ToMoBrush: Exploring Dental Health Sensing Using a Sonic Toothbrush

KUANG YUAN, Carnegie Mellon University, USA
MOHAMED IBRAHIM, Carnegie Mellon University, USA and Hewlett Packard Labs, USA
YIWEN SONG, Carnegie Mellon University, USA
GUOXIANG DENG, Carnegie Mellon University, USA
ROBERT A. NERONE, School of Dental Medicine, University of Pittsburgh, USA
SUVENDRA VIJAYAN, School of Dental Medicine, University of Pittsburgh, USA
AKSHAY GADRE, University of Washington, USA
SWARUN KUMAR, Carnegie Mellon University, USA

Early detection of dental disease is crucial to prevent adverse outcomes. Today, dental X-rays are currently the most accurate gold standard for dental disease detection. Unfortunately, regular X-ray exam is still a privilege for billions of people around the world. In this paper, we ask: "Can we develop a low-cost sensing system that enables dental self-examination in the comfort of one's home?"

This paper presents *ToMoBrush*, a dental health sensing system that explores using off-the-shelf sonic toothbrushes for dental condition detection. Our solution leverages the fact that a sonic toothbrush produces rich acoustic signals when in contact with teeth, which contain important information about each tooth's status. *ToMoBrush* extracts tooth resonance signatures from the acoustic signals to characterize the dental condition of each tooth. We further develop a data-driven signal processing pipeline to detect and discriminate different dental conditions. We evaluate *ToMoBrush* on 19 participants and dental-standard models for detecting common dental problems including caries, calculus, and food impaction, achieving a detection ROC-AUC of 0.90, 0.83, and 0.88 respectively. Interviews with dental experts further validate *ToMoBrush*'s potential in enhancing at-home dental healthcare.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing**; • **Applied computing** → *Health care information systems.*

Additional Key Words and Phrases: Mobile Health, Audio Signal Processing, Machine Learning, User Study

ACM Reference Format:
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. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3, Article 139 (September 2024), 27 pages. https://doi.org/10.1145/3678505

Authors' addresses: Kuang Yuan, Carnegie Mellon University, Pittsburgh, USA, kuangy@cmu.edu; Mohamed Ibrahim, Carnegie Mellon University, Pittsburgh, USA and Hewlett Packard Labs, Berkeley Heights, USA, ibrahim@hpe.com; Yiwen Song, Carnegie Mellon University, Pittsburgh, USA, yiwens2@andrew.cmu.edu; Guoxiang Deng, Carnegie Mellon University, Pittsburgh, USA, gdeng990101@163.com; Robert A. Nerone, School of Dental Medicine, University of Pittsburgh, Pittsburgh, USA, ron24@pitt.edu; Suvendra Vijayan, School of Dental Medicine, University of Pittsburgh, Pittsburgh, USA, suv16@pitt.edu; Akshay Gadre, University of Washington, Seattle, USA, gadre@uw.edu; Swarun Kumar, Carnegie Mellon University, Pittsburgh, USA, swarun@cmu.edu.

2474-9567/2024/9-ART139
https://doi.org/10.1145/3678505

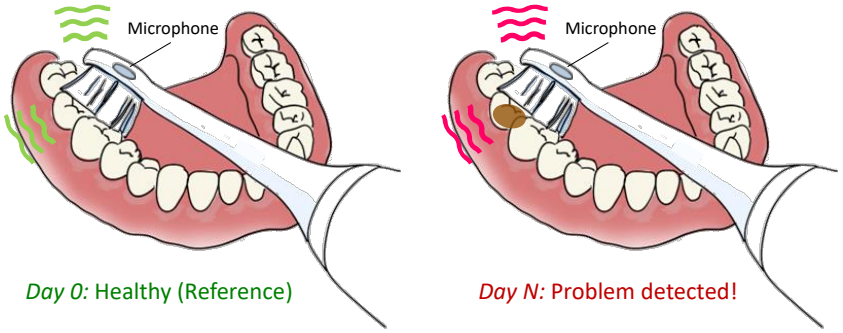


Fig. 1. *ToMoBrush* is a dental health sensing tool composed of an off-the-shelf sonic toothbrush with a microphone integrated close to the brush head. *ToMoBrush* analyzes the changes in the resonance behavior of teeth to enable dental health sensing.

1 INTRODUCTION

Dental disease is a major public health challenge, that can cause pain and infections which may lead to problems with eating, speaking, and even social interaction [10]. Early detection of dental disease is crucial, since it enables interventions that can halt disease progression and prevent the onset of adverse outcomes [22]. Due to the lack of awareness and availability of dental health monitoring, over $45 billion in US productivity is lost each year due to untreated dental disease [10]. Today, dental X-rays in dental clinics are the most accurate gold standard for dental disease detection [67]. Unfortunately, access to dental X-ray diagnostics at sufficient regularity is still a privilege for billions of people around the world, due to limited availability, high cost, and lack of awareness. In this paper, we explore a low-cost solution for dental health monitoring that users can use regularly in the comfort of their homes. We believe such a system can supplement the dental healthcare system even for those with access to professional dental care, by providing early warnings of potential issues in between the dental visits.

While there is rich prior work on dental health sensing using X-rays, advanced cameras or other infrastructure [31, 57], these are designed for clinical settings and are expensive. The few systems designed for at-home dental sensing largely focus on toothbrush localization to study brushing habits [11, 12, 32, 33, 38, 42]. Perhaps the closest work targeting low-cost dental health sensing is one prior system that uses smartphone cameras for dental self-exams [41]. However, this solution can only detect visually diagnosable diseases such as surface cavities and struggles with low-light conditions inside the mouth.

This paper presents *ToMoBrush*[1], which explores the potential of using a commodity electric sonic toothbrush with a minimum hardware modification for dental health sensing to enable regular, at-home dental self-examination. Rather than viewing a toothbrush purely as a cleaning instrument, *ToMoBrush* leverages the fact that an electric toothbrush is also a rich source of acoustic signals, that are generated by rapid automatic bristle vibrations. When the brush is in contact with a tooth, the tooth also vibrates along with the toothbrush and produces distinct acoustic signals depending on the tooth's state.

Fig. 1 visualizes the idea of *ToMoBrush*. An external microphone is integrated close to the brush head of an off-the-shelf sonic toothbrush, which captures the acoustic signals when the toothbrush is on. Whenever the user wants to perform an at-home dental self-exam, *ToMoBrush* instructs the user to brush their teeth one at a time along a pre-specified video-guided pattern. We note that this video-guided brushing procedure is performed in addition to regular oral hygiene and does not involve toothpaste. During the brushing procedure, the microphone on *ToMoBrush* captures the acoustic signals from each tooth of the user to detect changes in the tooth's status. These signals are compared with reference signals collected from the patient's healthy teeth (post any treatment)

[1]Tooth Monitoring Brush / Tomorrow's Toothbrush

at or shortly after the patient's most recent dental visit. Once a significant deviation is detected compared with the reference signals, *ToMoBrush* alerts the user about the potential disease with the tooth location, which enables an early follow-up with a dentist. We collaborated with dental experts to identify three common dental conditions that *ToMoBrush* can potentially detect at home and provides support to prevent more severe diseases, including dental caries (cavities from tooth decay), calculus (hardened plaque), and interdental food impaction (food between teeth). We perform an IRB-approved user study in a dental clinical center and in a home setting on 19 participants. *ToMoBrush* achieves a 0.87 average area under receiver operating characteristic curve (ROC-AUC) value for detecting these three dental conditions. We note that for caries and calculus detection, *ToMoBrush*'s accuracy is comparable to state-of-the-art camera-based systems [41].

ToMoBrush's approach relies on a key observation: Though every sonic toothbrush is designed to vibrate at some specific frequency (usually in 200-400 Hz) for the purpose of teeth cleaning, it also generates harmonics in higher frequencies up to around 20 kHz, due to the imperfections in the vibration motor. As a result, when the toothbrush is in contact with the teeth, it essentially serves as a wideband excitation signal source. The vibration from the toothbrush head traverses and reverberates across the teeth and bones, resulting in a frequency-dependent resonant response, which is unique to the shape, topology, as well as material of the teeth. Thus, when teeth experience health concerns, such as caries, calculus, or even food impaction, this acoustic resonance behavior changes accordingly. *ToMoBrush* seeks to analyze the difference in the resonance behavior to enable dental health sensing. To achieve this, *ToMoBrush* addresses two major challenges:

Extracting and Analyzing a Tooth Resonance Signature: *ToMoBrush* seeks to model the resonance behavior of the teeth when in contact with the excitation signal from a toothbrush. Indeed, in classical vibrometry, where one models the resonant behavior of objects (e.g. rail lines [20], rotten fruit [40, 55], cracked objects [7]) by exciting them with acoustic signals, one has a firm knowledge of the precise acoustic signal applied to the object. However, this is not the case for our system – we have no ability to control the vibration of an off-the-shelf electric toothbrush. Even worse, the toothbrush's vibration behavior is not stable during the brushing procedure. In contrast to static vibrations, brushing the teeth can also change the vibration pattern of the toothbrush. Furthermore, different strength of brushing can also induces different changes. This rules out a one-time calibration where one records the acoustic signal from the toothbrush before any brushing occurs.

To tackle this challenge, we carefully model the vibration system including toothbrush, tooth resonance, as well as brushing strength and movement. We propose an algorithm to separate these different factors and extract clean tooth resonance signatures based on a key observation: Though these factors share the same frequency band, their rates of change across the frequencies are different. In Sec. 5, we discuss how we adapt a technique that is widely used in speech processing to separate the glottal excitation and vocal tract resonances. Specifically, we convert the signal into the cepstrum domain where these distinct behaviors are easily separable (further detailed in Sec. 5). After obtaining the tooth resonance signature, we further develop a feature selection algorithm to select specific regions of signature that are specialized for detecting three different dental conditions (Sec. 6), and perform health detection by comparing the signatures with prior healthy reference measurements (Sec. 7).

Mitigating Noise Factors and Tooth Matching Errors during Brushing: *ToMoBrush* further addresses various sources of error that can impede its performance, such as ambient noise or user-specific brushing behavior. First, during the brushing procedure in real-world scenarios, the microphone on *ToMoBrush* captures not only the acoustic signal we desire from the vibrating tooth, but also the sounds directly from the toothbrush, as well as various ambient noise, which introduces significant error for the signature extraction algorithm. To address this, *ToMoBrush* applies Empirical Mode Decomposition (EMD) for noise suppression to mitigate environmental noise and the direct signal from the toothbrush (Sec. 8.1).

A second potential source of error occurs when users are instructed by *ToMoBrush* to brush their teeth in a pre-specified order, guided by a video, to collect signatures from individual teeth. However, given the low-light

conditions at some blind spots inside the mouth (such as inner upper molars), occasional matching errors can occur when a user collects measurements in these regions with poor visibility. For instance, a user may inadvertently miss brushing a tooth located in a blind spot. To limit the impact of matching errors in blind spots, *ToMoBrush* developed a sequence alignment algorithm based on Dynamic Time Warping (DTW) to align brushing sequences of the user's tooth with the references from a same quadrant of teeth, which allows to obtain signatures from each corresponding tooth more accurately (Sec. 8.2).

We implement the prototype of *ToMoBrush* using a Philips Sonicare ProtectiveClean 6100 electric toothbrush [8] with a Voice Technologies VT500X waterproof microphone [6]. We perform the experiments on dental-standard teeth models[2], as well as 19 users in a dental clinical center and in a home setting (IRB-approved). Our results show a health detection performance with ROC-AUC values of 0.90, 0.83, and 0.88 for detecting caries, calculus, and food impaction respectively. We further conduct a 21-day longitude study to demonstrate the performance repeatability over time and in different environments. Besides, an expert evaluation by interviewing two expert dentists (Sec. 9.2) suggests our data collection procedure is valid and system accuracy is promising, while also pointing out both the potential benefits of integrating *ToMoBrush* into the current dental healthcare system as well as potential barriers.

Contributions: Our main contributions include:

- The prototype of the *ToMoBrush* that uses an off-the-shelf sonic toothbrush for dental health sensing.
- A signal processing pipeline to extract tooth resonance signatures from raw audio recordings, and perform health detection to identify dental conditions, as well as mitigate real-world noise factors.
- Comprehensive experiments on dental models and user studies, and an expert evaluation to validate the performance of our system.

2 BACKGROUND OF DENTAL HEALTH PROBLEMS

This paper targets the early detection of dental conditions as they may evolve over time. Theoretically, *ToMoBrush* can potentially detect any dental health problem that induces significant acoustic changes to the resonant behavior of the teeth. We have collaborated with two dental experts to identify the following three dental conditions with high prevalence, and early at-home detection of these conditions can provide valuable support in preventing more severe dental diseases.

Caries: Tooth decay when bacteria in the mouth produce acids that demineralize the hard tissues of the teeth. Dental caries [48], also called cavities, occur when teeth get permanently damaged causing pain and infection. Such damage develops into small holes that enlarge over time and may result eventually in tooth extraction. Tooth decays have a high prevalence [5] (90%) among adults, which often occur in the back teeth. Moreover, they can form between teeth [19] or inside a tooth [60], complicating visual detection. Demineralized tissues absorb X-rays differently than hard tissues, allowing them to be identified in radiographs [1]. This paper aims to explore a solution for at-home early detection of such decay to mitigate more serious damage to the tooth.

Dental Calculus (Tartar): A step toward cavities and other dental diseases is the build-up of bacteria on teeth which starts with soft deposits resulting in calculus at the end. Dental calculus [24] is a hardened bacterial plaque forming below and above the gum line. Notably, the subgingival calculus that forms beneath the gumline is not immediately visible. If left untreated, dental calculus can irritate the gums and lead to gum disease over time.

Interdental Food Impaction: Food impaction is a very common cause of gingival and periodontal disease [37]. Removing the remains of food, especially those lodged between two neighboring teeth, is important to avoid the formation of plaque and eventually cavities. Food impaction frequently happens in hard-to-see areas. *ToMoBrush* can potentially serve as a tool to detect and localize the teeth with food impaction. By identifying these specific

[2]For ethical reasons, we choose to perform more controlled experiments on dental models

areas, brushing and cleaning can be optimized, allowing for targeted flossing at the sites of food impaction rather than relying on random flossing [53].

3 RELATED WORK

Dental Condition Monitoring: Currently, dental health monitoring is mainly available through infrequent dentist visits by using professional examination devices such as X-ray. There has been much clinical work in dentistry using pH-based and fluorescence-based tooth-decay sensing [13, 31, 49–51, 57, 71, 74]. Some recent work has explored incorporating sensing technologies into the regular toothbrushing procedure. However, most of these related systems focus on monitoring hygiene behaviors by detecting which part of the tooth a patient is currently brushing, such as using magnetic field sensing [33], sound being produced [38, 47], camera [17], and wrist-worn motion sensors [11, 12, 32, 42]. Besides, LumiO [73] integrates a blue-violet light intraoral camera into the toothbrush head to capture the intraoral images and monitor the progress of plaque removal, which does not focus on dental disease detection. More recently, OralCam [41] has proposed to use a smartphone camera for visible dental disease detection. However, OralCam can only detect visually diagnosable diseases and struggles in low-light conditions inside the mouth. To conclude, dental health monitoring using ubiquitous devices is an underexplored area, and *ToMoBrush* has the unique advantage of identifying invisible diseases such as inner cavity compared with existing camera-based solutions.

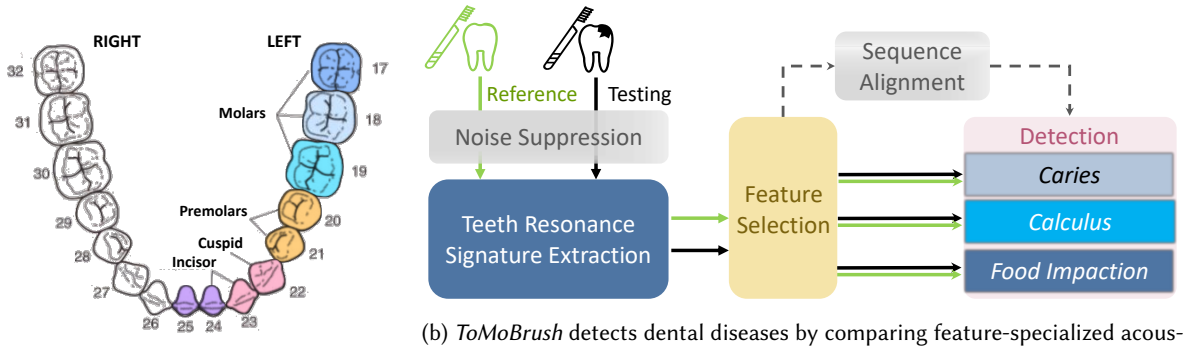
Acoustic Sensing on Humans: Acoustic sensing has been widely explored recently for gesture recognition [15, 46, 61] as well as health monitoring [23, 44, 45, 59, 64–66, 72, 75]. Specifically, acoustic signals have been utilized for monitoring heart rate [23, 65, 75], lung function[59], sleep apnea [45], and respiratory rate [64, 66, 72]. The majority of this line of work counts on detecting movements of the human skin surface for monitoring health conditions. On the other hand, *ToMoBrush* leverages the tooth resonance behavior caused by the electric toothbrush, for extracting unique signatures for monitoring the health status of the teeth.

Sensing with Acoustic Resonance: Measuring the acoustic resonance of an object can reflect its integrity and structure. Therefore, past work has explored acoustic vibrometry for detecting defects including cracks in cups [25], gear tooth breakage [58], railroad defects [28], and fruit ripeness [26, 55]. While this prior work has explored a wide set of applications using acoustic vibrometry, it detects mainly big objects and uses a dedicated transmitter as the excitation source. Compared to prior work, *ToMoBrush* uses an off-the-shelf sonic toothbrush as the excitation source, which is not precisely controllable, and can detect the defects on human teeth.

4 OVERVIEW

ToMoBrush is a dental health sensing system based on an off-the-shelf sonic toothbrush. *ToMoBrush* begins by collecting a set of reference signatures from the user's teeth when they brush for the first time with the system. Note that we require that new users visit a dentist and assume all teeth are in a relatively healthy status post-visit. After collecting the reference signatures, whenever the user wants to perform a self-examination, they brush their teeth to collect the signature again for testing. *ToMoBrush* compares this test signature with the reference signatures, and notifies the user if it detects potential dental problems. We further use the signature from different measurements across days to boost the detection accuracy, which enables a more accurate understanding of one's dental health status through long-term monitoring.

Data Collection: *ToMoBrush* instructs the user to perform the measurement through a video-guided pattern. The video guides the user to brush their teeth through the order of four teeth quadrants: upper right, upper left, lower right, and lower left. For the brushing in each quadrant, the video guides the user to brush the teeth one or two at a time, from the innermost molar to the outermost incisor, and stay on the chewing surface of each tooth for a few seconds. Fig. 2a shows the structure of the lower teeth, and the colors illustrate the brushing



(a) Chart of lower teeth of human

(b) *ToMoBrush* detects dental diseases by comparing feature-specialized acoustic signatures with pre-collected healthy references.

Fig. 2. System overview of *ToMoBrush*. (a) showcases the brushing segments of the lower-left quadrant in different colors. (b) shows the signal processing pipeline.

segments for the lower-left quadrant as an example. The user starts brushing from three molars (from 17 to 19) and brushes one at a time. For the premolars, cuspid, and incisors, as the size of the tooth is generally smaller than the regular toothbrush head, the video guides the user to brush two of the teeth together at a time (20&21, 22&23, and 24&25). *ToMoBrush* captures an audio clip for each single tooth/two teeth brushing and performs signal processing to extract the individual tooth signature. We also use the whole scanning sequence of each tooth quadrant to perform sequence alignment to mitigate the matching errors. To ensure the validity and consistency of the toothbrush head during long-term usage, the user is required to perform testing by collecting signatures on a reference tooth model before self-examination.

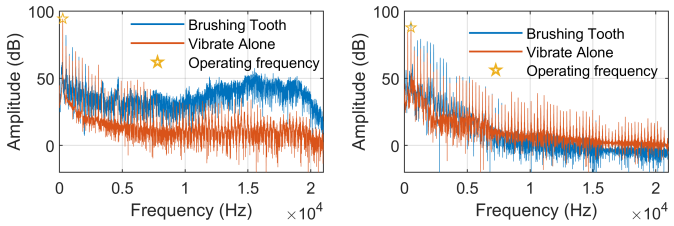
System Overview: Fig. 2b presents the processing pipeline of *ToMoBrush*. *ToMoBrush*'s primary contribution is it's health detection pipeline (Sec. 5-7) that consists of three key steps: (1) *Tooth Resonance Signature Extraction:* We slice the audio recording from the microphone into chunks of 50 milliseconds, for each of which we extract a resonance signature of the tooth that was vibrated over that specific time span. (2) *Feature Selection:* We then process this signature to further extract unique features that assist in identifying different dental conditions. (3) *Health Detection:* To detect each of the diseases, we aggregate and average the features from the same tooth among the 50-millisecond signal chunks, and finally compare them with the reference features captured from the corresponding tooth.

We further note two important processing steps for improving system resilience to errors and noise (Sec. 8). (1) *Noise suppression:* For every raw audio recording captured by the microphone on *ToMoBrush*, we first perform noise suppression to make the signal focused on the vibration behavior of the teeth and toothbrush. (2) *Sequence Alignment:* Before aggregating the features and performing health detection, *ToMoBrush* uses the features from each chunk to perform a time-series alignment algorithm to find the best match of each signature to the corresponding tooth, which mitigates the matching errors for brushing in blind spots.

In the following sections, we first discuss how *ToMoBrush* model the vibration behaviors during toothbrushing and extract tooth resonances signature (Sec. 5), then elaborate how we apply the feature selection (Sec. 6) and health detection (Sec. 7) to detect different dental conditions. We also detail the noise suppression sequence and alignment algorithms in Sec. 8.

5 TOOTH SIGNATURE EXTRACTION

In this section, we discuss how we extract tooth resonance signatures from raw audio recordings captured by the microphone during toothbrushing. *ToMoBrush* first applies Short-Term Fourier Transform (STFT) with 50



(a) Microphone signal (b) Vibration sensor signal

Fig. 3. Signal spectra when a sonic toothbrush is vibrating alone vs. when in contact with a tooth. (a) shows brushing induces the resonances of the tooth. (b) shows brushing also changes the excitation signal itself.

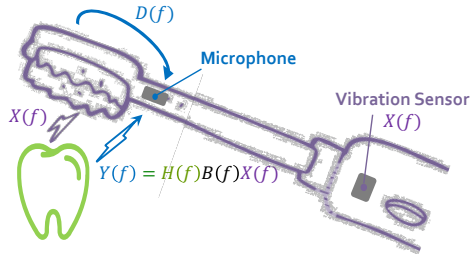


Fig. 4. The signal received by the microphone composed of the excitation signal directly from the toothbrush head as well as the tooth response

milliseconds as time window. Following this, we explain how our signature extraction extracts the resonance signature from the tooth (or two teeth) brushed upon over a single 50-millisecond time window.

5.1 Vibration Behavior Modeling

Sonic toothbrushes are a subset of electric toothbrushes with movement that is fast enough to produce vibration in the audible range. Typically, an off-the-shelf sonic toothbrush is designed to vibrate at a specific frequency in the range of 200-400 Hz. It also produces harmonics in higher frequencies which are integer multiples of the operating fundamental frequency. The red line in Fig. 3a shows the spectrum of the acoustic signal produced by the Philips Sonicare ProtectiveClean 6100 when it is turned on. While it is designed to operate at 260 Hz (see the yellow star), it also generates resonant tones at higher frequencies up to around 20 kHz.

The core idea of *ToMoBrush* leverages this special characteristic of the sonic toothbrush and uses it as a wideband excitation source for acoustic vibrometry for sensing the teeth. In a typical vibrometry system, the user applies excitation signal from sources such as piezoelectric transducers to induce vibrations in the object being studied [7, 20, 25]. The resonance behavior of the object can then be deduced and measured. Mathematically, we can describe the acoustic behavior of a simple probe-based system as [69]:

$$Y(f) = H(f)X(f) \tag{1}$$

where $X(f)$ is the excitation signal, $H(f)$ is the resonance signature of the object, and $Y(f)$ is the response signal being captured. Generally, one has a firm knowledge of $X(f)$ produced by a specialized source, so by simply deconvolving $X(f)$ from $Y(f)$ (division in frequency domain), the signature $H(f)$ can be obtained.

When a sonic toothbrush is in contact with a tooth, it also shows a similar resonance phenomenon. As shown in Fig. 3a, when the toothbrush is brushing a tooth, the resonant behavior of the tooth is induced, making the amplitude of the signal amplified at that frequency (2-20 kHz) compared to vibrating in isolation. However, compared to the typical vibrometry system, *ToMoBrush* faces three unique challenges when we want to extract the resonance signatures:

(1) The excitation signal especially the harmonic peaks in higher frequencies are constantly changing since they are essentially noise produced by the imperfect vibration motor.
(2) The excitation signal would be altered due to the contact with teeth. To confirm this, we attach a vibration sensor to the toothbrush body (illustrated in Fig. 4) to capture its source signal, the measurement of which is shown in Fig. 3b. We can clearly see the trend of amplitude changing across the spectrum when it is brushing (Blue line) versus vibrating alone (Red line).

(3) The toothbrush head and the tooth are not in perfect contact because of the soft bristles. As a result, different brushing strengths, tiny movements and slippages can produce slightly different acoustic responses from the teeth.

To formulate (1), we denote the excitation signal emitted by the motor $X(f)$, which is not a constant signal, but instead a harmonic source with constantly changing harmonics. To model (2) and (3) – the impact of bristle contact with the tooth on signal, we denote by $B(f)$, the overall change in the response signal this causes. These two terms in consonance allow us to re-write Eq. 1 as:

$$Y(f) = H(f)B(f)X(f) \tag{2}$$

Next, note that our microphone is not in direct contact with the teeth. Instead, it receives signals traveling through the air from both the teeth and the toothbrush. This means that in addition to $Y(f)$, the microphone also hears signals from the direct path $D(f)$ between the toothbrush and microphone, as well as other environmental noise $N(f)$. As shown in Fig. 4, the signal received by the microphone can be denoted as:

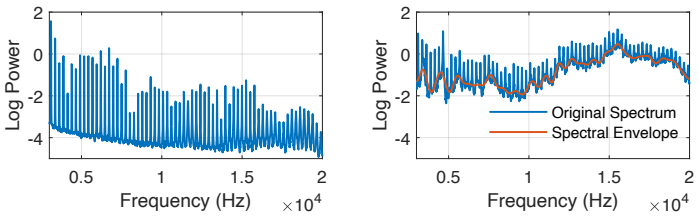
$$M(f) = Y(f) + D(f) + N(f) = H(f)B(f)X(f) + D(f) + N(f) \tag{3}$$

So far, we have modeled the vibration behavior of the system including toothbrush excitation $X(f)$, resonance behavior of the tooth $H(f)$, other effects caused by the contact $B(f)$, as well as direct path signal $D(f)$ and ambient noise $N(f)$. We will elaborate in Sec. 8.1 how we perform noise suppression based on EMD to mitigate the effect $D(f)$ and $N(f)$. For now, we assume that the signal we used to proceed with the signature extraction algorithm can be approximately represented by Eq. 2.

5.2 Extraction Algorithm

Given Eq. 2, our goal is to extract the teeth resonance signature $H(f)$ from the signal $Y(f)$. However, we also have two undesirable terms: $X(f)$, the constantly changing harmonics, and $B(f)$, related to strength, movement and other artifacts during the process of brushing. It is infeasible to measure either of these in real-time to divide them out.

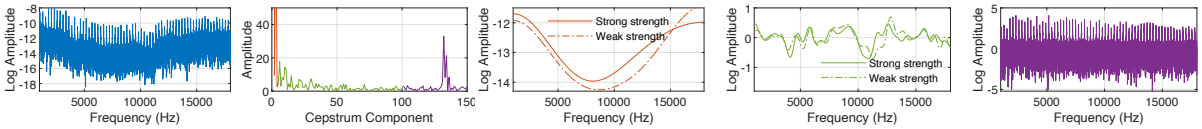
To tackle this challenge, we make a key empirical observation: while these three terms share the same frequency band, their rates of change across the frequencies are different.



(a) Vibrating alone (b) Brushing

Fig. 5. Examples of signal in log spectrum (denoised version) when the toothbrush is vibrating alone vs. brushing

To better understand this, consider the logarithmic spectrum of a signal shown in Fig. 5. We can clearly see that both the signal when vibrating alone (Fig. 5a) and the signal during brushing (Fig. 5b), have the periodic structure of a series of spikes, which correspond to the harmonics in the excitation signal. Importantly, we see that the signal produced during brushing (Fig. 5b) also has a macro-level structure, besides these harmonics, that owes its origin to the tooth's response. We can better observe this macro-structure by connecting the valleys of the harmonic structure. This is called the spectral envelope, which is composed of slow-changing peaks and



(a) Raw Log Spectrum (b) Cepstrum (c) Lower Components (d) Signature Components (e) Higher Components

Fig. 6. Signature Extraction algorithm by transforming the signal into cepstrum domain. The lower red components correspond to brushing artifacts such as strength and movement. The green components in the middle are the desired teeth resonance signatures. The higher purple components correspond to the excitation harmonics.

valleys (shown in red in Fig. 5b). By comparing Fig. 5a and Fig. 5b, we observe that the envelope corresponds to the tooth's resonance and is free of the excitations from the toothbrush.

One way of evaluating and separating these periodic structures in a signal changing at different scales/speeds is to use the Fourier Transform. Specifically, we take the Discrete Cosine Transform (DCT) of the log spectrum, to obtain the *cepstrum*[3] representation of the signal. This domain disentangles the three components described above cleanly across frequencies.

Mathematically, we transform the $M_p(f)$ into cepstrum domain as follows:

$$\mathcal{D}(\log|M_p(f)|) \approx \mathcal{D}(\log|X(f)| + \log|H(f)| + \log|B(f)|) = \mathcal{D}(\log|X(f)|) + \mathcal{D}(\log|H(f)|) + \mathcal{D}(\log|B(f)|)$$

where $\mathcal{D}$ represents the DCT. In this way, we convert the multiplicative components in the spectrum domain into additive components in the cepstrum domain where they are separable. The $\mathcal{D}(\log|H(f)|)$ term is the desired tooth resonance signature.

Fig. 6b shows an example of converting the signal in Fig. 6a into the cepstrum domain. Through detailed empirical analysis, we find that the three segments in different colors contain three different parts of information. First, we can clearly see a peak in higher cepstrum components colored in purple that represent the harmonic structure of the excitation signal $X(f)$, as it is composed of rapidly-changing spikes in the log spectrum. The location of the peak corresponds to the fundamental frequency of the harmonics. Next, the components at the middle cepstrum frequencies colored in green represent the tooth resonance signatures $H(f)$. We confirm this empirically, by observing that this region of the spectrum remains consistent when the same tooth is brushed multiple times, even with different strengths or slippages. Indeed, most brushing artifacts such as strength and movements mainly influence the few lowest cepstrum components colored in red, representing $B(f)$. Figs. 6c to 6e show the results of converting these three components back to the log spectrum domain. We can clearly see they represent the structure on different scales of changes in the spectrum and have measure the three different physical behaviors.

One might wonder if the extracted $H(f)$ (the green plot in Fig. 6d) is robust to differences in brushing behavior. To evaluate this, we illustrate a trace of the extracted components from the measurements on the same tooth but with different brushing strengths (we present a more robust multi-user evaluation later in results in Sec.9 and this is only a representative example to illustrate the point). In Fig. 6c and Fig. 6d, the dashed lines denote brushing with weak strength and the solid line for strong strength. We can clearly see the lower components in Fig. 6c have significant changes between the two brushing strengths, while the signatures in Fig. 6d are still very close to each other, especially when observing the location and amplitude of each peak and valley. This decomposition result showcases that the signature we extract is robust to differences in brushing behavior, and is robust to constantly changing excitation harmonics.

[3]a word-play on spectrum, commonly used in human speech analysis [68]

6 FEATURE SELECTION

The goal of *ToMoBrush* is to compare the tooth signature from any individual tooth being brushed with the corresponding healthy references, and compute a probability of each disease of interest for every single tooth. *ToMoBrush* seeks build a platform that can potentially detect a wide range of dental diseases. Different diseases are likely to induce changes to different parts of the signature. Using the signature as a whole from Sec. 5 may not be effective enough to distinguish different kinds of dental problems. One therefore desires more fine-grained signatures for different specific tasks (e.g. caries detection), to distinguish between healthy and unhealthy teeth.

In the data science community, researchers have developed contrastive learning frameworks based on Deep Neural Networks (DNN) to distinguish between similar or dissimilar pairs of data points in other contexts. However, DNN training of contrastive learning needs at least weakly labeled data [56] or data augmentations [18] from a relatively large dataset. Collecting a large real-world dataset using our new *ToMoBrush* platform is out of the scope of this work.

Instead, as the first system seeking to analyze the acoustic resonance of teeth for dental health sensing, we choose the most straightforward approach – to select a part of features among the extracted teeth resonance signature, which can best characterize if a specific disease occurs. The core idea of feature selection is to identify a sub-space of signature for each disease of interest where the separability of the healthy and unhealthy tooth is maximized, thereby providing fine granularity information as well as more resilience to noise and other factors for the classification task.

We modified the classical discriminant analysis (LDA [14]), a statistical algorithm modified to our specific feature selection problem. From Sec. 5, we extract tooth signature $\mathcal{D}(\log|H(f)|)$ represented by a series of DCT components (the points in the green plot of Fig. 6b), denoted as $(h_1, h_2, ..., h_s)$, where s is the length of a signature. Mathematically, we quantify the separability of each feature in the signature using the ratio of the between-class variance and the in-class variance. We define the gain for feature h_i:

$$G(h_i) = \frac{S_b(h_i)}{S_w(h_i)} = \frac{\sum_{k=1}^{K} N_k(\bar{h}_{i,k} - \bar{h}_i)^2}{\sum_{k=1}^{K} \sum_{n \in C_k} (h_i[n] - \bar{h}_{i,k})^2} \tag{4}$$

where $S_b(h_i)$ is the between-class variance that quantifies how far the collected data samples from different classes (teeth locations) stray from each other on feature h_i. $S_w(h_i)$ is the within-class variance that quantifies how compact the samples from the same class are on feature h_i. K is the number of classes (two classes: healthy/unhealthy in this case). C_k corresponds to the set of data samples in class k and N_k is the number of samples in class k. $\bar{h}_{i,k}$ is the mean value of feature i in class k, and $\bar{h}_i$ is the mean value of all samples.

Note that any feature h_i, offers different gains for different types of diseases based on the statistical distribution of our collected data. Thus, for every type of disease, we choose the set of features that provide the highest gain by maximizing:

$$\hat{M}, \hat{N} = \max_{M,N} \sum_{i=M}^{N} (G(h_i) - \alpha) \tag{5}$$

where $\alpha > 0$ is a tunable parameter to control the number of selected features. Instead of selecting individual features, we select a continuous range of DCT components in the cepstrum domain denoted as $(h_{\hat{M}}, h_{\hat{M}+1}, ..., h_{\hat{N}})$ with a starting index $\hat{M}$ and ending index $\hat{N}$, to preserve the physical meaning of the representation. We select this set of features as they best characterize disease occurrence on teeth. We also note that the feature selection is only performed on a partial validation dataset.

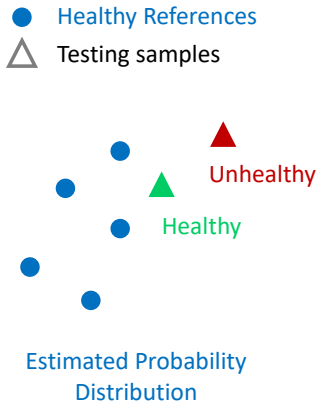


Fig. 7. Health detection by estimating the distribution of healthy references and thresholding the likelihood of the testing samples.

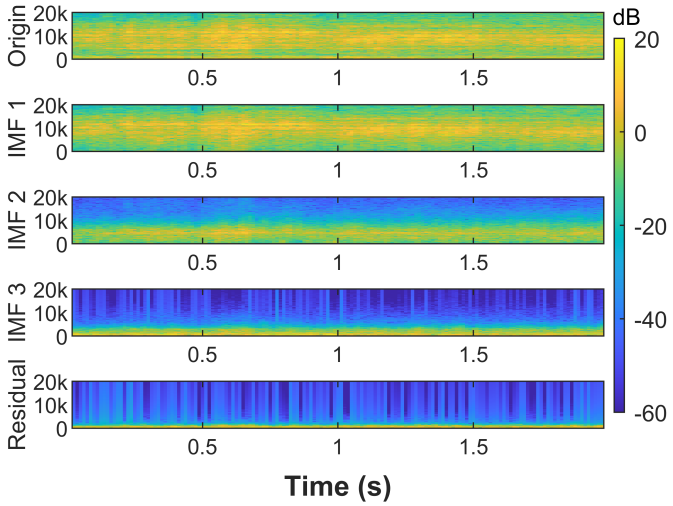


Fig. 8. Signal decomposition using EMD shown as spectrograms. IMF 1 and 2 correspond to resonance signals of teeth. IMF 3 corresponds to the excitation signal directly from the toothbrush, and the residual is environmental noise.

7 HEALTH DETECTION

After selecting features specialized for each kind of disease in every 50-millisecond time window, we aggregate and average the features from the same tooth among the time windows to improve the signal-to-noise ratio of the features. *ToMoBrush* then uses the aggregated features to perform health detection for diseases of interest. The most straightforward idea one may have is to use standard classification techniques such as SVM [21] to separate healthy and unhealthy teeth. However, since the tooth signature distributions of different teeth are very different, one would require training a classifier for every individual tooth from every different user. This of course is impractical because we will not have any unhealthy samples to train the classifier before the disease occurs!

Instead, given the healthy reference tooth signatures, *ToMoBrush* designs a one-class classifier for health detection using only healthy samples. Specifically, *ToMoBrush* estimates the probability density function of the healthy tooth signatures using the references, which essentially builds a profile for the original healthy tooth. The idea is visualized in Fig. 7. When a new measurement (testing sample) from the same tooth is recorded, we compute its likelihood given the estimated distribution. We consider the tooth as an unhealthy tooth if this computed likelihood is lower than a threshold. Note that every different disease on every single tooth requires a separate one-class classifier.

Mathematically, as the size of the reference samples is small and the underlying distribution is unknown, we model the probability distribution using Kernel Density Estimation (KDE) [30]. The probability density function can be estimated by:

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h})$$

where $X = (x_1, x_2, .., x_n)$ are the collected features from n healthy references. Note that we also perform normalization for each dimension of the feature before estimating the model. K is the kernel function that we empirically choose to be Gaussian. h is a smoothing parameter to control the variance of the Gaussian kernel. We note that using only a few reference samples, KDE can better characterize the probability distribution as a

non-parametric method. We present a comprehensive benchmark experiment by comparing KDE with other one-class classification algorithms.

To enable health detection, we evaluate the similarity of a new coming measurement with the healthy references using log-likelihood. Moreover, since the residual environmental noise can impact the tooth signature which may further produce false positive results, we bootstrap the detection accuracy using the data across different measurements and even across days. Specifically, we assume signatures across different measurements are independent. We can compute the aggregated log-likelihood:

$$\log p(x'_1, x'_2, ..., x'_m) = \sum_{i=1}^{M} \log \hat{f}_h(x'_i)$$

where $(x'_1, x'_2, ..., x'_m)$ are the features across different measurements. We present an evaluation result later in Sec. 9 to show multi-measurement bootstrapping significantly improves the health detection performance.

8 MITIGATING ERROR FACTORS

In real-world scenarios, *ToMoBrush* encounters various sources of error that can impede its performance. In this section, we first detail the noise suppression algorithm that was first mentioned in Sec. 5.1. We then present the DTW-based sequence alignment approach for mitigating tooth matching errors.

8.1 Noise Suppression

In real-world scenarios, the raw audio captured by the microphone is a mixture of various signals, including the resonance signal from the teeth, direct excitation signal from the toothbrush, environmental noise, and other artifacts. Although these components have different frequency distributions, they are likely to overlap in the frequency domain, making it harder to separate them simply through traditional band-pass filtering. As we modeled in Eq. 3, the raw acoustic signal captured by the microphone $M(f)$ can be denoted as:

$$M(f) = Y(f) + D(f) + N(f)$$

where $Y(f)$ is the response signal from the tooth, $D(f)$ is the direct path signal from the toothbrush, and $N(f)$ is other environmental noise. Only $Y(f)$ contains the tooth resonance behavior we desired which is required to be extracted for the later signature extraction step. To minimize the interference of other components and make the signal focus on $Y(f)$, we decompose the raw audio recording using Empirical Mode Decomposition (EMD) [34]. EMD is a data-adaptive multi-resolution technique that decomposes signals into physically meaningful components. These components are called Intrinsic Mode Functions (IMFs) that have well-defined frequency ranges and amplitudes. Prior systems leverage EMD-based algorithms for heartbeats [35, 62, 75] or breathing monitoring [72] by decomposing and removing the noise terms such as motion artifacts. Fig. 8 demonstrates a result of decomposing an audio recording of teeth scanning into three IMFs and residual components. Based on our detailed empirical analysis, IMF-1 and IMF-2 are resonance signals from the teeth mainly in the frequency range of 4-20 kHz, and 2-7 kHz respectively, which is consistent with the observation that the tooth mainly resonants in 2-20 kHz (shown in Fig. 3a). The IMF-3 is mainly the excitation signal directly from the toothbrush under 3 kHz ($D(f)$ in Eq. 3), and the remaining residual component is mainly environmental noise (e.g. surrounding speech) and other artifacts under 1 kHz. Thus, to perform noise suppression, *ToMoBrush* only retains the first two IMF components and adds them up for further processing. We note that this suppression is done once on the entire original audio signal, rather than the individual 50-millisecond time chunks.

As a result of eliminating residuals and IMF-3, the environmental noise in the raw audio signal is suppressed and the $D(f)$ term gets filtered out (detailed evaluation in Sec. 9.1.4). At the end of this process, we obtain the

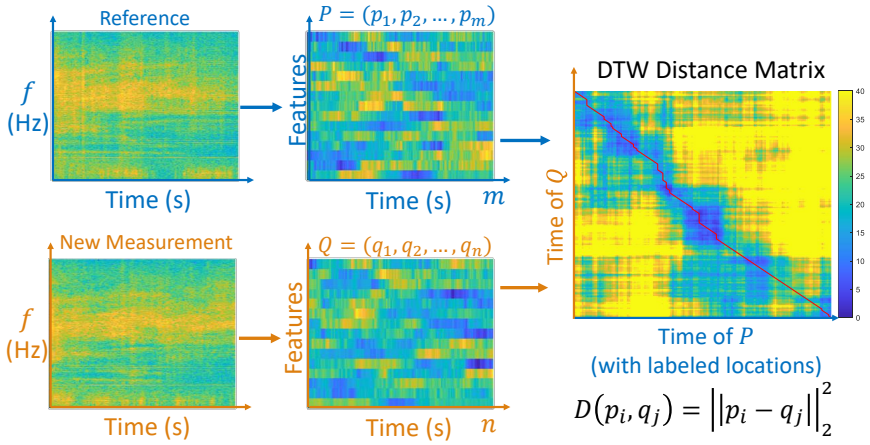


Fig. 9. *ToMoBrush* performs sequence alignment using DTW based on the pair-wise distances between extracted features.

processed signal as:

$$M_p(f) = \sum_{i=1,2} IMF_i(M(f)) \approx Y(f) = H(f)B(f)X(f) \tag{6}$$

8.2 Sequence Alignment

ToMoBrush instructs the users to brush their teeth in a pre-specified order, guided by a video, to collect signatures from individual teeth. However, given the low-light conditions at some blind spots inside the mouth (such as inner upper molars), occasional matching errors can occur when a user collects measurements in these regions with poor visibility. For instance, a user may inadvertently miss brushing a tooth located in a blind spot. To limit the impact of matching errors in blind spots, *ToMoBrush* develops a sequence alignment algorithm based on Dynamic Time Warping (DTW) to align brushing sequences of the user's tooth with the references from a same quadrant of teeth, which allows to obtain signatures from each corresponding tooth more accurately.

In Sec. 5, we obtain a tooth resonance signature in every 50 ms time window in the audio recording. Thus, as shown in Fig. 9, a time-series sequence of signatures is generated from the raw audio recording, which includes signatures from every individual tooth, as brushing is performed. The goal of the sequence alignment is to find the correct match of each signature to the corresponding tooth number (i.e. tooth location). To achieve this, *ToMoBrush* aligns the received acoustic testing sequence produced from the scan with prior reference sequences obtained in dental clinics, where the locations of teeth brushed are known (e.g. via camera imaging). In this manner, it can isolate acoustic signatures of specific individual teeth. Note that *ToMoBrush* only performs alignments for the sequences from the same quadrant of the same user.

ToMoBrush seeks to accurately align the sequence even under the condition that the tooth signatures are changed by health conditions. Therefore, to make the sequence alignment resilient to health condition changes and other residual environmental noise, *ToMoBrush* applies the same feature extraction algorithm as discussed in Sec. 6 to select a part of the signature as the input of the sequence alignment, instead of using the whole signature or raw audio recording. The features are selected by maximizing the separability between different teeth, and can best distinguish the location of the tooth brushed.

Essentially, *ToMoBrush* aims to align two time-series (the reference and the testing sequences) which may each involve brushing at a different speed (i.e. spending different durations on different teeth). We use Dynamic Time Warping (DTW), a standard technique known in time-series alignment (classically used in speech processing [43,

52]), to match and find similarities between two temporal sequences. For the distance metric in DTW, we use the square of Euclidean distance to evaluate the similarity between two feature samples. Specifically, for each pair of features in the two sequences $P = (p_1, p_2, ...p_m), Q = (q_1, q_2, ..., q_n)$, we obtain a distance:

$$D(p_i, q_j) = ||p_i - q_j||_2^2$$

Note that we perform normalization for each dimension of features using mean and standard deviation before the alignment. Fig. 9 shows an example of the distance matrix for two sequences and the curve in red indicates the alignment result. As the ground truth of the tooth location corresponding to each point in the reference sequence is known, we can map the points in the testing sequence to the corresponding tooth locations based on the alignment curve.

After obtaining the tooth locations for each 50 ms time window, we can aggregate the adjacent signatures from the same tooth in the sequence by averaging these values to improve signal-to-noise ratio. Thus, the user can spend a longer period on the tooth they suspect is unhealthy to obtain a more accurate detection result. We further present an evaluation in Sec. 9 to show the relationship between the evaluation accuracy and the duration of time the brush remains on an individual tooth.

9 IMPLEMENTATION AND EVALUATION

We implement the prototype of *ToMoBrush* on a Philips Sonicare ProtectiveClean 6100 electric toothbrush [8] and a Voice Technologies VT500X waterproof microphone [6]. For toothbrush heads, we choose Brushmo Compact Replacement Heads [4] as it has a smaller size compared with the standard one. As shown in Fig. 10a, the microphone is fixed 1.25 cm close to the toothbrush head using a tiny glass stick, which is integrated onto the toothbrush body using a hot-melt adhesive. The microphone is sampling at 44.1 kHz for audio recording and connected to a laptop through a 3.5 mm audio jack. For signal processing, we choose a 50 milliseconds time window and 75% overlap to perform STFT. For signals from users and dental models, we empirically choose the frequency range of 2-16 kHz, and 2-18 kHz respectively to obtain the cepstrum (illustrated in Sec. 5.2). We note that the only learnable parameter in our pipeline is the selected feature index presented in Sec. 6. To evaluate the generalizability of our method, the following evaluation result is presented by performing the leave-one-out cross-validation in the feature selection module.

In the following sections, we first present the technical evaluation of *ToMoBrush*, followed by an expert evaluation by interviewing two experienced dentists.

9.1 Technical Evaluation

In the evaluation in this section, we focus on the capabilities of *ToMoBrush* to detect three dental conditions: dental caries, dental calculus, and interdental food impaction. For the main result section (Sec. 9.1.2), we ensure that the tooth being brushed is always the tooth of interest (i.e. no matching errors). We further evaluate the capability of *ToMoBrush* to mitigate the matching errors in Sec. 9.1.5.

9.1.1 Data Collection Procedure.

Evaluation through user study: We conducted an IRB-approved user study on 20 participants in total (13 males and 7 females). Specifically, We evaluated the health detection for caries and calculus on 35 teeth with diseases from 9 participants in a dental clinical center, where we perform measurements on real patients under the guidance of professional dentists. Each measurement is taken by brushing from the chewing surface of the tooth and lasts for 3-4 seconds. For the evaluation of caries detection, we took measurements on specific teeth with caries (labeled by dentists, example shown in Fig. 10b) before and after the standard composite dental fillings [16]. For the evaluation of calculus detection, we perform measurements on teeth originally with calculus (example in Fig. 10c) before and after the standard prophylaxis dental cleaning [27]. For every tooth of interest

(a) Hardware (b) Dental Caries (c) Dental Calculus (d) Model Emulation

Fig. 10. (a) shows the hardware setup of *ToMoBrush*. (b) and (c) are examples of dental caries and calculus before and after the treatment, which are two diseases we seek to detect in the user study in the dental clinic center. (b) shows a dental model we used to emulate caries and food impaction.

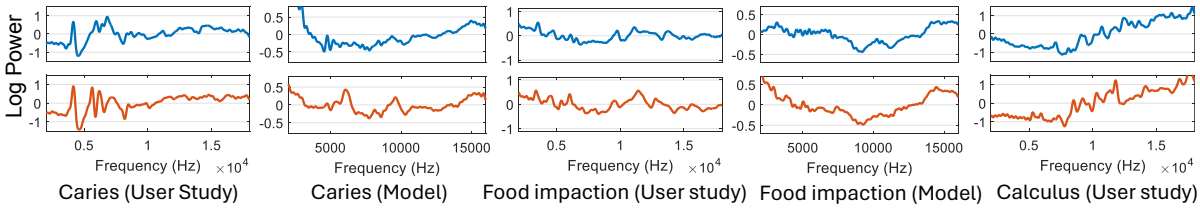
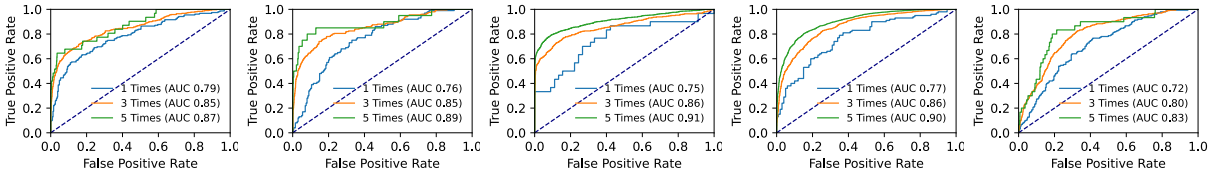


Fig. 11. Examples of signature visualization in the frequency domain. The blue plot is a signature from a healthy tooth, while the orange plot is a corresponding unhealthy signature from the same tooth.

from every patient, we collected 10 measurements each before and after the treatments (20 measurements in total). We assume the teeth after the treatments are healthy, use a subset of samples (5 measurements) after the treatments as references, and test on the rest of the samples (15 measurements). We use random combinations of 3 or 5 measurements among the 15 measurements to evaluate the multi-measurement bootstrapping as presented in Sec. 7. We note that all of caries and calculus we collected data on are in the early or mid stage. Specifically, the caries samples have the ICDAS scores [36] in the range of $1-4$ (out of 6), and the calculus samples have the CSSI [39] scores in the range of 1–2 (out of 3, thickness < 1mm).

We also evaluated detection for interdental food impaction in a home setting on 33 teeth from 11 participants. We collected measurements on the original normal teeth, teeth with food impaction after eating, and the teeth after flossing to clear out impacted food. We collected 10 measurements in each stage (30 measurements in total) for every tooth of interest from every participant. We use 5 measurements from the original normal teeth as references and test on the others.

Evaluation on dental models: In order to perform a comprehensive evaluation of our system in more controlled and flexible settings, such as evaluating under different brushing strengths, brushing duration, and damage to different parts of teeth, we choose to increase the scale of the evaluation by using dental models, mainly due to ethical reasons (i.e. avoiding possibility of damage to human teeth). We note that the properties of teeth material on dental models are not equivalent to those on humans. By collaborating with expert dentists, we select the dental models with as accurate as possible structures that are designed for professional dental training (e.g. filling, cleaning). Our following evaluation results demonstrate consistent health detection performance on user study and dental model, which validate the effectiveness of involving dental models into our experiments.



(a) Food impaction (U) (b) Food Impaction (M) (c) Caries (U) (d) Caries (M) (e) Calculus (U)

Fig. 12. *ToMoBrush* Overall Results (U - User Study, M - Dental Model Evaluation): (a) - (e) show the ROC curves of the three health detection tasks. Three solid lines colored in blue, orange, and green show the detection performance using different number of cumulative measurements. The blue dashed lines show the performance of random choosing.

To emulate dental caries, we use an electric drill to produce small holes on the surface of the model's teeth. Each hole has a less than 2mm diameter and less than 1mm depth, which is consistent with the size of general early-stage caries. Besides, we insert meat samples in between two teeth to emulate food impaction (shown in Fig. 10d). We collect measurements on the model teeth before and after the changes for our evaluation. Each measurement is taken by brushing from the chewing surface of the tooth and lasts for 5-6 seconds. The number of measurements on each tooth of interest is the same as in the user study.

For each of the experiments we conducted, we include an example signature to showcase the differences between a healthy and unhealthy sample from the same tooth, which is shown in Fig. 11. We convert the features in cepstrum domain back to the log spectrum for visualization. We note that the examples are just for illustration purposes. The signatures collected from different teeth can have totally different shapes.

9.1.2 Main Results of Health Detection.

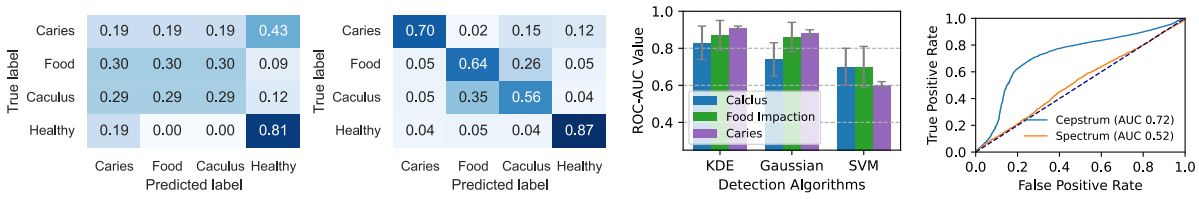
We evaluate the health detection performance of *ToMoBrush* using Receiver Operating Characteristic (ROC) curve, which shows the detection sensitivity (True Positive Rate) versus False Positive Rate by varying the decision threshold. The Area Under Curve (AUC) value of the ROC curve represents *ToMoBrush*'s overall capability to distinguish unhealthy samples from healthy samples.

Interdental Food Impaction: We evaluate *ToMoBrush* for the detection of interdental food impaction through the user study and on the dental model. We collected measurements on 33 teeth from 11 participants in total in the user study. As shown in Fig. 12a, using a single measurement for around 3-4 seconds, *ToMoBrush* achieves a 0.79 AUC value with a 95% confidence interval of (0.740, 0.834). With 3 and 5 measurements, *ToMoBrush* bootstraps the performance to 0.85 (0.822, 0.876), and 0.87 (0.787, 0.949) respectively.

We collected measurements on 20 teeth on the dental model with food impaction. As shown in Fig. 12b, *ToMoBrush* achieves a 0.76 (0.699, 0.821) AUC value using a single measurement. With three and five measurements, *ToMoBrush* bootstraps the performance to 0.85 (0.821, 0.888), and 0.89 (0.792, 0.982) respectively. We observe that the detection performance we obtain from user study and dental model emulation is very close to each other.

Caries Detection: We evaluate *ToMoBrush* for caries detection through the user study and on the dental model. We collected data on five teeth from four participants with caries from real patients during the user study and the results are shown in Fig. 12c. *ToMoBrush* achieves 0.75 (0.637, 0.870), 0.86 (0.834,0.884), and 0.91 (0.893, 0.922) AUC values with 1, 3, and 5 accumulative measurements respectively. Similarly, we collected data on 20 teeth with caries on the dental model the results are shown in Fig. 12d, where it achieves a performance very close to the user study – 0.77 (0.704, 0.835), 0.86 (0.848, 0.878), and 0.90 (0.894, 0.912) AUC values under 1, 3, and 5 measurements respectively.

Calculus Detection: We evaluate *ToMoBrush* for caries detection only through the user study. We collected data on 30 teeth from five participants with calculus from real patients and the results are shown in Fig. 12e. *ToMoBrush* achieves 0.72 (0.663,0.768), 0.80 (0.768, 0.829), and 0.83 (0.738,0.920) AUC values with 1, 3, and 5



(a) Without Feature Selection (b) With Feature Selection (c) Health Detection (d) Signature Extraction

Fig. 13. Benchmarking on different algorithm modules: (a) and (b) shows the confusion matrix of discriminating different diseases without and with the feature selection algorithm. (c) show the performance of using different algorithms for health detection. (d) shows the performance improvement of the signature extraction algorithms based on cepstrum.

accumulative measurements respectively. Note that the dental calculus on the users who get dental cleaning frequently is generally very tiny, which induces smaller changes to the tooth resonances compared with caries or food impaction, so the detection performance is slightly lower.

Accuracy vs. related work: OralCam [41] is the only related work to *ToMoBrush* that targets low-cost in-home dental health monitoring. OralCam reports 0.84, 0.79 AUC values for caries and calculus detection, which is close to the performance of *ToMoBrush* using 3 measurements (0.86, 0.80 AUC), but lower than our results with 5 measurements (0.91, 0.83 AUC). We note OralCam evaluates on a visible disease dataset, but *ToMoBrush* also has the unique advantage of detecting diseases at the most inner molars or even invisible corners where common cameras can hardly capture high-quality images.

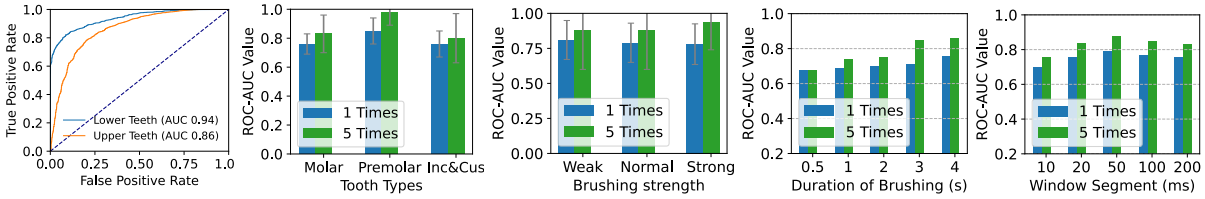
9.1.3 Benchmarking on Algorithm Modules.

In this section, we present the benchmark experiments on each of the algorithm modules in our data processing pipeline, including feature selection, health detection. and signature extraction.

Capability to Discriminate Different Diseases (Feature Selection Module): Besides detecting each individual disease, *ToMoBrush* is also able to discriminate between different diseases attributed to the feature selection algorithm presented in Sec. 6. We evaluate the performance of *ToMoBrush* to discriminate different diseases by combing the collected datasets from the user study. We first present the result without using the feature selection algorithm in a confusion matrix shown in Fig. 13a. We input the extracted tooth signature as a whole (green part in Fig. 6b) into the health detection module, and consider all three diseases to have equal probabilities for the positive outputs from the health detection. While the true negative rate is relatively high (0.81), the system can not distinguish different diseases without specialized features.

The confusion matrix shown in Fig. 13b demonstrates the performance of *ToMoBrush* to distinguish different diseases when the feature selection algorithm is integrated. We can observe that food impaction and calculus are more challenging for the system to discriminate, since both of them predominantly occur in interdental spaces, inducing changes of the resonance behaviors of the teeth in a similar way. We note that the results are generated by picking operating points on the ROC curves to maximize the overall accuracy. In actual deployment, *ToMoBrush* can choose a more conservative operating point to further reduce the false positive rate. An example is shown in Fig. 16, in which the overall false positive rate is reduced to 3% with a certain level of sacrifice to the true positive rate.

Health Detection Module: We conduct a benchmark experiment by substituting the health detection module with other one-class classification algorithms, and comparing the performance of detecting three dental conditions in our user study. Specifically, we compare the performance of KDE with two baseline algorithms, maximum likelihood estimation of a Gaussian distribution, and one-class SVM [54]. Similarly, we use the reference health samples to train the models, and test on other samples. The result is shown in Fig. 13c. For the detection of all



(a) Lower vs. Upper Teeth (b) Tooth Types (c) Brushing strength (d) Duration of brushing (e) Window Size

Fig. 14. Benchmarking on Experiment Setup: (a) shows the detection accuracy on lower teeth v.s. upper teeth. (b) shows the detection accuracy on different types of tooth. (Inc&Cus: Incisors and Cuspids). (c) compares the detection accuracy under different brushing strengths. (d) compares the detection accuracy of different durations of brushing on a single tooth. (e) compares the detection accuracy when using different window sizes.

three conditions, the KDE algorithm outperforms the two baseline algorithms, mainly because KDE employs a combination of multiple kernels, which can better characterize the distribution of the features from healthy teeth with only a few reference samples.

Signature Extraction Module: We evaluate the performance of the signature extraction algorithm on the detection of food impaction. Specifically, instead of using the cepstrum representation of the signal, we use the spectrum representation in the frequency domain. We remove the following feature selection module in this experiment for fair comparison. The result is shown in Fig. 13d. The system's end-to-end performance using spectrum representation obtains only 0.52 ROC-AUC value, which is close to the performance of random choosing. The main reason is that the spectrum representation of the signal is dominant by unstable excitation harmonics from the sonic toothbrush (example in Fig. 6a), which hinders the system from extracting reliable signatures to character the resonance of the tooth. We note that the performance of using cepstrum representation is lower than our main experiment, which is because of the removal of feature selection module.

9.1.4 Benchmarking on Experiment Setup.

Upper vs. Lower Teeth: We compare the detection accuracy of caries on the dental model for the upper teeth versus the lower teeth when using 5 times cumulative measurements. Fig. 14a shows *ToMoBrush* achieves a better detection performance on lower teeth (AUC = 0.94) compared with upper teeth (AUC = 0.86). Brushing upper teeth is usually more challenging as the applied force is in the opposite direction of gravity. As a result, the brush head generally has better contact with the lower teeth leading to better detection performance.

Different Types of Tooth: We compare the detection accuracy for different types of tooth on the detection for food impaction in the user study. Specifically, our dataset with 33 teeth in total consists of 15 molars, 8 premolars, 10 incisors and cuspids. The results in Fig. 14b show that the detection on premolars achieves the best performance, while the accuracy on the incisors and cuspids is slightly lower. We attribute the difference to the size of the tooth and the contact between the toothbrush head and the tooth. The sizes of premolars are generally small, so the resonances can be more easily affected by dental conditions. For incisors and cuspids, though the size of a part of them is also small (lower teeth), the area of the chewing surface of the teeth is very limited, which affects the quality of contact between the teeth and the toothbrush head.

Brushing Strength: In this experiment, we study how the brushing strength affects the detection accuracy of food impaction detection on the dental model in Fig. 14c. This study shows slight changes in accuracy when brushing strength varies, showing that *ToMoBrush* is robust to different brushing behaviors. As we discussed in Sec. 5.2, our signature extraction algorithm models and isolates the effects of different brushing behaviors, such as strength and tiny movement. During our experiments, the three different levels of brushing strength are

observable in the lower components of the cepstrum (an example shown in Fig. 6c). In our collected dataset, the total energy of the lower cepstral components when using normal strength is approximately 50% of the one using strong strength, and the energy when using weak strength is approximately 10% of the one using strong strength.

Interestingly, the detection accuracy increases as the brushing strength increases when *ToMoBrush* has enough measurements (5 times). As the strength increases, the brush bristles have closer and tighter contact with the brushed tooth, which in turn produces a response resonances with a higher signal-to-noise ratio.

Brushing Duration: We evaluate in this experiment how the duration of each brushing measurement on a single tooth affects the detection accuracy of food impaction in the user study. *ToMoBrush* seeks to average the tooth signatures from the same tooth across multiple time windows to improve the signal-to-noise ratio. As Fig. 14d shows the detection accuracy increases as the brushing duration increases for each tooth as expected, achieving relatively good performance when the duration reaches 3-4 seconds, which is also the chosen duration in our user study.

Length of Window Segment: We evaluate in this experiment how the configuration of window size affects the detection accuracy of food impaction in the user study. *ToMoBrush* averages the signatures from the same tooth across time windows to improve the SNR. If the window size is too small, the frequency resolution of each window would be low which hinders the signature extraction algorithm from obtaining fine-grained signatures with detailed information. On the other hand, if the window size is too large, the number of windows staying on a single tooth would be small, thus the averaging can not improve the SNR significantly. The result in Fig. 14e shows the detection accuracy when using different window sizes. The optimal window length in the experiment is 50 ms, which is also the chosen configuration in our main evaluation.

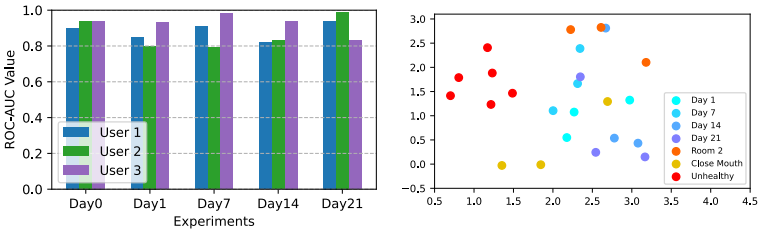
Different Sonic Toothbrushes: We further replicate our prototype on two other sonic toothbrushes, Philips Sonicare 4100 [9] and Oralvue Seebrush [2]. We evaluate the two prototypes for food impaction detection using dental models. Sonicare 4100 achieves 0.78 and 0.90 ROC-AUC values using 1 and 5 measurements respectively. Seebrush achieves 0.72 and 0.85 ROC-AUC values using 1 and 5 measurements respectively. The performance of both toothbrushes is within the margin of error of our main experiment prototype, which demonstrates the generalizability of our system to enable dental health sensing on different types of sonic toothbrushes.

9.1.5 Performance repeatability.

We further study the performance repeatability of *ToMoBrush* through a longitude study across 21 days, as well as under different environments. Three participants are included in the study. On day 0, we conduct the same experiment procedure as the evaluation for interdental food impaction illustrated in 9.1.1. We collect 10 measurements for the tooth with food impacted, and 10 measurements after the food is cleaned out. Next, we collect five measurements on the same teeth being tested on day 1, day 7, day 14, and day 21. Before taking the measurement, we ensure that the teeth being tested are clean.

We process the dataset by using the five measurements collected on each day as the healthy references, and testing on the measurements collected on day 0 (5-times boosting). The average ROC-AUC values under different setups on three participants in shown in Fig. 15a. From day 0 to day 21, we observe a certain level of performance variation, which could result from ambient noise, inconsistent brushing orientations, and other factors. Fortunately, as shown in the example in Fig. 15b, though the signatures collected from the same healthy tooth can be slightly inconsistent, they are still clustered around a center and can be well-separated with the unhealthy samples cluster colored in red. Thus, the overall ROC-AUC values across 21 days are consistently higher than 0.8, demonstrating a good performance repeatability of the system.

On day 21, the user also collects three measurements in another room (Room 2). We further instruct the user to control to minimize the opening of the mouth during brushing, and collect three measurements (Close Mouth). We can observe that the orange (Room 2) and yellow (Close Mouth) points in Fig. 15b drift from the center of the



(a) Performance across days (b) t-SNE scatter plot

Fig. 15. Performance repeatability: (a) shows the performances for food impaction detection by using the references collected across days. (b) visualize the features collected from a user using t-SNE [63] for dimension reduction.

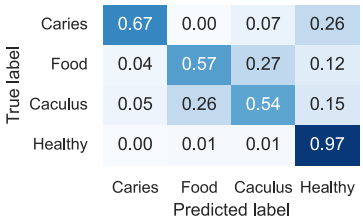
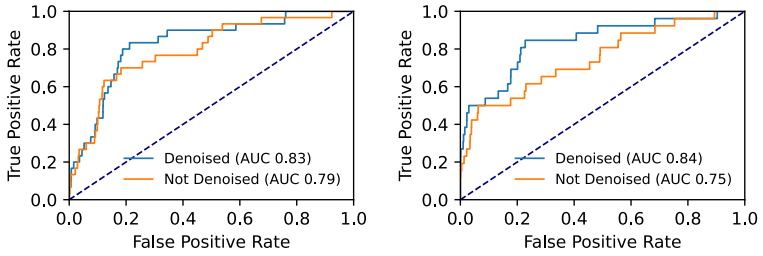


Fig. 16. Confusion Matrix generated by more conservative operating points.

original health samples cluster, but can still be separated from the unhealthy samples cluster. The experiment results indicate that, to ensure the optimal health sensing performance of the *ToMoBrush*, the user is encouraged to perform measurement in a consistent environment with a relatively consistent mouth openness. We leave a more comprehensive discussion of the multipath issue to Sec. 10.

9.1.6 Benchmarking on Error Mitigation Modules.

In this section, we present the evaluation results of the Noise Suppression module (Sec. 8.1) and the Sequence Alignment module (Sec. 8.2) to mitigate the errors.

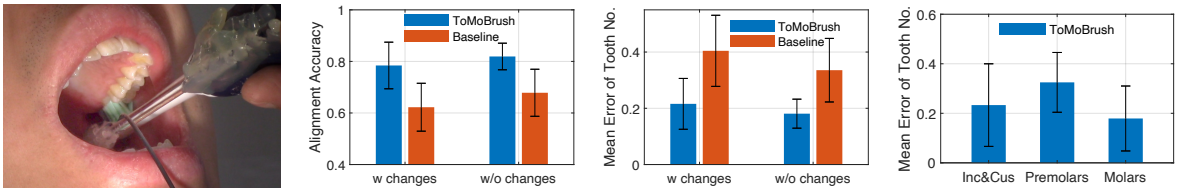


(a) Quiet environment (Calculus) (b) Noisy environment (Food Impaction)

Fig. 17. Noise Suppression Benchmark: The health detection performance in the user study with and without the denoised algorithms

Noise Suppression: We illustrate in this experiment how the Noise Suppression module (Sec. 8.1) affects *ToMoBrush*. We present the accuracy of the calculus and food impaction detection in our user study. The user study of dental calculus detection is conducted in a relatively quiet dental clinical room, while the user study of food impaction is conducted in a relatively noisy home setting with different noise sources, such as surrounding speech, air conditioner humming, and water gurgling from a faucet. As shown in Fig. 17a, the noise suppression algorithm improves the ROC-AUC of detecting calculus by 0.04, mainly because the direct path signal from the toothbrush is suppressed. For the detection of food impaction, the noise suppression module improves the performance more significantly by 0.09 ROC-AUC value, as it significantly suppresses both the direct path signal and other environmental noise.

Sequence Alignment: We conducted the user study on three participants along with the health detection experiments for interdental food impaction in a home setting. The user brushes the teeth one by one in a same



(a) A video frame (b) Alignment accuracy (c) Alignment Error (d) Types of Teeth

Fig. 18. Evaluation results of sequence alignment. (a) shows an example of a recorded video frame to capture the ground-truth location. (b) shows the accuracy of the sequence alignment to the correct tooth number. (c) shows the average alignment error of tooth number. (d) shows the alignment error on different types of teeth (Inc&Cus: Incisors and Cuspids).

quadrant, from the innermost molar to the outermost incisor. We collected measurements on the original teeth quadrant, the teeth quadrant with food impacted after eating, and the teeth quadrant after clearing out the food. We take 5 measurements in each stage (15 measurements in total from a teeth quadrant). We use the measurements from the original normal teeth as references and perform alignment on the others. A video camera is put in front of the user's mouth to capture the real-time ground truth locations of the toothbrush. A lamp with a narrow light beam to used to brighten up the video. Fig. 18a shows an example of a frame in the recorded video. We use an online video labeling tool [3] to label the tooth number being brushed in each video frame.

We evaluate the sequence alignment on 90 collected teeth scanning sequences from 6 teeth quadrants in total. The results are shown in Fig. 18b and Fig. 18c. We compare the results with a baseline algorithm which assumes the teeth scanning is performed at a uniform speed, equally distributed time to each tooth. We present the alignment performance when the status of some of the teeth in the quadrant changes (i.e. food impaction in our experiments) and without any changes compared with the status of the references. When the tooth status remains the same as the references, *ToMoBrush* achieves an alignment accuracy of 82% and an average error of tooth number 0.18, while the baseline algorithm has 68% and 0.33 respectively. When the status of some of the teeth changes, *ToMoBrush* achieves a 78% localization accuracy and 0.21 average error of tooth number. The baseline algorithm achieves 62% and 0.40 respectively. These results show the DTW algorithm in *ToMoBrush* outperforms the naive algorithm on sequence alignment of teeth scanning. Also, the alignment performance only degrades slightly when tooth status changes since the selected features for the sequence alignment algorithm are designed to be resilient to health condition changes.

Fig. 18d further shows how the different types of teeth, like Incisors, Cuspids, Premolars, and Molars, affect the alignment error. The figure shows the mean error of detecting the tooth number that the user is currently brushing, and how it varies over different types of teeth. Since the premolars are generally smaller and have irregular non-flat shapes compared with other teeth, it introduces slightly higher alignment errors.

9.2 Expert Evalution

To evaluate *ToMoBrush* from the expert's point of view, we interviewed 2 board-certified dentists (D1 & D2). D1 has 30 years of practice with expertise in restorative dentistry, and D2 has 8 years of practice with expertise in oral radiology. We investigated the following questions:

(1) Is our current data collection procedure valid? What would be your recommendations to improve the data collection procedure for our system?
(2) What is your opinion of the current performance of our system?
(3) What are the potential benefits of integrating this technology into the current dental healthcare system? What are the main limitations and barriers?

Data collection procedure: Both dentists agreed that our data collection procedure is sufficient and valid. For the recommendations, D2 points out the importance of collecting data from caries evolving at different stages, which can be left as future work. D2 mentioned: "*If they are really small caries, we don't care. If they're really big carries, we can actually see it and be fine. So this becomes very useful when we can detect those in-between carries. But that will be a much harder data collection.* " Besides, D1 suggested exploring the prototype of toothbrush heads that are designed for actual detection rather than existing ones designed for hygiene.

System Performance: Both dentists stated that the overall detection accuracy of our system is promising, especially considering our current system is just using a regular toothbrush without specialized hardware. Meanwhile, both dentists agreed that, from a clinical point of view, it is necessary to further optimize the system and improve its accuracy. D2 mentioned: "*I think the results are promising... When using cumulative 5 times measurements, the accuracy is consistently high... But we will have to tweak everything. The calculus one with one-time measurement seems to work at 0.72, which is not decent.*"

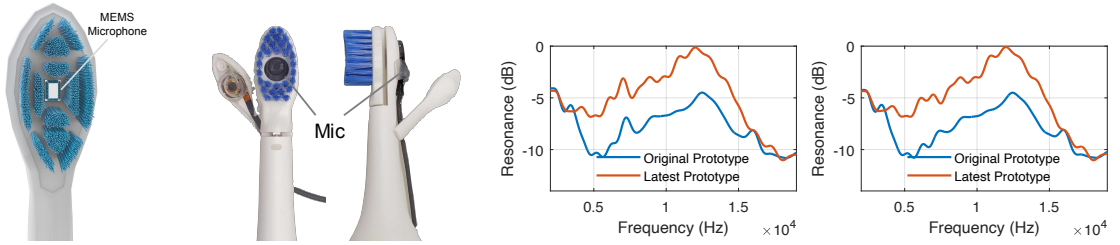
Benefits to current dental healthcare system and potential barriers: Both dentists considered that the potential of at-home early caries detection of our system is beneficial to the dental healthcare system, since it can catch caries earlier, prevent more severe diseases from happening and reduce the overall costs of the system. D1 also mentioned detecting unseen food impaction will have a direct impact on the patient's overall periodontal health. D2 emphasized the importance of at-home dental calculus detection: "*We recommend the patient do come for a checkup every 6 months, but in practicality that is not possible for a lot of people. So having that calculus detection system will also again help the patient understand that there's some issue going on, and they need to go to the dentist and get it fixed before it becomes a much bigger issue, such as periodontal bone loss.*"

For the potential barriers, D2 is concerned about the accessibility of our system to low-income families. Families without access to regular dental visits may also lack powerful enough smartphones to receive and process the data. D1 is concerned about the interpretation of data by the patients and suggests the actual system can present the result by incorporating the dentist's interpretation.

10 DISCUSSION AND LIMITATIONS

Next generation prototype: While our approaches and the proof-of-concept hardware setup (shown in Fig. 10a) demonstrates promising results, future areas of exploration still remain. One may be concerned that our current setup integrates the microphone with a separate glass stick, which may induce inconvenience and instability for everyday use. In the future, we envision building a next-generation prototype of *ToMoBrush* that integrates a tiny MEMS microphone into the center of the toothbrush head (as shown in the concept diagram in Fig. 19a). Such design not only ensures usability, but also shortens the distance between the vibrating teeth and the microphone, which has the potential to improve the Signal-to-Noise Ratio (SNR) and the overall sensing performance.

To demonstrate the feasibility of the proposed prototype, we conduct a proof-of-concept experiment by modifying the SeeBrush [2] (a camera-integrated toothbrush). As shown in Fig. 19b, we fixed the microphone to its original camera location on the toothbrush head. A preliminary experiment by brushing on a dental model shows promising results as in Fig. 19c. Comparing with our original setup (Fig. 10a), this new prototype obtains very similar resonance behaviors when observing peaks and valley locations. Moreover, in the main resonance frequency range of the object (4-16 kHz), the new prototype achieves a mean value of 3.75 dB signal power enhancement as well as captures more detailed variations, mainly attributed to the shorter signal path from the tooth to the microphone. We further damage the tooth using an electric drill and perform the measurement again. Similar changes in the resonance signal can be observed (e.g. the peak disappears at 7 kHz) for both the original prototype and the new prototype, which validates the potential of the new prototype for more convenient health



(a) Prototype rendering (b) Proof-of-concept Prototype (c) Resonance before Damaged (d) Resonance after Damaged

Fig. 19. A concept diagram of the next generation prototype of *ToMoBrush*. (a) shows the rendering of the ideal prototype. (b) shows a proof-of-concept prototype by modifying SeeBrush with a microphone. (c) and (d) show the resonance signal enhancement (harmonics removed by algorithms) by comparing with the original prototype shown in Fig. 10a. The experiment is conducted on a tooth from a dental model before and after damage.

sensing. For real deployment, we note that the resonance signatures obtained from different hardware are not comparable, i.e. the reference samples and testing samples should be obtained from a same hardware setup to ensure consistency.

Design choice of changes detection: One may wonder why *ToMoBrush* chooses to detect changes rather than detecting disease directly at once. We note that dental health sensing is a personalized task since the oral condition including the acoustic resonance of different people can vary widely. For example, one's healthy teeth may have a very similar resonance compared with another's teeth with caries. Building a normal discriminative model to judge the dental condition would require a large amount of labeled data and might not be accurate due to individual differences. Instead, *ToMoBrush* chooses the approach that compares the new measurements with the previous references collected from the same person in a healthy state. This approach can be beneficial as it only focuses on detecting changes to achieve personalized sensing.

Sensing during cleaning: Though *ToMoBrush* is based on off-the-shelf sonic toothbrushes, it requires an additional dedicated procedure to perform dental health sensing than regular oral hygiene and does not involve toothpaste. The video-guided procedure instructs the user to brush their teeth one or two at a time, staying on the chewing surface of each for 3-4 seconds. For future work, we aim to collect a larger dataset in the field and leverage data-driven approaches such as Connectionist Temporal Classification [29] based on deep learning, to support dental health sensing, despite an arbitrary brushing pattern.

Long-term monitoring to detect different severity of conditions: The current user study of *ToMoBrush* does not involve long-term monitoring to distinguish different severity of diseases (e.g. size of caries). However, the expert interviews pointed out the value of detecting caries at the transition stage. For future work, we aim to design and conduct long-term experiments incorporating dental clinics and large-scale populations susceptible to dental diseases, to validate and optimize the *ToMoBrush*'s detection performance for diseases at different stages.

Bristle deterioration on toothbrush head: One may be concerned that the captured tooth resonance signature would become inconsistent when the bristle on the toothbrush head undergoes wear and tear. To solve this problem, users can test the extracted signature of *ToMoBrush* on a reference object with a wide resonance band (e.g. a model tooth) before performing a self-exam. If a clear difference can be observed in the signatures from the same reference object, it indicates that the toothbrush head needs to be replaced.

Multipath: The audio signal generated by the tooth resonance propagates in the user's mouth taking both Line-of-Sight (LOS) and non-Line-of-Sight (nLOS) reflective paths. This causes potential interference to tooth

resonance signatures. *ToMoBrush* mitigates this problem by affixing the microphone as close as possible to the toothbrush (1.25 cm shown in Fig. 10a), which makes the LOS path much more dominant than other nLOS paths. Besides, as *ToMoBrush* only seeks to detect changes, we can make a reasonable assumption that the multipath changes inside one's mouth are relatively small even across days, as one's brushing habits (mouth openness) and the brushing environments are generally consistent for most of the users. Our latest design shown in Fig. 19a with a closer spacing between the tooth and the microphone can further remedy this issue. Furthermore, other vibrometry sensors such as laser doppler vibrometers [70] have the potential to leverage *ToMoBrush*'s signal processing techniques and develop a multipath-free solution. We leave further investigations for multipath mitigation to future work.

Microphone vs. Vibration Sensor: One may wonder why *ToMoBrush* chooses to use a microphone instead of a vibration sensor to capture the acoustic signals from the tooth, as the vibration sensor is more resilient to multipath and ambient noise. Indeed, based on the working principle of the vibration sensor, the sensor is required to have direct contact with the tooth to capture the resonance signals with enough SNR. However, touching the teeth with a hard sensor will cause discomfort to users, especially when the tooth is vibrating.

Other dental conditions: We believe that *ToMoBrush* can potentially be extended to detect dental conditions beyond those described in this paper. Technically, *ToMoBrush* can sense any dental disease that causes changes to the teeth resonances across a wide frequency range. For example, teeth problems such as abscesses and gum diseases such as periodontitis can potentially be diagnosed by *ToMoBrush*. We leave a more comprehensive evaluation of the potentials of sensing based on *ToMoBrush*'s platform for future work.

Braces, bridges and implants: We note that *ToMoBrush* was not comprehensively evaluated in the presence of braces, bridges, and implants. We acknowledge that these require more carefully targeted user studies that are challenging to mount for all types of dental attachments. In general, we believe that *ToMoBrush* can detect changes in acoustic signatures of a wide range of surfaces that the brush comes in contact with. That said, braces may pose a unique challenge given that slight fluctuations in their position may also influence teeth signature. We leave a thorough investigation of this to future work.

Incorporate other sensing modalities: As discussed in the related work section, multiple recent works propose to include sensing technologies into toothbrushing procedure, such as using IMU [11, 12, 32, 42] or magnetic field [33] for toothbrush tracking, as well as integrating blue-violet light camera on toothbrush head to monitor plaque removal [73]. We believe the platform *ToMoBrush* has great potential to incorporate these sensing modalities to support better toothbrush tracking as well as health sensing.

11 CONCLUSION

This paper presents *ToMoBrush*, an in-house toothbrush-based solution for detecting common dental diseases such as caries, calculus and food impaction, and achieves an average detection ROC-AUC of 0.91, 0.83 and 0.88 respectively on a user study with 19 participants. *ToMoBrush* achieves these objectives by disentangling the behavior of the teeth from the complex acoustic environment of the mouth and identifying important features that affect these conditions. There remain several open challenges for future researchers to build on *ToMoBrush* to detect new diseases. We further believe future work can develop data-driven inference techniques to improve accuracy and new smart toothbrush designs that co-optimize sensing and cleaning.

REFERENCES

[1] [n. d.]. ([n. d.]). https://www.ncbi.nlm.nih.gov/books/NBK574510/
[2] [n. d.]. ([n. d.]). https://www.kicktraq.com/projects/oralvue/smart-video-toothbrush-with-built-in-camera
[3] [n. d.]. ([n. d.]). https://labelbox.com/
[4] [n. d.]. Brushmo compact replacement toothbrush heads compatible for HX6023 to use with Philips Sonicare electric toothbrush, 8 pack. ([n. d.]). https://brushmo.com/products/brushmo-compact-replacement-toothbrush-heads-for-philips-sonicare-hx6023-proresults-white
[5] [n. d.]. Dental Caries in Adults (Ages 20 to 64 Years). ([n. d.]). https://www.nidcr.nih.gov/research/data-statistics/dental-caries/adults
[6] [n. d.]. Voice Technologies: Swiss precision for excellent audio quality - VT500X. ([n. d.]). https://www.vt-switzerland.com/en/vt-500x
[7] 2000. Acoustic resonance analysis in manufacturing. https://www.ndt.net/article/v05n01/hertlin/hertlin.htm. (2000).
[8] 2017. Buy the sonicare sonicare protectiveclean 6100 Sonic Electric Toothbrush HX6877/21 Sonic Electric Toothbrush. (Dec 2017). https://www.usa.philips.com/c-p/HX6877_21/sonicare-protectiveclean-6100-sonic-electric-toothbrush
[9] 2021. (May 2021). https://www.usa.philips.com/c-p/HX3681_23/4100-series-sonic-electric-toothbrush
[10] 2021. Oral Health Fast Facts. (Jan 2021). https://www.cdc.gov/oralhealth/fast-facts/index.html
[11] Sayma Akther, Nazir Saleheen, Mithun Saha, Vivek Shetty, and Santosh Kumar. 2021. MTeeth: Identifying Brushing Teeth Surfaces Using Wrist-Worn Inertial Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 53 (jun 2021), 25 pages. https://doi.org/10.1145/3463494
[12] 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. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1, Article 1 (mar 2019), 25 pages. https://doi.org/10.1145/3314388
[13] Keith Angelino, Pratik Shah, David A Edlund, Mrinal Mohit, and Gregory Yauney. 2017. Clinical validation and assessment of a modular fluorescent imaging system and algorithm for rapid detection and quantification of dental plaque. *BMC oral health* 17 (2017), 1–10.
[14] Suresh Balakrishnama and Aravind Ganapathiraju. 1998. Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing* 18, 1998 (1998), 1–8.
[15] Gaoshuai Cao, Kuang Yuan, Jie Xiong, Panlong Yang, Yubo Yan, Hao Zhou, and Xiang-Yang Li. 2020. Earphonetrack: involving earphones into the ecosystem of acoustic motion tracking. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 95–108.
[16] Keith HS Chan, Yanjie Mai, Harry Kim, Keith CT Tong, Desmond Ng, and Jimmy CM Hsiao. 2010. Resin composite filling. *Materials* 3, 2 (2010), 1228–1243.
[17] 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 (CHI '08)*. Association for Computing Machinery, New York, NY, USA, 363–372. https://doi.org/10.1145/1357054.1357115
[18] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. 2020. Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems* 33 (2020), 22243–22255.
[19] Caroline Chirichella. 2022. Cavity between teeth: What you should know. (Feb 2022). https://www.verywellhealth.com/cavity-between-teeth-5215707
[20] Robin Clark. 2004. Rail flaw detection: overview and needs for future developments. *Ndt & E International* 37, 2 (2004), 111–118.
[21] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning* 20 (1995), 273–297.
[22] Paul Deep. 2000. Screening for common oral diseases. *Journal-Canadian Dental Association* 66, 6 (2000), 298–299.
[23] Xiaoran Fan, Longfei Shangguan, Siddharth Rupavatharam, Yanyong Zhang, Jie Xiong, Yunfei Ma, and Richard Howard. 2021. HeadFi: Bringing Intelligence to All Headphones. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking (MobiCom '21)*. Association for Computing Machinery, New York, NY, USA, 147–159. https://doi.org/10.1145/3447993.3448624
[24] Roger Forshaw. 2022. Dental calculus-oral health, forensic studies and archaeology: a review. *British Dental Journal* 233, 11 (2022), 961–967.
[25] Akshay Gadre, Deepak Vasisht, Nikunj Raghuvanshi, Bodhi Priyantha, Manikanta Kotaru, Swarun Kumar, and Ranveer Chandra. 2022. MiLTOn: Sensing Product Integrity without Opening the Box using Non-Invasive Acoustic Vibrometry. In *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 390–402.
[26] N Galili, I Shmulevich, and N Benichou. 1998. Acoustic testing of avocado for fruit ripeness evaluation. *Transactions of the ASAE* 41, 2 (1998), 399.
[27] Nisha Garg and Amit Garg. 2010. *Textbook of operative dentistry*. Boydell & Brewer Ltd.
[28] Z Gong, EO Nyborg, and G Oommen. 1992. Acoustic emission monitoring of steel railroad bridges. *Materials evaluation* 50, 7 (1992).
[29] Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. 2006. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning*. 369–376.

[30] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.
[31] Raimund Hibst, Robert Paulus, and Adrian Lussi. 2001. Detection of occlusal caries by laser fluorescence: basic and clinical investigations. *Medical Laser Application* 16, 3 (2001), 205–213.
[32] 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 (SenSys '16)*. Association for Computing Machinery, New York, NY, USA, 202–215. https://doi.org/10.1145/2994551.2994563
[33] Hua Huang and Shan Lin. 2020. *MET: A Magneto-Inductive Sensing Based Electric Toothbrushing Monitoring System*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3372224.3380896
[34] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* 454, 1971 (1998), 903–995.
[35] Yandao Huang, Minghui Qiu, Lin Chen, Zhencan Peng, Qian Zhang, and Kaishun Wu. 2023. NF-Heart: A Near-Field Non-Contact Continuous User Authentication System via Ballistocardiogram. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 1, Article 16 (mar 2023), 24 pages. https://doi.org/10.1145/3580851
[36] Amid I Ismail, Woosung Sohn, Marisol Tellez, Ashley Amaya, Ananda Sen, Hana Hasson, and Nigel Berry Pitts. 2007. The International Caries Detection and Assessment System (ICDAS): an integrated system for measuring dental caries. *Community dentistry and oral epidemiology* 35, 3 (2007), 170–178.
[37] Mayur Khairnar. 2013. Classification of food impaction-revisited and its management. *Indian Journal of Dental Advancements* 5, 1 (2013), 1113–1119.
[38] 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 (UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 109–120. https://doi.org/10.1145/2750858.2804259
[39] Niklaus P Lang. 1982. Periodontal epidemiological indices for children and adolescents: II. Evaluation of oral hygiene; III. Clinical applications. *Pediatr Dent* 4, 1 (1982), 64–73.
[40] Sangdae Lee and Byoung-Kwan Cho. 2013. Evaluation of the firmness measurement of fruit by using a non-contact ultrasonic technique. In *2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 1331–1336.
[41] Yuan Liang, Hsuan Wei Fan, Zhujun Fang, Leiying Miao, Wen Li, Xuan Zhang, Weibin Sun, Kun Wang, Lei He, and Xiang 'Anthony' Chen. 2020. OralCam: Enabling Self-Examination and Awareness of Oral Health Using a Smartphone Camera. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376238
[42] 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*. 1234–1242. https://doi.org/10.1109/INFOCOM.2019.8737513
[43] Cory Myers, Lawrence Rabiner, and Aaron Rosenberg. 1980. Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28, 6 (1980), 623–635.
[44] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Jacob E Sunshine. 2019. Opioid overdose detection using smartphones. *Science translational medicine* 11, 474 (2019), eaau8914.
[45] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. 2015. Contactless sleep apnea detection on smartphones. In *Proceedings of the 13th annual international conference on mobile systems, applications, and services*. 45–57.
[46] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. 2016. Fingerio: Using active sonar for fine-grained finger tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1515–1525.
[47] 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*. 1–6. https://doi.org/10.1109/GLOCOM.2017.8255017
[48] Nigel B Pitts, Domenick T Zero, Phil D Marsh, Kim Ekstrand, Jane A Weintraub, Francisco Ramos-Gomez, Junji Tagami, Svante Twetman, Georgios Tsakos, and Amid Ismail. 2017. Dental caries. *Nature reviews Disease primers* 3, 1 (2017), 1–16.
[49] IA Pretty, WM Edgar, PW Smith, and SM Higham. 2005. Quantification of dental plaque in the research environment. *Journal of dentistry* 33, 3 (2005), 193–207.
[50] Iain A Pretty. 2006. Caries detection and diagnosis: novel technologies. *Journal of dentistry* 34, 10 (2006), 727–739.
[51] P Rechmann, Shasan W Liou, Beate MT Rechmann, and John DB Featherstone. 2014. SOPROCARE-450 nm wavelength detection tool for microbial plaque and gingival inflammation: a clinical study. In *Lasers in Dentistry XX*, Vol. 8929. SPIE, 42–48.
[52] Hiroaki Sakoe and Seibi Chiba. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing* 26, 1 (1978), 43–49.

[53] Nozha Sawan, Afnan Ben Gassem, Faisal Alkhayyal, Aroob Albakri, Nada Al-Muhareb, and Eman Alsagob. 2022. Effectiveness of Super Floss and Water Flosser in Plaque Removal for Patients Undergoing Orthodontic Treatment: A Randomized Controlled Trial. *International Journal of Dentistry* 2022 (2022).
[54] Bernhard Schölkopf, Robert C Williamson, Alex Smola, John Shawe-Taylor, and John Platt. 1999. Support vector method for novelty detection. *Advances in neural information processing systems* 12 (1999).
[55] Sarah Schotte, Nele De Belie, and Josse De Baerdemaeker. 1999. Acoustic impulse-response technique for evaluation and modelling of firmness of tomato fruit. *Postharvest biology and technology* 17, 2 (1999), 105–115.
[56] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 815–823.
[57] Manuja Sharma, Lauren K Lee, Matthew D Carson, David S Park, Se W An, Micah G Bovenkamp, Jess J Cayetano, Ian A Berude, Leonard Y Nelson, Zheng Xu, et al. 2022. O-pH: Optical pH Monitor to Measure Dental Biofilm Acidity and Assist in Enamel Health Monitoring. *IEEE Transactions on Biomedical Engineering* 69, 9 (2022), 2776–2786.
[58] Avinashchandra Singh, Donald R Houser, and Sandeep Vijayakar. 1998. Detecting gear tooth breakage using acoustic emission: a feasibility and sensor placement study. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 80388. American Society of Mechanical Engineers, V008T08A016.
[59] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. 2020. SpiroSonic: monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
[60] Staff and Staff. 2023. Are cavities inside teeth possible? (Feb 2023). https://dentalhealthsociety.com/oral-health/are-cavities-inside-teeth-possible/#:~:text=However%2C%20there%20are%20times%20when,an%20injury%20to%20the%20tooth.
[61] Ke Sun, Ting Zhao, Wei Wang, and Lei Xie. 2018. Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals. In *Proceedings of the 24th annual international conference on mobile computing and networking*. 591–605.
[62] 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.
[63] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).
[64] Anran Wang, Jacob E. Sunshine, and Shyamnath Gollakota. 2019. Contactless Infant Monitoring Using White Noise. In *The 25th Annual International Conference on Mobile Computing and Networking (MobiCom '19)*. Association for Computing Machinery, New York, NY, USA, Article 52, 16 pages. https://doi.org/10.1145/3300061.3345453
[65] Lei Wang, Kang Huang, Ke Sun, Wei Wang, Chen Tian, Lei Xie, and Qing Gu. 2018. Unlock with your heart: Heartbeat-based authentication on commercial mobile phones. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 2, 3 (2018), 1–22.
[66] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW based contactless respiration detection using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–20.
[67] Wikipedia contributors. 2022. Dental radiography — Wikipedia, The Free Encyclopedia. (2022). https://en.wikipedia.org/w/index.php?title=Dental_radiography&oldid=1114320192 [Online; accessed 19-November-2022].
[68] Wikipedia contributors. 2023. Cepstrum — Wikipedia, The Free Encyclopedia. (2023). https://en.wikipedia.org/w/index.php?title=Cepstrum&oldid=1144605797 [Online; accessed 15-March-2023].
[69] Wikipedia contributors. 2023. Frequency response — Wikipedia, The Free Encyclopedia. (2023). https://en.wikipedia.org/w/index.php?title=Frequency_response&oldid=1171077560 [Online; accessed 14-September-2023].
[70] Wikipedia contributors. 2023. Laser Doppler vibrometer — Wikipedia, The Free Encyclopedia. (2023). https://en.wikipedia.org/w/index.php?title=Laser_Doppler_vibrometer&oldid=1188672119 [Online; accessed 23-April-2024].
[71] Allen Wong, Paul E Subar, and Douglas A Young. 2017. Dental caries: an update on dental trends and therapy. *Advances in pediatrics* 64, 1 (2017), 307–330.
[72] Xiangyu Xu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. Breathlistener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. 54–66.
[73] Takuma Yoshitani, Masa Ogata, and Koji Yatani. 2016. LumiO: A Plaque-Aware Toothbrush. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 605–615. https://doi.org/10.1145/2971648.2971704
[74] Christian Zakian, Iain Pretty, and Roger Ellwood. 2009. Near-infared hyperspectral imaging of teeth for dental caries detection. *Journal of biomedical optics* 14, 6 (2009), 064047–064047.
[75] Fusang Zhang, Zhi Wang, Beihong Jin, Jie Xiong, and Daqing Zhang. 2020. Your Smart Speaker Can" Hear" Your Heartbeat! *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–24.