# TongueBoard: An Oral Interface for Subtle Input

**Richard Li, Jason Wu, Thad Starner**
Google Research & Machine Intelligence
Mountain View, CA, USA
{lichard,jsonwu,thadstarner}@google.com

## ABSTRACT

We present TongueBoard, a retainer form-factor device for recognizing non-vocalized speech. TongueBoard enables absolute position tracking of the tongue by placing capacitive touch sensors on the roof of the mouth. We collect a dataset of 21 common words from four user study participants (two native American English speakers and two non-native speakers with severe hearing loss). We train a classifier that is able to recognize the words with 91.01% accuracy for the native speakers and 77.76% accuracy for the non-native speakers in a user dependent, offline setting. The native English speakers then participate in a user study involving operating a calculator application with 15 non-vocalized words and two tongue gestures at a desktop and with a mobile phone while walking. TongueBoard consistently maintains an information transfer rate of 3.78 bits per decision (number of choices = 17, accuracy = 97.1%) and 2.18 bits per second across stationary and mobile contexts, which is comparable to our control conditions of mouse (desktop) and touchpad (mobile) input.

## CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; **Ubiquitous and mobile devices**;

## KEYWORDS

wearable devices, input interaction, oral sensing, subtle gestures, silent speech interface

**ACM Reference Format:**
Richard Li, Jason Wu, Thad Starner. 2019. TongueBoard: An Oral Interface for Subtle Input. In *Augmented Human International Conference 2019 (AH2019), March 11–12, 2019, Reims, France.* ACM, New York, NY, USA, Article 4, 9 pages. https://doi.org/10.1145/3311823.3311831



*AH2019, March 11–12, 2019, Reims, France*

ACM ISBN 978-1-4503-6547-5/19/03…$15.00
https://doi.org/10.1145/3311823.3311831

## 1 INTRODUCTION

Conversational technologies, such as Google Assistant, Siri, and Alexa, facilitate natural and expressive interaction with online services, connected Internet-of-Things (IoT) devices, and mobile computer interaction. While these conversational interfaces are popular due to their ease of use, they often rely on speech transcription technology that are ineffective in noisy environments, are often socially unacceptable, and present privacy concerns of the microphone inadvertently capturing nearby conversations or having nearby listeners hearing the user's commands.

We present TongueBoard, an oral interface for subtle gesture interaction and control using silent speech. TongueBoard is a retainer form-factor device that tracks the movement of the tongue, enabling silent speech recognition and accurate detection of tongue gestures. Our system's accurate characterization of the tongue's position and movement allows for subtle interaction through a combination of silent speech vocabulary sets and tongue gestures. We discuss how our system can be used to control common user interfaces by supporting the capabilities of existing input hardware. To evaluate the accuracy of our system, we test a lexicon consisting of 21 words (numerical digits, mathematical operations, and time descriptors) and 9 tongue gestures. We first evaluate the offline accuracy of the silent speech recognition with data from two American English speakers and two non-native deaf speakers. In addition, both interfaces are also compared in a live study that compares its accuracy and interaction speed to traditional inputs (mouse and touchscreen) in stationary and mobile contexts.

## 2 RELATED WORK

### Subtle Interfaces

While some gesture interfaces are expressive and afford fast input, they may not be practical for everyday use. Gestures that are subtle or that utilize everyday movements are more likely to be socially acceptable and willingly performed by users [31]. Many gesture interfaces seek to minimize their noticeability by requiring little or no obvious movement, and others allow for the interaction to be performed out of view. Despite the challenges of capturing signal from low-motion gestures using traditional sensors, electromyography (EMG) has been successfully used to capture neuromuscular signals

**Table 1: Wearable Silent Speech Interface Comparison**

| Silent Speech Interface | Modality | Proxy | Dictionary | Accuracy | Invisible | Walking |
|---|---|---|---|---|---|---|
| Bedri et al. 2015 [6] | Optical & Magnetic | Jaw and tongue | 11 phrases | 90.50% | No | No |
| Kapur et al. 2018 [21] | sEMG | Jaw and cheek | 10 words | 92.01% | Yes | No |
| Meltzner et al. 2018 [27] | sEMG | Face and neck | 2200 words | 91.10% | No | No |
| Sun et al. 2018 [36] | Camera | Lip movement | 44 phrases | 95.46% | No | No |
| Fukumoto 2018 [13] | Audio | Ingressive speech | 85 phrases | 98.20% | No | No |
| **TongueBoard** | **Capacitive** | **Tongue** | **15 words/2 gestures** | **97.10%** | **Partially** | **Yes** |

associated with different movements [8, 26]. Other input interfaces support subtle interaction by requiring movement of unseen body parts to decrease the visibility of the gesture. Eyes-free [2, 29] interfaces and in-attentive gestures allow interactions to take place out of view, such as below the table or inside the user's pocket [17, 32, 34]. Recently, the ear has been identified as a promising place for supporting subtle interactions and monitoring physiological signals such as heart rate, blood pressure, electroencephalography (EEG), and eating frequency [5, 15, 16, 30]. Motion of the *temporalis* muscle, responsible for moving the jaw, causes movement and deformation of the ear, which can be sensed through barometric and in-ear electric field sensing [1, 24, 25].

Hidden areas such as the inside of the mouth can also be sensed using a variety of methods, including neuromuscular signals (EEG, EMG), skin surface deformation (SKD) [28], bone conduction microphones [3], optical approaches (infrared proximity sensors) [4, 33], and wireless signals [14]. Retainer form-factor oral and tongue sensors have also been explored for enabling tetraplegics to control electronic devices and powered wheelchairs using a tongue-drive system [22]. However, many of the sensing approaches provide an incomplete or low resolution characterization of the inner mouth and are not suitable for more complex gestures and silent speech applications.

### Silent Speech

Silent speech interfaces allow the user to communicate with a computer system using speech or natural language commands without the need to produce a clearly audible sound. These interfaces allow users to communicate efficiently with computer systems without attracting attention or disrupting the environment using traditional voice-based interfaces. Denby et al. [10] and Freitas et al. [11] provide comprehensive surveys of silent speech interfaces used for speech signal amplification, recognition of non-audible speech, and "unspoken speech" (imagined speech without any explicit muscle movement). Broadly, they fall into the categories of physical techniques (which measure the vocal tract directly), inference from non-audible acoustics, and electrical sensing (which measures the activations of actuator muscle signals or command signals from the brain) [10].

More recently, improvements have been made in several areas of sensing described in Denby et al. by incorporating novel hardware [12] and more advanced machine learning techniques [18]. Several systems have incorporated deep learning techniques to lipread speech for the purposes of silent speech and augmenting automatic speech recognition (ASR) models [35, 37]. Most relevant to the work presented in this paper is research done on subtle physical sensing of silent speech. Table 1 shows several recent silent speech interfaces and situates our work with regard to sensing modality, proxy, dictionary size, accuracy, whether the sensed phenomenoni is visible to a bystander, and usability while walking.

We position our main contribution as an interaction that is robust to motion artifacts. To the best of our knowledge, the works noted in table 1 still require significant work to be able to withstand noise from external movements. Furthermore, with practice, TongueBoard can be used with no movements noticeable by an observer, since all motions are contained within the mouth.

## 3 TONGUEBOARD

### Interaction Design

We construct a silent speech interface with 21 words for application input and nine tongue gestures (Table 2). When used together, TongueBoard is able to support numerical input, navigation, and item selection, enabling a wide array of potential interactions.

- *Digits & Text* - While we designed our vocabulary primarily to support a calculator application, recognition of digits can be used for dialing phones, selecting between Smart Reply suggestions [20], or controlling a multitap-like telephone keypad (e.g., saying "2 2 2" to select a C), allowing slower but more expressive input.
- *Navigation* - TongueBoard emulates a d-pad (up/down/left/right) by allowing users to place their tongue on certain regions on the roof of the mouth. Pressing the tongue fully against the roof of the mouth is selection.

- *Editing* - Four swiping gestures (left-right, front-back, etc.) enable editing commands such as backspace.
- *Control* - Progressively disclosing interfaces on head worn displays such as Google Glass follow a hierarchical structure where each utterance displays a limited list of potential options ('OK Glass ... Send a message to ... John Smith ... Be there in 5 minutes"). TongueBoard can support similarly rapid and accurate interaction through mapping options to TongueBoard's lexicon. In practice, the most commonly used commands would be directly included in the default vocabulary set.

In this section, we have described the silent speech and subtle gesture components of TongueBoard's interaction technique. Moreover, we suggested subtle alternatives to existing input mechanisms for common interface paradigms using our system.

**Table 2: TongueBoard Lexicon**

| Vocabulary | Group |
|---|---|
| 0-9, oh | Digits (11) |
| add, subtract, multiply, divide, percent | Arithmetic (5) |
| AM, PM, hours, minutes, seconds | Time (5) |
| D-pad, Swipe Gestures, Full | Navigation (9) |

## System Overview

*Hardware.* TongueBoard uses the CompleteSpeech SmartPalate system to detect the location of the tongue inside the mouth. The CompleteSpeech SmartPalate system is a commercially available system intended for speech therapy. The SmartPalate system consists of an array of 124 capacitive touch sensors embedded in an oral mouthpiece supporting a sample rate of 100 Hz. A DataLink module reads the raw

**Figure 1: CompleteSpeech SmartPalate**

capacitive values of the sensors and transmits binary states for each of the sensors to a laptop or smartphone through a standard USB cable.

*Data Processing.* The SmartPalate sensor array provides consistent and accurate tracking of the user's tongue as it moves against the roof of the mouth, allowing clear differentiation between words and gestures in our vocabulary.

We use a classification algorithm that leverages both tongue placement information and temporal movement for word-level silent speech classification and tongue swipe detection. Specifically, we use a support vector machine (SVM) with a

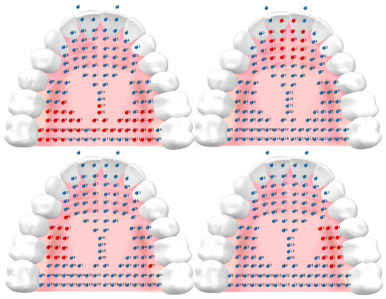

**Figure 2: Example visualization frames of TongueBoard sensor data. Red indicates the points where the tongue contacts the roof of the mouth, and blue indicates otherwise.**

global alignment kernel (GAK) [9], shown to be an effective discriminative model for time series data. Hyperparameter tuning was performed using an off-the-shelf optimizer [23]. A lightweight classifier such as an SVM is advantageous when compared to more sophisticated models such as deep neural networks due to its relatively lower processing requirements, which is crucial for mobile applications. The resulting advantages are lower latency in the output and less battery consumption.

Preliminary results revealed that the average example length was 1.29 s (or 129 samples at 100 Hz), and that resampling input segments (using average downsampling) to 10 samples was sufficient for accurate and fast classification. Maintaining the higher number of samples (100 samples/sec) would allow for greater temporal resolution and potentially better classification accuracy but would also increase the inference time of the system. Our aim was to maximize classification accuracy while minimizing inference time while running on a commodity laptop.

In contrast to the machine learning procedure required for classifying silent speech, recognizing tongue gestures is much simpler. Tongue placement is represented as a single point, computed as the centroid of the currently activated electrodes (Figure 2). Through some experimentation, we empirically set location boundaries for activating each button. We found that we are able to comfortably fit four activation regions without much overlap.

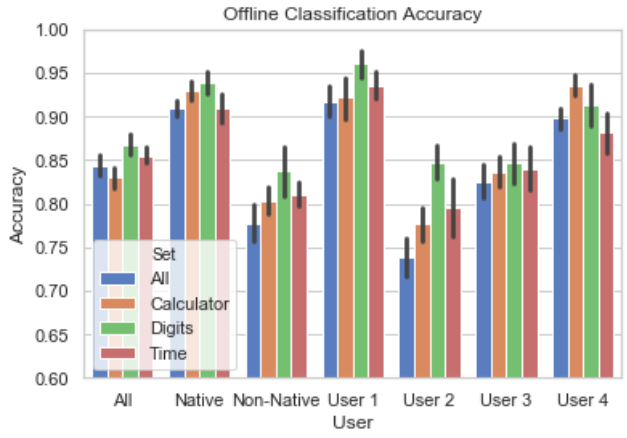


**Figure 3: Offline accuracy across users and vocabulary sets**

## 4 EVALUATION

### User Study - Offline Accuracy

We quantify our system's silent speech recognition accuracy by evaluating offline classification performance on 21 common words, including the numerical digits, arithmetic operators, and time descriptors (Table 2).

Due to the time and cost of fitting and producing a personalized Smart Palate per user, a limited number of four participants were recruited for this study. Two of the participants were males and native English speakers, while 1 male and 1 female were native Russian speakers with severe hearing disabilities that affect their speech. The inclusion of these non-native deaf speakers is due to on-going work on speech rehabilitation which compares SpeechPalate patterns between fluent and disfluent speakers [19].

Each participant collected 20 examples for each of the 21 words, presented in random order, forming a dataset of 20 examples * 21 words * 4 participants = 1680 utterances. Participants activated the silent speech interface using a "hold-to-gesture" mechanism, where the recognition system was fed data segmented by a key or button press. It is also possible to provide segmentation using a highly distinctive tongue gesture or train a completely live classifier. However, we were interested in quantifying the accuracy of the silent speech recognition alone and did not investigate these other options in our study.

Accuracy was calculated with respect to different user and vocabulary groupings. Datasets were created for each of the four users, along with the native speakers, non-native speakers, and all the users. Three additional vocabulary subsets were created for specific applications: Digits (numerical digits), Calculator (numerical digits and mathematical operations), and Time (numerical digits and time descriptors).

Using the data processing and machine learning techniques previously described, we evaluate each user/vocabulary set pair using ten iterations of shuffle split cross validation where random independent train/test sets of 75%-25% were generated.

Figure 3 shows the results of the offline accuracy experiments. When considering the entire 21-word vocabulary, the classifier accuracy across all users was $M_{all} = 84.36\%$. The native English speakers (Users 1 and 4) achieved much higher accuracy ($M_{native} = 91.01\%$) than the non-native, deaf users ($M_{non.} = 77.76\%$). The highest accuracy was achieved by User 1 ($M_1 = 91.71\%$) while User 2 had the lowest accuracy ($M_2 = 73.90\%$).

As the vocabulary set was limited to include only smaller, specialized subsets, the recognition rate of the system improved. User 1 achieved the highest accuracy on the Digits set (11 words) with $M_{digits,1} = 98.18\%$, and on average, the Digits set was the most distinctive ($M_{digits} = 88.75\%$). Using the Digits set, nonnative recognition accuracy was $M_{digts,non.} = 83.73\%$. Both the Calculator set and Time set contained 16 words, and had lower accuracy ($M_{calc} = 84.77\%$, $M_{time} = 86.10\%$).

Overall, our offline evaluation of TongueBoard shows it is able to recognize our mid-size vocabulary sets with high accuracy. For small vocabulary sets such as the Digits set, our system is able to achieve accuracies above 95%, making it suitable for use-cases described in the interaction design section. In particular, we find it encouraging that input queries from deaf and hard of hearing users can be recognized by our system with relatively high accuracy. As deaf speakers cannot use audio feedback to modulate their speech, they can be difficult for both human listeners and computers to understand. For example, the word error rate of audio-based transcription of 10 phrases is around 40% for deaf speech [7]. Our system's recognition operates by tracking the speaker's tongue and can ignore the acoustic differences of deaf speech. Perhaps TongueBoard may one day augment automatic speech recognition (ASR) systems or can be used as an interface to produce computer generated speech, allowing deaf users to operate speech-controlled devices [7].

### User Study - Live Input

The two native English speakers from the previous data collection effort also participated in a study to evaluate the system's performance in a real-time setting. Our aim was to measure the performance of the classifier in realistic contexts and to compare the speed and accuracy of the TongueBoard interaction with traditional input methods. The two participants first participated in a pilot study, and then engaged in an interaction study.

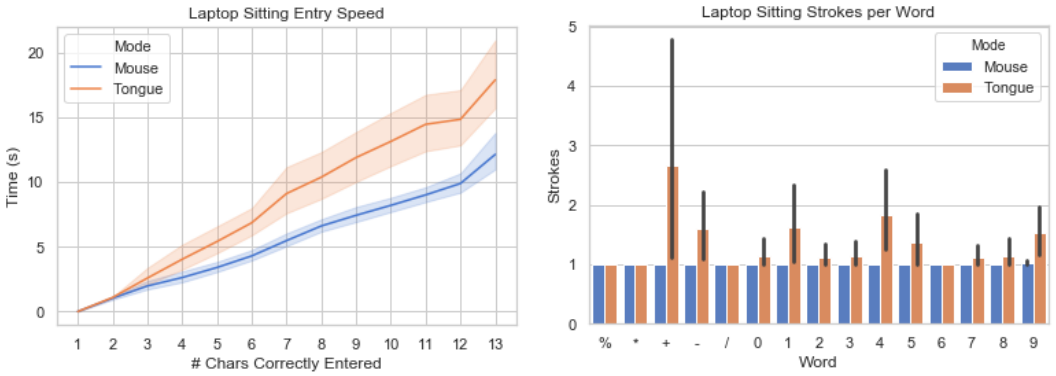


**Figure 4: Live laptop results.**

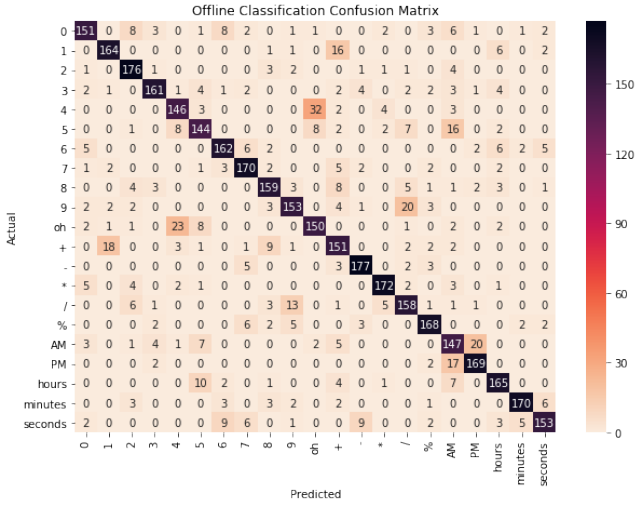


**Figure 5: Confusion matrix across all users and vocabulary**

Both of these studies involved operating a calculator application that accepted input from TongueBoard, using a button-segmented live classifier, and the default input modality for the respective device. Participants were asked to evaluate a series of mathematical expressions. In addition to silently mouthing for digit and arithmetic operation input, a "full" gesture (i.e., tongue pressed against the roof of the mouth attempting to activate all electrodes) was mapped to backspace and a "swipe up" gesture (i.e., running the tongue from the front of the mouth to the back) was mapped to the evaluate ("=") button.

*Pilot Study on a PC.* The pilot study was conducted on a Lenovo Thinkpad laptop computer, comparing the TongueBoard system with the laptop's pointing stick and trackpad buttons. The classifier was run live in the background of the calculator application and was trained on the Digits and Arithmetic data from the offline data collection. The expressions consisted of two randomly generated integers between -100000 and 100000, and an arithmetic operation. Each expression also prompted the user to use either the "tongue" modality or the mouse modality. Each user entered a total of 50 expressions.

Figure 4 compares the text entry speed of TongueBoard and traditional inputs. For each input method, we quantify the time needed to correctly enter the first $n$ characters of the prompt. The time needed to reach $n$ correctly typed characters includes the time taken by the user to perform the necessary backspace and retype operations. To eliminate variations caused by initial mouse seeking and prompt reading, we compute the speed with respect to the time the first character was successfully entered.

Overall, using the TongueBoard achieves an input accuracy of $M_{pilot,tongue} = 94.07\%$ for laptop input. Our analysis shows that TongueBoard is able to recognize most words in the vocabulary set with the exception of a small number of confused signals (Figure 4). In particular, we discovered that "+" (pronounced "add") and "1" (pronounced "one") produce very similar signals. Figure 5 shows this ambiguity in the offline confusion matrix, where there was a significantly higher number of misclassifications between the two words. These errors highlight the unavoidable flaw with our system: since production of speech involves more than just tongue movement (i.e. lip and jaw movements are also critical), the signals we obtain may not be necessarily unique per word. Additionally, the two participants, both native American English speakers, complained that it was unnatural to say "one add two" or "five subtract four." To improve the separability of the vocabulary set and make the interface more natural, a new dataset was collected for the following smartphone study that replaced "add" with "plus" and "subtract" with "minus".

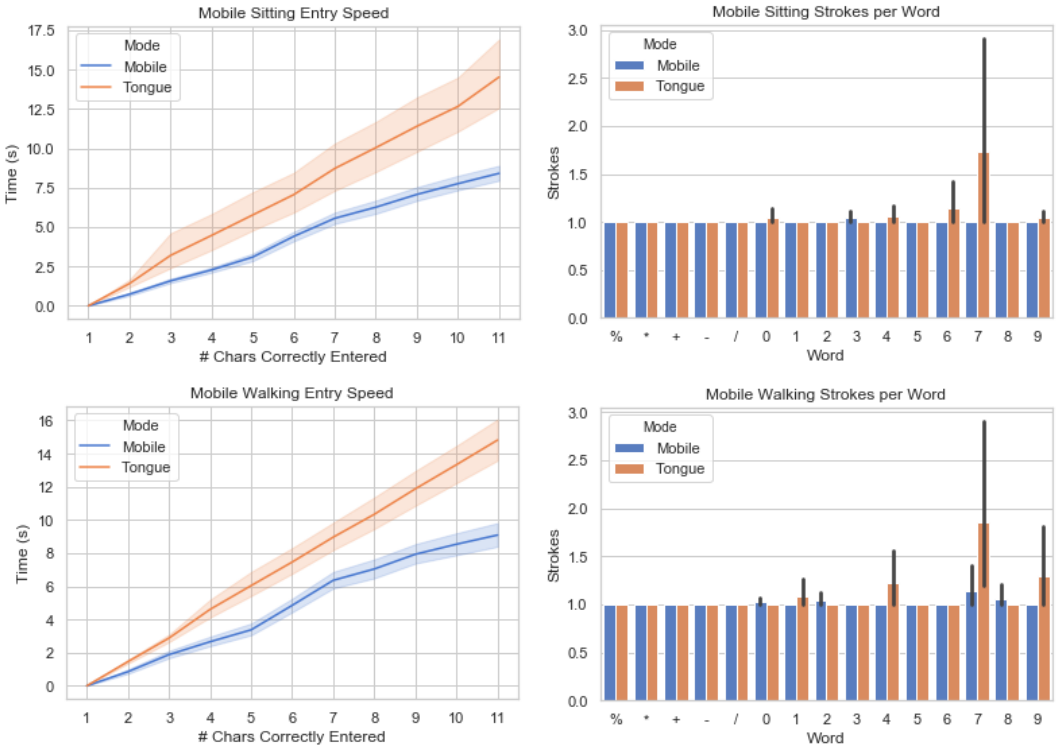


**Figure 6: Live Mobile Evaluation Results.**

*Interaction Study on a Smartphone.* The interaction study was conducted on the participants' personal Android phones (a Google Pixel and Pixel XL). The calculator application was designed to be as similar as possible to the desktop application, but offloaded the classification to a networked computer. In this interaction study, we investigated two conditions:

- *Sitting:* Users sat at a desk, comparing the TongueBoard system with an external mouse attached to the phone. This condition emulated the usage of a normal desktop computer (with a mouse) while keeping the rest of the experimental apparatus constant.
- *Walking:* Users walked up and down a hallway, comparing the TongueBoard system with the touchscreen.

During the walking condition, both participants reported that the TongueBoard system allowed them to walk more smoothly and bump into less people due to the nature of its operation requiring less visual search and attention. While our study design does not allow us to make claims about comparing the mobile and stationary conditions, users reported that overall they felt their performance suffered using the touchscreen in the mobile condition, while their performance using TongueBoard stayed consistent. One participant added that the rhythm of his walking helped him operate TongueBoard more consistently, both in tongue motion and in pace. This comment might explain why the error bar is smaller for the mobile walking condition than for the mobile sitting condition in Figure 6.

Since the mobile application required the classification to be made remotely, network lag contributed to the absolute amount of time taken, making it difficult to compare TongueBoard and the native modality in terms of speed.

In general, TongueBoard takes significantly more time to enter the entire expression (Figure 8, $p < 0.001$). We attribute the slower text entry speed of the system to classification accuracy of the model rather than the interaction technique itself. The majority of these misclassifications occur between one or two words in the vocabulary. Although we replaced "add" with "plus" from the previous study, Figure 7 shows that there are still words in the vocabulary set that produce similar signals. Specifically, "7" and "9" are confused with other two-syllable words and one-syllable words, respectively.

This confusion can be mitigated in several ways including increasing the temporal resolution of the sampling, designing applications to minimize the co-occurrence of conflicting words, and adding a language or word-transition model for free-form input.

While our live evaluation was designed to mimic free-form text input, we find that our system is suitable as a subtle alternative for low-bit item selection or notification

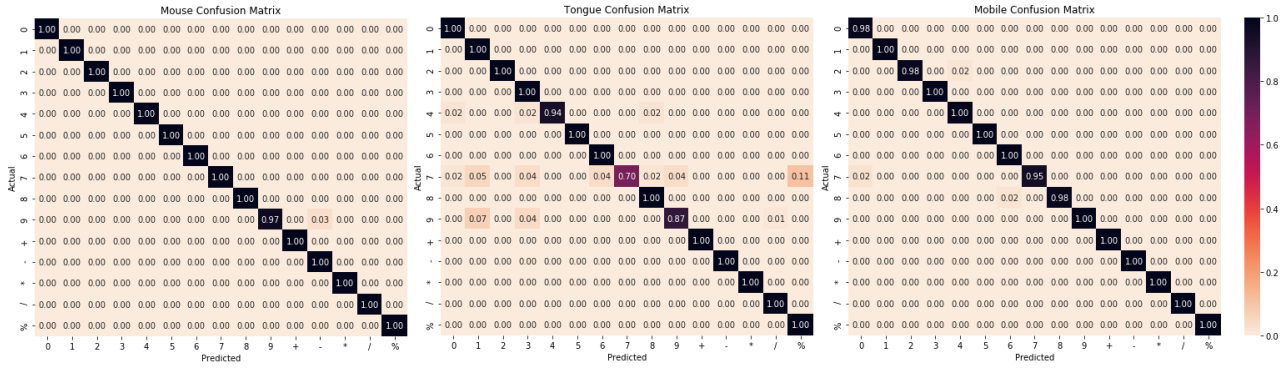


**Figure 7: Confusion Matrices by Input Mode**

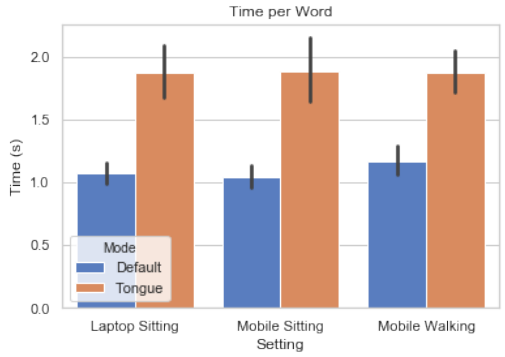


**Figure 8: Time needed to input a single word by setting and modality. Default refers to the pointing stick (Laptop), mouse (Mobile Sitting), and touchscreen (Mobile Walking).**

response interactions. This hypothesis is supported by our findings in Figure 6, where for small $n$, our method achieves similar speeds to that of traditional inputs. In addition, we find that the majority of misclassifications during the live evaluation was due to two similar words in the vocabulary set. By using a more limited vocabulary with distinctive words, the accuracy of the system can be further improved.

## 5 LIMITATIONS AND FUTURE WORK

Currently, TongueBoard is built using the CompleteSpeech SmartPalate system, which is a commercially available palate sensor used for speech therapy. TongueBoard relies on an array of 124 sensors positioned at the top of the user's mouth to detect tongue placement using a mechanism similar to that of orthodontic retainers. Reliable tongue placement sensing requires that the sensor array be held steadily in place while performing gestures and during speech. Each SmartPalate mouthpiece is custom built to a dental impression of the user, which makes it difficult to mass-produce and construct generalizable models for gesture detection and silent speech recognition. An alternative to this approach would be to mimic the design of mouth guards used for bruxism (teeth-grinding) treatment. These products require that users close their mouth to hold the guard in place with their teeth; however, this action is consistent with the applications of subtle gestures and silent speech, where minimal visible movement is desired.

While the sensing array itself is able to fit completely and invisibly inside the user's mouth, the mouthpiece is wired to an external DataLink unit that both decodes the sensor data and powers the sensor array. In the future, low-powered or self-powered sensing mechanisms such as piezoelectric film and forms of wireless power transfer (i.e. inductive coupling) can be used to remove the need for an external power source. To enable wireless transmission of data, a low-powered BLE (Bluetooth Low Energy) chip can be used to transmit sensor data to an external smartphone or computer.

Furthermore, the silent speech accuracy can be improved by using a more sophisticated recognition model. A language model can be implemented to improve free-form text input, and more data could be used to train a user-independent model. Specifically, additional data collected could be used to explore a number of implications, such as the feasibility of performing gestures that are entirely invisible to bystanders. From the users' perspective, performing the gestures was noted to be no more strenuous than normal talking. Indeed, while collecting data to train the recognizer, the users were asked to imagine speaking naturally while mouthing the words. TongueBoard may in fact be less strenuous than normal talking, since it is possible to keep the jaw closed while gesturing, requiring less muscle movement overall.

As the purpose of this study was to investigate the accuracy and expressiveness afforded by sensing of tongue-placement, these improvements that would make the system more appropriate for real-world use were not explored.

## 6 CONCLUSION

In this paper, we present TongueBoard, a tongue interface for subtle gestures and silent speech. We use a commercially available palate sensor intended for speech therapy and repurpose it to recognize a mid-sized silent speech and tongue gesture vocabulary.

We evaluate TongueBoard in an offline study, quantifying the classification accuracy of the system, and a live user study, comparing the system's input speed and accuracy to traditional input modalities. Our results show that the system achieves high accuracy for limited vocabulary sets - with an overall accuracy of 84.36% and 91.01% for native English speakers. In addition, for use cases such as subtle input and notification response, we find that our system is able to recognize speech from deaf users with much higher accuracy (77.76%) when compared to existing audio-based transcription methods.

We evaluate TongueBoard in a live user study to compare it against traditional input modalities for stationary and moving contexts. TongueBoard achieves an average information transfer rate of 3.78 bits per decision (number of choices = 17, accuracy = 97.1%) and 2.18 bits per second, enabling comparable interaction speeds to mouse and touchscreen interfaces.

## REFERENCES


[1] Toshiyuki Ando, Yuki Kubo, Buntarou Shizuki, and Shin Takahashi. 2017. CanalSense: Face-Related Movement Recognition System based on Sensing Air Pressure in Ear Canals. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. ACM, 679–689.

[2] Daniel Ashbrook, Patrick Baudisch, and Sean White. 2011. Nenya: subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2043–2046.

[3] 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.

[4] Abdelkareem Bedri, David Byrd, Peter Presti, Himanshu Sahni, Zehua Gue, and Thad Starner. 2015. Stick it in your ear: Building an in-ear jaw movement sensor. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, 1333–1338.

[5] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: using wearable sensors to detect eating episodes in unconstrained environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 37.

[6] Abdelkareem Bedri, Himanshu Sahni, Pavleen Thukral, Thad Starner, David Byrd, Peter Presti, Gabriel Reyes, Maysam Ghovanloo, and Zehua Guo. 2015. Toward silent-speech control of consumer wearables. *Computer* 48, 10 (2015), 54–62.

[7] Jeffrey P Bigham, Raja Kushalnagar, Ting-Hao Kenneth Huang, Juan Pablo Flores, and Saiph Savage. 2017. On How Deaf People Might Use Speech to Control Devices. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 383–384.

[8] Enrico Costanza, Alberto Perdomo, Samuel A Inverso, and Rebecca Allen. 2004. EMG as a subtle input interface for mobile computing. In *International Conference on Mobile Human-Computer Interaction*. Springer, 426–430.

[9] Marco Cuturi. 2011. Fast global alignment kernels. In *Proceedings of the 28th international conference on machine learning (ICML-11)*. 929–936.

[10] Bruce Denby, Thomas Schultz, Kiyoshi Honda, Thomas Hueber, Jim M Gilbert, and Jonathan S Brumberg. 2010. Silent speech interfaces. *Speech Communication* 52, 4 (2010), 270–287.

[11] João Freitas, António Teixeira, Miguel Sales Dias, and Samuel Silva. 2017. *An Introduction to Silent Speech Interfaces*. Springer.

[12] João Freitas, António JS Teixeira, and Miguel Sales Dias. 2014. Multimodal Corpora for Silent Speech Interaction.. In *LREC*. 4507–4511.

[13] Masaaki Fukumoto. 2018. SilentVoice: Unnoticeable Voice Input by Ingressive Speech. In *The 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, 237–246.

[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 *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 255–258.

[15] Valentin Goverdovsky, Wilhelm von Rosenberg, Takashi Nakamura, David Looney, David J Sharp, Christos Papavassiliou, Mary J Morrell, and Danilo P Mandic. 2017. Hearables: Multimodal physiological in-ear sensing. *Scientific reports* 7, 1 (2017), 6948.

[16] Christian Holz and Edward J Wang. 2017. Glabella: Continuously sensing blood pressure behavior using an unobtrusive wearable device. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 58.

[17] Scott E Hudson, Chris Harrison, Beverly L Harrison, and Anthony LaMarca. 2010. Whack gestures: inexact and inattentive interaction with mobile devices. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*. ACM, 109–112.

[18] Yan Ji, Licheng Liu, Hongcui Wang, Zhilei Liu, Zhibin Niu, and Bruce Denby. 2018. Updating the Silent Speech Challenge benchmark with deep learning. *Speech Communication* 98 (2018), 42–50.

[19] Dimitri Kanevsky, Sagar Savla, and Thad Starner. 2018. Self-managed Speech Therapy. *Technical Disclosure Commons* (August 2018).

[20] Anjuli Kannan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, Laszlo Lukacs, Marina Ganea, Peter Young, et al. 2016. Smart reply: Automated response suggestion for email. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 955–964.

[21] Arnav Kapur, Shreyas Kapur, and Pattie Maes. 2018. AlterEgo: A Personalized Wearable Silent Speech Interface. In *23rd International Conference on Intelligent User Interfaces*. ACM, 43–53.

[22] Jeonghee Kim, Hangue Park, Joy Bruce, Diane Rowles, Jaimee Holbrook, Beatrice Nardone, Dennis P West, Anne Laumann, Elliot J Roth, and Maysam Ghovanloo. 2016. Assessment of the tongue-drive system using a computer, a smartphone, and a powered-wheelchair by people with tetraplegia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 24, 1 (2016), 68–78.

[23] Davis E. King. 2009. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10 (2009), 1755–1758.

[24] Balz Maag, Zimu Zhou, Olga Saukh, and Lothar Thiele. 2017. BARTON: Low Power Tongue Movement Sensing with In-ear Barometers. In

*Parallel and Distributed Systems (ICPADS), 2017 IEEE 23rd International Conference on*. IEEE, 9–16.

[25] Denys JC Matthies, Bernhard A Strecker, and Bodo Urban. 2017. Earfieldsensing: a novel in-ear electric field sensing to enrich wearable gesture input through facial expressions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1911–1922.

[26] Edward F Melcer, Michael T Astolfi, Mason Remaley, Adam Berenzweig, and Tudor Giurgica-Tiron. 2018. CTRL-Labs: Hand Activity Estimation and Real-time Control from Neuromuscular Signals. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, D303.

[27] Geoffrey S Meltzner, James T Heaton, Yunbin Deng, Gianluca De Luca, Serge H Roy, and Joshua C Kline. 2018. Development of sEMG sensors and algorithms for silent speech recognition. *Journal of neural engineering* (2018).

[28] 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.

[29] Simon T Perrault, Eric Lecolinet, James Eagan, and Yves Guiard. 2013. Watchit: simple gestures and eyes-free interaction for wristwatches and bracelets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1451–1460.

[30] Ming-Zher Poh, Nicholas C Swenson, and Rosalind W Picard. 2010. Motion-tolerant magnetic earring sensor and wireless earpiece for wearable photoplethysmography. Institute of Electrical and Electronics Engineers.

[31] Julie Rico and Stephen Brewster. 2010. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 887–896.

[32] T Scott Saponas, Chris Harrison, and Hrvoje Benko. 2011. PocketTouch: through-fabric capacitive touch input. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 303–308.

[33] T Scott Saponas, Daniel Kelly, Babak A Parviz, and Desney S Tan. 2009. Optically sensing tongue gestures for computer input. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. ACM, 177–180.

[34] Jeremy Scott, David Dearman, Koji Yatani, and Khai N Truong. 2010. Sensing foot gestures from the pocket. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, 199–208.

[35] Brendan Shillingford, Yannis Assael, Matthew W Hoffman, Thomas Paine, Cían Hughes, Utsav Prabhu, Hank Liao, Hasim Sak, Kanishka Rao, Lorrayne Bennett, et al. 2018. Large-Scale Visual Speech Recognition. *arXiv preprint arXiv:1807.05162* (2018).

[36] Ke Sun, Chun Yu, Weinan Shi, Lan Liu, and Yuanchun Shi. 2018. Lip-Interact: Improving Mobile Device Interaction with Silent Speech Commands. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. ACM, New York, NY, USA, 581–593. https://doi.org/10.1145/3242587.3242599

[37] Michael Wand, Jan Koutník, and Jürgen Schmidhuber. 2016. Lipreading with long short-term memory. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 6115–6119.