Bitey: An Exploration of Tooth Click Gestures for Hands-Free User Interface Control

Daniel Ashbrook[1], Carlos Tejada[1], Dhwanit Mehta[1], Anthony Jiminez[1], Goudam Muralitharam[1], Sangeeta Gajendra[2], Ross Tallents[2]

[1]Golisano College of Computing and Information Sciences
Rochester Institute of Technnology
Rochester, NY, USA
{daniel.ashbrook, cet1318,
dmm8396, aj7794, gjm6993}@rit.edu

[2]Eastman School of Dental Studies
University of Rochester
Rochester, NY, USA
{sangeeta_gajendra, ross_tallents}
@urmc.rochester.edu

ABSTRACT

We present *Bitey*, a subtle, wearable device for enabling input via tooth clicks. Based on a bone-conduction microphone worn just above the ears, Bitey recognizes the click sounds from up to five different pairs of teeth, allowing fully hands-free interface control. We explore the space of tooth input and show that Bitey allows for a high degree of accuracy in distinguishing between different tooth clicks, with up to 94% accuracy under laboratory conditions for five different tooth pairs. Finally, we illustrate Bitey's potential through two demonstration applications: a list navigation and selection interface and a keyboard input method.

Author Keywords

Bio-acoustics; tooth input; gestures; wearable computing; subtle interfaces; audio interfaces.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: Input devices and strategies

INTRODUCTION

Modern mobile devices such as smartphones, smart watches, and other wearable computing devices, while enormously capable, require a high degree of explicit attention from the user for both input and output. A number of interfaces have been proposed to reduce the amount of interaction necessary for short tasks [4, 15, 33]; however, most of these interfaces still require the use of the hands, which may not be feasible in all cases (e.g., impairments, disabilities, or situational impairments [39] such as carrying objects [32]). One solution to this issue is speech recognition, such as used in the Apple Watch or Google Glass; however, in many situations, speech is socially disruptive. Speech may also not be appropriate for cases in which *microinteractions*—very short interactions such as dismissing a phone call [4]—are desired. Other options such as augmenting the surface of the body [16] are attractive but still require the user to have at least one hand unoccupied.

Figure 1. The bone-conduction microphone used in Bitey (bottom) and its nearly invisible positioning on the temporal bone just above the ear (top).

In this paper, we present Bitey, a system that detects and classifies tooth clicking sounds to allow hands-free input to an interface. Bitey allows a user to operate an interface instantly: to initiate a process such as a phone call, to respond to a notification, or to control an ongoing process such as music playback. Although other tooth click-based interfaces have been proposed previously (see *Related Work*), all detect only whether *any* pair of teeth was clicked; in contrast, Bitey greatly expands the interaction capabilities by classifying *which* pair of teeth was clicked.

To appear in the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI), September 6–9 2016, Florence, Italy.

As a simple example, imagine using Bitey to control music playback while exercising on a rowing machine. In this situation, speech is not practical due to the intense nature of the activity, and both hands are fully engaged in using the rowing machine. Using Bitey, one could simply click the right canine teeth together to switch to the next song, or click together the left canine and left bicuspid to lower the volume.

The hardware used in Bitey is minimal—simply a bone-conduction microphone which is worn discreetly above the ears and can easily be integrated into a head-worn display or a hat for complete invisibility. In contrast to speech-based interfaces, Bitey is always listening, requiring no activation sequence (e.g., "Okay, Glass"), is completely hands-free, and provides discrete access to multiple functions without the need for visual feedback.

The core contributions of our research are as follows:

1. we describe the use of inexpensive off-the-shelf hardware to reliably record tooth click sounds;
2. we describe methods for differentiating between the click sounds produced by different pairs of teeth;
3. we evaluate the performance of our recognition techniques with respect to the amount of training data needed, evolution of the characteristics of the input over time, and under non-ideal circumstances such as while the user is in motion or speaking;
4. we explore the sensitivity of our system to microphone placement;
5. and we demonstrate several example applications of our input technique.

Terminology

We briefly define some of the potentially unfamiliar terms used in this paper: *auscultation* is the act of listening to the internal sounds of the body, often for medial diagnosis; *gnathosonics* is the study of the sounds produced by the teeth and jaws [49]; *occlusion* is the position of the teeth when the jaws are closed; and, although the term *occlusogram* has been used for the visualization of the sound produced by occlusion [9], we use *gnathosonogram* (possibly coined in [30]) to connote the more general sense of a visualization of sounds made by the teeth.

RELATED WORK

Dentistry and gnathosonics

The field of dentistry has long been interested in detecting issues with dentition based on the sounds of tooth clicks. In 1953, Stewart introduced the idea of using auscultation to diagnose issues with occlusion [44], and noted that different teeth made different sounds based on their shapes. Starting in the 1960s [48, 49], Watt began an investigation of what he termed "gnathosonics," describing the study of sounds made by the teeth in order to diagnose issues with occlusion. In subsequent work [50, 51, 52], he defined a three classes based on the duration of the tooth click sound, essentially categorizing teeth contacts into "normal," "some abnormal" and "all abnormal". This early work and others contemporaneous [8, 10] were mainly limited to the dental professional directly listening to the sounds of tooth contact or visually inspecting an occlusogram.

In the late 1980s, gnathosonic research involving computers began to appear. Fuller and West investigated extracting simple features from tooth clicks in order to classify the clicks into Watt's categories [13]; they used the duration of the click, amplitude of the sound, and the duration of the initial high-frequency segment of the sound, but with only limited success. Teodorescu et al. made an early attempt at automatic analysis of gnathosonic sounds with analog circuitry [45]. Shi et al. [40, 41] used a piezoelectric transducer on the cheekbone to record tooth click sounds and classified them into Watt's categories using an autoregressive model.

Auscultation of the teeth has been used for other purposes beyond classification into Watt's categories: Hędzelek and Hornowski placed accelerometers on the side of the bone next to the eyes and analyzed the Fourier transform for three groups of patients with differing pathologies [17], while Prinz, using headphone as microphones, used visual inspections of different transformations of occlusograms to distinguish between single and multiple tooth clicks after dental work [35].

Tooth click interfaces

In the computing literature, there has been some limited work in using tooth clicking as input to human-computer interfaces, mainly with the aim to control assistive technology. None of this prior work attempts to differentiate among clicks from different teeth. Kuzume used an in-ear bone-conduction microphone to detect tooth click sounds using the FFT of the sound, also implementing a voice rejection algorithm [21, 24]. Zhong et al. detected tooth clicks and rejected voice via similar methods, also using a bone-conduction microphone [31, 56].

Multiple researchers [22, 23, 36, 42, 43, 55] have implemented tooth click-based interfaces for assistive technology, wherein a selection is made via some non-tooth click mechanism and then confirmed via a single or double tooth click. One exception is in Zhong et al. [56] where selection is via double-clicking and confirmation is via a single click. No prior work investigates disambiguating between teeth for a higher degree of control.

Although all of this work explores how to reject false positives due to speaking, none investigates the usability of a tooth-click interface in a non-lab setting such as while the user is walking, which can cause noise from movement of the microphone on the head. Additionally, while Zhong et al. acknowledge that microphone placement can be important [56], no previous research explores how changing the microphone location over time influences the performance of the system.

Auscultation-based interfaces

Gnathosonics, and therefore Bitey, is an example of auscultation, or the act of listening to the internal sounds of the body. Auscultation has a long history in the medical field, but also has history in the HCI and ubiquitous computing communities. Yatani and Truong [53] and Rahman et al. [37] experimented with auscultation hardware for recognition of various body-centric activities such as eating, drinking, speaking, coughing, and different breathing patterns.

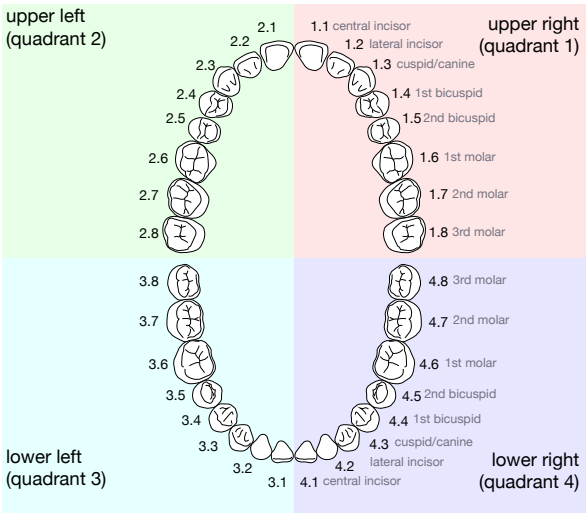


Figure 2. ISO 3950 system for tooth numbering. The view is from inside the mouth. Each tooth is encoded as <quadrant number>.<tooth number>.

(a) A (b) R33

(c) R34 (d) LR11

Figure 3. A selection of different types of clicks.

Specifically in the oral domain, Amft et al. used the sounds of chewing as transmitted through the ear canal to automatically detect eating activity and classify the type of food being consumed [2, 3]. Li et al. considered embedding sensors directly into teeth [26], and were able to recognize a number of mouth activities such as eating, speaking, and coughing.

Amento et al. [1] and Deyle et al. [12] both experimented with recognizing hand gestures via sounds transmitted through the bones of the hand to the wrist. Harrison et al. sensed taps on various arm locations via an upper arm-mounted resonator box [16], and Zhong et al. demonstrated the transmission of information through the human skeletal system via bone-conducted sound [56].

Hands- and eyes-free interfaces

Interfaces that may be used hands- and eyes-free, especially those avoiding the possible social acceptability issues with speech, are a common subject of research in the HCI and assistive technology communities. While much eyes-free interface work considers unobtrusive gesture [5] or touch [7,33,34,54], other work has, similar to Bitey, investigated face- or head-based input to enable entirely hands-free interaction.

Chin et al. used eye tracking for cursor control with a jaw clench for "click" activation [11], and Tuisku et al. used facial movements to activate a gaze-controlled cursor [46]; both of these projects required sensors to be attached directly to the user's face. Rather than controlling a cursor directly, Sahni et al. mounted a magnet to a user's tongue and recognized silent speech via a head-mounted magnetometer [38]. Goel et al. [14] and Li et al. [27] both developed systems to non-invasively capture the muscle movement of the tongue while performing various gestures. Although Bitey must also add sensors to the head in order to detect tooth click sounds, its hardware, located just above the ears, is much less intrusive and visible than those proposed in other face-based interfaces.

BITEY OVERVIEW

The idea behind Bitey is simple: to enable hands-free, subtle input via tooth clicks. Previous work on tooth clicking [21, 22, 23, 24, 42, 43, 55, 56] has enabled single-click interfaces, but has not investigated recognizing clicks from different teeth, rejecting non-speech-related sounds, nor the effect microphone placement has on system performance. In order to enable Bitey to be useful for a variety of tasks for both assistive and non-assistive uses, we have extended tooth click detection to encompass these additional capabilities.

Tooth click notation

To describe different possible tooth clicks, we adapt into a more compact format the ISO standard 3950:2009 notation for tooth numbering [19], illustrated in Figure 2. The syntax adopted in this paper is <side><top tooth number><bottom tooth number>. For example, a click of the right canines (1.3 and 4.3 in the Figure) becomes R33. Clicking the top left canine (1.3) to the bottom left first bicuspid (3.4) is L34. We use X to indicate either side (e.g., tapping two canines is X33), LR to indicate both sides (e.g., tapping all incisors is LR11), and A to indicate all teeth (i.e. an entire mouth "chomp"). Figure 3 illustrates the tooth contacts for some common clicks.

Although in this work we concentrate on characterizing Bitey's performance with simple clicks, this system of notation also lets us express more complex gestures, such as tapping the right canines then sliding the lower jaw forward so the canine touches the 1st bicuspid: R33–34. A multiple tap is prefixed with a number: 2R33 is a double click of the right canines.

SENSING

With Bitey, we wanted to enable an unobtrusive, inexpensive mechanism for tooth click detection. We used a monaural piezoelectric throat microphone sold online under various brand names (e.g., Zeadio ZP-AR201) for about USD 9.

Other tooth-click detection work has placed the microphone in a variety of locations. Dentists using gnathosonics for diagnosis have used the center of the forehead [52], the orbits of the eyes [17, 49], or the cheekbone [40, 41, 44, 48]; however, these locations are not socially subtle and could be uncomfortable for longer-term use. Prior computing interface research has used the throat [56], the tragus of the ear [42], and the ear canal [24]; again, however, these locations may suffer from the same problems for longer-term use as gnathosonic devices. With Bitey, we place the device such that the microphone rests just above the ears, on the temporal bone (Figure 1).

Previous research [17, 25, 35, 40, 56] suggested, and our pilot testing confirmed, that the dominant frequencies of tooth clicks occur well below 4000Hz; therefore, we set our recording rate to 8000Hz. To collect training data, we used the free audio recording program Audacity, while we implemented live testing using the Python package PyAudio.

SEGMENTATION AND FEATURE EXTRACTION

To detect and classify clicks in a stream of audio data from the bone-conduction microphone, we first segment the data into potential click candidates. We apply an empirically determined amplitude threshold A_c to the data stream to find places where the signal is loud enough to potentially constitute a click. We then back off a small amount T_b to ensure that we capture the beginning of the signal and then take a pre-determined duration of time T_d to form the potential click candidate.

The literature contains some agreement about the proper value of T_d. Considering the sound of occlusion (that is, the entire mouth closing), Watt considered "stable" (class A) occlusal sounds to have durations under 30 ms [51]. In their implementation of a tooth click interface, Zhong et al. used 23.3 ms as their click duration [56]. Kuzume and Morimoto [24] found click times under 5 ms, but only used the main body of the sound itself without considering time required for the amplitude to fall to baseline. Because the falloff of the amplitude and frequency of bone-conducted sound will vary with the placement of the recording device [20], the mild variation in the literature is unsurprising. Our own results are in general agreement, however: on average, the main body of the click is about 10 ms. To ensure we fully capture the sound, we set T_d to 20ms, or 160 samples. We set the back off time T_b to 2ms (16 samples). Figure 4(a)–(d) illustrates gnathosonograms extracted using this method.

For each candidate click, we generate a simple set of features: the FFT of the entire candidate (81 features) normalized by the maximum FFT value, and the mean, median, standard deviation, sum, minimum, and maximum of the raw signal (six features). We use these 95 features as input to a support vector machine (SVM) with an RBF kernel (implemented on a MacBook Pro laptop in Python using the scikit-learn library). Figure 4(e)–(h) shows FFT features for clicks from one tooth.

EXPERIMENTS

In order to evaluate the effectiveness of our system, we recruited twenty participants (ten female). Our goal was to understand the performance of Bitey, both under laboratory and more realistic conditions. The questions we investigated were:

- What is the best-case performance of the system—that is, the case where false positive-generating events are minimized, and a large amount of training data can be used? How many teeth pairs can be recognized and with what performance?
- How effectively can we reject false positive-causing events such as coughing, talking, chewing, moving, walking, and so forth?
- How much training data is necessary? In a system used in everyday life, we would want to minimize the amount of training data that we require the user to provide.
- How sensitive is the system to sensor placement? In actual use, a user would take off the microphone each day and put it back on the next; would the slight changes in placement cause degraded recognition performance? If so, how many different placements would be necessary to mitigate the performance decrease?

We do not test the system's user-dependent vs. user-independent performance (with one exception—see Section *False positives and false negatives*). The variety in tooth clicking abilities varies widely between people with the shape of the mouth, the motility of the jaw, and the condition of the teeth. Table 1 shows the variety of pairs of teeth our study participants were comfortable clicking.

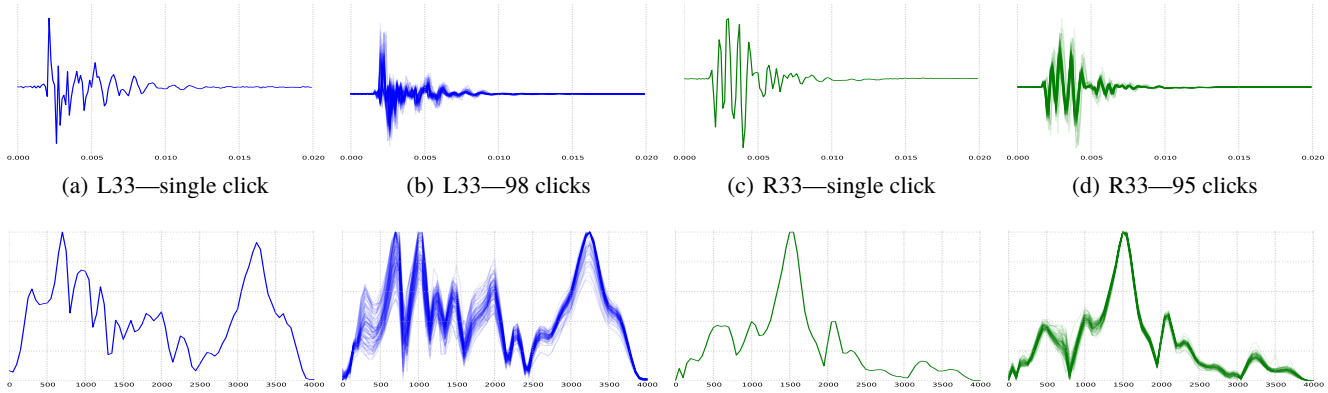
Data collection

We recorded data in quiet conditions in our lab. Before the initial recording, we informed each participant of the purpose of the study, the principle of tooth clicking for computer control, and explained how the recording apparatus worked. We asked the participant to experiment with clicking different teeth until they were able to find several pairs that they could comfortably click repeatedly, which we then noted; most participants could click three different pairs, although several could click four or five.

We then started a live waveform view of the recording and asked the participant to don the recording headset and click various teeth in order to see the effects, and to allow us to calibrate the recording volume (this functionality could be provided via auto gain control in a complete system). Once the participant felt comfortable with the system, we started recording. For each pair of teeth to be clicked, we asked the participant to tap them together repeatedly for about 20 seconds, saving the click waveforms for each pair of teeth in separate files. We ensured that the participant sat still in order to prevent noise artifacts in the data.

After each recording session, we asked the participant to remove the microphone, take a brief break, and then to replace the device for another session. Each session we refer to as a *placement*. For each participant, we collected between four and fifteen different placements (depending on participant availability), some over the course of several days.

We also wanted to test our system's robustness against noise; therefore, we also asked participants to generate data that might cause false positives—non-click sounds mistakenly recognized as tooth clicks. We requested them talk for 60 seconds, and



(a) L33—single click (b) L33—98 clicks (c) R33—single click (d) R33—95 clicks

(e) L33 FFT—single click (f) L33 FFT—98 clicks (g) R33 FFT—single click (h) R33 FFT—95 clicks

Figure 4. Images (a)–(d) show the gnathosonograms produced by tooth clicks (x-axis seconds) while images (e)–(h) illustrate the corresponding normalized FFTs (x-axis Hz). Images (a), (c), (e) and (g) illustrate single clicks, while (b), (d), (f) and (h) show multiple clicks overlaid.

to move around for 60 seconds. Both of these activities are very common and could be expected to cause issues for the system: talking—due to head-generated sound possibly containing spurious tooth clicks [28]—and movement due to sound transmitted through the microphone cable as it slides and moves against the wearer's clothing. On average, we cumulatively recorded approximately five minutes of noise data from each participant.

Table 1 summarizes the data we collected for each participant: the pairs of teeth the participant was able to click together comfortably and repeatedly, the number of different microphone placements in which we collected data from that participant, the number of segments of noise data found by the click detection algorithm (which could possibly be falsely classified as tooth clicks), and the amount of noise data we recorded.

RESULTS

Best-case performance

Although a system such as Bitey would normally be used under non-ideal circumstances—that is, during everyday life with attendant noise that could cause false positives—it is valuable to test the system under controlled conditions to get an idea of its best possible performance. Our data was collected in quiet lab conditions with no added noise from movement, speech, coughing, or other sounds. To evaluate the data in a realistic manner, we worked under the assumption that the user would train the system once and then use it later, after removing and replacing the microphone. Traditional N-fold cross validation divides the data into N folds, trains on $N-1$ and tests on the remaining N^{th}. However, with our data, this method has a high likelihood of training and testing with data from the same placement. Therefore, we treat each separate placement as a fold, leaving one out for testing, and training on the remaining placements. The results of this best-case test are illustrated in Table 1 (column Acc) and Figure 5. The results are promising: we see a mean accuracy of 78% ($STD = 15\%$), with some participants' data reaching above 90%. There appears to be no correlation between the number of teeth clickable or the number of placements and the accuracy achieved.

False positives and false negatives

We tested the system's performance with false positives (non-click data incorrectly recognized as clicking) and false negatives (clicks incorrectly recognized as noise). For each participant, we ran the click detection algorithm on that participant's noise data (movement and speaking) and combined the results into one "Noise" class. We included this Noise class with the other tooth click data and proceeded as in the best-case performance section above, testing via five-fold cross validation. The results are illustrated in the Acc_{Noise} column of Table 1 and in Figure 5.

Across all participants, we collected 1.64 hours of noise data, from which our click detection algorithm found 83,942 potential click candidates. Of these, just 674 (.8%) were incorrectly classified as clicks, while 5,255 of 44,324 intentional clicks (11.9%) were incorrectly classified as noise. This ratio of few false positives to more false negatives is desirable in an interactive system such as this: requiring the user to repeat an input is better than performing an action when no action as been requested. Table 1 summarizes this data per participant via *precision*—the ratio of the number of actual clicks recorded by the participant and detected by Bitey (true positives) to the total number of clicks (including false positives) detected by Bitey—and *recall*—the ratio of true positives to the total number of clicks recorded by the participant.

We also tested to see if the characteristics of the noise data differed between people. Having a pre-trained user-independent noise class would simplify the training procedure by not requiring users to train the system on noise. To determine the feasibility of this idea, we collapsed all of a participant's noise data into a single class with the label of the given participant. We then performed five-fold cross-validation on the noise classes from all participants. Our overall accuracy was surprisingly high at 69%, suggesting that noise is somewhat user-dependent. One possible reason is that, with a wide variety of body types represented amongst our participants, the distribution of absorption of different noise frequencies may vary predictably between people.

PN	*Tooth pairs*	N_{place}	N_{clicks}	N_{noise}	T_{noise}	*Acc*	Acc_{noise}	*P*	*R*
1	L88, LR11, R88	17	2721	3597	6:31	81.5	87.7	97.6	85.1
2	L33, L77, LR11, R33	11	2594	3691	2:36	95.9	96.4	99.7	94.8
3	L33, L44, LR11, R33, R44	6	1709	2700	8:03	93.5	96.4	99.1	95.7
4	L66, L77, LR11, R66	6	2860	12572	7:44	52.3	87.9	96.9	74.4
5	L33, LR11, R33	10	1355	1130	1:05	90.1	92.7	99.1	93.0
6	L33, L44, L66, LR11, R67	12	4871	1954	1:13	75.9	78.1	98.6	93.6
7	L33, L34, LR11, R33, R34	9	3209	716	2:02	93.0	92.9	99.7	96.0
8	L88, LR11, R88	8	2167	6737	3:57	74.5	91.0	97.3	86.9
9	L11, R11, R33	5	1186	7658	7:11	56.7	91.9	98.1	74.9
10	L33, L34, LR11, R33, R34	4	748	1441	1:35	68.9	87.4	99.7	91.6
11	A, L33, LR11, R33, R34	17	6011	1951	9:55	87.4	87.9	99.4	94.9
12	L33, LR11, R33	11	1542	3307	2:45	87.7	94.6	98.5	93.2
13	L34, L76, LR11	5	1205	11293	7:03	64.1	94.9	92.9	75.4
14	L88, LR11	8	1417	7312	3:52	92.4	90.7	91.8	69.7
15	L33, L88, LR11, R33, R88	10	4505	3657	7:13	41.6	66.3	99.8	94.9
16	L33, LR11, R33	4	1285	3937	6:40	78.4	91.7	98.5	86.4
17	L22, L33, R22, R33, R88	4	1159	834	1:14	70.5	76.6	99.4	83.8
18	L33, R11, R44	10	1439	3600	2:39	98.5	98.9	99.3	97.3
19	L33, LR11, R33	13	2891	3852	6:32	82.5	90.4	99.1	88.6
20	L22, L88, LR11, R33	12	2776	4662	8:19	87.5	93.7	99.4	94.3

Table 1. A summary of the data collected and best-case results for each participant in our data collection: the participant (*PN*), the pairs of teeth each study participant was comfortable clicking repeatedly, the number of different placements in which data was collected (N_{place}), the total number of detected clicks in each participant's data set (N_{clicks}), the total number of detected click candidates in each participant's noise set (N_{noise}), the amount of noise data collected for each participant (T_{noise}, minutes), the best-case accuracy without noise (*Acc*), the best-case accuracy with noise (Acc_{noise}), the precision (*P*) and the recall (*R*). Note that accuracy can improve with the Noise class included due to the large number of correctly classified noise examples.

Necessary training data

In a real-world usage scenario, we would prefer to minimize the amount of training data required from the user before the system is ready to use. How much training data is necessary to get a good level of performance? We combinatorially explored how the numbers of training examples and the number of placements training examples were taken from affected recognition performance. We experimented with training sizes T_S from each class of 1, 5, 10, 15, 20, 25, and 30 samples.

As discussed earlier, we have at least four placements of recorded data per study participant. For a given set of placements $P = \{p_1, p_2, \ldots, p_n\}$ of size n, we generated all possible *k-combinations* for k from 1 to $n-1$, by selecting k distinct placements from P. For example, for $n = 4$, we have:

$$\{1\}, \{2\}, \{3\}, \{4\},$$
$$\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\},$$
$$\{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{2,3,4\}$$

We take the k placements in each k-combination as training data and use the remaining $n-k$ placements as test data. However, the total number of k-combinations for n placements is

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

For 17 placements (e.g., participants 1 and 11) and $k = 8$ we have 24,310 possibilities, which is computationally prohibitive to work with. Therefore, we tested up to 100 randomly selected k-combinations of training sizes and groups of placements for k from 1 to $n-1$, in each randomly selecting T_S training samples from each class of the placements, training a model, and testing on all of the data not in the k-combination. For the four-placement example above, one such test might be selecting $T_S = 15$ training examples from each class in combination $\{1, 3, 4\}$, and testing on placement 2.

Following this procedure, we averaged the accuracy for all groups with the same number of placements, yielding a mean accuracy score for each combination of training size T_S and number of placements trained upon.

Our results varied quite widely between participants. In general, however, those with lower performance as reported in Table 1 required more placements with more training data to reach an acceptable level of accuracy. Overall, however, we detected no "magic" number of placements or number of training examples that are necessary. For some participants, accuracy stopped increasing significantly after adding 2–3 placements or 5–10 training examples, but for others accuracy continued to rise as far as we tested.

One possibility for this difference between participants may be in the morphology of the jaw and teeth. In future explorations, we intend to request participants to undergo a dental examination to help us understand what factors can influence the performance of Bitey. However, we speculate that the most

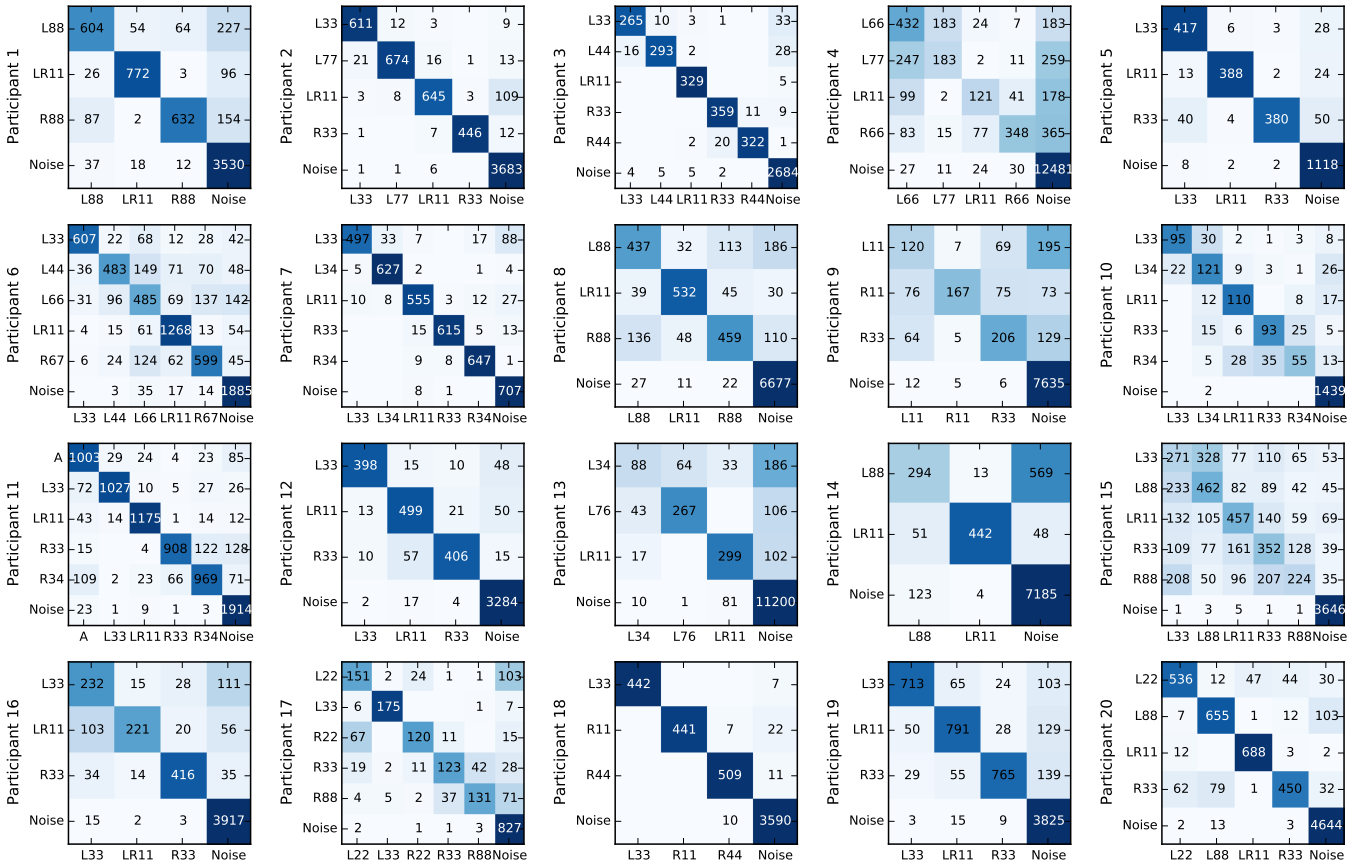


Figure 5. Confusion matrices for each participant based on leave-one-out validation based on placement, including the composite Noise class. The *y* axis on each matrix is the true label and the *x* axis is the predicted label.

improvement in this area will not come algorithmically, but via engineering the microphone to sit more reliably in the same place on the user's head, perhaps by incorporating it into a hat, glasses, or a head-worn display such as Google Glass.

SUPPLEMENTAL EXPERIMENTS

Live testing

Our experiments above were all performed with data collected via participants repeatedly clicking one pair of teeth before moving to the next pair. To test Bitey in a more externally valid scenario, we implemented two simple applications for live testing, which might be appropriate as assistive technology.

The most basic application was a simple list selection task inspired by Zhong et al.'s test, where participants used a single click to advance through a list and a double click to select [56]. Our software prompted the user to use three pairs of teeth to navigate to and select one item out of eleven (Figure 6). One click corresponded to *up*, one to *down*, and one to *select*. Once the user selected the requested item, another random item was requested. Table 2 shows the results for the list selection task.

Additionally, to explore the limits of what Bitey can accomplish, we implemented *BiteWrite*, a tooth-click based text input system. We used a version of MacKenzie's H4-Writer [29], using Huffman codes to assign minimal click sequences to

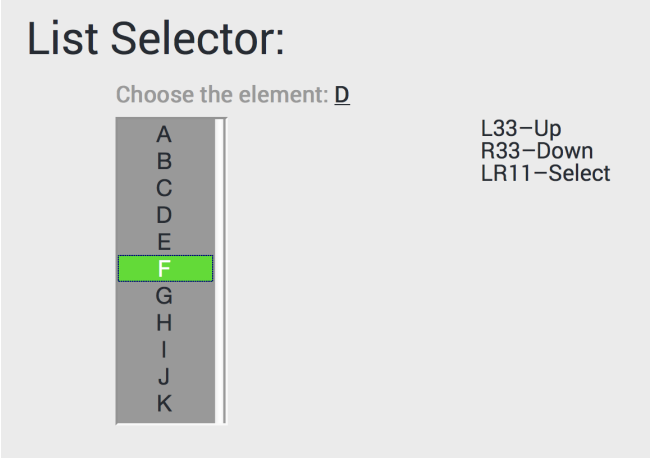


Figure 6. The list selection interface. The user clicks to move the cursor up and down and to select the desired item.

generate letters. For example, two clicks—represented as 0 and 1—generate sequences such as '010' for 'e' and '011001' for 'g'. For a participant able to click N pairs of teeth, we implemented the $N-1$-click version of the keyboard, with the N^{th} click used as "cancel" (in the case of a partially entered sequence) or "backspace" (otherwise). Table 3 displays the

	List Selection			BiteWrite		
PN	*N*	*CpM*	*SpM*	*N*	*SpC*	*Err*
1	5	1.2 (0.2)	2.7 (0.8)	3	9.7 (2.1)	13.0 (4.2)
2	5	1.6 (0.7)	2.2 (1.4)	3	8.3 (1.8)	15.7 (6.8)
6	5	1.6 (1.2)	2.0 (1.4)	—	—	—
8	7	2.9 (1.9)	3.7 (2.5)	1	24.8 (0.0)	10.8 (0.0)
11	5	1.3 (0.5)	2.1 (1.7)	3	6.6 (2.3)	16.6 (9.4)
12	5	1.1 (0.2)	1.4 (0.3)	2	21.1 (4.1)	7.3 (0.3)
13	9	1.1 (0.2)	3.9 (2.4)	—	—	—
19	5	1.3 (0.4)	3.9 (1.6)	3	12.2 (1.1)	10.8 (2.1)
20	5	1.3 (0.3)	1.6 (0.4)	3	12.2 (1.1)	10.8 (2.1)
Overall	*5*	*1.5 (1.0)*	*2.7 (1.9)*	*2*	*11.9 (5.5)*	*12.6 (5.3)*

Table 2. Statistics for the list selection and BiteWrite tests; *PN* is the participant number. For the list selection task, *N* is the number of selection tasks done by the participant, *CpM* shows the mean (std) clicks per move to get a correct selection, and *SpM* is the mean (std) seconds per move to get a correct selection. For BiteWrite, *N* is the number of sentences typed by the participant, *SpC* is the mean (std) number of seconds per character, and *Err* is the mean (std) error rate. The *Overall* row gives the mean (std) values over all participants. Note that participants 6 and 13 did not participate in the BiteWrite test, and 8 had a difficult time and elected to stop the test after a single sentence.

sequences used for $N = 5$. BiteWrite displays the sequences to click for each letter, a text input area, prompt text, and a record of what the last click was recognized as. Figure 7 illustrates BiteWrite for a participant able to click five pairs of teeth.

We recruited eight participants to return and informally test the list selection interface and BiteWrite. For each participant, we used all of their previously collected training data, as well as conducting a new training session in order to gain the best performance possible. The list selection interface worked well, with participants averaging 1.5 clicks to move between two list elements. BiteWrite, however, did not fare as well, with an average of almost 12 seconds to select each character, and a high error rate, with 12.6% of the clicks being the cancel/backspace click. These results are shown in Table 2.

DISCUSSION

Our preliminary testing of Bitey allows us to form some specific conclusions to guide future research.

Live performance

The performance of Bitey during live testing did not mirror its overall high accuracies in the best-case scenario, although we trained with a large amount of data. There are at least two factors that account for this discrepancy. During training, we asked participants to repeatedly click one pair of teeth for about 20 seconds. During this procedure, the impact of the teeth tends to happen at about the same location on each tooth; however, during live usage when the user is rapidly switching between teeth, variation will naturally occur. A second issue is that of fatigue: BiteWrite participants took on average 4.75 minutes ($STD = 2.4$) to type each sentence, and as the jaw fatigues, the impact of tooth on tooth will naturally vary.

Bitey can be improved for live performance in several ways. As a simple proof-of-concept, BiteWrite did not use prediction at all. Including predictive text or auto-complete would likely significantly speed up its operation. Additionally, the

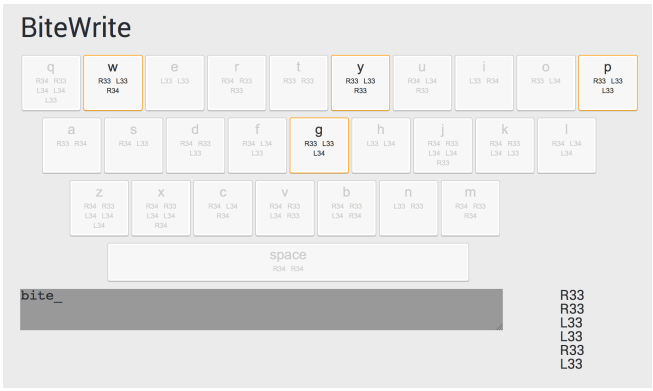


Figure 7. The *BiteWrite* interface. The keyboard view is for reference only, showing the sequence of clicks necessary to input each letter, and the possible keys given the input sequence so far. The last six tooth clicks detected are listed in the lower-right area; if the user taps R33, the letter 'y' will be generated.

Huffman coding method we borrowed from H4-writer [29] is optimized for information transmission [18], and the tree used to construct the coding is unbalanced. This property does not matter for machine information transmission, but for Bitey, it means that certain teeth are used more frequently than others. Reconstructing the tree to be less efficient but more balanced could help with fatigue. Despite these options, we consider BiteWrite to be an example of the possibilities of Bitey and not a useful text input technique; however, a similar method could be used for PIN input.

Another method for improving Bitey's performance is to change how users train the system. Rather than a repetitive tapping, we might ask users to tap their teeth in different sequences to take into account any "co-articulation" effects that occur. BiteWrite might be useful for this purpose—users could type a short sentence that uses all tap sequences.

System hardware

Although not expressly intended for bone-conduction applications, the inexpensive throat microphone we used worked remarkably well. Its major drawback was in the form factor—being shaped for the throat meant that it could be uncomfortably tight on the head. Re-bending the headband did mitigate this problem, however, and the hardware is easily adaptable to be placed in a hat or headband for improved wearability. Google Glass already incorporates a bone-conduction speaker, so adding a microphone could be a simple matter as well.

Improving performance

The best-case performance of the system (Table 1) was surprisingly variable, but tended to be close to 70% or higher. Participants 4 and 15 were exceptions, with 52.3% and 44.3% accuracy (without noise) respectively. We experimented with joining or removing some of the pairs of teeth for 4 and 15 to see if a reduced set would yield higher accuracy. We joined the two most-frequently confused pairs of teeth for each; for participant 4, we joined L66 and L67, and for participant 15, we joined L33 and L88 into a Left class and R33 and R88 into a Right class. The results, displayed in Figure 8, show the feasibility

22	e	33	t	34	a	31	o	24	i	23	n	42	s	21	h	433	r	411	l	432	d	414	c	413	u
412	f	434	m	324	w	323	y	322	p	321	g	4314	b	4313	v	4312	k	43114	x	43113	j	43112	q	43111	z

Table 3. H4-Writer sequences for generating letters. With BiteWrite, each number 1–4 corresponds to a particular pair of teeth for a user; the 5th tooth click is "cancel" or "backspace" if a sequence has not yet begun.

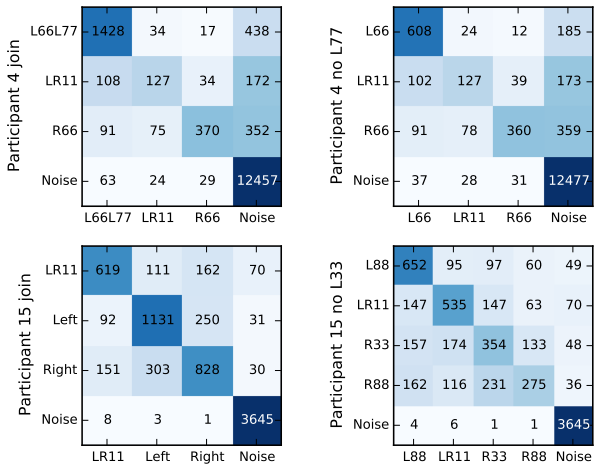


Figure 8. Confusion matrices for participants 4 and 15 with two tooth pairs joined (left column; accuracy 90.9% for 4 and 83.7% for 15) or one tooth pair removed (right column; 92.1% for 4 and 75.2% for 15).

of the approach, with joining confused pairs resulting in an average accuracy increase of 37%.

We also attempted to improve performance by using stereo data. Although our bone-conduction microphone is mono, it is relatively simple to construct a stereo version by removing the piezeoelectric microphone element and accompanying circuit board from one unit, inserting it into the other side of a second unit, and soldering on a new stereo headphone cable. We used a Griffin iMic USB sound card to enable stereo input into the computer. We recorded stereo data for six participants, but saw very little difference in accuracy, on the order of ±1–2%.

The surprisingly high performance of monaural audio can possibly be attributed to the attenuation of different frequencies as they pass via bone conduction through the mandible and skull [20]. Even without the amplitude differences afforded by two microphones, the spectral properties of a tooth clicked on the other side of the head from the microphone location will be different than those of a nearby click, allowing Bitey to differentiate between, for example, L33 and R33, even with monaural data.

Extended noise testing

To determine how Bitey might perform in an all-day scenario, we collected approximately six hours of non-click data from participant 7, including walking, eating (including chips, nuts, and ice), driving, and talking. Out of the six hours, Bitey detected 148,459 potential click candidates, but of these only 649 were incorrectly identified as clicks. While percentage-wise this is good performance, in terms of absolute numbers it leaves much to be desired, with a false positive on average every 30 seconds. One possible avenue of improvement is a more sophisticated segmentation algorithm, perhaps taking into account surrounding sounds—i.e., only passing a candidate on for consideration when a period of silence falls before and after the potential click.

BITEY IN DENTISTRY

Two of the co-authors are professors at a local school of dentistry, who assure the reader that the force of clicking used in Bitey is unlikely to cause damage to the teeth. In this section, we suggest some ways that Bitey could be used in modern clinical practice. A 1998 study conducted by Tyson found that gnathosonics can be a reliable method for monitoring occlusion [47], but still focused on manual inspection of the gnathosonogram. Bitey's simple and inexpensive recording hardware combined with machine learning techniques may have applications in dentistry.

The human mandible moves like a hinge. When a person closes the jaw, there is contact of the upper and lower teeth. Depending on the anterior and posterior relationship, teeth may contact evenly or unevenly. Bitey could allow dental practitioners to estimate the force of contact through the amplitude of the signal, or to determine differences between tooth contact. This is most important when evaluating patients having orthodontic treatment as well as crowns and bridges.

Bitey may also have applications in diagnosing conditions such as ankylosed teeth (teeth fused to the bone of the jaw) [6] or dental caries (cavities), or measuring the looseness of teeth, which will aid in the diagnosis of gum disease. We plan to conduct future studies to determine occlusion before and after orthodontic treatment and to compare temporomandibular joint disorder patients with TMJ pain to those without pain.

CONCLUSIONS AND FUTURE WORK

Bitey is a preliminary exploration of the possibilities of hands-free tooth click-based input. We have shown that it is possible to detect which pair of teeth is being clicked with a high degree of accuracy: up to 96% accuracy with five different tooth pairs.

Although in the present research we only evaluated single clicks, we plan to explore further possibilities for gnathosonic interaction. For example, a post-click slide from the canine to the bicuspid (R33-34) makes a double-clicking sound that is distinct from a double click of a single tooth (e.g., 2R33). Teeth also make sounds when they slide against one another, which could enable further expressivity.

We are currently working on implementing Bitey for controlling actual devices, such as watches and head-worn displays, and as part of the work will continue to improve the click detection and classification algorithms as well as the hardware itself.

Finally, in collaboration with other professors at our dental school, we plan to investigate the applicability of Bitey to dental diagnosis and monitoring.

REFERENCES

1. Brian Amento, Will Hill, and Loren Terveen. 2002. The sound of one hand: a wrist-mounted bio-acoustic fingertip gesture interface. In *CHI EA '02: CHI '02 Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, New York, USA, 724–725.
2. Oliver Amft, Martin Kusserow, and G Tröster. 2009. Bite Weight Prediction From Acoustic Recognition of Chewing. *Biomedical Engineering, IEEE Transactions on* 56, 6 (June 2009), 1663–1672.
3. Oliver Amft, Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. 2005. Analysis of Chewing Sounds for Dietary Monitoring. In *UbiComp 2005: Ubiquitous Computing*. Springer Berlin Heidelberg, Berlin, Heidelberg, 56–72.
4. Daniel Ashbrook. 2009. *Enabling Mobile Microinteractions*. Ph.D. Dissertation. PhD Thesis, Georgia Tech, Georgia Institute of Technology.
5. Daniel Ashbrook, Patrick Baudisch, and Sean White. 2011. Nenya: subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, New York, USA, 2043–2046.
6. William Biederman. 1962. Etiology and treatment of tooth ankylosis. *American Journal of Orthodontics* 48, 9 (Sept. 1962), 670–684.
7. Gábor Balázs Blaskó. 2007. *Cursorless Interaction Techniques for Wearable and Mobile Computing*. Ph.D. Dissertation. Columbia University.
8. H S Brenman. 1974. Gnathosonics and occlusion. *Frontiers of oral physiology* 1 (1974), 238–256.
9. H S Brenman and James S Millsap. 1959. A "Sound" Approach to Occlusion. *The bulletin of the Philadelphia County Dental Society* 24 (1959), 4–8.
10. H S Brenman, R C Weiss, and M Black. 1966. Sound as a diagnostic aid in the detection of occlusal discrepancies. *The Penn-Dental Journal* 69, 2 (1966), 33–49.
11. C A Chin, A Barreto, and J G Cremades. 2008. Integrated electromyogram and eye-gaze tracking cursor control system for computer users with motor disabilities. *Journal of Rehabilitation Research and Development* 45, 1 (2008), 161–174.
12. Travis Deyle, S Palinko, E S Poole, and T Starner. 2007. Hambone: A Bio-Acoustic Gesture Interface. In *Wearable Computers, 2007 11th IEEE International Symposium on*. IEEE, 3–10.
13. David J Fuller and Victor C West. 1987. The tooth contact sound as an analogue of the "quality of occlusion". *The Journal of Prosthetic Dentistry* 57, 2 (Feb. 1987), 236–243.
14. Mayank Goel, Chen Zhao, Ruth Vinisha, and Shwetak N Patel. 2015. Tongue-in-Cheek: Using Wireless Signals to Enable Non-Intrusive and Flexible Facial Gestures Detection. In *CHI '16: Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems*. New York, New York, USA, 255–258.
15. Sean Gustafson, Daniel Bierwirth, and Patrick Baudisch. 2010. Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. New York, New York, USA, 3.
16. Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: appropriating the body as an input surface. In *CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, New York, USA, 453–462.
17. Hędzelek and Hornowski. 1998. The analysis of frequency of occlusal sounds in patients with periodontal diseases and gnathic dysfunction. *Journal of Oral Rehabilitation* 25, 2 (Feb. 1998), 139–145.
18. David A Huffman. 1952. A method for the construction of minimum-redundancy codes. *Proceedings of the I.R.E.* 40, 9 (1952), 1098–1101.
19. ISO. 2009. *Dentistry—Designation system for teeth and areas of the oral cavity*. ISO 3950:2009. International Organization for Standardization, Geneva, Switzerland.
20. Krishan K Kapur. 1971. Frequency Spectrographic Analysis of Bone Conducted Chewing Sounds in Persons With Natural and Artificial Dentitions. *Journal of Texture Studies* 2 (1971), 50–61.
21. Koichi Kuzume. 2008. A Character Input System Using Tooth-Touch Sound for Disabled People. In *International Conference on Computers Helping People with Special Needs*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1157–1160.
22. Koichi Kuzume. 2011. Tooth-touch Sound and Expiration Signal Detection and Its Application in a Mouse Interface Device for Disabled Persons: Realization of a Mouse Interface Device Driven by Biomedical Signals. In *International Conference on Pervasive and Embedded Computing and Communication Systems*. SciTePress - Science and and Technology Publications, 15–21.
23. Koichi Kuzume. 2012. Evaluation of tooth-touch sound and expiration based mouse device for disabled persons. In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 387–390.
24. Koichi Kuzume and T Morimoto. 2006. Hands-free man-machine interface device using tooth-touch sound for disabled persons. In *Proceedings of the 6th International Confernece on Disability, Virtual Reality and Associated Technology*. 147–152.
25. Peter R L'Estrange, Alan R Blowers, Robert G Carlyon, and Stig L Karlsson. 1993. A microcomputer system for physiological data collection and analysis. *Australian Dental Journal* 38, 5 (Oct. 1993), 400–405.

26. Cheng-Yuan Li, Yen-Chang Chen, Wei-Ju Chen, Polly Huang, and Hao-hua Chu. 2013. Sensor-embedded teeth for oral activity recognition. In *the 17th annual international symposium*. ACM Press, New York, New York, USA, 41.

27. Zheng Li, Ryan Robucci, Nilanjan Banerjee, and Chintan Patel. 2015. Tongue-n-cheek: non-contact tongue gesture recognition. In *IPSN '15: Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. New York, New York, USA, 95–105.

28. Zicheng Liu, Amar Subramanya, Zhengyou Zhang, Jasha Droppo, and Alex Acero. 2005. Leakage Model and Teeth Clack Removal for Air- and Bone-Conductive Integrated Microphones. *(ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.* 1 (2005), 1093–1096.

29. I S MacKenzie, R W Soukoreff, and J Helga. 2011. 1 thumb, 4 buttons, 20 words per minute: Design and evaluation of H4-Writer. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*.

30. W D Mccall Jr., Antje Tallgren, and M M Ash Jr. 1979. EMG Silent Periods in Immediate Complete Denture Patients: A Longitudinal Study. *Journal of Dental Research* 58, 12 (Dec. 1979), 2353–2359.

31. Tamer Mohamed and Lin Zhong. 2006. *TeethClick: Input with Teeth Clacks*. Technical Report. Rice University.

32. Alexander Ng, Stephen A Brewster, and John Williamson. 2013. The Impact of Encumbrance on Mobile Interactions. In *Proceedings of The International Symposium on Open Collaboration*. Springer Berlin Heidelberg, Berlin, Heidelberg, 92–109.

33. Ian Oakley, Doyoung Lee, MD Rasel Islam, and Augusto Esteves. 2015. *Beats: Tapping Gestures for Smart Watches*. ACM, New York, New York, USA.

34. Jerome Pasquero, Scott J Stobbe, and Noel Stonehouse. 2011. A haptic wristwatch for eyes-free interactions. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, New York, USA, 3257.

35. J F Prinz. 2000. Computer aided gnathosonic analysis: distinguishing between single and multiple tooth impact sounds. *Journal of Oral Rehabilitation* 27 (2000), 682–689.

36. A. Prochazka. 2005. Method and apparatus for controlling a device or process with vibrations generated by tooth clicks. (Nov. 1 2005). US Patent 6,961,623.

37. Tauhidur Rahman, Alexander T Adams, Mi Zhang, Erin Cherry, Bobby Zhou, Huaishu Peng, and Tanzeem Choudhury. 2014. BodyBeat: a mobile system for sensing non-speech body sounds. In *MobiSys '14: Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. New York, New York, USA, 2–13.

38. Himanshu Sahni, Abdelkareem Bedri, Gabriel Reyes, Pavleen Thukral, Zehua Guo, Thad Starner, and Maysam Ghovanloo. 2014. The tongue and ear interface: a wearable system for silent speech recognition. In *ISWC '14: Proceedings of the 2014 ACM International Symposium on Wearable Computers*. New York, New York, USA, 47–54.

39. A Sears, M Lin, J Jacko, and Y Xiao. 2003. When computers fade. . . Pervasive computing and situationally-induced impairments and disabilities. In *International Conference on Human Computer Interaction*. HCI International.

40. C S SHI and Y MAO. 1993. Elementary identification of a gnathosonic classification using an autoregressive model. *Journal of Oral Rehabilitation* 20, 4 (July 1993), 373–378.

41. Chong-Shan Shi, Guan Ouyang, and Tian-wen Guo. 1991. Power spectral analysis of occlusal sounds of natural dentition subjects. *Journal of Oral Rehabilitation* 18, 3 (May 1991), 273–277.

42. Tyler Simpson, Colin Broughton, Michel J A Gauthier, and Arthur Prochazka. 2008. Tooth-Click Control of a Hands-Free Computer Interface. *Biomedical Engineering, IEEE Transactions on* 55, 8 (Aug. 2008), 2050–2056.

43. Tyler Simpson, Michel Gauthier, and Arthur Prochazka. 2010. Evaluation of Tooth-Click Triggering and Speech Recognition in Assistive Technology for Computer Access. *Neurorehabilitation and Neural Repair* 24, 2 (Feb. 2010), 188–194.

44. J M Stewart. 1953. Diagnosis of Traumatic Occlusion. *The Journal of the Florida State Dental Society* 24 (Oct. 1953), 4–9.

45. H N Teodorescu, V Burlui, and P D Leca. 1988. Gnathosonic analyser. *Medical and Biological Engineering and Computing* 26, 4 (July 1988), 428–431.

46. Outi Tuisku, Veikko Surakka, Toni Vanhala, Ville Rantanen, and Jukka Lekkala. 2012. Wireless Face Interface: Using voluntary gaze direction and facial muscle activations for human–computer interaction. *Interacting with Computers* 24, 1 (Jan. 2012), 1–9.

47. K W Tyson. 1998. Monitoring the state of the occlusion – gnathosonics can be reliable. *Journal of Oral Rehabilitation* 25, 5 (May 1998), 395–402.

48. David M Watt. 1963. A preliminary report on the auscultation of the masticatory mechanism. *Dental Practitioner* 14 (Sept. 1963), 27–30.

49. David M Watt. 1966. Gnathosonics—A study of sounds produced by the masticatory mechanism. *The Journal of Prosthetic Dentistry* 16, 1 (Jan. 1966), 73–82.

50. David M Watt. 1969. Recording the sounds of tooth contact: a diagnostic technique for evaluation of occlusal disturbances. *International Dental Journal* 2 (June 1969), 221–238.

51. David M Watt. 1970. Use of sound in oral diagnosis. *Proceedings of the Royal Society of Medicine* 63, 8 (Aug. 1970), 793.

52. David M Watt. 1981. *Gnathosonic Diagnosis and Occlusal Dynamics*. Praeger Publishers.

53. Koji Yatani and Khai N Truong. 2012. BodyScope: a wearable acoustic sensor for activity recognition. In *UbiComp '12: Proceedings of the 2012 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, New York, New York, USA, 341–350.

54. Shengdong Zhao, Pierre Dragicevic, Mark Chignell, Ravin Balakrishnan, and Patrick Baudisch. 2007. Earpod: eyes-free menu selection using touch input and reactive audio feedback. *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems* (April 2007), 1395–1404.

55. Xiaoyu Amy Zhao, Elias D Guestrin, Dimitry Sayenko, Tyler Simpson, Michel Gauthier, and Milos R Popovic. 2012. Typing with eye-gaze and tooth-clicks. In *ETRA '12: Proceedings of the Symposium on Eye Tracking Research and Applications*. New York, New York, USA, 341.

56. Lin Zhong, Dania El-Daye, Brett Kaufman, Nick Tobaoda, Tamer Mohamed, and Michael Liebschner. 2007. OsteoConduct: wireless body-area communication based on bone conduction. In *ICST 2nd international conference on Body area networks*. ICST.