Byte.it: Discreet Teeth Gestures for Mobile Device Interaction

Tomás Vega Gálvez
Massachusetts Institute of Technology
Cambridge, MA, USA
tomasero@mit.edu

Shardul Sapkota
Massachusetts Institute of Technology
Cambridge, MA, USA
sapkota@mit.edu

Alexandru Dancu
Augmented Human Lab
Singapore, Singapore
alex@ahlab.org

Pattie Maes
Massachusetts Institute of Technology
Cambridge, MA, USA
pattie@media.mit.edu

ABSTRACT

Byte.it is an exploration of the feasibility of using miniaturized, discreet hardware for teeth-clicking as hands-free input for wearable computing. Prior work has been able to identify teeth-clicking of different teeth groups. Byte.it expands on this work by exploring the use of a smaller and more discreetly positioned sensor suite (accelerometer and gyroscope) for detecting four different teeth-clicks for everyday human-computer interaction. Initial results show that an unobtrusive position on the lower mastoid and mandibular condyle can be used to classify teeth-clicking of four different teeth groups with an accuracy of 89%.

KEYWORDS
Discreet Interfaces; Teeth Gestures; Hands-free Interaction; Microgestures

CCS Concepts
Human-centered computing – Human computer interaction (HCI) – Interaction techniques – Gestural input
Human-centered computing – Ubiquitous and mobile computing – Ubiquitous and mobile computing systems and tools
Hardware – Communication hardware, interfaces and storage – Sensor devices and platforms

ACM Reference Format
Tomás Vega Gálvez, Shardul Sapkota, Alexandru Dancu, Pattie Maes, CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts), May 4–9, 2019, Glasgow, Scotland UK, ACM, New York, NY, USA, 6 pages, https://doi.org/10.1145/3290607.3312925

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05
https://doi.org/10.1145/3290607.3312925

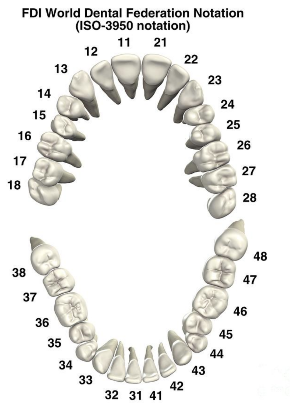

Figure 1: ISO 3950 system for tooth numbering [8]

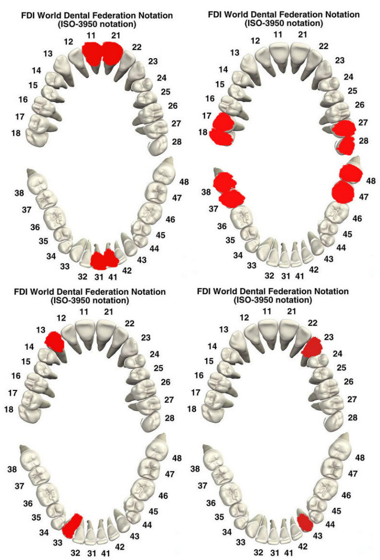

Figure 2: Corresponding teeth groups highlighted in red for (starting from the left) front click, back click, right click, and left click

INTRODUCTION

Microinteractions can reduce the amount of interaction necessary for short tasks. The disadvantage of microinteractions with today's interfaces is that they require the use of hands. This limits the use of mobile devices in dynamic, on-the-go contexts such as walking, running, and driving. Speech recognition systems might address this problem, but these can often be socially disruptive. In this paper, we present Byte.it – a system that uses accelerometer and gyroscope data to recognize teeth-clicking gestures for subtle, hands-free interface control.

RELATED WORK

Facial Gestures

Cheng et al. have explored the use of a non invasive pressure based tongue interface as an input modality for mobile and wearable devices [3]. Nguyen et al. explored non-invasive brain signals, muscle signals, and skin surface deformation (SKD) sensing to determine the relative location and interaction between the user's tongue and teeth [9].

In addition to tongue, researchers have also expanded the gesture recognition capabilities to the face. Lyons et al. introduced a new interaction modality using facial gestures and expressions to control musical sound [7]. A similar machine vision based technology was used by Lyons et al. as a new way for text entry method using coordinated motor action of hand and mouth [6]. While the use of tongue, mouth, and facial expressions provide novel ways to interface with mobile technologies in an non-invasive way, technologies that require a helmet, face mask, or a head mounted camera are arguably cumbersome for daily usage.

Teeth Gestures

Prior research has explored teeth-clicks as an input modality for assistive technology. People with limited motor abilities often have difficulties operating pointing devices for computer interaction. Using bone conduction microphones, researchers have classified sequences of teeth touch sounds and use these as commands for mouse control [5]. Bitey explored different teeth-click gestures for hands-free interface control using a wearable bone conduction microphone [1]. Simpson et al. explored head-tracking technology with an accelerometer attached on the side of the head for teeth-click/head-mouse control for people with severe upper limb paralysis. The system was found to be faster than dwell-time control and, although not faster, more reliable and less inconvenient than sip-and-puff [10]. Accelerometer and gyroscope data was also used to differentiate teeth clicks from general speech and movement artifacts. These studies concluded that teeth clicks were significantly faster at generating mouse button clicks than speech recognition technology [11]. However, these works were only able to recognize occurrences of teeth clicks and were not able to distinguish between the different forms of

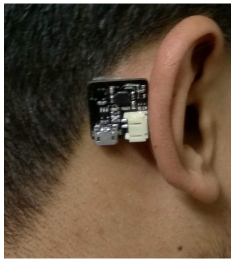

Figure 3: Position A of the GRU: upper mastoid

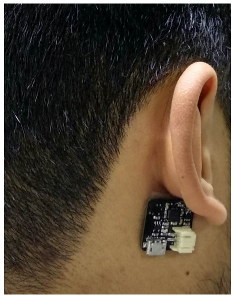

Figure 4: Position B of the GRU: PCB located on the lower mastoid touching constantly the mandibular condyle

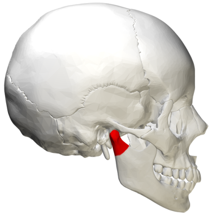

Figure 5: Mandibular condyle in red [2].

teeth clicks. Moreover, they put greater emphasis on interfacing with the computer mouse. Byte.it expands on this work by exploring the use of a smaller sensor suite (accelerometer and gyroscope) for detecting four different types of teeth clicks for everyday human-computer interaction.

SYSTEM

The Gesture Recognition Unit (GRU) is a compact (20x20mm) custom-made Bluetooth-enabled Inertial Measurement Unit (IMU) that measures acceleration and angular velocity in real-time. It contains an NRF52-based MCU (BC832), a 6-axis MEMS IMU (MPU6050), a low-dropout regulator (MIC5205), a Li-Ion/Li-Polymer charge management controller (MCP73831), and a female MicroUSB connector. The experimental setup consisted of a 13.3" laptop in front of the participant, the GRU attached using a double-sided skin-friendly tape, and an iPhone 7 to receive IMU data streamed by the GRU.

The current implementation contains two components: 1) an iOS app to collect and label sensor data and to export the recorded dataset via email 2) a gesture classifier system that runs a k-NN algorithm with Dynamic Time Warping (DTW) for IMU data classification.

We used the ISO system of tooth numbering for easier gesture identification and reference. The ISO 3950 system uses 2 digits to identify a tooth, where the first number refers to the quadrant and the second to the tooth within the quadrant [4]. Based on this system, the four teeth-clicking gestures are defined as follows (see Figure 2): front click, left click, right click, and back click.

STUDY

The purpose of our pilot study was to assess the possibility of using an IMU to detect different teeth-clicking gestures for human-computer interaction. Prior work has either used IMU data to classify teeth-clicking for assistive technology or used bone conduction microphone to recognize different teeth-clicking for everyday human-computer interaction [1, 10]. We tested the possibility of using an IMU to detect different teeth-clicking gestures for everyday interaction. Our study assesses the performance of a miniaturized wireless IMU placed in a discreet position to enable the able-bodied to interact with technology. This study is a first investigation of seamless hardware and positioning for exploring alternative input methods for mobile computing. Two hypotheses are driving this research:

(1) **H1**: An IMU can be placed in a discreet position to recognize distinct teeth-clicking gestures.
(2) **H2**: IMU data provides comparable classification accuracy to bone conduction microphones.

EXPERIMENT 1: FINDING THE OPTIMAL POSITION

Identifying the optimal location of the sensor on the user's head is a trade-off between having a socially acceptable position and locating a position that provides enough sensitivity to be able to distinguish between gestures.

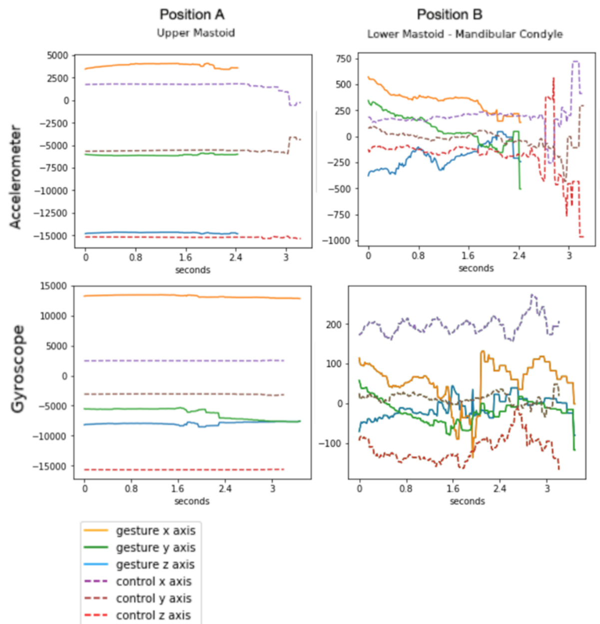

Figure 6: Gyroscope and accelerometer values for the back click gesture for the two positions. Control data points are shows in dashed lines and the time taken to perform one back click gesture lies within the range in the x-axis

Gestures	Position A	Position B
Back Click	2137.49	11384.75
Left Click	3722.48	14980.77
Right Click	2255.60	10728.42
Front Click	2527.07	7518.58

Table 1: DTW distance between gesture and control in A & B

We consider that mounting the GRU behind the ear is a socially acceptable position because the hardware is hidden from sight when looking straight at a person's face e.g. when talking with a person. We evaluated the location sensitivity of two positions: Position A is the upper mastoid depicted in Figure 3 and Position B location is between lower mastoid and mandibular condyle shown in Figure 4.

Procedure

In order to avoid an arbitrary choice of the positioning of the GRU, we placed the GRU in two positions and collected data samples to see which position gave the most distinct sensor reading. We chose these two positions by keeping in mind the visibility of the device to the person in front of the one on whom the sensor was mounted on. The two positions are shown in Figure 3, where first, the GRU was mounted at the position shown in Figure 3 and second, on the position shown in Figure 4. We did a pilot by placing the IMU on those two positions with the help of a dressing tape.

Since accelerometers are sensitive to minor vibrations, we normalized the detection of rotatory motion of the head by keeping the head at approximately the same position and also by fixing the orientation of the GRU while recording every new gesture sample for a gesture class. However, the length of each sample varied even for the same gesture.

The MPU6050 IMU has a sample rate of 10 Hz. In order to discount the high frequency noise in the input signal, we used a low pass filter with an α value of 0.2. We calculated a running threshold for each data point using the following formula:

$$current\_acc_x = \alpha \cdot acc_x + (1 - \alpha) \cdot current\_acc_x,$$

where x represents one of the three dimensions.

Results

We recorded a total of 100 gesture samples, across all teeth-click classes plus control, in the two positions discussed above. We obtained 10 samples per gesture class. To visualize the data, we made a 3d plot by averaging the x, y, and z values of the accelerometer and gyroscope data. We also recorded still gesture samples for both positions as our control. From Figure 7, it is clear that position B has less clustered data points than position A, suggesting a better sensitivity for teeth-clicking movements. The control data of the gyroscope shown in orange show that the angular movement is minimal. Accelerometer readings, on the other hand, showed that both the gesture and the control readings similarly formed data cluster, making it difficult to make a distinction between the two positions. This difference between accelerometer and gyroscope readings also shows how gyroscope readings were more sensitive for teeth gesture detection than accelerometer readings. We depict the accelerometer and gyroscope values for the back click gesture in the graphs in Figure 6 by taking the average of accelerometer and gyroscope data across all participants. Since the length of the accelerometer and

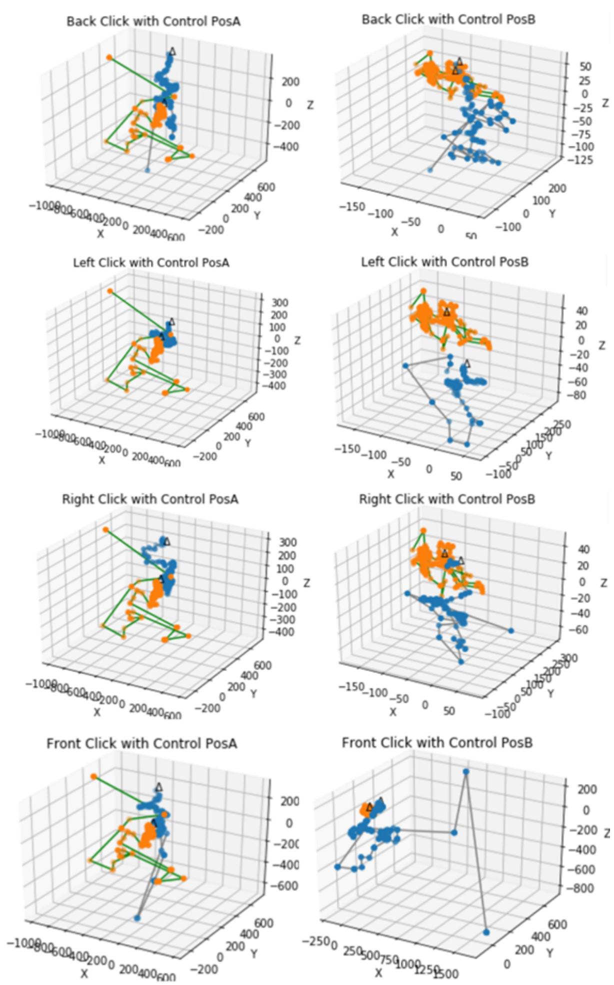

Figure 7: Gyroscope plots of gestures vs control samples. Orange data points are control and blue data points are gesture samples. Triangle shows the start of a gesture.

gyroscope values for the gesture samples vary even within the same gesture class from the same participant, the averaged values are not representative of the actual starting and stopping time of doing a single back click.

To further corroborate this result and to find numerical differences in the time series data between the control and the gestures in the two positions, a DTW distance measure was computed between the IMU readings of the gesture and control. Since DTW is a measure of similarity between two time series data, greater the DTW value, more distinct the two time series measures. These values can be seen in Table 1. Therefore, IMU readings from position B were, in general, more distinct than those from position A. Paired t-tests for each of the gestures revealed a significant effect of position on Position B over Position A ($t = -6.54$, $df = 3$, $p < 0.01$). Consequently, Position B has a better sensitivity for teeth-clicking movements. We have also shown that an IMU can be placed in a discreet position behind the ear, i.e., at Position B with permanent contact with the mandibular condyle which provides improved sensitivity, confirming our H1 hypothesis.

EXPERIMENT 2: TEETH-CLICKING CLASSIFICATION ACCURACY ON OPTIMAL POSITION

The experiment used a repeated-measures, single-factor design. All participants performed the 4 gestures (left, right, front, back). The order of the gestures was counterbalanced using randomization.

After determining the optimal position of the GRU at Position B, data samples were collected from participants for the four different gestures. We first informed each participant of the purpose of the study, explained them how the teeth-clicking sensor worked, and attached the sensor on Position B. We asked the participants to practice the four different gestures. Before starting recording, they were asked to remain as still as possible and keep their sight anchored at the computer screen to reduce the number of motion artifacts. Participants were shown the images in Figure 2 and asked to carry out the gesture corresponding to the teeth highlighted in red. The experimenter prompted the participant to initiate performing the target gesture and do a thumbs up when done. This was repeated 10 times per each of the four gestures. Using the mobile app, the experimenter created a class for each gesture and recorded the samples. At the end, participants were asked to report any feedback about the experiment. The experiment lasted 15 minutes in average.

Results

We used a k-nearest neighbors (k-NN) algorithm with DTW as a distance metric to classify the distinct teeth-clicking gestures. Due to our small dataset, we we opted for a leave one out (LOO) cross-validation method for testing, where we separated the data n times into a training set of size $n-1$ and a test set of size one. Since we did not do any hyperparameter tuning, we did not use a validation test set. We obtained a test accuracy score of 89%. Prior work from Ashbrook et al. showed that bone conduction microphone based teeth-click classification was possible with an accuracy score of 94% under laboratory setting [1]. However, given that this prior work made personalized classification models for each participant and had an average accuracy score of only 78%, our generalized classification model accuracy score using an IMU is comparable to that using bone conduction microphones, confirming our H2 hypothesis.

CONCLUSION

Byte.it is an exploration of the feasibility of using miniaturized, discreet hardware for hands-free teeth-clicking as input for wearable computing. Initial results show it is possible to classify teeth-clicking of four different teeth groups with an accuracy of 89%. Initial results show that system miniaturization, unobtrusive positioning, and high gesture recognition rate could enable seamless interactions with mobile devices.

REFERENCES

[1] Daniel Ashbrook, Carlos Tejada, Dhwanit Mehta, Anthony Jiminez, Goudam Muralitharam, Sangeeta Gajendra, and Ross Tallents. 2016. Bitey: An exploration of tooth click gestures for hands-free user interface control. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 158–169.

[2] BodyParts3D. 2013. Condyloid process. (Feb 2013). https://commons.wikimedia.org/wiki/File:Condyloid_process_-_lateral_view.png

[3] Jingyuan Cheng, Ayano Okoso, Kai Kunze, Niels Henze, Albrecht Schmidt, Paul Lukowicz, and Koichi Kise. 2014. On the tip of my tongue: a non-invasive pressure-based tongue interface. In *Proceedings of the 5th Augmented Human International Conference*. ACM, 12.

[4] ISO. 2009. Dentistry—Designation system for teeth and areas of the oral cavity. International Organization for Standardization, Geneva, Switzerland.

[5] Koichi Kuzume. 2012. Evaluation of tooth-touch sound and expiration based mouse device for disabled persons. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*. IEEE, 387–390.

[6] Michael J Lyons, Chi-Ho Chan, and Nobuji Tetsutani. 2004. Mouthtype: Text entry by hand and mouth. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1383–1386.

[7] Michael J Lyons, Michael Haehnel, and Nobuji Tetsutani. 2001. The mouthesizer: a facial gesture musical interface. In *Conference Abstracts, Siggraph 2001*. 230.

[8] Universal Images Group Medical Images. 2014. FDI Dentition. https://fineartamerica.com/featured/fdi-dentition-medical-images-universal-images-group.html

[9] Phuc Nguyen, Nam Bui, Anh Nguyen, Hoang Truong, Abhijit Suresh, Matt Whitlock, Duy Pham, Thang Dinh, and Tam Vu. 2018. Tyth-typing on your teeth: Tongue-teeth localization for human-computer interface. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 269–282.

[10] Tyler Simpson, Colin Broughton, Michel JA Gauthier, and Arthur Prochazka. 2008. Tooth-click control of a hands-free computer interface. *IEEE Transactions on Biomedical Engineering* 55, 8 (2008), 2050–2056.

[11] 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 (2010), 188–194.