Tongue Gestures for Hands-Free Interaction in Head Worn Displays

Tan Gemicioglu[*]
tgemici@gatech.edu
Georgia Institute of Technology
Atlanta, GA, USA

Mike Winters
mikewinters@microsoft.com
Microsoft Research
Redmond, WA, USA

Yu-Te Wang
yutewang@microsoft.com
Microsoft Research
Redmond, WA, USA

Ivan Tashev
ivantash@microsoft.com
Microsoft Research
Redmond, WA, USA

ABSTRACT

Head worn displays are often used in situations where users' hands may be occupied or otherwise unusable due to permanent or situational movement impairments. Hands-free interaction methods like voice recognition and gaze tracking allow accessible interaction with reduced limitations for user ability and environment. Tongue gestures offer an alternative method of private, hands-free and accessible interaction. However, past tongue gesture interfaces come in intrusive or otherwise inconvenient form factors preventing their implementation in head worn displays. We present a multimodal tongue gesture interface using existing commercial headsets and sensors only located in the upper face. We consider design factors for choosing robust and usable tongue gestures, introduce eight gestures based on the criteria and discuss early work towards tongue gesture recognition with the system.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Gestural input**; *Interaction paradigms.*

KEYWORDS

hands-free; non-intrusive; tongue gestures; tongue interface; multimodal interface; BCI

ACM Reference Format:
Tan Gemicioglu, Mike Winters, Yu-Te Wang, and Ivan Tashev. 2022. Tongue Gestures for Hands-Free Interaction in Head Worn Displays. In *Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp/ISWC '22 Adjunct), September 11–15, 2022, Cambridge, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3544793.3560363

1 INTRODUCTION

Progressive motor diseases such as Amyotrophic Lateral Sclerosis (ALS) and muscular dystrophy, as well as quadriplegic paralysis due to stroke or spinal cord injury, often greatly reduce users' ability to move their hands voluntarily. Moreover, many situations where head worn displays (HWDs) are used like warehouses and manufacturing require users' hands to be busy during the task. Hands-free interaction methods like voice recognition and gaze tracking have been used in HWDs to allow their accessible use in different environments. However, voice recognition is not usable in many environments due to noise or privacy limitations. Gaze and dwell [4], the most common approach for gaze-based interaction in HWDs is slow and requires continuous attention. Tongue gestures offer an alternative to these two methods that is more private than voice and less attention demanding than gaze and dwell.

Many researchers have explored the tongue as an input paradigm using various devices and sensing modalities. Where early work on tongue interfaces depended on intrusive retainers placed in the mouth [9], later work has used electromyography (EMG) sensors around the mouth, jaw and cheeks [10]. Some interfaces have even succeeded in detecting motion remotely through external sensors by using the Doppler shift effect [5]. Their applications have also

Figure 1: Interface hardware: Reverb G2 Omnicept Edition VR headset, Omnicept Edition face gasket, and Muse 2 EEG headband.

---

[*] Work done as intern at Microsoft Research.

UbiComp/ISWC '22 Adjunct, September 11–15, 2022, Cambridge, United Kingdom

ACM ISBN 978-1-4503-9423-9/22/09.
https://doi.org/10.1145/3544793.3560363

Domain	Design Factor	Question
Usability	Accessibility	Which muscles are necessary? With how much force?
	Privacy	Is the gesture silent? Is it visible visually?
	Feedback	Does the gesture provide haptic feedback for itself?
	Simplicity	Is the gesture simple enough to memorize and perform regularly?
Recognizability	Signal strength	Is the signal strong enough to separate from noise?
	Distinguishability	Is the gesture distinct from other gestures?
	Rarity	Does the gesture occur naturally in daily activity?

Table 1: Design factors and questions to help evaluate each criteria.

Gesture Name	Description
Single Tap	Tap front upper teeth once with tongue
Double Tap	Tap front upper teeth twice in a row with tongue
Shake	Swing tongue left and right repeatedly
Left Cheek	Tap left cheek with tongue
Right Cheek	Tap right cheek with tongue
Mouth Floor	Touch bottom of mouth, behind lower teeth with tongue
Curl Back	Curl tongue up and towards the back of the palate
Bite	Gently bite on tongue with front teeth

Table 2: Eight selected gestures based on design factors.

ranged from human activity recognition [2] to [10] gesture recognition and silent speech [6]. However, these tongue interfaces require custom hardware and have unique form factors that make them challenging to use on a daily basis. Instead, sensors in head worn displays can provide the benefits of tongue gestures in a familiar form factor. Sensors in HWDs can be placed in the ears [1], or at other facial contact points such as the face gasket of VR headsets [8]. Recently, Kæseler et al. found that 72% accuracy could be obtained while only using single-trial electroencephalography (EEG) comparing rest vs. a single gesture [7], demonstrating a new sensor that could be integrated to HWDs for tongue sensing.

Building on prior work using different non-intrusive sensors for tongue gestures, we sought to develop a tongue gesture interface that does not require any additional components beyond existing commercial hardware. Here, we show early progress towards a non-intrusive tongue gesture interface embedded into a head worn display using EEG and inertial measurement units (IMUs). We elaborate design factors considered in designing and selecting tongue gestures to enable a new approach for hands-free interaction with HWDs.

2 DESIGN CONSIDERATIONS

Previously, Chen et al. mapped out the larger space of mouth microgestures, finding that mouth gestures should be short and simple with few movements [3], while different from natural movements of the mouth. Divided into usability and recognizability, our criteria include both common interaction design and signal processing factors. In addition, there are some specific criteria for tongue-based interaction, such as the ability of the gesture to provide haptic feedback for itself using the teeth or other surfaces of the mouth. Our design factors, and questions used to decide whether a gesture fits those factors are shown in more detail in Table 1.

We put these design factors to practice by selecting eight gestures based on the criteria in addition to two visible control gestures. The control gestures were selected from common facial gestures in literature that could also easily be distinguished during the experimental procedure for verification: "Blink", where the user blinks once; and "Stick Out", where the user sticks out their tongue and eight subtle gestures using tongue movement with the mouth closed. These two provide a point of comparison for researchers in eye tracking, brain-computer interfaces, and tongue interaction. A description of the other eight gestures, based on the design criteria can be found in Table 2.

The gestures were selected by iteratively choosing gestures that met the seven design factors without failing any during a brainstorming activity until a total of eight gestures were selected. To evaluate accessibility and simplicity, we used a physiological muscle diagram to remove gestures using muscles below the neck and examined muscle strain after doing the gesture 50 times repeatedly. For privacy, we checked if the gesture was audible or distinctly visible. For feedback, the gesture had to contact the teeth, palate or another surface around the mouth. For the signal strength and distinguishability, a real-time visual feed of the EEG sensors was used to determine whether the gesture produced a clear signal. Finally, rarity was required by excluding natural interactions such as eating and smiling. Additionally, the gestures needed to be performed in less than a few seconds while aiming to cover a range of spatial motions to allow directional control during multi-gesture use cases. An evaluation of the recognition accuracy, performance, and usability of these gestures will be presented in future work.

3 TONGUE GESTURE INTERFACE

We selected a VR headset with multimodal sensing capabilities to explore the use of different sensing modalities in recognizing tongue gestures. The HP G2 Reverb Omnicept Edition features eye tracking, a heart rate sensor, an inertia measurement unit (IMU) and a mouth camera. Many of these are contained within the face gasket of the VR headset as shown in Figure 1. We supplemented sensors on this device with the Muse 2 to address the lack of EEG sensing. The Muse 2 contains four main EEG channels and an amplified auxiliary channel, located in the forehead and behind the ears.

The data from the HP G2 Reverb can be streamed to Unity via OpenXR for regular VR position sensing and using the HP Omnicept software for the additional sensors. Then, a Unity interface streams the data from these sensors to the Lab Streaming Layer (LSL), a standardized timestamped data streaming protocol for brain-computer interfaces. For Muse 2, the open source BlueMuse software allows streaming the data into LSL directly. The data can then be saved to an XDF file for analysis or streamed directly for real-time recognition. The proposed data pipeline is shown in 2.

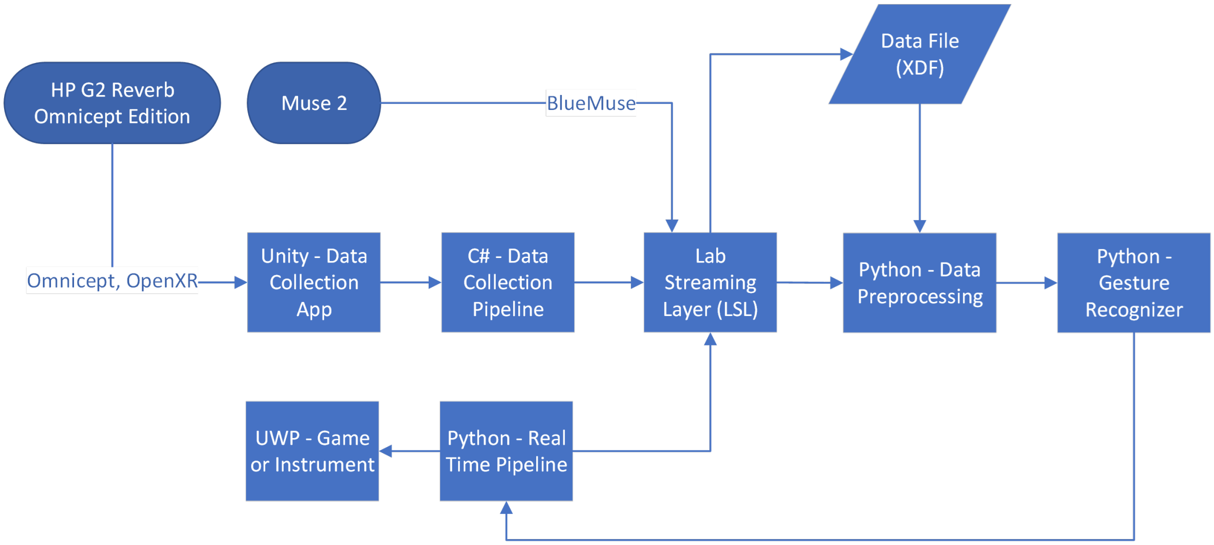

Figure 2: Pipeline for streaming and recording sensory data from HP Omnicept and BlueMuse with LSL.

4 CONCLUSION AND FUTURE WORK

We introduced tongue gestures as an alternative hands-free interaction method for head worn displays that can preserve user privacy without demanding intense visual attention. We evaluated factors important in designing tongue gestures that are usable in a wide range of settings and recognizable by classification algorithms. We also designed a set of gestures selected based on these factors. Finally, we showed a multimodal, multi-device system to evaluate the effectiveness of different sensors in distinguishing tongue gestures and proposed a pipeline to collect data from a prototype non-intrusive tongue gesture interface.

In a future paper, we plan to create a multimodal, multi-gesture dataset for enabling tongue gesture recognition. We plan to evaluate the efficacy of the gestures and interface proposed by studying the usability and building a machine learning classifier to recognize the gestures and different subsets of the gestures. We aim to use the interface in real-time to control different apps, games, and musical instruments in mixed reality.

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)*. Association for Computing Machinery, New York, NY, USA, 679–689. https://doi.org/10.1145/3126594.3126649

[2] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: Using Wearable Sensors to Detect Eating Episodes in Unconstrained Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (Sept. 2017), 1–20. https://doi.org/10.1145/3130902

[3] Victor Chen, Xuhai Xu, Richard Li, Yuanchun Shi, Shwetak Patel, and Yuntao Wang. 2021. Understanding the Design Space of Mouth Microgestures. In *Designing Interactive Systems Conference 2021 (DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1068–1081. https://doi.org/10.1145/3461778.3462004

[4] Morten Lund Dybdal, Javier San Agustin, and John Paulin Hansen. 2012. Gaze input for mobile devices by dwell and gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 225–228. https://doi.org/10.1145/2168556.2168601

[5] 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)*. Association for Computing Machinery, New York, NY, USA, 255–258. https://doi.org/10.1145/2702123.2702591

[6] Naoki Kimura, Tan Gemicioglu, Jonathan Womack, Richard Li, Yuhui Zhao, Abdelkareem Bedri, Zixiong Su, Alex Olwal, Jun Rekimoto, and Thad Starner. 2022. SilentSpeller: Towards mobile, hands-free, silent speech text entry using electropalatography. In *CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3491102.3502015

[7] Rasmus Leck Kæseler, Lotte N. S. Andreasen Struijk, and Mads Jochumsen. 2020. Detection and classification of tongue movements from single-trial EEG. In *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*. 376–379. https://doi.org/10.1109/BIBE50027.2020.00068 ISSN: 2471-7819.

[8] Richard Li and Gabriel Reyes. 2018. Buccal: low-cost cheek sensing for inferring continuous jaw motion in mobile virtual reality. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers (ISWC '18)*. Association for Computing Machinery, New York, NY, USA, 180–183. https://doi.org/10.1145/3267242.3267265

[9] 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)*. Association for Computing Machinery, New York, NY, USA, 177–180. https://doi.org/10.1145/1622176.1622209

[10] Qiao Zhang, Shyamnath Gollakota, Ben Taskar, and Raj P.N. Rao. 2014. Non-intrusive tongue machine interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2555–2558. https://doi.org/10.1145/2556288.2556981