Tongue-able Interfaces: Prototyping and Evaluating Camera Based Tongue Gesture Input System

Shuo Niu[a], Li Liu[b], D. Scott McCrickard[a]

[a] *Department of Computer Science,Virginia Tech,Blacksburg, USA*
[b] *Department of Computer Science, California State University, Los Angeles, USA*

Abstract

Tongue-computer interaction techniques create a new pathway between human mind and computer, with particular utility for people with upper limb impairment. The high dexterity and resilience of the tongue make it a good candidate for interacting with computers. This paper introduces a new interaction technique, camera-based tongue computer interface (CBTCI), which employs tongue without any direct physical contact required. Through a two-phase study, the CBTCI was evaluated and its interaction problems were identified and discussed. In the first phase, the performance of the CBTCI prototype was evaluated through two user tests. The participants behaviors were observed throughout the first phase and analyzed to scaffold the study of the design problems in gesture based tongue computer interaction. The Phase 2 study investigated the usability problems of CBTCI which were reflected through the user behavior and participants feedback; specifically the exploration of referential techniques to make users aware of their tongue position and adjust their gesture. Pros and cons of the referential strategies are discussed to foster future assistive tongue-computer interface design.

Keywords: Tongue interface, gesture, hands-free, camera, recognition, self-awareness

1. Introduction

A great many people worldwide have limitations in the use of their hands and arms. For example, it is estimated that 185,000 people in United States undergo an amputation of upper or lower limb each year (Ziegler-Graham et al., 2008), and the number of people living with upper body limitations

was 19,900,000 in 2010 (Brault et al., 2012). To assist people with upper-limb impairment access the computer, tongue-computer interfaces (TCIs) have been proposed to open a new avenue to build hand-free technologies (e.g. Kim et al., 2012; Miyauchi et al., 2013; Slyper et al., 2011; Struijk et al., 2017; Park and Ghovanloo, 2016; Mimche et al., 2016). Two ways have emerged to develop TCIs: plug sensors into the oral cavity, or the use of computer vision to detect tongue motion. Vision-based TCIs can avoid several problems of oral plug-in hardware. First, for any intrusive mouth device, hardware must be sterilized to meet hygienic requirement before use. Since diet and speech can be hindered by these kinds of in-mouth devices, they have to be taken out of the mouth between uses. Second, contact devices can easily cause intra oral trauma since mouth is sensitive to electronic and physical stimulation. Electromagnetic signal and mechanical force can hurt a users mouth if their strength is over a threshold. Third, it can be inconvenient for people with physical disabilities to put assistive hardware into their mouths.

In contrast to intrusive TCIs, tongue gesture input systems provide higher degrees of motional freedom (DOF). However, there are many questions regarding potential usability of such systems. Many of these questions stem from the lack of tangible feedback found in the more intrusive physical devices. Consider as a parallel the difference in feedback between a physical keyboard and a virtual one, in which careful consideration has to be given to how visual and haptic feedback on a virtual keyboard can replace the feeling of physical keys. Similarly, an input method that relies on tongue positioning requires feedback to the user to convey when proper input has been recognized contrasted with unwanted or non-recognizable input. Even if the tongue is in the recognition zone correctly, holding the correct gesture for the appropriate period of time is also challenging because slight tongue motion may influence the recognition rate and reduce system reliability. Further, because people are used to operating virtual objects in a similar manner as in the physical world, the human sensorimotor system affects tongue gestures significantly. Therefore the high flexibility of physically detached interaction increases the fluctuation and uncertainty of tongue recognition system. For example, pressing a virtual button on an interface by simulating the pressing of a physical button through moving the tongue with greater force or torque may not lead to faster response from the system; instead it may cause the system lose its recognition target, because the tongue may be out of the recognition zone or in the wrong position.

To maintain the high DOF and to reduce system fluctuation, self-awareness

is considered a common strategy in gesture-based research (e.g. Lin et al., 2010; Tu et al., 2005). Presented with a captured video window of oneself, a user can monitor motions by referring to the camera window. However, in this type of gesture-based interaction, a user must engage in multitasking, switching between performing the task and monitoring the camera, potentially resulting in performance degradation (McCrickard et al., 2003). We hypothesize that tracking feedbacks of the input and monitoring the motions simultaneously requires more mental effort, resulting in more errors.

To explore the potential utility of non-contact tongue computer interfaces, this paper introduces a novel non-contact camera-based tongue computer interface (CBTCI) and presents a two-phase study of its usability. Two referential awareness techniques are captured in the CBTCI: one that uses the captured video technique and another that uses an interpreted technique. A two-phase usability test compares performance and preferences using these techniques. In Phase 1, we introduce the design and implementation of CBTCI and (Liu et al., 2012, 2017) to explore the performance of tongue interface and the constraints of tongue operations in recognition-related tasks. By using computer vision users with upper body motor impairment can use their tongue as input without acquiring extra hardware. Phase 2 investigated how the two referential strategies (Niu et al., 2014) affect self-awareness of inputs during user tongue gesture adjustment.

The development and evaluation described in this paper contributes to the assistive technology community in three ways: (1) a novel and non-intrusive hand-free assistive technology is proposed and implemented; (2) user behaviors related to non-contact tongue interface are revealed and analyzed; (3) potential methods to improve the usability and their relevant upsides and downsides are discussed to foster future TCI design.

2. Related work

Dexterity impairment hinders computer access because ordinary input devices such as keyboard, mouse, and joystick are generally operated by hands. Alternate body parts can be used to operate a computer. Gaze tracking systems control computer by tracking the movement of the eyes and detecting gestures of the eye such as blinking (Rautaray and Agrawal, 2015). This kind of interaction methods may cause a headache if it is used continuously for a long time. Head control systems detect the position of the human face or track the orientation and position of the head to control computers. Neck

pain is the biggest problem for head-computer interaction. Speech control methods (Polacek et al., 2011) use speech recognition technology to transform human speech into computer instruction. The performance of these methods always becomes unstable because of ambient interference. Brain control is a promising method but it is expensive to be equipped (Mugler et al., 2010).

Recent findings suggest that tongue has superiority to build hand-free systems, since tongue is a dexterous muscular hydro-stat and usually remain unaffected even in severe injuries and neuromuscular diseases (Huo and Ghovanloo, 2010). In addition to acting as a substitution of hands when hands are temporally occupied or functionally impaired, tongue is also considered as a good option to build novel interaction experience. Engineers at Valve developed a tongue controller and demoed it with playing a popular 3D game. Considering that tongue is equal or more dexterous than hands and not get fatigued easily (Lau and OLeary, 1993), adopting tongue interface to gaming will enrich the way of entertainment for people either with or without physical impairment. Nam and Disalvo presented a technology which associate tongue interface and music to boost the emotional experience during kissing (Nam and DiSalvo, 2010). This work demonstrates that during some special time when looking at screen is not convenient, tongue interface is also useful to augment the affection of people.

In recent research of tongue and oral control technologies, various tongue interfaces are designed in a way by putting sensors in the users oral cavity. Chandra et al. proposed a method to use curved dipole antennas plugged on both sides of the teeth (Chandra and Johansson, 2011). The variations in the link loss of open and closed mouth were compared to recognize the mouth states. The electrical stimulation from the antennas has to be strictly constrained to protect the users oral cavity. Slyper and her colleagues presented a tongue joystick for maneuvering within a dialogue tree (Slyper et al., 2011). The joystick is used to create simple dialogue for costume performer. However, the joystick has to be held in mouth, it hinders the normal speaking of the user. Saponas et al. showed a technology for tongue gesture recognition using infrared optical sensors embedded within a dental retainer (Saponas et al., 2009). But manually customizing the in-mouth retainer for each user is complicated and it increases the financial burden of the user. Other tongue-based devices like the sip/puff switch and bite sensor also use mouth to provide interfaces for human-computer interaction. All the above-mentioned systems require an in-mouth device and therefore suffer the hygiene and in-

convenience problem.

Considering the risk of hygiene problem and ingestion by accident, recent research studied non-invasive approaches to detect tongue motion. Kim et al. introduced Tongue Drive System, a wireless assistive technology using permanent magnetic tracer glued in the middle of the tongue (Kim et al., 2012). The wireless transceiver is mounted on a headset and traces the variation of magnetic field when a user moves the tongue. However, accidental ingestion of the tracer is a potential threat for elderly and children when using this technology since the size of the trace is very small. Cheng et al. developed a pressure-based tongue interface by attaching a pressure detector to the users cheek (Cheng et al., 2014). Since the device need to be carefully attached and calibrated to get good recognition rate, it is not convenient for people with dexterity impairment to put on the device and do the calibration without an assist. Nam et al. utilized brain-wave based technology to detect four discreet states of the tongue (Nam and DiSalvo, 2010). The noise signal of electroencephalography and the requirement of concentrated attention are major concerns which might compromise the stability of the tongue interface. Miyauchi et al. introduced a system detecting tongue motion with Microsoft Kinect (Miyauchi et al., 2013). This technology can detect tongue motion when the tongue reached the depth baseline, but the face and tongue need to be placed at a relatively fixed position to the display and Kinect, which reduces the flexibility and comfort when operating the TCI.

Drawing from two major areas of research - tongue-based interfaces and vision-based gesture recognition - our project combines and extends prior work toward providing a software approach that provides a clean, safe, and relatively inexpensive and non-intrusive tongue interaction interfaces.

3. Creating CBTCIs

3.1. Design Rationales and Motivation

The acceptance of an assistive technology (AT) is determined by a combination of multi-dimensional factors. People usually depend on these perspectives to conclude their decision (Hurst and Tobias, 2011), 1) easiness of procuring the AT system, 2) user involvement in system selection, 3) system performance and 4) adaptiveness to a user's need. To develop an assistive system with high acceptance, we identify the following four factors as the basis of prototyping the interface.

3.1.1. Procuring a New Interface

Most AT systems with tongue interface use advanced electronic devices. Even though these devices can improve the reliability of the assistive technology, they are not wildly applied in daily life. So, it demands users to spend more time on learning. On the contrary, a new CBTCI based on a popular technology does not only bring familiarity to novice users during their learning process but also benefits those who are serving the community. It also promotes the acceptance of an AT system if it is built on popular hardware, in another way low-cost hardware implementation. Today, most computers are equipped with a webcam. As the online face-to-face communication is emerging it is convenience to use a webcam for an AT system. Meanwhile, our effort of computer vision makes most popular webcam fully meet the requirement of using a video stream of the tongue as an input signal during the interaction.

3.1.2. Cognitive Load

Cognitive load of using an AT is an essential factor in evaluating the usability. In the new CBTCI, we adopt the joystick manipulation to smooth the learning curve of a new AT and to provide a natural interaction paradigm. A joystick is an input device consisting of a stick that pivots on a base and reports the direction of the device it is controlling. It is widely utilized as a controller in many assistive systems, such as (Hinckley et al., 2014). The tongue has in common with a joystick as its protruding movement is similar with the manipulation of the joystick. Inspired by the design and function of the joystick, the four-directional gestures (Up, Down, Left and Right) of the tongue form the gesture set of input. However, the four-directional gestures are not enough to interact with a computer. There should be a gesture to be used as the confirm button of a joystick. So mouth close (Close) and the status of mouth open with the tongue placed in the oral cavity (Open) are added to the gesture set. Figure.1 illustrates the six tongue gestures. A user's tongue is in one of the six gestures when using the new interface.

3.1.3. Motor Fatigue

From a pilot study, we found that it causes motor fatigue if a user reaches out her or his tongue for a long time. So, the new interface should minimize both the time and the frequency that a tongue out of the mouth to avoid muscular fatigue. In the pilot study, we also learned that the motor fatigue

Figure 1: Six mouth and tongue gestures.

varies when a tongue moves towards different directions. Tongue action leading to higher motor fatigue, which usually takes more time should be cut in numbers of usage during the interaction. We also look into how long when a tongue is out mouth can trigger input events through computer vision.

3.1.4. Operability

Legacy pixel-based interaction paradigm between human and a computer display is designed for mouse-based pointing. Resulting from the fact that the scope of the tongue movement in an oral cavity is relatively small and the tongue is impossible to remain stable in one fixed position for a long time, it is hard to use the tongue interface in legacy interaction paradigm. In order to improve the operability of the tongue interaction method, a new interaction paradigm is desired to use tongue as an input device. The new interface adopts a non-pixel based interaction paradigm that associates internal structure and metadata of an interactive object with its graphical presentation to improve the operability of component selecting (Chang et al., 2011). In the new interface, we define each unit of cursor movement as an i*nteractive tile*, *tile* for short, which is a hybrid framework of pixels and metadata with the same event handler of user operation. Another effort to increase operability of using the tongue interface is to employ ambiguity input method for character entry. In (MacKenzie, 2009), an ambiguous input method is provided for n-key texting.

3.2. Implementation Method

The digital camera is becoming a standard accessory on most of laptops, tablets, mobile phones, wearables, and many other devices. Therefore,

camera-based interfaces can be implemented and disseminated without requiring extra hardware. CBTCI has another benefit in that the relative position of the user and device is flexible. The input video flow can be easily zoomed in or out to detect the tongue at different distances and positions. We believe that CBTCI is a promising assistive interface for people with upper-limb impairment to have a better access to digital technologies, worthy of closer investigation.

We also adopted a non-pixel based GUI to allow users to perform common computer operations with the CBTCI. PAX (Chang et al., 2011) is a hybrid framework combining pixels and accessibility APIs. It associates the representation of screen widgets with information from the operating system to boost the robustness and performance of the existing works. Inspired by PAX, we use a *tile* like graphics which is many pixels high and wide as the basic interactive component in the new interface because tongue-based interaction paradigm does not necessary to operated at single pixel level interaction as a mouse-based interaction does. Adjacent *tile*s can be selected with steps of tongue movement which benefits the directional operation of using a tongue. Pixels in one *tile* belong to one functional structure and handle the same tongue operation. In GUI, a *tile* may be a button, an icon or a word. To achieve this design, the system should analyze the meaning of all pixels and classify them into *tiles*. The relative positions of all *tiles* are tracked so that from one *tile* the user can access to the nearby up, down, left and right *tiles*.

Detecting the tongue protrusion is the core problem in classifying our six different tongue gestures (as visible in Figure.1). As discovered from initial observations, when instructed to gesture with the tongue, different people place the tongue in different positions. The size and color of the tongue also vary from person to person. All of these factors influence the accuracy of tongue gesture recognition. To overcome these limitations, we used a facial recognition algorithm, the adaptive boosting (AdaBoost) machine learning algorithm (adapted for other recognition tasks (Wang, 2012)), to solve the diversity problem of humans (Polikar, 2006). Adaboost extracts the common structures of the target components and background samples in a learning phase to build a series of weak classifiers, which are combined to form a strong classifier.

The tongue interface is implemented by a series of image processing procedures and pattern recognition. It classifies every tongue status into one of the gestures from the gesture set. The first step of tongue gesture detection

is face recognition. Using face recognition method the system locates the position of the user's face and zooms to the region of interest (ROI) in the original image. In the second step, the cavity of the mouth is enhanced by the Gray World Assumption algorithm (Van De Weijer et al., 2007) and Discrete Hartley Transform (Dalka et al., 2014). Between these two processes the blue channel of the original frame is set to zero to enhance the oral cavity. In the next step, as the region of the mouth is relatively the same proportion on the faces of different people, the oral cavity is clipped by a mask with the shape of the lower half ellipse whose long axis is the face image height, and short axis is half of the face image width. Next, the pattern recognition method is used to recognize the status of the tongue.

In the Phase 1 study, six strong classifiers generated by AdaBoost algorithm (Wang, 2012) classify the six tongue gestures. For one strong classifier of the tongue gesture, 1800 images of this tongue gesture from 6 people are used to form the training set. In the real-time recognition, the result of six classifiers is calculated, and the gesture whose corresponding classifier has the lowest error rate is chosen as the current tongue gesture. Using this method, the real-time processing cycle which starts from capturing the frame of the video and ends at outputting the recognition result, is approximate $80ms$. on the testing computer.

We notice using user's own tongue image as training samples can achieve better accuracy in tongue recognition. Therefore for Phase 2, a two-layer recognition method in the tongue recognition stage was adopted to increase accuracy. In the first layer, one classifier calculated the possibilities of all the six gestures with the captured images. The top three gestures with highest possibilities are reported to the second layer classifiers. In the second layer, the specific classifiers used to detect the top three gestures from the first layer were used to output the final decision (Figure.2). When a new real-time camera frame of the user is captured and to be processed, it first goes through the image processing pipeline and the region of oral area image is clipped. Then the two layer classifier processes the input image and categorizes it into one of the six tongue gestures. Just as a button press or joystick movement is passed on to an application when using other input devices, CBTCI passes on the recognized gesture to the application that is using it.

To evaluate the system, we designed usability tests focusing on accuracy, error rate, selection speed, and desired feedback. The next two sections present these experiments and results.

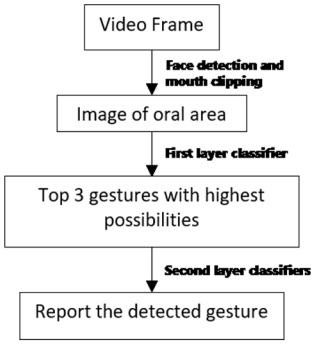

Figure 2: Image processing pipeline of the CBTCI

4. Phase 1: Pointing and Selecting

Two user tests were conducted on university campus (4 males and 6 females aging from 20 to 30) to evaluate the performance of the CBTCI system and investigate the usability related to the non-contact tongue computer interface in Phase 1 of the study. This section highlights our findings in Liu et al. 2012, 2017 and focuses on how the knowledge gleaned from the experiments leads to the design of Phase 2 study. We conducted two experiments to the following factors that characterize biomechanical properties of CBTCI:

- 1) The relationship between the speed of cursor movement and time of pointing task.
- 2) Motor fatigue of the four-directional gestures.
- 3) The error rate of the four-directional gestures.
- 4) Reaction time of the user when using tongue for pointing and texting.

4.1. Selecting Task

In the first experiment, an animal selection test was designed to examine the error rate and reaction time when using the CBTCI prototype (Slyper et al., 2011). This experiment aims to evaluate factor 2) and 3). The experiment includes a testing program and a questionnaire survey. In the test, four different animal avatars (chicken, horse, fish and dog) were displayed around an indicating box (Figure.3). The computer randomly spoke the name of one animal. A user sat in front of an RGB camera and protruded his/her tongue to the corresponding direction of the indicated animal. After the animal was

selected by the participant, there was a 4-second break before the computer speaks the name of the next animal. In one trial, every participant made the animal selection for 30 times. The test application treated close gesture as the default gesture and took the first detected directional gesture (either up, down, left or right) as the selection of the participant. All participants were asked to respond correctly to all the animal indications. If a participant produced any cognitive errors (e.g. reached out the tongue to the horse while the system said fish), he/she would be asked to start over again.

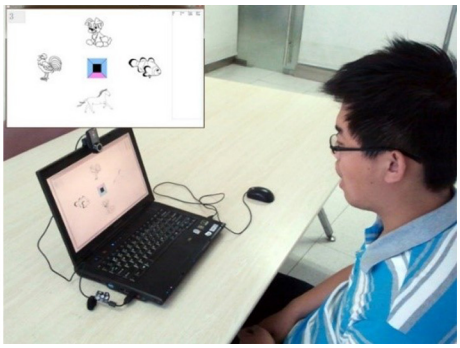

Figure 3: Animal selection test

4.2. Pointing and Texting Task

The second test was designed to evaluate factor 1) and 4) with a task of using the CBTCI prototype to move a cursor on the screen. 30 tiles aligned horizontally without overlap on the screen (Figure.4). The pointing task is to move the on-screen cursor to a target *tile*. Since the degree of freedom of using a tongue is limited comparing to using a hand or fingers, moving strategies are designed differently from legacy cursor movement. Both tongue's movement and eye-tongue coordination determine the travel speed of cursor. Therefore a pointing task is translated to the time of the cursor's trajectory across a screen plus the time of positioning the cursor onto the target. Under this hypothesis, the moving speed of the cursor should be set carefully. A fast movement will make it difficult to adjust a cursor's position and a slow movement will increase the time of cursor's trajectory. The basic element of interaction is *tile* in this interaction. The speed of the cursor is measured in ms. per *tile* stay (*mspt*), which means the time of the cursor stay on one *tile* before moving to the next. It should be noted that when the value of speed

in *mspt* is greater, the slower the cursor moves. We use constant cursor speed to evaluate user behavior.

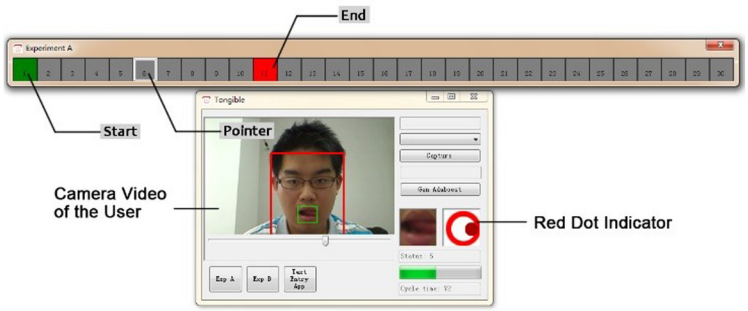

Figure 4: Cursor movement test

A cursor starting from the leftmost (green) moved left or right according to the detected left or right gesture. If the cursor stopped at the target (red), the participant would need to hold the open gesture for 2 seconds to finish the trial. Every participant moved the cursor from the start to the target for 9 times. Across the 9 trials, the time of holding a gesture to move the cursor by one tile increased from 50 milliseconds to 450 milliseconds with an equal increase of 50 milliseconds. Movement time and adjustment time were recorded by the system. Movement time is the time from the start of the trial to the time when the cursor is first time over the target. Adjustment time is the time from the first time over target to the finish of the trial.

Before the start of the two tasks, all participants did a 5-minute training to learn how to perform the test. Then they did a calibration session to find their most recognizable tongue positions. After the experiment, all participants were asked to finish a questionnaire to rate the easiness of tongue protrusion with a scale of 1 to 5 with 1 being very easy and 5 being very hard.

4.3. Results

In the animal selection test, the average error rate of the directional protrusion was 16.89 % (SD=0.08). The average respond time was 1512 milliseconds (SD=788). All subjects choose the Up position as the most difficult task. In the subjective rating, the mean of the fatigue of using Up position is the highest (3.2). Therefore, using the Up position should

Table 1: Error rate and rating of the four-directional gestures

	Up	Down	Left	Right
Error Rate	18.6%	18.4%	16.3%	14.5%
Mean Score	3.2	2.2	1.6	1.4

be eliminated in texting. In the error rate study, the up-down targeted protrusion has the highest error rate (18.6% for Up and 18.4% for Down). Table.1 shows the result. From the result of the experiment B and the assumptions in the previous section, we suggest that the character selection should be controlled by the gesture of Left and Right gestures and word selection should be controlled by the Up and the Down.

Errors in selecting are caused by two reasons. The first reason is system recognition error and mistaken tongue gesture. From the statistical analysis, the influence of gesture length on the value of error rate with this reason(e_l) can be ignored. So the error rate can be roughly expressed as the mean value of the error rate of the left gesture (e_l) and right gesture (e_r). This value of e_l from experiment B is,

$$e_1 = \frac{e_l + e_r}{2} = \frac{(16.3\% + 14.5\%)}{2} = 15.4\% \quad (1)$$

In the second experiment, with the cursor speed moving from 50 millisecond/tile to 450 millisecond/tile, the adjustment time decreased while the movement time increased. The average task time which equals adjustment time plus movement time reached the lowest value of 5.86 seconds when the cursor moves one tile every 250 milliseconds (Figure.5). Figure.6 is a histogram shows the range of the reaction time and the number of the blocks with which the user can properly react during different reaction time. For simplicity, the upper bound of one reaction time interval is chosen as the gesture length of all the blocks within this range.

4.4. Findings from Phase 1 Study

The result of Phase 1 user tests shows that the recognition rate is encouraging considering that the classifiers were trained with only five peoples sample oral images. From the perspective of human factor, it is interesting to notice that humans tongue protrusion abilities are not the same when reach

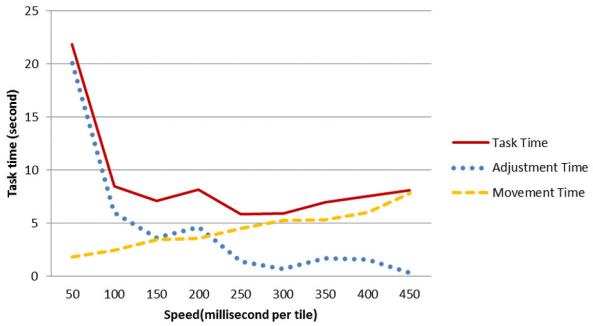

Figure 5: Movement time and adjustment time of the cursor moving task

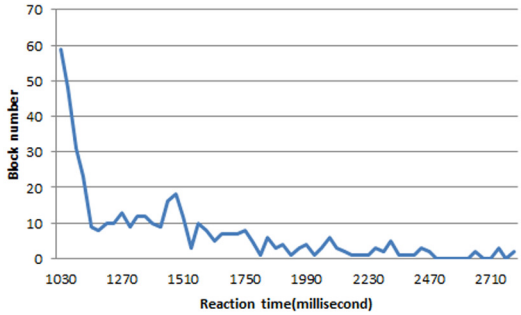

Figure 6: Histogram of blocks projected to reaction time range.

out to different directions. We also found that the parameters of interface such as the speed of cursor moving also impact the usability of CBTCI. This results from the fact that there exists a disparity between the time of gesture recognition and reaction time of human being, it is hardly possible for a participant to stop immediately when the cursor is right over the target, especially when the cursor is moving fast. Therefore in the non-contact TCI design, it is essential to strike a balance between the efficiency of the tongue instructions and the ability to tolerant the delay of the user reaction.

Through our observation, we found there are other gaps between the system implementation and the participants behavior while using the TCI. The first gap was that the participants expect the system can recognize their tongue gestures as long as they reached out their tongue. But in fact they had to adjust the tongue to a shape similar to the one in the sample images. Another gap was when the participants tilt or rotate their heads too much, which make the system unable to recognize the tongue gesture, some of the participants thought that keeping the gesture and waiting for a longer time would help the system recognize the gesture. The third gap was when users expected the cursor could move faster, they generally turned their head to the direction they protrude and reach their tongue as out as possible. This type of user gesture lowered the tongue recognition rate because less part of the tongue was captured by the camera.

Even though the prototype achieve an accuracy of more than 80 percent, the participants still showed some concerns about the recognition errors. From participants feedback we found that their satisfaction about of the system usability largely depended on the recognition rate. Incorrectly recognition of their gestures made them confused and even paused the ongoing task. Most of the comments of the system were about which direction cannot be recognized well and what part of the interface design helped them to realize the better manners to place their tongue. The accuracy of the assistive recognition technology shoed its impact on usability since a tiny error may result in large dissatisfaction of the system (Norman, 2002).

Sampling the tongue gestures with different degrees of head rotation and different strength of protrusion might improve the recognition rate. But as the human head and tongue have very high level of dexterity, it is hardly feasible to sample all the tongue gestures at all the possible head and tongue states. It is also impractical to require the participant to stop the task, collect the undetected gesture, update the classifiers and go back to do it again if the recognition error occurs.

However, as a common comment to our prototyping design, the participants noted that the referential techniques played an important role to help them place their tongue in a more recognizable way. As a participant mentioned that showing the camera video of himself is helpful, because he get aware of what the tongue gesture looked like when the system recognized his tongue gesture well. Another participant said: the red dot here is good, I can quickly make a tongue adjustment if it shows the direction I dont want. The red dot mentioned by this participant was the indicator window with a red dot notifying the detected tongue direction (Figure.4). All those comments implied that besides the improvement of recognition methods, referential strategies also played a meaningful role in reducing errors and improving the usability of CBTCI.

5. Phase 2: Referential Feedback

As mentioned in the first phase of the study, we noticed that referential information played an essential role in gesture-based tongue-computer interface. Referential information helps users adjust their behavior by providing feedback about the user input. Most input devices have some sort of referential information; e.g., keys on a laptop keyboard depress and click, a phone virtual keyboard lights up the letter that is pressed.

CBTCI presents referential information in a referential window. We seek to study how awareness and functional representation as referential techniques provide insights into the improvement of user experience when using CBTCI. We employed two techniques to provide referential feedback: a self-awareness mirror view and a functional representation gesture indicator.

The common practice to provide referential informations when using gesture-based technology is enabling self-awareness (Lin et al., 2010; Tu et al., 2005). Self-awareness strategy refers to presenting a camera video of a user himself and showing the real-time status of the tongue motion. Duval and Wicklund proposed that when a mirror or a camera is presented to people in self-awareness states, a user often self-identifies as an object to be evaluated and adjusts behaviors accordingly (Duval and Wicklund, 1973). Geller further indicated that false actions can be reduced according to the degree of self-awareness (Geller and Shaver, 1976). But as moving a cursor or pressing a button on the screen with CBTCI also require users mental process, monitoring the motions of oneself cause somewhat distractions from the main tasks.

Another method to help users monitoring the system status is through the functional representation. Functional representation is represented by abstract window components such as progress bar or symbols that indicate the status of critical parameters in the system. As those window components can be set at any size and be placed close to the working area, it is reasonable to assume that the functional representation is less intrusive to the major task. But as they are abstract form of system status, their ability to convey the system information still need examine. Building on the knowledge from the Phase 1 study, the next stage of the research explored referential feedback in CBTCI, exploring how the two referential strategies can help the user find a better way to position their tongues and overcome recognition errors.

The experimental task is designed to explore how the referential techniques influence the behavioral adaption text input with CBTCI. The user interface comprises a text pad and a text area. The text pad included two letter keys and a word list (Figure.7). The letters from A to N are assigned to the left key and the rest of the letters are assigned to the right key. The left and right tongue gestures are used to choose the corresponding letter-selection keys. After a tongue gesture being hold for 1800 milliseconds (slightly longer than the reaction time obtained from Phase I), the corresponding key will be selected. During a gesture, if the tongue moves, which leads to missed recognition, the timer to monitor the gesture holding time gradually rolls back to 0. If the gesture is detected again, the timer increases from its current value. After it reaches 1800, the selection is registered and the timer resets to 0.

After the key-strike sequence is finished, matched words will appear vertically in the word list. We use MacKenzies two key text-entry technique to generate the potential word list (MacKenzie, 2009). Up and down tongue gestures will be used to select a word. To confirm a word selection, the user will open his/her mouth and the selected word will appear in the text area. The mouth close gesture is the default gesture and no action is taken with it. When any of the four tongue gestures is maintained for 1.8 seconds, the matching instruction will be executed.

5.1. Self-awareness Strategy

The self-awareness strategy is presented in a different interface window (Figure.8 left). A mirror window showing the video stream captured by the camera is placed under the text pad. The user can watch the real-time video to get real-time feedback. Other facial components including face, eyes, nose

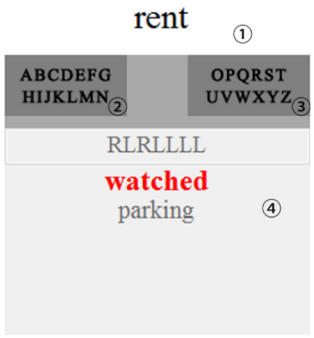

Figure 7: The text input application. The text pad locates in the middle top of the screen and the text area is full screen. ①Text area. ②Left key. ③Right key. ④Word list.

and mouth are outlined to illustrate the recognition. A dot/circle shape is also shown in the mouth frame to reflect the gesture by appearing in different places in the mouth frame.

5.2. Functional Representation Strategy

To implement the functional representation in the task, four progress bars are used to indicate the time of each tongue protrusion and a gesture icon representing the current recognized tongue gesture is included in the text pad (Figure.8 right). The four progress bars located on the left, right, top and bottom in the text pad are to indicate the time of corresponding directional tongue gestures. The icon between the left and right keys changes based on the tongue gesture.

5.3. Action model of the text-input task

The variables explored in the study include the potential errors in the input, and peoples ability to reach and hold the correct gesture. Text entry with the tongue-computer interface is accomplished with a series of tongue actions. Regarding the parameters to be studied, an action is defined as one letter-key press, word selection, or word confirmation. Actions can be correct or incorrect. For each correct action, there are two stages to finish the action: reaching the proper gesture Adjustment Time (AT) and holding the gesture until execution Effective Time (ET). The first stage starts from the completion of the previous actions, to the intended gesture is detected. The second stage starts from the detection of the gesture, to the computer execution of the instruction. The incorrect actions are actions unrelated to

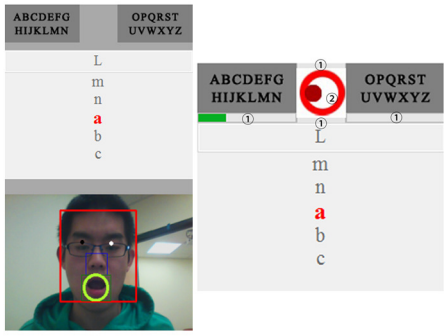

Figure 8: Referential feedback approaches. Left: Self-awareness window. A mirror window at the bottom of the screen shows the current video frame of the user. Right: Functional representation window. At the top center, ① four progress bars reflect the time of directional tongue gestures, and ② Gesture indicator points out the current recognized tongue position.

spelling the intended sentence, such as incorrectly pressing left and right keys, extra up or down movement on the word list or confirming an incorrect word. The roles of self-awareness and functional representation in correct and incorrect actions in two stages are studied.

5.4. Experiment Protocol

Three female and nine male college students (mean age=22.67, SD=3.26) at CS department of Virginia Tech participated in the Phase 2 experiments. The participants had no motion or vision impairments, because at this stage we were interested in usability of the system by the general population. Participants were randomly assigned into one of two groups (A and B). One Group A participant and one Group B participant did not finish the experiment due to personal time limitations; their data were not included. Before the experiment, we collected training data from all participants. Before the task began, all participants took part in a 5-minute training session to get familiar with the CBTCI, in which participants from both groups finished the baseline task (Task 1) three times without any referential strategies. During the Task 1, participants were asked to type rent watched the entire movie using the tongue interface. Afterwards, participants were asked to type as fast as possible using the tongue interface with two types of interfaces (Task 2).

We chose to use a between-subjects design to minimize the learning curve for the participants important because of the significant time demands al-

ready placed on the participants due to the machine learning training phase of the study. Participants from Group A used the functional representation strategy and those from Group B used self-awareness strategy. Each participant performed the task using the tongue interface with one of the referential strategy three times. Before undertaking the baseline task and the referential task, participants had a chance to practice with their respective interface. Before starting, participants acknowledged that all letters are in lower case and spaces are automatically added between words. They were also informed that recognition error might occur. If any error occurs they were told to use the mouth open gesture to pop the word to the text area and try the same word again. They did not need to delete or correct the misspelling. After the participants finished the two tasks, they were asked to finish a 10-entry questionnaire to evaluate their experience in using the interface. We also probed participants for open-ended qualitative feedback on their experience.

5.5. Results

Our experiment showed that without the referential strategy, the variance of AT in Task 1 was very large. Participants varied widely in their understanding and chosen strategies when adjusting their tongue position. According to the normality tests, none of the data obtained from the study followed normal distribution. Therefore, nonparametric analyses were used in this study instead (Kolmogorov-Smirnov Test (Lilliefors, 1967)). The Kolmogorov-Smirnov Test results indicated that participants from both groups do not have significant difference in ET during the baseline trials, but showed significant difference in AT during the baseline trials.

The results also showed that without the referential information, participants effective time does not differ significantly for the last baseline trials, but their ET differed significantly during the trials with different referential strategies (Figure.9). The average ET of Group A of all three trials in Task 2 is 2.25s (SD=1.01), while with self-awareness strategy, participants in Group B achieve the ET of 2.17s (SD=0.86), which is significantly less than Group A ($p<0.0001$). The ET values of Group B are smaller than Group A in all three trials with last two trials in significant difference ($p=0.0468$ and 0.0300).

The numbers of errors in the task with referential strategy were counted in the three trials (#1, #2, #3) for the five participants in Group A (e.g., A01, A03) and the five in Group B (see Table.2). The number in the table was obtained by subtracting the number of necessary actions from the total

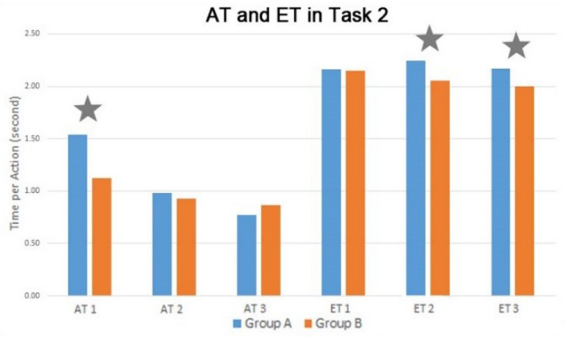

Figure 9: Adjustment time and effective time of the 3 trials in the Task 2. AT 1-3 on the left data of the three trials in Task 2, ET 1-3 on the data of three trials in Task2. The trial marked with star is with significant difference between Group A and Group B.

Table 2: The number of errors for different referential strategies in the task for the five participants in each of the two groups (A and B).

	A01	A03	A04	A06	A07	B01	B02	B04	B06	B07
#1	0	0	2	0	0	1	0	12	0	21
#2	0	0	0	0	0	1	11	10	8	8
#3	0	0	2	2	1	0	0	24	0	8

number of actions taken in the system in the corresponding trial. Unnecessary actions include misspelling, moving the words list before the letter sequence is finished, and passing the correct word and moving back.

5.6. Findings from Phase 2 Study

The statistics revealed that the self-awareness showed no significant advantages in assisting user find the proper tongue gesture over functional representation, as the time to reach the correct gesture was similar when using two referential strategies. However, the self-awareness assisted the user in holding the intended gesture more effectively than does the functional representation. The main effect of self-awareness in gesture interaction is in helping people monitor the changes of the gesture to make necessary adjustments so that the gesture does not unconsciously go out of a recognizable shape. The statistics of error actions also indicated that when using the self-awareness strategy, users were more likely to produce errors than when

using the functional representation. As when evaluating the tongue gesture through camera window, it was easier for people to neglect the execution of the instruction. These errors may have highly negative consequences on the operations that require a series of actions because the actions after the first error might be useless.

The results of questionnaire reflected rational behind our findings. Most notably in one question The additional indicator window helped me avoid errors I made during in trials without them (1 is almost none, 7 is almost all) participants in Group A gave 5.601.1 (mean SD) and B answered less enthusiastically 5.172.3 (p=0.0161). This suggests self-awareness did not show advantages over the functional representation in avoiding making less recognizable gestures when using CBTCI.

6. Discussion and Conclusion

Through our two-phase study, we demonstrate that camera-based tongue computer interface is a promising hand-free assistive technology. CBTCI enables people with dexterity impairment access to various digital devices without the trouble of carrying and wearing special devices. With the pervasive use of camera, CBTCI can be easily migrated to mobile devices like smartphones and tablets.

Intrusiveness highly influences users intention and willingness to use an TCIs. People who needs hands-free input shall consider using tongue as the motor part to operate assistive devices or computers. However, oral hygiene is a common concern when using in-mouth devices, existing hardware-based tongue technologies may not be suitable for long-term use without regular clean-ups. Computer vision based approach does not require special assistance in implanting the device. Our computer vision based approach does not rely on special hardware or oral installation: the user can interact with the interface directly without the help of caregivers.

Usability is another factor we considered when designing and implementing the tongue-based interface. The directional tongue gestures are intuitive and easy to perform. From our user study, participants successfully finished tasks with relatively low reaction time and error rate. Most of other tongue-based interaction techniques either ask the user to remember different tongue gestures and corresponding computer commands such as (Slyper et al., 2011), or use the binary-style operation and only allow the user to do the yes-or-no task such as (Struijk et al., 2009). Though the user study shows the error

rate of our device is higher than some other in-mouth devices, we noticed that the error was mostly caused by false recognition of the tongue gesture, rather than the user performing the wrong tongue action. We see the room where the accuracy and efficiency of our approach can increase by improving the recognition algorithm.

Besides the intrusiveness and usability, the affordances of the assistive technologies also decide their applications in people's daily life. Our approach advances in that users can perform multiple types of interaction. The directional operation can be used for both characters/words selection and cursor movement. Therefore the user can perform texting and pointing tasks with one single interaction technique. These affordances make the CBTCI suitable to operate different types of digital devices. With the camera becoming a standard device on most laptop computers and mobile devices, our approach is promising in supporting tongue interaction with different types of digital devices.

The joystick-like operation of the prototype also proved to be an easy and direct scheme which does not require special learning efforts. In the questionnaire of Phase 2 study, two groups of participants gave 4.2 and 4.8 (SD=0.45 and 1.17) for the training session, with 1 being too short and 7 being too long. This indicates that most participants easily understand how to use the CBTCI and easily get familiar with the interaction technology.

The system achieves an accuracy of 83% with a training set collected from 5 participants. There is definitely a room of improvement in the recognition rate. But during the user test of the CBTCI prototype, we discovered that some factors besides the recognition algorithm influence the recognition rate and further impact the usability of CBTCI. It is to our curiosity realizing that one of those factors is referential information, which affect the process of gesture adjustment and reduce the fluctuation of the user behavior. We maintain that the adjusting process is necessary for gesture recognition system like CBTCI, and the knacks to perform the tongue gesture more recognizable need the user to discover by trials and errors. To scaffold the process of gesture adjustment, referential information is important. It acts as a notification of improper tongue gestures which cannot be recognized correctly by the CBTCI. Additionally, after a correct tongue gesture is performed, referential techniques such as self-awareness window can help users maintain the gesture and overcome the unconscious motion that compromise accuracy.

Another important finding in our study lies in that different referential strategies also have different influences to the gesture adapting process.

Showing a camera window of users themselves can make them aware of their gestures. Participants manipulated themselves through tracing their behaviors. Comparing to functional representation, self-awareness showed no significant benefit to adjust the gesture to a recognizable state. But the statistics showed that participants use the self-awareness tool to maintain their gestures until the execution of the instruction. When a participant tried to press a button, he/she used the referential tools to find the more recognizable manner. Participants with camera window tended to look at themselves to hold the gesture. This strategy showed value to help the user keep the gesture, which reduce the effective time for tongue instruction. However, monitoring oneself through a camera window also brings more distraction than using extracted window components. The experiment showed that participants tend to produce more unnecessary or wrong actions when using self-awareness referential method. When attempted to press the left button once during a trial, one of the participants even held the left gesture and looked at himself until he realized the left button has already been pressed for five times. This kind of distraction resulted from the self-awareness should be carefully considered in the future CBTCI design.

Overall, the new possibilities brought by the tongue-computer interface benefit the diversity group by providing an easier access to digital technology and novel experience of human-computer interaction. To foster future tongue-related research and product design, our study of CBTCI probes the fundamental opportunities and challenges lie in the tongue computer interface. With the improvement of the recognition techniques accompanied with proper referential strategies, we believe that non-contact tongue interfaces will keep emerging and become a widely used technology to develop hand-free interaction applications.

7. Future Work

This work shows how tongue gestures can be used in building an non-contact human-computer interface. The paper demonstrates the design and implementation of a camera-based tongue computer interface and our usability study of pointing and text input tasks. Moving forward, it is important to develop and test applications that leverage these findings for people with disabilities; e.g., browsing accessible webpages and documents and reading SMS and social media feeds. As we develop and explore a broader set of applications used by larger numbers of participants, we expect to see differ-

ent strategies emerge for fast, accurate, and easy gesturing reflecting how people learn how to gesture in ways that are recognizable by the system. As our understanding grows regarding ways that tongue gesturing can meet the needs of people with disabilities, we look forward to the improved well-being and increased independence of users.

8. Reference

Brault, M. W., et al., 2012. Americans with disabilities: 2010. US Department of Commerce, Economics and Statistics Administration, US Census Bureau Washington, DC.

Chandra, R., Johansson, A. J., Aug 2011. In-mouth antenna for tongue controlled wireless devices: Characteristics and link-loss. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 5598–5601.

Chang, T.-H., Yeh, T., Miller, R., 2011. Associating the visual representation of user interfaces with their internal structures and metadata. In: Proceedings of the 24th annual ACM symposium on User interface software and technology. ACM, pp. 245–256.

Cheng, J., Okoso, A., Kunze, K., Henze, N., Schmidt, A., Lukowicz, P., Kise, K., 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. ACM, New York, NY, USA, pp. 12:1–12:4.
URL http://doi.acm.org/10.1145/2582051.2582063

Dalka, P., Bratoszewski, P., Czyzewski, A., 2014. Visual lip contour detection for the purpose of speech recognition. In: Signals and Electronic Systems (ICSES), 2014 International Conference on. IEEE, pp. 1–4.

Duval, S., Wicklund, R. A., 1973. Effects of objective self-awareness on attribution of causality. Journal of Experimental Social Psychology 9 (1), 17 – 31.
URL http://www.sciencedirect.com/science/article/pii/0022103173900590

Geller, V., Shaver, P., 1976. Cognitive consequences of self-awareness. Journal of Experimental Social Psychology 12 (1), 99 – 108.
URL http://www.sciencedirect.com/science/article/pii/0022103176900895

Hinckley, K., Jacob, R. J., Ware, C., Wobbrock, J. O., Wigdor, D., 2014. Input/output devices and interaction techniques.

Huo, X., Ghovanloo, M., 2010. Evaluation of a wireless wearable tongue-computer interface by individuals with high-level spinal cord injuries. Journal of Neural Engineering 7 (2), 026008.
URL http://stacks.iop.org/1741-2552/7/i=2/a=026008

Hurst, A., Tobias, J., 2011. Empowering individuals with do-it-yourself assistive technology. In: The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility. ACM, pp. 11–18.

Kim, J., Park, H., Ghovanloo, M., Aug 2012. Tongue-operated assistive technology with access to common smartphone applications via bluetooth link. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 4054–4057.

Lau, C., OLeary, S., 1993. Comparison of computer interface devices for persons with severe physical disabilities. American Journal of Occupational Therapy 47 (11), 1022–1030.
URL + http://dx.doi.org/10.5014/ajot.47.11.1022

Lilliefors, H. W., 1967. On the kolmogorov-smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association 62 (318), 399–402.
URL http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1967.10482916

Lin, S. Y., Lai, Y. C., Chan, L. W., Hung, Y. P., Aug 2010. Real-time 3d model-based gesture tracking for multimedia control. In: 2010 20th International Conference on Pattern Recognition. pp. 3822–3825.

Liu, L., Niu, S., McCrickard, S., July 2017. Non-contact human computer interaction system design and implementation. In: 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). pp. 312–320.

Liu, L., Niu, S., Ren, J., Zhang, J., 2012. Tongible: A non-contact tongue-based interaction technique. In: Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility. ASSETS '12. ACM, New York, NY, USA, pp. 233–234.
URL http://doi.acm.org/10.1145/2384916.2384969

MacKenzie, I. S., 2009. The one-key challenge: Searching for a fast one-key text entry method. In: Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility. Assets '09. ACM, New York, NY, USA, pp. 91–98.
URL http://doi.acm.org/10.1145/1639642.1639660

McCrickard, D. S., Chewar, C. M., Somervell, J. P., Ndiwalana, A., Dec. 2003. A model for notification systems evaluation—assessing user goals for multitasking activity. ACM Trans. Comput.-Hum. Interact. 10 (4), 312–338.
URL http://doi.acm.org/10.1145/966930.966933

Mimche, S., Ahn, D., Kiani, M., Elahi, H., Murray, K., Easley, K., Sokoloff, A., Ghovanloo, M., 2016. Tongue implant for assistive technologies: Test of migration, tissue reactivity and impact on tongue function. Archives of Oral Biology 71, 1–9.

Miyauchi, M., Kimura, T., Nojima, T., 2013. A tongue training system for children with down syndrome. In: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology. UIST '13. ACM, New York, NY, USA, pp. 373–376.
URL http://doi.acm.org/10.1145/2501988.2502055

Mugler, E. M., Ruf, C. A., Halder, S., Bensch, M., Kubler, A., 2010. Design and implementation of a p300-based brain-computer interface for controlling an internet browser. IEEE Transactions on Neural Systems and Rehabilitation Engineering 18 (6), 599–609.

Nam, H. Y., DiSalvo, C., 2010. Tongue music: The sound of a kiss. In: CHI '10 Extended Abstracts on Human Factors in Computing Systems. CHI EA '10. ACM, New York, NY, USA, pp. 4805–4808.
URL http://doi.acm.org/10.1145/1753846.1754235

Niu, S., Liu, L., McCrickard, D. S., 2014. Tongue-able interfaces: Evaluating techniques for a camera based tongue gesture input system. In: Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility. ASSETS '14. ACM, New York, NY, USA, pp. 277–278.
URL http://doi.acm.org/10.1145/2661334.2661395

Norman, D. A., 2002. The psychopathology of everyday things. Foundations of cognitive psychology: core readings. MIT Press, Cambridge, MA, 417–443.

Park, H., Ghovanloo, M., 2016. A wireless intraoral tongue–computer interface. Wireless Medical Systems and Algorithms: Design and Applications 56, 63.

Polacek, O., Mikovec, Z., Sporka, A. J., Slavík, P., 2011. Humsher: a predictive keyboard operated by humming. In: The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility. ACM, pp. 75–82.

Polikar, R., 2006. A tutorial article on ensemble systems including pseudocode, block diagrams and implementation issues for adaboost and other ensemble learning algorithms. IEEE Circuits and Systems Magazine 6, 21–45.

Rautaray, S. S., Agrawal, A., 2015. Vision based hand gesture recognition for human computer interaction: a survey. Artificial Intelligence Review 43 (1), 1–54.

Saponas, T. S., Kelly, D., Parviz, B. A., Tan, D. S., 2009. Optically sensing tongue gestures for computer input. In: Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology. UIST '09. ACM, New York, NY, USA, pp. 177–180.
URL http://doi.acm.org/10.1145/1622176.1622209

Slyper, R., Lehman, J., Forlizzi, J., Hodgins, J., 2011. A tongue input device for creating conversations. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. UIST '11. ACM, New York, NY, USA, pp. 117–126.
URL http://doi.acm.org/10.1145/2047196.2047210

Struijk, L. A., Bentsen, B., Gaihede, M., Lontis, E., 2017. Error-free text typing performance of an inductive intra-oral tongue computer interface for severely disabled individuals. IEEE Transactions on Neural Systems and Rehabilitation Engineering PP (99), 1–1.

Struijk, L. N. A., Lontis, E. R., Bentsen, B., Christensen, H. V., Caltenco, H. A., Lund, M. E., 2009. Fully integrated wireless inductive tongue computer interface for disabled people. In: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 547–550.

Tu, J., Huang, T., Tao, H., May 2005. Face as mouse through visual face tracking. In: The 2nd Canadian Conference on Computer and Robot Vision (CRV'05). pp. 339–346.

Van De Weijer, J., Gevers, T., Gijsenij, A., 2007. Edge-based color constancy. IEEE Transactions on image processing 16 (9), 2207–2214.

Wang, R., 2012. Adaboost for feature selection, classification and its relation with svm, a review. Physics Procedia 25, 800–807.

Ziegler-Graham, K., MacKenzie, E. J., Ephraim, P. L., Travison, T. G., Brookmeyer, R., 2008. Estimating the prevalence of limb loss in the united states: 2005 to 2050. Archives of Physical Medicine and Rehabilitation 89 (3), 422 – 429.
URL http://www.sciencedirect.com/science/article/pii/S0003999307017480