds the user to brush the right-lower-chewing surface, they overbrushed this area. The user said it is because he was concerned about not brushing enough, leading him to continue brushing this area longer a bit. This suggests that the user may need to continue using the system over time in order to become more accustomed and confident in following the audio feedback accurately.

Usefulness and Comfort. Participants rated the usefulness and comfort of the SmarTeeth system on a 7-point Likert scale [65], where 1 indicates "very low" and 7 indicates "very high". As shown in Fig. 31, the average usefulness score was a strong 6.0/7.0, indicating that users think the system is highly beneficial for improving their toothbrushing habits due to the timely reminder and brushing score evaluation. However, the comfort score of hardware was lower, averaging 4.86/7.0. The lower comfort rating is primarily due to the current prototype design, which includes wired over-ear earphones that are heavier and more cumbersome than commercial wireless earbuds. In this paper, we focus on the technical feasibility validation of using earphones to enhance manual toothbrushing. In the future, iterations of our system will focus on miniaturizing the hardware or collaborating with commercial earphone companies to integrate our techniques into their more ergonomic and lightweight earbuds. This strategy is feasible due to the widespread availability of in-ear microphones in commercial ANC earbuds.

Workload Evaluation. We combined a 7-point Likert scale and The NASA Task Load Index (NASA-TLX) [35] to evaluate the workload of users when using our system. NASA-TLX is a widely used tool to assess perceived workload across six dimensions: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Each dimension is rated on a scale from 1 to 7, with higher scores indicating greater perceived demand or effort. Note that, contrary to other items, for performance, 1 represents a high level, and 7 represents a low level.

As shown in Fig. 32, participants rated the overall workload score as 2.46/7.0, indicating a low level of usage workload. Specifically, participants rated the mental demand of using SmarTeeth at a low level (2.42), suggesting that the system is reasonably straightforward to understand and use. However, the physical demand received a higher score (3.99) compared to the mental demand. This is due to the prototype design, which requires users to adjust the earphones to achieve a proper occlusion state. Because of the additional weight and wired constraints, some users said, *"I feel weird when I brush my teeth with earbuds, because they may drop."* For the same reason, the effort score is 3.13 slightly higher than ideal, mainly due to the discomfort of wearing the prototype earphones. We believe that

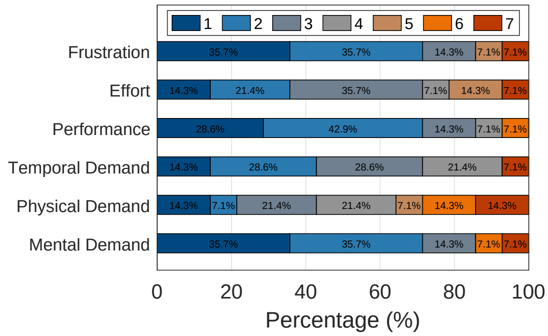


Figure 32: Workload evaluation of SmarTeeth.

optimizing the hardware size and weight can significantly alleviate this issue. Additionally, we will integrate existing fit detection technologies [3, 27] to ensure that users achieve proper sealing and placement and avoid constant adjustments. Users feel the time pace of using SmarTeeth was manageable (2.92), indicating that the system does not significantly impact or slow down their toothbrushing routine. Participants also rated their performance very high (2.3) while using SmarTeeth, indicating they felt successful in following the system's recommendations and effectively improving their brushing habits. Additionally, the frustration level is rated lower (2.35), indicating that while there were some discomforts and adjustments needed, the overall experience was not highly frustrating for most users due to the short brushing time per day.

User Suggestions. At the end of the questionnaire, we have a blank to let participants provide feedback on the SmarTeeth system. Common themes included concerns about the hardware size and earphones potentially falling out during use. One user said, "*A way to know when proper sealing of earbuds and ear canal has happened* (is needed)." Another user found our system particularly useful: *"I took three big tooth surgeries in the past ten years. And I can't use the electric toothbrush. The toothbrush I use has certain requirements, and the brushing time is controlled according to the state of the teeth every day."*

Feedback from the Dentist. After evaluation, we conducted an interview with the dentist again. She indicated that our survey findings are consistent with her clinical experience, noting that many people have dental disease due to the lack of enough coverage and timing of toothbrushing. She was impressed with the system's performance, particularly its ability to monitor fine-grained brushing surfaces and ensure sufficient brushing time, which is a critical issue for many kids and manual toothbrush users.

7 Limitation and Future Work

Hardware Improvement. While effective in detecting brushing surfaces, the current SmarTeeth prototype needs to improve the comfort level and sealing check. The wired overear design was noted by users as inconvenient for daily use. Moving forward, we will optimize the form factor by miniaturizing the hardware. Collaboration with established earphone manufacturers could allow us to leverage existing advancements in wireless, lightweight, and comfortable earbuds, which would ensure good sealing and significantly enhance user experience [3].

Toothbrush Degradation. In our current study, participants used new toothbrushes over a period of one week, but toothbrush bristles naturally degrade, becoming softer and more worn over time. Although SmarTeeth relies primarily on channel-related features, some audio features, such as MMFCC, may be affected by the bristle condition. Dentists generally recommend replacing toothbrushes every three months [11]. Yet, bristle wear may impact performance within this period. Therefore, we suggest periodic recalibration of the model after extended use, such as one month, using video-guided brushing instructions with minimal user effort. Furthermore, by employing continuous learning techniques [70], the model could be iteratively updated to adapt to gradual toothbrush degradation.

User Applicability. In the current stage, we have only evaluated SmarTeeth with participants who have normal oral health conditions. However, our system must be carefully considered across different user demographics. For instance, older adults and children, who may have missing teeth or dentures, could benefit from a more flexible scheme for determining proper brushing duration. Additionally, for users with orthodontic braces, the presence of metal wires and brackets could alter the propagation of sound within the oral cavity. Therefore, these users may need to update their model after prosthetic dental work to ensure accurate brushing surface detection.

Potential Applications. Beyond toothbrushing tracking, the in-ear brushing sound used in SmarTeeth has the potential to be applied to other areas of oral health. For example, during the post-evaluation interview, the dentist suggested that incorporating brushing pressure detection would make the system even smarter, with which we can prevent gum recession. We consider this for future work by exploring the correlation between pressure and sound intensity, aiming to bring even more intelligent and powerful functionality to manual toothbrush/low-end electric toothbrush users. In addition, root canals create cavities within the tooth that need to be diagnosed by X-ray imaging which is not always accessible. However, these cavities within the tooth can alter the way sound propagates through the dental and surrounding bone structures. By analyzing differences in sound patterns, it might be possible to identify root canals and predict brushing pressure. Such capabilities could make SmarTeeth a comprehensive tool for daily oral health monitoring, and we leave them for future work.

8 Conclusion

We propose SmarTeeth, a pioneering fine-grained toothbrushing monitoring system utilizing in-ear microphones of earbuds, revolutionizing manual toothbrushes by providing users with detailed brushing tracking information. By integrating cross-channel features and modified audio features, SmarTeeth achieves high accuracy in tracking dental regions and surfaces after refinement through temporal constraints. The user feedback is also promising: an majority of our participants have shown enthusiasm toward adopting SmarTeeth, anticipating its seamless integration into their daily toothbrushing routine. This system empowers the vast community of users who use manual toothbrushes and low-end electric toothbrushes to attain the benefits of high-end smart toothbrushes using just a pair of earphones, which holds significant promise for promoting public dental health. Since daily toothbrushing is conducted naturally near the ear canal, the audio captured in the ear could be extended to detecting more complex dental conditions such as root canal infections and tooth cavities, further enhancing oral healthcare practices with earables.

Acknowledgments

This work is supported by ERC through Project 833296 (EAR) and EPSRC grant EP/Y035925/1.

References

[1] 2024. 3 Tooth Brushing Techniques | Colgate®. https://www.colgate.com/en-us/oral-health/brushing-and-flossing/3-tooth-brushing-techniques-that-get-the-job-done. (Accessed on 05/01/2024).

[2] 2024. ANC Headphones Market Size | Mordor Intelligence — mordorintelligence.com. https://www.mordorintelligence.com/industry-reports/active-noise-cancellation-headphones-market/market-size.

[3] 2024. earbud fit | MindMics. https://www.mindmics.com/earbud-fit. (Accessed on 07/30/2024).

[4] 2024. HiFiBerry DAC+ ADC Pro | HiFiBerry. https://www.hifiberry.com/shop/boards/hifiberry-dac-adc-pro/. (Accessed on 05/01/2024).

[5] 2024. iO Series 10 Rechargeable Electric Toothbrush | Oral-B. https://oralb.com/en-us/products/electric-toothbrushes/io-series-10-rechargeable-electric-toothbrush-cosmic-black/. (Accessed on 05/01/2024).

[6] 2024. Oral-B iO8 Electric Toothbrush, Violet Ametrine | Oral-B. https://oralb.com/en-us/products/electric-toothbrushes/io-series-8-electric-toothbrush-violet-ametrine/. (Accessed on 04/28/2024).

[7] 2024. SPU1410LR5H-QB Knowles | Audio Products | DigiKey. https://www.digikey.co.uk/en/products/detail/knowles/SPU1410LR5H-QB/3621629. (Accessed on 05/01/2024).

[8] 2024. Toothbrushes | American Dental Association. https://www.ada.org/resources/ada-library/oral-health-topics/toothbrushes. (Accessed on 04/24/2024).

[9] 2024. Use Oral-B App to Improve your Dental Routine | Oral-B UK. https://www.oralb.co.uk/en-gb/product-collections/oral-b-app. (Accessed on 07/30/2024).

[10] 2024. XIAOMI MIJIA T300 Electric Toothbrush IPX7 Waterproof Smart Sonic Brush Ultrasonic Whitening Teeth Tooth Brush For Toothbrushes. https://www.aliexpress.com/item/1005005497209597.html. (Accessed on 05/01/2024).

[11] ADA. 2024. Toothbrushes | American Dental Association. https://www.ada.org/resources/ada-library/oral-health-topics/toothbrushes. (Accessed on 11/12/2024).

[12] Sumio Akifusa, Ayaka Isobe, Kanako Kibata, Akinori Oyama, Hiroko Oyama, Wataru Ariyoshi, and Tatsuji Nishihara. 2020. Comparison of dental plaque reduction after use of electric toothbrushes with and without QLF-D-applied plaque visualization: a 1-week randomized controlled trial. *BMC Oral Health* 20 (2020), 1–6.

[13] Sayma Akther, Nazir Saleheen, Mithun Saha, Vivek Shetty, and Santosh Kumar. 2021. mteeth: Identifying brushing teeth surfaces using wrist-worn inertial sensors. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 5, 2 (2021), 1–25.

[14] Sayma Akther, Nazir Saleheen, Shahin Alan Samiei, Vivek Shetty, Emre Ertin, and Santosh Kumar. 2019. mORAL: An mHealth model for inferring oral hygiene behaviors in-the-wild using wrist-worn inertial sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–25.

[15] Amazon. 2023. Amazon.com: Oral-B iO Series 9 Rechargeable Electric Toothbrush, Aquamarine with 4 Brush Heads and Travel Case - Visible Pressure Sensor to Protect Gums – 7 Cleaning Modes - 2 Minute Timer : Health & Household. https://www.amazon.com/dp/B088CK6XJ4?th=1. (Accessed on 04/03/2024).

[16] Takashi Amesaka, Hiroki Watanabe, and Masanori Sugimoto. 2019. Facial expression recognition using ear canal transfer function. In *Proceedings of the 2019 ACM International Symposium on Wearable Computers*. 1–9.

[17] Oliver Amft, Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. 2005. Analysis of chewing sounds for dietary monitoring. In *UbiComp 2005: Ubiquitous Computing: 7th International Conference, UbiComp 2005, Tokyo, Japan, September 11-14, 2005. Proceedings 7*. Springer, 56–72.

[18] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, and Cecilia Mascolo. 2023. heart: Motion-resilient heart rate monitoring with in-ear microphones. In *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 200–209.

[19] Jiani Cao, Chengdong Lin, Yang Liu, and Zhenjiang Li. 2022. Gaze tracking on any surface with your phone. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. 320–333.

[20] Yetong Cao, Chao Cai, Fan Li, Zhe Chen, and Jun Luo. 2024. Enabling Passive User Authentication via Heart Sounds on In-Ear Microphones. *IEEE Transactions on Dependable and Secure Computing* (2024).

[21] Yetong Cao, Chao Cai, Anbo Yu, Fan Li, and Jun Luo. 2023. Earace: Empowering versatile acoustic sensing via earable active noise cancellation platform. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (2023), 1–23.

[22] Ana Caraban, Maria José Ferreira, Rúben Gouveia, and Evangelos Karapanos. 2015. Social toothbrush: fostering family nudging around tooth brushing habits. In *Adjunct proceedings of the 2015 acm international joint conference on pervasive and ubiquitous computing and proceedings of the 2015 acm international symposium on wearable computers*. 649–653.

[23] Yu-Chen Chang, Jin-Ling Lo, Chao-Ju Huang, Nan-Yi Hsu, Hao-Hua Chu, Hsin-Yen Wang, Pei-Yu Chi, and Ya-Lin Hsieh. 2008. Playful toothbrush: ubicomp technology for teaching tooth brushing to kindergarten children. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 363–372.

[24] Ching-Han Chen, Chien-Chun Wang, and Yan-Zhen Chen. 2021. Intelligent brushing monitoring using a smart toothbrush with recurrent probabilistic neural network. *Sensors* 21, 4 (2021), 1238.

[25] Kaixin Chen, Lei Wang, Yongzhi Huang, Kaishun Wu, and Lu Wang. 2023. LiT: Fine-grained Toothbrushing Monitoring with Commercial LED Toothbrush. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–16.

[26] Kaiyan Cui, Qiang Yang, Yuanqing Zheng, and Jinsong Han. 2023. mmRipple: Communicating with mmWave radars through smartphone vibration. In *Proceedings of the 22nd International Conference on Information Processing in Sensor Networks*. 149–162.

[27] Berken Utku Demirel, Ting Dang, Khaldoon Al-Naimi, Fahim Kawsar, and Alessandro Montanari. 2024. Unobtrusive air leakage estimation for earables with in-ear microphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 4 (2024), 1–29.

[28] Geertruida W Dijkstra, Andor WJM Glaudemans, Paola A Erba, Marjan Wouthuyzen-Bakker, Bhanu Sinha, David Vállez García, Luc WM van der Sluis, and Riemer HJA Slart. 2020. Relationship between 18F-FDG uptake in the oral cavity, recent dental treatments, and oral inflammation or infection: a retrospective study of patients with suspected endocarditis. *Diagnostics* 10, 9 (2020), 625.

[29] Di Duan, Zehua Sun, Tao Ni, Shuaicheng Li, Xiaohua Jia, Weitao Xu, and Tianxing Li. 2024. F2Key: Dynamically Converting Your Face into a Private Key Based on COTS Headphones for Reliable Voice Interaction. In *Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services*. 127–140.

[30] Stefanie Ebel, Helen Blättermann, Ulrike Weik, Jutta Margraf-Stiksrud, and Renate Deinzer. 2019. High plaque levels after thorough toothbrushing: what impedes efficacy? *JDR Clinical & Translational Research* 4, 2 (2019), 135–142.

[31] Mahmoud Essalat, Oscar Hernan Madrid Padilla, Vivek Shetty, and Gregory Pottie. 2023. Monitoring brushing behaviors using toothbrush embedded motion-sensors. *Authorea Preprints* (2023).

[32] Andrea Ferlini, Dong Ma, Robert Harle, and Cecilia Mascolo. 2021. EarGate: gait-based user identification with in-ear microphones. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 337–349.

[33] Ming Gao, Lingfeng Zhang, Leming Shen, Xiang Zou, Jinsong Han, Feng Lin, and Kui Ren. 2023. Exploring practical acoustic transduction attacks on inertial sensors in MDOF systems. *IEEE Transactions on Mobile Computing* 23, 5 (2023), 3539–3557.

[34] Yang Gao, Wei Wang, Vir V Phoha, Wei Sun, and Zhanpeng Jin. 2019. EarEcho: Using ear canal echo for wearable authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.

[35] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[36] David Harwood, Muralidhara Subbarao, Hannu Hakalahti, and Larry S Davis. 1987. A new class of edge-preserving smoothing filters. *Pattern Recognition Letters* 6, 3 (1987), 155–162.

[37] Changshuo Hu, Thivya Kandappu, Yang Liu, Cecilia Mascolo, and Dong Ma. 2024. BreathPro: Monitoring Breathing Mode during Running with Earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–25.

[38] Hua Huang and Shan Lin. 2016. Toothbrushing monitoring using wrist watch. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*. 202–215.

[39] Hua Huang and Shan Lin. 2020. Met: a magneto-inductive sensing based electric toothbrushing monitoring system. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.

[40] Zawar Hussain, David Waterworth, Murtadha Aldeer, Wei Emma Zhang, Quan Z Sheng, and Jorge Ortiz. 2021. Do you brush your teeth properly? an off-body sensor-based approach for toothbrushing monitoring. In *2021 IEEE International Conference on Digital Health (ICDH)*. IEEE, 59–69.

[41] Sijie Ji, Xuanye Zhang, Yuanqing Zheng, and Mo Li. 2023. Construct 3D Hand Skeleton with Commercial WiFi. In *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems*. 322–334.

[42] S.M. Kay. 1988. *Modern Spectral Estimation: Theory and Application*. Prentice Hall. https://books.google.co.uk/books?id=wutQAAAAMAAJ

[43] Steven M Kay. 1993. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc.

[44] Kazuki Kobayashi and Tomoaki Ohtsuki. 2015. User identification based on toothbrushing information using three-axis accelerometer. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. 129–132.

[45] Joseph Korpela, Ryosuke Miyaji, Takuya Maekawa, Kazunori Nozaki, and Hiroo Tamagawa. 2015. Evaluating tooth brushing performance with smartphone sound data. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 109–120.

[46] Joseph Korpela, Ryosuke Miyaji, Takuya Maekawa, Kazunori Nozaki, and Hiroo Tamagawa. 2016. Toothbrushing performance evaluation using smartphone audio based on hybrid HMM-recognition/SVM-regression model. *Journal of Information Processing* 24, 2 (2016), 302–313.

[47] Young-Jae Lee, Pil-Jae Lee, Kyeong-Seop Kim, Wonse Park, Kee-Deog Kim, Dosik Hwang, and Jeong-Whan Lee. 2011. Toothbrushing region detection using three-axis accelerometer and magnetic sensor. *IEEE Transactions on Biomedical Engineering* 59, 3 (2011), 872–881.

[48] Haicui Li and Lei Jing. 2023. 3D Monitoring of Toothbrushing Regions and Force Using Multimodal Sensors and Unity. *IEEE Access* (2023).

[49] Yang Liu, Kayla-Jade Butkow, Jake Stuchbury-Wass, Adam Pullin, Dong Ma, and Cecilia Mascolo. 2024. RespEar: Earable-Based Robust Respiratory Rate Monitoring. *arXiv preprint arXiv:2407.06901* (2024).

[50] Yang Liu, Zhenjiang Li, Zhidan Liu, and Kaishun Wu. 2019. Real-time arm skeleton tracking and gesture inference tolerant to missing wearable sensors. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. 287–299.

[51] Yang Liu, Chengdong Lin, and Zhenjiang Li. 2021. WR-Hand: Wearable armband can track user's hand. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–27.

[52] Chengwen Luo, Xingyu Feng, Junliang Chen, Jianqiang Li, Weitao Xu, Wei Li, Li Zhang, Zahir Tari, and Albert Y Zomaya. 2019. Brush like a dentist: Accurate monitoring of toothbrushing via wrist-worn gesture sensing. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 1234–1242.

[53] Dong Ma, Andrea Ferlini, and Cecilia Mascolo. 2021. Oesense: employing occlusion effect for in-ear human sensing. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*. 175–187.

[54] Alexis Martin and Jérémie Voix. 2017. In-ear audio wearable: Measurement of heart and breathing rates for health and safety monitoring. *IEEE Transactions on Biomedical Engineering* 65, 6 (2017), 1256–1263.

[55] Medicalnewstoday. 2024. How many teeth do adults have? Anatomy and function. https://www.medicalnewstoday.com/articles/how-many-teeth-do-adults-havefunction. (Accessed on 04/24/2024).

[56] Ulrich Oberst, Martin Scheicher, and Ingrid Scheicher. 2020. *Linear Time-Invariant Systems, Behaviors and Modules*. Springer.

[57] Alan V Oppenheim. 1999. *Discrete-time signal processing*. Pearson Education India.

[58] Oral-B. 2024. The Oral-B App Will Improve your Brushing Routine | Oral-B. https://oralb.com/en-us/oral-b-app/. (Accessed on 04/24/2024).

[59] World Health Organization. 2023. Global oral health status report: towards universal health coverage for oral health by 2030. https://www.who.int/publications/i/item/9789240061484. (Accessed on 04/03/2024).

[60] Zhenchao Ouyang, Jingfeng Hu, Jianwei Niu, and Zhiping Qi. 2017. An asymmetrical acoustic field detection system for daily tooth brushing monitoring. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 1–6.

[61] Marco A Peres, Lorna MD Macpherson, Robert J Weyant, Blánaid Daly, Renato Venturelli, Manu R Mathur, Stefan Listl, Roger Keller Celeste, Carol C Guarnizo-Herreño, Cristin Kearns, et al. 2019. Oral diseases: a global public health challenge. *The Lancet* 394, 10194 (2019), 249–260.

[62] Jay Prakash, Zhijian Yang, Yu-Lin Wei, Haitham Hassanieh, and Romit Roy Choudhury. 2020. EarSense: earphones as a teeth activity sensor. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–13.

[63] Tobias Röddiger, Christopher Clarke, Paula Breitling, Tim Schneegans, Haibin Zhao, Hans Gellersen, and Michael Beigl. 2022. Sensing with earables: A systematic literature review and taxonomy of phenomena. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–57.

[64] Yang Shao, Zhaozhang Jin, DeLiang Wang, and Soundararajan Srinivasan. 2009. An auditory-based feature for robust speech recognition. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 4625–4628.

[65] Paul E Spector. 1992. *Summated rating scale construction: An introduction*. Vol. 82. Sage.

[66] Statista. 2024. U.S.: usage of manual toothbrushes 2020 | Statista. https://www.statista.com/statistics/278116/us-households-usage-of-manual-toothbrushes/. (Accessed on 04/03/2024).

[67] Xue Sun, Jie Xiong, Chao Feng, Wenwen Deng, Xudong Wei, Dingyi Fang, and Xiaojiang Chen. 2023. Earmonitor: In-ear motion-resilient acoustic sensing using commodity earphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–22.

[68] Xavier Valero and Francesc Alias. 2012. Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification. *IEEE transactions on multimedia* 14, 6 (2012), 1684–1689.

[69] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).

[70] Liyuan Wang, Xingxing Zhang, Hang Su, and Jun Zhu. 2024. A comprehensive survey of continual learning: theory, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024).

[71] Yang Wang, Feng Hong, Yufei Jiang, Chenyu Bao, Chao Liu, and Zhongwen Guo. 2024. ToothFairy: Real-time Tooth-by-tooth Brushing Monitor Using Earphone Reversed Signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 4 (2024), 1–19.

[72] Zi Wang, Yili Ren, Yingying Chen, and Jie Yang. 2022. Toothsonic: Earable authentication via acoustic toothprint. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–24.

[73] Zi Wang, Sheng Tan, Linghan Zhang, Yili Ren, Zhi Wang, and Jie Yang. 2021. Eardynamic: An ear canal deformation based continuous user authentication using in-ear wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–27.

[74] Peter Welch. 1967. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics* 15, 2 (1967), 70–73.

[75] Kaishun Wu, Qiang Yang, Baojie Yuan, Yongpan Zou, Rukhsana Ruby, and Mo Li. 2020. Echowrite: An acoustic-based finger input system without training. *IEEE Transactions on Mobile Computing* 20, 5 (2020), 1789–1803.

[76] Wentao Xie, Qingyong Hu, Jin Zhang, and Qian Zhang. 2023. EarSpiro: Earphone-based Spirometry for Lung Function Assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–27.

[77] Huanqi Yang, Mingda Han, Mingda Jia, Zehua Sun, Pengfei Hu, Yu Zhang, Tao Gu, and Weitao Xu. 2023. XGait: Cross-Modal Translation via Deep Generative Sensing for RF-based Gait Recognition. In *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems*. 43–55.

[78] Qiang Yang, Yang Liu, Jake Stuchbury-Wass, Kayla-Jade Butkow, Dong Ma, and Cecilia Mascolo. 2024. BrushBuds: Toothbrushing Tracking Using Earphone IMUs. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 655–660.

[79] Qiang Yang and Yuanqing Zheng. 2023. Aquahelper: Underwater sos transmission and detection in swimming pools. In *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems*. 294–307.

[80] Zhijian Yang, Yu-Lin Wei, Sheng Shen, and Romit Roy Choudhury. 2020. Ear-ar: indoor acoustic augmented reality on earphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.

[81] Shiming Yu, Xianjin Xia, Ningning Hou, Yuanqing Zheng, and Tao Gu. 2024. Revolutionizing LoRa Gateway with XGate: Scalable Concurrent Transmission across Massive Logical Channels. In *ACM MobiCom*. 482–496.

[82] Shiming Yu, Xianjin Xia, Ziyue Zhang, Ningning Hou, and Yuanqing Zheng. 2024. FDLoRa: Tackling Downlink-Uplink Asymmetry with Full-duplex LoRa Gateways. In *ACM SenSys*. 281–294.

[83] Kuang Yuan, Mohamed Ibrahim, Yiwen Song, Guoxiang Deng, Robert A Nerone, Suvendra Vijayan, Akshay Gadre, and Swarun Kumar. 2024. ToMoBrush: Exploring Dental Health Sensing Using a Sonic Toothbrush. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 3 (2024), 1–27.

[84] Tao Zhang, Yangyang Shao, Yaqin Wu, Yanzhang Geng, and Long Fan. 2020. An overview of speech endpoint detection algorithms. *Applied Acoustics* 160 (2020), 107133.

[85] Yongpan Zou, Qiang Yang, Yetong Han, Dan Wang, Jiannong Cao, and Kaishun Wu. 2019. AcouDigits: Enabling users to input digits in the air. In *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*. IEEE, 1–9.