A Mouth Gesture Interface Featuring a Mutual-Capacitance Sensor Embedded in a Surgical Mask

Yutaro Suzuki(✉), Kodai Sekimori, Yuki Yamato, Yusuke Yamasaki,
Buntarou Shizuki, and Shin Takahashi

University of Tsukuba, Tsukuba, Japan
{ysuzuki,sekimori,yamato,yusukeyamasaki,shizuki,
shin}@iplab.cs.tsukuba.ac.jp

Abstract. We developed a mouth gesture interface featuring a mutual-capacitance sensor embedded in a surgical mask. This wearable hands-free interface recognizes non-verbal mouth gestures; others cannot eavesdrop on anything the user does with the user's device. The mouth is hidden by the mask; others do not know what the user is doing. We confirm the feasibility of our approach and demonstrate the accuracy of mouth shape recognition. We present two applications. Mouth shape can be used to zoom in or out, or to select an application from a menu.

Keywords: Mouth gesture interface · Surgical mask · Mutual-capacitance sensor · Wearable device · Non-verbal input

1 Introduction

Touch is the most popular input to mobile devices. This requires one or both hands, which are not available in a crowded train or when holding luggage. Voice input is a useful alternative but may not work well in noisy public spaces [1,2]. Non-verbal mouth/tongue gestures are not affected by noise, but are nonetheless useful inputs. Prior studies have recognized mouth/tongue gestures using a smartphone camera [3], pressure sensors [4], and myoelectric potential sensors [5].

We developed a mouth gesture interface featuring a mutual-capacitance sensor embedded in a surgical mask. In East Asia, mask-type interfaces are acceptable; many people wear surgical masks on a daily basis. Our interface recognizes non-verbal mouth gestures. Thus, it is robust for acoustic noise, and others cannot eavesdrop on anything the user does with the user's device. The mask covers the user's mouth; others do not know what the user is doing. Here, we introduce the mouth gesture interface and its implementation. We evaluate mouth shape recognition accuracy and offer some useful applications.

M. Kurosu (Ed.): HCII 2020, LNCS 12182, pp. 154–165, 2020.
https://doi.org/10.1007/978-3-030-49062-1_10

2 Related Work

We employed mouth gesture interfaces featuring mutual-capacitance sensors to control mobile/wearable devices. We explored mouth shapes, hands-free inputs to mobile/wearable devices, and mutual-capacitance sensing.

2.1 Recognizing Mouth Shapes

Cameras recognize mouth shape and position. Azh et al. [6] operated a mobile device using camera-captured lip shapes. Lyons et al. [7] combined lip shapes and Japanese character inputs to control mobile devices. Vowels were obtained from lip shapes and consonants from keystrokes. Chan et al. [8] detected mouth shapes using a head-mounted camera; mouth movement and a hand-operated pen were used to draw pictures. Koguchi et al. [9] used touch-free lip shape inputs; a camera recognized shaped vowels.

In addition, some studies have exploited tongue movements. Miyauchi et al. [10] identified mouth regions by reference to Kinect depth and RGB data and evaluated tongue protrusion during training of children with Down's syndrome. Crawford et al. [11] identified mouth areas using a web camera and recorded tongue protrusions in real-time by reference to color and textural characteristics. Tongue movement has also been evaluated without a camera. Cheng et al. [4] used a fabric, pressure sensor array attached to the outside of the cheek to this end. Sasaki et al. [5] estimated tongue movement by measuring the myoelectrical potentials of multichannel electrodes attached to the lower jaw. Similarly, Zhang et al. [12] recorded tongue movements using six myoelectric potential sensors attached to the chin and two attached to the cheeks. Goel et al. [13] used a headset featuring three (forward, left, and right) X-band motion detectors to record tongue movement. Li et al. [14] placed three micro-radar (forward, left and right) sensors around the mouth to detect tongue movement using the Doppler effect.

2.2 Mutual-Capacitance Sensing

Capacitive sensing is important in the field of human-computer interaction (HCI) [15]; such sensing is employed by mobile, wearable, and stationary devices. Zimmerman et al. [16] used capacitive sensing to detect humans. Dietz et al. [17] developed the DiamondTouch system that simultaneously detects the touches and gestures of many people; the sensors are arrayed. Hinckely et al. [18] attached capacitive touch sensors to a mouse and a trackball. Rekimoto [19] developed the SmartSkin system for detecting changes in capacitance at multiple positions when the human body touched electrode meshes on a horizontal plane. Sato et al. [20] developed the Touché system that recognizes grip using only a single electrode; the impedance frequency characteristics change by the touching mode employed. Tsuruta et al. [21] developed a single-connection, RootCap capacitive sensor that was activated when multiple electrodes were touched. Wang et al. [22] distinguished the driver from a passenger when an in-vehicle screen was touched, exploiting the capacitances of sensors in the seats and the screen. No study has yet used capacitive sensing to identify mouth shape.

3 A Mouth Gesture Interface Featuring a Mask-Type Sensor

Figure 1 shows a schematic of our mouth gesture interface. The user wears a surgical mask featuring a mutual-capacitance sensor that recognizes mouth gestures and maps them to commands controlling applications. As the mouth is hidden by the mask, others do not know what the user is doing. For example, a user can unlock a smartphone without password leakage.

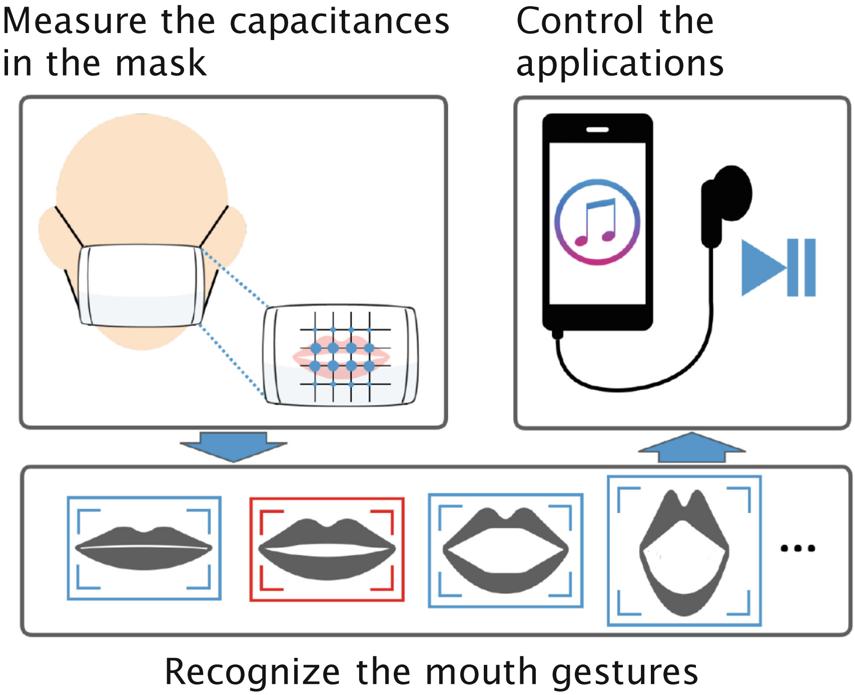
Measure the capacitances in the mask
Control the applications
...
Recognize the mouth gestures

Fig. 1. An overview of our approach. The mask sensor measures capacitances that correspond to user mouth gestures; these control the applications.

3.1 Mutual-Capacitance Sensing

We use mutual-capacitance sensing to recognize mouth shape. When a user moves the mouth, the capacitances of touched or approached intersections change; the interface thus recognizes mouth gestures.

A mutual-capacitance sensor features multiple intersecting electrodes [19,23]; those facing in one direction collectively serve as a transmitter delivering a sine wave and those facing in the opposite direction as the receiver. The transmitter and receiver create an electrical field. When an element approaches the intersections, that element interacts with the electrical field and the intersection capacitances change. The voltages at each receiver electrode thus also change.

3.2 Mouth Gestures

Mouth gestures can serve as interface inputs in many ways. One of the simplest mouth gesture sets includes only the open and closed mouth; these gestures are robustly recognized. The extent of mouth-opening (to which a numerical value may be assigned using a slider) may serve as an input. We use multiple mouth shapes as inputs. The gesture set contains six mouth shapes: ‘n’ (the neutral state) and the Japanese vowels ‘a,’ ‘i,’ ‘u,’ ‘e,’ and ‘o’ (Fig. 2). This allows simple character input triggering a command or operation. The five gestures other than ‘n’ can select and move a cursor up, down, left, or right; the gesture set can also execute an application that is pre-planned by sequentially changing mouth shape. For example, the sequence ‘a-e-a’ activates a smartphone camera (‘camera’ in Japanese is pronounced ‘ka-me-ra,’ thus with the vowel sequence ‘a-e-a’).

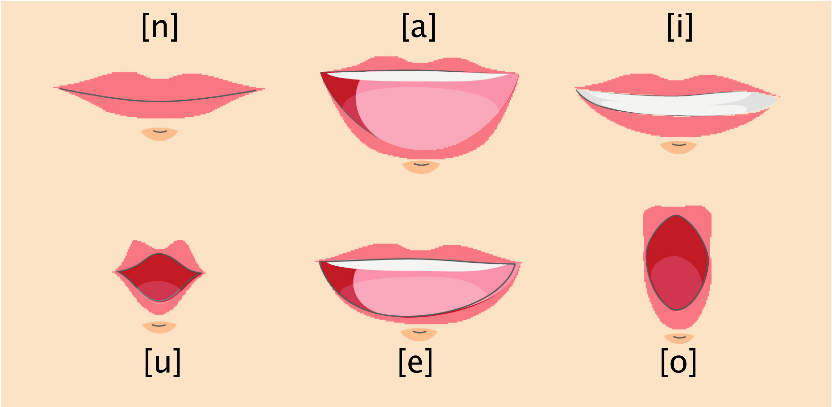

Fig. 2. The six mouth shapes. ‘n’ is the neutral state and ‘a,’ ‘i,’ ‘u,’ ‘e,’ and ‘o’ are Japanese vowels.

4 Implementation

We confirmed the feasibility of mouth gesture recognition via a mutual-capacitance sensor embedded in a surgical mask. The system features a sensing circuit and software (Fig. 3). In the sensing circuit, a sine wave is applied to the transmitter electrodes; then the voltages at the receiver electrodes are measured.

We used an Analog Discovery 2 instrument[1] to both apply the sine wave and measure voltages. The analyzer is connected to a laptop via a cable and sends defined voltages to the software. The laptop performs signal pre-processing and recognizes mouth gestures using a machine-learning algorithm.

Four horizontal wires serve as transmitter electrodes and five vertical wires serve as receiver electrodes; 49 1 cm^2 copper foils are attached to the wires (Fig. 4). The horizontal wires are connected to a waveform generator via a multiplexer and the vertical wires are connected to an A/D converter via another multiplexer. The generator and converter are included in the Analog Discovery 2 instrument. The wire intersection points are insulated using thin transparent tape. The sensor is covered with plastic wrap to prevent direct mouth contact. A sine wave of 10 $V_{\text{peak-to-peak}}$ is delivered at 100 kHz to the transmitter electrodes; the voltages of all intersections are sampled 1,000 times and the data sent to Python software.

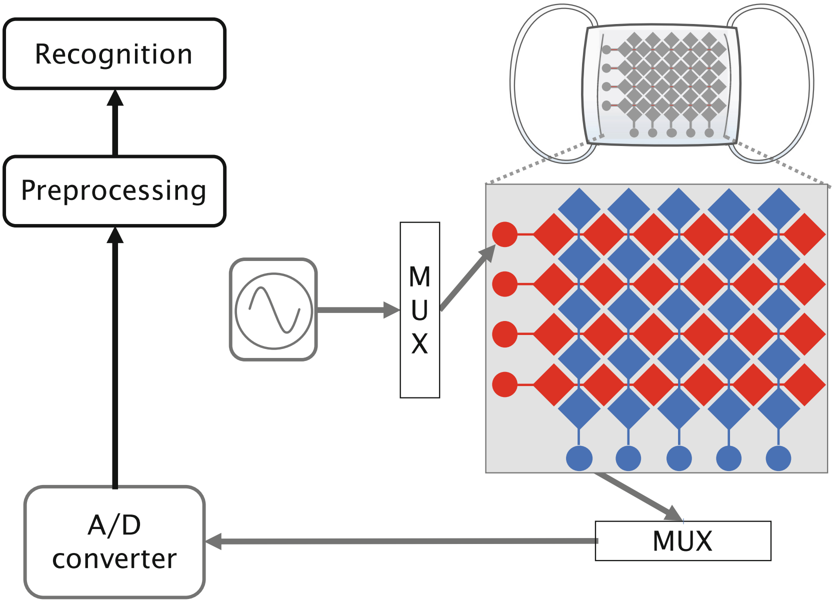

Fig. 3. Our implemented system. A mutual-capacitance sensor is embedded in the mask. The transmitter electrodes are shown in red and the receiver electrodes in blue. (Color figure online)

The signals are first pre-processed and then recognized. During preprocessing, the software removes noise using a band-pass filter that blocks signals

[1] https://reference.digilentinc.com/reference/instrumentation/analog-discovery-2/start.

of frequencies other than 90 to 110 kHz. We used the SciPy[2] "buttord" function to this end. Next, the software calculates the root mean square voltage (V_{rms}) at each intersection using the 1,000 sets of voltage values corresponding to capacitances. The V_{rms} values of all 20 intersections are considered a single frame; the process requires about 0.18 s. Recognition employs the Random Forest (RF) classifier of the scikit-learn library[3].

5 Preliminary Experiment

We performed a preliminary experiment to determine where the mouth touched the sensor and to evaluate the accuracy of mouth shape recognition. Three male volunteers (mean age 23.3 years) participated.

5.1 Procedure

We used the six mouth shapes shown in Fig. 2 (those made while mouthing the neutral 'n' and the five vowels 'a,' 'i,' 'u,' 'e,' and 'o'). As participants wore the mask, we told them to shape their mouths as instructed and to hold the shapes

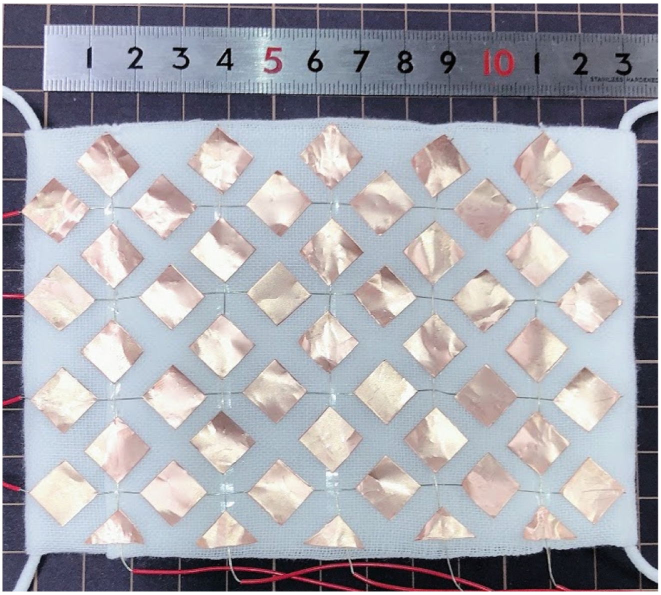

Fig. 4. The mutual-capacitance sensor used in the preliminary experiment. The horizontal electrodes served as transmitters and the vertical electrodes as receivers. Copper foil was cut into 1 cm^2 and placed on the wires at 2 cm intervals.

[2] https://docs.scipy.org/.

[3] https://scikit-learn.org/.

for about 3 s. We randomized the mouth shape order and acquired 20 frames for each shape. All participants completed five consecutive sessions each featuring all six mouth shapes. We thus acquired 600 frames [$5(sessions) \times 6(shapes) \times 20(frames)$] per participant.

5.2 Results and Analyses

Figure 5 shows heatmaps representing the sums of the $20(frames) \times 5(sessions)$ V_{rms} values at each intersection for the mouth shapes of Participant 1. The deeper blue points reflect higher V_{rms} values; the mouth often touched these points strongly.

We used principal component analysis (PCA) to calculate the contributions of all intersections (Fig. 6). The most significant points were (in order) [0, 2], [2, 2], [0, 3], [3, 3], [1, 2], and [3, 1]; points [1, 0], [1, 1], [1, 3], and [1, 4] made only small contributions. Thus, the central regions of the vertical wires well-captured mouth shapes; the edges of the wires did not. Thus, the edges of the vertical wires and the second horizontal wire from the top are redundant; their removal would accelerate data collection and implementation.

We trained the RF classifier to recognize mouth shapes. We randomly chose 80% of the data (480 frames) for training and used the remaining 20% (120 frames) to evaluate recognition accuracy; the average value was 99.2% (Table 1).

Then we performed *Leave-One-session-Out Cross-Validation* (*LOOCV*); the average recognition accuracy was 75.4% (Table 1), less than that of random validation. Changes in mask positions between sessions may explain the difference. We will collect more training data when the mask is worn in different positions.

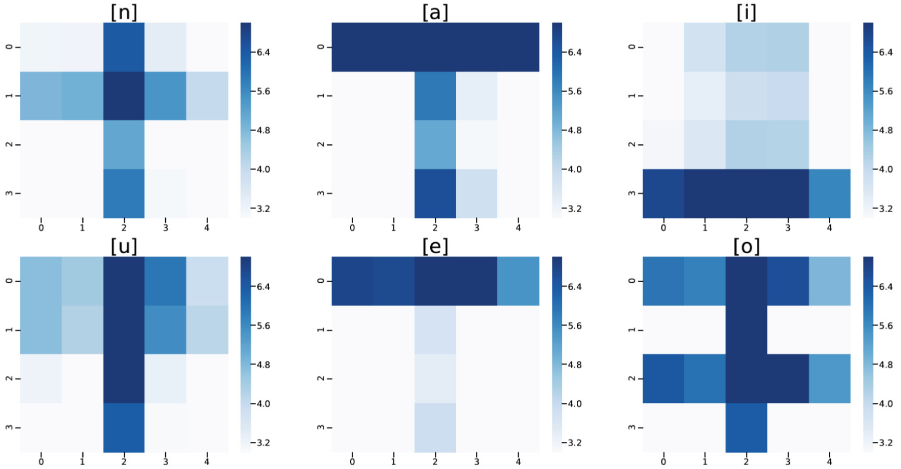

Fig. 5. Heatmaps showing the 100-frame V_{rms} sums at each intersection for each mouth shape of Participant 1. The darker blue points are associated with higher V_{rms} values; the mouth often touched these points strongly. (Color figure online)

Table 1. The recognition accuracies for each participant.

Participant	1	2	3	Average
Random (%)	99.17	98.33	100.0	99.17
LOOCV (%)	83.67	78.83	63.83	75.44

6 Application Examples

6.1 Zooming in or Out

Zooming in or out is very common; mobile devices employ pinch-in/-out systems. In our application, the user zooms in using the mouth shape ‘u’ and zooms out employing ‘o.’ In Fig. 7, the user mouths different vowels, but others do not know what he is doing.

6.2 Executing Commands

Application or command selection is often desirable. We used the mask-type interface to move through menus. We assigned the mouth shapes to commands. The user moves the cursor up by forming a ‘u,’ down by forming an ‘o,’ right by forming an ‘i,’ left by forming an ‘a,’ and selects the item using ‘e’ (Fig. 8). Thus, a desired application or command can be chosen in a hands-free manner.

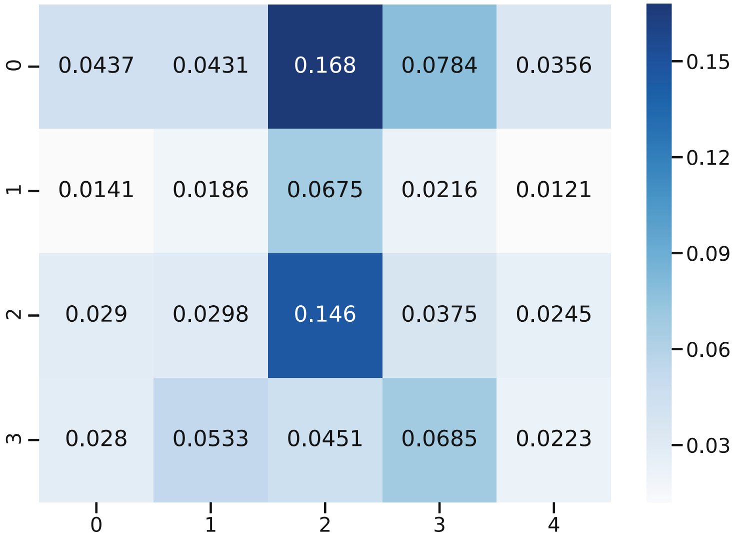

Fig. 6. The contribution ratio of each PCA intersection to the mouth shapes of Participant 1.

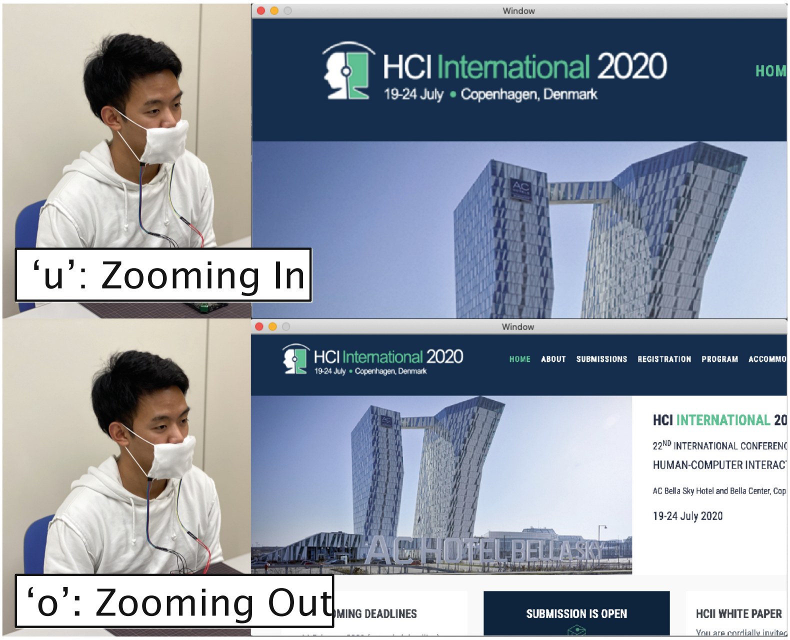
Window
HCI International 2020
19-24 July • Copenhagen, Denmark
HOM
'u': Zooming In
Window
HCI International 2020
19-24 July • Copenhagen, Denmark
HOME ABOUT SUBMISSIONS REGISTRATION PROGRAM ACCOMMO
HCI INTERNATIONAL 20
22ND INTERNATIONAL CONFERENC
HUMAN-COMPUTER INTERAC
AC Bella Sky Hotel and Bella Center, Cop
19-24 July 2020
'o': Zooming Out
MING DEADLINES
SUBMISSION IS OPEN
HCII WHITE PAPER
You are cordially invited

Fig. 7. A zooming application. The user zooms in using the mouth shape 'u' and zooms out employing 'o.'

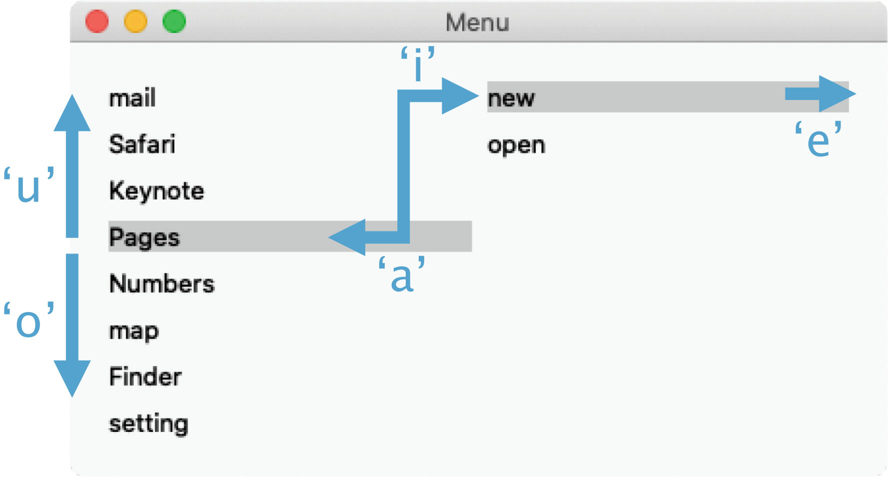

Fig. 8. Command selection. The user moves the cursor up by forming a 'u,' down by forming an 'o,' right by forming an 'i,' left by forming an 'a,' and selects the item using 'e.'

6.3 Preliminary Evaluation of the Applications

Zooming three-class classification worked well in terms of command execution; six-class classification did not. Real-time recognition accuracy must be improved. The vowel mouth shapes are not completely different: ‘a’ and ‘e’ are similar. Subtle variation in mouth shape caused by changes in facial expression and breathing patterns triggered misrecognition even when the mouth shape was correct. We will collect more data. We plan to use a sliding window to stabilize real-time classification, and to identify mouth shapes that facilitate robust classification.

7 Conclusion and Future Work

We developed a mouth gesture interface; a mutual-capacitance sensor is embedded in a surgical mask. This wearable hands-free interface recognizes non-verbal mouth gestures; others cannot eavesdrop on anything the user does with the user’s device. As the mouth is hidden by the mask, others do not know what the user is doing. We confirmed the feasibility of our approach, and showed that the mutual-capacitance sensor performed well. We explored two applications. Mouth shape was used to zoom in and out and to select from a menu of applications. In the future, we will develop a low-cost, disposable mutual-capacitance sensor. We will also replace the copper foils and wires with conductive cloth and threads. Real-time recognition accuracy will be improved. We will also define a recognizable set of mouth shapes.

References

1. Fukumoto, M.: SilentVoice: unnoticeable voice input by ingressive speech. In: Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology, UIST 2018, pp. 237–246. ACM, New York (2018)
2. Ronkainen, S., Häkkilä, J., Kaleva, S., Colley, A., Linjama, J.: Tap input as an embedded interaction method for mobile devices. In: Proceedings of the 1st International Conference on Tangible and Embedded Interaction, TEI 2007, pp. 263–270. ACM, New York (2007)
3. Sun, K., Yu, C., Shi, W., Liu, L., Shi, Y.: Lip-Interact: improving mobile device interaction with silent speech commands. In: Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology, UIST 2018, pp. 581–593. ACM, New York (2018)
4. Cheng, J., et al.: On the tip of my tongue: a non-invasive pressure-based tongue interface. In: Proceedings of the 5th Augmented Human International Conference, AH 2014, pp. 12:1–12:4. ACM, New York (2014)
5. Sasaki, M., et al.: Tongue interface based on surface EMG signals of suprahyoid muscles. ROBOMECH J. **3** (2016). Article number: 9. https://doi.org/10.1186/s40648-016-0048-0
6. Azh, M., Zhao, S.: LUI: lip in multimodal mobile GUI interaction. In: Proceedings of the 14th ACM International Conference on Multimodal Interaction, ICMI 2012, pp. 551–554. ACM, New York (2012)

7. Lyons, M.J., Chan, C.-H., Tetsutani, N.: MouthType: text entry by hand and mouth. In: CHI 2004 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2004, pp. 1383–1386. ACM, New York (2004)
8. Chan, C., Lyons, M.J., Tetsutani, N.: Mouthbrush: drawing and painting by hand and mouth. In: Proceedings of the 5th International Conference on Multimodal Interfaces, ICMI 2003, pp. 277–280. ACM, New York (2003)
9. Koguchi, Y., Oharada, K., Takagi, Y., Sawada, Y., Shizuki, B., Takahashi, S.: A mobile command input through vowel lip shape recognition. In: Kurosu, M. (ed.) HCI 2018. LNCS, vol. 10903, pp. 297–305. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91250-9_23
10. Miyauchi, M., Kimura, T., Nojima, T.: A tongue training system for children with down syndrome. In: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, UIST 2013, pp. 373–376. ACM, New York (2013)
11. Crawford, C.S., Bailey, S.W., Badea, C., Gilbert, J.E.: Using Cr-Y components to detect tongue protrusion gestures. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA 2015, pp. 1331–1336. ACM, New York (2015)
12. Zhang, Q., Gollakota, S., Taskar, B., Rao, R.P.N.: Non-intrusive tongue machine interface. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2014, pp. 2555–2558. ACM, New York (2014)
13. Goel, M., Zhao, C., Vinisha, R., Patel, S.N.: 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 2015, pp. 255–258. ACM, New York (2015)
14. Li, Z., Robucci, R., Banerjee, N., Patel, C.: Tongue-n-cheek: non-contact tongue gesture recognition. In: Proceedings of the 14th International Conference on Information Processing in Sensor Networks, IPSN 2015, pp. 95–105. ACM, New York (2015)
15. Grosse-Puppendahl, T., et al.: Finding common ground: a survey of capacitive sensing in human-computer interaction. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI 2017, pp. 3293–3315. ACM, New York (2017)
16. Zimmerman, T.G., Smith, J.R., Paradiso, J.A., Allport, D., Gershenfeld, N.: Applying electric field sensing to human-computer interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1995, pp. 280–287. ACM Press/Addison-Wesley Publishing Co., New York (1995)
17. Dietz, P., Leigh, D.: DiamondTouch: a multi-user touch technology. In: Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology, UIST 2001, pp. 219–226. ACM, New York (2001)
18. Hinckley, K., Pausch, R., Goble, J.C., Kassell, N.F.: Passive real-world interface props for neurosurgical visualization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1994, pp. 452–458. ACM, New York (1994)
19. Rekimoto, J.: SmartSkin: an infrastructure for freehand manipulation on interactive surfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2002, pp. 113–120. ACM, New York (2002)
20. Sato, M., Poupyrev, I., Harrison, C.: Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2012, pp. 483–492. ACM, New York (2012)

21. Tsuruta, M., Nakamae, S., Shizuki, B.: RootCap: touch detection on multi-electrodes using single-line connected capacitive sensing. In: Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces, ISS 2016, pp. 23–32. ACM, New York (2016)
22. Wang, E.J., Garrison, J., Whitmire, E., Goel, M., Patel, S.: Carpacio: repurposing capacitive sensors to distinguish driver and passenger touches on in-vehicle screens. In: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, UIST 2017, pp. 49–55. ACM, New York (2017)
23. Pourjafarian, N., Withana, A., Paradiso, J.A., Steimle, J.: Multi-touch kit: a do-it-yourself technique for capacitive multi-touch sensing using a commodity microcontroller. In: Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, UIST 2019, pp. 1071–1083. ACM, New York (2019)