Optically Sensing Tongue Gestures for Computer Input

T. Scott Saponas[1], Daniel Kelly[2], Babak A. Parviz[2], and Desney S. Tan[3]

[1]Computer Science & Engineering
DUB Group
University of Washington
ssaponas@cs.washington.edu

[2]Electrical Engineering
University of Washington
{kellyd22, parviz}@u.washington.edu

[3]Microsoft Research
desney@microsoft.com

ABSTRACT

Many patients with paralyzing injuries or medical conditions retain the use of their cranial nerves, which control the eyes, jaw, and tongue. While researchers have explored eye-tracking and speech technologies for these patients, we believe there is potential for directly sensing explicit tongue movement for controlling computers. In this paper, we describe a novel approach of using infrared optical sensors embedded within a dental retainer to sense tongue gestures. We describe an experiment showing our system effectively discriminating between four simple gestures with over 90% accuracy. In this experiment, users were also able to play the popular game Tetris with their tongues. Finally, we present lessons learned and opportunities for future work.

ACM Classification: H.1.2 [User/Machine Systems]; H.5.2 [User Interfaces]: Input devices and strategies; B.4.2 [Input/Output Devices]: Channels and controllers

General terms: Design, Human Factors

Keywords: Tongue-Computer Interface, infrared, gestures.

INTRODUCTION

Traumatic brain and spinal cord injuries as well as medical conditions such as amyotrophic lateral sclerosis (also known as Lou Gehrig's disease) often leave patients severely paralyzed. Many of these patients retain highly functional cognitive abilities and there is great value in creating alternate input modalities that allow them to interact with computers and with the world around them.

Fortunately, the cranial nerves, which control organs such as the eyes, jaw, and tongue, often go unaffected even in severe injuries and neuromuscular diseases. While there has been quite a bit of work applying technologies such as eye-tracking and speech recognition in such scenarios [2, 3], much less effort has been placed in exploring the use of explicit tongue gestures for communication and control.

The most obvious way to exploit direct control with the tongue is to provide physical transducers the tongue can actuate or manipulate. For example, both Peng et al. and Salem et al. create what amounts to a joystick that the user can control with their tongue [5, 6]. Similarly, a commercial device from New Abilities Systems embeds pressure sensitive buttons into a dental retainer placed on the roof of the user's mouth (www.newabilities.com).

These devices treat the tongue much as a finger and do not exploit its unique ergonomic abilities. For example, continuously curling the tongue to push or tap buttons located on the palette is potentially awkward and tiring. In fact, the tongue is a highly flexible skeletal muscle most often used for generating speech as well as manipulating and swallowing food. Both these activities require a high degree of control over tongue shape and position, and suggest opportunities for designing more natural and richer gesture spaces.

Researchers who have realized this opportunity have tried to track complex movements by instrumenting the tongue with metallic piercings or magnetic attachments [4, 6]. The movement of the attached elements within the mouth can then be detected either by a dental retainer worn in the mouth or by a separate device worn outside the mouth. Unfortunately, these tongue augmentations are quite obtrusive and while they might be marginally acceptable to the disabled population with few other options, do not make them appealing to otherwise healthy individuals.

In our work, we embed optical sensors into orthodontic dental retainers worn in the mouth. These sensors provide us with the potential to robustly sense explicit and complex tongue movement. Building the sensing device into a dental retainer creates a form factor that is both easy to don, but also largely undetectable to an observer. This is important to disabled individuals, who typically prefer to maintain a sense of normalcy, as well as healthy individuals when traditional forms of computer control are inadequate. Furthermore, many people already wear dental retainers or

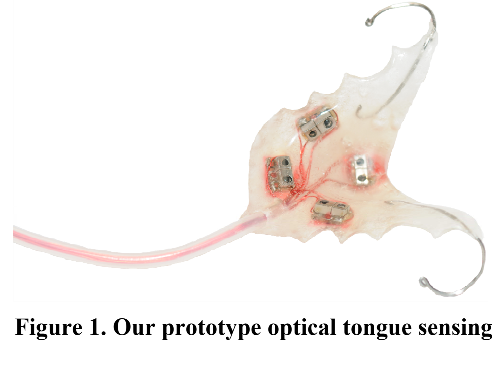

Figure 1. Our prototype optical tongue sensing retainer.

UIST'09, October 4–7, 2009, Victoria, British Columbia, Canada.

dentures for medical reasons, and providing technological functionality could be a relatively appealing proposition.

TONGUE-SENSING DENTAL RETAINER

Our prototype for sensing tongue gestures consists of an orthodontic dental retainer with four embedded optical proximity detectors (see Figure 1) and a custom data acquisition board with an amplifier and microcontroller.

To build the dental retainer, we first take a physical impression of the mouth and teeth using a biocompatible material called alginate. While this is not difficult and takes less than 5 minutes, we had a local dentist assist with this. From the alginate mold, we make a plaster casting, which serves as the fabrication platform for the retainer.

We place a separating material called tinfoil substitute on the mold to ease releasing the retainer when it is completed. We then apply a thin coat of dental acrylic which forms the part of the retainer that will touch the user's palette. It is important to us that our augmented device be as biocompatible and safe as its medical counterpart. As such, we ensured that the technology would be entirely embedded within the acrylic and that the only object the user would ever come in to contact with was the retainer itself. To hold the retainer against the user's palette, we bend medical grade stainless steel wires to grasp the back molars.

We then place four proximity sensors, one each on the right, left, front, and back within the acrylic. We oriented the sensors such that their sensing rays intersected loosely at a point in the middle of the tongue and slightly above its resting position at the bottom of the mouth. We believe this provides maximal discriminability in the combined signal.

The first prototype of our retainer design, and the one we tested in this paper, was a wired device. We fed ten 30-gauge wires running from the sensors out the front of the mouth in a small bundle that was as unobtrusive as we could get it. The wires are a temporary solution and we will discuss construction of a wireless version later in this paper. We use more dental acrylic to encapsulate the technology. Once cured, we grind the retainer to the approximate shape and then sand and polish to the final form.

Each proximity sensor is made up of an infrared light emitting diode (LED) and an infrared photodiode. We used Avago Technologies HSDL-9100 surface mount proximity sensors (www.avagotech.com), which measured 7.1mm × 2.75mm × 2.4mm (tall). As an object approaches the sensor, variable amounts of light from the LED reflects back into the photodiode, allowing us to sense proximity of the tongue to each sensor. Since the mouth stays relatively dark, the signal remains extremely robust, even when the user has their mouth open in sunlight.

Our data acquisition board consists of a microcontroller that activates the LED and reads the signal off the photodiode for each sensor in sequence. It samples the sensors at approximately 90Hz. The board then amplifies the signal from the photodiode by a level controlled by a potentiometer on the acquisition board. We tuned these potentiometers individually for each retainer so that the sensors would detect the tongue as it got in range. Because the timing between the left and right sensors is critical to gesture detection, we also took care to set the left and right sensors to equal sensitivity. Finally, we pass the signal through the microcontroller's analog-to-digital converter, and then over a serial connection to the desktop computer for processing.

Data Processing and Gesture Recognition

We created desktop software to process data from the retainer, recognize tongue gestures, provide feedback, and control a real-time system. The system currently recognizes four tongue gestures: a left swipe, a right swipe, tap up, and hold up (see Figure 2). We chose these gestures because they can be performed easily from the tongue's natural resting position and without applying pressure to any part of the mouth or retainer. Our system recognizes these gestures based on simple timing and duration relationships as the tongue is detected by the proximity sensors.

We process the raw signal by smoothing the proximity sensors' signals using a sliding average with a window size of 100ms. From this we continually determine a binary tongue-presence state using a threshold of 50% of the possible value. We employ this state in each of the sensors to recognizing gestures using a set of timing constraints.

Left and right swiping gestures are detected by the left sensor and the right sensor being activated between 40ms and 500ms of each other. In working with the system, we identified several ways people could comfortably make the left and right swiping gestures. One method is to swipe the tip of the tongue across the back of the bottom teeth. When

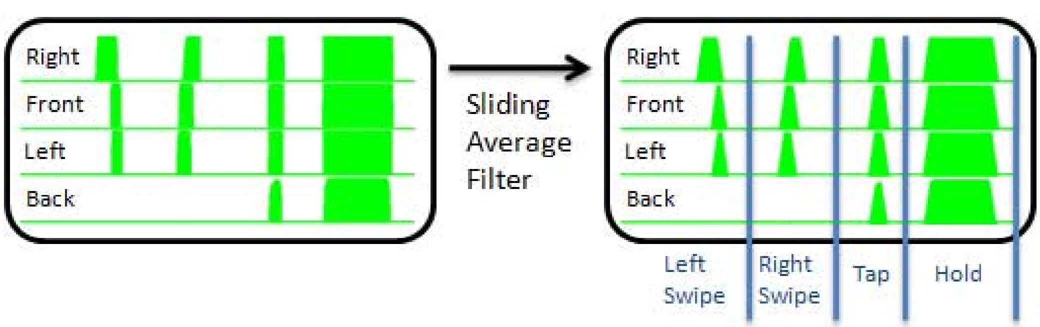

Figure 2. Example proximity sensor signal and processing. Each row represents the proximity data from a different sensor. The box at right shows the effect of a sliding average filter and examples of our four tongue gestures.

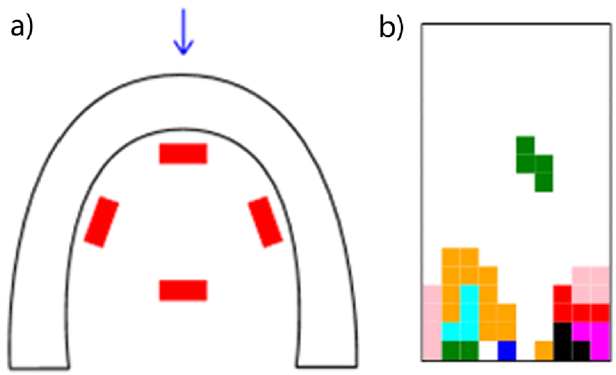

Figure 3. a) Sensor state visualization and recognized gesture feedback b) Tetris game

this is done, the middle section of the tongue rises toward the roof of the mouth, and is detected by the proximity detectors. Interestingly, a left to right swipe is actually manifested as the middle of the tongue rising in the opposite order. The other way people made the left to right gesture was to swipe the surface of their tongue from left to right on the top of their mouth or to swipe the tip of their tongue along their upper teeth. The proximity data can be used to determine how close the tongue is to the retainer and which of these gestures are being performed.

A tap gesture consists of all four sensors activating within 40ms of each other and then deactivating less than a second later. If the user continues to activate the sensors for longer than a second, a hold gesture is recorded. Both of these gestures could be physically performed by raising the tongue toward the roof of the mouth, for example making a gesture that would generate an exaggerated "L" sound. Samples of the signals can be seen in Figure 2.

While we empirically selected the timing and durations to work for the small set of our users, a more systematic calibration methodology that takes individual differences into account remains future work.

EXPERIMENT

We conducted a laboratory study to explore the feasibility of our approach to sensing tongue gestures for input. Four people (2 female) from our research organization volunteered for the experiment. Each received a small gratuity.

Prior to the experiment, we took participants to a dentist to have their impressions taken, so we could fabricate custom retainers. In a pre-experimental session, we checked that their retainers correctly fit their mouths and let each briefly test the retainer as an input device. The back sensor in one participant's retainer did not work very well, so we modified the logic to use only the remaining three sensors.

We broke the experiment into three distinct phases. In the first phase, the user donned their retainer and freely explored various tongue gestures while visualizing the raw and processed signals, as well as the gesture recognition on the four gestures (see Figure 3a). This was intended both to allow them to get used to the system, but also to solicit feedback on the gestures that were natural and comfortable, and that seemed to get recognized, given our algorithms.

In the second phase, we used a stimulus-response paradigm in which the system prompted the participant with a gesture to perform. Once the participant performed a recognized gesture, the system provided feedback and moved to the next gesture. We used a block design with each block containing the four gestures presented in random order. Participants practiced the task on sets of two blocks until they felt comfortable. None required more than six blocks of practice. The actual test comprised two sets of six blocks each, for a total of 48 gesture trials. The goal of this phase was to understand the gesture recognition accuracy when participants were focused solely on performing the gestures.

In the third phase of the experiment, we had each participant use the four tongue gestures they had learned to play the popular game Tetris (see Figure 3b). In Tetris, players rotate and move falling pieces so that they fit with existing pieces. In our tests, pieces fell by a row once every second. We chose this game because it has both a cognitive and a time-pressure aspect. In order to successfully play the game, the user must not only make the correct gestures, but must concentrate on deciding on the correct moves as well. In our implementation, left and right swipe gestures moved pieces left and right, tapping the tongue rotated the current piece clockwise, and holding the tongue against the palette dropped the current piece to the bottom. Our participants had varied experience playing Tetris. The entire experiment lasted between 30 and 50 minutes.

Results

Qualitative feedback

We learned several things while participants were experimenting with the retainers. First, the shape of the tongue at various points in the mouth is not often under direct control of the user. For example, while swiping their tongues across their teeth, participants noticed that even subtle differences sometimes affected our recognition system. Depending on whether they tried to do this at the base of the teeth or on top of the teeth, the middle of the tongue would deform in slightly different ways, causing our binary recognition scheme to perform slightly differently. We believe that understanding these ergonomics and deformations carefully is critical in designing more accurate and higher bandwidth recognition systems.

We also found quite a range of individual differences, both in terms of comfort and preference for gestures, but also in how these gestures would culminate in tongue deformations. Our participants were able to learn the intended gestures very quickly, but the differences points to a need for personalized models, especially for complex gestures.

Quantitative Stimulus-Response Performance

In the stimulus-response phase, participants specified gestures with a relatively high mean accuracy of 92.2% (S.D.=15.6). They missed only 15 gestures out of 192 total trials. Unfortunately, we cannot differentiate between participant and system errors. On average, each gesture took

1.5 seconds (S.D.=0.22) to perform and recognize. This is encouraging as it suggests that all four of our users were able to learn and control the system with relative ease.

In-place Use

In playing Tetris, two participants were able to play the game immediately, while the third required several warm-up games to get used to the tongue gestures. These three participants played successfully for five minutes without blocks stacking to the top, at which point we stopped them. In these five minutes, they used an average of 180 moves (S.D.=30) to place 56 pieces (S.D.=4) and successfully clear 10 lines each (S.D.=5). This is roughly 1.67 seconds per move, which is again encouraging. Our fourth participant, who's back sensor did not work, was not able to play the game well enough to clear lines. Misrecognitions and non-recognitions happened frequently enough that they misplaced many of their pieces.

DISCUSSION AND FUTURE WORK

Though we are encouraged by our preliminary findings, which suggest that there is potential in the optical approach to sensing tongue gestures, much work remains to be done.

Retainer Fabrication and Sensors

One point to ponder with retainer-based devices is the need for custom fabrication. Since each person's mouth and teeth are shaped differently, this is an inherent limitation. That said, we have consulted with various dentists and orthodontic laboratories and do not think such a proposition would be out of the question. More importantly, we would have to show high utility for such devices, and we believe this work is only a small first step towards this.

One limitation to the infrared proximity sensors we used was that the air-saliva-acrylic interface caused some infrared light to reflect back before reaching the tongue. As such, the effective range of the sensors was short, which led to the choice of binary classification in our gesture recognition. As we design more complex gestures requiring higher resolution proximity data, we will explore custom sensors as well as sensor placements that derive more information.

Our current prototype was wired, which is obviously not acceptable for real world use. We are now working on a wireless version of this device. The design mounts a ZigBee system-on-chip, IR sensors, and amplification circuit onto a flexible substrate that is embedded with a small inductively rechargeable battery in the acrylic of the retainer.

Exploring More Complex Gestures

The four gestures we decided on for this experiment were a small step towards testing the technology for sensing complex gestures. We have learned through this experiment that there is opportunity in understanding much more thoroughly how the tongue moves and how the user can directly and indirectly control various parts of it. In this paper we only employ *new* tongue gestures. Potentially, a tongue input system could also leverage the tongue movements we are already familiar with from speech.

One issue that will arise when continuously sensing the tongue is that the tongue is active when the user speaks and eats. We think that an explicit gesture that is not frequently naturally invoked could be used to toggle recognition and avoid the "midas touch."

Other Potential Uses of Technology Embedded in Retainers

While we have focused on decoding tongue movements for computer input in this paper, there is also an opportunity to use the platform for output. For example, researchers have explored using tactile stimulation to deliver relatively high resolution imagery to the tongue [1]. We are also currently exploring bone conduction technologies that allow us to deliver audio to the inner ear through bones in the skull.

Additionally, we could place sensors to monitor other activity in the mouth, such as jaw tension, movement, or even chemical changes in the saliva. Research has shown enzymatic changes in the saliva to be indicative of factors such as stress, and we hypothesize could also be an interesting marker for hunger, perhaps even before the user senses it.

CONCLUSION

In this paper, we have presented a novel approach of using infrared optical sensors embedded within an orthodontic dental retainer in order to sense tongue gestures. We have shown relatively promising results suggesting that the approach is feasible for a range of interactions. Additionally, we have presented lessons learned and opportunities for future work both with the retainer as an input platform, but also for output as well as other forms of sensing.

ACKNOWLEDGMENTS

We thank Dodi Nov, DDS, and Penn-Brookside Orthodontics Laboratory for assistance fabricating retainers.

REFERENCES

1. Bach-y-Rita, P., Kaczmarek K., Tyler M., & Garcia-Lara, J. (1998). Form perception with a 49-point electrotactile stimulus array on the tongue. *J Rehabil Res Dev*, 35, 427–430.
2. Chin, C. A., Barreto, A., Cremades, J. G. and Adjouadi, M. 2008. Integrated electromyogram and eye-gaze tracking cursor control system for computer users with motor disabilities. *J Rehabil Res Dev,* 45(1), 161-174.
3. Fried-Oken, M. 1985. Voice recognition device as a computer interface for motor and speech impaired people, *Archives of Physical Medicine and Rehabilitation,* 66(10), 678-81.
4. Huo, X., Wang, J., and Ghovanloo, M. 2008. A magneto-inductive sensor based wireless tongue-computer interface, *IEEE Trans on Neural Sys Rehab Eng 16(5)*, 497-504.
5. Peng, Q., Budinger, T.F. ZigBee-based wireless intraoral control system for quadriplegic patients. *Conf Proc IEEE Eng Med Biol Soc. 2007*, 1647-50.
6. Salem, C. and Zhai, S. 1997. An isometric tongue pointing device. In *Proc. of CHI '97 Conference*, 538-539.
7. Struijk, L. 2006. An inductive tongue computer interface for control of computers and assistive devices, *IEEE Trans on BioMed Eng,* 53(12): 2594-2597.