HearForce: Force Estimation for Manual Toothbrushing with Earables

QIANG YANG, University of Cambridge, United Kingdom
YANG LIU, University of Cambridge, United Kingdom
JAKE STUCHBURY-WASS, University of Cambridge, United Kingdom
MATHIAS CILIBERTO, University of Cambridge, United Kingdom
TOBIAS RÖDDIGER, Karlsruhe Institute of Technology, Germany
KAYLA-JADE BUTKOW, University of Cambridge, United Kingdom
ADAM PULLIN, University of Cambridge, United Kingdom
EMELI PANARITI, King's College London, United Kingdom
DONG MA, Singapore Management University, Singapore
CECILIA MASCOLO, University of Cambridge, United Kingdom

Excessive tooth brushing force can contribute to oral health issues such as gum recession and bleeding. While some electric toothbrushes offer force feedback, manual toothbrushes remain widely used due to their affordability and accessibility but do not give this sort of information to users. Therefore, they often lack awareness of the force they apply while brushing. In this paper, we introduce HearForce, the first system to estimate tooth brushing force from manual toothbrushing using widely available earbuds. Unlike existing solutions that require significant modifications to toothbrushes, HearForce leverages in-ear microphones on commercial earbuds to capture bone-conducted toothbrushing sound propagating from the oral cavity to the ear canals. Our key insight is that variations in brushing force modulate these toothbrushing sounds due to the friction effect, allowing us to infer force levels through deep learning. However, individual habitual and anatomical differences introduce significant challenges for force estimation. To mitigate this, we propose a self-supervised representation learning network with a cross-attention mechanism to suppress user-dependent variability and a heuristic calibration strategy to adapt the model to different brushing habits. Through extensive evaluation, HearForce demonstrates force estimation capabilities with a Mean Absolute Error (MAE) of 37.3g in user-independent settings, corresponding to 11.4% of the typical force dynamic range. Our study makes the first step in the use of everyday earbuds for manual toothbrushing force monitoring, paving the way for accessible solutions to improve brushing habits for manual toothbrush users globally.

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

Additional Key Words and Phrases: Toothbrushing monitoring, Acoustic sensing, Earable devices

ACM Reference Format:
Qiang Yang, Yang Liu, Jake Stuchbury-Wass, Mathias Ciliberto, Tobias Röddiger, Kayla-Jade Butkow, Adam Pullin, Emeli Panariti, Dong Ma, and Cecilia Mascolo. 2025. HearForce: Force Estimation for Manual Toothbrushing with Earables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 4, Article 232 (December 2025), 22 pages. https://doi.org/10.1145/3770688

Authors' Contact Information: Qiang Yang, qiang.yang@cl.cam.ac.uk, University of Cambridge, United Kingdom; Yang Liu, yl868@cam.ac.uk, University of Cambridge, United Kingdom; Jake Stuchbury-Wass, js2372@cam.ac.uk, University of Cambridge, United Kingdom; Mathias Ciliberto, mc2514@cam.ac.uk, University of Cambridge, United Kingdom; Tobias Röddiger, tobias.roeddiger@kit.edu, Karlsruhe Institute of Technology, Germany; Kayla-Jade Butkow, kjb85@cam.ac.uk, University of Cambridge, United Kingdom; Adam Pullin, alp78@cam.ac.uk, University of Cambridge, United Kingdom; Emeli Panariti, k21163606@kcl.ac.uk, Faculty of Dentistry, Oral and Craniofacial Sciences, King's College London, London, United Kingdom; Dong Ma, dongma@smu.edu.sg, Singapore Management University, Singapore; Cecilia Mascolo, cecilia.mascolo@cl.cam.ac.uk, University of Cambridge, United Kingdom.

ACM 2474-9567/2025/12-ART232
https://doi.org/10.1145/3770688

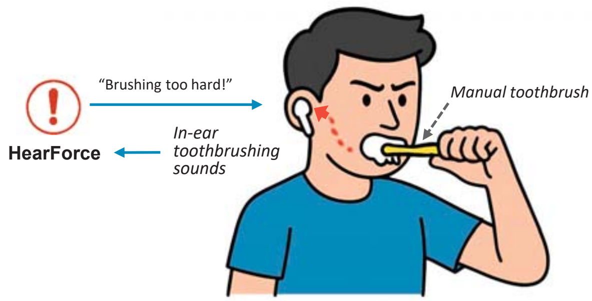


Fig. 1. HearForce utilizes in-ear microphones embedded in earbuds to monitor the brushing force for manual toothbrushes.

1 Introduction

Oral health is essential to overall well-being, yet it remains a widespread public health concern [40]. According to the World Health Organization (WHO), oral diseases affect nearly 3.5 billion people globally, with untreated dental cavities being the most common health condition worldwide [37]. Among the many contributing factors, daily toothbrushing is widely recognized as a foundational preventive measure. Specifically, both the duration and the applied force during brushing play a critical role in determining its effectiveness [1]. Insufficient brushing (both in time and force) may leave plaque behind, while excessive brushing can damage tooth enamel and irritate the gums, potentially leading to conditions such as gingival recession and tooth sensitivity [1].

To help users brush effectively, modern high-end smart electric toothbrushes often include built-in force/motion sensors and provide real-time feedback on brushing duration and force levels [2]. However, manual toothbrushes still dominate the global market due to their affordability, simplicity, and wide accessibility [47]. As a result, most users lack access to toothbrushing guidance during their daily brushing routine. Recent research has explored the use of commodity sensors with cameras [7, 10], inertial measurement units (IMUs) [8, 20, 31], and microphones [25, 55] to monitor brushing coverage and duration. Despite this progress, *brushing force* has received surprisingly limited attention. Some prior studies [6, 28] attempted to estimate brushing force by instrumenting the toothbrush with multiple force sensors attached to the handle. However, this approach requires significant hardware modification and imposes restrictions on how the toothbrush is held, making it impractical to use. This leads us to a fundamental research question: *Can we estimate brushing force for manual toothbrushes using only off-the-shelf devices, without modifying hardware?*

In this paper, we explore this possibility by presenting HearForce, the first system that estimates continuous toothbrushing force using commodity earable devices, such as off-the-shelf earbuds and hearing aids. Unlike prior work that relies on instrumented toothbrushes, HearForce requires no hardware modification and leverages existing microphones built into widely available earbuds, enabling tooth brushing force estimation without any additional hardware or modifications. This enables a low-cost, unobtrusive, and highly accessible solution for continuous brushing force monitoring in daily life. The key idea behind HearForce is that earable devices are naturally positioned close to the oral cavity and can pick up toothbrushing sounds transmitted through bone conduction [42]. As shown in Fig. 1, HearForce exploits the in-ear microphones on earbuds, typically used for active noise cancellation (ANC), to capture subtle vibrations generated during brushing. Since brushing force affects the contact pressure between bristles and teeth, it alters the frictional interactions and thus modulates the resulting acoustic signal [17]. These variations are captured by the in-ear microphones and can be mapped to force levels using deep-learning models, enabling toothbrushing force monitoring for manual toothbrushes.

However, building such a system presents several challenges. First, the relationship between brushing force and the resulting in-ear sound is highly nonlinear and influenced by many factors, including toothbrush type, bristle stiffness, and the brushing location [8]. These variables make it difficult to construct a deterministic model. Second, since brushing sounds originate from physical contact inside the mouth, they are heavily distorted by individual behavioral features like brushing speed and dynamics, as well as anatomical features such as jaw structure, skull thickness, and oral cavity geometry [8]. These factors introduce significant user-specific distortions, making it challenging for learning-based models to extract force-relevant features that generalize across users. Third, unlike smart electric toothbrushes, manual toothbrushes do not provide built-in force measurement. This makes it difficult to personalize models using standard user registration procedures, as no ground-truth data is available for calibration.

To address these challenges, we propose the following technical approaches. First, we develop a physical understanding of how brushing force translates to frictional sound and how that sound propagates through the head before reaching the ear canal. This model guides our feature extraction and helps isolate non-force-related variations, particularly those related to user anatomy and behavior. We also further employ a self-supervised strategy on a large number of unlabeled in-ear brushing audio to learn robust feature representations. Second, we design an attention-based mechanism that explicitly extracts and disentangles user-dependent characteristics from force-related features, suppressing variability biases during learning. Finally, to enable user adaptation without access to ground-truth force values, we introduce a lightweight heuristic calibration procedure. During the first-time setup, users are asked to brush different regions of the mouth with varying self-applied forces. This produces a relative distribution of signals, which the model uses to adapt to individual brushing styles and toothbrush differences.

We implement HearForce on a prototype platform by modifying a commercial electric toothbrush to record brushing force ground truth while disabling its vibration motor to mimic manual brushing. We also developed an open-source force-aware toothbrush for evaluation. Using this setup, we conduct a user study with 16 participants, collecting brushing audio from prototype earphones along with reference force data across different toothbrush types, brushing regions, and force levels. Our experimental results demonstrate the effectiveness of HearForce in brushing force estimation, with a Mean Absolute Error (MAE) of 37.3g in a user-independent setting, corresponding to 11.4% of the typical force dynamic range.

Since our system runs with earphones, it can naturally deliver real-time audio feedback when the predicted brushing force is too high or too low, helping users adjust their force accordingly. We further present a case study in which HearForce generates a post-brushing force distribution report across different regions of the mouth. Feedback from a collaborating dentist confirms its clinical relevance in identifying brushing imbalances and providing personalized oral care guidance. Our goal is to bring manual toothbrushes closer to electric toothbrushes in terms of providing real-time brushing force feedback. Beyond the fixed-threshold feedback used by commercial electric toothbrushes, our system provides continuous force monitoring, which can be seamlessly integrated with existing work on earable brushing coverage monitoring [52, 55] to deliver more personalized, fine-grained, and context-aware oral health guidance.

To summarize, this paper makes the following contributions:

- To the best of our knowledge, HearForce is the first system to enable continuous brushing force estimation for manual toothbrushes using everyday earables, without requiring any modification to the toothbrush.
- We develop a set of technical approaches to tackle key challenges, including a self-supervised feature encoder, a cross-attention-based network, and a heuristic user calibration method.
- The experiments validate the effectiveness of HearForce in brushing force estimation. A case study further demonstrates its clinical potential, marking a first step toward accessible and personalized brushing force monitoring for manual toothbrush users.
- We develop and open-source a toothbrush with integrated force sensing to encourage broader community engagement in oral health monitoring.

2 Preliminaries

2.1 Motivation: Clinical Perspectives on Brushing Force

Brushing force is a critical factor in maintaining oral health. Excessive force can cause enamel abrasion and gingival recession [53], while insufficient force may lead to inadequate plaque removal [35]. Despite its importance, there is currently no universally accepted clinical standard that defines an optimal brushing force for all individuals [43]. Dental associations typically offer qualitative guidance. For example, the American Dental Association (ADA) recommends using a soft-bristled brush and brushing "gently" with short strokes rather than scrubbing hard [4]. Our discussions with collaborating dental professionals further highlight that, in clinical practice, there is currently no universally established threshold for defining optimal or excessive brushing force. Instead, it should be tailored to the patient's oral health condition. For example, patients with gingival recession require significantly lighter brushing force, whereas patients who smoke may require greater force to effectively remove stains. In commercial practice, some electric toothbrushes incorporate fixed force thresholds (*e.g.*, 2.5N) to trigger pressure warnings [23, 33]. However, these thresholds are not clinically standardized and are normally conservative and follow a one-size-fits-all approach across the general population [33]. In addition, clinical practice also cares about the balance of force across different regions, as users tend to apply excessive force to the first region brushed or the region opposite to their dominant hand, leading to localized enamel wear and periodontal damage [49]. However, manual toothbrush users currently have no built-in force sensing or feedback mechanisms to assist them during brushing. Motivated by these findings, our work aims to advance continuous brushing force monitoring for manual toothbrush users, providing accessible feedback where none currently exists. We hope that enabling force estimation in everyday brushing will promote further research attention and clinical development in this under-addressed yet critical aspect of oral health.

Note that we focus on the *force* exerted on the teeth during brushing in grams (g), omitting the gravitational constant $\mathbf{g}$ for simplicity. Although *pressure* (the force per unit area) might be more physically accurate, the contact area of a toothbrush is difficult to estimate, hence we use force as a practical proxy like other studies [28].

2.2 From Brushing Force to Toothbrushing Sounds

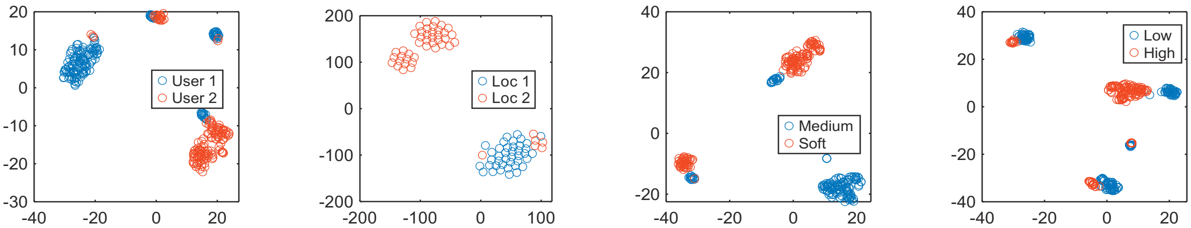
The generation of tooth brushing sound is a complex process influenced by several physical factors, including the applied brushing force, the frictions between the toothbrush bristles and the tooth surface, the subsequent vibrations induced in the bristles and dental structure, and the propagation of the sound from the teeth to the two ear canals [42]. In this section, we build a physical model to understand this process and identify key sources of variability that significantly affect the audio signals. As such, we can utilize the physical understanding to guide our model design.

Brushing Sound Generation. The interaction between toothbrush bristles and the tooth surface can be viewed as a friction-driven vibration source [17]. When we apply a certain level of force $F(t)$ to brush our teeth and move with a relative velocity $\boldsymbol{v}$, the interaction between the toothbrush bristles and the tooth surface generates a frictional force $F'(t)$, which can be expressed as:

$$F'(t) = f(\mu, \mathbf{v}, F(t)) \tag{1}$$

where μ is the friction coefficient between the bristles and the tooth surface, denoting a collective representation of all relevant material and surface characteristics, such as bristle stiffness, tooth roughness, and toothpaste conditions [17]. This friction force acts as an excitation to the mass-spring-damper system [6], leading to bristle vibrations, which in turn contribute to the sound produced during brushing. Therefore, the resulting vibration $x(t)$ of bristles can be described by the following second-order differential equation:

$$m\frac{d^2x(t)}{dt^2} + c\frac{dx(t)}{dt} + kx(t) = F'(t) \tag{2}$$



(a) Different users. (b) Different locations. (c) Different toothbrushes. (d) Different speeds.

Fig. 2. t-SNE visualization of in-ear brushing audio features under different factors.

Here, m represents the effective mass representing the inertia of the bristle tips. c and k are the damping coefficient and the stiffness of the system, respectively, which are related to the material characteristics of the brush bristle.

From Teeth to Ears. The brushing sound is generated by pressure waves caused by vibration $x(t)$ and then propagates through the bones and facial soft tissues to both ear canals. Therefore, we can model this propagation process as a Linear Time-Invariant (LTI) system [55] and present the toothbrushing sound in ear canals $y(t)$ as:

$$y(t) = x(t) * h(t) + n(t) \tag{3}$$

where $h(t)$ is the impulse response of the propagation channels between the tooth surface and both ear canals, which are associated with the physiological structure of humans. $n(t)$ is the noise.

From this model, we can observe that the in-ear toothbrushing sound involves a sequence of mechanical interactions that start with the applied brushing force. This force creates friction between the toothbrush bristles and the tooth surface, which in turn induces bristle vibrations. As the bristle vibrates, the brushing sounds are transferred into the pressure waves, which travel from the dental surface to both ear canals through the bones and facial muscles. We can combine Equations 1 to 3 and express the resulting in-ear audio as:

$$y(t) = S(\underline{F(t)}, \underline{h(t)}, \underline{I(\mu, m, c, k, \boldsymbol{v})}) \tag{4}$$

we omit the noise term n for simplicity, as its impact is relatively minor compared to other dominant components.

Remarks. From Eq. 4, we can identify three key sources of variability that significantly affect the in-ear audio signals for toothbrushing: (1) the applied brushing force $F(t)$, (2) the user's physiological structure $h(t)$ (*e.g.*, skull and oral cavity geometry), and (3) a composite term I that captures the effects of brushing friction dynamics and brushing behavior. This abstraction helps guide our algorithm design to estimate the brushing force from the in-ear audio.

2.3 Empirical Study

To validate our modeling that in-ear brushing audio is affected by multiple confounding factors beyond brushing force, we conducted an empirical study analyzing the variation in brushing sounds under controlled changes in users, brushing locations, toothbrush types, and brushing speeds. We collected a brushing audio dataset under different conditions and extracted acoustic features using the OpenSMILE toolkit [13]. We then applied t-SNE [50] to project the features into a 2D space for visualization and analysis.

As shown in Fig. 2(a), the brushing sounds from different users form well-separated clusters, even when using the same toothbrush and brushing the same area. This indicates that individual anatomical differences significantly affect in-ear acoustic signals. In Fig. 2(b), we observe distinct clustering patterns when brushing different regions of the mouth (*i.e.*, left *v.s.* right molars) for the same user. This result shows that the difference in surface geometry and oral cavity resonance results in location-specific sound patterns. Fig. 2(c) illustrates the

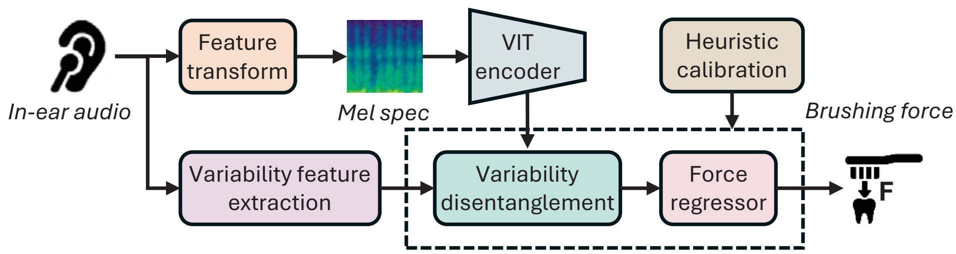


Fig. 3. The system overview of SmarTeeth.

influence of bristle stiffness. Sounds generated using soft and medium toothbrushes are clearly distinguishable in the feature space, as stiffer bristles produce stronger vibrations and contact noise. We then quantify brushing speed by analyzing the interval between adjacent brush strokes. We select segments from the upper and lower tails of the speed distribution to define high-speed and low-speed brushing examples. As shown in Fig. 2(d), although the clusters for different speeds are closer compared to other factors, we still observe noticeable shifts in the feature space, suggesting that brushing speed may cause the change in vibration frequency and friction dynamics.

Design Guidelines. These results and Eq. 4 indicate that in-ear brushing audio is influenced by a range of physiological, physical, and behavioral factors, making it infeasible to map audio signals to brushing force using a simple or deterministic model. These insights guided the design of our framework in three key ways. **First**, they highlight the need to enhance generalizability by capturing the broad variability introduced by different toothbrush materials and brushing behaviors (Sec. 3.3). **Second**, they suggest that we need to eliminate the impact of user-specific physiological differences, such as variations in skull and oral cavity structure (Sec. 3.4). **Third**, they underscore the need for a calibration step to adapt to personal brushing behaviors (Sec. 3.6). Together, these strategies aim to improve the robustness and cross-user consistency of brushing force estimation, as detailed in the following section.

3 System Design

3.1 System Overview

As illustrated in Fig. 3, HearForce takes binaural toothbrushing audio signals recorded by in-ear microphones as input. The raw audio is first converted into Mel spectrograms through the feature transform block (Sec. 3.2). A feature encoder, pretrained on unlabelled toothbrushing sounds, is then used to obtain high-level representations from the spectrograms (Sec. 3.3). However, in-ear audio inevitably contains user-dependent physiological and behavioral variability, which may degrade model performance. Hence, we extract variability features from the audio using signal processing techniques in the variability feature extraction module (Sec. 3.4.1) and then introduce a cross-attention mechanism to explicitly disentangle user-dependent components from the learned toothbrushing representations (Sec. 3.4.3). To further mitigate variability caused by individual brushing habits and toothbrush types, we incorporate a heuristic calibration module to align the feature distributions across users (Sec. 3.6). Finally, a force prediction module estimates the brushing force levels as the final result (Sec. 3.5).

3.2 Preprocessing and Feature Transformation

To process the binaural in-ear audio signals captured during toothbrushing, we first segment the raw waveforms into fixed-length windows of 1 second using a sliding window with a stride of 0.5s. Within this time frame, the toothbrush typically remains on the same or adjacent tooth surfaces, allowing the window to capture a temporally

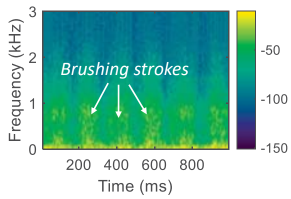


Fig. 4. Brushing sounds.

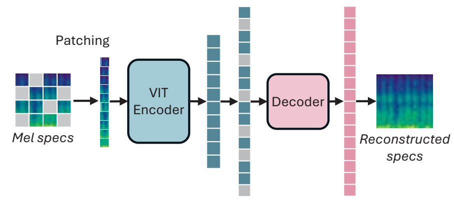


Fig. 5. VIT encoder pretraining.

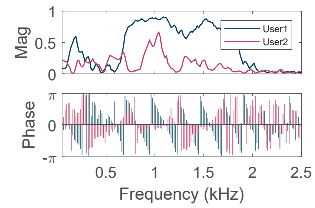


Fig. 6. Anatomical features.

consistent brushing force and acoustic pattern. Figure 4 illustrates the spectrogram of a toothbrushing audio clip. We observe that most toothbrushing sounds are concentrated below 2.5 kHz, while frequencies below 50 Hz are often contaminated by hardware-induced noise. To suppress such low-frequency interference and retain informative signals, we apply a 20th-order bandpass Butterworth filter [36] with cutoff frequencies set to [50, 2500] Hz. The filtered signals are then transformed into Mel spectrograms, which better capture the perceptual characteristics of audio [44]. Since the traditional Mel filter banks are typically designed for speech signal processing and often span up to 8 kHz, we specifically re-designed the Mel filter bank using 64 half-overlapping triangular filters equally spaced on the Mel scale within the [50, 2500] Hz range. We used a short-time Fourier transform (STFT) with a window size of 1024 samples, and a 64-dimensional Mel filter bank is applied to compress the spectral representation, balancing temporal-spectral resolution and computational efficiency. The resulting Mel spectrograms $\mathbf{M} \in \mathbb{R}^{64\times64\times2}$ serve as the input to the subsequent representation learning module. The first two dimensions correspond to frequency bins and time frames, respectively, and the last dimension represents the two in-ear microphone channels (left and right).

3.3 Sound Representation Learning

To improve the generalizability of brushing force estimation across different toothbrush materials and user behaviors (*i.e.*, the first guideline as discussed in Sec. 2.3), we begin by learning robust representations from in-ear audio. This task is highly challenging, as the acoustic signals are influenced by various confounding factors beyond force, such as user anatomy, brushing location, bristle stiffness, brushing speed, and angle. Directly training a supervised model from scratch may lead to overfitting, especially when labeled data is limited. However, collecting enough labeled data is inherently difficult because measuring ground-truth brushing force requires specialized hardware (*e.g.*, modified toothbrushes), dedicated software, and a controlled lab environment. This setup is impractical for daily use as users cannot perform such procedures at home. In contrast, unlabeled in-ear audio can be easily collected as users simply wear their earbuds and brush their teeth as usual. This ease of data acquisition enables us to build a large and realistic brushing audio corpus, which is well-suited for self-supervised learning to pretrain a feature encoder [29]. This encoder captures high-level representations that are more resilient to noise and sound variability [58].

As shown in Fig. 5, we employ a Vision Transformer-based Masked Autoencoder (ViT-MAE) architecture [15] to learn informative representations from Mel spectrograms on an in-house unlabelled dataset (we will introduce it in Sec. 4). Specifically, we chose AudioMAE [21] as our pretraining backbone due to its proven effectiveness on time-frequency representations of audio. AudioMAE introduces local window attention into ViT-MAE to better capture local time-frequency correlations, leading to superior representation learning. Since the traditional Mel filter banks are typically designed for speech signal processing and often span up to 8 kHz, we specifically re-designed the Mel filter bank using 64 half-overlapping triangular filters equally spaced on the Mel scale within the [50, 2500] Hz range. The input spectrogram is first divided into non-overlapping patches of size 4×4, and then each patch is linearly projected into a latent embedding. We add learnable absolute positional encodings to preserve temporal-frequency structure. During pretraining, 75% of the patches are randomly masked, and the model is trained to reconstruct the missing spectrogram regions from the remaining visible patches. This masked prediction task encourages the encoder to learn contextual and semantically meaningful patterns in toothbrushing audio. The overall architecture includes an encoder for processing masked patches and a decoder for reconstructing the full spectrogram. The encoder consists of 4 transformer layers, each with 256-dimensional embeddings and 4 attention heads. The decoder has a more lightweight structure with 2 transformer blocks. After pretraining, we discard the decoder and use only the encoder as the feature extractor in downstream tasks.

3.4 Variability Feature Extraction and Disentanglement

To address the second challenge identified in Sec. 2.3 (*i.e.*, the impact of user-specific physiological differences), we further explore how to mitigate the effects of individual anatomy and behavior on in-ear audio signals. Although self-supervised pretraining improves the robustness of the learned audio representations, in-ear brushing signals are still significantly affected by user-specific variability. This variability arises from both *anatomical factors*, such as ear canal shape, jawbone structure, skull thickness, and oral cavity geometry, as well as *behavioral factors*, including brushing speed and motion dynamics. These characteristics modulate how sound propagates from the mouth to the in-ear microphones via bone and tissue conduction, introducing substantial signal variations that are unrelated to brushing force. To address this challenge, we explicitly extract variability features from the in-ear audio, including both anatomical and behavioral components. We then introduce a cross-attention-based disentanglement module [51] that isolates and suppresses these user-specific features from the learned brushing representations. This design encourages the downstream force estimation model to focus on brushing-induced acoustic patterns, rather than being misled by individual anatomical or habitual artifacts.

3.4.1 Anatomical Feature Extraction. We observe that one of the primary sources of inter-user variation in in-ear brushing audio lies in the propagation path from the teeth to the ear canal. This path is shaped by individual anatomical structures, such as skull shape, bone density, and tissue composition, which affect how sound waves are attenuated, delayed, or distorted as they travel through the head.

Inspired by prior work on binaural acoustic modeling [55], we capture these differences by analyzing the brushing sound received simultaneously by the left and right in-ear microphones. Since the brushing sound originates from the same source (*i.e.*, the toothbrush on the teeth), any consistent differences between the two signals must come from the different anatomical paths the sound travels through. By comparing the frequency difference between two channels, we extract a pair of user-specific features: one captures how much the sound energy differs across frequencies (*i.e.*, magnitude response), and the other captures the timing difference across frequencies (*i.e.*, phase response) between the two sides. Together, these features form a compact representation of the user's anatomical transfer characteristics, *i.e.*, how the user's anatomical structure shapes the sound propagation paths. We refer to them as an anatomical feature.

As shown in Fig. 6, the normalized magnitude vector captures the relative energy difference between the two ears across frequencies, and the phase vector reflects the relative delay and phase shift. We observe that different users exhibit significantly different transfer functions, which are shaped by their individual anatomical structures. Since we applied a 1024-point FFT on the audio window to generate the Mel spectrogram in Sec. 3.2, we obtain two 158-dimensional vectors from magnitude and phase within the valid frequency band of [50, 2500]Hz. This feature is later used in our model to help isolate the brushing force from the confounding effects of different users' physical structures.

3.4.2 Behavioral Feature Extraction. In addition to anatomical factors, users exhibit substantial behavioral variability in their brushing style, particularly in terms of motion dynamics such as brushing speed and stroke

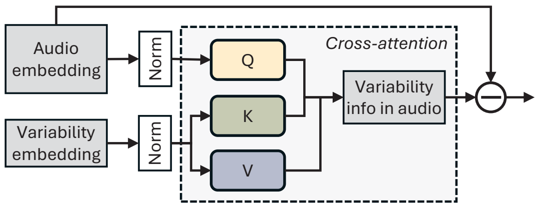


Fig. 7. Cross-attention-based variability disentanglement module.

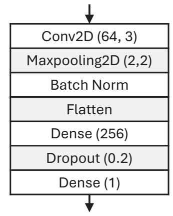


Fig. 8. Force estimation head.

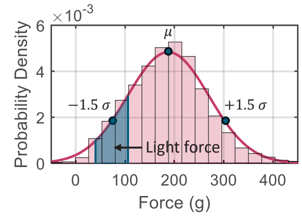


Fig. 9. Force distribution and pseudo-label generation.

patterns. From visual inspection of the in-ear audio (Fig. 4), we observe that brushing signals within a short window often contain multiple periodic strokes, reflecting the repetitive back-and-forth brushing motion.

To capture these dynamics, we compute the autocorrelation of the audio signal within each window. Autocorrelation characterizes the signal's periodicity and provides a proxy for brushing speed and motion rhythm. Intuitively, a faster brushing speed results in shorter stroke intervals and thus different autocorrelation profiles compared to slower, more deliberate brushing. According to the Wiener–Khinchin theorem [11], the power spectral density (PSD) of a signal serves as the frequency-domain equivalent of its autocorrelation function. Therefore, we compute the PSD of the in-ear audio signals as a compact and informative representation of motion dynamics. For each audio window, we extract PSD features from both the left and right in-ear microphone channels. Each channel produces a 158-dimensional behavioral feature vector. These vectors, together with the anatomical features, are later incorporated into the variability disentanglement module to separate user-specific artifacts from brushing force-related representations.

3.4.3 Variability Disentanglement. To remove user-specific variability interference from the learned audio representations, we design a cross-attention module [59] that adaptively suppresses variability components while preserving brushing-related features. The key idea is to use the variability feature vector to probe the learned brushing representations (*i.e.*, the encoder outputs) to identify and filter out the user-dependent components.

As illustrated in Fig. 7, let $\mathbf{A} \in \mathbb{R}^{4\times 158}$ denote the variability feature vector, and $\mathbf{M}' \in \mathbb{R}^{T\times d}$ the sequence of brushing audio representations output by the encoder, where $T = 256$ is the number of tokens and $d = 32$ is the feature dimension. We compute the cross-attention as follows:

$$\text{Attention}(\mathbf{M}', \mathbf{A}, \mathbf{A}) = \text{Softmax}\left(\frac{(\mathbf{M}'\mathbf{W}_Q)(\mathbf{A}\mathbf{W}_K)^\top}{\sqrt{d}}\right)(\mathbf{A}\mathbf{W}_V), \tag{5}$$

where $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$ are learnable projection matrices for the query, key, and value, respectively. The attention weights indicate how strongly each brushing token is influenced by the variability pattern, and the output provides a user-dependent component that can be removed from the original representation. The final cleaned representation is obtained by subtracting the attended interference from the original audio embedding [16]:

$$\mathbf{M}'_{\text{clean}} = \mathbf{M}' - \text{Attention}(\mathbf{M}', \mathbf{A}, \mathbf{A}). \tag{6}$$

This residual operation enables the model to retain force-relevant brushing patterns while explicitly discarding components aligned with the variability feature. By training this module end-to-end along with the force estimation objective (Sec. 3.5), the model learns to automatically suppress user-specific distortions and generalize better across different individual anatomies and behaviors.

3.5 Force Estimation

After variability disentanglement, the cleaned brushing representations $\mathbf{M'}_{\text{clean}}$ are passed into a lightweight convolutional neural network to estimate the applied brushing force. As illustrated in Fig. 8, we adopt a three-layer Convolutional Neural Network (CNN) regression model that captures local temporal-frequency patterns in the representation and maps them to a scalar force value. Each convolutional layer is followed by a batch normalization and ReLU activation to stabilize training and improve nonlinearity. The final output is obtained via a global average pooling layer followed by a linear regression head.

To account for the large variations in brushing force values and to mitigate the impact of outliers on training, we adopt the Huber loss [22] as our regression objective instead of the standard mean squared error (MSE) loss. Due to the influence of diverse user differences, brushing locations, and toothbrush types, the force distribution exhibits considerable variation and extreme force values, although rare, could disproportionately dominate the optimization process if using MSE. The Huber loss offers a compromise between MSE and MAE:

$$\mathcal{L}_{\text{Huber}}(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{if } |y - \hat{y}| \leq \delta, \\ \delta \cdot (|y - \hat{y}| - \frac{1}{2}\delta) & \text{otherwise}, \end{cases} \tag{7}$$

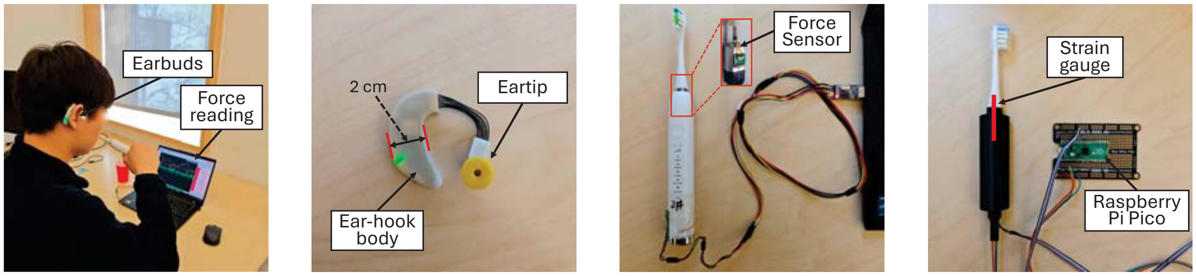
where y is the ground-truth force and $\hat{y}$ is the predicted value. We empirically set $\delta = 50$ in our experiments. This regression model is trained end-to-end alongside the variability disentanglement module, enabling joint optimization of interference suppression and force estimation under realistic noisy conditions.

3.6 Heuristic Calibration

To address the third challenge identified in Sec. 2.3 (*i.e.*, the need to adapt to individual brushing habits), we incorporate a lightweight calibration mechanism to personalize force estimation. At this stage, our system incorporates self-supervised pretraining to improve generalization to behavioral patterns (*e.g.*, brushing method and rhythm), and variability disentanglement to suppress anatomical differences (*e.g.*, bone structure). However, completely eliminating user-specific variation remains challenging due to differences in individual habitual force levels and toothbrush types. While our pretraining dataset spans 13 users with diverse brushing habits, real-world brushing behavior can vary widely across individuals. Moreover, unlike smart electric toothbrushes, manual toothbrushes do not provide real-time force measurements, making it impossible for end users to self-calibrate the system using ground-truth data.

To address this, we propose a heuristic calibration procedure that enables lightweight model adaptation for new users without requiring any special hardware or labeled force data. During the initial setup, new users are asked to brush their teeth on sextants [39] (*i.e.*, 6 equal parts in the mouth) in three self-reported force conditions: light, normal, and strong, corresponding to typical usage scenarios in daily life. Each part is brushed for about 30 seconds. Instead of collecting actual force values, which would require specialized equipment, we assign pseudo-labels to these samples using a heuristic derived from the population-wide force distribution observed in the training data, as shown in Fig. 9.

Specifically, we empirically choose three representative force ranges corresponding to light, normal, and strong brushing levels, centered at $\mu - 1.5\sigma$, μ, and $\mu + 1.5\sigma$, respectively. For example, a light brushing trial may be assigned a pseudo-label sampled from $[\mu - 2\sigma, \mu - 1\sigma]$. Within each range, the local distribution follows a truncated normal pattern, which unavoidably skews pseudo-label sampling toward the denser side. To reflect the coarse-grained nature of user-perceived force distinction, we sample uniformly within each segment, ensuring a broader and more even coverage of plausible force values within that level. To introduce variability, we randomly perturb each label within a 0.5σg range. We then perform lightweight fine-tuning of the trained force estimation head using these data, minimizing the Huber loss defined in the previous section. This approach requires no additional hardware and can be completed in about three minutes (under the App guidance). It enables the model



(a) Data collection (b) Earbuds (c) Modified toothbrush (d) Opensource toothbrush

Fig. 10. System implementation and experiment setting.

to adapt to user-specific brushing habits and toothbrush types, thereby improving force estimation accuracy under personalized conditions.

4 Evaluation

Implementation. Current commercial earbuds do not provide direct access to raw in-ear microphone signals. To collect in-ear audio, we use the OpenEarable 1.4 platform [46] (Fig. 10(b)), an open-source earable device to record the in-ear audio at 16KHz. Likewise, commercial electric toothbrushes do not offer access to internal force sensor readings. To obtain ground-truth force measurements, we modify a commercial electric toothbrush by disabling its motor and externally wiring its built-in force sensor to a USB ADC connected to a laptop (as shown in Fig. 10(c)). This modification allows us to simulate manual brushing while continuously recording force values at a sampling rate of 30Hz. To further promote community engagement, we also design a fully open-source toothbrush prototype OpenBrush [56] (Fig. 10(d)). Specifically, we use a strain gauge mounted on the toothbrush neck to detect force during brushing. The signal is amplified by an ADC module, and then collected with a Raspberry Pi Pico [41]. Before data collection, we calibrated all force sensors using standard calibration weights to ensure accurate and consistent readings across devices. We implement our model using TensorFlow. We use the Adam optimizer with a batch size of 256, and train the model for 128 epochs. This training setting is also used during ViT encoder pretraining. For fine-tuning, we use a batch size of 16 and fine-tune the model for 64 epochs. On a ThinkPad X1 laptop with an Intel i7-1370P 1.90 GHz CPU, inference takes approximately 52 ms per 1-second audio window, which is enough to support real-time feedback on most current smartphones.

Data collection. We recruited 16 participants (8 male and 8 female, aged 23~36) to participate in our data collection study, which was approved by the Ethics Committee of the University. Each participant was instructed to brush their teeth twice daily for five consecutive days, using our modified toothbrush system and wearing OpenEarable devices (as shown in Fig. 10(a)). One session lasts for about 2 minutes. Participants used interchangeable brush heads of different stiffness levels across sessions. To reflect realistic usage, they were told to brush according to their natural habits without any guidance on force or motion. For pretraining, we use an in-house in-ear toothbrushing audio dataset collected from 13 participants who brushed their teeth twice daily over five consecutive days. Seven of these participants also took part in the evaluation dataset collection. To train our model effectively under limited data, we adopt a stage-wise training with a progressive freezing strategy. We first pre-train the AudioMAE encoder using our unlabeled dataset to learn general-purpose acoustic representations. After this pretraining stage, the encoder is frozen to preserve the learned features. For each evaluation split, we then train the disentanglement module and the force estimation head using data from 15 participants. Once this stage is complete, we freeze the disentanglement module as well. In the final step, we fine-tune the force

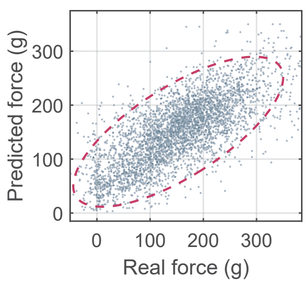


Fig. 11. Overall Performance.

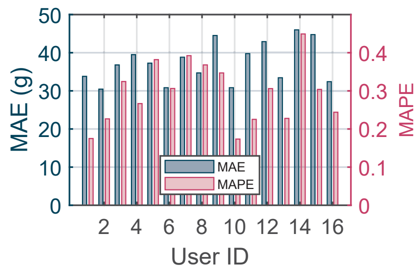


Fig. 12. Impact of different users.

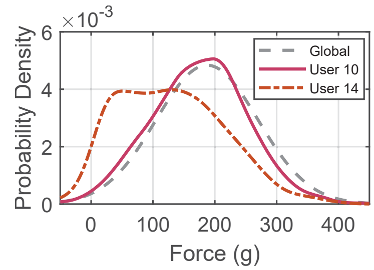


Fig. 13. Force distribution of different users.

estimation head using a small amount of calibration data from the held-out participant to adapt to user-specific brushing patterns.

Evaluation and Metrics. We adopted a leave-one-subject-out (LOSO) cross-validation scheme to assess the model's generalization ability to unseen users. During encoder pretraining, we exclude the data from the held-out subject to prevent data leakage. In addition, we conducted controlled analyses across different variables, such as toothbrush types, different users, and sessions. We evaluate the system using three standard metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Pearson correlation coefficient (r). MAE provides an intuitive measure of average difference in grams between predicted and ground-truth forces, with a lower value indicating better absolute accuracy. MAPE further normalizes this error by the ground-truth magnitude, reflecting relative prediction accuracy across varying force levels. Lastly, the Pearson correlation coefficient quantifies the linear association between predictions and ground-truth, capturing whether the model preserves the overall trend and ranking of force levels. We also conducted a user study to reveal users' imbalanced brushing patterns. A professional dentist was invited to review the force estimation distribution visualizations generated by our system and the ground truth and provide qualitative feedback.

4.1 Overall Performance

Figure 11 presents a scatter plot comparing the predicted and ground-truth brushing force values, overlaid with a 95% confidence ellipse. The ellipse captures the central distribution of data points, revealing a strong linear trend along its major axis. The overall Pearson correlation coefficient is 0.71, indicating an effective correlation between predictions and actual force values. This suggests that our model successfully captures the underlying force-related acoustic features.

Most data points lie within the 95% confidence region, demonstrating the model's consistent performance across a wide range of brushing conditions. The predictions, however, show a broader spread, particularly at the upper and lower extremes. This is expected, as force values in these regions are underrepresented in the training data. For instance, in the high-force range, the model is more error-prone due to data scarcity. At the lower end, some ground-truth force values appear slightly negative, which can occur when the toothbrush is obstructed by the lips during brushing. These negative values are not physically meaningful, and our model correctly predicts most of these segments as near-zero or positive force.

The overall MAE is 37.3g, corresponding to 11.4% of the typical dynamic range (defined as 95% confidence interval of the force distribution). The MAPE is 0.29. Given the complex and highly variable nature of real-world brushing, affected by factors such as bristle stiffness, contact area, brushing angle, and user habits, this performance is acceptable and comparable to the only existing system [28], despite that their system relies on external sensors and does not evaluate real-world brushing behavior. While our approach incorporates self-supervised learning and anatomical disentanglement to suppress the most significant user-specific variation, there is residual variability that cannot be fully eliminated. Additionally, various unmodeled factors, such as brush head shape, oral moisture, and even incidental tongue contact, may further contribute to prediction variability, highlighting the complexity of capturing force-related audio cues in real-world settings. Many commercial smart toothbrushes adopt a fixed threshold of approximately 255g to trigger overpressure warnings [33]. Building on this practice, our system can similarly define a conservative threshold, *e.g.*, 220g, and deliver real-time audio feedback to help users avoid overbrushing even when using standard manual toothbrushes.

Overall, the results demonstrate the practical feasibility of estimating brushing force using in-ear audio signals. Despite the challenges, the model captures meaningful trends and delivers reliable performance for downstream applications such as imbalance detection and brushing habit feedback (shown in Sec. 4.9). These findings highlight the promise of leveraging earable devices for accessible and non-intrusive toothbrushing force monitoring in everyday settings.

4.2 Impact of Users

To better understand how individual differences affect model performance, we visualize the per-user MAE and MAPE in Fig. 12. We can observe that, although there is some variation in performance across users, the differences remain relatively minor, with standard deviations of 5.27g for MAE and 0.08 for MAPE. The best-performing user 10 achieves an MAE of 29.8g and an MAPE of 0.17, while the worst case (user 14) reaches an MAE of 46.0g and an MAPE of 0.45. This disparity persists despite our user-specific variability disentanglement efforts, suggesting that factors beyond physiology, such as brushing style and force habit, also play a critical role.

As shown in Fig. 13, we plot these two users' force distribution compared to the population-level global distribution. Our analysis reveals that individual force usage patterns vary substantially. For example, User 10 exhibits brushing force distributions that closely align with the overall dataset distribution. In contrast, user 14 consistently applies much lighter force, with most values concentrated around 100g, leading to a higher MAPE. Since our heuristic calibration uses a globally defined pseudo-label range centered at 187g, such users receive misaligned calibration targets, which degrade estimation performance. These findings suggest that a one-size-fits-all calibration scheme may not be optimal. This also helps explain why expanding the calibration dataset did not improve and, in some cases, even worsened model performance: increasing data volume under a misaligned distribution may reinforce incorrect learning targets and lead to overfitting on user-inconsistent force patterns. Future work could explore personalized calibration priors by actively clustering users into force-behavior archetypes. Additionally, incorporating uncertainty estimation in predictions may help the system adapt more cautiously when encountering unfamiliar brushing styles and habits.

4.3 Impact of Brushes

To evaluate the model's robustness across different toothbrush types, we tested the system using four distinct brush-head settings: Instrumented/Modified brush-medium (I-med), Instrumented brush-soft (I-soft), OpenBrush-medium (O-med), and OpenBrush-soft (O-soft). As shown in Fig. 14, the performance across brush types is very close, indicating that the model can generalize well to bristle stiffnesses. This suggests that the brush hardness variation is likely captured during the self-supervised pretraining stage, in which diverse brushing sounds allow the encoder to learn brush-invariant features. Additionally, we observe that the MAE and MAPE on our custom-built OpenBrush are slightly higher compared to the instrumented commercial brush. We infer the result is that the handle of OpenBrush is shorter, resulting in slightly altered grip dynamics. When users switch from the commercial Brush to OpenBrush during the data collection, they may unconsciously exert slightly more force due to the leverage difference.

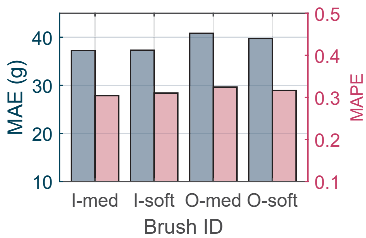


Fig. 14. Impact of different types of brushes.

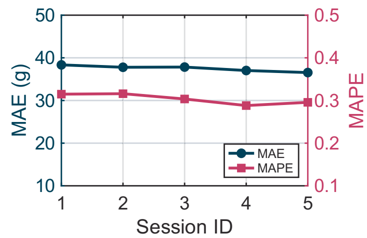


Fig. 15. Impact of different brushing sessions.

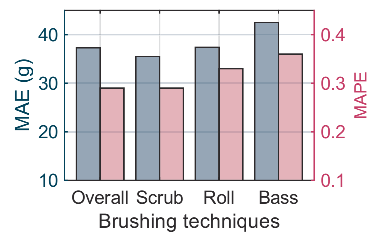


Fig. 16. Impact of different brushing techniques.

4.4 Impact of Sessions

We also investigate whether the model's performance varies across different sessions (days) by evaluating the system on data collected over five consecutive days. As shown in Fig. 15, we observe no significant performance differences between days, with an MAE of 0.71g and MAPE of 0.01. We found that the force distributions remain consistent over time for most users. This consistency is likely due to the fact that many participants already have established brushing habits. These results also suggest that, in the absence of active intervention, brushing style tends to be stable over long timescales. This is encouraging for our system design, as for users with unhealthy or imbalanced force patterns, timely feedback and intervention may be necessary. In Sec. 4.9, we present a case study demonstrating how our system can identify such brushing force imbalances, providing brushing feedback to users.

4.5 Impact of Brushing Techniques

To evaluate the impact of different brushing techniques, we examined the brushing techniques used by participants in our study and observed interesting variations. Among the 16 users, 3 participants (18.8%) adopted a scrub technique (*i.e.*, horizontal brushing), 12 participants (75%) used the roll technique (*i.e.*, vertical brushing), and only 1 participant (6.25%) employed the Bass method. Although the Bass technique is widely recommended by dental professionals due to its effectiveness in removing plaque at the gum line, it remains underused, likely due to its operational complexity. This observation aligns with previous population-level statistics [14]. The scrub method is not generally recommended by dentists, as it tends to introduce aggressive horizontal motions that can cause gum abrasion. However, due to its simplicity, a portion of users still adopt it. The roll technique serves as a middle ground, offering a reasonable trade-off between ease of use and clinical effectiveness, which may explain its widespread use in participants [34].

We analyzed the force estimation performance across different techniques. As shown in Fig. 16, the MAE for scrub, roll, and Bass users were 35.5g, 37.4g, and 42.5g, respectively, with corresponding MAPE values of 0.29, 0.33, and 0.36. We observe that scrub and roll showed comparable estimation performance, but scrub users exhibited slightly lower error, possibly due to the clearer brushing sounds. The Bass technique yielded the lowest performance. This is likely attributed to the inherently lighter force of Bass brushing caused by subtle vibrations around the gumline, leading to larger calibration-induced distribution shifts.

The Bass technique is clinically recommended due to its minimal risk of soft tissue damage and superior effectiveness in removing plaque [45]. However, it is not yet widely adopted in everyday brushing routines, likely due to the complex and subtle motions it requires. Similar to current electric toothbrushes, our system is designed to provide brushing force feedback even under natural brushing behaviors. Currently, only one participant in our experiment used the Bass method. Looking forward, including more Bass-style brushing may improve the system's performance in low-force regimes and broaden its applicability across brushing techniques.

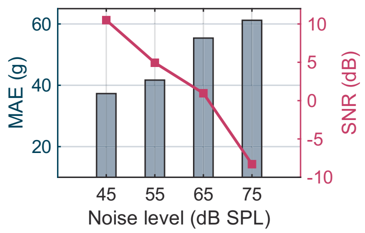


Fig. 17. Impact of different noise levels.

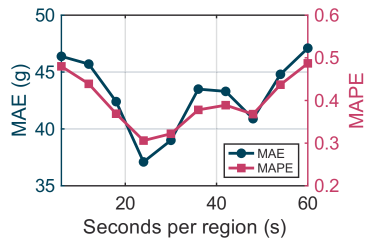


Fig. 18. Impact of different sizes of calibration data.

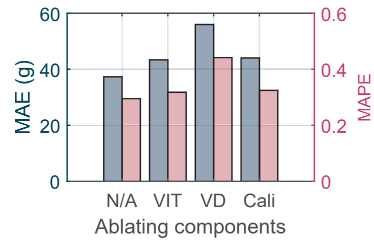


Fig. 19. Ablation study on different system components.

4.6 Impact of Different Noise Levels

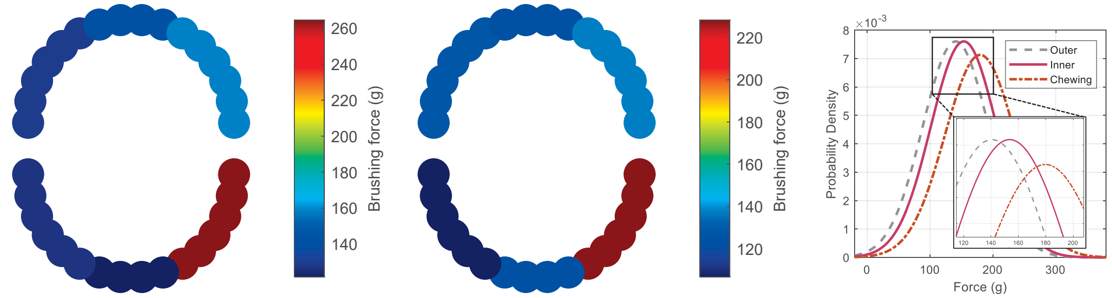
We evaluated the robustness of HearForce under varying noisy environments. We conducted a study to analyze the environmental conditions. The average RT60 (*i.e.*, Reverberation Time) value in the rooms where most brushing sessions took place was 0.24 seconds, indicating a moderate level of reverberation. Ambient noise levels averaged around 45 dB SPL under normal conditions, typically including background sounds from ventilation fans. Thanks to the occlusive seal of in-ear eartips and the dominance of bone-conducted signals during brushing, the recorded in-ear audio retains a relatively high signal quality. Aided by the occlusion effect [32] that naturally amplifies low-frequency bone-conducted sounds, empirical measurements show that brushing signals can reach a Signal-to-Noise-Ratio (SNR) of up to 10.5 dB under these conditions.

To simulate more challenging environments, we conducted controlled tests in which participants brushed their teeth while music and speech audio were played nearby at 45–75 dB SPL. As shown in Fig. 17, this led to a gradual degradation in signal quality, with SNR dropping to as low as −8.3 dB at the highest noise level. We observed that performance remained relatively stable under moderate noise. At 55 dB, the MAE increased only slightly to 41.7g. However, when the noise rose to 65 dB, the SNR dropped sharply to 0.96 dB, and the MAE increased substantially to 55.4g. Accordingly, the MAPE also rose from 29.0% to 61.2%. Further increases in noise level led to continued performance degradation.

These results suggest that HearForce remains robust under typical household noise conditions but begins to break down in environments exceeding 65 dB SPL levels that are rarely encountered during normal brushing activities [18]. It is worth noting that the current experiments were conducted with an earable prototype. We anticipate that commercial-grade earbuds with better acoustic isolation and optimized in-ear microphone design would provide even greater noise resilience.

4.7 Impact of calibration data sizes

To evaluate the impact of different data sizes for heuristic calibration, we varied the calibration data size by adjusting brushing time per region from 6s to 60s. As shown in Fig. 18, we found that MAE initially decreased as sample size increased, but plateaued around 30s per region. Beyond this point, both MAE and MAPE began to rise. This is because user-provided labels are inherently coarse-grained, and excessive data may cause the model to overfit to a misaligned force distribution. Thus, we selected 30s per region (about 3 mins total) as a practical and effective calibration duration. We also experimented with increasing force levels from three to five. However, this led to a noticeable MAE increase, as users found it difficult to reliably distinguish and reproduce finer levels of brushing force. Since our pseudo-labels are derived from a population-level distribution, increasing the number of levels further amplifies inconsistencies, resulting in greater calibration noise.



(a) Real brushing force distribution across regions

(b) Estimated force distribution across regions

(c) Force distribution across different surfaces

Fig. 20. A case study of brushing force monitoring.

4.8 Ablation Study

We conducted an ablation study to evaluate the contribution of three key components in our framework: ViT-based self-supervised pertaining (VIT), variability feature disentanglement (VD), and heuristic calibration (HC) by removing each component individually and measuring the resulting performance. When we first remove the ViT pretraining and instead train the encoder from scratch, the MAE increases from 37.4g to 43.4g. This performance drop indicates that the model benefits from large-scale brushing audio pretraining, which helps capture generalizable acoustic structures and reduces overfitting under limited labeled data. Then, we found that removing the variability feature disentanglement module and using the audio spectrogram embeddings only leads to a much larger increase in error, with MAE rising to 55.9g, similarly to a naive mean predictor, indicating that user-specific noise overwhelms force-related information. Without explicitly isolating and suppressing these factors, the model fails to generalize across users, leading to large prediction errors. We also remove the heuristic calibration step, then the MAE increases to 44.0g. Although the pseudo-labels used for calibration are approximate, they provide helpful alignment between individual brushing forces and the model's internal force representation space. Overall, these results show that all three components contribute meaningfully to final performance.

4.9 Case Study

To explore potential real-world applications of our system, we discussed with our collaborating dentist and conducted a case study that visualizes regional brushing force across different parts of the mouth. Following dental clinical practice, we divide the oral cavity into six sextants, a standard used in periodontal assessment. We adopt an earable toothbrushing location-tracking method [55] to estimate which sextant the user is brushing over time. We then compute the average predicted force per region using our model (Fig. 20(b)), and compare it to the ground-truth force collected from the instrumented toothbrush (Fig. 20(a)).

As shown in Fig. 20, we found that although the predicted and true force values differ numerically, the relative force distribution across regions remains consistent. That is, areas where the user tends to apply imbalanced pressure are accurately reflected by the model's predictions. Specifically, the user tends to apply stronger forces on the side opposite to their dominant hand. For this left-handed user, the right lower sextant shows noticeably higher force, especially when brushing the molar region. In contrast, the left sextant, which is harder to reach with the dominant hand, receives significantly lighter brushing force.

This asymmetry reflects a common behavioral pattern in daily brushing habits and aligns with clinical observations [24]. Our collaborating dentist highlighted this as a promising use case for brushing force prediction. While there is ongoing debate regarding the exact threshold between “safe” and “excessive” force, there is general consensus that imbalanced pressure across different regions can contribute to enamel wear and gum recession over time. By providing users with a region-specific force summary, our system enables personalized feedback. Users in the study also expressed positive feedback, noting that this visualization made them more aware of their brushing patterns and motivated them to adjust balanced brushing force in future sessions.

Beyond region-level analysis, to further understand surface-level brushing force differences, we analyzed the users' average applied force across these areas. As shown in Fig. 20(c), the average force on the outer and inner surfaces was approximately 140.7g and 153.6g, respectively, while the chewing surface received the highest average force at around 179.7g. The similarity between outer and inner surfaces suggests users apply comparable pressure on both sides, though slightly more force is used on the outer surface, potentially due to easier access and visibility. The higher force on chewing surfaces may reflect their structural robustness and users' natural tendency to apply greater pressure during occlusal cleaning. These findings point to the potential for future personalized guidance not only at the regional level, but also tailored to specific surface types.

5 Related Work

5.1 Toothbruhsing Location Tracking

Toothbrushing location tracking has been widely explored using various sensing modalities. Vision and light-based approaches use cameras or optical sensors to monitor brushing regions. For instance, Playful Toothbrush tracks brushing position using an LED extension and a webcam placed in front of the user's mouth [10], while LiT [12] embeds photosensors into LED toothbrushes. IMU-based approaches [8, 20, 31] monitor brushing gestures either by attaching IMUs to the toothbrush, smart watches, or relying on ear-worn devices. Systems like Social Brush [9] and mTeeth [8] achieve fine-grained brushing region recognition using data from the hand or wrist. BrushBuds [54] utilizes the IMU sensor in earbuds to track the brushing locations. Audio-based approaches record brushing sounds with smartphones [25], throat microphones [38], or in-ear microphones from earbuds [42]. In addition to physiological sensing [19, 30, 48], earbuds have recently been explored for toothbrushing monitoring. ToothFairy [52] leverages acoustic attenuation models using earphone-reversed signals for outer-teeth tracking. TomoBrush [57] explored detecting dental health issues with a microphone built into electric toothbrushes. While these works provide extensive coverage tracking, few have explored brushing force estimation, which is equally crucial for oral health. SmarTeeth [55] further utilizes in-ear audio to support 16-region toothbrushing monitoring across both inner and outer surfaces. While we adopt the physiological feature extraction method from SmarTeeth [55], it is used solely as one input to our cross-attention-based variability disentanglement module, which additionally incorporates behavioral features and brushing audio embeddings specifically designed to suppress non-force-related variance. Unlike SmarTeeth, which focused on discrete brushing location classification, our work addresses the more challenging task of continuous force estimation under signal ambiguity, anatomical variability, and behavioral diversity, which was not addressed in prior work.

5.2 Toothbrushing Force Estimation

While toothbrushing location tracking for manual toothbrushes has been extensively studied, brushing force estimation has received surprisingly little attention. Smart electric toothbrushes like Oral-B iO Series [3] equipped with pressure sensors have become increasingly common, providing real-time feedback to help users use appropriate force levels. For example, when the applied force exceeds 2.5N, the brush lights up red to warn users of excessive pressure [33]. However, these devices are much more expensive (~550 USD) than manual toothbrushes (~2 USD), which limits their adoption. Moreover, research shows that brushing force tolerance varies significantly across individuals [26]. For example, people with sensitive teeth or bleeding gums require much gentler pressure, making fixed thresholds insufficient for personalized oral care. This observation was also supported by our collaborating dentist. In fact, manual toothbrushes still dominate the global market due to their affordability and ease of use [47]. To compensate for the lack of force sensing in manual brushing, prior research has explored augmenting toothbrushes with external sensors. Some studies [5, 6] embed force sensors directly beneath the bristles. A recent system [27, 28] used five force sensors on a fixed toothbrush to indirectly estimate the force applied to the brush head. However, this set-up required the user to maintain a predefined grip and did not test the system in actual brushing scenarios; instead, it relies on simulated brushing against a weight scale. Moreover, such approaches require custom hardware modifications on manual toothbrushes, making them impractical for wide-scale deployment. In contrast, our work leverages the in-ear microphones already embedded in modern earables, proposing a solution that requires no modification to the manual toothbrush or the user's grip and enables continuous brushing force monitoring that is accessible to everyone, seamlessly integrating into users' normal daily routines.

6 Limitation and Discussion

6.1 Dataset size and System Performance

Although our dataset covers a realistic range of users and brushing scenarios, its scale could still be expanded given the high variability in individual brushing styles and force habits. A more diverse cohort spanning different age groups, oral conditions, and cultural brushing habits would be necessary to fully capture population-wide variability. In practice, brushing audio is affected by numerous factors such as tooth alignment, enamel condition, toothpaste, brushing posture, and even daily mood, making data-driven modeling particularly sensitive to sample diversity. Despite our design of self-supervised pretraining, variability disentanglement, and calibration to help mitigate data scarcity, residual variability remains. Our experimental results in Sec. 4.2 reveal substantial differences in brushing force distributions across users, even under similar experimental protocols. This variability highlights the intrinsic challenges in generalizing force estimation models and suggests that achieving higher accuracy and robustness will likely require the collection of a significantly larger and more diverse dataset. We envision that after open-sourcing our hardware, broader participation can be enabled. Once the commercial earbuds provide the data access, users can contribute unlabeled brushing audio collected via earable devices, which remains extremely valuable for self-supervised pretraining. Expanding the dataset in this way will be key to improving generalization and performance. Additionally, the pressure sensors themselves are subject to a certain degree of measurement error. Despite this, our system successfully validates the existence and significance of the relationship between in-ear audio and brushing force. This lays a foundational step for future work to build upon, including larger-scale deployment, personalization, and clinical validation.

6.2 Brushing Force Interpretation

One of the core challenges in evaluating brushing force estimation lies in the lack of universally accepted clinical standards. What constitutes "light," "normal," or "strong" brushing varies widely between individuals, and no clinically endorsed thresholds exist that apply universally. As a result, it is difficult to define a precise target or evaluate the model using conventional accuracy metrics. Our goal is to initiate this line of research, as no existing clinical tools currently provide such force estimation. Like most existing electric toothbrushes, our current system focuses on estimating brushing force uniformly across all regions. However, the clinical relevance of force is highly context-dependent, varying with brushing location, method, and the target being cleaned (*e.g.*, enamel *v.s.* gumline). As existing work [52, 55] has demonstrated the feasibility of using earables to detect brushing location and surfaces, our system can augment these efforts by adding force-awareness as a complementary sensing dimension to deliver fine-grained, context-aware oral health guidance. We hope that our system can serve as a first step toward helping establish clinical measurement methods and guidelines, and raise broader awareness about the need for personalized brushing force monitoring.

6.3 Device Stability and Usage Practicality

The stability of earable devices during toothbrushing is an important consideration due to potential jaw and head movements. In our study, we used the OpenEarable 1.4 platform, which features foam ear tips with strong rebound elasticity and an over-ear hook design for improved mechanical stability (Fig. 10(b)). While some participants with smaller ears reported mild discomfort during initial fitting, the device remained securely worn throughout brushing sessions. However, larger head movements could lead to minor shifts, but normal brushing posture generally keeps motion within a limited range. Modern commercial earbuds are typically ergonomically designed and provide interchangeable ear tip sizes, suggesting that proper fit would further improve sensing reliability in real-world deployment. Our results demonstrate the feasibility of force sensing using the in-ear microphone. Currently, most commercial earbuds (*e.g.*, AirPods) already include inward-facing microphones for built-in noise cancellation. Although access to this microphone is not currently exposed to developers, once access is available, our approach can be directly applied without additional hardware. We are actively engaging with industry partners to advocate for broader microphone access. Brushing the gums and tongue gently is also commonly recommended during toothbrushing. We tested these scenarios and found that in-ear signals remained detectable despite slightly lower, suggesting potential for future target-specific brushing monitoring.

7 Conclusion

In this paper, we presented HearForce, the first system that estimates continuous manual toothbrushing force using only off-the-shelf earable devices, without requiring any modification to the toothbrush. By leveraging in-ear microphones and the bone-conducted sound produced during brushing, HearForce enables a low-cost and accessible solution for manual toothbrushing force monitoring in daily life. Looking forward, we aim to expand the scale and diversity of our dataset through open-source release, and collaborate with dental researchers to define clinical ground-truth standards for brushing force interpretation. We believe HearForce offers a promising foundation for developing personalized, clinically meaningful oral health feedback, and opens up new opportunities in earable health sensing.

Acknowledgments

This research was supported by the ERC project 833296, EPSRC grants EP/Y035925/1 and EP/Z53447X/1, and the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (T2EP20124-0046).

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. 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).
[3] 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).
[4] ADA. 2025. Brushing Your Teeth — mouthhealthy.org. https://www.mouthhealthy.org/all-topics-a-z/brushing-your-teeth. [Accessed 27-04-2025].
[5] Md Akhtaruzzaman. 2019. Prototype of a force-sensitive smart toothbrush. In *Proc. EICT, 4th Int. Conf.*
[6] Md Akhtaruzzaman. 2021. Force-Sensitive Classic Toothbrush: System Analysis, Design, and Simulation. *Electrica* 21, 2 (2021).
[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] Chris Chatfield and Haipeng Xing. 2019. *The analysis of time series: an introduction with R*. Chapman and hall/CRC.

[12] 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.

[13] Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*. 1459–1462.

[14] Carolina Ganss, Nadine Schlueter, Susanne Preiss, and Joachim Klimek. 2009. Tooth brushing habits in uninstructed adults—frequency, technique, duration and force. *Clinical oral investigations* 13, 2 (2009), 203–208.

[15] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. 2022. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 16000–16009.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[17] Maria A Heckl and ID Abrahams. 1996. Active control of friction-driven oscillations. *Journal of Sound and Vibration* 193, 1 (1996), 417–426.

[18] Marie-Eve Héroux, Wolfgang Babisch, Goran Belojevic, Mark Brink, Sabine Janssen, Peter Lercher, Marco Paviotti, Göran Pershagen, Kerstin Persson Waye, Anna Preis, et al. 2015. WHO environmental noise guidelines for the European Region. In *Euronoise*, Vol. 2015. 2589–2593.

[19] Changshuo Hu, Qiang Yang, Yang Liu, Tobias Röddiger, Kayla-Jade Butkow, Mathias Ciliberto, Adam Luke Pullin, Jake Stuchbury-Wass, Mahbub Hassan, Cecilia Mascolo, et al. 2025. A Survey of Earable Technology: Trends, Tools, and the Road Ahead. *arXiv preprint arXiv:2506.05720* (2025).

[20] 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.

[21] Po-Yao Huang, Hu Xu, Juncheng Li, Alexei Baevski, Michael Auli, Wojciech Galuba, Florian Metze, and Christoph Feichtenhofer. 2022. Masked autoencoders that listen. *Advances in Neural Information Processing Systems* 35 (2022), 28708–28720.

[22] Peter J Huber. 1992. Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution*. Springer, 492–518.

[23] Karen Janusz, Bruce Nelson, Robert D Bartizek, Patricia A Walters, and Aaron R Biesbrock. 2008. Impact of a novel power toothbrush with SmartGuide technology on brushing pressure and thoroughness. *J Contemp Dent Pract* 9, 7 (2008), 1–8.

[24] Mahdi Kadkhodazadeh, Amin Khodadustan, Reza Amid, and Ayoob Darabi. 2012. Plaque removal ability in left-and right-handed patients in different parts of the oral cavity. *Journal of Advanced Periodontology & Implant Dentistry* 4, 1 (2012), 24–28.

[25] 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.

[26] Santhosh Kumar, Pratibha Gopalkrishna, Ayman K Syed, and Abishikka Sathiyabalan. 2025. The Impact of Toothbrushing on Oral Health, Gingival Recession, and Tooth Wear—A Narrative Review. In *Healthcare*, Vol. 13. MDPI, 1138.

[27] Haicui Li and Lei Jing. 2022. An indirect method of brushing force detection with five force sensors and RF algorithm. In *2022 IEEE Sensors*. IEEE, 1–4.

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

[29] Xiao Liu, Fanjin Zhang, Zhenyu Hou, Li Mian, Zhaoyu Wang, Jing Zhang, and Jie Tang. 2021. Self-supervised learning: Generative or contrastive. *IEEE transactions on knowledge and data engineering* 35, 1 (2021), 857–876.

[30] Yang Liu, Kayla-Jade Butkow, Jake Stuchbury-Wass, Adam Pullin, Dong Ma, and Cecilia Mascolo. 2025. Respear: Earable-based robust respiratory rate monitoring. In *2025 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 67–77.

[31] 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.

[32] 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.

[33] Vanshika Maini, Rupanjan Roy, Gargi Gandhi, Aditi Chopra, and Subraya G Bhat. 2025. Artificial-Intelligence-Based Smart Toothbrushes for Oral Health and Patient Education: A Review. *Hygiene* 5, 1 (2025), 5.

[34] Stefano Mastroberardino, Maria Grazia Cagetti, Fabio Cocco, Guglielmo Campus, Jessica Pizzocri, and Laura Strohmenger. 2014. Vertical brushing versus horizontal brushing: a randomized split-mouth clinical trial. *Quintessence International* 45, 8 (2014).

[35] GI McCracken, J Janssen, M Swan, N Steen, M De Jager, and PA Heasman. 2003. Effect of brushing force and time on plaque removal using a powered toothbrush. *Journal of clinical periodontology* 30, 5 (2003), 409–413.

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

[37] 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).

[38] 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.

[39] Claire Palvetzian. 2023. *Dental & Wellness Office Administration 2nd ed.* Conestoga College. https://doi.org/10.58067/JG0G-VX79

[40] 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.

[41] Raspberry Pi. 2023. Raspberry pi pico. *Obtenido de https://www. raspberrypi. com/products/raspberry-pi-pico* (2023).

[42] 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.

[43] Patr Pujarern, Pirasut Rodanant, Eakkachai Warinsiriruk, and Kanyawat Rattanasuwan. 2021. Evaluation of the optimum brushing force on dental plaque removal: an in vitro study. *Mahidol Dental Journal* 41, 1 (2021), 75–82.

[44] Lawrence Rabiner and Ronald Schafer. 2010. *Theory and applications of digital speech processing*. Prentice Hall Press.

[45] Andrea Rani Rajwani, Sophia Nancy Diana Hawes, Amanda To, Alessandro Quaranta, and Julio C Rincon Aguilar. 2020. Effectiveness of manual toothbrushing techniques on plaque and gingivitis: a systematic review. *Oral health & preventive dentistry* 18, 4 (2020), a45354.

[46] Tobias Röddiger, Jake Stuchbury-Wass, Mathias Ciliberto, Philipp Lepold, and Michael Beigl. 2024. OpenEarable 1.4: Dual Microphones Earpiece to Capture In-Ear and Outer-Ear Audio Signals. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 930–933.

[47] Statista. 2024. U.S.: usage of manual toothbrushes 2011-2024 | Statista. https://www.statista.com/statistics/287357/usage-of-manual-toothbrushes-in-the-us-trend/. (Accessed on 04/03/2024).

[48] Jake Stuchbury-Wass, Yang Liu, Kayla-Jade Butkow, Josh Carter, Qiang Yang, Mathias Ciliberto, Ezio Preatoni, Dong Ma, and Cecilia Mascolo. 2025. WalkEar: holistic gait monitoring using earables. In *2025 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 99–109.

[49] Tooth and Go. 2025. Tooth Abrasion: Causes, Symptoms, and Effective Treatments — dentist-manila.com. https://www.dentist-manila.com/glossary-abrasion/. [Accessed 27-04-2025].

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

[51] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[52] 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.

[53] Annette Wiegand, John Patrik Matthias Burkhard, Florin Eggmann, and Thomas Attin. 2013. Brushing force of manual and sonic toothbrushes affects dental hard tissue abrasion. *Clinical oral investigations* 17 (2013), 815–822.

[54] 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.

[55] Qiang Yang, Yang Liu, Jake Stuchbury-Wass, Kayla-Jade Butkow, and Cecilia Mascolo. 2025. SmarTeeth: Augmenting Manual Toothbrushing with In-ear Microphones. (2025).

[56] Qiang Yang, Jake Stuchbury-Wass, Thomas Bytheway, and Cecilia Mascolo. 2025. OpenBrush: An Open-Source Manual Toothbrush for Force and Motion Sensing. (2025).

[57] 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.

[58] Yuwei Zhang, Tong Xia, Jing Han, Yu Wu, Georgios Rizos, Yang Liu, Mohammed Mosuily, J Ch, and Cecilia Mascolo. 2024. Towards open respiratory acoustic foundation models: Pretraining and benchmarking. *Advances in Neural Information Processing Systems* 37 (2024), 27024–27055.

[59] Zixin Zheng, Yumeng Liang, Rui Lyu, Junjie Bao, Yiwen Huang, Anfu Zhou, Huadong Ma, Jingjia Wang, Xiangbin Meng, Chunli Shao, et al. 2024. BP3: Improving Cuff-less Blood Pressure Monitoring Performance by Fusing mmWave Pulse Wave Sensing and Physiological Factors. In *Proceedings of the 22nd ACM Conference on Embedded Networked Sensor Systems*. 730–743.