DUMask: A Discrete and Unobtrusive Mask-Based Interface for Facial Gestures

Arpit Bhatia*
arpit16229@iiitd.ac.in
Weave Lab, IIIT-Delhi
New-Delhi, Delhi, India

Aryan Saini*
aryan@exertiongameslab.org
Exertion Games Lab, Department of Human-Centred Computing, Monash University
Melbourne, Australia
Weave Lab, IIIT-Delhi
New-Delhi, Delhi, India

Isha Kalra
isha16152@iiitd.ac.in
Weave Lab, IIIT-Delhi
New-Delhi, Delhi, India

Manideepa Mukherjee
manideepam@iiitd.ac.in
Weave Lab, IIIT-Delhi
New-Delhi, Delhi, India

Aman Parnami
aman@iiitd.ac.in
Weave Lab, IIIT-Delhi
New-Delhi, Delhi, India

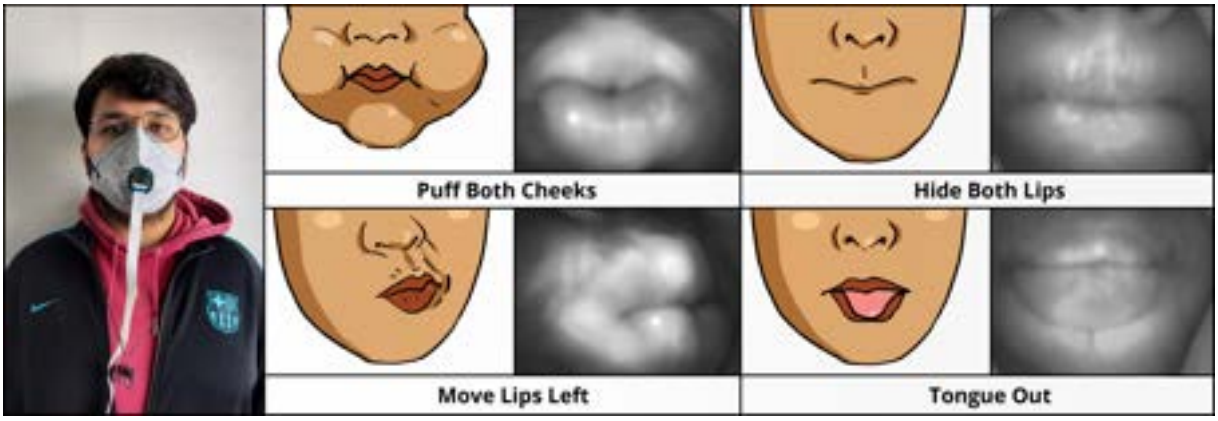

Figure 1: The DUmask interface along with a subset of the gestures it enables. How each gesture is being performed under the mask (left) and how it is captured by our camera (right) are both displayed.

ABSTRACT

Interactions using the face, not only enable multi-tasking but also enable us to create hands-free applications. Previous works in HCI used sensors attached directly to the person's face or inside their mouth. However, a mask, which has now become a norm in our everyday life and is socially acceptable, has rarely been used to explore facial interactions. We designed, "DUMask", an interface that uses face parts covered by a mask to discretely enable 14 (+1 default) interactions. DUMask uses an infrared camera embedded inside an off-the-shelf face mask to recognize the gestures, and we demonstrate the effectiveness of our interface through in-lab studies. We conducted two user studies evaluating the experience of both the wearer and the onlooker, which validated that the interface is indeed inconspicuous and unobtrusive.

CCS CONCEPTS

• **Human-centered computing** → **Interaction devices**; *Gestural input*.

KEYWORDS

Mask, Facial Gestures, Wearables

ACM Reference Format:
Arpit Bhatia, Aryan Saini, Isha Kalra, Manideepa Mukherjee, and Aman Parnami. 2023. DUMask: A Discrete and Unobtrusive Mask-Based Interface for Facial Gestures. In *Augmented Humans Conference (AHs '23), March 12–14, 2023, Glasgow, United Kingdom.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3582700.3582726

*Both authors contributed equally to the paper

AHs '23, March 12–14, 2023, Glasgow, United Kingdom

ACM ISBN 978-1-4503-9984-5/23/03…$15.00
https://doi.org/10.1145/3582700.3582726

1 INTRODUCTION

Exploration and facilitation of facial interactions is a prevalent topic in recent HCI literature. Facial interactions are usually attributed to improving multi-tasking efficiency [49] and enabling hands-free applications [8]. The latter has been extensively explored in building assistive technologies for people with physical impairments [1, 24].

While multiple interfaces have been proposed to leverage the tongue, lips, and cheeks for enabling facial interactions, they suffer from various issues. For example, interfaces attached to the face are usually bulky [15] and/or conspicuous [22, 57]. Non-attached interfaces usually require augmentation of a specific body part [19] or are invasive [13, 45]. Performing facial interactions in public is also of concern as making faces or sticking the tongue out may be perceived as rude, disgusting, playful, or sexually provocative.

To overcome these issues, we propose a wearable, face mask-based interface, DUMask. While this work was initially inspired by the increasing social acceptability of face masks due to rapidly rising air pollution levels in countries like China, India, Australia, and Pakistan; the recent COVID-19 pandemic has made masks even more commonplace and acceptable, sometimes even required by law. Furthermore, masks have also been treated as fashion accessories with designers creating unique design patterns on their surface [9, 12]. DUMask conceals electronics within the mask, allowing it to create a non-intrusive way to capture facial gestures (tongue, lips, and cheeks) as input. The form factor of the mask completely occludes the mouth which hides gestures that may seem unnatural or awkward to onlookers. Our work makes the following contributions:

- We present a novel interface that recognizes tongue, lip, and cheek gestures in a form factor that is socially acceptable, discreet as well as non-intrusive
- We also present the results and insights gained from a comprehensive evaluation of the usability of our interface (from the wearer's perspective) and its social acceptability (from an observer's perspective).

2 RELATED WORK

Previous research focused on detecting facial gestures either by integrating a sensor inside the mouth leveraging the tongue [35] and oral muscle movement [13], or through a wearable device on the head leveraging facial muscle movement [15]. With face masks becoming increasingly commonplace these days, there is some recent research on augmenting them for different purposes as well. In this section, we discuss research relevant to Oral Input Interfaces, Facial Input Interfaces, and Mask-based Wearables.

2.1 Oral Input Interfaces

Gesture recognition systems to capture oral activity have been explored widely for mobile [39, 54] and personal smart devices [16, 40]. An on-body gesture sensing technique using acoustic interferometry was described by Iravantchi et al. [20], which can identify eleven hand and nine face gestures. While touch-based gestures have been accepted for ubiquitous devices, in-the-air gestures have not been widely integrated into commercial systems due to the uncertainty about their social acceptability [4]. In ChewIt [13], researchers showed an intra-oral interface for discrete interactions which was carefully designed to not be obtrusive and can detect several tongue-based gestures such as flipping, biting, etc. A sensor embedded inside artificial teeth was designed by Li et al. [32] which uses a small accelerometer for detecting oral activities such as chewing, drinking, speaking, and coughing. Another teeth-based interaction system called TYTH [42] was developed which is based on the location of the tongue with respect to the teeth. Here, the tongue act as a finger and the teeth as a keyboard. By sensing the brain and muscle signals from behind the user's ear along with capturing the skin surface deformation caused by tongue movement, it detects the interaction between the tongue and teeth and enables typing. In TongueBoard [35] researchers placed a capacitive sensor in the roof of the mouth for recognizing non-vocalizing speech. Though these systems show promising results and claim to create minimal discomfort to the users, placing a wearable inside the mouth may need some habit formation time. Moreover, the characteristics of oral features such as the tongue and the speaking style may vary on a per-user basis, thus designing for a wider scale and comfort may be a challenge. Therefore in DUMask, we augment a face mask, which has been increasingly ubiquitous, to capture oral gestures without intruding on the oral cavity.

2.2 Facial Input Interfaces

Similar to oral activity, facial gestures have been leveraged extensively to facilitate hands-free interaction. Matthies, et al. also underline the opportunity of leveraging facial expression in their survey about wearable sensing of facial expressions [38]. Specifically, a tongue-based gesture detection system was developed by Goel et al. [15], which uses Doppler radar units around the user's face to identify gestures. Although the system achieves an acceptable gesture identification accuracy, its social acceptability of it is of concern due to its visual prominence. Electromyography (EMG) signals have been previously used in interfaces to sense tongue, mouth, and cheek gestures but they require prominent electrodes to be stuck on a user's face. An AI-enabled silent speech headset was developed by Kapur et al. [41, 57] to converse with a computing device silently. The device is an ambulatory wearable system that connects through Bluetooth with the computing device. Lukaszewicz has used ultrasound images [36] to recognize the selected region of the tongue surface for speech synthesis. Although the author concluded that the position of the tongue surface is not sufficient for directly steering speech on the level of phonemes, they showed that this can be used to create simple oral input to control different devices. A sound-based input recognition system was designed by Ashbrook et al. to detect tooth clicks [8]. It uses a bone conduction microphone to recognize the tooth click sounds from five different pairs of teeth. In Lip-Interact [26], Sun et al. leverage the front camera of a smartphone to propose silent-speech input and also evaluate the social acceptability of their system compared to voice commands. Mugeetion [27] describes a facial gesture-based music-playing system where the authors have shown how the music played changes with the facial expression and emotion of the user.

Facial expression and gesture identification has also been explored by sensing ear [31], jaw [55] and facial muscle [43] movements. A jaw, face, and head movement recognition system called CanalSense has been developed by Ando et al. [7] which uses

barometer-based earphones to recognize the air pressure change inside the ear canal. Amesaka et al. designed a facial expression detection system using an ear canal transfer function to detect twenty-one facial expressions in [5]. Buccal [34] augments a Mobile VR headset with five IR proximity sensors to recognize lip and jaw movement by measuring the deformation of cheeks and temples. Similar work was presented in CheekInput [56] where optical sensors were used on a Head-mounted Display to measure the deformation of the skin for sensing touch gestures. Besides gestures, previous works have also explored tracking facial movements in the context of food intake monitoring [50].

While these interfaces successfully capture facial gestures to create novel methods of interaction, they struggle with social acceptability and/or require instrumentation. Future developments in technology may allow these interfaces to improve their social acceptability and reduce their obtrusiveness, but our focus with this work is to leverage the omnipresent face mask to design a system that is socially acceptable and relatively easy to set up.

2.3 Mask-Based Wearables

While air pollution has been a major stakeholder in the promotion of face masks, COVID-19 contributed to them becoming a mandatory wearable in most parts of the world, eventually presenting an opportunity for innovation. As a response to the COVID-19 outbreak, researchers augmented face masks with sticker strips to gauge a user's exposure to the virus throughout the day [21]. Genç et al. [14] have added electrochromic displays at the center of regular face masks to display a visualization representing the wearer's expression. This idea is aimed at compensating for the lack of facial expressions while communicating with other people. TransEmotion [29] is a full-color display in the form factor of a mask that replaces the lower half of the wearer's face with a photo-realistic virtual face. This display can be used to help people who have difficulties in making appropriate expressions based on a given situation. Adhikary et al. [3] have created a prototype that embeds a microphone and a Carbon Monoxide sensor inside a surgical mask to monitor lung health and ambient air quality respectively. Adhikary, et al. recently extended this work in SpiroMask, by augmenting consumer-grade face masks with a microphone for continuous lung health monitoring through spirometry [2]. In "Giving Up Control" [47], researchers explored the implications of a face mask that only opens and closes automatically based on pollution levels and does not allow the user any control over their exposure to pollution. Similarly, with FaceBit [10], researchers have also explored augmenting face masks for heart and respiration rate monitoring along with several other health metrics in an energy-efficient system for health management and frontline workers.

Takagi, et al. [48] developed a horizontal projection system for projecting lip animation onto the user's face masks for better communication during physical meetups. Kashino, et al. [23] further explored a new normal for face-to-face interactions by augmenting a user with AR markers to facilitate a virtual full-face mask while envisioning a future with augmented vision. Further, Kunimi, et al, created E-MASK by using flexible and highly sensitive strain sensors for silent speech interaction with 21 Alexa operations [30]. Additionally, with a premise for reducing stress, Xie, et al. [53], created custom face masks for encouraging as well as gauging eye contact in children with ASD to reduce stress while making eye contact. Masui [37] also proposed a closed-loop dynamic facial expression augmentation method for creating masks for theatrical performances by using thermochromic ink that leverages temperature change to change colors while the artist is engaged in a performance.

While these existing works combine face masks with technology, none of them aims to leverage this form factor specifically for novel gestural input nor discuss the advantages of doing so. DUMask augments a face mask to present a novel interface that enables oral and facial interaction.

3 DUMASK

In this section, we first describe the design criteria of DUMask before discussing its implementation and the gestures it enables.

3.1 Design Criteria

The aim of building DUMask is to create an interface to enable facial gestures while mitigating issues identified in previously explored form factors. An additional hurdle that we wish to overcome is the awkward seeming appearance of facial gestures which has limited their use while out in public. Below we describe the design guidelines we enforced while creating DUMask to achieve these goals.

- *Discreet Interactions:* Interacting with DUMask should not attract attention from onlookers. Due to this, audio-based interactions such as tooth clicks and tongue flicks are out of the scope of this work. The interface should actually obscure the awkward seeming gestures from onlookers.
- *Discrete Interactions:* A person wearing a mask will perform many compound facial movements such as while talking, laughing, or licking their lips. Our gestures must be distinct from such actions and hence DUMask focuses on detecting short simple events which are not generally performed in daily life.
- *Non-Contact Sensing:* Besides parts of the mask touching the wearer's face, there should not be any additional contact to make the system as noninvasive as possible.
- *Maintain Protection:* The core functionality of the mask is to protect the wearer from inhaling airborne microbes and particulate matter. Our alterations must not compromise this functionality.

3.2 Setup

Our setup consists of a standard N95 mask with a respirator and fasteners attached to the strings allowing the users to make adjustments to the mask according to their comfort. The electronics are mounted on top of the mask by cutting a hole of 1.5 cm radius. The hole is patched up with a cloth that maintains the mask's protection. The weight of the whole setup is 15.75g, including the sensor and LEDs, while regular N95 masks usually weigh around 10g. While this is more than a 0.5 times increase, wearing DUMask was rated above average when it comes to comfort in our user study (Section 5.2), and this new weight is also along the lines of the weight when

double masking with a surgical and N95 mask (10g + 4g), which has been a prevalent practice during some parts of the pandemic.

Figure 2: DUMask Prototype

Our setup employs a Pi NOIR camera v2[1] with a Sony IMX219 8-megapixel sensor and a 73.8-degree wide field-of-view. Owing to the absence of an IR filter in the NoIR camera, we were able to view inside the mask in a low/no-light setting with the help of IR illumination provided by an IR LED attachment[2]. We were able to capture infrared images of the view inside the mask at a resolution of 1024 x 768 pixels. The IR illuminator is configured to be switched on at all times to provide uniform illumination across the samples captured in different lighting settings. Since the LEDs operate in the IR spectrum (880 nm), the illumination is not visible to either the user or an onlooker through naked eyes. The camera communicates with a standalone Raspberry Pi 3 Model B[3] via a 15-pin ribbon cable attached to the Pi's camera port. DUMask is powered by a 5V, 10000 mAh battery bank via a USB cable and consumes approximately 350 mA current on average when operational. The IR camera setup allows us to capture images without the presence of visible light which would have been extremely difficult with a generic RGB camera operating in the visible spectrum. Further, in this case, adding extra visible lighting inside the mask to capture clear images would illuminate the inside of the mask making it very noticeable causing DUMask's claim of being discreet to no longer be valid.

While we repurposed an existing N95 mask when creating DUMask, we only use it as a representative for face masks. We acknowledge that DUMask cannot achieve the same level of protection an N95 mask does due to features such as electrostatically charged mask fibers not being possible. However, our system can still provide the level of protection a regular cloth mask does, similar to other examples of masks with electronics inside them such as the AirPop Active+ Halo Smart Mask [4], Razer's Project Hazel [5] and the MASKFONE [6].

DUMask classifies 14 different gestures (Figure 4) based on the movement of three parts of the face which are covered by a mask: the tongue (T), cheeks (C), and lips (L). As mentioned in our design goals, we skip gestures involving jaw movements to keep our gesture set distinct from everyday facial movements made while talking. We consider the degrees of freedom available for translation (Figure 3) for these parts which leads us to the following gesture set:

[1]https://www.raspberrypi.org/products/pi-noir-camera-v2/
[2]https://elementztechblog.wordpress.com/2017/09/20/ir-illuminator-for-raspberry-pi-noir-camera/
[3]raspberrypi.org/products/raspberry-pi-3-model-b/
[4]https://www.airpophealth.com/us/airpop-active-smart-black-yellow
[5]https://www.razer.com/concepts/razer-project-hazel
[6]https://maskfone.com/

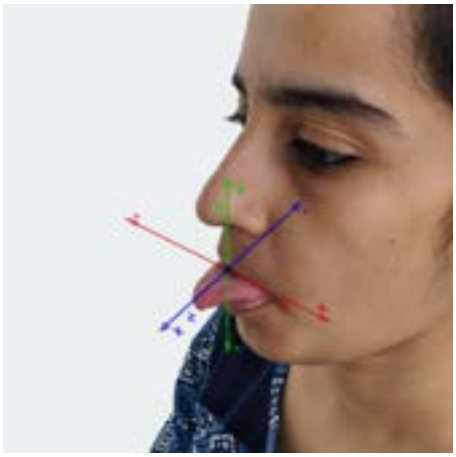

Figure 3: Defining facial gestures in the three-dimensional space

(1) Move lips left, MLL (L:+x)
(2) Move lips right, MLR (L:−x)
(3) Hide upper lip, HUL (L:+y)
(4) Hide lower lip, HLL (L:−y)
(5) Hide both lips, HBL (L:−z)
(6) Pout, P (L:+z and Left C:−x, Right C:+x)
(7) Puff both cheeks, PF (Left C:+x, Right C:−x)
(8) Puff left cheek, PL (C:+x)
(9) Puff right cheek, PR (C:−x)
(10) Tongue-left, TL (T:+x)
(11) Tongue-right, TR (T:−x)
(12) Tongue-up, TU (T:+y)
(13) Tongue-down, TD (T:−y)
(14) Tongue-out, TG: (T:+z)

3.3 Gesture Set

The same set of 5 tongue gestures has also been the standard used in previous works [15, 57] while the cheek gestures have been explored before in the form of sip and puff interfaces [1]. While the dexterity of the tongue also allows for a Tongue-In (T: -z) gesture by retracting the tongue further inside the mouth, this often led to the tongue being obscured or not properly visible due to lack of light inside the mouth. This makes it indistinguishable from a normal no-gesture state and hence we do not include it in our gesture set. Pushing inside the mouth with the tongue was another subset of gestures we considered but we had issues with distinguishing these gestures from the puff cheek gestures and hence do not include them in the final DUMask gesture set.

4 GESTURE RECOGNITION

Our goal in this section is to prove the technical feasibility of detecting oral gestures a user performs by recognizing the IR images of the lower half of their face. Our focus is hence not to find the most optimized model to recognize the gestures but to demonstrate high

Figure 4: The set of gestures enabled by DUMask

performance using generic machine learning models trained on image features. The input features we consider are Flattened image, HOG descriptor and LBP descriptor. The classifiers we consider are Support Vector Machines, Decision Trees, and Random Forests. In addition, we also consider a standard pre-trained CNN model, AlexNet, which was proposed by Krizhevsky et al. [28] with raw image inputs for our experiments. Parameters and intuition for each of the algorithms can be found in the supplemental materials. We use sklearn, skimage, openCV and PyTorch python libraries for the implementations of these algorithms and feature extractors. For the purpose of our experiments, we work with a set of 15 gesture classes: 14 tongue, cheek, and lip gestures defined in Gesture Set along with a "no gesture" or rest state.

We consider a single session as the period in which the user wears the mask and performs the gestures without making any major changes to the positioning of the mask. Whenever the user wears the mask for a new session, there are changes in the position of the mask relative to the user's mouth which lead to affine transformations (scaling, rotation, shearing, and translation) in different sessions. Due to these differences in inter-session data, we also consider how our interface performs on data from unseen sessions.

4.1 Data Collection Methodology

For collecting the data, the participants perform each gesture with breaks. The gesture images are manually captured by an assistant researcher after a two-second interval, in which the user must return to the no gesture state and perform the ongoing gesture again. This is done to capture as much variance as possible in the gesture images of the subject. Multiple gesture performances, coupled with pauses account for real-world differences in the participant's expressions. An alternative approach would be to capture multiple frames in a few seconds such as in a video recording, however, we do not take that approach as the frames would be very closely related and would prove to be redundant information. The participants were asked to notify the researcher by raising their thumbs when they are performing the gesture. Although the user's mouth was covered by a mask, a real-time feed of the camera was also available to be used at their discretion.

As our setup involves augmenting a face mask that too at the time of a pandemic, safety and hygiene was major concern while collecting data from the participants. Each subject was given a new mask to perform the studies, which were duly approved by our institute's IRB. As a precautionary measure, the researcher creating the setup actively sanitized their hands while wearing a face mask and gloves themselves. Social distancing was maintained and a sanitizer and disinfectant were kept near the participants to be used at will.

4.2 Baseline

Data Collection. We recruited 12 (6F/6M) participants aged between 19 and 41 for data collection for a session each. For each subject, a total of 15 images are captured for each of the 15 gestures, thus giving us a total of 2700 (12*15*15) images. For each of the 15 gestures, we have 180 images.

Experiment. For our baseline experiment, we train classifiers with data from multiple users collected over a single session. Our goal is to train the classifier with some images of each pose or gesture and then accurately predict the pose in unseen test images. For our traditional machine learning algorithms (SVM, DT, RF), we use 12 of the 15 images per pose per subject as the training images and the remaining 3 images as test images. For each pose, we have a total of 144 training images and 36 testing images. For these algorithms, we resize the images from their original resolution of 1024*768 pixels to 512*384 pixels. i.e, scaled by 50% along each dimension. We also convert the image from RGB to grayscale i.e. 3-channel to a single channel. We observed minimal to no loss in recognition accuracies due to this scaling, and it considerably reduced the training and testing durations. We use an 80-20 train-test split and perform 10-fold cross-validation. For our CNN algorithm, for each pose, we use 11 training images, 1 image for validation, and 3 test images. We resize the images to 224*224 pixels and retain the RGB channel information.

Results. Table 1 presents the accuracy of the baseline experiments. It is observed that Random Forests, with flattened images as input, provide the highest accuracy of 98.7%, with pre-trained

Table 1: Gesture Recognition Accuracy for the Baseline Experiment

Features	Classifier		
	SVM	DT	RF
Flattened Image	0.9486	0.7722	**0.9870**
Histogram of Oriented Gradients (HOG)	0.8842	0.6805	0.9638
Local Binary Pattern (LBP)	0.7464	0.4402	0.8333

Transfer learning based AlexNet	
Training Accuracy	0.9696
Validation Accuracy	0.9166
Testing Accuracy	**0.9722**

Table 2: Gesture Recognition Accuracy for Session Dependence (15 training images per pose)

Participant	Random Forest+Flattened Image		
	Minimum	Maximum	Average
Participant 1	0.6266	0.9911	0.8818
Participant 2	0.5733	0.9955	0.8274
Participant 3	0.6533	0.8888	0.8133
			0.8408
Participant	**Pre-trained AlexNet**		
	Minimum	**Maximum**	**Average**
Participant 1	0.6888	0.9624	0.8464
Participant 2	0.4572	0.9214	0.7471
Participant 3	0.5884	0.8268	0.7644
			0.7859

AlexNet close behind at 97.2%. In terms of practical applicability, this experiment verifies that for a single session, given 11 or 12 images per gesture for a subject, we can nearly perfectly recognize the gesture being performed by them in a testing environment.

4.3 Session Dependence

Data Collection. We recruited 3 (1F/2M) participants aged 22 (F), 22, and 35 for data collection over multiple sessions. For each subject, we collect data for 12 different sessions, spread across multiple days. In each session, 15 images are captured for each gesture, with a 2-second gap in between, wherein the subject relaxes their mouth before resuming the gesture.

Experiment. For each of the 3 participants, we perform leave-one-out experiments, with data from 11 sessions as training data and data from the 12th session is used as test data. We perform 12-fold cross-validation and our final goal is to evaluate the average leave-one-out accuracy over data from multiple sessions. If the model performs well on unseen test sessions, we can claim that our model is session-independent. For our algorithms, we use the Random Forest and AlexNet which had the highest accuracies in our previous experiment.

Results. Table 2 presents the accuracy of the session dependence experiment. We evaluated the average leave-one-out accuracy for the protocol and it comes out to be 84.08% for Random Forest+Flattened Image, and 78.59% for pre-trained AlexNet. These

Table 3: Gesture Recognition Accuracy for Session Independence (5 training images per pose)

Participant	Accuracy (5 training images per pose)		
	Minimum	**Maximum**	**Average**
Participant 1	0.6533	1.0000	0.8533
Participant2	0.5066	0.9600	0.7755
Participant 3	0.6533	0.8667	0.7744
			0.8010

models can be considered to be session independent and their performances are likely to improve if more session data is available. To build a session-independent model, the amount of data required is a bottleneck. While this experiment establishes that building a session-independent model is possible, ideally our model would be one that requires less input from the user for training data. In the next experiment, we thus focus on how to reduce the user effort to produce a session-independent model. In lieu of the results of this experiment, we also finalize the Random Forest+Flattened Image learning algorithm over pre-trained AlexNet, as Random Forest outperforms AlexNet, and takes less time to train. While the performance of AlexNet seemed to be promising, the limited amount of data and the dissimilarity between the ImageNet dataset and our gesture set limited the performance of our model.

4.4 Minimum Required Training Data

Motivation. We seek to find the minimum number of training images per gesture that would allow us to create a good session-dependent model for the initial 10-12 usages, after which the setup can be used in a completely session-independent manner. The initial 10-12 session-dependent usages would provide us with sufficient user data to create a session-independent model.

4.4.1 *Experiment 3.1: Session Dependent model with reduced input.*

Data Collection. We use the same data as the baseline experiment to find the minimum number of training images required per gesture to create a robust session-dependent recognition system.

Experiment. We train the best-performing random forest classifier with 1-5 training images per gesture for all the participants and use 5 test images for each of the gestures.

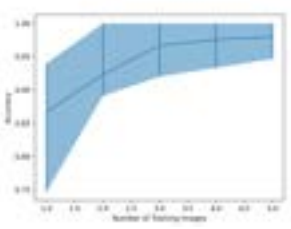

Figure 5: Graph of Number of Training Images vs. Testing accuracy, with errorbar denoting the maximum and minimum accuracies

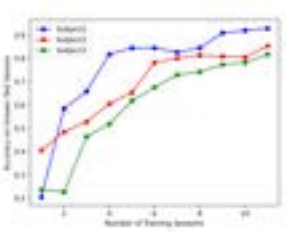

Figure 6: Graph of Number of Training Sessions vs. Accuracy on Unseen Test Sessions

Results. Figure 5 shows a curve representing the Number of Training Images vs Test Accuracy. From the graph, we observe that given 5 training images per gesture, the average testing accuracy comes out to be 98%, and the minimum and maximum accuracies for individual participants are 94.66% and 100% respectively. This result indicates that we can reduce the number of training images used per session, with a bearable loss in accuracy and minimum effort from the user.

4.4.2 Experiment 3.2: Session Independent model with reduced input.

Data Collection. We use the multi-session data for 3 participants, as used in the previous subsection to verify whether fewer training images per session can still create a reasonably well-performing session-independent model.

Experiment. We train our session-independent model (random forest with 100 trees and min max scaler) from experiment 2 with 5 training images per gesture for each session instead of the original 15 images per gesture.

Results. Table 3 demonstrates the performance of our model when 5 training images are used per gesture per session, instead of the original 15. While we suffer some loss in average accuracy: 84.08% to 80.10%, our model still performs well with an average accuracy of 80.1%. These results verify that we can reduce the number of training images required in each session, without any drastic changes in performance. In conclusion, our setup can be used in a session-dependent format for the first 10-12 times, during which our model can be incrementally trained to finally obtain a session-independent model.

5 USER EVALUATION

5.1 Study 1: Spectator's Perspective

We conducted a survey to evaluate the social acceptability of the DUMask gestures both with and without the mask. Our approach consists of asking questions based on video recordings of the gestures because it allows us to show the gestures being performed in a real-world scenario rather than a controlled setting and is a standard practice accepted in the community [6, 25]. We use a between-participants design for the survey consisting of one between-participants factor, whether the gestures were being performed while wearing or not wearing a mask. We chose this design as looking at a gesture being performed without a mask first and then with a mask (or vice versa) can make the responders conscious of what changes to expect. Thus, we wanted to avoid any ordering effects. An unmodified mask was used in this study as we did not want our augmentation to be a factor in how performing the interactions looked. On manufacturing, DUMask can easily be made to look like a regular mask by adding an extra layer of cloth to hide the circuitry.

5.1.1 Procedure. We created two surveys, one corresponding to each gesture with a mask and the other without. Both surveys consisted of the same exact questions with only the videos being different. Responders were shown 15 videos (average duration: 4 seconds) in each scenario, 14 corresponding to one of the gestures being performed once and 1 corresponding to a no gesture state where all actors kept a straight face (which was our control condition). Each survey consisted of the following sections:

(1) **Public space scenario:** This section is based on a scenario that consists of two people inside a lift with one person performing a DUMask gesture while the other does not and keeps a straight face. Each video was followed by the below question: *Which one of the two do you think is making faces?*

• *Left* • *Right* • *Both* • *None*

(2) **Single person scenario:** This section is based on a scenario that consists of one person walking towards the camera while performing a DUMask gesture. Each video was followed by the below question:

Do you think that this person is making faces? • *Yes* • *No*

(3) **Opinion on making faces:** This section did not consist of any videos and was thus the same for both the with-mask and without-mask surveys. It consisted of the following 4 questions:

Do you find it weird when others make faces in public?
• *Yes* • *No*
Do you find it weird when others make faces in public while wearing a mask?
• *Yes* • *No*
Would you be comfortable making faces in public?
• *Yes* • *No*
Would you be comfortable making faces in public while wearing a mask?
• *Yes* • *No*

The order of the gestures in both scenarios was randomized for each responder. Our initial approach with the survey consisted of a general question about anything standing out in the scene with a Likert scale answer. However, we faced issues with this approach in our pilot testing where responders didn't notice anything in the scene or picked on subtle observations unrelated to the gesture (actors' clothing, movement of lift etc.). They also had difficulties quantifying the degrees of weirdness required for the Likert scale answer. Thus, we intentionally asked a leading question to shift their focus to the face which had a binary answer. Though this question is loaded, it works favorably in this specific context as we can verify that even when the participants are biased towards looking for specific actions around the face, they are unable to clearly identify the gestures. We also explored alternate ways to

Figure 7: Scenarios covered in our study on spectator perspective[L-R]: a public space, a lift with a mask; same lift without a mask; a single person with a mask; the same person without a mask. (Screenshots from the videos used)

word the questions such as "gesturing with the mouth" and "moving parts of the face" but "making faces" was the easiest for participants to understand. To not lead the responses around "weirdness", we split the task by first having the responders identify if a gesture was being performed and only then asking if it was weird or not.

5.1.2 Results.

Survey 1: Without a Mask. This survey was filled by 30 people (21 males and 9 females) with ages ranging from 19 to 52 years. The results show that most people do classify all our gestures as making faces i.e. actions different from those seen in daily life. The gestures which some people felt as normal are the lip ones. We attribute this to the fact that they involve very little movement and thus they can be confused with just having the mouth closed if someone does not look too carefully. There was also some confusion in the single-person scenario regarding the tongue-down gesture as only 46.7% classified it as making faces. We speculate that this may be due to the actor not protruding her tongue enough while performing the gesture. We discuss such notes on the videos in the section below.

Survey 2: With a Mask. This survey was filled by 30 people (23 males and 7 females) with ages ranging from 19 to 57 years. The overall responses are ambiguous as to who is performing a gesture (public space scenario) and if a gesture is being performed (single-person scenario). This suggests that people were unable to clearly perceive whether something was happening behind the mask or not and only speculated because our question asked them to.

Two differences that are observed in the results are for the Puff Left gesture in both scenarios (60%, 63.3%) and the Hide Lower Lip in the single-person scenario (56.7%). Looking at the videos for these cases we find a slight jaw movement in all three cases which leads to a movement of the mask. The hide lower lip gesture does involve a slight jaw movement and the actor seems to have taken a deep breath before performing the puff left gesture to fill their cheeks with air.

Our intention with the videos was to record in a natural setting without over-instructing the actor on how to perform the gestures. This seems to have led to some artifacts in the recordings due to possible variations in how one can perform a gesture e.g. sucking in air/blowing air from the lungs to puff a cheek. While the results are overall in favor of a mask being able to occlude facial gestures, finding the most discreet way to perform a gesture can be a future exploration.

Social Acceptability of Making Faces. Merging the responses from both survey groups for the last section we see that a majority of responders are averse to making faces in public (76.67%) and find it strange when others make faces in public (61.67.7%). However, making faces while wearing a mask is deemed to be acceptable by a similar majority (68.33%). These results when combined with the fact that most people were not able to identify gestures while a mask is worn prove the social acceptability of the DUMask interface.

5.2 Study 2: User's Perspective

To understand the usability of our system, we conducted a user study with the actual DUMask interface. 12 individuals (6 male and 6 female) participated in the study and their ages ranged from 19-41 years.

5.2.1 Procedure. We introduced the participants to DUMask by describing its purpose and functionality. The participants then had to wear DUMask and one of the researchers demonstrated the gestures one by one while not wearing a mask themselves. After each demonstration, the participants had to try out the gestures themselves and rate them on a 5-point Likert scale how 1) **Comfortable**, 2) **Easy to perform** and 3)**Unobtrusive** they were. After trying out all the gestures, we took a short semi-structured interview on the participant's overall experience with the DUMask system.

5.3 Analysis

We observe that tongue out was the most preferred gesture with a high average rating and small standard deviation for all three factors. Participants felt "used to" (P6) this gesture having performed it before. Next come the hide lip gestures, out of which the hide upper and lower lip ones were rated highest for unobtrusiveness ($\mu = 4.92, \sigma = 0.29$ for both) as they required "very less movement" (P1). Conversely, the puffing gestures were rated lowest for unobtrusiveness as they lead to a large change in the shape of the lower half of the face.

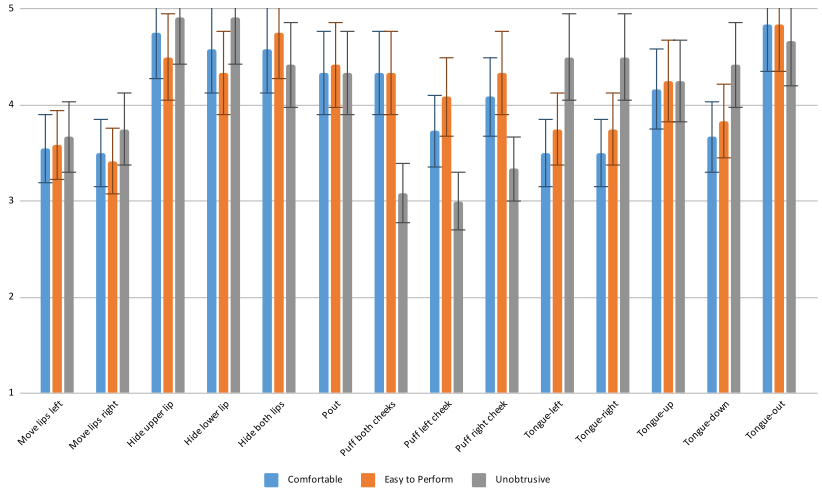

Figure 8: Summary of Likert Scale Ratings for how 1) Comfortable, 2) Easy to Perform and 3) Unobtrusive each Gesture was

The move lips and tongue left/right gestures were the least favored amongst the participants with one commenting that "moving my lips sideways felt unusual" (P3). Both move lips gestures were rated lower than the tongue gestures for unobtrusiveness due to more movement involved in performing them. However, they are rated similarly on comfort and ease of performance. An interesting observation is that none of the participants favored one side for facial gestures, with the tongue left and right gestures having the exact same rating for each participant. Although Move Lips Left/Right and Puff Left/Right also have similar ratings for each side, they are not identical. We had expected the tongue-based gestures to have a higher easy-to-perform rating than the others due to the dexterity of the tongue but no such trend was observed.

6 DISCUSSION

6.1 Potential Applications

Along with being used as a hands free interface, DUMask's design also facilitates additional applications enabled through a mask-based interface.

Private Interactions. Since DUMask is designed to hide the interactions being performed by the user, it adds a layer of privacy by not letting anyone know if and what gesture is being performed. For example, one can send an SOS to alert their family about their location without inducing risk in a dangerous situation. Further, a series of gestures supported by DUMask may be used as a password for sensitive information or to access restricted areas in a building.

Touchless/Contactless Interactions. With the world surrounded by the COVID-19 pandemic, an obvious application is using DUMask to interact with the interfaces for public infrastructure without touching them. For example, a user can use DUMask to navigate in an elevator. Before entering, the user can choose between going up or down with the tongue up and down gesture. Once inside the elevator, the user can choose which floor to go to by performing the assigned gesture for that specific floor.

Assistive Device. Persons with disabilities often make use of facial gestures to perform particular tasks. The discreet nature of DUMask enables these individuals to use facial gestures in public without fear of judgement from onlookers or the discomfort of holding something in their mouth. Another use case could be mute individuals using DUMask gestures to construct sentences which could then be used to communicate with people around them by using a speaker embedded in the mask and a text-to-speech service.

Leveraging DUMask's Camera. Smartphone cameras have been extensively leveraged to propose healthcare solutions [11, 51, 52]. Owing to its design and placement, the DUMask camera has a complete view of the user's mouth offering an opportunity to monitor their oral activity associated with lips, tongue, and teeth. DUMask can be used to diagnose oral para functions such as bruxism, clenching, and lip biting. Additionally, the camera feed can also be used to train a silent speech interface.

6.2 Using Clothing to Enable Discreet Interactions

While DUMask is a unique non-intrusive socially acceptable oral interface, the idea of using a face mask can be used to enhance existing works in this domain. For example, if discreetness is not a requirement, microphones can be embedded inside the mask to support audio-based interactions as in [8]. Intrusive interfaces such as [13, 55] can possibly be made socially acceptable by hiding the gestures performed and the hardware inside the mask. Similarly, other articles of clothing can also be used to hide interactions such as performing hand gestures inside mittens and pockets, toe wiggling gestures inside shoes or ear wiggling inside a beanie.

6.3 Towards DUMask as a Consumer Product

In the current setup, the camera and IR LEDs are tethered to the processing unit (RPi) via a ribbon cable for power and data transfer. For future iterations, we imagine a USB C cable, running through the ear strap, to the camera and IR illuminator inside the mask to a smartphone for power and data processing (similar to market available AR glasses[7]). This prototype would completely hide the camera inside the mask and offload all the processing to the mobile device. Alternatively, a wireless interface sporting a battery and a micro-controller unit can be created by adding a custom strap situated at the back of the user's head. While we acknowledge that masks are an essential wearable in the current world scenario, the addition of a custom strap would add weight to the mask which could cause minor discomfort. Custom straps have become quite prevalent for wireless VR headsets[8] to improve the battery life as well as to balance the front-heaviness of the interface. Further, this would allow DUMask to offer a completely wireless interface capable of capturing and recognising gestures via the in-strap setup.

When it comes to hygiene, we imagine having masks custom stitched for DUMask that have openings to insert the electronics that also have flaps to conceal what is inside. These custom stitched masks can be then washed or disposed after use similar to regular masks. For ensuring the electronics remain contagion free and clean, they will need to be disinfected with methods used to sanitize electronics such as UVC Light [9].

7 LIMITATIONS AND FUTURE WORK

This work aims to introduce the possibilities enabled by the mask form factor and validate its feasibility through a research prototype. Our approach to validating the performance of our interface in a controlled lab setting is based on similar papers on oral interactions published at top HCI conferences [7, 8, 15, 22, 32, 35, 45, 46, 57]. While DUMask is a fully functional research prototype, it is meant to be a proof-of-concept and is currently not optimised for daily usage. However, it is essential to discuss how it may perform in real-world settings and how we can handle potential issues that may occur. Till we are able to add these specific enhancements to the interface, we reserve testing DUMask in more ecological settings for future work.

7.0.1 Moisture in the Air Between the Mask and the Face. Misting on the DUMask lens may occur due to condensation of the water vapour from the wearer's breath preventing the camera from detecting any gestures. With the increased mask usage due to the Covid-19 pandemic, research has been conducted on preventing fogging on spectacles while wearing masks [17, 18, 33] and any of the proposed methods such as antifogging agents or iodophor can be applied on the camera lens to prevent fogging.

7.0.2 False-positives with Other Facial Movements. Although the DUMask gesture set by design consists of movements that are not performed usually in everyday life, certain measures can be taken to further prevent the detection of false positives. In the current setup, the interface has a three sample check which considers 3 samples evenly distributed over a two-second window. For a gesture to be detected, three consecutive samples would have to yield the same result when processed by our model. Further, a microphone can be installed inside the interface such that DUMask does not detect gestures when the wearer is speaking. Another alternative could be to use one of the puff cheek gestures (as they involve the most movement) as a "Double Flip" gesture [44] to signal to DUMask to start detection.

7.0.3 User Independent Recognition. The issue of low person independent accuracy is often noted in oral interaction papers with researchers building user-dependent systems for teeth [8, 32], tongue [45, 57], and silent speech [22] gestures. Since user-dependent models are an accepted practice for mouth-based gestures in the community, we use a similar approach to prove the feasibility of the mask form factor and to demonstrate that oral gestures can be accurately detected even across different alignments of the mask. A large amount of training data may enable us to produce a transfer learning-based model, that is pre-trained on data from a large number of participants, and needs fewer images from the users. A greater amount of data would also open the possibility of using a deeper network and training it from scratch. However, this would require a huge data collection effort on people with different facial features making it out of scope for our current work.

8 CONCLUSION

In this work, we present DUMask, a face mask-based oral interface which captures facial gestures. We embedded an IR camera inside a mask to recognize the gestures performed by a user while wearing a mask. We evaluate our interface with a preliminary set of 14 gestures performed by manipulating Cheeks, Tongue and/or Lips. Our evaluation of the user's as well as on looker's perception of performing the gestures while wearing a mask shows the feasibility of the interface and that it is suitable to be used even in public.

ACKNOWLEDGMENTS

This work is supported by the Centre for Design and New Media, a TCS Foundation initiative, supported by Tata Consultancy Services.

REFERENCES

[1] [n.d.]. Sip/Puff Switch. https://www.orin.com/access/sip_puff/
[2] Rishiraj Adhikary, Dhruvi Lodhavia, Chris Francis, Rohit Patil, Tanmay Srivastava, Prerna Khanna, Nipun Batra, Joe Breda, Jacob Peplinski, and Shwetak Patel. 2022. SpiroMask: Measuring Lung Function Using Consumer-Grade Masks. *ACM Transactions on Computing for Healthcare* (Nov. 2022). https://doi.org/10.1145/3570167 Just Accepted.
[3] Rishiraj Adhikary, Tanmay Srivastava, Prerna Khanna, Aabhas Asit Senapati, and Nipun Batra. 2020. Naqaab: Towards Health Sensing and Persuasion via Masks. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers* (Virtual Event, Mexico) *(UbiComp-ISWC '20)*. Association for Computing Machinery, New York, NY, USA, 5–8. https://doi.org/10.1145/3410530.3414403
[4] David Ahlström, Khalad Hasan, and Pourang Irani. 2014. Are You Comfortable Doing That? Acceptance Studies of around-Device Gestures in and for Public Settings. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Toronto, ON, Canada) *(MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 193–202. https://doi.org/10.1145/2628363.2628381
[5] Takashi Amesaka, Hiroki Watanabe, and Masanori Sugimoto. 2019. Facial Expression Recognition Using Ear Canal Transfer Function. In *Proceedings of the 23rd International Symposium on Wearable Computers* (London, United Kingdom)

[7] https://epson.com/For-Work/Wearables/Smart-Glasses/Moverio-BT-40-Smart-Glasses-with-USB-Type-C-Connectivity-/p/V11H969020
[8] https://www.oculus.com/accessories/quest-2-elite-strap-battery/
[9] https://cleanboxtech.com/

(ISWC '19). Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3341163.3347747

[6] Fraser Anderson, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2015. Supporting Subtlety with Deceptive Devices and Illusory Interactions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 1489–1498. https://doi.org/10.1145/2702123.2702336

[7] 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* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 679–689. https://doi.org/10.1145/3126594.3126649

[8] Daniel Ashbrook, Carlos Tejada, Dhwanit Mehta, Anthony Jiminez, Goudam Muralitharam, Sangeeta Gajendra, and Ross Tallents. 2016. Bitey: An Exploration of Tooth Click Gestures for Hands-Free User Interface Control. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 158–169. https://doi.org/10.1145/2935334.2935389

[9] Irina Blok. [n.d.]. Fashion Surgical Masks. https://www.irinablok.com/fashionmasks

[10] Alexander Curtiss, Blaine Rothrock, Abu Bakar, Nivedita Arora, Jason Huang, Zachary Englhardt, Aaron-Patrick Empedrado, Chixiang Wang, Saad Ahmed, Yang Zhang, Nabil Alshurafa, and Josiah Hester. 2022. FaceBit: Smart Face Masks Platform. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (Dec. 2022), 151:1–151:44. https://doi.org/10.1145/3494991

[11] Lilian de Greef, Mayank Goel, Min Joon Seo, Eric C. Larson, James W. Stout, James A. Taylor, and Shwetak N. Patel. 2014. BiliCam: Using Mobile Phones to Monitor Newborn Jaundice. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. ACM, New York, NY, USA, 331–342. https://doi.org/10.1145/2632048.2632076

[12] Rose Eveleth. 2019. Will air-filtering face masks be the new sunglasses? https://www.vox.com/the-goods/2019/3/19/18262556/face-mask-air-filter-pollution-vogmask-airpop

[13] Pablo Gallego Cascón, Denys J.C. Matthies, Sachith Muthukumarana, and Suranga Nanayakkara. 2019. ChewIt. An Intraoral Interface for Discreet Interactions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article 326, 13 pages. https://doi.org/10.1145/3290605.3300556

[14] Çağlar Genç, Ashley Colley, Markus Löchtefeld, and Jonna Häkkilä. 2020. Face Mask Design to Mitigate Facial Expression Occlusion. In *Proceedings of the 2020 International Symposium on Wearable Computers* (Virtual Event, Mexico) *(ISWC '20)*. Association for Computing Machinery, New York, NY, USA, 40–44. https://doi.org/10.1145/3410531.3414303

[15] 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* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 255–258. https://doi.org/10.1145/2702123.2702591

[16] M. Hamedi, Sh-Hussain Salleh, T. S. Tan, K. Ismail, J. Ali, C. Dee-Uam, C. Pavaganun, and P. P. Yupapin. 2011. Human facial neural activities and gesture recognition for machine-interfacing applications. *International journal of nanomedicine* 6 (2011), 3461–3472. https://doi.org/10.2147/IJN.S26619 22267930[pmid].

[17] J Hu and J Zhao. 2020. Anti-fog skills for medical goggles during the period of prevention and control of Coronavirus disease 2019. *Chin Nurs Res* 34, 4 (2020), 573.

[18] Yuli Hu, Lan Wang, Sanlian Hu, and Fang Fang. 2020. Prevention of fogging of protective eyewear for medical staff during the COVID-19 pandemic. *Journal of Emergency Nursing* 46, 5 (2020), 564–566.

[19] X. Huo, J. Wang, and M. Ghovanloo. 2008. A Magneto-Inductive Sensor Based Wireless Tongue-Computer Interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 16, 5 (Oct 2008), 497–504. https://doi.org/10.1109/TNSRE.2008.2003375

[20] Yasha Iravantchi, Yang Zhang, Evi Bernitsas, Mayank Goel, and Chris Harrison. 2019. Interferi: Gesture Sensing Using On-Body Acoustic Interferometry. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article 276, 13 pages. https://doi.org/10.1145/3290605.3300506

[21] Jesse Jokerst. 2021. *Making Masks Smarter and Safer against COVID-19*. https://jacobsschool.ucsd.edu/news/release/3206

[22] Arnav Kapur, Shreyas Kapur, and Pattie Maes. 2018. AlterEgo: A Personalized Wearable Silent Speech Interface. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. Association for Computing Machinery, New York, NY, USA, 43–53. https://doi.org/10.1145/3172944.3172977

[23] Zendai Kashino, Daisuke Uriu, Ziyue Zhang, Shigeo Yoshida, and Masahiko Inami. 2022. A New Mask for a New Normal: Investigating an AR Supported Future under COVID-19. In *Augmented Humans 2022 (AHs 2022)*. Association for Computing Machinery, New York, NY, USA, 243–253. https://doi.org/10.1145/3519391.3519409

[24] Jeonghee Kim, Hangue Park, Joy Bruce, Erica Sutton, Diane Rowles, Deborah Pucci, Jaimee Holbrook, Julia Minocha, Beatrice Nardone, Dennis West, Anne Laumann, Eliot Roth, Mike Jones, Emir Veledar, and Maysam Ghovanloo. 2013. The tongue enables computer and wheelchair control for people with spinal cord injury. *Science Translational Medicine* 5, 213 (27 11 2013). https://doi.org/10.1126/scitranslmed.3006296

[25] Marion Koelle, Swamy Ananthanarayan, and Susanne Boll. 2020. Social Acceptability in HCI: A Survey of Methods, Measures, and Design Strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3313831.3376162

[26] Yuto Koguchi, Kazuya Oharada, Yuki Takagi, Yoshiki Sawada, Buntarou Shizuki, and Shin Takahashi. 2018. A Mobile Command Input Through Vowel Lip Shape Recognition. In *Human-Computer Interaction. Interaction Technologies*, Masaaki Kurosu (Ed.). Springer International Publishing, Cham, 297–305.

[27] Eunjeong Stella Koh and Shahrokh Yadegari. 2018. Mugeetion: Musical Interface Using Facial Gesture and Emotion. *CoRR* abs/1809.05502 (2018). arXiv:1809.05502 http://arxiv.org/abs/1809.05502

[28] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.), Vol. 25. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[29] Ryoga Kumazaki and Akifumi Inoue. 2019. Development and Evaluation of a Mask-Type Display Transforming the Wearer's Impression. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction* (Fremantle, WA, Australia) *(OZCHI'19)*. Association for Computing Machinery, New York, NY, USA, 568–571. https://doi.org/10.1145/3369457.3369533

[30] Yusuke Kunimi, Masa Ogata, Hirotaka Hiraki, Motoshi Itagaki, Shusuke Kanazawa, and Masaaki Mochimaru. 2022. E-MASK: A Mask-Shaped Interface for Silent Speech Interaction with Flexible Strain Sensors. In *Augmented Humans 2022 (AHs 2022)*. Association for Computing Machinery, New York, NY, USA, 26–34. https://doi.org/10.1145/3519391.3519399

[31] M. Lee, S. Je, W. Lee, D. Ashbrook, and A. Bianchi. 2019. ActivEarring: Spatiotemporal Haptic Cues on the Ears. *IEEE Transactions on Haptics* 12, 4 (Oct 2019), 554–562. https://doi.org/10.1109/TOH.2019.2925799

[32] Cheng-Yuan Li, Yen-Chang Chen, Wei-Ju Chen, Polly Huang, and Hao-hua Chu. 2013. Sensor-Embedded Teeth for Oral Activity Recognition. In *Proceedings of the 2013 International Symposium on Wearable Computers* (Zurich, Switzerland) *(ISWC '13)*. Association for Computing Machinery, New York, NY, USA, 41–44. https://doi.org/10.1145/2493988.2494352

[33] L Li and X Cai. 2020. The observation of the application and effect of using different anti-fogging methods in the intensive care isolation units during Covid-19. *J Nurs Train (China)* 35, 9 (2020), 817–9.

[34] 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* (Singapore, Singapore) *(ISWC '18)*. Association for Computing Machinery, New York, NY, USA, 180–183. https://doi.org/10.1145/3267242.3267265

[35] Richard Li, Jason Wu, and Thad Starner. 2019. TongueBoard: An Oral Interface for Subtle Input. In *Proceedings of the 10th Augmented Human International Conference 2019* (Reims, France) *(AH2019)*. Association for Computing Machinery, New York, NY, USA, Article 1, 9 pages. https://doi.org/10.1145/3311823.3311831

[36] Konrad Lukaszewicz. 2003. The Ultrasound Image of the Tongue Surface as Input for Man/Machine Interface. In *INTERACT*.

[37] Motoyasu Masui, Yoshinari Takegawa, and Keiji Hirata. 2022. Dynamic Appearance Augmentation Method that Enables Easy Prototyping of Masks for Performance. In *Augmented Humans 2022 (AHs 2022)*. Association for Computing Machinery, New York, NY, USA, 267–275. https://doi.org/10.1145/3519391.3522751

[38] Denys Matthies, Nastaran Saffaryazdi, and Mark Billinghurst. 2022. Wearable Sensing of Facial Expressions and Head Gestures. In *NordiCHI'22 Workshop*. https://doi.org/10.13140/RG.2.2.26960.38408/2

[39] Denys J. C. Matthies, Bodo Urban, Katrin Wolf, and Albrecht Schmidt. 2019. Reflexive Interaction: Extending the Concept of Peripheral Interaction. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction* (Fremantle, WA, Australia) *(OZCHI'19)*. Association for Computing Machinery, New York, NY, USA, 266–278. https://doi.org/10.1145/3369457.3369478

[40] Jess McIntosh, Asier Marzo, and Mike Fraser. 2017. SensIR: Detecting Hand Gestures with a Wearable Bracelet Using Infrared Transmission and Reflection. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 593–597. https://doi.org/10.1145/3126594.3126604

[41] Takuro Nakao, Yun Suen Pai, Megumi Isogai, Hideaki Kimata, and Kai Kunze. 2018. Make-a-Face: A Hands-Free, Non-Intrusive Device for Tongue/Mouth/Cheek Input Using EMG. In *ACM SIGGRAPH 2018 Posters* (Vancouver, British Columbia,

Canada) *(SIGGRAPH '18)*. Association for Computing Machinery, New York, NY, USA, Article 24, 2 pages. https://doi.org/10.1145/3230744.3230784

[42] Phuc Nguyen, Nam Bui, Anh Nguyen, Hoang Truong, Abhijit Suresh, Matt Whitlock, Duy Pham, Thang Dinh, and Tam Vu. 2018. TYTH-Typing On Your Teeth: Tongue-Teeth Localization for Human-Computer Interface. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services* (Munich, Germany) *(MobiSys '18)*. Association for Computing Machinery, New York, NY, USA, 269–282. https://doi.org/10.1145/3210240.3210322

[43] Francesca Nonis, Nicole Dagnes, Federica Marcolin, and Enrico Vezzetti. 2019. 3D Approaches and Challenges in Facial Expression Recognition Algorithms—A Literature Review. *Applied Sciences* 9, 18 (2019), 3904.

[44] Jaime Ruiz and Yang Li. 2011. DoubleFlip: A Motion Gesture Delimiter for Mobile Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2717–2720. https://doi.org/10.1145/1978942.1979341

[45] Himanshu Sahni, Abdelkareem Bedri, Gabriel Reyes, Pavleen Thukral, Zehua Guo, Thad Starner, and Maysam Ghovanloo. 2014. The Tongue and Ear Interface: A Wearable System for Silent Speech Recognition. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers* (Seattle, Washington) *(ISWC '14)*. Association for Computing Machinery, New York, NY, USA, 47–54. https://doi.org/10.1145/2634317.2634322

[46] 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* (Victoria, BC, Canada) *(UIST '09)*. Association for Computing Machinery, New York, NY, USA, 177–180. https://doi.org/10.1145/1622176.1622209

[47] Britta F. Schulte, Zuzanna Lechelt, and Aneesha Singh. 2018. Giving up Control - A Speculative Air Pollution Mask to Reflect on Autonomy and Technology Design. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems* (Hong Kong, China) *(DIS '18 Companion)*. Association for Computing Machinery, New York, NY, USA, 177–181. https://doi.org/10.1145/3197391.3205432

[48] Luna Takagi, Toshiki Sato, Shio Miyafuji, and Hideki Koike. 2021. Real-time Projection of Lip Animation onto Face Masks using OmniProcam. In *ACM SIGGRAPH 2021 Posters (SIGGRAPH '21)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3450618.3469171

[49] Kathryn G. Tippey, Elayaraj Sivaraj, and Thomas K. Ferris. 2017. Driving While Interacting With Google Glass: Investigating the Combined Effect of Head-Up Display and Hands-Free Input on Driving Safety and Multitask Performance. *Human Factors* 59, 4 (2017), 671–688. https://doi.org/10.1177/0018720817691406 arXiv:https://doi.org/10.1177/0018720817691406 PMID: 28186420.

[50] Tri Vu, Feng Lin, Nabil Alshurafa, and Wenyao Xu. 2017. Wearable food intake monitoring technologies: A comprehensive review. *Computers* 6, 1 (2017), 4.

[51] Edward Jay Wang, William Li, Doug Hawkins, Terry Gernsheimer, Colette Norby-Slycord, and Shwetak N. Patel. 2016. HemaApp: Noninvasive Blood Screening of Hemoglobin Using Smartphone Cameras. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. ACM, New York, NY, USA, 593–604. https://doi.org/10.1145/2971648.2971653

[52] E. J. Wang, W. Li, J. Zhu, R. Rana, and S. N. Patel. 2017. Noninvasive hemoglobin measurement using unmodified smartphone camera and white flash. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2333–2336. https://doi.org/10.1109/EMBC.2017.8037323

[53] Xiaoqian Xie, Jiashuo Cao, Jiayu Yao, Lunyu Shang, Qianru Liu, Jinlun Lin, and Qin Wu. 2020. MaskMe: Using Masks to Design Collaborative Games for Helping Children with Autism Make Eye Contact. In *Companion Proceedings of the 2020 Conference on Interactive Surfaces and Spaces (ISS '20)*. Association for Computing Machinery, New York, NY, USA, 29–32. https://doi.org/10.1145/3380867.3426206

[54] Kele Xu, Yuxiang Wu, and Zhifeng Gao. 2019. Ultrasound-Based Silent Speech Interface Using Sequential Convolutional Auto-Encoder. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 2194–2195. https://doi.org/10.1145/3343031.3350596

[55] Xuhai Xu, Chun Yu, Anind K. Dey, and Jennifer Mankoff. 2019. Clench Interface: Novel Biting Input Techniques. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article 275, 12 pages. https://doi.org/10.1145/3290605.3300505

[56] Koki Yamashita, Takashi Kikuchi, Katsutoshi Masai, Maki Sugimoto, Bruce H. Thomas, and Yuta Sugiura. 2017. CheekInput: Turning Your Cheek into an Input Surface by Embedded Optical Sensors on a Head-Mounted Display. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (Gothenburg, Sweden) *(VRST '17)*. Association for Computing Machinery, New York, NY, USA, Article 19, 8 pages. https://doi.org/10.1145/3139131.3139146

[57] 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* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2555–2558. https://doi.org/10.1145/2556288.2556981