TYTH-Typing On Your Teeth: Tongue-Teeth Localization for Human-Computer Interface

Phuc Nguyen
University of Colorado Boulder
vp.nguyen@colorado.edu

Nam Bui
University of Colorado Boulder
nam.bui@colorado.edu

Anh Nguyen
University of Colorado Boulder
anh.tl.nguyen@colorado.edu

Hoang Truong
University of Colorado Boulder
hoang.truong@colorado.edu

Abhijit Suresh
University of Colorado Boulder
abhijit.suresh@colorado.edu

Matt Whitlock
University of Colorado Boulder
matthew.whitlock@colorado.edu

Duy Pham
University of Colorado Denver
duy.pham@ucdenver.edu

Thang Dinh
Virginia Commonwealth University
tndinh@vcu.edu

Tam Vu
University of Colorado Boulder
tam.vu@colorado.edu

ABSTRACT

This paper explores a new wearable system, called *TYTH*, that enables a novel form of human computer interaction based on the relative location and interaction between the user's tongue and teeth. *TYTH* allows its user to interact with a computing system by tapping on their teeth. This form of interaction is analogous to using a finger to type on a keypad except that the tongue substitutes for the finger and the teeth for the keyboard. We study the neurological and anatomical structures of the tongue to design *TYTH* so that the obtrusiveness and social awkwardness caused by the wearable is minimized while maximizing its accuracy and sensing sensitivity. From behind the user's ears, *TYTH* senses the brain signals and muscle signals that control tongue movement sent from the brain and captures the miniature skin surface deformation caused by tongue movement. We model the relationship between tongue movement and the signals recorded, from which a tongue localization technique and tongue-teeth tapping detection technique are derived. Through a prototyping implementation and an evaluation with 15 subjects, we show that *TYTH* can be used as a form of hands-free human computer interaction with 88.61% detection rate and promising adoption rate by users.

CCS CONCEPTS

• **Human-centered computing** → **Keyboards**; **User interface design**; *Text input*;

KEYWORDS

Human Computer Interaction (HCI), Tongue-Teeth Typing, Wearable Devices, Brain-Muscles Sensing, Skin Deformation Sensing

1 INTRODUCTION

First invented in the 18th century [20], typing on a keyboard has been arguably the most common form of human-machine interface [19] thanks to its intuitiveness, convenience, efficiency, and reliability. Since this form of interface relies on the dexterity of hand movement to hit the correct key on a keyboard, it works best when the user is stationary, can see the keyboard, has at least one hand available for typing, and is not fully cognitively occupied. Though these conditions are commonly satisfied when a user is using stationary devices (e.g. a personal computer), it is often not the case in many mobile scenarios. For example, typing on a keypad of a mobile device, especially on a virtual one, while walking or driving often causes typos and distractions [30]. This is because users have to hold the phone, locate the key, and coordinate their hand to hit the correct key all at the same time. In addition to mobile scenarios, typing on a keyboard is not a viable solution for users whose hands are fully occupied (e.g. operating a machine or carrying objects), who have difficulty in coordinating their hand
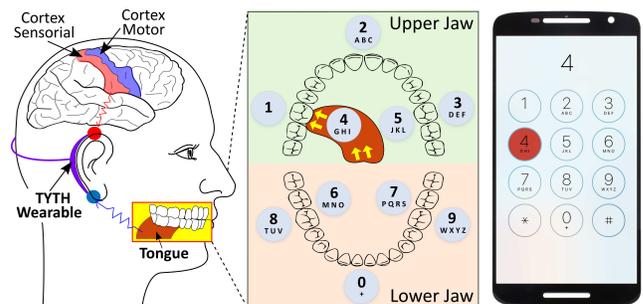
Figure 1: *TYTH*'s Overview

(e.g. quadriplegic, Parkinson, and Amyotrophic Lateral Sclerosis (ALS) patients), or whose physical contact with a keyboard can pose significant risk (e.g contamination for surgeons and nurses during surgery [67]). Therefore, hands-free interfaces are desirable. An alternative is a speech-to-text interface as pervasively found in Apple Siri and Amazon Alexa, among other commercial systems. However, this technique cannot be used in a noisy environment or in situations where speaking out loud to a device can compromise a user's privacy such as in a meeting, in a carpool, and other public settings.

We aim to develop a hands-free computer interface in which a user can privately interact with machines through a wearable device that he or she can comfortably wear and use in everyday life. To that end, we take the first step to build a tongue-on-teeth typing system, called *TYTH* (Typing on Your TeetH)[1], a head-mounted device that can be worn from behind the user's ears to capture the relative location and interaction between user's tongue and teeth. This form of interaction is analogous to using a finger to type on a keypad except that the tongue substitutes for the finger and the teeth is for the keypad. *TYTH* can potentially benefit many applications in which it serves as (1) an accessible means of input for ALS, quadriplegic, and non-oral Parkinson patients, (2) a private text entry and operating input method in public settings, (3) a less distracting phone interaction method while driving, (4) an authentication and input method for VR systems, to mention a few.

Given the potential advantage of a tongue-based interface, a large body of research has been devoted to the topic. The majority of existing approaches including TongueDrive [21] and Tongible [35] require a foreign object to be inserted inside user's mouth. While this design yields a high level of accuracy and is tolerated by some patients, it is not desirable for most users to have a foreign object inside their mouth. TongueSee [80] makes a tongue-based interface more practical by solving that problem. It is able to recognize 6 tongue gestures by capturing electromyography (EMG) signal using 8 sensors mounted on the lower jaw location. However, its usability is limited due to the need to mount such electrodes on the jaw. Tongue-in-Cheek [13] uses a 10 GHz wireless signal to detect four tongue directions (left, right, up, and down) and two modes (tap and hold). The wearable device requires three pairs of 10 *GHz* transceiver antennas placed at three locations around the user's cheeks. Such sensor sizes and placements may prevent the system from being socially acceptable for daily uses. In addition, these techniques can only capture a fixed set of trained gestures as opposed to localizing the interaction between teeth and tongue. We defer an extensive discussion of other existing tongue-computer interfaces and hands-free technologies together with their limitations to a later section (Sec. 9).

This work explores a new wearable system to capture the interaction between tongue and teeth from behind the user's ears. We answer the following key questions in order to realize such a system: what kind of bio-signals can we capture from behind the ear that are valuable for sensing the tongue-teeth relationship? what are the best locations behind the ears to reliably sense those biosignals? how can we derive a technique to detect when a tongue taps on teeth? and how can we locate the area on the teeth that the

tongue taps on? In particular, we study the neurological and anatomical structures of the tongue and teeth to design *TYTH* so that the obtrusiveness and social awkwardness caused by the wearable is minimized while maximizing its accuracy and sensing sensitivity. We identify the positions behind the ears to use for sensing the brain signals (EEG) and muscle signals (EMG) that control tongue movement sent from the brain and to capture the miniature skin surface deformation (SKD) caused by tongue movement. We use traditional bio-electrical sensing technique to capture EEG an EMG signals and propose to use capacitive sensing technology to capture the low-frequency signal from the skin surface deformation (SKD). We model the relationship between tongue movement and the signals recorded, from which a tongue tap locating technique and tongue-teeth tapping detection technique are derived. Through a prototyping implementation and evaluation on 15 subjects, we show that *TYTH* can be used as a from of hands-free human computer interaction with 88.61% detection rate with potentially high adoption rate by users.

In summary, we make the following contributions through this paper:

- Identifying a set of biosignals including EEG, EMG, and SKD captured from behind the ear that are valuable for sensing the tongue-teeth relationship through a thorough study of the neurological and anatomical structures of the tongue movement and teeth (Sec.2)
- Identifying and validating the positions behind the ears to sense these three signals based on which a design of an unobtrusive wearable device is developed to capture tongue movement with minimum social awkwardness (Sec. 6).
- Developing a set of multi-modality sensing algorithms to locate tongue tapping areas and detect tongue-teeth tapping events based on the three signals captured. We do so by analyzing and modeling the relationship between tongue movements and their corresponding muscle contractions and the brain activity signals captured by our sensors (Sec. 5)
- Prototyping *TYTH* using 3D printed components, a custom built electronic circuit, and off-the-shelf electronics to prove the feasibility of capturing EEG, EMG, and SKD from the proposed locations. (Sec. 6).
- Evaluating the system with 15 users to confirm the feasibility of the proposed solution and conducting a user study with 9 users to learn the usability of the system (Sec. 7).

2 BIO-SIGNALS RELATED TO TONGUE MOVEMENTS

Tongue Movement's Anatomy. The tongue is one of the most complex anatomical structures of the human body. Its movement can be classified into three broad categories: *Retract*, *Protrude* and *Articulate* which can be further classified into different categories as summarized in Fig. 2 [61].

The retract pattern is characterized by the posterior movement of the tongue with limited change in shape. It is mediated by the extrinsic muscles, specifically the Hyoglossus (*HG*) and Styloglossus (*SG*). Both of these muscles receive input from the Hypoglossal (CN XII) cranial nerve. This retraction results in the shortening of the blade of the tongue which is oriented lengthwise and can

---

[1]*TYTH* is pronounced similar to "teeth" - /tēth/
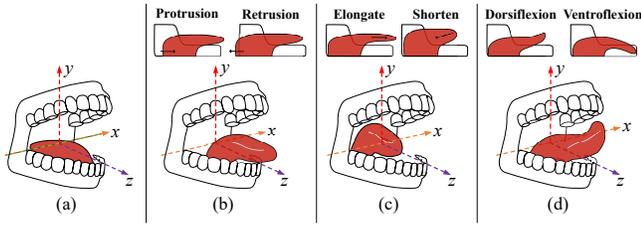
Figure 2: Tongue gestures: (1) Rest, (2) Protrude, (3) Retract, (4) Artriculate

be attributed to the superior longitudinal (*SL*) and inferior longitudinal (*IL*) muscles. While the posterior movement of the tongue refers to retraction, the anterior movement of the muscles refers to protrusion.

The second pattern, protrusion, can be caused by the posterior fascicles of the genioglossus muscle (*GG*) muscle. The *GG* muscle is the largest muscle in the tongue musculature [71] and is innervated by the hypoglossal (CN XII) cranial nerve. Elongation of the tongue can be caused by contraction of the vertical and transverse (*T|V*) muscles.

The third type of pattern is the *Articulate* pattern. As the name suggests, these movements are mostly associated with the motor function of articulation. Dorsiflexion, which is characterized by the superior bending of the tongue tip, is caused by the contraction of the superior longitudinal (*SL*) muscles. Ventroflexion, on the other hand is the inferior bending caused by contractions of inferior longitudinal (*IL*) muscles. Retroflexion is the superior movement of the tongue base, which combines the superior and posterior pull of the *SG* muscle aided by depression of the mid tongue by vertically oriented muscle fascicles of the *GG*. Understanding this anatomical knowledge helps us derive which muscles are are needed for a certain tongue's posture or tongue typing location on the teeth. We will discuss this relationship in the upcoming section.

Bio-signals generated by tongue movements. A single movement of the tongue can trigger different types of bio signals. These bio signals include different types of electric potentials emanating from brain and muscles cells. The electrical activity in the brain is called Electroencephalogram (EEG) signal. Neurons are the fundamental processing units of the brain. Neurons consist of dendrites, cell body (the soma) and axon hillock. When a group of neurons is active at any particular moment in time, they produce an electrical field potential which can be measured on the skin (*EEG signal*) [42].

Secondly, though muscles are regularly used in daily activity, we often do not consciously control their actions. However, skeletal muscles work under voluntary control. The skeletal muscles are composed of bundles of muscle fibers. Muscle fibers are long cylindrical cells containing several nuclei muscles will contract or relax when they receive signals from the nervous system. The muscles involved in the movement of tongue include but not limited to *HG, GG, SG, T|V*. The electrical activity in response to a nerve's stimulation of the muscle or just the muscle response is *EMG signal*.

Thirdly, skin surface deformation, namely *SKD signal*, represents the change of the surface of the skin during muscle activities. In addition, during tongue movement, the skin surface drastically changes at some major locations on the human face. This skin deformation information is useful in identifying tongue location and
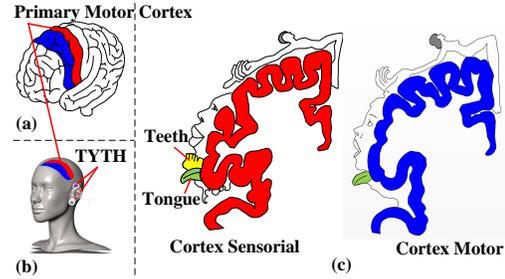
Figure 3: (a) Human brain, (b) Primary Motor Cortex, and (c) Brain regions that control tongue and teeth

direction. We introduce a novel way to capture such surface deformation information using capacitive sensing technology. The idea is to find the capacitance variations created by skin deformation. We will discuss the technique in the upcoming section.

Understanding this knowledge helps us to identify some potential locations to place the sensor for capturing tongue movement. In the next section, we will discuss our choice on selecting the back of the ear location to capture the signal from the muscle and brain signal related to tongue movement.

3 SENSING TONGUE MOVEMENTS AT THE BACK OF THE EAR

We utilize the EEG, EMG, and SKD signals generated by tongue movement to recognize the teeth area tapped by the tongue. These three main signals are captured at different places on the human head. However, we found that the back of the ear is the best location to place our sensor because it is close to both the brain signal source (EEG) that controls the tongue (Primary Cortex), and the muscles that involved for tongue movement (*HG, SG*).

Tongue's EEG signal. Primary motor cortex plays a vital role in controlling tongue movement by innervating hypoglossal, vagus and facial cranial nerves [17]. It is located in the frontal lobe of the brain along the precentral gyrus. Any movement of the tongue is coupled with firing of neurons on the primary motor cortex which receives somatosensory information through the efferent fibres of the tongue. The firing of neurons creates an electric field potential which can be measured using the EEG technique. Vanhatalo et al. [77] designed a study to characterize the EEG potentials related to tongue movement. Results indicate that different types of tongue movements can be characterized differently in terms of electrical potential. In addition, the study also illustrates that significant scalp potential is caused even by modest tongue movements meaning it is possible to differentiate tongue movements based on change on scalp potential. The scalp current source was identified [77] to have the highest density near ear canals and orbital fossae. In addition, our goal of placing EEG sensors at the top of outer ear location is to capture the teeth signal generated by cortex sensorial brain where the tongue is pressing against the teeth. Hence, we believe that EEG data from back of the ear is crucial to classify different tongue movements. The EEG signal related to human tongue movement are found to be in the range of 10 *Hz* [41] to 40 *Hz* [56].

Tongue's EMG signal. Many of the tongue's extrinsic muscles are attached to the hyoid bone, located in the anterior midline of the neck between the chin and thyroid cartilage. This makes the hyoid
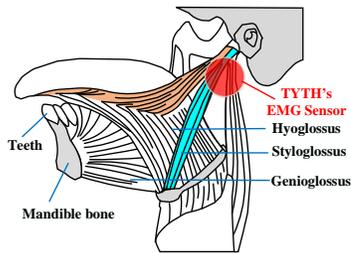
Figure 4: Tongue extrinsic muscles.

bone a prime site for all tongue's EMG studies [63, 80]. However, EMG studies are limited to measuring the electrical potential generated by the muscle cells. In order to accurately measure tongue movement, the electrical activity in the cranial nerves in addition to EMG must be measured.

The sensor locations should be concentrated on locations where we can identify the evoked potential due to the signals that innervate the different muscle movements. In addition to identifying the signal, we should classify them into the corresponding movement pattern. One of the most important regions of interest is the back of the ear where the different nerves originating from the brain start converging towards the hyoid bone. As illustrated in Figure 4, we found that the back ear location is where we could capture the EMG signals from *HG* and *SL* muscles. Even though these two are not fully connected with the tongue intrinsic muscle as the *GG* muscle, the EMG signals generated by tongue movement can be clearly captured by the sensor below to the ear canal.

Skin surface deformation signal. The muscle contraction during tongue movement creates a minor changes on skin surface. More specifically, the relaxation/contraction of *HG* and *SL* muscles expands/compresses the skin surface at the ear location where the jaw bone is connected to the human head. Such surface deformation happens strongly at the gap between the lower jaw and the head, where the *TYTH*'s SKD sensor is placed.

4 *TYTH*'S SYSTEM OVERVIEW

Above analysis suggests to have 3 sensors at each ear to capture the signal from the brain activity, muscle contraction, and skin surface deformation. Therefore, we could place 6 sensors at behind the human ear to capture different information that are related to tongue movements. We found that these 6 sensors can be placed at 4 locations. The EEG sensor is placed at the top of the outer ear to capture the EEG signal. The EMG and SKD sensors can be co-located at the bottom of the outer ear area to capture the EMG signal and skin surface deformation signals. However, realizing *TYTH*'s idea is difficult due to the following challenges:

- **Unknown relationship between tongue typing and generated bio-signals.** While existing literature tries to map between the EMG signal and the tongue movement from the tongue captured at the lower location, there is no work on conducting the relationship between the tongue movement/pressing and the brain/muscle activities or skin deformation.
- **Brain and muscle signals are weak**. Capturing these signals is extremely challenging because their amplitudes are too weak (at the μV scale). Accurately capturing them requires carefully

design of the sensing hardware and signal processing algorithms to capture them.

- **Skin surface deformation is difficult to capture by vibration sensor, i.e., piezoelectric.** The skin surface deformation is difficult to capture, we have tried with different types of vibration sensors but none of them provides sufficient level of sensitivity. The SKD signal requires a specific design and sensor to capture such tiny movement of the skin surface.

TYTH system

The overall design of *TYTH* is illustrated as in Figure 5. The *TYTH* system includes a wearable device for users to wear to their ear. This device has ability of sensing a very small electrical potential and skin surface deformation. These signals are then go through an analog amplifier to amplify the signal of interest. This analog amplifier usually has gain of 24. Also, the device is implemented a notch filtering to remove the impact of electricity noise. The *TYTH* wearable device use bluetooth communication to transmit the sensing data to a host computer.

The host computer receives the streaming data from all the sensors, analyzes them, and predict the ID of the teeth area where the tongue is pressing on. There are three main components inside the *TYTH*'s software architecture including (1) Pre-processing, (2) Typing Detector, and (3) Typing Recognizer.

- **Pre-processing components** are used to filter out the environment noises using Notch and Band-pass filters. We then apply a low-rank decomposition analysis to extract the main structure of the bio-electrical signals to obtain EEG and EMG data (Sec. 5.1).
- **Typing detector** is used to detect the tongue movement and tongue typing event. We apply wavelet analysis to capture the tongue movement event and utilize Short Time Fourier transform (STFT) to detect the typing events (Sec. 5.2).
- **Typing recognizer** is used to recognize which teeth the tongue is tapping on. Then, the recognized input will be mapped into the key map to generate the input key and feedback to user (Sec. 5.3, 5.4).

■ **Turn ON/OFF *TYTH*'s input**. To differentiate between input data using *TYTH*'s device and other interference movement such as talking and eating, we design a triggering mechanism to allow user to turn on/off *TYTH*'s input. We build a binary classifier to detect the "gritting the teeth" event. When the user grits his/her teeth, the combined EEG, EMG, and SDK signals has a unique signature. This signature is used to start and stop *TYTH*'s input.

In the next section, we will present the key algorithms of *TYTH* including the low-rank decomposition technique to remove the noise component of the signal, followed by an algorithm to detect the tongue pressing event. We will also discuss our classification and localization algorithm in details.

5 ALGORITHMS

5.1 Low-rank Decomposition for Signal Denoising and Extraction

Most bio-signals are compact and condense representation in some domains called sparsity [3]. Low-rank spare matrix decomposition (low-rank recovery, rank-sparsity incoherence, etc.) is well-known for signal reconstruction in the presence of low SNR. Distorted
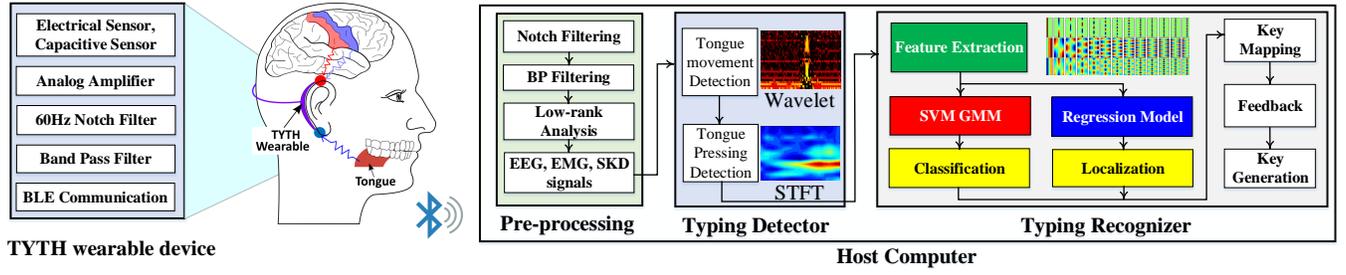
Figure 5: *TYTH*'s system overview

data (noisy EMG, EEG signals) are sparse distributed representation among the interested signals which are the electrograms (EMG/EEG) in our tongue typing system. The missing data points can be well reconstructed by stacking multiple EMG/EEG samples together. If the number of samples is much smaller than the dimension, the observed matrix M is low-rank. Mathematically, it is formulated as the summation of a low-rank cascaded bio electrical matrix L_b and a sparse noise S_n. To derive the close forms of L_b and S_n from the given matrix M, [6] minimizes an in-equality constraint ill-posed problem as follows.

$$minimize \ \|L_b\|_* + \|S_n\|_1 \ s.t. M = L_b + M_n, \quad (1)$$

where $\|L_b\|_*$ defines the rank of matrix L_b and $\|*\|_1$ is the L1 norm. The key idea is to recover L_b and S_n by solving a complex optimization problem. One might concern about the latency of the low-rank computation. We will show that as low-rank computation only happens on single dimension, rather than multiple dimension in image processing, it results in millisecond response even the algorithm is implemented in compute-intensive software Matlab (Sec. 7). We apply RPCA to all input channels including 2 EEG sensors, 2 EMG sensors, and 2 skin deformation sensors. The EEG and EMG signals are very weak and require a dedicated signal processing technique to carefully remove the noise why avoiding losing the brain and muscle bio-electric signal activity.

The RPCA technique removes most of the high frequency noise in the signal, it, however, is not sufficient to extract the EEG and EMG information because these two signals are fall into the same category that cannot be distinguished by RPCA. To overcome this issue, the spare vector representing the atoms and their associated weights for the best EEG/EMG signals can be recovered by solving the optimization problem. The signal is extracted based on the characteristics of the recovered sparse vector. The signatures of the bio signals belonging to the same class are assumed to approximately lie in a low-dimensional subspace. In order words, the EEG signals that related to each movement will be lied on one subspace, and the EMG signal is lied into another subspace.

Every sequence of bio-signal $f(x)$ can be represented using a basic function with Gabor atoms as following:

$$f(x) = \sum_{i=1}^{N_D} \delta_i g_i, \quad (2)$$

where N_D is the number of atoms in the Gabor dictionary and g_i is one of the atoms in the dictionary where δ_i is the coefficient

of corresponding to g_i computed by Matching Pursuit (MP) algorithm [37]. In other words, mixed bio-signals are sparse in the Gabor dictionary. From MP computation results, the first component of the results would include the main structure of the data and the rest presents the details of the data.

$$f(x) = f(x_{\text{main structure}}) + f(x_{\text{detail structure}})$$
$$\Leftrightarrow f(x) = f(x_{\text{EEG signal}}) + f(x_{\text{EMG signal}}) + f(x_{\text{noise}})$$
$$\Leftrightarrow f(x) = \sum_{\theta_1}^{\theta_M} \delta_{\theta_i} g_{\theta_i} + \sum_{\theta_{M+1}}^{\theta_N} \delta_{\theta_i} g_{\theta_i} + \sum_{\theta_{N+1}}^{\theta_{N_D}} \delta_{\theta_i} g_{\theta_i} \quad (3)$$

In the above equation, the EEG signal is filled into the main structure of the signal and at lower frequency. The EMG and the remaining noise after RPCA represents the detail structure of the signal. The EEG excluded signal is then put into another analysis to extract EMG out of the noise signal. The key idea here is to design the EEG and EMG dictionaries to make sure the sensor capture the proper desired signals. In other words, the dictionaries must make sure that the signal extracted through Matching Pursuit implementation will only keep the low frequency components (EEG, EMG).

In the upcoming section, these main structures will be used to detect the tongue pressing event (Sec. 5.2), and to recognize the teeth region (Sec. 5.3).

5.2 Tongue Pressing Detection

One of the key components in the system is to detect the moment at which the user's tongue is pressed against the teeth. There are two key signatures that are used to detect this movement including (1) the tongue movement and (2) the presence of the brain signal that control the tongue. Firstly, we apply wavelet transformation to detect the discontinuity of the signal where the tongue is moving and pressing against the teeth. Secondly, the system detects the available of the brain signal that control the tongue which varies from around 10 *Hz* [41] to 40 *Hz* [56]. We observed from our data that most of our participants creates a signal from 8 to 12 *Hz* from their brain when pressing again the teeth, which is also matched with the literature [41].

■ **Tongue Movement Detection.** The processed signals from 6 sensors are then put to a movement detection algorithm to confirm the tongue movement event. The spectrogram of the signal captured at an EEG sensor location are shown in Figure 6 when the user moves the tongue 1 time (LEFT), and 2 times (RIGHT). We use wavelet coefficient analysis to detect the movement from the tongue
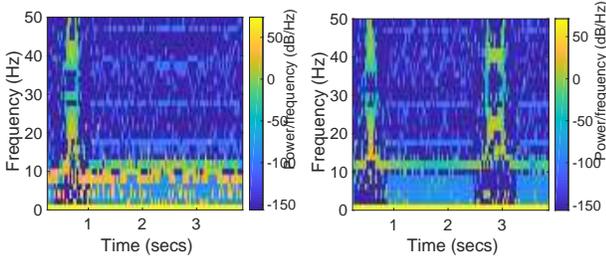
Figure 6: Spectrogram of the EEG signal when there is one and two tongue movements

Figure 7: Tongue movement and typing event captured by EEG sensor

at different sensor. A majority vote mechanism is then used to validate whether the tongue is pressed or not.

In particular, a wavelet, denoted by $w(t)$, maintains local information in both the time and frequency domains. It is defined as a waveform that satisfies the following condition: $\int_{-\infty}^{+\infty} w(t)dt = 0$. The Wavelet Transform [73] uses as the wavelet that satisfies the condition of dynamic scaling and shifting function, $w_{s,p}$,

$$w_{s,p}(t) = \frac{1}{\sqrt{s}} w\left(\frac{t-p}{s}\right) \qquad (4)$$

where $w_{s,p}(t)$ are the integrated and integral transformation signal, s is the scale and p is the shift parameter, which can also be the central location of the wavelet in the time domain. The wavelet can be stretched and translated with flexible windows by adjusting s and p, respectively. The wavelet transform of the wireless received samples $\tilde{r}(t)$ using transform coefficient $W(s,p)$ is calculated as following:

$$\begin{aligned} W(s,p) &= \int_{-\infty}^{+\infty} \tilde{r}_f(t)\overline{w_{s,p}}(t)dt \\ &= \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} \tilde{r}_f(t)\overline{w_{s,p}}\left(\frac{t-p}{s}\right)dt \end{aligned} \qquad (5)$$

where $\overline{w_{s,p}}(t)$ represents the complex conjugate of $w_{s,p}(t)$. The result of the wavelet transform gives us a correlation function of the template signal at different scales (frequency bands) in both the time and frequency domains. As in Equation 5, the correlation function $W(s,p)(t)$ has two main features as follows. (1) The time resolution is high with high frequencies while the frequency resolution is high with low frequency signals. When multiplying the high frequency component of the signal with the high frequency of the wavelet, the correlation result will indicate the exact location where it happens. This can be used to identify the very first tongue movement event. (2) As the wavelet has local existence in both time and frequency domain, the point of discontinuity in the signal can be detected with high sensitivity. As the discontinuity (generated by tongue movement) is considered as an event and happens quickly in time, the result of correlation with high frequency wavelet will be readily captured. A tongue movement event is detected if majority of the sensors detect the movement from their own wavelet analysis. Each sensor will mark the data as tongue movement when the coefficient from the wavelet continuous analysis is over a designed threshold. ■ **Tongue Typing Detection.** As seen earlier in Figure 6, the brain activity signal creates a periodic signal that is well-reflected in the
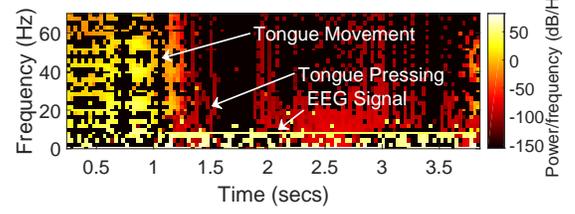
FFT-based spectrogram at 8 - 12 Hz. Conversely, a wavelet transform that is better-suited for capturing transitory phenomena such as a tongue movement is not well-suited for tongue pressing detection. We formulate the brainwave signal in a form of $X sin(2\pi f t + \phi)$. From the received EEG signal $x(t)$, an efficient approximation of the brain signal activity is to identify the dominant frequency E that has maximum power spectrum density (PSD) through the STFT. Then, the approximation of the brain activity frequency f_{brain} is as follows:

$$f_{brain} = \max_{[f_{min} \to f_{max}]} \left(\left| \sum_{k=1}^{N} x(t)e^{-j2\pi f t k} \right|^2 \right) \qquad (6)$$

where N is the number samples. After f is estimated, it can be used to estimate the amplitudes and phases of different signals using the following: $X = \frac{2}{N} | \sum_{k=1}^{N} x(t)e^{-j2\pi f t k}|$, and $\phi = actan$ $\frac{-\sum_{k=1}^{N} x(t)sin(2\pi f t k)}{\sum_{k=1}^{N} x(t)cos(2\pi f t k)}$. In this way, the system obtains the desired quantities X, ϕ, f. The presence of brainwave signal is observable through a short-time Fourier analysis. This event is confirmed when the maximum power distribution of the peak frequency belongs to the range of 8 - 12 Hz.

5.3 Classifying the Typing Area

In this subsection, we present an algorithm to accurately recognize correct pressing areas. We first extract the features from the collected data using MFCC feature extraction technique. We then estimate apply Gaussian Mixture Model (GMM) to extract the mean vector and final descriptor representation. Finally, a Supported Vector Machine (SVM) algorithm with KBF kernel is designed to classify the data. We will evaluate the impact of GMM and KBF in terms of boosting up *TYTH*'s accuracy in Sec. 7.

The pattern of tongue pressing on each teeth location is defined by signals captured from 6 sensors (1500 samples collected for each second). The data are sliced into overlapping chunks using a Hamming window. Then each chunk goes through the feature extraction process where it is convolved with a filter bank to obtain the coefficients as a feature vectors. The Mel filter bank or Mel-frequency cepstral coefficients (MFCC) that we adopt into our system is commonly used in most of signal processing application [15, 44]. At this point, one particular signal generates a matrix of MFCC features where number of rows is the number of chunks and the columns corresponding to the dimension of MFCC features. Here, we not only use the Mel-coefficients but also adopt their first and second order derivative to extend our feature space. The combination of
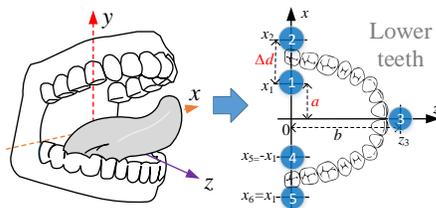
Figure 8: Tongue Localization

Mel-coefficients, delta and double delta shows a very strong improvement in distinguishing the subjects' speech comparing with only Mel-coefficients alone [15].

Each sample now is represented by a set of MFCC feature, we then estimate the distribution of these feature points using the Gaussian Mixture Model (GMM) and extract the mean vector for our final descriptor representation. The common theme of estimating the GMM from a set of data points is by (1) initializing the parameters of distribution, (2) using the Expectation Maximization (EM) algorithm to adjust the model to fit our feature points.

Random initialization is not effective for the EM to converge to the optimal point, especially in the context of using bio-electrical signals which are highly ambiguous and are easily dominated by other types of signals. We adopt the Universal Background Model (UBM) [58] for the initialization step. The UBM model is a GMM model but is trained on the entire dataset. Therefore, using this model for initialization helps to generalize our problem characteristics, and thus helps the GMM adaptation to quickly converge. The processes of getting GMM model based on UBM can be summarized in two stages: (1) using the EM algorithm for the whole training samples to obtain the UBM model, (2) using the Maximum a Posteriori Estimation to fit the UBM with the feature points of each sample to crate the specific GMM model for this particular sample. Recall that we have 6 channels for one posture and the process above is applied for each of them separately.

Finally, the data is classified by Support Vector Machine with different kernels. We evaluate our system using 3 basic kernels: linear, cosine and RBF. The purpose of using multiple sensors to produce multiple channel is important. In our experiment, we show that one sensor alone generates very low accuracy but when all of them are fused together, it significantly improves the performance. We basically average the kernels of 6 channels.

5.4 Tongue Typing Localization

The above classification algorithm provides a fixed set of "pressing areas". In this subsection, we describe a technique to fine-grained locate untrained pressing locations. We developed a localization algorithm that enables TYTH to continuously track the tongue pressing locations even at untrained area.

Different teeth requires different activation muscles as well as brain activity. *TYTH*'s localization algorithm relies on that characteristic. The key idea is to build a regression function that represents the correlation between the input bio-signal and the output x,y,z coordinate of each teeth. Figure 8 illustrates the coordinate system of each of the 6 locations of the lower teeth. Let assume the root of the tongue has the coordinate of $0(0,0,0)$, the distance between the middle of the tongue to the left teeth is a, to the front teeth
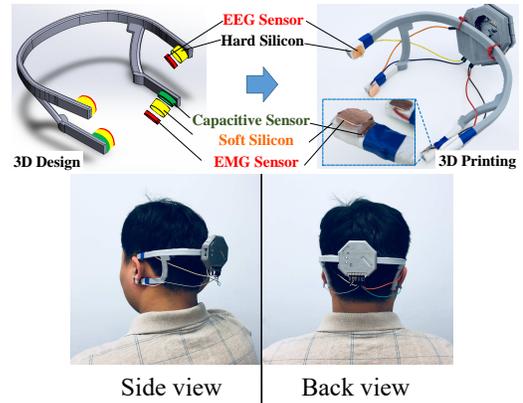
Figure 9: *TYTH*'s 3D design

is b, to the upper teeth is c, and the distance between inside and outside teeth is Δd, the coordinate of each location of interest is as following $1(x_1, 0, 0) \approx 1(a, 0, 0)$. Similarly, we can convert the location of the main 5 sensors at the upper teeth with an assumed distance c from the lower teeth.

Using such information as a prior knowledge, we build a regression model that approximate the x,y,z coordinate of a certain tongue pressing location based on the input from featured signal. It is impossible to have a direct mapping between the input GMM features and the 3 dimensional location because the GMM has 42 dimension in total. Therefore, there needs to be a intermediate step that transform a high dimensional feature GMM into a 3D coordinates. In order to do that, the informative level of each feature dimension is compared to select the best three representative ones. Using the Principal Compnents Analysis (PCA), we can select the coordinates that represent our data, in our problem, we only choose the top 3 coordinates.

First, the whole features are used to extract the coefficience matrix of the PCA. Then each feature vector will be multipled with the first three columns of the coefficience matrix to project it onto the 3D space. We have successfully constructed a reference projection between feature coordinate and real world coordinate for regression. We then apply a linear regression model to interpolate the relation between our ground truth data and the mapping feature location. Note that although the coordination between the biosignal feature are seems to be nonlinear, it is proved that every non-linear regression model can be approximated by a linear model using Taylor's theorem [12]. Also, the regression model only can be applied to a small fixed set of feature from the original data, it is also proved that regression model cannot be done on raw data.

6 SYSTEM IMPLEMENTATION

In this section, we present the current design and prototype including the 3D printed design, the bio-electrical, and skin deformation sensors implementation.

3D printed design. We design wearable ear pieces as shown in Fig. 9 (TOP). The 3D model is printed with flexible material to make sure that the sensors always have a good contact with the human skin for a reliable measurements. Figure 9 (BOTTOM) the actual device was wear by one of our participants.

EEG/EMG sensors. We use copper tape electrode to make a direct contact to the human skin at the top and bottom of the outer ears location as illustrated in Figure. There are four sensors in total. While 2 top electrodes are used to capture the signal from the human brain, the two bottom electrodes are used to capture the EMG signal generated by tongue's extrinsic muscles. These sensors are placed on top of a silicon layer 1 mm to create a conformable contact between the sensor and the human skin. To measure the EEG, EMG signal, a combination differential input and output amplifier techniques are usually preferred. Figure 10 (LEFT) shows an example programmable gain amplifier implementation of and low noise CMOS chip named ADS1299 [25]. ADS1299 is a well-known instrument amplifier as its key analog component to measure the electric potential generated by brain and muscle contraction activities. ADS1299 supports programmable gain amplifier (PGA) of 1, 2, 4, 6, 8, 12, and 24). The ADCs can support 250 sample/s to 16 ksample/s.

Skin surface deformation sensor. Capacitive sensing has been used to estimate physical properties such as touch, proximity, and deformation by measuring the capacitance between two or more conductors. These conductors are originally made by conductive material including metal, foils, transparent films, plastics, rubbers, textiles, inks, paints, or human body [14, 23, 49, 74–76]. We exploit capacitive sensing characteristics to measure the skin surface deformation caused by tongue movement in this paper. A capacitance exists whenever two electrodes are separated each other by a distance Δd. We placed a copper tape at separate with a human skin by a soft and deformable silicon Ecoflex 00-10 from Smooth On [52] at 1 mm. At the bottom electrodes on each ear, there is another electrode at another side of the flexible form capture the skin surface deformation caused by tongue movements. The tongue movement will create a distance changes between the two side of the flexible form. Such movement can be captured using piezoelectric sensor or accelerometer. We experimentally found that the signal obtained from these two traditional methods are not sensitive enough to measure the skin surface deformation. We propose to use capacitive sensing to capture the tiny movement of the skin caused by the tongue behavior.

The key idea is to measure the distance changes between the two electrodes, one is on the skin and another one is on the wearable device. At a stable condition, the capacitance created by two metal plates can be calculated as $C = \frac{\epsilon_0 \epsilon_r A}{d}$, where C is the capacitance in Farads, A is the area in meters square, d is distance between 2 plates in meters, and ϵ is dielectric constant, which is the product of free space ϵ_0 and relative dielectric constant of the material, ϵ_r. When the tongue movement happens, the skin surface deforms, the flexible material in the middle of two copper plate create a change in their area and distance. This generates a change in capacitance which is possibly measurable by capacitive sensor.

Relaxation oscillator is one of a well-known technique to measure capacitance due to its simplicity of operation. Fundamentally, the schematic of relaxation oscillator technique is presented in Figure 10 (b). Any change in capacitance at the measurement pin C_{sensor} is captured using internal Timer_A of the MSP4305969 circuit. The R ladder network creates a reference for comparator that changes with its input when Px.y is high. This reference is
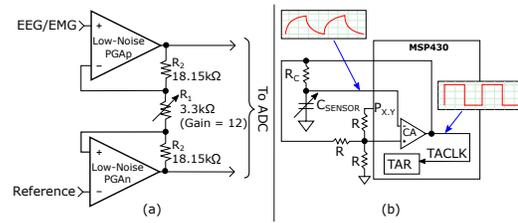
Figure 10: (a) A basic schematic of a channel to measure EEG signal using ADS1299 for the EEG, EMG sensors and (b) Relaxation oscillator capacitance measurement circuit for skin surface deformation sensor.

opposite in polarity to the charge or discharge of the C_{sensor}, resulting in a continuous oscillation. With equal R, the frequency of oscillation is obtained by $f_{osc} = 1/[1.386 \times R \times C]$. f_{osc} can be obtained by counting the oscillation periods over a fixed duration. Then, C_{sensor} is measured through f_{osc}. $Rc = 100k\Omega$ is used.

■ **Putting together.** We prototyped *TYTH* system on a openBCI board [53] for EEG and EMG data collection. We also use MSP430-FR5969 [24] to measure the capacitance variation created by the skin surface deformation. Both device are communicate to a Lenovo ThinkPad T570 [31] laptop through Bluetooth Low Energy device at 115200 baudrate. The openBCI is sampled at maximum sampling rate at 250 Hz, and the MSP430FR5969 is sampled at 10 Hz. The data from openBCI is streamed to laptop through Lab Streaming Layer (LSL) network protocol written on python. The pre-processing and algorithms are implemented on Matlab R2017b. The Matlab and Python data are exchanged using a basic TCP protocol. The signal de-noising, extraction, classification (SVM GMM), localization algorithm are implemented on Matlab.

7 PERFORMANCE EVALUATION

7.1 Experimental Methodology

To evaluate the performance of *TYTH*, we conducted our experiments over 15 participants in a normal office environment. The participants' demographic is summarized in Table 1. In the experimental setup, we first helped the user to wear the device shown in Figure 9 on the back of their head. Specifically, as shown in Figure 9 (RIGHT), the 2 pairs of copper-tape sensors were placed behind the participant's ears in the left and right sides to collect the signals of interest (i.e., EEG, EMG, and SKD signals). *TYTH*'s does not generate any signal to affect the user, it just passively listens to the bioelectrical signals generated by the brain and the tongue muscles, and the capacity change caused by the skin surface deformation. Beyond our system (*TYTH*), we used a camera to record the user's tongue gestures for groundtruth data.

Table 1: Participants' demographic description

Participant Demographics	
Age (years)	18 - 35 years old
Gender Ratio	Male: 11, Female: 4
Head size use	Small: 3, Medium: 8, Large: 4

After the user had *TYTH* correctly situated, we started the study in which the user sat in front of a monitor that instructed him

Figure 11: Different sizes of *TYTH*'s prototype

Teeth Area ID	Description
1,2	inside/outside the lower teeth on the left
3	outside the lower teeth in the front
4,5	inside/outside the lower teeth on the right
6,7	inside/outside the upper teeth on the left
8	outside the upper teeth in the front
9,10	inside/outside the upper teeth on the right
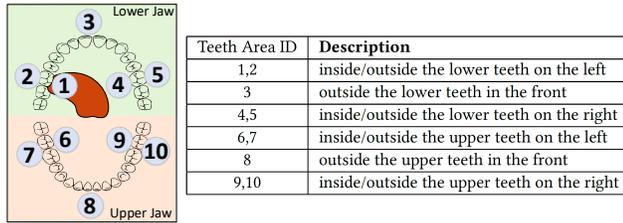
Figure 12: Teeth areas used during our experiment

on when to perform a gesture and when to rest his tongue. From our pilot studies with team members, for each gesture, we found that the users need approximately 3 *s* for performing a gesture. Hence, for each user, s/he will be asked to perform 10 gestures, 20 times for each gesture, 3 *s* each time. The gestures are described in Figure 12 in detail. The users are required to consciously follow the instruction on the screen and the performs the gestures. When the screen says *Press*, the user presses to a requested location after that. When the screen says *Release*, the user releases her/his tongues. The duration between *Press* and *Release* is 1.5 *s*. This is the actual meaningful data of the whole 3 *s*. However, we extend the time to help user easier to follow the instruction, 1.5 *s* data pressing action can be performed flexible shift within 3 *s*. The data collection duration of each user varies from 45 *mins* to 75 *mins*.

After the experiment, we gave the participants a questionnaire to evaluate the feasibility of the *TYTH* and the comfort of doing the suggested gestures. The questionnaire includes (1) *How to like to perform input using TYTH?*, (2) *How difficult to perform tongue gestures using TYTH?*, (3) *How comfortable the tongue gestures were performed?*, (4) *How long can you use TYTH continuously with current form factor?*. To understand the target of *TYTH*'s input system, we ask the users to perform 10 gestures with randomly order as fast as they could. This result will be used as our target for future development of *TYTH* given its rate of 1.5 s/input as its current performance.

7.2 *TYTH*'s System Performance

In this section, we design the experiments to answer the following questions (1) *How accurate TYTH can recognize a trained area?*, (2) *How accurate TYTH can locate the untrained typing location?* (3) *What are the key factors that affect the performance of TYTH?*, and (4) *How do users like TYTH?* We start by validating the performance of *TYTH* in recognizing the ten areas (See Figure 12). Then, we then validate with areas of the teeth the *TYTH* could obtain better performance. We also present the *TYTH*'s performance in localizing the untrained location on the teeth. Furthermore, we discuss the accuracy of the pressing detection algorithm and the energy consumption of the current prototype. We aim to continue develop a system and improve its accuracy and the form factor of the prototype before recruiting ALS patients for clinical trials. In terms of

accuracy, the key errors are from the missing of detailed structures of bio-signals captured by our current sensors. In particular, the current sampling rate for EEG/EMG sensors is 250 *Hz*, and the surface deformation is only 10 *Hz*. This is the limitation of our off-the-shelf components, if we could improve the sampling rate of our analog sensor, the accuracy could be boosted up significantly because more detailed features are captured.

■ **Classification Different Teeth Areas.** We evaluate the system on the data collected from 15 participants. Each user is required to perform 10 typing and 1 resting gestures. Each gestures is repeated for 20 times. The total samples are $20 \times 11 \times 15 = 3300$. Each epoch contains a matrix of 6 columns representing signal from 6 sensors (2 EEG, 2 EMG, and 2 SKD). 75% of data is used for training and the remaining 25% of data is used for testing. The results in the paper are the average accuracy for the whole data set that we collected. We haven't evaluated how the system performance change over time. We reserve temporal performance analysis for future work.

Figure 16 shows the results of the teeth typing area classification algorithm where Teeth Area ID 0 is where the user is at rest. . *TYTH* performance can obtain up to 96.9% of accuracy in detecting a specific area while the overall accuracy is 88.61%. More specifically, the results suggest that the performance of the system depends on the location on the teeth the user one to type, the performance of the location outside the teeth including 2, 4, 6, 8, 10 obtains better performance compared with insides areas. We then conduct a follow up experiment to validate how accurate the system can perform if we independently evaluate the system with only the outside and inside areas. Figure 14 illustrates the *TYTH*'s performance in performing 6 outside area (ID 2, 4, 6, 8, 10 in Figure 12). The system obtains much higher performance with average accuracy of 93.02%. More interestingly, when we only consider the inside teeth locations (ID 1, 3, 5, 7), the system only perform 89.57%. This matches to our studies because most of users state that performing the typing at outside areas are easier than the inside ones. Notes that these 10 locations are used to evaluate the accuracy of our recognition algorithm. To develop an end-to-end application to laptop input or mobile device, one can translate these area ID to keyPress event using Java class (java.awt.Robot) developed by Oracle [54]. For example, key "1" can be map to character "A" by `Robot.keyPress(KeyEvent.VK_A)`.

Confident Level. We validate the robustness of the system based on the confident level of the classification algorithm. We confirm the confident level of *TYTH*'s system by 95% with the accuracy of 86.02% and only varies within 0.62 interval (Figure 16).

■ **Tongue Typing Localization.** Our tongue localization model (built from Sec. 5.4) performs a good results in terms of tracking the tongue location in fine-grained. Given there are only 10 trained location, the results of regression is shown in Figure 17. Assuming we are using normal human mouth size as the input for the regression, let us assume that the distance from the root of the tongue to each location is around 3 *cm*. Based on the results in Figure 17, the system can locate pressing location with 4.5 *cm* error range (3 x 3 x 3 is the maximum error range) at the accuracy of 90%. Given this preliminary result, we believe the more fined grain localization could be obtained when we have more training locations.
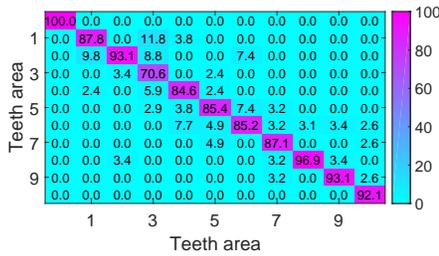
Figure 13: Localizing the tongue typing areas at 10 locations
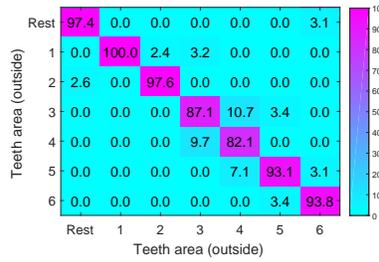
Figure 14: Predicting the tongue typing areas at 6 outside locations
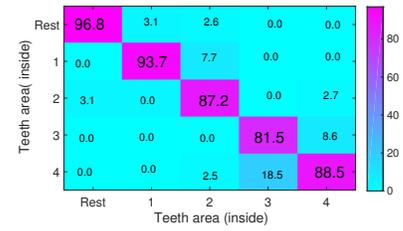
Figure 15: Predicting the tongue typing areas at 4 inside locations
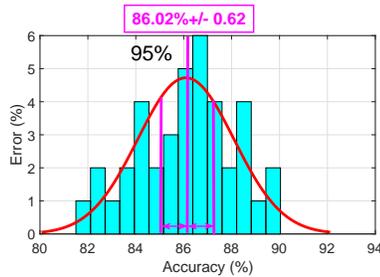
Figure 16: Confident level
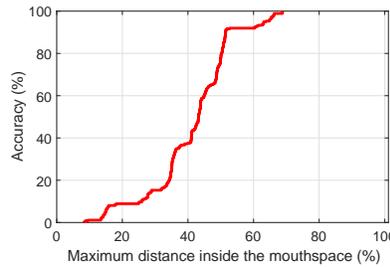
Figure 17: Tongue localization results
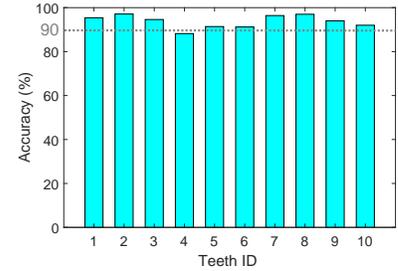
Figure 18: Tongue pressing detection

■ **Tongue Pressing Detection.** We also evaluate the performance of the system in terms of detecting the tongue pressing event. We combine the data that contains the tongue pressing event and the tongue relax data with 50-50 ratio and use it for evaluation. The system obtains up to 97% of accuracy in terms of detecting the typing event using the binary classifier built from wavelet and short-time Fourier transform as illustrated in Figure 18.

■ **Energy Profiling.** To validate the energy consumption of our wearable prototype, we use Moonson power measurement tool [40] to measure the power consumption of our prototype. The openBCI, which is used to sense EEG and EMG signals at 4 channels consumed the average power is 232.82 mW, the average current is 55.53 mA. The MSP430FR5969 and the bluetooth component requires 167.57 mW power, and 36.89 mA current. Figure 19 shows the power consumed by both devices. The power consumption of the prototype is quite high due to the lack of optimization in the hardware selection. For example, *TYTH* uses only 4 analog pins over 16 pins supportted by openBCI [53] board for EMG and EEG measurements. We use only 2 ADC pins over 16 pins supported by MSP430FR5969 [24] for SKD sensing. The optimized circuit include an Analog front-end ADS1299-4PAGR [25] to gather the data from EEG, EMG, and SKD sensor. ADS1299 chip is the main component inside the openBCI [53], we basically remove all unnecessary components to cut down the power consumption. ADS1299 provides an on-chip bandpass and low-pass filter that allows us to map its output directly to computer for processing. This analog chip consumes only 6 mW. To upload the data to the host device, BLE module utilizing C2640R2FPHBT [26] from Texas Instrument can be used. This BLE chip consumes less than 9.9 μW at standby mode, and 550 μW at idle mode, and 1.45 mA + 61 $\mu A/MHz$ at 3.3 V. The radio RX
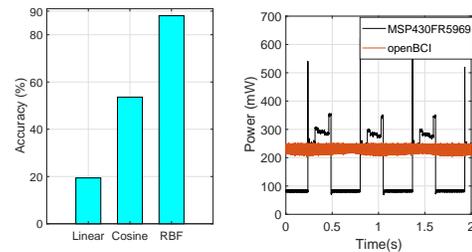
Figure 19: Impact of kernel selection and energy profile of the developed prototype

requires 6 mA, the radio TX requires 6 mA at 0 dBm output and 9 mA at 5 dBm output. The whole system consumes less than 20 mW of power compared with 400.39 mW as our current prototype.

7.3 TYTH's Sensitivity Analysis

■ **Impact of Gaussian Model and Training Size**. The impact of Gaussian Model is shown in Figure 21. In that, we fixed the size of the testing data and vary the size of the training data. The system is currently obtained best performance when the training size is three times larger than the testing size. In addition, without GMM, the system could only obtains upto 67% compared with 88.6% when using GMM. Also, the selection of kernel SVM is also important and defines the accuracy of the system. We used different type of kernel for our classification purposes including (1) Linear, (2) Cosine, and (3) RBF [47]. Figure 19 (a) illustrates the performance of different kernel where RBF is the best fit for our application.
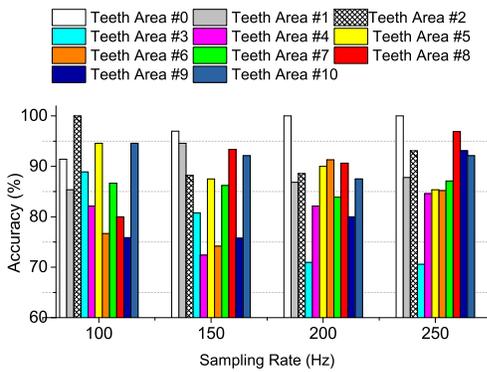
Figure 20: Impact of sampling rate
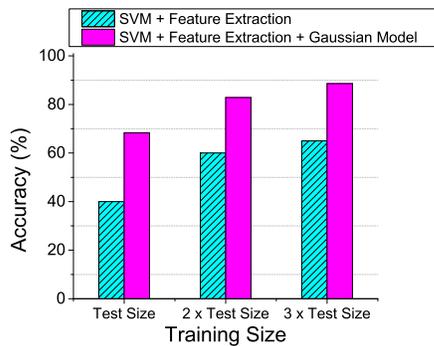
Figure 21: Impact of training size and Gaussian model

◼ **Impact of Sampling Rates** The sampling rate represents the details of signal that the system could capture. Higher sampling rate also means to capture more noise from the environment. We vary the sampling rate from 100 *Hz* to 250 *Hz* and report the sensitivity of the system. As seen in Figure 20, the system's performance varies at different sampling rates but the overall performance is very constant (86-88%). Note that the system doesn't converge when we slower down the sampling rate to 50 *Hz*.

7.4 User study

We asked the participants to answer our questionnaire on their experience in controlling their tongue following different gestures suggested in *TYTH*. Primarily, our questions are about their comfort of wearing our *TYTH* prototype and performing the tongue gestures as well as the feasibility of *TYTH* accepted by our society in the future. Overall, based on the scale of 1 to 5 corresponding to "Strongly Uncomfortable" and "Strongly Comfortable", respectively, the result shows that the users strongly support us on the *TYTH*'s prototype and idea. We also found that the current form factor is not well-accepted by users. Most of users can only wear the devices and perform tongue typing from around 15-30 minutes. We are improving the form factor to make it more comfortable with users. From the collected data, the average typing performed as illustrated in Figure 22(5). All users agreed that the outer gestures are easier to perform compared with inner ones. Note that the above results are the average accuracy from our collected data, we plan to analyze how the accuracy falls when a user continuously uses the system.
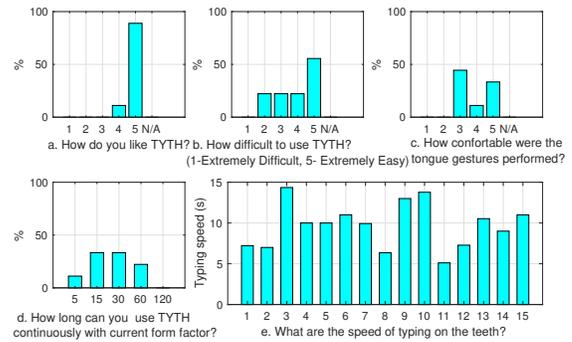
Figure 22: User study results.

As mentioned, at the end of the study, we asked user to type to all the 10 location as fast as they could, the results is obtained as in the bottom right figure. Most of user can perform their tongue typing with less than 1 touch area/second. This result is the target of our development for upcoming version of *TYTH*.

8 RELATED WORKS

There is a high risk of bacterial infection when people have to directly control external devices in public using common touch-based interfaces such as touchscreen, mouse, joystick, keyboard, etc. [67]. Assistive technologies using hands-free gestures for human-computer interaction (HCI) [33] have been enabling individuals to implicitly communicate to such the contaminated devices using bioelectrical signals generated by human organs. Prior work has also studied voice input as an accessible interaction to people with motor impairments [16]. Recently, researchers have explored the feasibility of using the human brain [10, 32, 66], eyes [18], chin [27, 28], arms [59, 63], neck and shoulders [7, 39, 79], and head orientation [8, 43, 79], as alternative interaction channels for various intentions through EEG, EOG, ECoG, GSR, and EMG signals. These approaches require user to wear a large head mounted devices, eyes glasses, chin's joiystick, or neck and shoulders' devices. These devices are often big and visible to public, which prevents these systems from being used in practice.

Similar to *TYTH*, others have studied the use of teeth as input, registering jaw or tooth vibrations [57, 68] or bone conductive sounds [2] as interactions. Prior studies have also explored tongue or oral computer interfaces, utilizing optical sensing [35, 62] or, as we implemented, magnetoresistive sensing [50, 55, 65, 72, 78] to infer tongue location. Most of these organ-operated interfaces require cumbersome devices (e.g. a head-mounted device for brain signal capture as in Tongue-n-Cheek [34]) or visible control (e.g., eye and arm movements) that make them awkward to use. Additionally, wearing the equipment and/or performing the interactions often gets in the way of normal activities.

HCI devices providing invisible interaction with sophisticated control tasks and reliable performance have become a new hi-tech trend. Specifically, the tongue, teeth, and their combination have been considered as suitable candidates for these requirements [21, 29, 62]. Previous studies have considered in-ear wearables as a solution for facial movement recognition [1] and biosignal colletion

[48]. Prior work including Tongue Drive [21, 29], TongueWise [5], and Sahni et. al. [60] suggests attaching a magnetic tracer the size of a grain of rice to the tongue. Alternatively, a tongue computer interface could utilize infrared optical sensors embedded within an orthodontic dental retainer [62]. Research on inductive tongue computer interfaces proposes invisible, in-mouth interactions to reduce negative social stigma [4, 36, 69, 70]. A magnetic implant [22], piezo-ceramic material [51], and piezoelectic film sensors [38] are other candidates to adapt the tongue for computer input. Such solutions of extracting rich tongue gestures are intrusive, requiring that sensors be embedded within the mouth both to determine the relative position between the tongue and the teeth and to sense tongue motion.

Due to the uncomfortable feeling of in-mouth sensors for long term use, non-obtrusive approaches leveraging various bioelectrical signals have been proposed for tongue movement detection. Common inputs for representing muscle activities of the tongue include EEG [11], EMG [64], glossokinetic potential (GKP) [45, 46], and pressure [9]. The TongueSee [80] system recognizes a set of 6 different movements through the suprahyoid muscles correlated to tongue gestures. Tongue-in-Cheek [13] uses 10 GHz wireless signal to detect the four tongue directions (left, right, up, and down) and two modes (tap and hold). In our opinion, the sensor size and and its placement may prevent the system from socially acceptable for daily uses. Other systems applied the concept of surface electromyography (sEMG) which requires the sensors be mounted on the facial skin close to the tongue [64]. These systems, however, input only a single type of signal, which provides limited information to extract spatial, temporal, and spectral features.

9 DISCUSSION

Form factor. In the current prototype, the form factor is still very large due to our limited expertise in building the wearable and the lack of miniaturization skills. However, since all electrical components used to create every single element of *TYTH* are off-the-shelf basic electrical components, it can be made significantly smaller so that *TYTH* can be hidden behind the user's ears. The connection between the two ears can be a thin electrically shielded cable to improve the aesthetic appearance. We also wish to improve the hardware design to reduce the power consumption, increase the accuracy by implementing our optimized hardware design. The optimized design would reduce significantly the power consumption by omitting unnecessary components and obtains higher accuracy by capturing more details featured.

Improving the localization granularity. With the current tongue tapping resolution, *TYTH* can merely locate tapping areas representing about 10 different key on a keypad. As a future work, we would like to improve localization granularity by improving the spatial resolution and sensitivity of the EEG, EMG, and SKG sensors. One possible direction to accomplish that is to use a 2-D array of tiny electrodes as opposed to the single electrode that we currently using so as *TYTH* can support a larger collection of key.

TYTH's ease of use. Just like with any new interaction method, users will need to take sometime to familiar themselves with this form of interaction. We acknowledge that this form of interface is harder to user than typing since the dexterity of hand is arguably

higher than that of tongue. However, we believe that users can be trained to use this form of interface when the benefit it brings overcome the learning hardship. In addition, one of our participant took around 30 minutes to learn how to perform inner gesture as his tongue is quite large. We will also consider this scenario in our future design.

Impact of talking and eating behaviors. Talking, eating, chewing and other tongue-related movements might unexpectedly trigger the system. We wish to develop a more sophisticated classification algorithm allowing the system to differentiate between *TYTH*'s typing events and these noises. In addition, we would like to reduce the level of the user's consciousness needed to use *TYTH*. This would also allow the user to use *TYTH* while they are talking/eating.

Running *TYTH*'s algorithms on mobile environment. Deploying the system on mobile environment is our next logical step. We have successfully implement our solution on Android is possible by writing our python code on Android Scripting Environment (ASE/SL4A) project for real-time data streaming from the wearable device to Samsung Galaxy S5. We found that most of processing algorithm including band-pass, low-pass, notch-filter, STFT, SVM, GMM, mapping prediction output to keyPress can be implemented on mobile environment. We haven't successfully implemented wavelet transform on mobile environment yet. We wish to finish this task in the near future.

10 CONCLUSION

In this work, we envision a future hands-free computer interface in which a user can privately interact with machines through a wearable device that he or she can comfortably wear and use in everyday life. To that end, we take the first step towards building a tongue-on-teeth typing system (called *TYTH*) that can be worn from behind user's ears to capture the relative location and interaction between user's tongue and teeth. From the fundamental understanding of the neurological and anatomical structures of the tongue, we design *TYTH* so that the obtrusiveness and social awkwardness caused by the wearable is minimized while maximizing its accuracy and sensing sensitivity. We modeled the relationship between tongue movement and the signals recorded, from which a tongue localization technique and tongue-teeth tapping detection technique are derived. An evaluation with a 15 subjects using our custom-built prototype shows that *TYTH* can be used as a from of hands-free human computer interaction with 88.61% detection rate and promising adoption rate by users. *TYTH* is a promising interface for assisting Parkinson, ALS, mutism, quadriplegic patients who lost control of their limbs' muscle and verbal communication. It could also be adopted for other applications serving as, for example, a private text entry method in public settings, authentication and input method for VR system, among others.

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 (UIST '17)*. 679–689.

[2] 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 (MobileHCI '16)*. 158–169.

[3] R. G. Baraniuk. 2007. Compressive Sensing [Lecture Notes]. *IEEE Signal Processing Magazine* 24, 4 (July 2007), 118–121.

[4] H. A. Caltenco, E. R. Lontis, S. A. Boudreau, B. Bentsen, J. Struijk, and L. N. S. Andreasen Struijk. 2012. Tip of the Tongue Selectivity and Motor Learning in the Palatal Area. *IEEE Transactions on Biomedical Engineering* 59, 1 (2012), 174–182.

[5] H. A. Caltenco, L. N. S. A. Struijk, and B. Breidegard. 2010. TongueWise: Tongue-computer interface software for people with tetraplegia. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. 4534–4537.

[6] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. 2011. Robust Principal Component Analysis? *J. ACM* 58, 3, Article 11 (June 2011), 37 pages.

[7] Gwo-Ching Chang, Wen-Juh Kang, Jer-Junn Luh, Cheng-Kung Cheng, Jin-Shin Lai, Jia-Jin J. Chen, and Te-Son Kuo. 1996. Real-time implementation of electromyogram pattern recognition as a control command of man-machine interface. *Medical Engineering & Physics* 18, 7 (1996), 529 – 537.

[8] Yu-Luen Chen, Te-Son Kuo, W. H. Chang, and Jin-Shin Lai. 2000. A novel position sensors-controlled computer mouse for the disabled. In *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No.00CH37143)*, Vol. 3. 2263–2266.

[9] 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 (AH '14)*. 12:1–12:4.

[10] F. Galan, M. Nuttin, E. Lew, P.W. Ferrez, G. Vanacker, J. Philips, and J. del R. Millan. 2008. A brain-actuated wheelchair: Asynchronous and non-invasive Brain-computer interfaces for continuous control of robots. *Clinical Neurophysiology* 119, 9 (2008), 2159–2169.

[11] L. Y. Ganushchak, I. K. Christoffels, and N. O. Schiller. 2011. The Use of Electroencephalography in Language Production Research: A Review. *Frontiers in Psychology* 2 (2011), 208.

[12] "Charles D. Ghilani". 2010. "Appendix C: Nonlinear Equations and Taylor's Theorem". In *"Adjustment Computations: Spatial Data Analysis"*, "Jan Fagerberg, David C. Mowery, and Richard R. Nelson" (Eds.).

[13] 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 (CHI '15)*. ACM, New York, NY, USA, 255–258. https://doi.org/10.1145/2702123.2702591

[14] Tobias Grosse-Puppendahl, Christian Holz, Gabe Cohn, Raphael Wimmer, Oskar Bechtold, Steve Hodges, Matthew S. Reynolds, and Joshua R. Smith. 2017. Finding Common Ground: A Survey of Capacitive Sensing in Human-Computer Interaction. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. New York, NY, USA, 3293–3315.

[15] B. A. Hanson and T. H. Applebaum. 1990. Robust speaker-independent word recognition using static, dynamic and acceleration features: experiments with Lombard and noisy speech. In *International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2. 857–860.

[16] Susumu Harada, James A. Landay, Jonathan Malkin, Xiao Li, and Jeff A. Bilmes. 2008. The Vocal Joystick: Evaluation of voice-based cursor control techniques for assistive technology. *Disability and Rehabilitation: Assistive Technology* 3, 1-2 (2008), 22–34.

[17] Olaf Hauk, Ingrid Johnsrude, and Friedemann Pulvermüjller. 2004. Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron* 41 (Jan. 2004), 301–307.

[18] Jeong Heo, Heenam Yoon, and Kwang Suk Park. 2017. A Novel Wearable Forehead EOG Measurement System for Human Computer Interfaces. *Sensors* 17, 7 (2017), 1485.

[19] Computer Hope. 2017. Input Device. https://goo.gl/4tSXEU. (2017). [Online; accessed Dec 08, 2017].

[20] Computer Hope. 2017. When was the first keyboard invented? https://goo.gl/cu9ia1. (2017). [Online; accessed Dec 08, 2017].

[21] X. Huo and M. Ghovanloo. 2012. Tongue Drive: a wireless tongue-operated means for people with severe disabilities to communicate their intentions. *IEEE Communications Magazine* 50, 10 (2012), 128–135.

[22] X. Huo, J. Wang, and M. Ghovanloo. 2008. A Magneto-Inductive Sensor Based Wireless Tongue-Computer Interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 16, 5 (2008), 497–504.

[23] Manh Huynh, Phuc Nguyen, Marco Gruteser, and Tam Vu. 2015. POSTER: Mobile Device Identification by Leveraging Built-in Capacitive Signature. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15)*. ACM, New York, NY, USA, 1635–1637. https://doi.org/10.1145/2810103.2810118

[24] Texas Instrument. 2017. TI-MSP430FR5969 chip. https://goo.gl/1dBTyD. (2017). [Online; accessed Nov 08, 2017].

[25] Texas Instruments. 2017. ADS1299. https://goo.gl/8cdSsi. (2017). [Online; accessed Nov 08, 2017].

[26] Texas Instruments. 2017. CC2640. https://goo.gl/whf92S. (2017). [Online; accessed Nov 08, 2017].

[27] Reinhilde Jacobs, Elke Hendrickx, Isabel Van Mele, Kirstie Edwards, Mieke Verheust, Arthur Spaepen, and Daniel Van Steenberghe. 1997. Control of a trackball by the chin for communication applications, with and without neck movements. *Archives of Oral Biology* 42, 3 (1997), 213 – 218.

[28] Chin Joystick. 2003. Chin Joystick. http://www.pulsar.org/archive/eyal/fal97/fal97_headset.html. (2003). [Online; accessed Oct 18, 2017].

[29] G. Krishnamurthy and M. Ghovanloo. 2006. Tongue Drive: a tongue operated magnetic sensor based wireless assistive technology for people with severe disabilities. In *2006 IEEE Int'l Symp. on Circuits and Systems*.

[30] Ja Young Lee, Madeleine C. Gibson, and John D. Lee. 2016. Error Recovery in Multitasking While Driving. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5104–5113.

[31] Lenovo. 2017. Cyton Biosensing Board. https://goo.gl/Di9Gce. (2017). [Online; accessed Nov 08, 2017].

[32] Eric C Leuthardt, Gerwin Schalk, Jonathan R Wolpaw, Jeffrey G Ojemann, and Daniel W Moran. 2004. A brain computer interface using electrocorticographic signals in humans. *Journal of Neural Engineering* 1, 2 (2004), 63 – 71.

[33] Tianxing Li, Qiang Liu, and Xia Zhou. 2017. Ultra-Low Power Gaze Tracking for Virtual Reality. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems CD-ROM (SenSys '17)*.

[34] Zheng Li, Ryan Robucci, Nilanjan Banerjee, and Chintan Patel. 2015. Tongue-n-Cheek: Non-contact Tongue Gesture Recognition. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks (IPSN '15)*. 95–105.

[35] Li Liu, Shuo Niu, Jingjing Ren, and Jingyuan Zhang. 2012. Tongible: A Non-contact Tongue-based Interaction Technique. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '12)*. 233–234.

[36] M. E. Lund, H. V. Christiensen, H. A. Caltenco, E. R. Lontis, B. Bentsen, and L. N. S. Andreasen Struijk. 2010. Inductive tongue control of powered wheelchairs. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. 3361–3364.

[37] S. G. Mallat and Zhifeng Zhang. 1993. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 41, 12 (Dec. 1993), 3397–3415. https://doi.org/10.1109/78.258082

[38] L. Monczak, D. Shapiro, A. Borisenko, O. Draghici, and M. Bolic. 2011. Characterization of piezoelectric film sensors for tongue-computer interface. In *2011 IEEE International Symposium on Medical Measurements and Applications*. 492–497.

[39] Inhyuk Moon, Kyunghoon Kim, Jeicheong Ryu, and Museong Mun. 2003. Face direction-based human-computer interface using image observation and EMG signal for the disabled. In *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*, Vol. 1. 1515–1520.

[40] Moonson. 2017. Moonson Power Management Tool. https://goo.gl/f9J3Dz. (2017). [Online; accessed Nov 08, 2017].

[41] Valerie Morash, Ou Bai, Stephen Furlani, Peter Lin, and Mark Hallett. 2008. Classifying EEG signals preceding right hand, left hand, tongue, and right foot movements and motor imageries. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology* 119, 11 (Nov. 2008), 2570–2578.

[42] G. Moruzzi and H. W. Magoun. 1949. Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology* 1, 1 (Jan. 1949), 455–473.

[43] Denis Anson M.S., O.T.R./L., Gretchen Lawler O.T.S., Alexis Kissinger O.T.S., Megan Timko O.T.S., Jason Tuminski O.T.S., and Brian Drew O.T.S. 2002. Efficacy of Three Head-Pointing Devices for a Mouse Emulation Task. *Assistive Technology* 14, 2 (2002), 140–150.

[44] Lindasalwa Muda, Mumtaj Begam, and I. Elamvazuthi. 2010. Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. *CoRR* abs/1003.4083 (2010).

[45] Yunjun Nam, Bonkon Koo, Andrzej Cichocki, and Seungjin Choi. 2016. Glossokinetic Potentials for a Tongue-Machine Interface: How Can We Trace Tongue Movements with Electrodes? *IEEE Systems, Man, and Cybernetics Magazine* 2, 1 (2016), 6–13.

[46] Y. Nam, Q. Zhao, A. Cichocki, and S. Choi. 2012. Tongue-Rudder: A Glossokinetic-Potential-Based Tongue Machine Interface. *IEEE Transactions on Biomedical Engineering* 59, 1 (2012), 290–299.

[47] Tan Dat Trinh Ngoc Nam Bui, Jin Young Kim. 2014. A non-linear GMM KL and GUMI kernel for SVM using GMM-UBM supervector in home acoustic event classification. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* 97, 8 (2014), 1791–1794.

[48] Anh Nguyen, Raghda Alqurashi, Zohreh Raghebi, Farnoush Banaei-kashani, Ann C. Halbower, and Tam Vu. 2016. A Lightweight and Inexpensive In-ear Sensing System For Automatic Whole-night Sleep Stage Monitoring. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM (SenSys '16)*. 230–244.

[49] Phuc Nguyen, Ufuk Muncuk, Ashwin Ashok, Kaushik R. Chowdhury, Marco Gruteser, and Tam Vu. 2016. Battery-Free Identification Token for Touch Sensing Devices *(SenSys '16)*. ACM, 109–122. https://doi.org/10.1145/2994551.2994566

[50] Takuya NIIKAWA and Ryosuke KAWACHI. 2006. Human-Computer Interface Using Mandibular and Tongue Movement. *Transactions of Japanese Society for Medical and Biological Engineering* 44, 1 (2006), 94–100.

[51] Wolfgang Nutt, Christian Arlanch, Silvio Nigg, and Gerhard Staufert. 1998. Tongue-mouse for quadriplegics. *Journal of Micromechanics and Microengineering* 8, 2 (1998), 155.

[52] Smooth on. 2017. Smooth On Ecoflex 00-10. https://goo.gl/1dBTyD. (2017). [Online; accessed Nov 08, 2017].

[53] OpenBCI. 2017. Cyton Biosensing Board. http://openbci.com/. (2017). [Online; accessed Nov 08, 2017].

[54] Oracle. 2017. Oracle keyPress. https://goo.gl/4aTQFc. (2017). [Online; accessed Dec 08, 2017].

[55] H. Park, B. Gosselin, M. Kiani, H. M. Lee, J. Kim, X. Huo, and M. Ghovanloo. 2012. A wireless magnetoresistive sensing system for an intra-oral tongue-computer interface. In *2012 IEEE International Solid-State Circuits Conference*. 124–126.

[56] Gert Pfurtscheller, Doris Flotzinger, and Christa Neuper. 1994. Differentiation between finger, toe and tongue movement in man based on 40 Hz EEG. *Electroencephalography and Clinical Neurophysiology* 6 (June 1994), 456–460. https://doi.org/10.1016/0013-4694(94)90137-6

[57] A. Prochazka. 2005. Method and apparatus for controlling a device or process with vibrations generated by tooth clicks. (Nov. 1 2005). https://www.google.com/patents/US6961623 US Patent 6,961,623.

[58] Douglas A. Reynolds. 2015. Universal Background Models. In *Encyclopedia of Biometrics, Second Edition*. 1547–1550.

[59] R. Rosenberg. 1998. The biofeedback pointer: EMG control of a two dimensional pointer. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*. 162–163.

[60] 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 *Proceedings of the 2014 ACM International Symposium on Wearable Computers (ISWC '14)*. 47–54.

[61] Ira Sanders and Liancai Mu. 2013. A Three-Dimensional Atlas of Human Tongue Muscles. *The Anatomical Record* 296, 7 (2013), 1102–1114.

[62] 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 (UIST '09)*. 177–180.

[63] T Scott Saponas, Desney S. Tan, Dan Morris, and Ravin Balakrishnan. 2008. Demonstrating the Feasibility of Using Forearm Electromyography for Muscle-computer Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. 515–524.

[64] M. Sasaki, K. Onishi, T. Arakawa, A. Nakayama, D. Stefanov, and M. Yamaguchi. 2013. Real-time estimation of tongue movement based on suprahyoid muscle activity. In *2013 EMBC*. 4605–4608.

[65] M. Sasaki, K. Onishi, A. Nakayama, K. Kamata, D. Stefanov, and M. Yamaguchi. 2014. Tongue motor training support system. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 3582–3585.

[66] G. Schalk and E. C. Leuthardt. 2011. Brain Computer Interfaces Using Electrocorticographic Signals. *IEEE Reviews in Biomedical Engineering* 4 (2011), 140–154.

[67] Maureen Schultz, Janet Gill, Sabiha Zubairi, Ruth Huber, and Fred Gordin. 2003. Bacterial Contamination of Computer Keyboards in a Teaching Hospital. *Infection Control & Hospital Epidemiology* 24, 4 (2003), 302âĂŞ303.

[68] 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.

[69] L. N. S. Andreasen Struijk. 2006. An Inductive Tongue Computer Interface for Control of Computers and Assistive Devices. *IEEE Transactions on Biomedical Engineering* 53, 12 (2006), 2594–2597.

[70] Lotte N. S. Andreasen Struijk, Eugen R. Lontis, Michael Gaihede, Hector A. Caltenco, Morten Enemark Lund, Henrik Schioeler, and Bo Bentsen. 2017. Development and functional demonstration of a wireless intraoral inductive tongue computer interface for severely disabled persons. *Disability and Rehabilitation: Assistive Technology* 12, 6 (2017), 631–640.

[71] Hironori Takemoto. 2001. Morphological analyses of the human tongue musculature for three-dimensional modeling. *Journal of Speech, Language, and Hearing Research* 44, 1 (2001), 95–107.

[72] Hui Tang and D. J. Beebe. 2006. An oral tactile interface for blind navigation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 14, 1 (2006), 116–123.

[73] Carl Taswell and Kevin C. McGill. 1994. Algorithm 735: Wavelet Transform Algorithms for Finite-duration Discrete-time Signals. *ACM Trans. Math. Softw.* (1994), 398–412.

[74] Hoang Truong, Phuc Nguyen, Nam Bui, Anh Nguyen, and Tam Vu. 2017. Demo: Low-power Capacitive Sensing Wristband for Hand Gesture Recognition. In *Proceedings of the 9th ACM Workshop on Wireless of the Students, by the Students, and for the Students (S3 '17)*. ACM, New York, NY, USA, 21–21. https://doi.org/10.1145/3131348.3131358

[75] Hoang Truong, Phuc Nguyen, Anh Nguyen, Nam Bui, and Tam Vu. 2017. Capacitive Sensing 3D-printed Wristband for Enriched Hand Gesture Recognition. In *Proceedings of the 2017 Workshop on Wearable Systems and Applications (WearSys '17)*. ACM, New York, NY, USA, 11–15.

[76] Hoang Truong, Phuc Nguyen, Viet Nguyen, Mohamed Ibrahim, Richard Howard, Marco Gruteser, and Tam Vu. 2017. Through-body Capacitive Touch Communication. In *Proceedings of the 9th ACM Workshop on Wireless of the Students, by the Students, and for the Students (S3 '17)*. ACM, New York, NY, USA, 7–9. https://doi.org/10.1145/3131348.3131351

[77] Sampsa Vanhatalo, Juha Voipio, A Dewaraja, MD Holmes, and JW Miller. 2003. Topography and elimination of slow EEG responses related to tongue movements. *NeuroImage* 20, 2 (2003), 1419–1423.

[78] Yikun K. Wang, Martyn P. Nash, Andrew J. Pullan, Jules A. Kieser, and Oliver Röhrle. 2013. Model-based identification of motion sensor placement for tracking retraction and elongation of the tongue. *Biomechanics and Modeling in Mechanobiology* 12, 2 (2013), 383–399.

[79] M. R. Williams and R. F. Kirsch. 2008. Evaluation of Head Orientation and Neck Muscle EMG Signals as Command Inputs to a Human-Computer Interface for Individuals With High Tetraplegia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 16, 5 (2008), 485–496.

[80] Qiao Zhang, Shyamnath Gollakota, Ben Taskar, and Raj P.N. Rao. 2014. Nonintrusive Tongue Machine Interface. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, 2555–2558.