A Scalable Solution for Signaling Face Touches to Reduce the Spread of Surface-based Pathogens.

CAMILO ROJAS, MIT Media Lab, USA
NIELS POULSEN, EPFL, Switzerland
MILEVA VAN TUYL, MIT Media Lab, USA
DANIEL VARGAS, University of Costa Rica, Costa Rica
ZIPPORAH COHEN, Wellesley College, USA
JOE PARADISO, MIT Media Lab, USA
PATTIE MAES, MIT Media Lab, USA
KEVIN ESVELT, MIT Media Lab, USA
FADEL ADIB, MIT Media Lab, USA

Fig. 1. Working principle of Saving Face, a mobile technology that employs ultrasound signals to detect and alert a user when they touch their face.

Hand-to-Face transmission has been estimated to be a minority, yet non-negligible, vector of COVID-19 transmission and a major vector for multiple other pathogens. At the same time, as it cannot be effectively addressed with mainstream protection measures, such as wearing masks or tracing contacts, it remains largely untackled. To help address this issue, we have developed Saving Face - an app that alerts users when they are about to touch their faces, by analyzing the distortion patterns in the ultrasound signal emitted by their earphones. The system only relies on pre-existing hardware (a smartphone with generic earphones), which allows it to be rapidly scalable to billions of smartphone users worldwide. This paper describes the

Authors' addresses: Camilo Rojas, camilorq@media.mit.edu, MIT Media Lab, USA; Niels Poulsen, niels.poulsen@epfl.ch, EPFL, Switzerland; Mileva Van Tuyl, mvantuyl@wellesley.edu, MIT Media Lab, USA; Daniel Vargas, daniel.vargasdiaz@ucr.ac.cr, University of Costa Rica, Costa Rica; Zipporah Cohen, zc1@wellesley.edu, Wellesley College, USA; Joe Paradiso, joep@media.mit.edu, MIT Media Lab, USA; Pattie Maes, pattie@media.mit.edu, MIT Media Lab, USA; Kevin Esvelt, esvelt@media.mit.edu, MIT Media Lab, USA; Fadel Adib, fadel@media.mit.edu, MIT Media Lab, USA.

2474-9567/2021/3-ART31
https://doi.org/10.1145/3448121

design, implementation and evaluation of the system, as well as the results of a user study testing the solution's accuracy, robustness, and user experience during various day-to-day activities (93.7% Sensitivity and 91.5% Precision, N=10). While this paper focuses on the system's application to detecting hand-to-face gestures, the technique can also be applicable to other types of gestures and gesture-based applications.

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

Additional Key Words and Phrases: behavior change, wearables, signal processing, machine learning, mobile devices

ACM Reference Format:
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

1 INTRODUCTION

By January 28, 2021, more than 100 million people had contracted COVID-19, almost half in the last 2 months alone, and over 2.1 million had lost their lives [10]. While vaccine development has been rapidly advancing, it will likely be many months before immunization solutions become available to the broad population. In the meantime, containing transmission and guarding against the next pandemic are paramount.

The CDC and researchers analyzing infection clusters estimate that hand-to-face transmission accounts for 5-10% of COVID-19 cases [46] due to the virus' ability to survive on commonly used surfaces for multiple days [41]. At the same time, vulnerable populations and front-line workers who cannot work remotely, continue to use public transport, and may not be able to frequently sanitize their hands, find themselves particularly at risk.

While a minority transmission vector, unlike person-to-person contact, hand-to-face transmissions cannot be effectively mitigated by using masks, or detected through contact tracing, thus remaining largely untackled [6]. Ever since the beginning of the pandemic, institutions such as the WHO have advised us to avoid touching our faces. Yet, this is easier said than done. Most of us touch our faces frequently throughout the day, usually without thinking about it, making it a very difficult habit to break and requiring a significant amount of conscious effort.

Using Awareness Enhancement Devices (AEDs) in the form of wearables (e.g., necklaces, watches) that notify users when they are about to touch their faces could potentially help. Studies performed in the context of other behavioral changes, such as Trichotillomania (i.e., compulsive hair pulling) [14] and self-harm in people with mental disabilities [22] have demonstrated they can be effective at bringing a marked and lasting reduction in the harmful behavior. In order to have a significant impact, AEDs would nonetheless need to: i) effectively detect hand-to-face touching in a wide variety of environments, ii) seamlessly integrate with day-to-day activities and iii) reach as many people as possible. Nonetheless, existing AEDs typically rely on custom hardware, making them difficult to rapidly manufacture and largely inaccessible to the broad population.

To help address these limitations, we propose Saving Face - a system that tracks hand-to-face movements and alerts users when they touch their faces, while only relying on their smartphones and a set of generic headsets (see Figure 1). At a high level, Saving Face leverages the reflections of acoustic signals off a user's hand in order to track hand-to-face gestures. It analyzes distortion patterns in the non-audible ultrasound signal emitted by earphones to detect hand-to-face gestures and alert the user with a vibration or an audible nudge. Saving Face's functionality can be summarized in the following 4 steps:

(1) When the user launches the Saving Face App,[1] an audio file is played, enabling the connected earset to emit an ultrasound signal through the left earphone that is then continuously captured by the embedded microphone. In this way, the earset is practically activated to operate as a sonar.

[1]Note that Saving Face can also operate as a background app.

(1) (2) (3)

Fig. 2. To use Saving Face, a user (1) downloads and opens the app, (2) attaches the earphones as per the guide in the app and then (3) receives an alert (audio or vibration) when touching their face.

(2) When the user's hand approaches their face, the ultrasound signal bounces off the hand creating distinctive distortion patterns, which are then captured by the microphone.
(3) The Saving Face algorithm analyzes the signal distortions with the help of custom-made signal processing and machine learning algorithms, to decide if the distortion can be attributed to a face-touching gesture.
(4) If the pattern is deemed characteristic of a face-touching gesture, the user receives feedback in the form of either a vibration or an audio nudge. By exposing the user to repeated feedback, through a prolongued use of the app, we hope to trigger a learning effect and ultimately reduce their frequency of face touching.

Figure 2 summarizes the experience of a Saving Face user. Beyond saving the users from acquiring new hardware, we also seek to facilitate on-boarding by making the user experience as simple and seamless as possible. The Saving Face app has a minimalistic, user-friendly interface, as shown in Figure 2.1, only requiring the user to press one intuitively positioned button to turn Saving Face on and then leave it running in the background.

The two most significant challenges in developing Saving Face are (1) achieving a high detection accuracy while only using inaudible acoustic (or ultrasonic) signals, and (2) ensuring the app can be used while the user is moving, while limiting false positives due to motion interference. Moreover, Saving Face must overcome these challenges while relying entirely on existing smartphones and earsets, i.e., without requiring any hardware or firmware modifications and without rooting the phone. This is necessary in order to build a scalable and user-friendly solution for mitigating pathogen surface transmissions. To achieve this, Saving Face incorporates the following techniques:

(1) In order to avoid having to make adjustments to users' hardware, we had to rely on the ultrasound signals that the earphones are already able to emit. To avoid annoying the user, we only used signals outside the audible range. However, operating in the inaudible range using off-the-shelf earsets restricts Saving Face to a small bandwidth of less than 2 kHz (18 - 20 kHz in the frequency spectrum). Unfortunately, such a limited bandwidth constrains the resolution of wireless sensing [1, 27], making it difficult to detect face touches with high accuracy. To address this, Saving Face implements two complementary techniques and fuses their information: wideband (Frequency Modulated Carrier Wave - FMCW) and narrowband (Doppler). FMCW gives a rich set of features from the spectral analysis (as it is normally used to detect distances), while Doppler provides speed indicators. The combination resulted in a significant improvement in the detection accuracy, from 81.2% and 89.4% (F1-Score for only Doppler and only FMCW) to 94.1% (F1-score for FMCW and Doppler together).
(2) For wearers to be able to use the system during day-to-day life, Saving Face needs to be robust to movements, while at the same time distinguishing small hand-to-face movements from motion interference. To address

this, we have built a machine learning model to help distinguish the variety of distortion patterns typical of face-touching gestures from unrelated movements and perturbations.

Saving Face has been prototyped and tested in user studies emulating real-life activities such as sitting on a chair, walking around the room, or stacking cans. It has shown an average sensitivity of 93.7% (correctly detected face touches / total face touches) and an average precision (number of nudges due to a face-touch / total number of nudges) of 91.5%. At the same time, most users have stated that they would be willing to use the App in real life.

Given its high detection accuracy, relatively seamless user experience (can be used in the background even when the user listens to music), and high scalability (it leverages hardware the user already owns and the App can easily be downloaded from the online stores), we believe Saving Face can become an effective, widely adopted solution for mitigating the surface transmission of COVID-19 and other infectious diseases. At the same time, while our tests have mainly focused on hand-to-face applications, the solutions we have developed can be applicable to multiple other gesture recognition issues, where a set of speakers and a microphone are available. This could span to areas such as behavior-reversal therapy (e.g., against compulsive hair pulling), security (e.g., device owner recognition), safety (e.g., falling asleep behind the wheel) or health monitoring (e.g., detecting movement patterns preliminary to a heart attack).

In the frame of this paper, we will describe 4 contributions that best summarize our work on Saving Face so far, namely:

(1) Introducing a novel and rapidly scalable approach to behavioral change to lower the hand-to-face transmission of COVID-19 and other pathogens.
(2) Designing the first acoustics-based system to track hand-to-face movements, based on signal processing (Frequency Modulated Carrier Wave - FMCW - and Doppler Shift) and machine learning (Logistic Regression) algorithms.
(3) Implementing the system in a set of user friendly iPhone and Android apps, available for anyone who owns a smartphone and a set of commercial earphones.
(4) Evaluating the system's performance in multiple experimental environments - to observe an average sensitivity of 93.7% and precision of 91.5%, as well as an overall positive user experience and adoption intention.

In the next sections, we will give additional details on the most relevant related work (Section 2), our main design choices (Section 3), signal processing (Section 4) and machine learning solutions (Section 5), platform implementation (Section 6) and user study results (Section 7). We will also discuss our current limitations and future research direction (Section 8) and highlight our main conclusions and acknowledgements (Sections 9 and 10 respectively).

2 RELATED WORK

Developing techniques for reducing face-touching has been an object of research in multiple fields, including epidemiology and behavioral change. Before COVID-19, studies were focused on limiting the spread of viral and bacterial diseases relying on hand-to-face contact as their main transmission vector [13], such as influenza [26]. The proposed techniques were also mainly aimed at training medical personnel to reduce face-touching [17], given their particularly high exposure to pathogens. Other use cases focused on the treatment of compulsive behaviors, such as Trichotillomania (i.e., hair pulling) [14], nail biting and thumb sucking in adults [4], smoking [29], acne-inducing touching [19] and self-harm in people with mental disabilities [22].

Awareness Enhancement Devices (AEDs), such as custom made wearables (e.g., wristbands, necklaces, etc.) that emit alerts as a hand approaches the face, have emerged as promising solutions, showing potential to lead to significant reductions in face and mouth touching [3] [4] [15] [35] [45]. However, to date, their use remains

relatively isolated and inaccessible to large audiences. This is primarily due to the fact that they involve custom hardware, which tends to be expensive and difficult to deliver, or require in-person support from specialized personnel, as well as laboratory settings that create a clear disconnect from daily use. Addressing these limitations has been a key objective of Saving Face, hence our striving to only leverage the capabilities of hardware that users already own and use in their day-to-day activities.

In the remainder of this section, we review in further detail the existing technologies that can be used to track hand-to-face movements through AEDs. We follow the categorization of previous work in three broad domains as established by [27], namely 1) Acoustic localization, 2) RF-based gesture systems and 3) Near-device interaction.

2.1 Acoustic Gesture Recognition

Acoustic Gesture Recognition systems use acoustic signals emitted by speakers and recorded by a microphone to obtain information from the movements of the human body. Examples include SoundWave [12], Dolphin [34], AudioGest [36], SonicOperator [21], Strata [47], FingerIO [27], CAT [23] and the work from Watanabe *et al.* [44] and Mao *el al* [24]. While these past systems have made significant advances in acoustic sensing, they cannot *simultaneously* satisfy two key design features of Saving Face: operating entirely with existing smartphones and headsets, and sensing small hand movements in the presence of significant mobility. In particular, users can wear the Saving Face gesture detection system (smartphone, earset) during their daily activities without being constrained to remain in the same place, while the previous examples require the gesture detection system to remain static during operation (e.g., placed over a table) and the user in its close vicinity. Our system thus addresses the challenge of delivering a high detection accuracy despite the movements of the system relative to the body (i.e., bouncing while walking) and having to deal with a significantly lower signal-to-noise ratio provided by the earphones (compared to the speaker/microphone).

2.2 Radio Frequency Based Gesture Systems

Gesture detection based on Radio Frequency (RF) relies on analyzing the reflection of electromagnetic waves on the human body. Examples of detection systems specific to hand gestures include: WiSee [32], AllSee [18], SideSwipe [49], and WiTrack [1]. These systems typically require the user to remain within the range of a deployed infrastructure (e.g., WiFi router or specialized radar) in order to track a user's gestures. As a result, they are less suitable for Saving Face's deployment scenario, where one needs to track users throughout their daily activities (including walking outdoors, in elevators, etc.)

2.3 Near Device Interaction

A number of technologies for gesture detection are available in a wearable format and seek to support mobility and integration with daily activities. For example, NoFaceContact [48] employs Near-Field Communication technology using a custom-made earpiece and wristwatch to warn users not to touch their faces. Pulse [25] also makes use of wearable sensors in the form of a pendant, which detects face touches. Immutouch [16] pairs a custom wristband equipped with an accelerometer and an Android and iOS app to detect and track face touches. The previous technologies require the user to acquire custom-made hardware. In contrast, Saving Face proposes a solution that leverages commonly available devices (smartphones and earphones).

There are several gesture detection solutions that are implemented as Smartwatch apps. For example, FaceOff [7] analyses accelerometer data to detect the face touch gesture. However, this approach only detects face touches from a single hand (the one wearing the smartwatch) and it is prone to false positives due to its operation in dead-reckoning mode (i.e., without a position reference, such as the face). No Face-touch [30] relies on the magnetometer of a smartwatch to detect proximity to a magnetic pendant. This solution can only detect face touches from one hand and requires additional hardware (the pendant). In comparison, Saving Face can detect face

touches from both hands with commonly used hardware. Additionally, smartphones are more widely available and accessible than smartwatches.

A non-contact alternative is Do Not Touch Your Face [5] that uses the computer webcam to detect face touches. This solution, however can only be used while the user is in front of the camera, and it not compatible with mobility. Moreover, it could potentially raise privacy concerns from users.

3 DESIGN

Prior work demonstrates Awareness Enhancing Devices can help reduce face-touching behaviors. However, existing technologies often show limited effectiveness and typically require attaching custom hardware to people's hands, wrists, and neck, limiting their convenience and accessibility, and thus likely their adoption potential. To overcome these challenges, we have focused the development of Saving Face on the following 3 design objectives:

- Effectiveness: accurately detect hand-to-face touching behavior and transmit feedback to the user.
- Scalability: only rely on hardware that users already posess, minimizing economic and practical barriers.
- Convenience: seamlessly integrate with user's regular activities and behaviors.

3.1 Selecting the Platform

Given their intrinsic ability to transmit and detect ultrasound and their widespread use, smartphones, together with earphones, were an intuitive starting point for our choice of platform. Earphones are capable of transmitting frequencies over 18 kHz (beginning of the ultrasound band [28]). While their frequency responses are usually poor outside a 20-30cm range, this range is sufficient for the detection of face touches when the earphone and microphone are placed near the face. At the same time, smartphones also allow a quick deployment of a software solution through dedicated Apps.

Additionally, smartphones and earphones are objects that users carry with them everywhere, all the time. This enhances Saving Face's effectiveness for reducing face touching compared to typical AEDs, enhances its ecological validity, and maximizes the time of use of the intervention. Relying on objects of daily use also makes the solution convenient since the users will not need to carry and charge additional devices. That said, it is important to minimize any interference with the regular use of the hardware, for example by operating in the background of the operating system and enabling the user to continue using the device for its primary purpose. Both these features are displayed by Saving Face.

Additionally, the earphones can be easily attached in multiple positions depending on the use context (see Figure 2). For example, a bank cashier might prioritize concealing the system to enhance social acceptance and might prefer a discreet position hanging from the ear, while a cargo-lift operator will have to comply to safety regulations stating that earphones cannot be used in the ears.

3.2 Designing the System

As shown in Figure 3, the Saving Face system consists of four main building blocks:

(1) The sonar system - The solution ultimately relies on transforming the user's smartphone-earset pair into a sonar system. When the user launches the Saving Face App, a WAV file is played and enables the connected earset to emit an ultrasound signal through the left earphone that is continuously captured by the embedded microphone. With the proposed configurations of the earphones, the left earphone is never positioned in the ear, so the ultrasound signal will never be playing directly into a user's ear. When the user's hand then approaches their face, the ultrasound signal bounces against the hand creating distinctive distortion patterns captured by the microphone

(2) The signal processing - The signal captured by the microphone is then processed with the help of FMCW and Doppler to extract information related to the perceived signal distortions and generate the spectrogram that will be fed into the gesture recognition machine learning algorithm
(3) The machine learning algorithm - Face-touching gestures appear to generate a unique signal distortion footprint. The Saving Face machine learning algorithm continuously analyzes the signal received by the microphone to identify patterns consistent with face-touching gestures that can then be fed back to the user interface
(4) The user interface - From the user's viewpoint Saving Face is a smartphone App that he or she activates and then leaves running in the background while using their smartphone and earphones. When the App detects a hand-to-face movement, it provides a nudge in the form of a vibration or audio signal. At the same time, the counter on the App displays the number of times the user has touched their face.

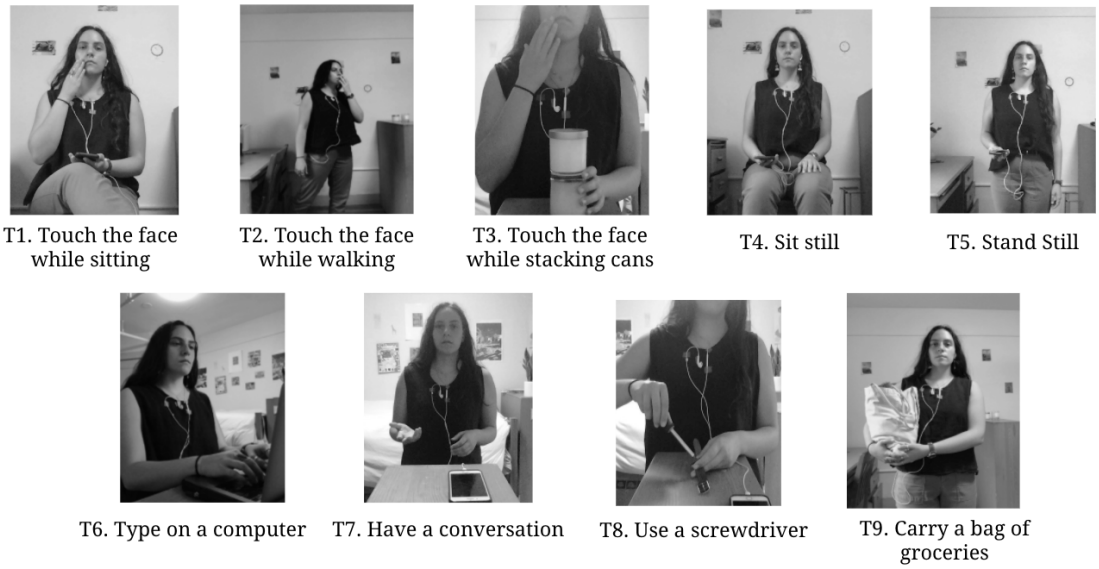


Fig. 3. The design and implementation of Saving Face center around four building blocks: (a) ultrasound sensing, which involves emitting an ultrasound signal from the earphones, (b) signal processing to compute the velocity and position of the moving hand, (c) gesture detection to identify a face touch event, and (d) user feedback to alert a user when a face touch occurs.

In the following two sections we will provide more details about the second and third building blocks, namely the signal processing techniques and the machine learning algorithm. This is due to the fact that these two components represented a stronger technical challenge to our development.

4 SONAR SIGNAL PROCESSING

The fact that neither the smartphone nor the earphones have been manufactured for the purposes of our application raises a set of challenges. In particular, reliable detection of fine-grained movements like hand-to-face gestures necessitates a large bandwidth (2 cm resolution requires typically > 8kHz).[2] However, commercial

[2]This is because the range resolution is c/2B where c is the speed of propagation of acoustic signals and B is the bandwidth [1].

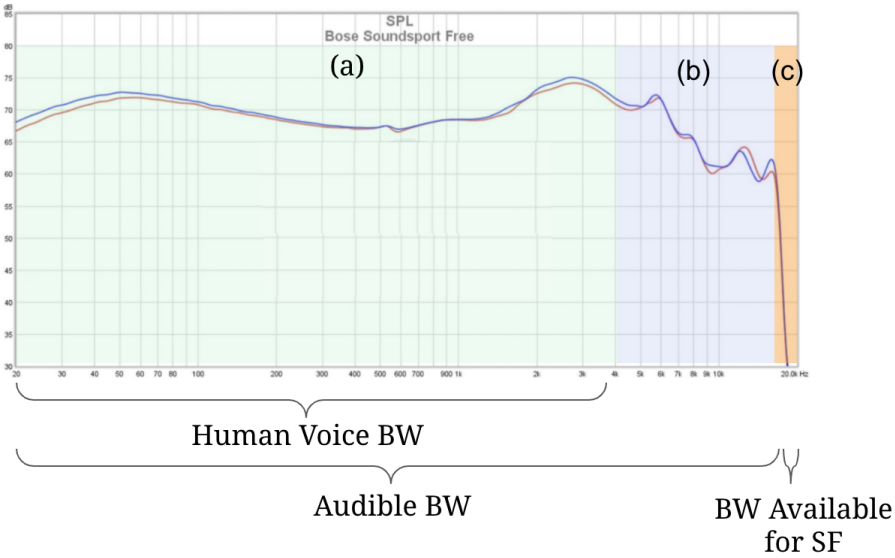


Fig. 4. Typical response curve (gain vs. frequency) of commercial headphones (two measurements, with the same brand and model) [40] divided in three Bandwidth (BW) sections: (a) is the range of the human voice (100 - 4 kHz), (b) is the audible range (4 kHz - 18 kHz) and (c) is the ultrasound range (beyond 18 kHz). Saving Face is constrained to operate in the ultrasound range.

earphones only offer a narrow range of frequencies, as shown in Figure 4, displaying the typical frequency response (gain vs. frequency) for commercial earphones. The gain is maximum and relatively flat in the range of the human voice (0 - 4 kHz, region A), progressively decreases through the range of audible frequencies (up to 18 kHz, region B), and sharply decreases for ultrasound (region C). In practice, the sharp decrease of the gain reduces the sensing distance and increases the Signal-to-Noise Ratio (SNR). This means frequencies beyond 20 kHz are not usable, due to negligible gains (notice the sharp decrease projected beyond region (C) in Figure 4).

At the same time, for the system to be usable, it must not disturb users with audible sounds. Therefore, it should operate with frequencies above 18 kHz (i.e., ultrasound). This constrains Saving Face to a slim operating bandwidth of 2 kHz, increasing the difficulty of extracting reliable information with existing signal processing techniques.

4.1 Choice of Sonar Techniques

The design of the signal processing for Saving Face seeks to optimize the sensing accuracy of fine-grained movements within a 2 kHz bandwidth. Upon evaluating multiple sonar techniques ([43] offers a comprehensive review) and building and testing prototypes, we have identified FMCW and Doppler Shift as the only techniques capable of providing speed and position information while also being robust to external interference, inaudible (not introducing audible sounds due to discontinuities in the signal), and compatible with the narrow band of available frequencies. The techniques evaluated, as well as the conclusions are summarized in Table 5.

As the two techniques rely on distinct physical phenomena, they provide complimentary informational value. We have thus decided to try to corroborate them to maximize the processing accuracy.

Sonar Method (output)	Analysis
Amplitude (Distance)	Due to a poor signal-to-noise ratio, very low precision and no robustness to background noise.
OFDM (Distance)	Audible harmonics are introduced.
FMCW (Distance and Speed)	No audible harmonics are introduced. Robust to background noise.
Doppler Shift (Speed)	No audible harmonics are introduced. Robust to background noise.

Fig. 5. Analysis of different sonar systems. FMCW and Doppler shift prove the most suitable combination for tracking the speed and position of a person's hand.

Constant Frequency Component - The Constant Frequency (CF) Component (C_{CF}) provides information about the speed of objects moving around Saving Face's earphone-microphone pair based on the Doppler effect. This physical phenomena produces a change in frequency of the wave when reflected by a moving object (the amount of the frequency shift is proportional to the speed). Mathematically, the doppler shift can be expressed as $f_{Doppler} = v f_c / c$, where v is the speed of movement of the reflecting object (i.e., hand) with respect to the microphone-earset pair, f_c is the frequency of the transmitted acoustic signal and c is the speed of propagation of the acoustic signal in air.

Figure 3 (b) (bottom) shows the frequency shifts caused by the movement of a hand hovering over Saving Face's earphone-microphone. The hand approaching the pair triggers a shift to higher frequencies, while the hand moving away from the pair triggers a shift to lower frequencies.

We selected the 18 kHz frequency for the C_{CF} (continuous sinusoidal wave), because it has the highest gain in the available spectrum. We prioritize on allocating the highest gain to C_{CF}, instead of C_{FMCW} because C_{CF} can provide information while requiring significantly less bandwidth.

FMCW Component - The FMCW Component (C_{FMCW}) provides information about the position and speed of objects. It operates by transmitting a frequency modulated signal and evaluating the distortions in the components reflected by neighboring objects. Figure 3 (b) (top) shows the distortions in the spectrogram caused by the movement of a hand hovering over Saving Face's earphone-microphone.

The combination of FMCW (for C_{FMCW}) and Doppler (for C_{CF}) signal processing offers Saving Face a multi-resolution tracking capability. On the one hand, since Doppler relies on a single frequency, Saving Face can track the Doppler shift with sample-level granularity. On the other hand, since FMCW operates over longer time periods (as it requires transmitting a full sweep over multiple samples), it can track location more robustly, but at a lower effective sampling rate than Doppler. As we demonstrate empircally in our results, this combination yields superior performance over using each technique independently.[3]

4.2 Signal Processing Steps

Figure 6 shows the signal processing steps starting when the signal, comprised of the C_{CF} and C_{FMCW} components, is transmitted. The steps are detailed as follows:

[3]It is also worth noting that it is possible to extract Doppler shift from FMCW sweeps; however, this typically suffers from a well-known problem called range-Doppler ambiguity. Moreover, it would reduce Saving Face's ability to acquire Doppler shifts with fine temporal resolution as it would be limited to a single Doppler estimate each FMCW sweep.

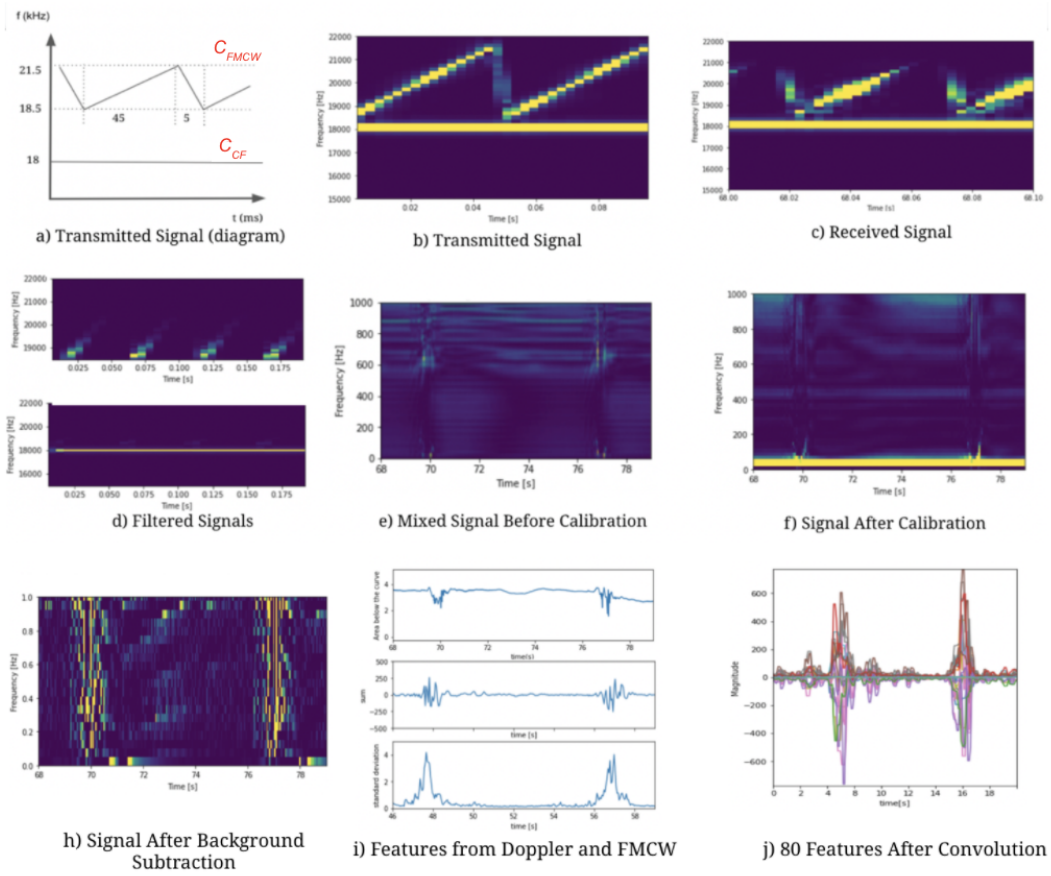
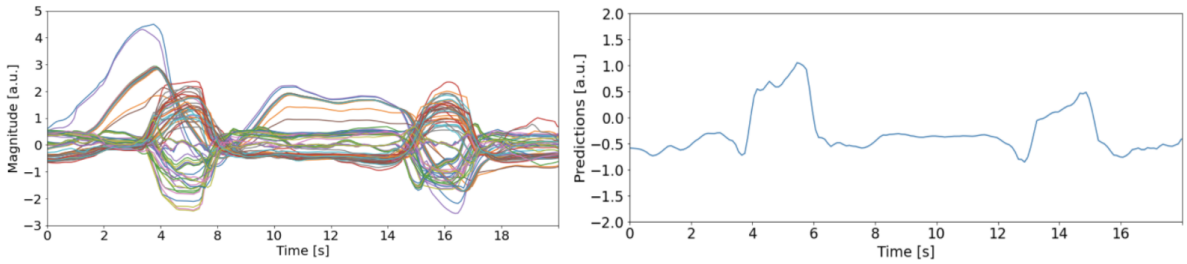


Fig. 6. Signal processing pipeline from the transmission of the signal from the earphone until the features that are used for detecting the face touch.

(a) The diagram of the signal to be transmitted shows the timing and frequency parameters of C_{FMCW} (top) and C_{CF} (bottom).
(b) The spectrogram of the signal emitted by the earphones (frequency vs. time, the gradient of color represents the amplitude of the signal where blue is minimum and yellow is maximum).
(c) The spectrogram of the signal received by the embedded microphone. Note the significant reduction in amplitude for frequencies above 20 kHz, due to the decay in the microphone and earphone frequency response shown in Figure 4.
(d) The received signal is filtered using highpass and bandpass filters to isolate the C_{CF} and C_{FMCW} components of the signal from each other. Only C_{FMCW} will follow steps (e) and (f).
(e) The transmitted and received signals are then mixed. The figure shows the result of the mixing when the upchirp of the transmitted signal does not match the upchirp of the received signal. Face touches are vaguely distinguishable.



(a) (b)

Fig. 7. Feature Extraction and Model Training: (a) Doppler and FMCW Features after standardization, (b) Predictions from the Logistic Regression model

(f) The upchirp of the transmitted signal is aligned with the upchirp of the received signal during calibration. The mixing of the aligned signals display traces associated with face touches that can be better distinguished.
(g) Background substraction is used to highlight the changes in the spectrogram due to moving objects.
(h) The final processing steps involve extracting features for the machine learning model. Specifically, the C_{CF} component is passed through two bandpass filters after step (d) to extract relevant information about speed. The area under each bandpass filter along with the sum and standard deviation across each FFT in the FMCW spectrogram (h) are used as features for the machine learning model. The features from (h) are passed through four additional convolutional filters.
(i) The resulting 80 features (overlapped).

The resolution of FMCW increases with the bandwidth. We selected the range of frequencies 18.5 kHz to 21.5 kHz seeking to maximize the bandwidth while abiding by the constraints of the system. The lower bound was selected with a safety margin of 500 Hz above the C_{CF} component to avoid interference with reflections of the C_{FMCW} component shifted towards higher frequencies. The upper bound was selected at 21.5 kHz, seeking to center the signal at 20 kHz.

In order to optimize the information that can be extracted from FMCW, we implemented the following design choices:

- Extending the bandwidth beyond 20 kHz: the range 18.5 to 21.5 kHz enables the extraction of information between 18.5 and 20 kHz when the gain has high values, and it also enables the potential extraction of useful information between 20 and 21.5 kHz in case the hand moves very close to the earphone and microphone pair, or the hardware happens to have a better gain in this frequency region.
- Designing the FMCW signal as a sawtooth, instead of the common triangle waveform, to concentrate the radiated energy in the upchirp (section with increasing frequency) and maximize the SNR of the position information. We selected a distribution of 90% upchirp and 10% downchirp to avoid abrupt changes in the resulting sine wave that can introduce audible harmonics.
- Selecting a period of 100ms for the sawtooth modulating signal, since it brings a good trade-off between being short enough to provide sufficient measurements to describe the dynamics of a face touch (10 measurements / second), and large enough to provide a sufficient SNR to distinguish patterns in the FMCW and respond to face touches.

5 ML-BASED GESTURE RECOGNITION

Figure 7 shows that it is possible to observe a distinct pattern in the spectrogram of the received signal, which is associated with the face-touching gesture. However, it also displays a number of challenges for using the proposed techniques. First, the SNR is very low, leading to imprecise measurements of speed and position deriving from both CW and FMCW. Second, due to the relatively close distance between the hand and the sonar sensors (earset and mic), one cannot approximate the hand as a single point reflector (or scatterer); this is because acoustic signals may reflect off multiple sections of the user's hand. Addressing this problem is particularly challenging in the context of Saving Face since it cannot employ arrays for imaging [38] and must rely entirely on a single sensor (earset-mic pair).

The machine learning algorithm was thus built to augment the system's capacity to detect distortion patterns in the ultrasound signal spectrogram that are unique to a face-touching gesture. As shown in figure 7, the 80 features obtained from the signal processing had to be standardized (a) before being passed to the logistic regression model. The logistic regression model then returned a set of predictions as to whether or not a face-touching gesture had occurred (b).

Three development steps were required to build the machine learning algorithm. The first two involved building the dataset and building the model using a PC, while the third step consisted of deploying the algorithm on the smartphones for its actual use.

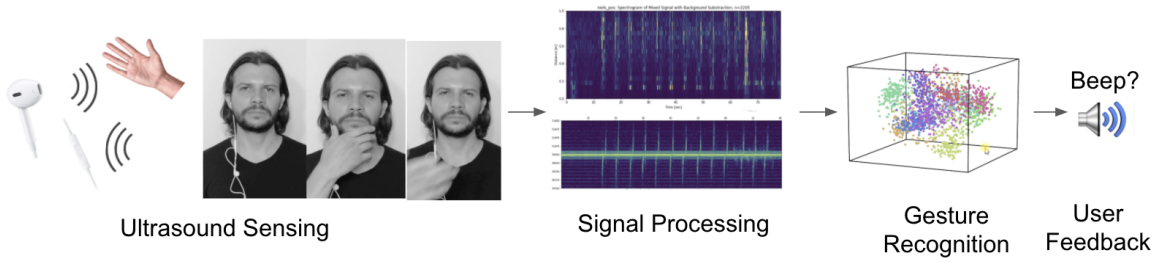


Fig. 8. We have used nine different tasks to train the machine learning model and evaluate the system. The tasks represent common activities in three different use cases: office, manufacturing, supermarket.

5.1 Building the Dataset

In order to train the machine learning model that will operate in real-time on the device, we had to first build a dataset. For this, we recruited 29 participants. This experiment enabled us to collect data from a variety of environments, users, and hardware (phone models and headphones). The protocol was approved by the IRB board of our university.

Due to the limitations on person-to-person contact during the pandemic, we designed an evaluation that could be performed by the users at home while guided by the app. Figure 8 shows a user performing the different tasks. The users followed the protocol below:

(1) Download the app. The App will display a series of instructions informing the user to perform a set of 9 simple tasks.
(2) Video record themselves while performing the tasks.
(3) Perform Tasks 1 to 9 shown in Figure 8. The duration of each task is 30s, except for 1, 2, and 3, which lasted 60s.

(4) In Tasks 1, 2 and 3, the users were instructed to touch their faces upon hearing a "beep" sound. The app will emit ten beeps interleaved by an time delay of 6-10s (randomly distributed). These tasks provided the positive samples (i.e., regions in the spectrogram corresponding to face-touching events) needed for training the machine learning model. The remaining tasks are intended to provide the negative samples for the model (i.e., regions in the spectrogram where no face-touching events occur).
(5) Share the files produced by the app with recorded traces with the experimenters via e-mail.

The positive samples were labelled by members of the team using the video recording to identify the timestamps when the user's hand touched their face. We used an initial beep in the app to align the data traces in the phone with the video recording. The negative samples were drawn randomly from the recordings of tasks with no face touches.

5.2 Building the Model

To detect a face touch gesture, we built a binary classifier trained on the Doppler and FMCW features described in Section 4. In total, we had 29 datasets with 379 positive samples and 358 negative samples. Before extracting any features, we applied a subject-wise split on the dataset randomly assigning 75% of subjects to the training set and the remaining 25% of subjects to the testing set. We also balanced the training set to ensure it contained an equal number of positive and negative samples for model training.

We then extracted features for each sample by obtaining 500ms of data surrounding the sample, applying four convolutional filters and a rolling standardization, and calculating the minimum, maximum, mean, and standard deviation to extract the final features for the machine learning model.

w_r (s) \ w_t (s)	**1**	**1.5**	**2**	**2.5**
2	0.77	0.94	0.82	0.82
4	0.83	0.89	0.82	0.81
6	0.83	0.87	0.84	0.77
8	0.78	0.87	0.86	0.84
10	0.80	0.87	0.86	0.82
12	0.81	0.88	0.86	0.83

Fig. 9. F1- Metric in test sets during a parametric sweep of standardization windows

To determine the parameters for the rolling standardization, we conducted a parametric sweep to determine i) the length of the reference window w_r used to obtain the mean and standard deviation for the z-score, and ii) the length of the averaging window w_t used to determine the most recent feature value. The results of applying feature standardization are shown in Figure 7. Additionally, Figure 9 demonstrates that a 2-second reference window and 1.5-second averaging window produced the best performing machine learning model with an F1-score of 94.1%.

We compared three logistic regression models each trained on a different set of features deriving from the signals: Doppler (Shift), FMCW, and the combination of Doppler and FMCW. For each model, we applied leave-one-subject-out cross validation (LOOCV) on the training set to set the model hyperparameters. By tuning the hyperparameters with only the train and validation sets from LOOCV, we ensured that the hyperparameters were not unintentionally tuned to the test set. In addition, subject-wise splits were used for both LOOCV and the initial train-test split so that samples from the same user were not used to both train and evaluate the model in order to prevent issues of overfitting. The LOOCV mean accuracies for the Doppler, FMCW, and Doppler and FMCW features were 76.7% (σ = 12.1%), 78.8% (σ = 12.9%), and 79.5% (σ = 13.3%) respectively.

Throughout the process, we focused on reducing the number of false positives and false negatives while maintaining a high true positive rate. Thus, our primary metric was the F1-score. The logistic regression model had an F1-score of 81.2% with Doppler and an F1-score of 89.4% with FMCW when evaluated on the test set. Our final model combines both Doppler and FMCW and yields an F1-score of 94.1% on the test set as shown in Figure 10. The model has a precision and recall of 95.5% and 92.7% respectively. These results validate the joint use of Doppler and FMCW to enhance the robustness of the model.

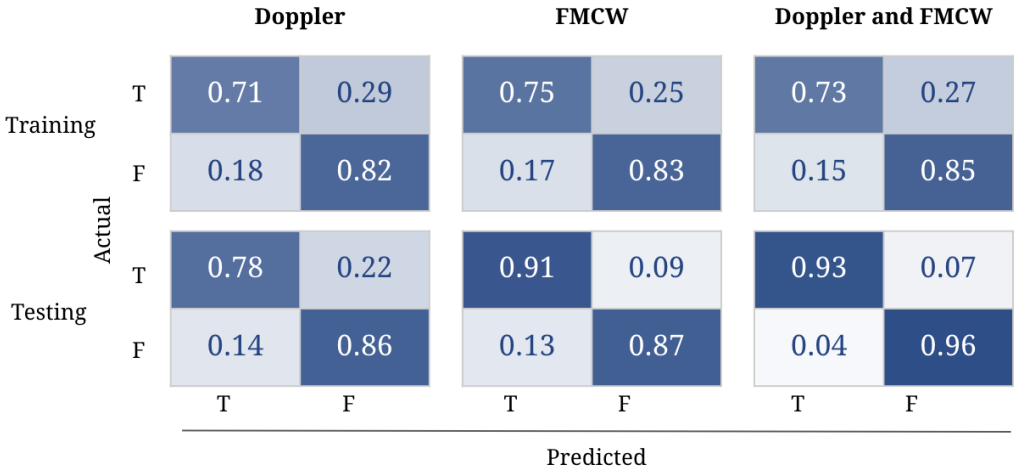


Fig. 10. Evaluation of the logistic regression model on three sets of features: Doppler, FMCW, and the combination of Doppler and FMCW. The combination of Doppler with FMCW feature provides higher precision and recall than the separate use of each technique. The label True (T) represents a face touch and False (F) represents the lack of a face touch.

5.3 Deploying on the Smartphone

Since the algorithm was trained using a computer, it then had to be deployed to the smartphone app. As the user's microphone captures the ultrasound system, the algorithm performs three core steps (1) recovering the spectrogram, (2) extracting the features, and (3) evaluating the machine learning model.

Step 1. Recovering the spectrogram: The ultrasound signal shown is emitted and received by the earphones that are connected to the phone. The received signal is then passed through three filters to separate the Doppler and FMCW portions from the signal. To generate the FMCW spectrogram, the received signal is mixed with the transmitted signal, which is followed by applying a low pass filter to remove high frequency noise. A Fast Fourier Transform (FFT) is applied to the mixed signal to extract the frequency content. By subtracting consecutive FFTs, we can isolate the reflected signals created by the moving hand from those caused by stationary objects in the surroundings. This produces the spectrogram in Figure 3, which depicts the position of a user's hand.

Step 2, Extracting features: As the hand approaches the face, we notice characteristic distortions in both the constant tone and the FMCW spectrogram. In the case of the constant tone, the area under the signal is distorted. In addition, in the Doppler spectrogram, we see a vertical line above 18kHz as the hand approaches the face followed by a vertical line below 18kHz as the hand leaves the face (Figure 3). Meanwhile, in the FMCW spectrogram we see two vertical bars from around 0.2 - 1.0m when a face touch occurs (Figure 3). To capture these distortions, we extract the area under the signal for Doppler and the sum and standard deviation values of each FFT for FMCW.

Step 3, Evaluating the model: Every 100ms, we combine the given Doppler and FMCW features for the past 500ms and evaluate the logistic regression model to determine the probability that a face touch event occurred. If the probability of a face touch event is less than 50%, then a face touch is not predicted and nothing happens. Otherwise, a face touch event is predicted and the system emits a beep or vibration to alert the user.

We have used the libraries Audiokit [31] and vDSP [8] for signal acquisition and processing in iOS. In Android, we rely on the library Tarsos [39].

A key to our goal of designing the system so that it seamlessly integrates with day-to-day activities, is that all the libraries and operations mentioned in this section can operate in the background of the smartphone operating system. This would enable other applications that use audible sound to operate in parallel. For example, if the user is listening to music by keeping the right earphone in the ear (see Fig. 2.2.A), the left one can be used for ultrasound emission by keeping it out of the ear. No ultrasound is emitted from the right earphone [4]. We have done a preliminary test of this feature in an iPhone 12, using Saving Face and Spotify, and have not noticed any differences in performance due to the music reproduction. However, extensive testing is required to verify compatibility with other audio-related functions and operating systems. In the event of incompatibility with other app, a potential strategy would be to pause Saving Face and automatically resume after the other app is closed.

6 REAL-LIFE USER STUDIES

This section shows early results of evaluating Saving Face in-the-wild, distributed to users via the app store and tested in their own home with their phones and earphones.

The user study consisted of a 30-minute zoom call between an experimenter and each participant. Prior to the user study, each participant downloaded the Saving Face [5] application and was given instructions to attach the microphone and left earphone to the neck of their shirt using adhesive tape (see Fig. 11). At the start of the study, the experimenter confirmed the microphone and earphone positioning, clarified questions, and prompted the participant to begin the Saving Face application on their phone.

The experimenter then provided oral instructions asking the user to touch their face every 10 seconds while performing a series of day-to-day activities such as sitting on a chair, walking around the room and stacking cans. In these instructions, the experimenter specified that the user should alternate which hand they use to touch their face, but did not provide any additional specifications about how and where the user should touch their face. For each activity, the user performed 15 face touches and the experimenter quantified the beeps from the system to obtain a ground truth about the number of true positives and false positives that occurred for each activity. At the end of the study, the experimenter performed a semi-structured interview to study the user experience.

The experiment included 10 participants (mean age 30.2 years old, σ = 15.9) recruited via email lists and social networks. The inclusion criteria required participants to be 18 - 65 years old. The experiment focused on iPhone

[4]We would also like to make sure that the user does not place the wrong earpiece into their ear. A potential strategy to avoid this is to set up the system so that if the ultrasound is not detected by the microphone, it prompts the user to correct the positioning

[5]The real-life user studies were done with a version of the ML model trained and tested with k-Fold cross validation with 5 splits, instead of one-leave-out (described in Sec. 5.2). We did not observe significant changes in the resulting precision (94.1%) and recall (92.8%) (compare to 95.5% and 92.7%, reported in Sec. 5.2).

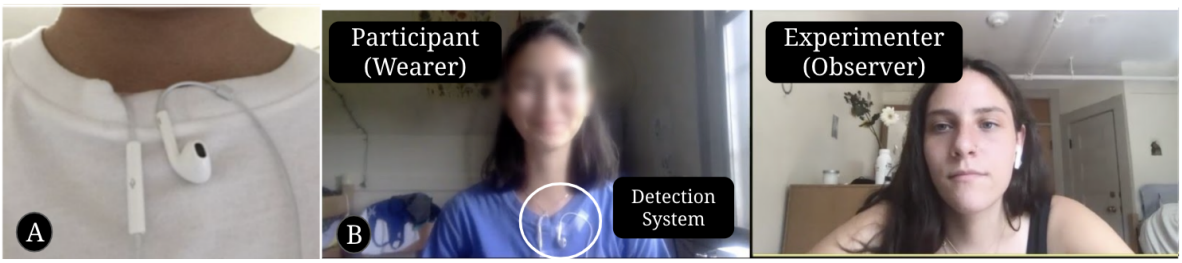


Fig. 11. We conducted 10 user studies. (A) depicts the set-up where the earphones are taped to the collar of the shirt. The participant (B) then performs a series of tasks while using the system as the observer (C) notes the number of false negative and false positives that arise during the duration of the study.

users. Prior to the experiment, the users were instructed to sign a consent form securely delivered via DocuSign [9]. The study was approved by the Institutional Review Board of our university.

6.1 Sensitivity and Precision Detecting Face Touches

Figure 12 shows the per-task Sensitivity (defined as correctly detected face touches / actual number of face touches) and the Precision (nudges linked to face touches / total number of nudges), averaged over the ten users. The experimenter quantified the number of correctly- or erroneously-detected face touches during the session by listening to the beeps emitted by the app through the video-call and corroborating with the actual action seen on the screen (of the participant touching their face or not). The participant was also asked to indicate if they noticed any erroneously-detected face touches after being presented with the basic operating principles of the system.

Across all activities and users, the overall Sensitivity was 93.7% percent and the Precision was 91.5 %. The metrics were highest when users were sitting on a chair (average Sensitivity of 96.0% and Precision of 98.3%). When participants started walking around the room, the Sensitivity became 95.0% and Precision 90.0%. When participants started stacking cans, the Sensitivity became 90.0% and Precision 86.5%.

The results suggest that the performance of Saving Face is at a maximum for situations when there is little bouncing of the microphone-earphone pair. These tasks are typical of use cases such as working in an office/home, attending a lecture in a classroom, driving a car, etc.). The performance suffers moderately when the tasks involve abrupt movements or manipulating objects close to the torso region.

We expect the performance of the system to be improved in the next phases of experimental results especially by increasing the amount of data used for training the machine learning models and the number of features extracted from Doppler and FMCW. This will help the model to better distinguish the finger print of face touching and differentiate it better from movements that involve a large surface in proximity of the headphone-microphone pair (e.g., doing jumping jacks, with the system bouncing on the chest, or manipulating a box).

We believe that the actual performance of the system may already be at an acceptable level for practical application on behavioral change. In this direction, the existence of false positives and negatives could be leverages in the behavioral change intervention to avoid dependency of the user on the device.

6.2 Qualitative Insight

In addition to quantifying the sensitivity and precision of the tool, we have also asked participants to share qualitative insight into their user experience, intent to use such a system to reduce face-touching behavior and ways the system could be improved. Out of the 10 users, 7 chose to answer the qualitative questions about their

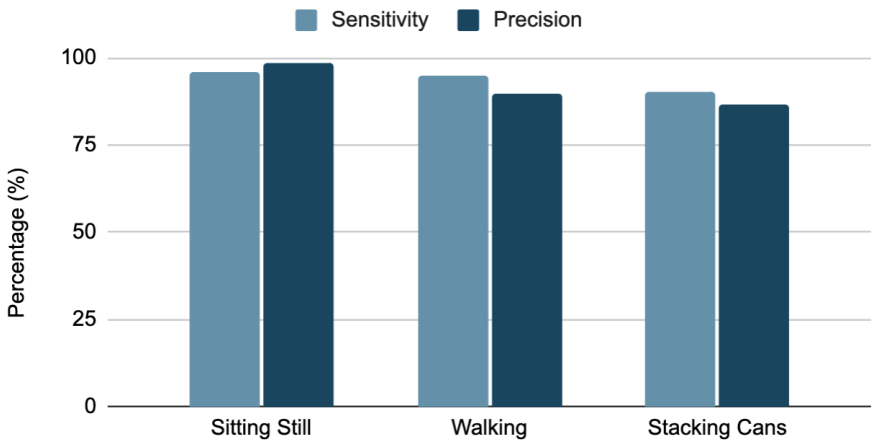


Fig. 12. Sensitivity and Precision measured during user studies of multiple tasks.

user experience. Of the 7 respondents, 4 indicated they would be willing to use the App in certain instances of their day to day lives. The main improvement points that users noted were no longer needing to tape the left earphone to their shirts and not using the audio nudge (the tests were performed with the audio instead of the vibrating nudge, in order for the experimenter to be able to also notice the nudges via zoom). One participant also noted that the issue of protecting herself from the surface transmission of COVID-19 is simply not top of mind for her.

In the next batch of user studies, we will aim to also involve participants outside the student population, to best emulate the target users of Saving Face. While the attaching of the earphone to the shirt is an ongoing area of improvement, we hope it may play a less significant role for users such as manufacturing workers. At the same time, a group of front line workers could potentially also give a higher level of priority to protecting themselves from the hand-face transmission of COVID than college students, mostly due to their higher levels of exposure.

7 DISCUSSION, LIMITATIONS AND FUTURE WORK

The conducted experiments have allowed us to generate early evidence of the system's effectiveness in detecting face touches, as well as insight into the real user experience and potential improvement areas. We have also identified several potential limitations of the system that motivate a more extensive validation, notably:

(1) Safety considerations due to exposure to low-intensity, low-frequency ultrasound. Saving Face have been designed to operate at frequency levels of 18-21.5 kHz, which is above the audible range of the average human ear (situated at a maximum of 15-17 kHz [33]). Nonetheless, some individuals with a particularly sensitive sense of hearing, as well as human infants and pets, could be able to hear the emissions if the headset is to be placed in very close proximity to their ear. Given that Saving Face is not used directly in the ear, we would expect the ability to hear it to be a rare occurrence. In the case of human adults, we would expect this to result in a natural reluctance to use Saving Face [20]. When it comes to infants, while we would not expect them to be among Saving Face users, we would recommend avoiding the use of the App for people below 18 years old, until we will have been able to perform more thorough testing. At the same time, given the sharp decay on the ultrasound signal intensity with distance (significantly sharper

than the case of audible frequencies [37]), being in the proximity of a Saving Face user should result in a very low probability of incidentally hearing the signal.
To date, low frequency ultrasound is being used in consumer products where the source of sound is close to the head, for example, Xiaomi's Mi 10T Lite smartphone uses ultrasound emissions from the speaker to detect proximity to the head and turn off the screen (with frequencies between 23kHz and around 35kHz) [42]. Some studies report potential harmful effects of low frequency ultrasound exposure at intensities >190 dB (cavitation), or >140dB (slight heating) [2]. Nonetheless, there is an acknowledged need for more research on its health impacts, standardisation and potential regulation of the field [2]. Currently, the U.S. Occupational Safety and Health Administration establishes an ultrasound exposure limit of 105 dB – which also sets the standard for the design limitation of commercial headphones, and an implicit upper bound to Saving Face's intensity levels [11]. Given Saving Face is not meant to be used directly in the ear, and the sharp decay of the intensity with distance from the earphone, we do not expect the intensity perceived by the human ear to be able to approach a level associated with any studied harmful effect. Nonetheless, we acknowledge the need to engage in dedicated research of the phenomenon, including and beyond any legal requirement before deploying the solution.

(2) Some types of face touching might not be detected by the system. Face touching events where the hands or arms of the user do not move in the proximity of the earphone-microphone pair may not cause the distortions in the ultrasound signal that the system is designed to detect (e.g., scratching the right ear with the right hand). Additionally, the system has not yet been tested with gestures that involve an extended contact between the hand and the face (e.g., scratching the nose, head propping, etc.).
(3) Some activities might be prone to confound the system and trigger false positives. For example, eating involves constant approaching of the hands to the face without actual contact (similar to drinking or adjusting the eyeglasses).
(4) Impact of the wearing position. The earphone and microphone can be placed in multiple positions (Fig. 2.2.A-D shows several examples). The evaluation in this paper focused on the case where the ultrasound-emitting earphone (the left one) and the microphone are attached to the neck of the shirt (Fig. 2.2.C). An option with a more simple setup is to place the right earphone in the ear (the one that does not emit ultrasound) and keep the microphone and the other earphone hanging from the ear (Fig. 2.2.A.). We have done preliminary testing of the last option and observed that gestures from both hands can be detected. However, it is expected that the precision and recall will decrease due to movement and the partial occlusion of the ultrasound signal by the head.
(5) Limitations of the ML model due to the small dataset size. Our current results are still preliminary due to the relatively small dataset size (29 users for training the model and 10 for subject testing, refer to Secs. 5 and 6), therefore the sensitivity and precision might be affected in other testing conditions (e.g., type of smartphone, earphones, operating system, etc.). The size of the dataset should be increased to represent this variability and further improve the model, as well as confirm the results obtained.
(6) Impact on the battery life. Preliminary testing on a full-charged iPhone 12 running only Saving Face has shown that its battery level decreases 5% after 4h (the phone is kept with the screen off and in flight mode, methodology inspired by [27]). Going forward, a thorough analysis of the energy consumption must be performed, since the current implementation has not considered maximizing battery-life as a design goal. We expect that the most power-consuming operations in the app are the real-time signal processing and the emission and reception of ultrasound. We also expect that the battery consumption due to ultrasound emission will be lower than in most other always-on mobile gesture sensing techniques that rely on the speaker of the phone (e.g., FingerIO [27]), since the radiated power of the earphones is significantly lower.
(7) Exploring the use of wireless headphones. The research and testing of the system has focused only in wired headphones. The reason is that expanding to wireless headphones (e.g., Apple Airpods) brings additional

development complexity, since the wireless communication protocol might filter out the ultrasound frequencies in the signal as a strategy to optimize power consumption (we have noted this behavior during prototype testing). This might be addressed by adapting the wireless protocol and updating the firmware in the earphones.

For Saving Face to ultimately become a broadly available solution, we are focusing on further advancing and testing its robustness in a wider variety of situations (e.g., underlying activities), as well as on improvements of the user experience. Concrete improvement avenues include:

(1) Enriching the ML dataset, to increase robustness of gesture detection. The classification accuracy of the ML model could be further improved by increasing the number of users in the training set. Additionally, the datasets will also be extended to more challenging scenarios by including activities with more drastic movements (e.g., riding a bicycle, jumping jacks, etc.) and that involve manipulation of objects around the torso (e.g., cooking, cleaning, collecting garbage, etc.).
(2) Further advancing the signal processing algorithms for extracting additional features, such as the sum and standard deviation of the column values in spectrograms within small frequency bands (we currently use the entire column), or the output of additional convolutional filters. This would provide more information that the ML model can use to distinguish the face touches. It would also be interesting to explore the possibility of relying solely on ultrasound frequencies beyond 20 kHz. Based on our testing, a substantial fraction of the earphones (typically the high-end ones) have the capability of emitting such frequencies. This would reduce the likelihood of users hearing the system.
(3) Exploring additional wearing positions. The current evaluation required users to attach the earphones to the neck of their t-shirts using adhesive tape (see Fig. 2). That said, we aim to also evaluate other ways of wearing the system to first improve the user experience, but also account for other regulations and social factors. For example, bank cashiers and average users may prefer to hang the system from one's ear while this position might not be allowed for a heavy machine operator or meat-packing employees due to safety concerns.
(4) Testing the system with a batch of 50 users, other than university students. This study focused on the performance of the system while users were performing basic tasks such as sitting on a chair, walking or stacking cans. In the next set of tests, we will aim to test additional activities and validate the safe operation of the system. This will nonetheless require a consistent recruitment effort outside the university premises.

In addition to our immediate priorities of improving the system's robustness and ensuring that it is safe for use in a wide variety of conditions, we will explore the following expansions in the near-term:

(1) Advancing the Android platform. This evaluation was performed on users employing the iOS application because (to date) our research and development has focused on this platform. We also have a version for Android devices, but it is at the prototype stage to demonstrate technical feasibility. We would like to continue the work in extending our evaluation to Android users. This development will enable us to ensure robustness across mobile platforms and make the system accessible to both sets of users.
(2) Deploying Saving Face for longer time periods in ambulatory settings. In order to prove the behavioral change power of the solution, we aim to deploy it for longer periods of time in ambulatory settings with an increased number of users. Voluntary participants would be asked to regularly use the App for approximately 1 week and report on the evolution of their face-touching behavior, as well as on the user experience over more extended periods.
(3) Exploring additional gestures. While in this paper we focus on detecting hand-to-face gestures, in the mid-term, we would also aim to explore its applicability to other gestures (e.g., air-swiping, hair pulling) and other applications where a set of speakers and microphone available could be used as a sonar (e.g., the sound system within a car).

8 CONCLUSION

The Saving Face App is an accessible and easily scalable system, aimed at reducing the risky hand-to-face movement that contributes to disease transmission in the age of COVID-19, and beyond. Our solution relies on the novel idea of turning existing smartphones and ear sets into a sonar system, to then leverage advanced signal processing and ML techniques to detect hand to face touches and alert the users.

Preliminary user test results show Saving Face to have an 90.0 - 96.0% sensitivity in detecting face touches, as well as a 86.5 - 98.3 % precision, depending on the underlying activity of the user. At the same time, most users appear willing to try the App in their daily lives. Our future work will focus both on further increasing the robustness of the solution (e.g., through larger data sets for the ML, extracting additional features from the Doppler and FMCW signals) and enhancing the user experience (e.g., by exploring alternate ways of wearing the headset). Given its high detection accuracy, potential for a seamless user experience, and high scalability (it leverages hardware the user already owns and the App can easily be downloaded from the online stores), we believe Saving Face can become an effective, widely adopted solution for tackling the surface transmission of COVID-19.

At the same time, while our tests have mainly focused on hand-to-face applications, the solutions we have developed can be applicable to multiple other gesture recognition issues, where a set of speakers and a microphone are available. This could span to areas such as behavior-reversal therapy (e.g., against compulsive hair pulling), security (e.g., device owner recognition), safety (e.g.,falling asleep behind the wheel) or health monitoring (e.g., detecting movement patterns preliminary to a heart attack).

ACKNOWLEDGMENTS

We are grateful to all the participants of our user study for sharing their valuable feedback. We would also like to thank the anonymous reviewers for their insightful comments. The project has been advanced with the help of a team that contributed their time and effort during the challenging time of the COVID-19 pandemic, including: Irmandy Wicaksono, Cedric Honnet, Guadalupe Babio, Nicolas Ayoub, Corina Stoenescu, Junhao Xu, Zhi Wei Gan, Erica Radler, Susanna Chen, Richter Jordaan, Javier Araya, Sarah Uriarte, Jaffette Solano, David Esquivel, Soo Park, Korrawat (James) Pruegsanusak, Shellie Hu, Carol Mai, Tarini Banerji, Franklin Zhang, Noa Schwartz and Aaron Stinnett. A full list can be found in the website media.mit.edu/projects/saving-face/overview/.

The project has received support by the Swiss National Science Foundation (grant P2ELP2.184528) and the U.S. National Science Foundation (RAPID award CNS-2032704). We would like to extend our gratitude to the MIT Media Lab community and its member companies for their continuous support.

REFERENCES

[1] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C. Miller. 2014. 3D Tracking via Body Radio Reflections. In *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*. USENIX Association, Seattle, WA, 317–329. https://www.usenix.org/conference/nsdi14/technical-sessions/presentation/adib

[2] Farzaneh Ahmadi, Ian V McLoughlin, Sunita Chauhan, and Gail Ter-Haar. 2012. Bio-effects and safety of low-intensity, low-frequency ultrasonic exposure. *Progress in biophysics and molecular biology* 108, 3 (2012), 119–138.

[3] Asaph Azaria., Brian Mayton., and Joseph Paradiso. 2016. Thumbs-Up - Wearable Sensing Device for Detecting Hand-to-Mouth Compulsive Habits. (2016), 54–65. https://doi.org/10.5220/0005680900540065

[4] Karina S Bate, John M Malouff, Einar T Thorsteinsson, and Navjot Bhullar. 2011. The efficacy of habit reversal therapy for tics, habit disorders, and stuttering: a meta-analytic review. *Clinical Psychology Review* 31, 5 (2011), 865–871.

[5] Mike Bodge, Brian Moore, and Isaac Blankensmith. 2020. Do Not Touch Your Face. https://donottouchyourface.com/

[6] William J Bradshaw, Ethan C Alley, Jonathan H Huggins, Alun L Lloyd, and Kevin M Esvelt. 2020. Bidirectional contact tracing is required for reliable COVID-19 control. *medRxiv* (2020).

[7] Xiang'Anthony' Chen. 2020. FaceOff: Detecting Face Touching with a Wrist-Worn Accelerometer. *arXiv preprint arXiv:2008.01769* (2020).

[8] Apple Co. 2020. vDSP: Perform basic arithmetic operations and common digital signal processing routines on large vectors. https://developer.apple.com/documentation/accelerate/vdsp

[9] DocuSign. 2020. Electronic Signature and Agreement. https://www.docusign.com/

[10] Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. 2020. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 368, 6491 (2020).

[11] U.S. Center for Disease Control. 2019. What Noises Cause Hearing Loss? | NCEH | CDC. https://www.cdc.gov/nceh/hearing_loss/what_noises_cause_hearing_loss.html

[12] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. SoundWave: Using the Doppler Effect to Sense Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1911–1914. https://doi.org/10.1145/2207676.2208331

[13] 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.

[14] 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.

[15] Joseph A Himle, David M Perlman, and Laura M Lokers. 2008. Prototype awareness enhancing and monitoring device for trichotillomania. *Behaviour research and therapy* 46, 10 (2008), 1187–1191.

[16] Immutouch. 2020. Protect yourself from germs break bad habits. https://immutouch.com/

[17] Günter Kampf, Daniel Todt, Stephanie Pfaender, and Eike Steinmann. 2020. Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. *Journal of Hospital Infection* 104, 3 (2020), 246–251.

[18] Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. 2014. Bringing Gesture Recognition to All Devices. In *Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation* (Seattle, WA) *(NSDI'14)*. USENIX Association, USA, 303–316.

[19] National Health System United Kingdom. 2018. Skin picking disorder. https://www.nhs.uk/conditions/skin-picking-disorder/

[20] T. G. Leighton. 2017. Comment on 'Are some people suffering as a result of increasing mass exposure of the public to ultrasound in air?'. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 473, 2199 (March 2017), 20160828. https://doi.org/10.1098/rspa.2016.0828 Publisher: Royal Society.

[21] X. Li, H. Dai, L. Cui, and Y. Wang. 2017. SonicOperator: Ultrasonic gesture recognition with deep neural network on mobiles. In *2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. 1–7.

[22] Ethan S Long, Raymond G Miltenberger, Sherry A Ellingson, and Shelley M Ott. 1999. Augmenting simplified habit reversal in the treatment of oral-digital habits exhibited by individuals with mental retardation. *Journal of Applied Behavior Analysis* 32, 3 (1999), 353–365.

[23] Wenguang Mao, Jian He, and Lili Qiu. 2016. CAT: high-precision acoustic motion tracking. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. 69–81.

[24] Wenguang Mao, Mei Wang, Wei Sun, Lili Qiu, Swadhin Pradhan, and Yi-Chao Chen. 2019. RNN-Based Room Scale Hand Motion Tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.

[25] Rafael Martinez. [n.d.]. JPL COVID-19 PULSE Pendant. https://medeng.jpl.nasa.gov/covid-19/pulse/

[26] American Society For Microbiology. 2003. Survey: Many Travelers Coast Through U.S. Airports Without Washing Their Hands. https://www.sciencedaily.com/releases/2003/09/030916074111.html

[27] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. 2016. Fingerio: Using active sonar for fine-grained finger tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1515–1525.

[28] Minister of National Health and Government of Canada Welfare. 1991. Guidelines for the Safe Use of Ultrasound - Safety Code 24. https://www.canada.ca/content/dam/hc-sc/documents/services/environmental-workplace-health/reports-publications/radiation/safety-code_24-securite-eng.pdf

[29] Pavlok. 2020. The Most Effective Solution To Living Life Smoke Free For Those Who Have Tried Everything Else – Guaranteed. https://pavlok.com/quit-smoking-with-pavlok/

[30] Domenico Prattichizzo, Tommaso Lisini, Gianluca Paolocci, Nicole D'Aurizio, Sara Marullo, and Annino De Petra. 2020. https://sites.google.com/unisi.it/noface-touchapp/

[31] Aurelius Prochazka. 2020. AudioKit Developer Documentation Site. https://audiokit.io/

[32] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. 2013. Whole-Home Gesture Recognition Using Wireless Signals. In *Proceedings of the 19th Annual International Conference on Mobile Computing amp; Networking* (Miami, Florida, USA) *(MobiCom '13)*. Association for Computing Machinery, New York, NY, USA, 27–38. https://doi.org/10.1145/2500423.2500436

[33] Dale Purves, George J. Augustine, David Fitzpatrick, Lawrence C. Katz, Anthony-Samuel LaMantia, James O. McNamara, and S. Mark Williams. 2001. The Audible Spectrum. *Neuroscience. 2nd edition* (2001). https://www.ncbi.nlm.nih.gov/books/NBK10924/ Publisher:

Sinauer Associates.
[34] Yang Qifan, Tang Hao, Zhao Xuebing, Yin Li, and Zhang Sanfeng. 2014. Dolphin: Ultrasonic-Based Gesture Recognition on Smartphone Platform. 1461–1468. https://doi.org/10.1109/CSE.2014.273
[35] John T Rapp, Raymond G Miltenberger, and Ethan S Long. 1998. Augmenting simplified habit reversal with an awareness enhancement device: Preliminary findings. *Journal of Applied Behavior Analysis* 31, 4 (1998), 665–668.
[36] Wenjie Ruan, Quan Z. Sheng, Lei Yang, Tao Gu, Peipei Xu, and Longfei Shangguan. 2016. AudioGest: Enabling Fine-Grained Hand Gesture Detection by Decoding Echo Signal. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 474–485. https://doi.org/10.1145/2971648.2971736
[37] Dan Russel. [n.d.]. Absorption and Attenuation of Sound in Air. https://www.acs.psu.edu/drussell/Demos/Absorption/Absorption.html
[38] Sheng Shen, Daguan Chen, Yu-Lin Wei, Zhijian Yang, and Romit Roy Choudhury. 2020. Voice localization using nearby wall reflections. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
[39] Joren Six, Olmo Cornelis, and Marc Leman. 2014. TarsosDSP, a real-time audio processing framework in Java. In *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*. Audio Engineering Society.
[40] Sam Vafaei. 2018. Raw Frequency Response. https://www.rtings.com/headphones/tests/sound-quality/raw-frequency-response
[41] Neeltje Van Doremalen, Trenton Bushmaker, Dylan H Morris, Myndi G Holbrook, Amandine Gamble, Brandi N Williamson, Azaibi Tamin, Jennifer L Harcourt, Natalie J Thornburg, Susan I Gerber, et al. 2020. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *New England Journal of Medicine* 382, 16 (2020), 1564–1567.
[42] Chris Velazco. [n.d.]. Ultrasound and software could replace a phone's proximity sensor. https://www.engadget.com/2016-01-19-elliptic-labs-beauty-ultrasound.html
[43] Z. Wang, Y. Hou, K. Jiang, C. Zhang, W. Dou, Z. Huang, and Y. Guo. 2019. A Survey on Human Behavior Recognition Using Smartphone-Based Ultrasonic Signal. *IEEE Access* 7 (2019), 100581–100604.
[44] Hiroki Watanabe and Tsutomu Terada. 2018. Improving Ultrasound-Based Gesture Recognition Using a Partially Shielded Single Microphone. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers* (Singapore, Singapore) *(ISWC '18)*. Association for Computing Machinery, New York, NY, USA, 9–16. https://doi.org/10.1145/3267242.3267274
[45] Douglas Woods and Raymond Miltenberger. 2007. *Tic disorders, trichotillomania, and other repetitive behavior disorders: Behavioral approaches to analysis and treatment*. Springer Science & Business Media.
[46] Chaojun Xie, Hongjun Zhao, Kuibiao Li, Zhoubin Zhang, Xiaoxiao Lu, Huide Peng, Dahu Wang, Jin Chen, Xiao Zhang, Di Wu, Yuzhou Gu, Jun Yuan, Lin Zhang, and Jiachun Lu. 2020. The evidence of indirect transmission of SARS-CoV-2 reported in Guangzhou, China. *BMC Public Health* 20, 1 (2020), 1–9. https://doi.org/10.1186/s12889-020-09296-y
[47] Sangki Yun, Yi-Chao Chen, Huihuang Zheng, Lili Qiu, and Wenguang Mao. 2017. Strata: Fine-grained acoustic-based device-free tracking. In *Proceedings of the 15th annual international conference on mobile systems, applications, and services*. 15–28.
[48] 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*. 468–469.
[49] Chen Zhao, Ke-Yu Chen, Md Tanvir Islam Aumi, Shwetak Patel, and Matthew S. Reynolds. 2014. SideSwipe: Detecting in-Air Gestures around Mobile Devices Using Actual GSM Signal. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 527–534. https://doi.org/10.1145/2642918.2647380