FaceSense: Sensing Face Touch with an Ear-worn System

VIMAL KAKARAPARTHI[†], University of Colorado Boulder
QIJIA SHAO[†], Dartmouth College
CHARLES J. CARVER, Dartmouth College
TIEN PHAM, University of Texas at Arlington
NAM BUI, University of Colorado Boulder
PHUC NGUYEN, University of Texas at Arlington
XIA ZHOU, Dartmouth College
TAM VU, University of Colorado Boulder and Oxford University

Face touch is an unconscious human habit. Frequent touching of sensitive/mucosal facial zones (eyes, nose, and mouth) increases health risks by passing pathogens into the body and spreading diseases. Furthermore, accurate monitoring of face touch is critical for behavioral intervention. Existing monitoring systems only capture objects approaching the face, rather than detecting actual touches. As such, these systems are prone to false positives upon hand or object movement in proximity to one's face (e.g., picking up a phone). We present FaceSense, an ear-worn system capable of identifying actual touches and differentiating them between sensitive/mucosal areas from other facial areas. Following a multimodal approach, FaceSense integrates low-resolution thermal images and physiological signals. Thermal sensors sense the thermal infrared signal emitted by an approaching hand, while physiological sensors monitor impedance changes caused by skin deformation during a touch. Processed thermal and physiological signals are fed into a deep learning model (TouchNet) to detect touches and identify the facial zone of the touch. We fabricated prototypes using off-the-shelf hardware and conducted experiments with 14 participants while they perform various daily activities (e.g., drinking, talking). Results show a macro-F1-score of 83.4% for touch detection with leave-one-user-out cross-validation and a macro-F1-score of 90.1% for touch zone identification with a personalized model.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Computer systems organization** → **Embedded systems**.

Additional Key Words and Phrases: face touch detection, thermo-physiological sensing, multimodal deep learning.

ACM Reference Format:
Vimal Kakaraparthi[†], Qijia Shao[†], Charles J. Carver, Tien Pham, Nam Bui, Phuc Nguyen, Xia Zhou, and Tam Vu. 2021. FaceSense: Sensing Face Touch with an Ear-worn System. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 110 (September 2021), 27 pages. https://doi.org/10.1145/3478129

† Both co-primary authors contributed equally to this work. Listed alphabetically.

Authors' addresses: Vimal Kakaraparthi[†], University of Colorado Boulder, Boulder, CO, 80309, venkata.kakaraparthi@colorado.edu; Qijia Shao[†], Dartmouth College, 9 Maynard St, Hanover, NH, 03755, Qijia.Shao.GR@dartmouth.edu; Charles J. Carver, Dartmouth College; Tien Pham, University of Texas at Arlington; Nam Bui, University of Colorado Boulder; Phuc Nguyen, University of Texas at Arlington; Xia Zhou, Dartmouth College; Tam Vu, University of Colorado Boulder and Oxford University.

2474-9567/2021/9-ART110
https://doi.org/10.1145/3478129

1 INTRODUCTION

Touching the face is an unconscious habit linked to key aspects of day-to-day human existence (e.g., irritation, anxiousness, and lack of focus) [8, 26, 34, 42, 58]. Face touching has been observed to be as frequent as 23 times per hour [43], increasing health risks related to touching sensitive areas. For example, touching the eyes, nose, or mouth (mucosal areas) with hands covered in microbes can pass pathogens into the body and spread diseases (e.g., COVID-19, Ebola, influenza) [53, 85]. In addition, frequent face touching is often associated with obsessive-compulsive behaviors such as chronic eye-rubbing [55] and compulsive touching [52], both which have long-lasting effects on the body [47, 69]. To facilitate changing face touching behavior, it is critical to monitor face touches accurately and continuously. The occurrence of face touch events can serve as a metric for evaluating the progress of behavioral change. The monitoring results can also be integrated into awareness enhancement devices (e.g., phones, watches, headphones) to alert users when face touches occur via various in-situ feedback such as vibrations and sounds [7, 37].

In the past year, many feedback devices/technologies have emerged for monitoring face touches. However, existing systems only capture approaching hands rather than detecting actual touches. For example, FaceOff [15] uses a wrist-worn accelerometer to identify potential hand movement patterns that may lead to face touches. Unfortunately, activities like drinking water, eating food, and scratching the neck cause false feedback. In [4, 68, 90], hand-to-face proximity is monitored using ultrasound, magnets, and radio signals aided by body worn devices (e.g., earphones, smartphones, smartwatches). However, since hand-to-face proximity is used to categorize face touch events, false feedback is inevitable for the same reasons above. Additionally, false positives occur from everyday objects that have similar properties as the body worn devices (e.g., magnetic objects, Bluetooth connected devices) or objects that reflect ultrasound waves. Attempting to control face touching using Habit Reversal Therapy (HRT) is therefore ineffective given the unreliable feedback from existing systems [7, 37]. Researchers investigating HRT found that such therapy is only effective when participants accept the models that are used to create awareness of the habit [38, 74]. In the case of face touching, existing devices use a behavior (hand-to-face proximity) that has significant overlap with essential daily activities, as well as benign trivial activities. This could eventually lead to mistrusting the model of the device. Consequently, there is a need for a system capable of detecting actual hand-to-face contact with high fidelity.

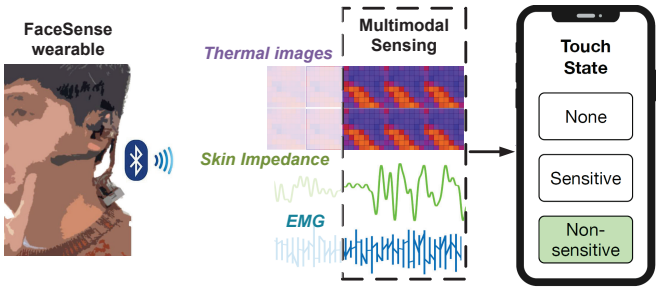


Fig. 1. FaceSense's concept: fusing physiological and thermal streams for touch detection and touch zone identification.

We propose FaceSense, a wearable headset that can detect face touch events and identify the facial zone of the touch. FaceSense follows a multimodal sensing approach by integrating thermal image sensors and physiological sensors. Thermal sensors sense the thermal infrared signal emitted by approaching hands, and physiological sensors monitor skin deformation caused by face touches. These two sensing modalities aid each other to detect not only actual touches, but also whether the touches have landed in sensitive zones (eyes, mouth, nose). During the development of FaceSense, we face a series of challenges. First, face touch events generate micro-movements on the skin's surface which is difficult to capture with low-cost, low-power sensors. Second, although thermal cameras are a privacy-preserving, small, and low-power solution for detecting the presence of a hand, they suffer from strong environmental noise caused by objects with any thermal signature. Third, existing algorithms are

insufficient at taking heterogeneous inputs from the thermal and physiological sensors and returning a unified output for face touch detection and touch zone identification.

We address these challenges as follows. (1) To extract skin deformation information caused by face touches, we explore an asymmetry-based impedance and EMG analysis. (2) To deal with interference from other thermal objects, we leverage a GMM-based model to extract the foreground object from low-resolution thermal images, allowing hand information to be extracted in the presence of background thermal objects. (3) We propose TouchNet, a multimodal, user-independent, deep learning model which recognizes useful cross-modality relationships in addition to modality-specific features for accurate face touch detection and touch zone identification.

We designed and fabricated a pair of earpieces as a proof-of-concept to evaluate the efficacy of FaceSense. Using 3D printed, thermoplastic urethane, the earpieces are elastic, lightweight, and soft. Each earpiece hosts utmost 3 electrodes for physiological sensing and a thermal image sensor with 8×8 pixels. The sensor data is transmitted back to a computing device (e.g., a smartphone or a computer) via Bluetooth for further processing. We conducted a user study with 14 participants where each user touches their face naturally throughout various daily activities (e.g., talking, eating, drinking). We obtain the following key findings:

- FaceSense achieves a macro-F1-score of 83.4% for face touch detection with leave-one-user-out cross validation, indicating the feasibility of a face touch detection system without any training overhead.
- FaceSense obtains a macro-F1-score of 90.1% on average for face touch zone identification with a personalized model trained with 5-minutes' worth of data from each user, demonstrating the ability to identify the touch zone with only a small training overhead.
- FaceSense can reliably differentiate between face touch events and other daily activities involving hand-to-face motion (e.g., eating, drinking), outperforming existing face touch monitoring systems.

Contributions. We summarize our contributions as follows.

- To the best of our knowledge, FaceSense is the first system capable of detecting actual hand-to-face contact and identifying the facial zone of the touch. It can therefore help to prevent/reduce health and/or behavioral issues related to face touching.
- We propose a thermo-physiological sensing methodology to jointly capture the thermal infrared signal emitted by approaching hands and the skin deformation caused by face touches.
- We design and implement TouchNet, a multimodal deep learning model which fuses thermal images and physiological data streams, extracting effective representations of face touch detection and touch zone identification.
- We build a FaceSense prototype and validate its performance with 14 participants in various environmental settings and throughout daily activities.

2 BACKGROUND AND MOTIVATIONS

Before discussing FaceSense, we examine key concepts underlying face touch, existing approaches to detect face touch, and the motivations of developing a thermo-physiological technique to detect face touch and identify the facial zone of the touch.

Frequent Face Touch: A Common Mistake. Since hands are an integral part of typical human interactions with their surroundings, it is of critical importance to monitor face touch events in the wake of viral outbreaks such as the COVID-19 pandemic. Many studies delve into the subject of self-touching, of which face touch events are a significant part. Apart from skin irritations, face touch is linked to an individual's emotional and cognitive states as this activity was observed to increase attentiveness, pressure, and anxiety as a tool of attentional disruption [8, 58]. In [42], face touch is observed to occur during the events of incompatible responses. Face touch is also used as a tool of feedback during performance of tasks [58]. All the research regarding face touch habits

suggest the involuntary nature of the activity [26, 34]. Apart from this, there are conditions such as chronic eye-rubbing [55], compulsive touch [52] and several such Obsessive Compulsive behaviors [47, 69] that can lead to increased frequency of face touch events in an individual. A behavioral study [43] conducted on 26 participants during a lecture, shows that face touch event can be as frequent as 23 times per hour. In these face touch events, the fraction of touch events happening in the mucosal areas (referred as *sensitive touch* in this paper) such as eyes, nose, mouth are 44% while the remaining touch events happened at non-mucosal areas (referred as *non-sensitive touch* in this paper) such as chin, cheek, ear etc. Another study [71] observed 1000 people in public places during the COVID-19 pandemic. It is observed that the face touch event occurred 10 times per hour on average, where, sensitive touch was happening at an average of 5.5 times an hour. Such studies show the necessity of a reliable, non-invasive face touch detection and touch zone identification system that is socially acceptable to wear. FaceSense is the first system to detect and categorize the face touch events into mucosal (sensitive) and non-mucosal (non-sensitive) types. In addition, the device can be used in Habit Reversal Therapies (HRT) for various compulsive face touch behaviors.

Detecting Hand-face Contact, Not Hand Proximity. Researchers designed multiple hand-to-face proximity detection systems to identify the possibility of a hand touching the face and alert user timely. Wrist-worn devices and camera-based solutions are two main approaches to solve this problem. In particular, wrist-worn devices like accelerometers, magnets etc. are paired with smartphone attributes to identify passive signatures of face touch events [4, 15, 25, 56]. Such approaches suffer from false positives generated by hand movements similar to face touch events (drinking water, eating, picking up a phone call etc.) and interference noises from everyday devices such as computers, smartphones and general household appliances [15, 25]. In [91], researchers use a SONAR approach that sends ultrasonic waves using off-the-shelf earbuds and receives the waves reflected from an approaching hand using a microphone. Another approach to measure hand proximity is to measure the strength of BLE communication between hand-worn smartwatch and face-worn earbuds (RSSI) [68]. These approaches are bound to have many false positives from everyday surrounding obstacles like (smartphones, laptops etc.) and similar hand movements that do not involve face touch. Camera-based technologies leverage computer vision to detect face touch [5, 6, 12, 35, 59], but they raise strong privacy concerns. These systems also need the camera to be capturing the user in a front-face orientation during their usage, which would restrict user mobility significantly even while being used in an indoor setting. In summary, all of the above systems can only detect whether there is a hand approaching the face or not, but they are not designed to capture the 'actual face touch' events. To fill the gap, we build a system to not only detect hand-to-face proximity, but also to detect hand-to-face contact and their contact zone on a human face.

3 SYSTEM OVERVIEW AND CHALLENGES

3.1 System Overview

The main objective of this project is to design a wearable device that is able to detect hand-face contact and identify the contact zone for face touch activities monitoring. First, the system needs to detect actual hand-to-face contact to build the user trust in the model of the device, which is necessary for an effective habit reversal. Second, the device needs to be able to detect facial zones of touch, sensitive (mucosal) and non-sensitive (non-mucosal) zones in our case, to have more relevant feedback during viral outbreaks. Third, the system needs to be a compact wearable that is effective in diverse conditions and environments without raising any public concerns.

Based on these requirements, we design and implement FaceSense, an ear-mounted device with two sensing modalities: *thermal images* and *physiological signals* (impedance and EMG). Thermal image sensors sense the thermal infrared radiated from our body to detect an approaching hand, while the physiological sensing detects the actual contact between the hand and face. These two sensing modalities aid each other by providing a variety of temporal and spatial information to accurately detect the face touch event and identify whether or not the touch

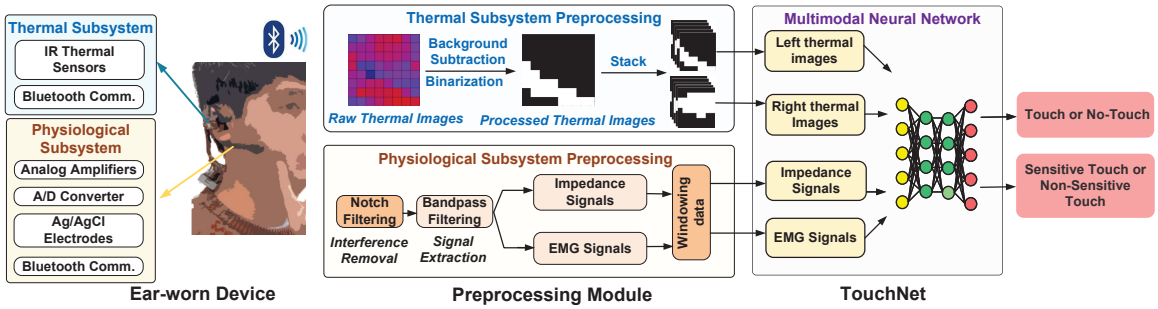


Fig. 2. FaceSense System Overview.

has landed at sensitive facial zones (i.e., eyes, mouth, nose), fulfilling the first two objectives of FaceSense design. Impedance and EMG signals are obtained from physiological sensors by applying appropriate band-pass filters of 1-10 Hz and >10 Hz respectively [78] after undergoing RF and power line interference removal via notch filtering [23]. Thermal images are captured through two miniature low-resolution thermal cameras. Subsequently the images undergo background subtraction and binarization (as detailed in Section 5). While we choose thermal image sensors for our current design, they can also be replaced by RGB image sensors and the rest of the system design still apply. In comparison, thermal image sensors are more privacy-preserving since they capture only the thermal infrared emitted by the body, rather than detailed color information with RGB cameras [70].

The overview of FaceSense is shown in Fig. 2. The ear-worn device includes an integrated thermo-physiological monitoring system that captures signals generated by face touch events and streams the collected signals to the host computing device via Bluetooth. The collected signals will be passed through pre-processing modules, that precisely window and stack the data, before feeding into the proposed multimodal deep neural network to detect face touch and identify the touch zone. The design of the multimodal deep neural network is detailed in Section 6.

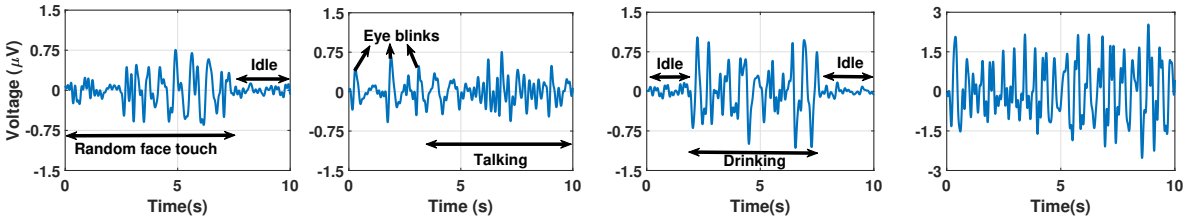
FaceSense system is designed to perform two main tasks:

- **Face Touch Detection.** Face touch event would cause micro skin surface deformations, which can be captured by the physiological sensors in the form of skin-electrode impedance variations. However, other daily activities (e.g., talking, drinking, eating) would also lead to similar skin surface deformations on faces. In practice, those activities could happen at the same time as a face touch event. To handle the complexity and provide more information specific to face touch, we add the EMG signal for analyzing the facial muscle movement and leverage low-resolution thermal sensing to capture the hand proximity.
- **Face Touch Zone Identification.** Since the distance between the touching point to the physiological sensors determines the magnitude of the physiological signals, touching different facial zones would lead to recognizable magnitude variance of the physiological signals. Intuitively, a thermal camera would also provide spatial hints on how the hand is oriented with respect to the face. We build on these two rationales to identify if a face touch event happens at the sensitive facial zones.

3.2 Challenges

Developing FaceSense presents challenges on three fronts:

Impedance Variations Are Buried Under the Noise Floor. On the front of physiological sensing, extracting skin-electrode impedance variances caused by face touch is challenging due to various confounding activities



(a) Face touch (b) Talk (c) Drink (d) Eat

Fig. 3. Impedance signals collected from the right side of the face.

and noises. Activities like talking (Fig. 3b), eating (Fig. 3c), and drinking (Fig. 3d) are facilitated by facial muscle movements that deform the attached skin tissue. Thus they lead to skin impedance signals similar to face touch. (Fig. 3a). Other activities like eye movements, head shakes also introduce formidable noises that make the isolation of face touch signature hard. When a human moves their eyes, a charge is induced in the physiological channel, generating signals in the skin impedance frequency range [78]. Head movements cause cable movements that lead to an increase in channel impedance similar to face touch. The influence of other well-documented artifacts/noises in electrograms, such as interference noises and electronic noises on touch artifacts, is an unexplored problem. To address this problem, we present asymmetry-based impedance and EMG signal analysis to detect face touch events which is discussed in Section 4.

Thermal Background Is Hard to Avoid. Detecting the approach of a hand is non-trivial using an extremely low resolution, low-power, off-the-shelf thermal cameras. The low-resolution makes it possible for background thermal objects to share similar thermal profiles as human hands.

As shown in Fig. 4, the thermal image of a hand touching a face (a) is similar to a heater (b) in front of a user when we set the heater temperature the same as the participant's hands. Although thermal cameras only capture the temperature information, which is privacy preserving, it also makes the detection of an approaching hand more challenging. When there exist other thermal objects in the background, the thermal profile of the hand (e.g., shape information) would be ruined by other thermal objects (Fig. 4c). To solve this problem, we present background detection and binarization techniques discussed in Section 5.

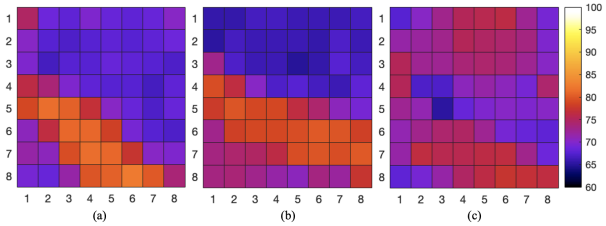


Fig. 4. (a) A hand in a clear background(F°). (b) A heater in a clear background(F°). (c) A hand and a computer(F°).

Designing Robust Heterogeneous Touch Sensing Algorithm Is Difficult. It is challenging to build a model that can effectively fuse heterogeneous sensors' information and generalize well to different users and environments. Though the two modalities are collecting data concurrently, sampling rates differ substantially between them. This adds further heterogeneity to the statistics extracted from different sensing modalities. Also, the collected sensors data carry substantial environment-specific (especially for thermal sensors) and user-dependent information (especially for physiological sensors), making it more challenging to design an face touch detection/localization algorithm that generates well for different users and environments. We present a custom-built deep learning algorithm that addresses these issues (See Section 6).

4 IMPEDANCE SENSING

In this section, we first present the rationale of how the skin impedance can be changed by face touch and how we can measure the impedance variation with the appropriate circuit arrangement. Then we explore two unique patterns of physiological signal for face touch detection and localization.

Impact of Touches on Facial Skin Impedance Variations. Skin impedance is often used to analyze the condition of the skin body such as skin hydration [20, 30], the thickness of the stratum corneum [54], wound monitoring [63], and so on. Face touch events create tiny deformations on the skin surface where electrodes are placed and generate skin impedance fluctuations, causing voltage fluctuations in the sensor readings. As shown in [76], at the frequency range of physiological signals (1-500 Hz), the effective contact area of an electrode increases as the pressure applied on the surface electrode increases, decreasing the impedance value. Similarly, during a face touch event, skin surface deformations cause a temporary decrease in the effective contact between surface electrode and skin, thereby increasing the value of skin impedance before returning to normal after the touch event concludes.

4.1 Detecting Impedance Variations Caused by Face Touch in Idle Settings

Tracking Impedance Variations Caused by Face Touch. To measure the skin impedance Z_{SE}, we use the *lead-off detection* mode of ADS1299 [17]. In this mode, a constant current source injects a small known current to the positive terminal of a physiological sensor channel. The current passes from the positive terminal to the current sink via Measurement Electrode(ME), human body, Ground Electrode(GE) and known resistors R_s (Fig. 5).

Z_{total} is the total impedance of the circuit calculated using: $Z_{total} = (V_{RMS})(\sqrt{2})\ Volts/\text{I}$, where V_{RMS} is the Root Mean Square of sliding windows of output voltage data after applying a bandpass filter from 1 – 30 Hz and I is the known injected current. The bandpass filter range was selected to capture the skin impedance signals which are known to be in the low frequency range of physiological signals [78]. The reference electrode is placed in proximity to the ear (electrode touching the Targus of the ear). The measurement electrode is placed at the junction of the upper and lower jaws (superficial masseter muscle surface). The arrangement is intended to record skin surface deformations of a touch event that might be happening anywhere on the right side of the face. Z_{SE} is calculated only from the ME as the current does not pass through the reference electrodes (RE) owing to a high input impedance (Fig. 5).

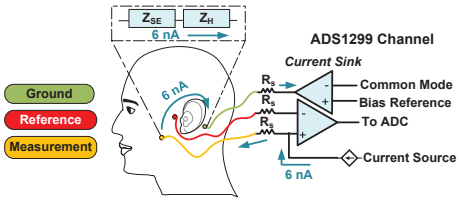
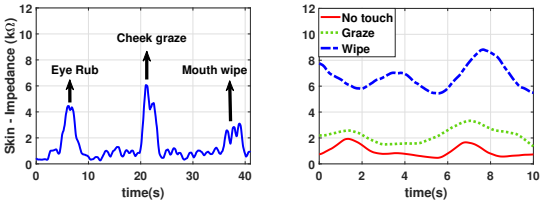


Fig. 5. Skin impedance measurement circuit.

With this arrangement, the impedance variation can be observed when a participant rubs their eyes, grazes their cheek and wipes their mouth, as shown in Fig. 6a. When a participant was asked to lightly graze their face and then wipe their face for a period of 10s, we observed that in comparison with no-touch, a light graze still has higher impedance measurement due to subtle skin-surface deformations as illustrated in Fig. 6b. Moreover, wiping action in the same face area has a bigger impedance change because of larger skin-surface deformations. We repeated the experiment 10 times, and similar results were observed. These preliminary results

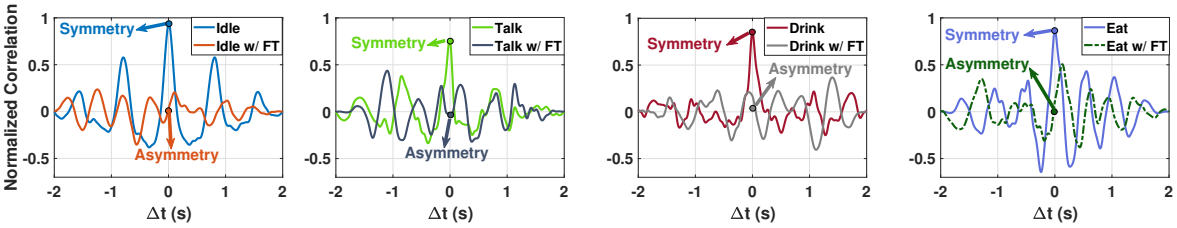


(a) (b)

Fig. 6. Skin impedance monitoring experiments.

confirm that impedance sensing can be used to capture tiny skin impedance variations caused by face touch events.

4.2 Detecting Face Touch in Practical Settings



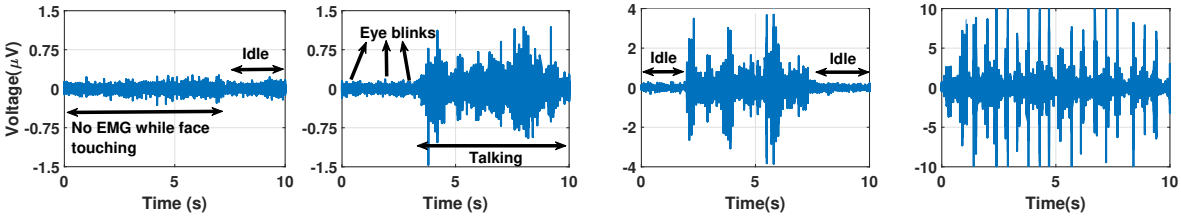
(a) Idle (1-10 Hz) (b) Talk (1-10 Hz) (c) Drink (1-10 Hz) (d) Eat (1-10Hz)

Fig. 7. Correlation of physiological sensor channels placed on two sides of the face while performing facial activities (i.e., talk, drink, eat) with and without face touch.

While the previous results are promising, the data was obtained when the user does not perform any activities (e.g., talking, eating, drinking, moving head). In practice, actions like head movements and eye movements also generate similar signature with face touch signals. To address this challenge, we utilize asymmetry of impedance signals and EMG analysis. Specifically, we present two unique patterns of the physiological signals that are captured from symmetric sensor locations on both sides of the face to detect face touch: strong asymmetry of impedance signals and non-existence of the EMG signal. The algorithm's intuition is simple. Face touch generates asymmetry in impedance signals, but they do not generate EMG signal, while, facial activities generate symmetric impedance signals together with symmetric EMG.
More specifically, most facial activities are facilitated by muscles on both sides of the face [27, 88]. Hence, they create symmetric skin surface deformations, whose impacts are felt equally on both sensor channels. face touch is observed to be done on one side of the face most of the time [58]. Hence, its effects are felt only on one of the sensor channels. Even if a face touch is performed on both sides of the face, the intensities and patterns are observed to be different. Fig. 7 illustrates the symmetric impedance signals of facial activities and the asymmetric impedance signals of facial activities when confounded with face touch activity in the form of normalized cross-correlation scores [14]. However, the EMG signals of the confounded activities stay highly correlated. This observation guides our neural network architecture design for extracting physiological features as detailed in Section.6.1.

Confirming the Non-Existence of EMG. Facial activities generate EMG proportional to the muscular activity (Figs. 8b, 8c, 8d). On the other hand, face touch does not generate EMG signals on the face since facial muscles are not facilitating it(Fig. 8a). As a result, the EMG behavior of a facial activity stays the same (symmetric on both facial sides) when confounded with face touch. To reliably detect face touch events, our algorithm captures the asymmetry of impedance signals from two channels using correlation and confirm the non-existence of EMG by implementing a layer of Depthwise Convolution [19]. This layer takes input from both the channels in impedance and EMG frequency range to learn effective correlation filters. The details of this implementation are presented in Section 6.1.

Localization of Face Touch. The physiological sensors capture face touch events at different facial areas (eyes, cheek, mouth etc.) with a magnitude that is inversely proportional to the distance between them. This is



(a) Face touch (b) Talk (c) Drink (d) Eat

Fig. 8. EMG signals collected from the same experiment for generating Fig. 3

because the skin deformations become more passive as the distance increases from the face touch event. The sensor arrangement is capable of differentiating touch events that happen at sensitive areas (eyes, nose, mouth) and non-sensitive areas (cheeks, side face etc.) of the face using this property. For example, a face touch at the cheek has stronger signal than a face touch at the eyes as eyes are located farther from the sensors. However this applies only when a user touches their face with one intensity. However, it was observed that humans touch their face in certain ranges of intensity depending on the stimulus of the touch (attention disruption, irritation, conversational response, stress etc.) [24]. Hence differentiating the face touch location solely based on magnitude of the impedance fluctuations can be inaccurate and we need more information to localize an face touch. The thermal subsystem provides useful information regarding a user's hand position and orientation relative to their face while performing face touch. Each user has a tendency of using a particular hand orientation while touching a certain facial area. Hence, thermal sensor data coupled with impedance data is effective to localize a face touch.

5 THERMAL SENSING

As discussed in the previous section, physiological sensing is able to handle the detection and touch zone identification in scenarios where users are idle or perform daily activities (e.g., talking, drinking) that would introduce unique physiological patterns. But in some activities which involve asymmetric facial muscle movement (e.g., sneering, chewing on one side), physiological sensing alone cannot produce accurate results. To make our system applicable in most daily activities, we added a thermal sensing module to obtain the proximity information of hands to the face, which can help to not only identify face touch events involving asymmetric facial muscle movement, but also provide spatial information for touch zone identification.

While the problem of hand sensing/tracking with wearable devices has previously been explored, we found that existing approaches fall short in one way or another in our case: vision-based methods [22, 44] require an on-body camera to stream the front facing video in real-time, which is obtrusive and power-hungry; acoustic-based sensing systems cannot differentiate between approaching hands and other items (e.g., an approaching spoon when drinking soup, an approaching book when reading).

We propose to leverage thermal sensing with two extremely low resolution (8×8 pixels) miniature thermal cameras. Thermal sensing leverages the infrared radiation from human beings, which is purely passive. The passive sensing rationale and the low resolution ensure that it is low-power and privacy-preserving. The small size of the camera allows them to be seamlessly embedded into an ear-worn device. However, at the same time, these properties also amplify the influence of the thermal background noise as illustrated in Section 3. In order to mitigate the influence of ambient thermal objects, our thermal image prepossessing technique leverages two unique observations: (1) **mobility:** when a hand approaches the face, it is moving into the view of the thermal

camera, which differs from static thermal objects in the ambient environment that are always in the scene. (2) **temperature:** Humans give off infrared radiation within a certain range (92°F – 99°F), which is usually stronger than the ambient indoor environment (68°F – 86°F)[83]. Although the temperature range of hands would change in different environments, they are stable in one environment. This feature can be used to separate them out from other background objects. We next elaborate more on how we leverage these two observations for extracting the hand-shape information out of raw thermal images.

5.1 Thermal Cameras Position Selection

Pre-experiment Consideration. Since the miniature thermal camera we used [1] has a limited viewing angle of 60° both horizontally and vertically, we need two cameras for capturing the left and right hands separately. Different camera positions capture different areas. With a suitable arrangement, the vertical 60° should be able to guarantee to capture both eyes and mouth. The horizontal view can be easily blocked by a user's face when the camera tilts towards face. In general, the closer the camera tilts to the face, more front area of the face is captured but also more occlusion is induced by the user's face.

Experiment on Camera Position. To explore the best thermal camera position for hand proximity detection, we built a headphone prototype with three adjustable pieces that can easily change the position and tilt angle of the thermal camera. We tested different combinations of locations and tilt angles. In each setting, we recruited two users to perform three daily activities: touching face, drinking, siting still. We found that , the best location is around a user's earlobe, which captures the most front area of the face while keeping the face out of the field of view of the camera. However, we found that even with the same thermal camera location and tilt angle, there are variations of the occlusion due to the shape difference of users' faces. Thus, it suggests a calibration stage for each user when wearing FaceSense for the first time.

5.2 Preprocessing of Thermal Images

To generate environment-independent handshape information, we conducted a Gaussian mixture model (GMM)-based preprocessing proposed by Zivkovic [92] on the captured thermal images to separate the moving object (foreground object) out of the ambient environment.

Background Subtraction and Binarization. According to our first observation, if we can extract the moving objects' thermal profile out, we can significantly mitigate the influence of the ambient thermal objects as shown in Fig. 4 (b) and (c). Detecting an intruding object in a static scene is a well studied case in vision community [28, 84]. A common assumption is that the images of the scene without the intruding objects exhibit some regular behavior that can be well described by a statistical model [92]. An applicable bottom-up approach is to assume the scene model has a probability density function for each pixel separately. An intruding object can be detected by finding out the pixels which do not fit the statistical model of the background scene.

In our case, when hands approach the face, they are moving into a relatively static ambient thermal scene. Therefore, we leverage the same idea to extract the all the moving thermal pixels out of the ambient thermal background in a thermal picture. However, thermal pixel values often have complex distributions rather than a fixed probability density function. We applied a more elaborate model proposed by [92] as follows. The thermal value of a pixel at time t in a thermal image is denoted by $x(t)$. Pixel-based ambient thermal background subtraction involves a decision of whether the pixel belongs to background (BG) or some foreground thermal object (FG). Bayesian decision D is made by [92]:

$$D = \frac{p(FG|x(t))}{p(BG|x(t))} = \frac{p(x(t)|FG) * p(FG)}{p(x(t)|BG) * p(BG)} \tag{1}$$

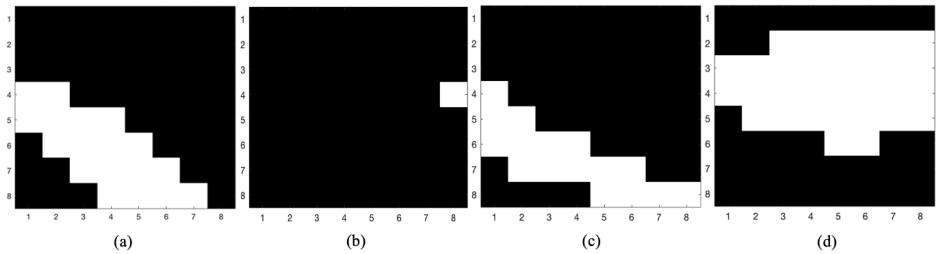


Fig. 9. The extracted binary shape image after applying the background subtraction and binarization to different thermal images: (a) A hand in a clear background (static) (b) A heater in a clear background (static) (c) A hand and a computer (static) (d) a user shaking head quickly in front of a computer (the same as moving a computer quickly in front of a user)

In general each pixel has the same probability of being FG or BG when we do not have other information. Thus we can set $p(FG) = p(BG)$. When $p(x(t)|FG) > p(x(t)|BG)$, we would determine this pixel as FG. Otherwise it belongs to BG. With our second observation that the temperature of human hands is usually in a stable range, we can assume $p(x(t)|FG))$ as a uniform distribution C_{FG}. And we can efficiently estimate the background model and adapt it to possible changes by the improved adaptive GMM proposed in [92]. Once the estimated background model is computed, we can determine if the pixel belongs to FG or BG. Then we do the binarization by setting all FG pixels as 1 and BG pixels as 0. Fig. 9 shows the result after applying this algorithm to the raw thermal images with different settings. We can see from the result that the binary image array only contains the shape information of the moving objects, which solves the problem of ambient thermal objects cluttering the thermal profile of human hands. We then feed the binary images array together with the physiological signals to the multimodal deep learning model for face touch detection and touch zone identification.

6 MULTIMODAL DEEP LEARNING MODEL

The processed physiological data (i.e., impedance and EMG) and thermal images are still heterogeneous. They are characterized by distinctive statistical properties, representation and correlation structures. As a result, it is difficult to systematically recognize useful cross-modality relationships in addition to modality-specific ones. Traditional machine learning algorithms either combine the features from each modality into a single feature vector (feature concatenation) or train separate classifiers on each modality and yield an overall classification (ensemble learning). As shown in literature [64], feature concatenation would easily overlook inter-sensor relationships with the number of explored feature combinations limited by the curse of dimensionality and ensemble learning is hard to find cross-sensor relationships due to the late-stage fusion.

Inspired by DeepMV [86] and domain-adversarial training [29], we propose TouchNet, a multimodal deep learning model for accurate face touch detection and touch zone identification with physiological signals and thermal images as inputs. TouchNet consists of three main modules: a **feature extractor**, a **touching detector** and a **user discriminator**. As shown in Fig. 10, the feature extractor would first extract the modality-specific representations separately and then integrate them together to fuse information between different modalities. Collectively, all neurons contribute to the learning of a joint representation of both sensing modalities. The learned joint representation is then fed to the touch detector for face touch detection and touch zone identification.

The thermal preprocessing removed most ambient environment information, leaving the thermal images mainly the foreground object shape information. Human handshapes are almost the same, producing similar

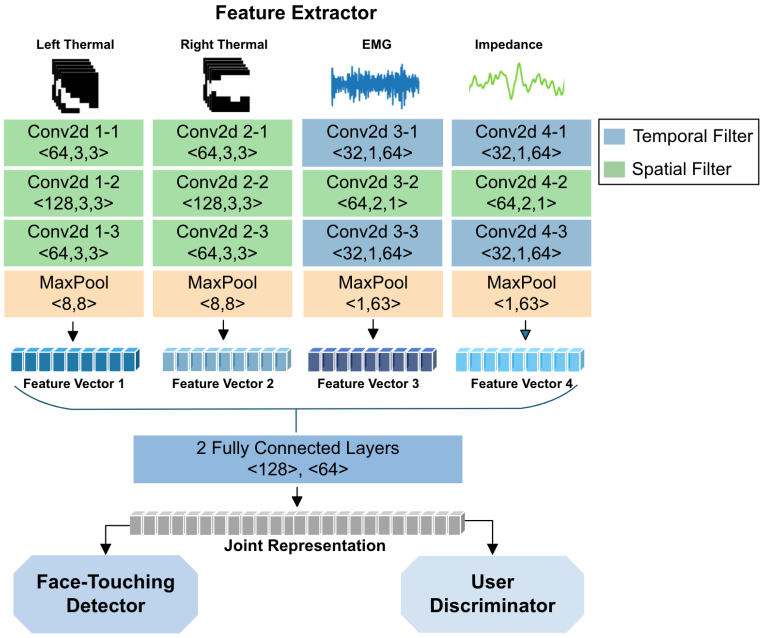


Fig. 10. Illustration of TouchNet. The processed thermal images, impedance signals and EMG signals first go through several CNN-based spatial/temporal filters for generating modality-specific features. Then they are merged together to learn a cross-modality joint representation, which would be fed into the following touching detector and user discriminator. The hyperparameters of Conv2d are shown as $< channel, kernel_x, kernel_y >$; Max-Pooling layer's kernel size is $< kernel_x, kernel_y >$; The Fully Connected layer's hidden units size is $< size >$.

thermal reading when touching faces. So it can be seen as a transferable feature across different users. However, the physiological is user-dependent (e.g., the magnitude of the physiological signals varies across different users). Therefore, to ensure the learned joint representation contains little user-specific information, we add a user discriminator module connected to the feature extractor via a gradient reversal layer[29]. In this way, we embed the user adaptation into the process of learning representation, which would enforce the feature extractor to learn a user-independent joint representation for better cross-user performance. We explain each module in detail in the following subsections.

6.1 Feature Extractor

We design two convolutional neural network (CNN)-based sub-network architectures for the thermal images and the physiological signals. Left and right thermal images share the same structure while impedance variance signal and EMG signal share the same structure.

Extracting Physiological Features. As discussed in Sec. 4, the raw physiological signals are processed into 2 channels of impedance variance signals and EMG signal. Taking the processed physiological signals as input, we choose a window size of 0.5 seconds with a 0.1-second moving step (The sampling rate of physiological signal is 250Hz, thus one input sample of each signal is a 2×125 vector.) Similar to EEGNet [46], we perform three convolutional steps in sequence. First, we feed the physiological time series to a 2d-CNN with the kernel size of (1, 64) with paddings to learn meaningful temporal features (The padding size is set as (0, 32) such that the

output maintains the same dimensions as the inputs.) In Sec.4.2, the symmetry analysis shows the importance of a correlation filter. Thus, the second 2d-CNN layer is a Depthwise Convolution [19] of size (2, 1), which can learn effective spatial/correlation filters. We add another temporal 2d-CNN with the same kernel size of (1, 64) to further learn the temporal filter for each spatial filter.

Extracting Thermal Features. The processed thermal input is a stack of 5 consecutive thermal images (The sampling rate of the thermal camera is 10Hz, thus 5 consecutive thermal images correspond to the same window size of 0.5s as one input sample of physiological signals.) CNN has been proved to perform extremely well in extracting the features in images[22, 41, 48, 72], so the thermal sub-network basically consists of three 2d CNN layers to extract effective spatial features. Since the input image resolution is extremely low (8×8), a human hand usually would occupy half the field of view. We set the kernel size of all the 2d CNN layers to be (3,3) as used in most literature for image feature extraction. At the end we add a max pooling layer, which would perceive the whole thermal image.

Merging Feature Vectors. After extracting four feature vectors separately from 4 sub-networks, we concatenate them together and feed to two fully connected layers (with the hidden units size of 128 and 64.) These layers fuse features extracted from different modalities together and output a joint representation for the following touch detector and user discriminator modules.

For both sub-network architectures, we add a batch normalization layer after each convolutional layer to stabilize the training process and prevent overfitting. For each fully connected layer, we add the batch normalization layer and a dropout layer to introduce non-linearity and prevent overfitting. ReLU[31] is used as the activation function in the model.

6.2 Touching Detector

Taking the joint representation from the feature extractor, the touch detector module is to determine whether it represents a face-touching event (binary classification) and further classify the touching event into sensitive touch or non-sensitive touch (facial zone identification; 3-class classification). This module consists of three stack fully connected layers (wit hidden size of 64, 32 and 2/3). The final output is used by the LogSoftmax function to compute the probabilities of belonging to each class (x_i). Since our dataset is highly unbalanced, we used the weighted cross-entropy loss as our loss function:

$$Loss_1 = -\frac{1}{N}\sum_{k=1}^{N}\frac{\sum_{i=1}^{m} W_i \log(\frac{e^{x_i}}{\sum_{j=1}^{m} e^{x_j}})}{\sum_{i=1}^{m} W_i}, \tag{2}$$

where W_i, N and m are the weight of i-th class, the number of samples with labels and the number of classes, respectively. All the weights are inversely proportional to the class frequencies.

6.3 User Discriminator

In order to improve the cross-user performance, we apply a domain adaptation technique proposed by [29]. The original purpose of [29] is to leverage unlabeled data from the target domain to learn domain-independent features from the source domain, so it is called unsupervised domain adaptation. Our purpose is to learn user-independent features, and since we already had the user label, we can build a similar supervised user discriminator.

The user discriminator takes the joint representation as input data, the user index as the label, and predicts which user this joint representation is coming from. The user discriminator has a similar neural network architecture as the touching detector: two layers of fully connected layers (with the hidden units size of 32 and 11). And the loss

function is also a cross-entropy function with equal weight for each user:

$$Loss_2 = -\frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{m}\log(\frac{e^{x_i}}{\sum_{j=1}^{m}e^{x_j}}), \tag{3}$$

where N and m are the number of samples with labels and the number of users, respectively. Ideally we want the accuracy of the user discriminator to be low, which means that the joint representation extracted from the feature extractor contains little user-specific information. To achieve this, we added a gradient reversal layer proposed by [29] as the first layer of the user discriminator, which directly passes the joint representation to the discriminator network during the forward process but multiplies the gradient by a certain negative constant λ during the backpropagation stage. Then the whole training process of TouchNet aims to minimize the face-touch classification loss plus the user discriminator loss ($Loss_1 + Loss_2$).

7 FACESENSE WEARABLE DESIGN

This section goes into detail about the design of FaceSense prototype, which takes all the previously mentioned noise factors as well as user compatibility into consideration. The two main units of the prototype are: (a) Flexible Earpieces and (b) Sensing Circuit. The earpieces house the thermal cameras and the physiological sensing channels while the sensing circuit consists of the bio-amplifier, thermal sensors and the Bluetooth module.

7.1 Flexible Earpieces

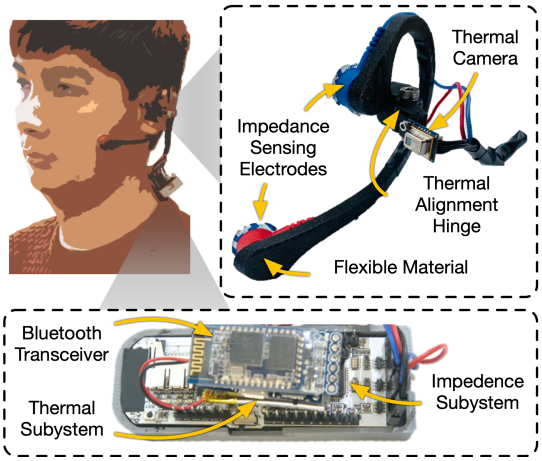


Fig. 11. FaceSense's Prototype

The central piece of the wearable system are the earpieces that house the electrodes and the thermal cameras. There are two earpieces going on each side of the face. The main criteria for the design of the earpiece are: (a) User comfort and social acceptability (b) Positioning and contact quality of the electrodes (c) Positioning of the thermal camera. The design of the earpiece is shown in Fig. 11.

First, the earpiece has a center that enters the ear-canal analogous to earphones/earbuds available in the market. This part helps to affix the earpiece without rotation throughout the usage of the wearable device and it also holds the thermal camera in place. Second, the earpiece has a portion that goes around a user's ear (Ear-cling). It is designed to be elastic to hold the electrodes in place for a wide demographic of population. There is also an elastic bend introduced into the ear-cling (face-cling). Face-cling is necessary to maintain the contact quality of the MEs despite being used on wide range of facial geometries. The initial angle of the face-cling is 30° which is derived from our experiments on participants with lean face geometries. As a result, face-cling angle would reduce when used on other facial geometries, ensuring good contact-quality. Finally, we mount the thermal sensor on the earpiece with a rotatable hinge. This provides both pitch and yaw adjustments of the thermal sensor, allowing manifold face shapes/geometries for a single earpiece.

The earpieces are 3D printed using a thermoplastic urethane. The material is flexible and lower in density, making the earpieces soft, adjustable and lightweight. The earpieces are made such that putting them on and removing them can be done with ease and the use of the wearable is comfortable over extended periods of time.

7.2 Sensing Circuit

Impedance sensing. We use a bio-amplifier circuit with ADS1299, a multi-channel Σ-Δ ADC, at its center [17]. Each channel consists of three electrode connections to the body: a) Measurement Electrode (ME), b) Reference electrode (RE), and c) Ground/Bias Electrode (GE). The ADC consists of built in Low-noise Programmable Gain Amplifiers (PGAs) for each channel that multiply the voltage difference between the ME and RE before the Analog to Digital conversion. The GE is connected to a current sink that removes the common-mode body voltage signals from the channels. Anatomical address of the placements of ME and RE can be given as: Junction of upper and lower jaws (superficial masseter muscle surface) and Temporal bone surface (touching the tragus part of the ear) respectively. One such channel is placed on each side of the face, totalling 4 electrode connections and one common GE placed on the skull surface at the back of the right ear (due to low neurological activity at that location). This arrangement covers skin deformations on the entire face. We use customized Ag/AgCl surface electrodes due to their high conductivity and reliable contact quality.

Thermal sensing. To minimize the size of the thermal subsystem, we utilize a TinyCircuits WirelingZero MCU [3]. This MCU implements an Atmel SAMD21 processor in a tiny, 15 mm x 32 mm package. Our I2C ports are implemented on the MCU using the SAMD21's various SERCOMS. For our thermal sensors, we use two Panasonic AMG8833 thermal sensors [1] which interface with the SAMD21 over I^2C. To wirelessly transfer the thermal data to the host computer, we reconfigure one of the I^2C ports as a serial interface. We connect a single HC-06 Bluetooth transceiver [2] to this serial port, and power the entire system using a small, 70 mAh battery.

8 SYSTEM EVALUATION

In this section, we evaluate the FaceSense's performance for face touch detection and touch zone identification. We first describe the experiment setups for data collection and annotation of the data streams. Then we introduce several face touch detection models as baselines for comparison. After that we show FaceSense's performance with various evaluation metrics (i.e., detection sensitivity and reliability, comfort level, and robustness.)

8.1 Setup and Data Collection

Participants and Environments. We recruited 14 participants with various backgrounds from our local institution (9 male, 5 female, age 27±6). During the data collection session, the participants wore the FaceSense device while being seated in a typical office setup with various ambient backgrounds. The data collection sessions took place over a period of 50 days in 3 different office environments. The room temperature during the sessions ranged from 65°F - 75°F, which is typical for the indoor office setting [83]. Thermal energy emitters such as displays, laptops, phone screens, room heaters, lights that affect the data from thermal modality were present in the surroundings. The surroundings also have regular electronic and RF interference noises arising from WiFi and electronic devices, which affect the data from the impedance modality.

Procedure. Once participants were given an overview of the research and sign the consent form, they enter the data collection area and were instructed to wear the FaceSense wearable. Then an instructor helped participants calibrate the physiological sensor locations and the orientation of thermal cameras, which takes around 1 minute. We also set up a web camera in front of participants to record videos as ground truth. Then participants were asked to perform face touch events every 20 seconds under 4 common daily scenarios (idle, talking, eating, drinking). We collect 5-minute data for each scenario:

- Idle: Participants listen to music or watch a video on their phone.
- Talking: Participants converse with the instructor or read an article out loud.
- Eating: Participants eat a food item that they are comfortable with.
- Drinking: Participants drink two cups of water.

No other instruction was provided at any point. Participants can perform face touch in various ways (i.e., touching both sensitive facial zones (eyes, nose and mouth) and non-sensitive zone with different magnitudes.) as they perform in daily life and can move freely while being in the video frame. The sample rate of thermal stream is 10 FPS, while physiological signal sample rate is 250 Hz. The ground truth video is captured at 30 FPS. After the preprocessing, we obtained around 15-minutes effective data from each user. In total, the face touch dataset contains 210 minutes (15×14), which corresponds to 3150k physiological data points and 126k thermal frames.

Annotation. To minimize the error in face touch labeling, physiological data (with highest sampling rate) was annotated by trained annotators. The annotation process has two steps: (1) frame-level labeling: Since physiological signals have a higher frequency, we label them using the video to get higher resolution labels for each data point first. A data point is labeled as a touch if the hand is contacting with the facial skin at the corresponding video frame. Touch events are further classified into sensitive and non-sensitive touches when the hand is in contact with mucous (eyes, nose, mouth) and non-mucous areas of the face respectively. (2) training-sample-level labeling: As described in Section 6, one training sample corresponds to 0.5 seconds, which consists of 125 points of physiological data. We annotate one sample as sensitive touch when there are more than 25 data points (0.1 seconds) are labeled as sensitive touch; annotate it as non-sensitive touch when it is not a sensitive touch sample but consists of more than 25 frames of non-sensitive touch; the rest samples are annotated as non-touch. In total, we have 12.6k samples: 68.4% no-touch, 16.1% non-sensitive touch, and 15.5% sensitive touch.

8.2 Baseline Models and Implementation.

We compare our system's performance with the following baselines:

Multilayer Perceptron (MLP). One straightforward approach to use multimodal data in a deep model is concatenating the raw sensor streams right at the input layer. The rest deep learning model consists of several stacked fully-connected layers. This Vanilla neural network is also called MLP. In our implementation, we set the number of fully-connected layers as 7 with hidden units of 512, 512, 256, 128, 64, 32, 3, respectively.

TouchNet with Single Inputs. To show the necessity of the thermal modality, we also tested the performance of our model with only physiological inputs. Specifically, we only feed feature vector 3 and 4 shown in Fig. 10 into the following fully connected layers to learn the joint representation.

TouchNet w/o Adversarial. In order to show the effectiveness of the adversarial training for handling cross-user performance, we simplify TouchNet by removing the user discriminator module as another baseline. This is also the final model used in personalized FaceSense. In this model, the cost function is only the weighted cross-entropy loss shown in Equation. 2.

We used PyTorch [62] to implement all deep learning models, and trained the models using three GeForce RTX 2080 Ti. The batch size is set to be 64. For all models, we used an Adam optimizer with a dynamic learning rate schedule which started from 0.0005 and repeatedly reduced to half once the validation accuracy did not improve after 5 epochs. We set the epoch number to 50, with an early stop number as 30. The parameter size of TouchNet is 4.48 MB.

8.3 Evaluation Protocols

8.3.1 Touch Detection. For repetitive behavior control and facial hygiene applications, a binary detection of whether a hand is actually touching the face achieves the purpose. We examine the binary touch detection performance with 14 participants by applying leave-one-user-out cross-validation. Each time we train the model, we leave one user's data out as testing data and then rotate the dataset. We use recall, precision and F1 score (all

are macro average) as the classification performance metrics, which represent the sensitivity (capture all face touch events/less false negatives), reliability (less false positives) and overall performance, respectively.

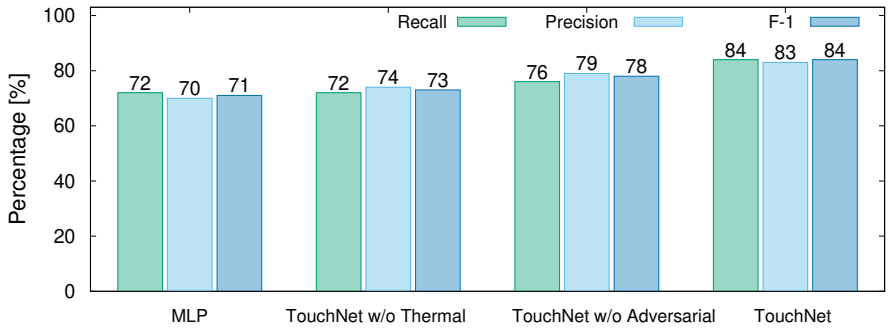


Fig. 12. Average leave-one-user-out face touch detection results across all participants.

Comparison with Baselines. Fig. 12 presents the average performance across all participants. The average cross user recall, precision and F-1 score of TouchNet with both modalities inputs are 84%, 83%, 83%, which shows that our system is accurate and robust to various environments and different users. TouchNet outperforms MLP in all three metrics by a clear margin, which verified the effectiveness of the feature extractor module. TouchNet is at least 10% higher than TouchNet without thermal input in all three metrics. This confirmed the necessity of adding the thermal modality. Comparing to TouchNet without the adversarial training, TouchNet has around 5% improvement on each metric. Therefore, the user discriminator indeed helped to learn a representation containing less user-specific information.

Error Analysis. While TouchNet achieves a reasonable detection accuracy, there are some cases leading to wrong classification results. (1) False Positives. False positives mainly arise from drinking and eating scenarios where a hand is proximate to the face but does not touch the face and at the same time facial skin is deformed. For example, when a user deforms their facial skin with their tongue (impedance asymmetry, no facial EMG) while keeping a hand proximate to the face (hand presence detection), which would be classified as a touch. (2) False Negatives. There are three main causes of false negatives. The first is the limited field of view of the thermal camera: when participants touch their lower jaw area (under the lips) with only one or two fingers, it cannot be captured by the thermal camera. The second cause is the imperfect background subtraction: When a user holds their hand still around the face for a while, the hand would be treated as background. Then when they touch their face, the event would not be detected. Thirdly, there is an area in the mid face region leading from the nasal bridge to the forehead region, where faint touches do not register an impedance signal. This happens as the cartilage density confines the skin deformations to a small area that does not reach the physiological sensors around the ear.

Comparison with Existing Face Touch Detection Methods. A recent work [68] developed an app that alerts users when they are about to touch their face by using an ultrasound signal emitted by earphones. Similar to other work, it is for detecting the approaching of hands rather than actual face touch. We experimented with the released online version of Saving Face [91] as an example to show the difference between our system and hand approaching detection systems. We carefully set up and calibrate both the Saving Face system and our FaceSense system as instructed and recruited one participant to wear both systems separately. Then the participant performs the same face touch tasks twice under the same 4 scenarios (idle, talking, eating, drinking)

Table 1. face touch detection performance with SavingFace and FaceSense (generic model).

	SavingFace [91]			**FaceSense**		
Scenarios	Recall	Precision	F1	Recall	Precision	F1
Idle	0.437	1.00	0.608	0.896	0.912	0.903
Talking	0.500	1.00	0.667	0.842	0.851	0.846
Drinking	0.750	0.667	0.700	0.822	0.811	0.814
Eating	0.714	0.612	0.659	0.814	0.801	0.807
Overall	0.538	0.736	0.621	**0.847**	**0.842**	**0.844**

as our data collection procedure described in Section 8.1 for 12 minutes (3 minutes for each scenario. We planned to collect the same duration [10 minutes each] of data as our dataset, but the released version of SavingFace kept sending out a high-frequency audible sound, and the participant was feeling uncomfortable after 3 minutes.) As shown in Table. 1, the overall F-1 score of the released version of SavingFace is 62.1% (This released version is different from what is used in [68] but the latest available version.) and our system (trained with other 14 users' data) achieves an F-1 score of 84.4%. Specifically, SavingFace's precision in idle and talking scenarios is 100% but at the same time it missed half of the face touch events with a recall lower than 50%. While in drinking and eating scenarios, SavingFace misclassified more than half of the drinking and eating motions as face touch. FaceSense performs consistently across all scenarios (F1 scores are always higher than 80%).

8.3.2 Touch Zone Identification. For diseases like COVID-19, it would be better if we are not only able to know if hands touch the face but also if they touched sensitive areas. We use micro-accuracy and macro-F1 score as the touch zone identification performance metrics. (A micro-average aggregates the contributions of all classes to compute the average metric and a macro-average computes the metric independently for each class and then takes the average. In multi-class classification problems, micro-precision, micro-recall, micro-F1 score, and micro-accuracy are always the same.) In an unbalanced dataset, the micro metric indicates more how the classifier perform on the majority class of the dataset while the macro F1 score can better reflect the model's classification performance on each class.

Table 2. Average leave-one-user-out face touch zone identification results.

Models	Micro-Accuracy	Macro-F1
MLP	0.582	0.561
TouchNet w/o Thermal	0.611	0.621
TouchNet w/o adversarial	0.642	0.646
TouchNet	**0.684**	**0.701**

Generic Model. We begin also with examining the leave-one-user-out cross-validation performance for touch zone identification, which corresponds to generic models. As shown in Table. 2, while TouchNet still outperforms all the rest models, the Macro-F1 score is 0.701. The uniqueness of the signals generated when touching different facial areas comes from the magnitudes of the impedance signal and the hand position/orientation. There are two potential reasons why cross-user touch zone identification performance is not as good as face touch detection: (1) Different participants have different skin fat densities, which would influence the magnitude of the physiological signals. (2) Participants tend to touch sensitive areas of the face in specific hand orientations. Since our dataset

is relatively small, the training data is likely to miss the skin and behaviour profile of the left-out users. We acknowledge this as one of the limitations and discuss the potential ways to build a generic model with more details in Section 10.

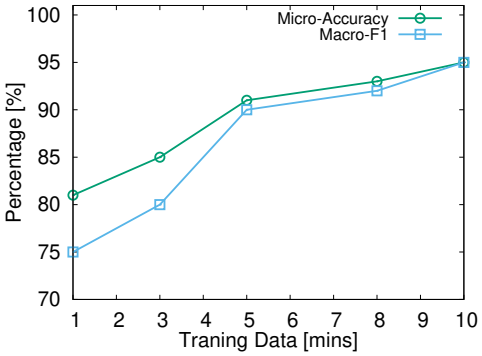


Fig. 13. Average personalized model performance with different training overhead.

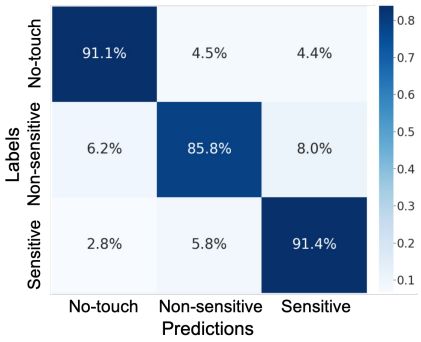


Fig. 14. Confusion matrix with 5-min training.

Personalized Model. As the generic model's F1 score is below 80%, we explored building a personalized model for each participant. As we are training a personalized model, we deactivate the user discriminator module (TouchNet w/o Adversarial is the final model). One critical aspect of a personalized model is the training overhead. We thus use different portions of a participants data for training and testing to explore the trade-off between the training overhead and zone identification accuracy. Fig. 13 shows the average personalized model performance across all participants with different training overhead. It is as expected that the performance keeps rising with the increase of the training data . We can see from the figure that the increase speed is obviously lower after the training time reaches 5 minutes. The micro-accuracy and macro-F1 are 90.8% and 90.1% when taking 5-minute data as training data. Therefore, we think 5-minute strikes a good balance between training overhead and accuracy.

We then plot the confusion matrix with this strategy in Fig. 14. We can observe that TouchNet achieves slightly different performance in each category. Since no-touch occupied most in the dataset (69%), thus the better performance for no-touch also results in better micro-metric comparing to macro-metric, which is consistent with Fig. 13. Non-sensitive touch is relatively harder to classify among the three and the false positives are almost half no-touch and half sensitive touch. There are areas of the face which are ambiguous when localizing a touch (e.g., facial area close to nose and mouth). We also noticed that during the experiment, some participants would touch the chin area and area close to the ears. They are either out of the field of view of the thermal or saturate the thermal camera due to the close distance.

Table 3. Average face touch zone identification results with personalized models.

Models	Micro-Accuracy	Macro-F1
MLP	0.664	0.644
TouchNet w/o adversarial	**0.908**	**0.901**

Lastly, we compared the TouchNet (w/o adversarial) touch zone identification performance with MLP using 5-minute training data for each user. As shown in Table. 3, TouchNet still achieves much better performance.

While we removed the adversarial user discriminator module, the feature extractor module can extract more representative features than the MLP model. The MLP model performs worse with the potential reason of not enough training data, which also reflects that our feature extractor design guided by our experiment findings indeed helped the feature learning process.

8.3.3 Qualitative Evaluation. As a wearable device, the wearability and comfort level are important features. So we asked participants to evaluate our system's wearability and comfort level with a short questionnaire. Additionally, participants also wrote down their comments and suggestions on the system at the end. The questionnaire asks participants to rate on the standard 5-point Likert scale: (1) Strongly Disagree, (2) Disagree, (3) Neutral, (4) Agree, and (5) Strongly Agree to the following statements:

- Q1: I was comfortable wearing the device.
- Q2: I like the experience of FaceSense device overall.

We average the score of each question cross all participants. FaceSense scored 3.8 out of 5 for the overall using experience (Q2). FaceSense obtained 3.5 out of 5 on comfort level (Q1), which is a moderate performance. Most participants agree that the overall user experience is similar to wearing a normal earphone but the one user commented in the open question that the prototype is a bit heavy for long-time wearing. In addition, there were 4 out of 10 participants felt the prototype is rigid and not soft enough. Another potential reason that participants were not fully satisfied with the current prototype is that we only provided one size. There were 3 out of 10 participants mentioned that the prototype cannot perfectly fit their ears. They all have relatively small ears, which suggested that we need further improve the adjustable earpiece design or fabricate it with different sizes. We acknowledge this as one of our limitations and discuss it with more details in Section. 10.

8.4 Robustness Analysis

We also conduced controlled experiments to examine system robustness against body movement and touches in mid-face regions. For both studies, we recruited one participant wearing the FaceSense prototype.

8.4.1 Sensitivity to Body Movements. To assess system robustness against body movements, we instructed the participant to perform all the steps in the experiment procedure while brisk walking and moving. We observed that the touch detection had a precision, recall and F-1 score of 82%, 84% and 83% respectively. This result proves that FaceSense is robust against brisk walking and other less rapid body movements, thanks to the prototype design that ensures good contact quality of the surface electrodes. However, when a participant performed the experiments while running, the touch detection's precision, recall and F-1 score dropped to 63%,62% and 62% respectively. There are some potential reasons behind this performance. First, high relative motion between the face and the wearable device introduces formidable motion artifacts [82] into physiological sensor channels. Second, thermal cameras deviate from their calibrated positions due to prototype vibrations to thrust forces, which adversely affect the field of view. Third, the background subtraction is less effective due to rapidly changing/evolving environments. Most of these problems can possibly be solved by exploring the integration of accelerometer into the system, which can assess and negate motion artifacts.

8.4.2 Mid-Face Touches. To examine system performance for mid-face touches, we instructed the participant to perform feather touch on the face with a significance on the nasal bridge to forehead area as we identified this region to be vulnerable to light touches owing to the cartilage density and distance from the physiological sensor channels. The touch detection had a precision, recall and F-1 score of 71%, 72% and 72% respectively while using the leave-one-user-out evaluation. However, we consider this to be unrepresentative of the usual face touch behavior which includes all intensities of touch at all regions of the face. Possible hardware improvements for this phenomenon could be to move the sensor channels closer to the mid-face region while preserving the usability and social acceptability by using transparent electrodes and connections. We could also improve the results by

recruiting more participants to perform this controlled experiment to have a more representative dataset. We plan to pursue these tasks in a future work.

8.5 Power Consumption

We measure the energy consumption of FaceSense using a power monitoring device (Monsoon High Voltage Power Monitor [57]). All measurements were conducted at 60°F when the system was powered by a 3.7V Lithium ion battery. The power consumed by each component of the system is showed in Fig. 15. More specifically, in the idle mode, when the physiological subsystem is not transmitting data via Bluetooth, the power consumption is 105.4 mW on average. In the transmission mode, where the physiological subsystem transmits data via Bluetooth, the power consumption is 210.4 mW on average. The bio-amplifier in the subsystem currently uses traditional Bluetooth to transmit data to the pre-processing module reliably at high data rate. Similarly the power consumption for thermal subsystem in the idle mode is 48.3 mW, while in transmission mode it is 78.8 mW on average. The total system consumes 289.2 mW on average in the working mode. This amounts to a total run time of 6 hours 23 minutes on a 3.7 V 500 mAh Lithium ion battery which was used for our experiments. The power consumption of FaceSense can be reduced significantly by using the BLE transceiver of the thermal subsystem as the prime mode of data transmission. This optimization would get rid of the power budget allotted for the transceiver module of the physiological subsystem and improve the battery life of the device to 10 hours 2 minutes. We plan to pursue the power optimization task in a future work.

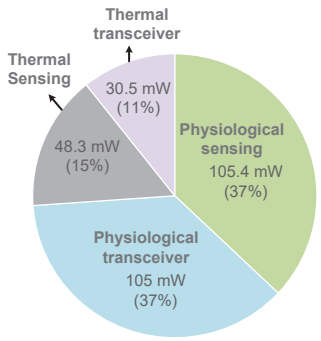


Fig. 15. Power distribution of FaceSense

9 RELATED WORK

Approaches to Detect Face Touch Events. Researchers approached the problem of face touch detection in many ways. These approaches can be grouped into 3 main categories:

(1) *Motion sensors* such as accelerometers, magnets etc. have been the major focus of wearable sensing devices in detecting face touch events [4, 15, 25, 56]. These sensors are usually integrated into bracelets or smart watches to identify passive signature of face touching events and develop a warning system to prevent this movement in the wake of Covid-19 pandemic. These approaches usually suffer from false positives generated by hand movements similar to face-touching events (drinking water, eating, picking up a phone call etc.) and interference noises from everyday electronic devices such as computers, smartphones and general household appliances [15, 25]. However, for the wrist-based approaches, the device should be worn on both hands to have optimal feedback from the device. This problem makes it more uninspiring for these techniques to be implemented on a smartwatch.

(2) *Wireless sensing* have been utilized to detect the face touch event based on the interference of hand movement and the transmission signal. In [68], researchers use a SONAR inspired approach that sends ultrasonic waves using off-the-shelf earbuds and receives the waves reflected from an approaching hand using a microphone. This approach suffers many false positives from surrounding obstacles like (smartphones, laptops etc.) and similar hand movements that don't involve face-touching.

(3) *Camera*-based face touch classification models using MS Azure's Custom Vision [5], Tensorflow were built using transfer learning and deep learning techniques [6, 35, 59]. In [12], a large set of audio-visual recordings were collected and labelled for touching and no-touching events to have over 2 million images of classification data. The data is fed into a fine-tuned ResNet152 model to achieve 84% accuracy. All camera-based methods

are constrained by lighting conditions, camera positioning, and the need to constantly run the camera, which raises privacy concerns and power consumption. None of the works in this domain reports the sensitivity of their models to such constraints. Most of these camera-based models expectedly use a significant amount of processing power and time [6, 35, 59].

Thermal-based Sensing Techniques. Pyroelectric infrared sensing is a common technique used to detect the presence of humans or trigger alarms. Passive Infrared (PIR) sensors have also been explored for human localization and tracking[51, 87], thermal imaging[50], gesture recognition[32]. Commercial PIR sensors are typically tuned for human detection by adding a bandpass filter window, which convert the thermal radiation into an electrical current when a thermal object presents. Most prior work focused on detecting large body movements or micro finger gestures without the ability to differentiate the presence of different thermal objects. Recently thermal image classification has been explored for object classification [67, 81], material type recognition [18], and pedestrian detection [33]. Thermal-image-based classification works efficiently without the need for ambient light sources and consume relatively less power and computational overhead when compared to the RGB images. With only one channel information, however, the granularity and the detection accuracy would decrease. Most of the prior works in this space leverage thermal camera with high resolution, which are too big to be integrated on an earable device and consume too much energy for a small wearable device. In our work, we explore to use the extremely-low resolution thermal camera (8×8 pixel) for detecting the presence of human hands.

Skin Impedance and Electrograms. Electrograms are biosignals including EEG, ECG, EMG, EOG, HRV, etc. are used widely in health monitoring. Various researchers use EEG and EOG to detect brain disorders such as Epilepsy [40, 45, 49, 77], Dementia [60, 66], Stroke [65, 79, 80], etc. In these works, the abnormalities of biosignals are recognized and indicate the state of human brain. Biosignals are also applied for human computer interaction [10, 61, 75]. The signals are generated by moving eyes, tongue, blinking, grinding teeth. Skin impedance, like biosignals, is commonly used for monitoring the condition of human body. Electrical impedance spectroscopy is proposed to detect skin cancer cell [89]. They proposed a method to sense the electrical properties of the cells via impedance sensing devices and distinguish between normal and cancer cells. Radio frequency skin tightening treatments leverage skin impedance to analyze radio frequency energy [36]. The treatments need to maintain the constant energy via the balance of current applied on skin and its impedance. In our work, EMG and skin impedance are leveraged to detect events of face touch.

10 DISCUSSIONS AND FUTURE WORK

User Study. We recognize that our current user study is limited by the small user group and short-term user experiences, because of the difficulty of recruiting participants during the pandemic. In future work, we will expand our user group to include users with a wider age range and more diverse background (e.g., different skin fat densities) and test the robustness of FaceSense with a longer-term study. With a dataset collected from a larger user group, we potentially can also improve the cross-user accuracy for face touch zone classification.

Advanced Machine Learning Models. Another direction of future work is to seek more advanced machine learning models to lower the labeling overhead of training data and improve the performance of a generic model to eliminate personalized training. In particular, to avoid the tedious labeling process, we can explore semi-supervised learning (SSL) [9, 13], a classical sub-field of machine learning [9, 13]. SSL provides an effective way of leveraging unlabeled data to improve a model's performance. Specifically, we can apply techniques proposed in recent works that leveraged data augmentation to (1) propagate label information from a small amount of labeled data to unlabeled data [21] and (2) construct similar and dissimilar data for training [11, 73]. To reduce the overhead of personalized training, we can consider applying meta-learning [39] and few-shot

learning [16] so that each user only needs to provide a minimal amount of data (e.g., a few seconds of data) to fine-tune the model quickly.

Usability. The physical design of FaceSense can be optimized to improve its comfort level for long-term wear. We can further miniaturize the form factor of FaceSense by implementing the physiological subsystem using transparent electrodes and connections as well as using customized fixtures for thermal cameras instead of adjustable hinges. A smaller ear-worn system is also less intrusive and will be more socially acceptable. Additionally, we are interested in integrating FaceSense with existing headsets/earbuds. From the user study, some users also suggested that the system can be integrated into headbands or caps to reduce the ear fatigue.

Thermal Sensing as a Trigger. In the current implementation, the thermal and physiological module are both always-on to simultaneously receive two sensor data streams for multimodal sensing. A future improvement is to leverage thermal sensing as a trigger so that only when the thermal sensing detects an approaching hand, the physiological sensing module starts to record data. It will enable the system to be used as a hand-to-face proximity detection system in addition to the current functionality of detecting hand-to-face contact solely. To realize the trigger, we will need to minimize the latency of starting the physiological sensing module, which currently requires a 1-second 'wake up' time to calibrate and output accurate reading. We will also need to improve the accuracy of the thermal sensing component in detecting the approaching hand. It currently detects hands approaching with 81% accuracy in various ambient background, which is not ideal for a trigger.

11 CONCLUSION

In this work, we designed, implemented, and evaluated FaceSense, an ear-worn system to detect and differentiate face touches in sensitive/mucosal zones (i.e., eyes, nose, and mouth) from other facial zones. FaceSense leverages low resolution thermal cameras to sense the thermal infrared signal emitted by approaching hands and physiological sensors to monitor the skin deformation cased by touches. After preprocessing, both thermal images and physiological signals are fed to the proposed deep learning model (TouchNet) to extract temporal and spatial features, detecting face touches and further identifying touch zones. Experiment results with 14 participants indicate FaceSense is a capable face touch detection system without any training overhead and can additionally provide touch zone identification with minimal training overhead. Unlike existing hand-face proximity detection solutions, FaceSense is the first system that detects and localizes actual hand-to-face contact, illustrating the potential for the system to prevent/reduce health and/or behavioral issues related to face touching activities.

ACKNOWLEDGMENTS

We sincerely thank the reviewers for their insightful comments that helped improve the paper. This work is in part supported by the National Science Foundation under CNS-1552924, CNS-1846541, SenSE-2037267, DGE-1840344 and Alfred P. Sloan Research Fellowship 2020. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the funding agencies or others.

REFERENCES

[1] [n. d.]. AMG8833 Thermal Sesnor. https://media.digikey.com/pdf/Data%20Sheets/Panasonic%20Sensors%20PDFs/Grid-EYE_AMG88.pdf. ([n. d.]). Accessed: 2021-05-10.

[2] [n. d.]. HC-06 Bluetooth Transceiver Module. http://wiki.sunfounder.cc/index.php?title=Bluetooth_Transceiver_Module_HC-06. ([n. d.]). Accessed: 2021-05-10.

[3] [n. d.]. WIRELINGZERO PROCESSOR. https://tinycircuits.com/products/wirelingzero. ([n. d.]). Accessed: 2021-05-10.

[4] 04/01/2020. Magnet-based face touch detection app for coronavirus: Use your phone's compass and a magnet on your wrist to detect face touching, including audio and vibration alerts. (04/01/2020). https://matter.childmind.org/face-guardian.html

[5] 2020. Custom Vision: An AI service and end-to-end platform for applying computer vision to your specific scenario. (2020). https://azure.microsoft.com/en-us/services/cognitive-services/custom-vision-service/

[6] April 2, 2020. How YOU Can Use Computer Vision to Avoid Touching Your Face! (April 2, 2020). https://medium.com/microsoftazure/how-you-can-use-computer-vision-to-avoid-touching-your-face-34a426ffddfd

[7] Nathan H Azrin, R Gregg Nunn, and SE Frantz. 1980. Treatment of hairpulling (trichotillomania): a comparative study of habit reversal and negative practice training. *Journal of Behavior Therapy and Experimental Psychiatry* 11, 1 (1980), 13–20.

[8] Felix Barroso, Norbert Freedman, and Stanley Grand. 1980. Self-touching, performance, and attentional processes. *Perceptual and Motor Skills* 50, 3_suppl (1980), 1083–1089.

[9] Mikhail Belkin, Irina Matveeva, and Partha Niyogi. 2004. Regularization and Semi-supervised Learning on Large Graphs. In *Learning Theory*, John Shawe-Taylor and Yoram Singer (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 624–638.

[10] Guillermo Bernal, Tao Yang, Abhinandan Jain, and Pattie Maes. 2018. PhysioHMD: a conformable, modular toolkit for collecting physiological data from head-mounted displays. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers (ISWC '18)*. Association for Computing Machinery, New York, NY, USA, 160–167. https://doi.org/10.1145/3267242.3267268

[11] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin Raffel. 2019. Mixmatch: A holistic approach to semi-supervised learning. *arXiv preprint arXiv:1905.02249* (2019).

[12] Cigdem Beyan, Matteo Bustreo, Muhammad Shahid, Gian Luca Bailo, Nicolo Carissimi, and Alessio Del Bue. 2020. Analysis of Face-Touching Behavior in Large Scale Social Interaction Dataset. In *Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20)*. Association for Computing Machinery, New York, NY, USA, 24–32. https://doi.org/10.1145/3382507.3418876

[13] Avrim Blum and Tom Mitchell. 1998. Combining Labeled and Unlabeled Data with Co-Training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory (COLT' 98)*. Association for Computing Machinery, New York, NY, USA, 92–100. https://doi.org/10.1145/279943.279962

[14] John R Buck, Michael M Daniel, and Andrew C Singer. 1997. *Computer explorations in signals and systems using MATLAB*. Prentice-Hall, Inc.

[15] Xiang 'Anthony' Chen. 2020. FaceOff: Detecting Face Touching with a Wrist-Worn Accelerometer. (2020). arXiv:cs.HC/2008.01769

[16] Yinbo Chen, Xiaolong Wang, Zhuang Liu, Huijuan Xu, and Trevor Darrell. 2020. A new meta-baseline for few-shot learning. *arXiv preprint arXiv:2003.04390* (2020).

[17] Chip. April 10, 2014. OpenBCI: Measuring Electrode Impedance. (April 10, 2014). http://eeghacker.blogspot.com/2014/04/openbci-measuring-electrode-impedance.html

[18] Youngjun Cho, Nadia Bianchi-Berthouze, Nicolai Marquardt, and Simon J. Julier. 2018. Deep Thermal Imaging: Proximate Material Type Recognition in the Wild through Deep Learning of Spatial Surface Temperature Patterns. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173576

[19] François Chollet. 2017. Xception: Deep Learning with Depthwise Separable Convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1800–1807. https://doi.org/10.1109/CVPR.2017.195

[20] EJ Clar, CP Her, and CG Sturelle. 1975. Skin impedance and moisturization. *J Soc Cosmet Chem* 26 (1975), 337–353.

[21] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116* (2019).

[22] N. H. Dardas and N. D. Georganas. 2011. Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques. *IEEE Transactions on Instrumentation and Measurement* 60, 11 (2011), 3592–3607. https://doi.org/10.1109/TIM.2011.2161140

[23] Scott Day. 2002. Important factors in surface EMG measurement. *Bortec Biomedical Ltd publishers* (2002), 1–17.

[24] Kyra Densing, Hippokrates Konstantinidis, and Melanie Seiler. 2018. Effect of stress level on different forms of self-touch in pre-and postadolescent girls. *Journal of motor behavior* 50, 5 (2018), 475–485.

[25] N. D'Aurizio, T. L. Baldi, G. Paolocci, and D. Prattichizzo. 2020. Preventing Undesired Face-Touches With Wearable Devices and Haptic Feedback. *IEEE Access* 8 (2020), 139033–139043. https://doi.org/10.1109/ACCESS.2020.3012309

[26] Paul Ekman and Wallace V Friesen. 1972. Hand movements. *Journal of communication* 22, 4 (1972), 353–374.

[27] Paul Ekman, Joseph C Hager, and Wallace V Friesen. 1981. The symmetry of emotional and deliberate facial actions. *Psychophysiology* 18, 2 (1981), 101–106.

[28] Nir Friedman and Stuart Russell. 2013. Image Segmentation in Video Sequences: A Probabilistic Approach. (2013). arXiv:cs.CV/1302.1539

[29] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-Adversarial Training of Neural Networks. 17, 1 (Jan. 2016), 2096–2030.

[30] Norm Gitis and Raja Sivamani. 2004. Tribometrology of skin. *Tribology Transactions* 47, 4 (2004), 461–469.

[31] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Deep Sparse Rectifier Neural Networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research)*, Geoffrey Gordon, David

Dunson, and Miroslav Dudík (Eds.), Vol. 15. PMLR, Fort Lauderdale, FL, USA, 315–323. http://proceedings.mlr.press/v15/glorot11a.html

[32] Jun Gong, Yang Zhang, Xia Zhou, and Xing-Dong Yang. 2017. Pyro: Thumb-Tip Gesture Recognition Using Pyroelectric Infrared Sensing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. Association for Computing Machinery, New York, NY, USA, 553–563. https://doi.org/10.1145/3126594.3126615

[33] T. Guo, C. P. Huynh, and M. Solh. 2019. Domain-Adaptive Pedestrian Detection in Thermal Images. In *2019 IEEE International Conference on Image Processing (ICIP)*. 1660–1664. https://doi.org/10.1109/ICIP.2019.8803104

[34] Jinni A Harrigan, John R Kues, John J Steffen, and Robert Rosenthal. 1987. Self-touching and impressions of others. *Personality and Social Psychology Bulletin* 13, 4 (1987), 497–512.

[35] Chad Hart. March 19, 2020. webrtcH4cKS: Stop touching your face using a browser and TensorFlow.js. (March 19, 2020). https://webrtchacks.com/stop-touching-your-face-with-browser-tensorflow-js/

[36] Yoram Harth and Daniel Lischinsky. 2011. A novel method for real-time skin impedance measurement during radiofrequency skin tightening treatments: Multisource radiofrequency skin tightening and skin impedance. *Journal of Cosmetic Dermatology* 10, 1 (March 2011), 24–29. https://doi.org/10.1111/j.1473-2165.2010.00535.x

[37] Megan R Heinicke, Jordan T Stiede, Raymond G Miltenberger, and Douglas W Woods. 2020. Reducing risky behavior with habit reversal: A review of behavioral strategies to reduce habitual hand-to-head behavior. *Journal of applied behavior analysis* 53, 3 (2020), 1225–1236.

[38] Joseph A. Himle, Deborah Bybee, Lisa A. O'Donnell, Addie Weaver, Sarah Vlnka, Daniel T. DeSena, and Jessica M. Rimer. 2018. Awareness enhancing and monitoring device plus habit reversal in the treatment of trichotillomania: An open feasibility trial. *Journal of Obsessive-Compulsive and Related Disorders* 16 (2018), 14–20. https://doi.org/10.1016/j.jocrd.2017.10.007

[39] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. 2020. Meta-learning in neural networks: A survey. *arXiv preprint arXiv:2004.05439* (2020).

[40] M. Shamim Hossain, Syed Umar Amin, Mansour Alsulaiman, and Ghulam Muhammad. 2019. Applying Deep Learning for Epilepsy Seizure Detection and Brain Mapping Visualization. *ACM Transactions on Multimedia Computing, Communications, and Applications* 15, 1s (Feb. 2019), 1–17. https://doi.org/10.1145/3241056

[41] 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., 1097–1105. https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[42] Maurice H. Krout. 1954. An Experimental Attempt to Produce Unconscious Manual Symbolic Movements. *The Journal of General Psychology* 51, 1 (1954), 93–120. https://doi.org/10.1080/00221309.1954.9920209 arXiv:https://doi.org/10.1080/00221309.1954.9920209

[43] Yen Lee Angela Kwok, Jan Gralton, and Mary-Louise McLaws. 2015. Face touching: a frequent habit that has implications for hand hygiene. *American journal of infection control* 43, 2 (2015), 112–114.

[44] O. Köpüklü, A. Gunduz, N. Kose, and G. Rigoll. 2019. Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks. In *2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019)*. 1–8. https://doi.org/10.1109/FG.2019.8756576

[45] Zakareya Lasefr, Ramasani Rakesh Reddy, and Khaled Elleithy. 2017. Smart phone application development for monitoring epilepsy seizure detection based on EEG signal classification. In *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*. IEEE, New York City, NY, 83–87. https://doi.org/10.1109/UEMCON.2017.8248992

[46] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. 2018. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering* 15, 5 (jul 2018), 056013. https://doi.org/10.1088/1741-2552/aace8c

[47] James F Leckman, Dorothy E Grice, James Boardman, Heping Zhang, Amy Vitale, Colin Bondi, John Alsobrook, Bradley S Peterson, Donald J Cohen, Steven A Rasmussen, et al. 1997. Symptoms of obsessive-compulsive disorder. *American Journal of Psychiatry* 154, 7 (1997), 911–917.

[48] Yann LeCun and Yoshua Bengio. 1998. *Convolutional Networks for Images, Speech, and Time Series*. MIT Press, Cambridge, MA, USA, 255–258.

[49] Jiseon Lee, Junhee Park, Sejung Yang, Hani Kim, Yun Seo Choi, Hyeon Jin Kim, Hyang Woon Lee, and Byung-Uk Lee. 2017. Early Seizure Detection by Applying Frequency-Based Algorithm Derived from the Principal Component Analysis. 11 (2017), 52. https://doi.org/10.3389/fninf.2017.00052

[50] Haili Liu, Ya Wang, Kevin Wang, and Hanbing Lin. 2017. Turning a pyroelectric infrared motion sensor into a high-accuracy presence detector by using a narrow semi-transparent chopper. *Applied Physics Letters* 111, 24 (2017), 243901. https://doi.org/10.1063/1.4998430 arXiv:https://doi.org/10.1063/1.4998430

[51] Xuefeng Liu, Tianye Yang, Shaojie Tang, Peng Guo, and Jianwei Niu. 2020. From Relative Azimuth to Absolute Location: Pushing the Limit of PIR Sensor Based Localization. In *Proc. of MobiCom*. Association for Computing Machinery, New York, NY, USA, Article 1, 14 pages. https://doi.org/10.1145/3372224.3380878

[52] Christine Lochner, Annerine Roos, and Dan J Stein. 2017. Excoriation (skin-picking) disorder: a systematic review of treatment options. *Neuropsychiatric disease and treatment* 13 (2017), 1867.

[53] Melika Lotfi, Michael R Hamblin, and Nima Rezaei. 2020. COVID-19: Transmission, prevention, and potential therapeutic opportunities. *Clinica chimica acta* (2020).

[54] Fei Lu, Chenshuo Wang, Rongjian Zhao, Lidong Du, Zhen Fang, Xiuhua Guo, and Zhan Zhao. 2018. Review of stratum corneum impedance measurement in non-invasive penetration application. *Biosensors* 8, 2 (2018), 31.

[55] Charles W McMonnies. 2008. Management of chronic habits of abnormal eye rubbing. *Contact Lens and Anterior Eye* 31, 2 (2008), 95–102.

[56] Allan Michael Michelin, Georgios Korres, Sara Ba'ara, Hadi Assadi, Haneen Alsuradi, Rony R Sayegh, Antonis Argyros, and Mohamad Eid. 2021. FaceGuard: A Wearable System To Avoid Face Touching. *Frontiers in Robotics and AI* 8 (2021), 47.

[57] Inc. 2021 Monsoon Solutions. 2021. High Voltage Power Monitor. (2021). https://www.msoon.com/high-voltage-power-monitor

[58] Stephanie Margarete Mueller, Sven Martin, and Martin Grunwald. 2019. Self-touch: contact durations and point of touch of spontaneous facial self-touches differ depending on cognitive and emotional load. *PloS one* 14, 3 (2019), e0213677.

[59] Raphael Mun. July 13, 2020. Face Touch Detection with TensorFlow.js Part 1: Using Real-Time Webcam Data With Deep Learning. (July 13, 2020). https://www.codeproject.com/Articles/5272773/Face-Touch-Detection-with-TensorFlow-js-Part-1-Usi

[60] Katrin Müller, Stephanie Fröhlich, Andresa M. C. Germano, Jyothsna Kondragunta, Maria Fernanda del Carmen Agoitia Hurtado, Julian Rudisch, Daniel Schmidt, Gangolf Hirtz, Peter Stollmann, and Claudia Voelcker-Rehage. 2020. Sensor-based systems for early detection of dementia (SENDA): a study protocol for a prospective cohort sequential study. *BMC Neurology* 20, 1 (Dec. 2020), 84. https://doi.org/10.1186/s12883-020-01666-8

[61] 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*. ACM, Munich Germany, 269–282. https://doi.org/10.1145/3210240.3210322

[62] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.). Curran Associates, Inc., 8024–8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[63] Xiachuan Pei, Hao Jin, Shurong Dong, Dong Lou, Lie Ma, Xingang Wang, Weiwei Cheng, and Hei Wong. 2019. Flexible wireless skin impedance sensing system for wound healing assessment. *Vacuum* 168 (2019), 108808.

[64] Valentin Radu, Catherine Tong, Sourav Bhattacharya, Nicholas D. Lane, Cecilia Mascolo, Mahesh K. Marina, and Fahim Kawsar. 2018. Multimodal Deep Learning for Activity and Context Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 157 (Jan. 2018), 27 pages. https://doi.org/10.1145/3161174

[65] Rajiv Ranganathan, Rui Wang, Rani Gebara, and Subir Biswas. 2017. Detecting Compensatory Trunk Movements in Stroke Survivors using a Wearable System. In *Proceedings of the 2017 Workshop on Wearable Systems and Applications*. ACM, Niagara Falls New York USA, 29–32. https://doi.org/10.1145/3089351.3089353

[66] M. Raza, M. Awais, W. Ellahi, N. Aslam, H.X. Nguyen, and H. Le-Minh. 2019. Diagnosis and monitoring of Alzheimer's patients using classical and deep learning techniques. *Expert Systems with Applications* 136 (Dec. 2019), 353–364. https://doi.org/10.1016/j.eswa.2019.06.038

[67] C. D. Rodin, L. N. de Lima, F. A. de Alcantara Andrade, D. B. Haddad, T. A. Johansen, and R. Storvold. 2018. Object Classification in Thermal Images using Convolutional Neural Networks for Search and Rescue Missions with Unmanned Aerial Systems. In *2018 International Joint Conference on Neural Networks (IJCNN)*. 1–8. https://doi.org/10.1109/IJCNN.2018.8489465

[68] Camilo Rojas, Niels Poulsen, Mileva Van Tuyl, Daniel Vargas, Zipporah Cohen, Joe Paradiso, Pattie Maes, Kevin Esvelt, and Fadel Adib. 2021. A Scalable Solution for Signaling Face Touches to Reduce the Spread of Surface-Based Pathogens. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 31 (March 2021), 22 pages. https://doi.org/10.1145/3448121

[69] Steven Schlosser, Donald W Black, Nancee Blum, and Rise B Goldstein. 1994. The demography, phenomenology, and family history of 22 persons with compulsive hair pulling. *Annals of Clinical Psychiatry* 6, 3 (1994), 147–152.

[70] Md. Haidar Sharif. 2021. Laser-Based Algorithms Meeting Privacy in Surveillance: A Survey. *IEEE Access* 9 (2021), 92394–92419. https://doi.org/10.1109/ACCESS.2021.3092687

[71] Ramin Shiraly, Zahra Shayan, and Mary-Louise McLaws. 2020. Face touching in the time of COVID-19 in Shiraz, Iran. *American Journal of Infection Control* 48, 12 (2020), 1559 – 1561. https://doi.org/10.1016/j.ajic.2020.08.009

[72] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. (2015). arXiv:cs.CV/1409.1556

[73] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *arXiv preprint arXiv:2001.07685* (2020).

[74] Melinda A. Stanley and Suzanne G. Mouton. 1996. *Trichotillomania Treatment Manual*. Springer US, Boston, MA, 657–687. https://doi.org/10.1007/978-1-4899-1528-3_17

[75] Tanuja Subba and Tejbanta Singh Chingtham. 2019. A Survey: EMG Signal-Based Controller for Human–Computer Interaction. In *Advances in Communication, Cloud, and Big Data (Lecture Notes in Networks and Systems)*, Hiren Kumar Deva Sarma, Samarjeet Borah, and Nitul Dutta (Eds.). Springer, Singapore, 117–125. https://doi.org/10.1007/978-981-10-8911-4_13

[76] Bahareh Taji, Adrian DC Chan, and Shervin Shirmohammadi. 2018. Effect of pressure on skin-electrode impedance in wearable biomedical measurement devices. *IEEE Transactions on Instrumentation and Measurement* 67, 8 (2018), 1900–1912.

[77] Kaat Vandecasteele, Thomas De Cooman, Jonathan Dan, Evy Cleeren, Sabine Van Huffel, Borbála Hunyadi, and Wim Van Paesschen. 2020. Visual seizure annotation and automated seizure detection using behind-the-ear electroencephalographic channels. *Epilepsia* 61, 4 (April 2020), 766–775. https://doi.org/10.1111/epi.16470

[78] A.T. Vehkaoja, J.A. Verho, M.M. Puurtinen, N.M. Nojd, J.O. Lekkala, and J.A. Hyttinen. 2005. Wireless Head Cap for EOG and Facial EMG Measurements. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. 5865–5868. https://doi.org/10.1109/IEMBS.2005.1615824

[79] Alexander von Luhmann, Jessica Addesa, Sourav Chandra, Abhijit Das, Mitsuhiro Hayashibe, and Anirban Dutta. 2017. Neural interfacing non-invasive brain stimulation with NIRS-EEG joint imaging for closed-loop control of neuroenergetics in ischemic stroke. In *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, Shanghai, China, 349–353. https://doi.org/10.1109/NER.2017.8008362

[80] Kyle B Walsh. 2019. Non-invasive sensor technology for prehospital stroke diagnosis: Current status and future directions. *International Journal of Stroke* 14, 6 (Aug. 2019), 592–602. https://doi.org/10.1177/1747493019866621

[81] W. Wang, J. Zhang, and C. Shen. 2010. Improved human detection and classification in thermal images. In *2010 IEEE International Conference on Image Processing*. 2313–2316. https://doi.org/10.1109/ICIP.2010.5649946

[82] John G Webster. 1984. Reducing motion artifacts and interference in biopotential recording. *IEEE transactions on biomedical engineering* 12 (1984), 823–826.

[83] Wikipedia contributors. 2021. Room temperature — Wikipedia, The Free Encyclopedia. (2021). https://en.wikipedia.org/w/index.php?title=Room_temperature&oldid=1016698070 [Online; accessed 5-May-2021].

[84] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. 1997. Pfinder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 7 (1997), 780–785. https://doi.org/10.1109/34.598236

[85] Chaojun Xie, Hongjun Zhao, Kuibiao Li, Zhoubin Zhang, Xiaoxiao Lu, Huide Peng, Dahu Wang, Jin Chen, Xiao Zhang, Di Wu, et al. 2020. The evidence of indirect transmission of SARS-CoV-2 reported in Guangzhou, China. *BMC public health* 20, 1 (2020), 1–9.

[86] Hongfei Xue, Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shiyang Wang, Ye Yuan, Shuochao Yao, Aidong Zhang, and Lu Su. 2020. DeepMV: Multi-View Deep Learning for Device-Free Human Activity Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 34 (March 2020), 26 pages. https://doi.org/10.1145/3380980

[87] D. Yang, W. Sheng, and R. Zeng. 2015. Indoor human localization using PIR sensors and accessibility map. In *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*. 577–581. https://doi.org/10.1109/CYBER.2015.7288004

[88] Uldis Zarins. 2018. *Anatomy of Facial Expressions*. Exonicus, Incorporated.

[89] Fan Zhang, Tongyu Jin, Qingqing Hu, and Pingang He. 2018. Distinguishing skin cancer cells and normal cells using electrical impedance spectroscopy. *Journal of Electroanalytical Chemistry* 823 (Aug. 2018), 531–536. https://doi.org/10.1016/j.jelechem.2018.06.021

[90] Junbo Zhang and Swarun Kumar. 2020. NoFaceContact: Stop Touching Your Face with NFC. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services (MobiSys '20)*. Association for Computing Machinery, New York, NY, USA, 468–469. https://doi.org/10.1145/3386901.3396603

[91] camilorq Zhi Wei Gan. 2020. Saving Face App. https://github.com/camilorq/SavingFaceApp. (2020).

[92] Z. Zivkovic. 2004. Improved adaptive Gaussian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, Vol. 2. 28–31 Vol.2. https://doi.org/10.1109/ICPR.2004.1333992