PPGface: Like What You Are Watching? Earphones Can "Feel" Your Facial Expressions

SEOKMIN CHOI, University at Buffalo, State University of New York, USA
YANG GAO, Northwestern University, USA
YINCHENG JIN, University at Buffalo, State University of New York, USA
SE JUN KIM, University at Buffalo, State University of New York, USA
JIYANG LI, University at Buffalo, State University of New York, USA
WENYAO XU, University at Buffalo, State University of New York, USA
ZHANPENG JIN*, University at Buffalo, State University of New York, USA

Recognition of facial expressions has been widely explored to represent people's emotional states. Existing facial expression recognition systems primarily rely on external cameras which make it less accessible and efficient in many real-life scenarios to monitor an individual's facial expression in a convenient and unobtrusive manner. To this end, we propose PPGface, a ubiquitous, easy-to-use, user-friendly facial expression recognition platform that leverages earable devices with built-in PPG sensor. PPGface understands the facial expressions through the dynamic PPG patterns resulting from facial muscle movements. With the aid of the accelerometer sensor, PPGface can detect and recognize the user's seven universal facial expressions and relevant body posture unobtrusively. We conducted an user study (N=20) using multimodal ResNet to evaluate the performance of PPGface, and showed that PPGface can detect different facial expressions with 93.5 accuracy and 0.93 f1-score. In addition, to explore the robustness and usability of our proposed platform, we conducted several comprehensive experiments under real-world settings. Overall results of this work validate a great potential to be employed in future commodity earable devices.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile devices**; **Human computer interaction (HCI)**.

Additional Key Words and Phrases: Photoplethysmogram, PPG, Facial Expression, Ear Canal, Blood Vessel Deformation

ACM Reference Format:
Seokmin Choi, Yang Gao, Yincheng Jin, Se jun Kim, Jiyang Li, Wenyao Xu, and Zhanpeng Jin. 2022. PPGface: Like What You Are Watching? Earphones Can "Feel" Your Facial Expressions. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 48 (June 2022), 32 pages. https://doi.org/10.1145/3534597

*This is the corresponding author.

Authors' addresses: Seokmin Choi, University at Buffalo, State University of New York, Department of Computer Science and Engineering, Buffalo, NY, 14260, USA; Yang Gao, Northwestern University, Department of Computer Science, USA; Yincheng Jin, University at Buffalo, State University of New York, Department of Computer Science and Engineering, USA; Se jun Kim, University at Buffalo, State University of New York, Department of Computer Science and Engineering, USA; Jiyang Li, University at Buffalo, State University of New York, Department of Computer Science and Engineering, USA; Wenyao Xu, University at Buffalo, State University of New York, Department of Computer Science and Engineering, USA; Zhanpeng Jin, University at Buffalo, State University of New York, Department of Computer Science and Engineering, USA, zjin@buffalo.edu.

2474-9567/2022/6-ART48 $15.00
https://doi.org/10.1145/3534597

1 INTRODUCTION

Video streaming services and on-demand streaming have been all the rage the past decade. It was reported that the global video streaming market is expected to grow at a 21% compound annual growth rate (CAGR) from 2021 to 2028, with a forecasted market revenue of $223.98 billion [39]. Another study also suggests that U.S. viewers have already moved on from traditional pay-TV to streaming services with related streaming expenses outpacing that of pay-TV by 2024 [36]. Coinciding this market trend, many media and service providers have explored various approaches to capture users' behaviors, preferences, and emotional reactions on the fly, through video watching habits. This allows for sharply tuned personalized recommendations and accurate target advertising. Unfortunately, such behavioral user information was acquired through manual methods, such as questionnaires, which aren't taken seriously by consumers and often skipped or given random answers to move on to the content they desire. Hence, it is imperative to investigate alternative recognition methods which are both effective and efficient.

Facial expressions have been deemed as the universal non-verbal language to express internal emotional states. Moreover, studies revealed that spontaneous facial expressions are more correlated with the true emotions of users than posed expressions [69, 87]. In the late 1970s, Dr. Ekman proposed the universality of six facial expressions: happiness, sadness, anger, disgust, fear, and surprise [26, 28]. Contempt, the least researched emotion, was later recognized as a universal facial expression[62]. Each facial expression can be coded by the Facial Action Coding System (FACS) from specific muscles that are involved in particular facial expressions. These facial expressions are also accompanied by unconscious body postures. For example, when people face moments of surprise, they make the fear expression, as well as tend to lean their bodies backwards which is considered a natural reflex. However, many facial expression recognition systems in literature are based on when the user is in a static state, i.e., users are required to refrain from moving their bodies. However, if users make a facial expression in a static state that doesn't allow them to change their body postures, they would be potentially obstructed from eliciting true emotions. Therefore, it is imperative to capture the user's body postures while recognizing their facial expressions, which will be elaborated in Section 3.2.

Among existing solutions for recognizing and understanding facial expressions, camera-based methods using computer vision techniques are arguably the most prominent approaches [16, 89, 96]. However, due to privacy issues, people are reluctant to be video recorded when watching movies or engaging in other activities. Moreover, cameras are severely affected by various lighting conditions and need to be adjacent to the user, which is not always applicable in real-world applications. Another well-known approach for facial expression recognition leverages WiFi signals [17, 43], which always needs at least an emitter and a receiver, while being in close proximity to the user. Researchers have also investigated the potential usage of wearable sensors for recognizing facial expressions [15, 16, 61], which however, still mostly rely on cameras which can't capture the entire face in a convenient way.

The photoplethysmography (PPG) uses a simple optical technique that detects blood volume changes from the wrist [29] or fingertip [44] from different light sources: green, red, or IR light. The PPG sensor has an edge over other modalities in being more cost-effective, and portable, making it suitable for many wearable devices with very small form factors. Several prior studies utilized PPG signals to recognize facial expressions: the PPG signals were acquired using a camera positioned in front of the user's face [46] or a camera embedded in a glass-style device [53]. Given the vulnerability of motion artifacts, most prior studies, for PPG-based emotion detection, collected PPG data under strictly controlled environments [56, 77]. To this end, we propose a new design form factor that integrates a miniature PPG sensor into earable devices, which have seen significant growth and popularity in the market. Actually, in a recent patent filed by Apple [74], a PPG sensor was added to the earbud so that the device can monitor biometric data collected, which suggests the availability of such PPG sensors in future commercial earable devices.

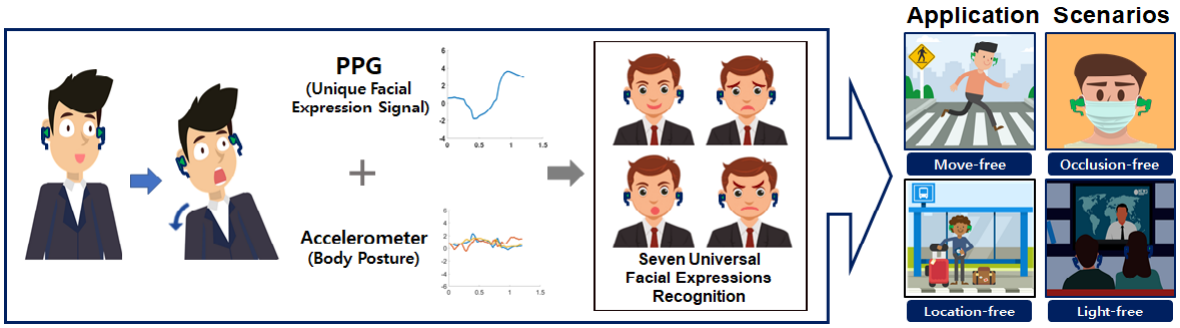


Fig. 1. PPGface Blueprint : Facial expression recognition system using re-interpreted PPG signals and an accelerometer, captured during video watching. When making a facial expression, PPGface captures unique PPG signals induced by the elastic deformation of blood vessels and spontaneous body postures with the aid of an accelerometer. PPGface is able to be employed ubiquitously: anytime, anywhere, any occlusions such as mask-wearing.

To enable a more accessible, easy-to-use, and privacy-preserving facial expression recognition system, this study proposes PPGface — an ubiquitous facial expression recognition platform. As shown in Fig. 1, leveraging in-ear PPG signals with complementary accelerometer data, PPGface aims to detect and recognize the user's facial expressions without interference (e.g., occlusions, lighting conditions, and motions). Specifically, PPGface intends to understand various facial expressions through the dynamic PPG patterns that are collected inside the ear and altered by the facial muscle motions. As illustrated in Fig. 1, the proposed PPGface could be potentially applicable to a wide variety of application scenarios in peoples' daily lives. For example, the WiFi-based facial expression recognition system requires additional hardware and is unsuitable for outdoor activities, and performance is negatively affected when the user wears a mask. The camera-based facial expression recognition system is vulnerable to ambient light situations and raises serious privacy concerns. Different from those aforementioned modalities, PPGface instead understands facial expressions from facial muscle movements detected inside the ear. When people make a facial expression, facial muscles called mimetic muscles control facial expressions supplied by the facial nerves. When making different facial expressions, different muscles contract and relax to make a certain expression and these muscles directly or indirectly affect not only the physical deformation of the ear canal, but also different skin layers, arterial blood vessels, and muscles. Unlike other modalities such as acoustic sensing or pressure sensing which can only utilize ear canal deformation, PPGface expands the one dimensional deformation to a higher dimensional deformation by employing the data from multiple skin layers. PPGface works as different wavelengths of light that can penetrate different skin layers. We utilize green, red, and IR light sources to collect mixed physiological signals from different skin layers. To complement the PPG signals, we leverage the accelerometer sensor to capture spontaneous body postures which indicates certain emotions, such as the act of leaning backwards when surprised. Then, we designed a multimodal Residual neural network (ResNet) to extract meaningful features from both the PPG and accelerometer signals, and proceed to classify the seven universal facial expressions shown in Fig. 6: contempt, happiness, disgust, anger, surprise, fear, and sadness. To evaluate the proposed solution, we conduct experiments with 20 human participants.

In summary, we highlight the following contributions in this work:

(1) This paper introduces PPGface, a novel facial expression recognition system, which establishes a user-friendly, ubiquitous system using earphones that recognizes facial expressions and circumvents any user involvement. It also captures the user's body postures with the accelerometer sensor. To the best of our knowledge, this work is the first of its kind leveraging in-ear PPG signals with the assistance of an accelerometer to recognize the seven universal facial expressions.

(2) This study addresses the following technical challenges: 1) capturing fine-grained facial expression signals, 2) distinguishing major facial expressions from other irrelevant activities (e.g., yawning or swallowing saliva), and 3) localizing the exact start and end points of the irregular facial expression signals with varying lengths.
(3) This paper conducts several comprehensive experiments (N=20) and validates system performance under different real-world conditions, including different body motions, wearing a mask, sensor mount positions. Specifically, a longer-term usability study, which reproduces the real-world environment without any restrictions, and user assessment implicate PPGface's ability of distinguishing seven major facial expressions in daily life.

The rest of the paper is organized as follows. We review the related works in Section 2 and portray a detailed background of PPGface in Section 3. Section 4 demonstrates the proof-of-concept feasibility studies and results. Section 5 describes system design, detailed design and algorithms of the PPGface. We present an extensive performance evaluation from Section 6 to 8. Section 9 and 10 will discuss comparisons with existing work, limitations, future works, application scenarios, and conclusions of this study.

2 RELATED WORK

In this section, we will review the related work of this study. Specifically, four different perspectives, including facial expression recognition, facial expression recognition with wearable devices, in-ear sensing, and motion-invoked PPG deformation are discussed.

2.1 Facial Expression Recognition

The intuitive way to recognize people's facial expressions is by looking at their faces. In this manner, computer vision (CV) based approaches are one of the most pervasively studied areas in facial movement recognition. Visual-based approaches employ various kinds of cameras, including the RGB camera [81], IR camera [51], or depth camera [49]. Noticeable advancement in deep learning, especially the CNN which specializes in image classification, supported researchers in developing several deep-learning based facial expression recognition systems. Zeng *et al.* [95] proposed the deep sparse autoencoders (DSAE) which extracts meaningful features in an unsupervised manner by preserving crucial data after removing input redundancies. Zhao *et al.* [98] extracts the face landmarks and textures of the eye region to detect fatigue in users with a single camera. They train the two stacked autoencoder neural networks to construct a bi-modal deep learning network. Moreover, in health care, pain recognition is considered an important task. Rodriguez *et al.* [82] constructs long short-term memory (LSTM) to detect the pain in facial expressions. By inputting raw images to the LSTM model, their system outperforms other approaches in pain intensity estimation. However, as aforementioned in the previous section, visual-based approaches have a critical weakness of raising privacy concerns. Moreover, cameras usually need to be placed in front of the users to continuously track their facial expressions, which is impractical.

Besides the CV-based approach, few studies have adopted WiFi signals to recognize facial expressions. Chen *et al.* [17] proposed a WiFi based facial expression recognition system to classify six different emotions by using a laptop and three antennas, which were placed 90cm away from the user. However, this study disregarded the impact of body motion, which is one of the major causes of signal interruption between antennas and laptops. Also, tracking facial expressions with extra hardware devices is unrealistic. Compared with those existing approaches, our proposed PPGface doesn't require any additional device setup, while it can be easily used in daily life due to its higher usability and easier accessibility.

2.2 Facial Expression Recognition with Wearables Devices

Since facial expressions are mainly concerned with facial muscles, wearable devices typically used in studies include eyewear devices, earable devices, neckband devices, or devices attachable to the face. Sheirer *et al.* [83] adopted a glass-type device with piezoelectric sensors to recognize facial expressions, and this was believed to be one of the first attempts to deploy wearable devices for facial emotion recognition. Furthermore, other studies use photo reflective sensors on eyewear devices to capture the facial expression from the upper part of the face [60, 61]. Xie *et al.* [93] on the other hand, proposed a glass-mounted acoustic sensing system for upper facial action recognition and applied an orthogonal frequency division multiplexing (OFDM)-based channel state information (CSI) estimation to recognize six upper facial actions. Although they use contact-free sensors, the sensors are vulnerable to ambient light, a critical drawback as a wearable commodity device. When recognizing facial expressions, earable devices are good candidates because ear canal deformation occurs when users make facial expressions, which in turn is captured by the earable device. Amesaka *et al.* [2] utilizes a speaker and microphone by incorporating the microphone into the earphones and used acoustic sensing to capture the ear canal deformation when the user makes facial expressions. Lee *et al.* [57] proposes a kinetic sensor mounted ear device to recognize facial expressions and use the 6-axis inertial measurement unit (IMU) to capture two facial expressions; happiness and frowning. However, they only classified two facial expressions and didn't provide a comprehensive analysis under different environments and disregarded the issue of unwanted motion artifacts. ExpressEar [88] recognizes 32 different AUs by employing accelerometer and gyroscope, and achieves 89% accuracy. As facial expressions are controlled by a set of facial muscles, the EMG sensor can easily record muscle activities using electrodes attached to the skin surface. Gruebler *et al.* [42] proposes a wearable device embedded with an EMG sensor and shows that the device can continuously track smiling and frowning. Hamedi *et al.* [45] recognizes 8 facial expressions by employing three channels in the bi-polar configuration surface EMG (sEMG) sensors, which are attached to the face. These showed relatively good performance levels, but embedding the sensors into commercial products are hard, and people usually feel uncomfortable wearing these sensors on their faces.

2.3 In-Ear Sensing

In the past two decades, wearable and mobile devices have become daily necessities. The ubiquity of wearable devices has created a wealth of new opportunities that leverage the teeming information provided from sensors. As advancements are made in miniaturized and ultra-low-power electronics, society is engrossed in a new group of smart wearable devices. "Hearables" [22], going beyond the traditional scope of sound-listening earphones, are defined as next-generation, wireless, in-ear computational earpieces that are capable of multiple functions ranging from fitness tracking, medical monitoring, hearing enhancement and aid. With a rich set of sensors, hearables could potentially become an essential pervasive sensing platform, not only for health and activity monitoring, but also for security and authentication.

Bui *et al.* [13] demonstrated the "eBP" as an in-ear blood pressure monitoring device with a pulse sensor and air pump. Nakamura *et al.* [67] designed an in-ear EEG device and explored the applications in sleep monitoring and user authentication [68]. Goverdovsky *et al.* [37, 38] constructed the "Hearables", consisting of an earpiece made of earplug foam with two EEG electrodes and a microphone. Using multimodal sensors, the earpiece can measure the user's respiratory, brain and cardiac activities. Gao *et al.* [33] demonstrated an in-ear device "EarEcho", embedding the microphone into commercial earpieces to leverage the physical and geometric uniqueness of an individual's ear canal as biometrics. Wang *et al.* [90] proposed "EarDynamic" with ear canal deformation based user authentication using a speaker and microphone. It analyzes the reflected signals from the speaker and extracts fine-grained discriminatory features for user authentication. Bi *et al.* [8] demonstrated the "Auracle", which recognizes eating behaviors by analyzing the chewing sounds that pass through the bones and tissues

of the face, using commercial microphones placed behind the ears. Moreover, Passler *et al.* [71] explored the possibility of measuring the in-ear PPG pulse rate as a valid alternative to the ECG-derived heart rate, using two commercial in-ear devices. All research efforts mentioned have laid the foundation for this work, to seek a more convenient, unobtrusive, inexpensive, and secure authentication solution leveraging the emerging hearable sensing technology.

2.4 Motion-Invoked PPG Deformation

PPG signals are normally used to measure the heart rate, blood pressure, or respiration; namely cardiac activities. PPG signals are also prone to motion artifacts, so data is collected from people in an immobile stance. Therefore, many studies have proposed solutions to remove motion artifacts to enhance the quality of data [52, 78, 79]. Recently, studies have started to explore the potential use of PPG motion artifacts as a discriminating signal serving as an input to classify different gestures. Zhao *et al.* [99] proposed a gesture recognition system using PPG signals and motion sensors embedded in a wrist worn device that can discriminate finger movements. Boukhechba *et al.* [10] introduced an activity recognition system from wrist-worn devices, applying CNN and recurrent neural network (RNN). They extract useful features from the wrist PPG and predict five different activities. In addition, Papapanagiotou *et al.* [70] showed a chewing detection system based on PPG signals, audio and accelerometer. A prototype of chewing sensors was developed and PPG sensors were placed on both sides of the ear concha. Besides the classification problem, Shang *et al.* [84] studied whether the PPG motion artifact can be used as an authentication system. Observing that muscle and tendon movements affect blood flow, they leveraged wrist worn PPG signals to authenticate the user. Likewise, PPGface employs in-ear PPG signals with unique patterns and information for different facial expressions of the user.

3 BACKGROUND

3.1 Muscles of Facial Expression and Elastic Deformation of Blood Vessels

Ekman and Friesen [31] proposed the facial action coding system (FACS), given a set of different action units (AU), to explain muscle movements and represent the subtleties of facial expressions clearly. For example, controlled by the zygomatic major muscle, the lip corner puller (AU-12) pulls the lip corners obliquely towards the cheekbone and is typically observed from a happy or contempt expression. Given the inter-correlation of the human muscular system, some facial expressions involve concurrent contractions of different muscles while others encompass asynchronous contraction and relaxation of several facial muscles. This enables endless facial expression possibilities; among them are the seven universal facial expressions of emotions that come with muscle and jaw movements. Several studies have proved that the ear canal moves along with certain muscles and the jaw. Grenness *et al.* [40] used a reflex microscope to prove the variability of the ear canal movement, which went against previous studies that concluded the ear canal was widened only when there were jaw movements.

Fig. 2a and Table 1 show facial muscles associated with different facial expressions. Additionally, Berzin [7] used EMG signals to detect spontaneous motions. Strong action potential was observed when the user opens the mouth and makes a happy face from the *Auricular* muscles, which consist of three different muscles around the ears next to the ear canal. As interrelated facial muscles and jaw movements induce ear canal deformation, the blood vessels around the ear canal, located above the muscle layer, are affected accordingly. To support our hypothesis, we can see the work by Cotofana *et al.* [19], which focused on a facial expression (i.e., smiling) influencing the blood vessels. Their work deployed the Doppler ultrasound imaging of facial blood vessels to measure the blood flow. It was observed that, from most of the subjects, the blood flow stopped during smiling after the 3 second retrograde flow interval, suggesting that the blood flow in the facial area can be influenced by contractions of the zygomaticus major muscle during smiling. Moreover, it was also reported [20] that, the facial blood vessels around the cheek area located subjacent to the zygomaticus major muscle lies in contact with the

Table 1. Facial expression associated AU, muscles, and joints. [21, 23, 94]

Facial Expression	Action Unit (AU)	Related Muscles / Joints
Happiness	AU 6, 12, 25	Orbicularis oculi, Pars orbitalis, Zygomaticus major, Depressor labii inferioris, Mentalis, Orbicularis oris, Auricular
Contempt	AU 6, 12, 14, 25	Orbicularis oculi, Pars orbitalis, Zygomaticus major, Depressor labii inferioris, Buccinator, Mentalis, Orbicularis oris, Auricular
Disgust	AU 6, 12, 20, 25	Orbicularis oculi, Pars orbitalis, Zygomaticus major, Risorius, Platysma, Depressor labii inferioris, Mentalis, Orbicularis oris
Anger	AU 4, 6, 9, 25, 26	Corrugator supercilii, Depressor supercilii, Orbicularis oculi, Pars orbitalis, Levator labii superioris alaeque nasi, Depressor labii inferioris, Mentalis, Orbicularis oris, Masseter, Medial pterygoid, Lateral pterygoid, Temporalis
Sadness	AU 1, 4, 15, 17	Frontalis, Pars medialis, Corrugator supercilii, Depressor supercilii, Triangularis, Mentalis
Fear	AU 1, 2, 5, 7, 27	Frontalis, Pars medialis, Pars lateralis, Levator palpebrae superioris, Pars palpebralis, Pterygoids, Digastric, Temporomandibular joint (TMJ)
Surprise	AU 1, 2, 5, 26, 27	Frontalis, Pars medialis, Pars lateralis, Levator palpebrae superioris, Masseter, Temporalis, Pterygoid, Digastric, Temporomandibular joint

maxilla, and is able to be compressed by the overlying muscle upon its contraction. The above explorations in literature demonstrates that, facial muscle activities associated with facial expressions can cause blood vessel deformation.

It is observed that, for the happy, contempt, disgust, and anger expressions, the set of AU 6, 12 and 25 dominantly affect the ear canal and elastic blood vessel deformation, causing PPG signal fluctuations. For the remaining expressions, AU 1, 26 and 27 primarily affect the ear canal geometry and PPG signals. At first glance, some PPG signal fluctuations seem visually similar for different facial expressions. However, different duration and intensities result in unique signal patterns; steeper and shorter (contempt), or stronger and longer duration signals (disgust), which are illustrated in Fig. 2b.

Thus, it is argued in this study that PPG variations resulting from an individual's facial expressions possess important and intrinsic patterns with regard to the universal facial expressions because they stem from different physical conditions of the human muscular and circulatory system. In this respect, we confirm that this physical deformation could be leveraged for recognizing individuals' facial expressions in this study.

3.2 Emotions with Coupled Facial Expressions and Body Postures

Reading facial expressions is one of the most direct and accurate ways to understand people's emotions. However, without certain body postures, it may be hard to identify the intended facial expression. For example, people may be confused between a fear and surprise expression as both have raised eye brows and an open mouth, which shows unexpectedness and novelty [24, 97]. Researchers have shown that body posture helps understand facial expressions more accurately. Reed *et al.* [80] examined how a restrained body posture can influence emotional expression and concluded that this decreases the recognition of the viewer's emotion, as it becomes difficult to differentiate between two similar emotions, such as disgust and anger. Veenstra *et al.* [64] showed that people recall more negative emotions when they are sitting in a stooped, not upright posture. Therefore, previous research shows employing a certain body posture is just as important as the facial expression itself to convey more accurate and truer emotions.

In this paper, contrary to previous studies that focused on facial expressions without the consideration of body postures, we incorporated the accelerometer sensor to capture spontaneous and natural body postures that measures directional movement from acceleration of the device. Zhao *et al.* [99] demonstrated through a comprehensive study that the IMU sensor (using both the accelerometer and gyroscope) is insufficient at

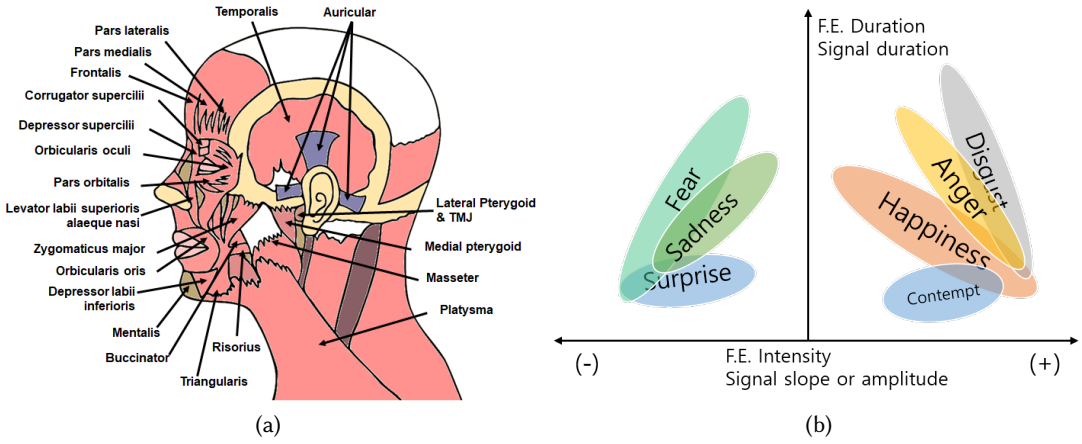


Fig. 2. Illustrations of (a) facial muscles of the side view, (b) facial expression category based on the intensity and duration. (F.E. : Facial Expression)

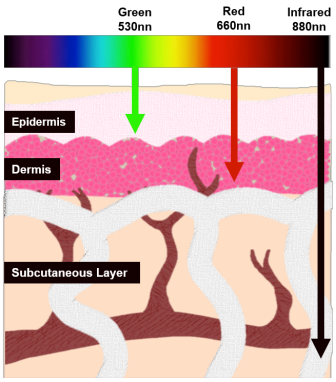


Fig. 3. Diverse skin penetration rates for different wavelengths of light

capturing finger-level micro movements, but works for fine-grained wrist movements. Typically, a wearable device's accelerometer sensor is placed close to the skin surface, which helps acquire clear and accurate signals. However, for earable devices, the accelerometer sensor is mostly embedded inside the device due to a risk of injury to the ear canal's inner skin. This makes it difficult for the earable device's accelerometer sensor to capture micro-level facial muscle movement, which occurs deep down in the skin. Lee *et al.* [57] observed that only a few among the 30 basic AUs show noticeable signal variations using both the accelerometer and gyroscope sensor. This demonstrates that solely using the accelerometer cannot successfully capture micro movements of the muscles. Even if we can somehow capture facial expressions using this, it can be heavily corrupted by subtle head motions, which we will elaborate on in the next section.

3.3 PPG Signals and Blood Vessel Deformation

The PPG sensor captures circulatory signals with at least one LED and photodiode that measures the pulsatile volume of blood circulation. The LED is used to emit light within the skin layers, and the increased pulse of the heartbeat causes different light patterns that are reflected back to the photodiode. PPG signals have been commonly measured from various body locations: fingertip, wrist, toe, and forehead [34]. Researchers have recently started to investigate alternative ways of acquiring stable PPG signals inside or around the ear [11, 12, 71, 72], and a similar idea has also been adopted by industry manufacturers in future smart earbuds, as shown in Apple's most recent patent application [47, 75].

As shown in Fig. 3, human skin can be divided into three layers: the epidermis which is the superficial skin layer that provides skin color and contains no blood vessels [66], the dermis which contains capillary blood vessels and arteries, and lastly the subcutaneous layer which contains fat, connective tissues and larger blood vessels. Muscles are located below these skin layers, therefore affecting all skin layers whenever there are muscle activities. Researchers discovered that different wavelengths of light have different penetration rates [3]. Lights with shorter wavelengths, such as the green light can only penetrate the upper layers of the skin, i.e., the epidermis or the upper part of the dermis. Contrarily, lights with longer wavelengths such as red, or IR lights can penetrate deeper into the skin. As light travels through different skin layers and is reflected back to the photodiode sensor, various light sources contain a unique mixture of information from different blood vessels, as well as subtle movements of tissues and muscles.

It has been shown in prior studies that PPG signals are affected by the blood vessel deformation and tissue motions [30, 48]. Instead of normal PPG signals captured from the resting stage, PPGface re-interprets the fluctuated PPG signals induced by facial muscle movements under the skin. Hence, it is imperative to examine the variations of all skin layers discussed above and involve all light color channels. To this end, in this study we propose to incorporate three different wavelengths of light to capture the unique deformation patterns of each skin layer and blood vessel.

4 PROOF-OF-CONCEPT FEASIBILITY STUDY

We provided a comprehensive background in the previous section to describe the research rationale of our hypothesis that the underlying facial muscle movements and corresponding blood vessel deformation can be captured and gauged using in-ear PPG and accelerometer sensing. To further verify the validity and feasibility of the proposed approach in recognizing facial expressions, it is necessary to investigate how the facial expressions are related to the in-ear PPG signals. Hence, we conducted two feasibility studies: 1) to quantify the Action Units from in-ear PPG signals when the user makes expressions, and 2) to explore the feasibility of using PPG and accelerometer sensors to recognize facial expressions.

In all of the following studies, including the feasibility study, we used the MAXIM86161EVSYS evaluation kit [50], which consists of a battery, a PPG sensor, an accelerometer sensor, the PCB board, and a Bluetooth dongle for wireless communication. The sampling frequency was set to 128 Hz, integration pulse width to 58.7 μs which controls the exposure time of the LED, and three LEDs (i.e., green, infrared (IR), and red) to 10.21 mA, consuming 559.8 μs LED supply current in total. The wavelengths of green, IR, and red light sensors are 530 nm, 660 nm, and 880 nm, respectively. Along with the LEDs, they are equipped with 3-axis accelerometer sensors (i.e., x-, y-, and z-axis), synchronized with PPG sensors. Fig. 4 illustrates sensor orientation and the depiction of the accelerometer axes of the sensor, and Fig. 5 shows the experimental setup of the PPGface prototype worn by the user. As can be seen from Fig. 4, we wrap the sensor with a normal silicone earbud tip with ear loops to provide better comfort and make it similar to market products, so that the PPG and accelerometer sensors don't irritate the ear canal skin. Please note that the sensor is flipped on the z-axis in Fig. 4 for purpose of showing the front layer of the sensor, which means participants actually wore the device towards the back of the ear in this study.

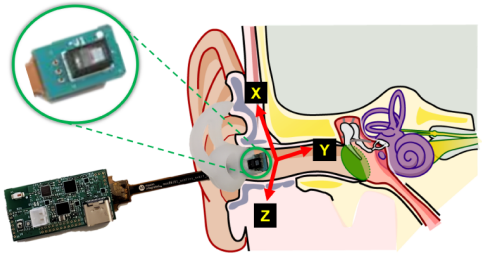


Fig. 4. The sensor orientation and the depiction of the accelerometer axes of the sensor

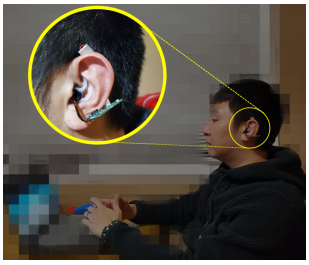


Fig. 5. Experimental setup of the PPGface prototype worn by the user

The anterior part of the ear is supplied by the anterior auricular artery originating from the superficial temporal artery [73]. The posterior auricular artery, originating from the external carotid artery, supply the posterior part of the ear, but frequently reach the anterior part [91].

4.1 Quantification of Action Units

For the first sub-study, we recruited two participants (one male and female) for data collection. Participants wore the device sitting on a chair comfortably. Since they are not professional actors, we asked them to practice until they were familiar with making designated facial expressions. We then asked the participants to make different action units from Table 1 ten times each and recorded the corresponding PPG signals. To quantify the relationship between different AUs and PPG signals, we calculated the signal energy of the PPG signals and mapped the intensities to the corresponding facial regions. Please note that this is just to help understand how the signals are related to the facial muscles. As can be seen from Fig. 7, we quantified different AUs from PPG signals by highlighting related facial muscles with different colors (Color-bar represents the signal intensity ranging from -1 to 1, where the sign and numerical value indicate the shape – convex down or up – and intensity of the signal).

As expected, AUs that are not directly related to the ear canal show lower energy (AU1, AU4, AU5, AU7, AU9, and AU17). In particular, AUs around the eye and eyebrow region show relatively weaker strengths than the mouth and cheek regions. For happiness, contempt, disgust, and anger expressions, it is observed that AU 12 (lip corner puller) and AU 25 (lips part), which involve *Zygomaticus major*, *Depressor labii inferioris*, *Mentalis*, and *Orbicularis oris*, show stronger signal relations than other AUs. On the other hand, AU 26 (jaw drop) and 27 (mouth stretch), which involve *TMJ*, *Pterygoids*, *Masseter*, *Temporalis*, and *Digastric* muscle, affect PPG signal fluctuations for the surprise and fear expressions. For the sadness expression, AU 15 (lip corner depressor) and AU 17 (chin raiser) alter PPG signals. Although the signal intensity of each individual AU is not sufficiently distinguishable, when a set of AUs are combined together to form facial expressions, PPG signals are amplified and exhibit unique patterns for different expressions due to the inter-correlation characteristics of the muscular system. This sub-study also proves that facial expressions sharing the same AUs can be categorized together as illustrated in Fig. 2b.

4.2 Distinguishing Facial Expressions from Motion Artifacts

4.2.1 Fusion of PPG and Accelerometer Data. This study is to investigate the feasibility of using both the PPG and accelerometer sensors in commercial earable devices. To show consistency, we involved one participant in making seven universal facial expressions 10 times, respectively. We then input the signals into the PPGface system, which will be elaborated in Section 5. Fig. 6 illustrates a user's mesh-grid face, PPG and accelerometer signals for each facial expression. It is observed that, PPG signals show distinct patterns for different facial expressions.

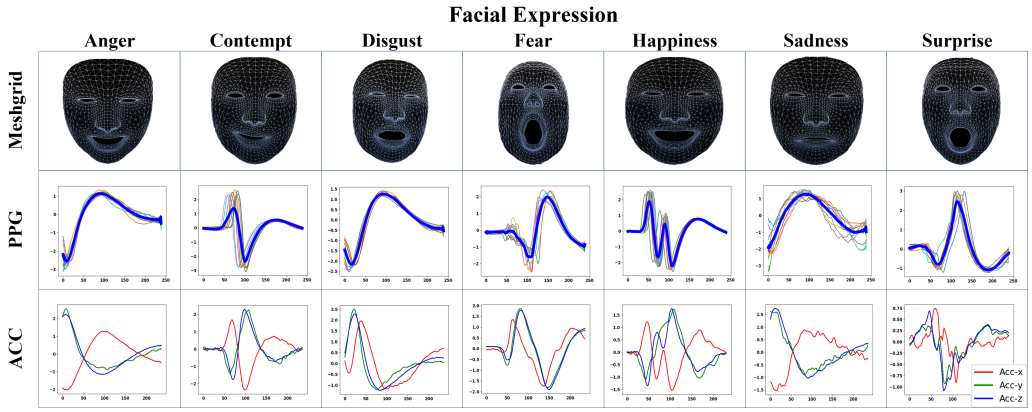


Fig. 6. Meshgrid face image, PPG, and IMU signals of 7 universal facial expressions. (x-axis and y-axis show time in seconds and normalized signals, respectively)

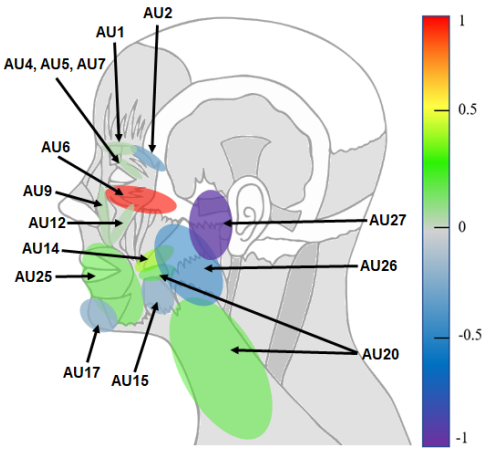


Fig. 7. Illustration of AU quantification, where color bar represents the signal intensity

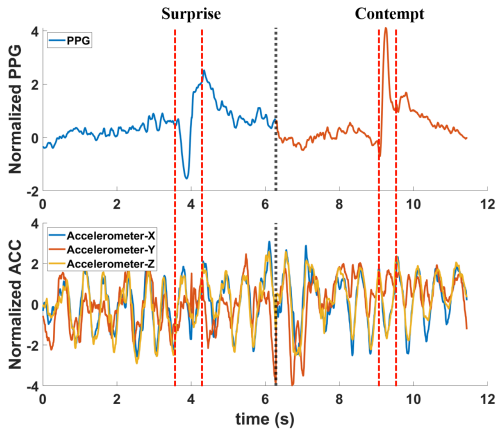


Fig. 8. Example of two facial expression with continuous head motions

From the second row in the figure, an average of 10 PPG signals are plotted with a thicker blue-colored line, where 10 PPG signals are plotted with narrower lines shown in back. For the accelerometer signals, we also plotted the average 10 signals from the third row. The 30 lines of x, y, and z-axis accelerometer signals are not shown in a single plot as it becomes messy, distracting, and hard to recognize. As mentioned before, although some expressions share one or more AUs, signal details show different steepness and duration of the signals. For some expressions, the accelerometer signal also shows noticeable changes, however those signals show almost the same patterns.

4.2.2 Impact of Head Motions. In daily-life scenarios, facial expressions are often accompanied by voluntary head motions, which could be much more significant than the facial muscle motions. Hence, we sought to explore whether the proposed combination of PPG and IMU data could be robust to detecting facial expressions

interfered by various degrees of head motions. We asked the user to exhibit two facial expressions (i.e., surprise and contempt) and keep moving the head naturally, such as turning the head to find the direction when walking or checking the flight time. As shown in Fig. 8, PPG signals are captured accurately regardless of any head motions. The accelerometer on the other hand, only captures body motions over facial expressions as the size of motions is much larger. Since the accelerometer is directly affected by the device's directional movement, it is inevitable for the accelerometer to capture all kinds of movement-related signals in addition to facial muscle motions. This implies that the accelerometer is better at capturing bigger motions. In contrast, as head motions are affected mainly by neck muscles and PPG signals only reflect the blood vessel deformation around the ear canal caused by facial muscles, PPG signals are less vulnerable to motion artifacts given its unique device location (i.e., inserted and fixated on the ear canal surface), which are dominantly affected by the facial expressions. This finding indicates that: 1) the in-ear PPG sensor is more robust to head motions in recognizing facial expressions, while the accelerometer sensor is only suitable under restricted scenarios without significant head motions, and 2) the accelerometer can be a perfect auxiliary modality which can help capture spontaneous body posture.

Verma *et al.*[88] captured different AUs using a 6-axis IMU sensor, consisting of an acclerometer and gyroscope inside the ear canal. One of the critical limitations of this work is the vulnerability of the IMU sensor to motion artifacts, especially head motions. These corrupted signals are not easy to be segmented and separated, which severely weakens the practicality and ubiquitousness of this system to be integrated in real life. For real world applications, the sensing modality and device needs to be selected carefully.

5 SYSTEM DESIGN

5.1 System Overview

The proposed PPGface system aims to recognize the user's facial expression by unifying both the physiological (PPG signals from the muscle movement) and behavioral (IMU signals from the body posture) patterns into one synergistic process. As shown in Fig. 9, this system is mainly composed of four parts: facial expression detection, signal filtering, dynamic signal preprocessing, and facial expression classification via multimodal ResNet.

When a user makes an expression, the device measures the in-ear PPG and accelerometer signals. The measured PPG exhibits clear differences between the resting stage and when the user is making a facial expression. If the user is either in the resting state or performing other physical activities such as swallowing saliva or yawning, the irrelevant PPG segments will be detected and removed for signal preprocessing.

Following the event detection stage, the PPGface signals are screened out and the detrend technique is applied to remove the DC components due to respiration, and then put through the bandpass filter, which filters out both the high and low frequency components to preserve the characteristics of both the facial expressions and related body postures.

The dynamic signal preprocessing stage includes two sub-stages, segmentation and detection. First, the continuous wavelet transform (CWT) technique and peak are used to segment all the possible sets of start and end points. Then, by analyzing these signals, the final start and end points of each segment corresponding to the facial expression activities are determined from the previous steps.

Finally, features are extracted from the preprocessed signals which will be the inputs of the classification module. In this stage, multimodal ResNet is used to extract the features and classify seven different facial expressions. During the training stage, the user profile is created and trained to distinguish the facial expressions. During the testing stage, for any unseen data, PPGface goes through feature profiling to classify seven different expressions based on pre-trained multimodal ResNet.

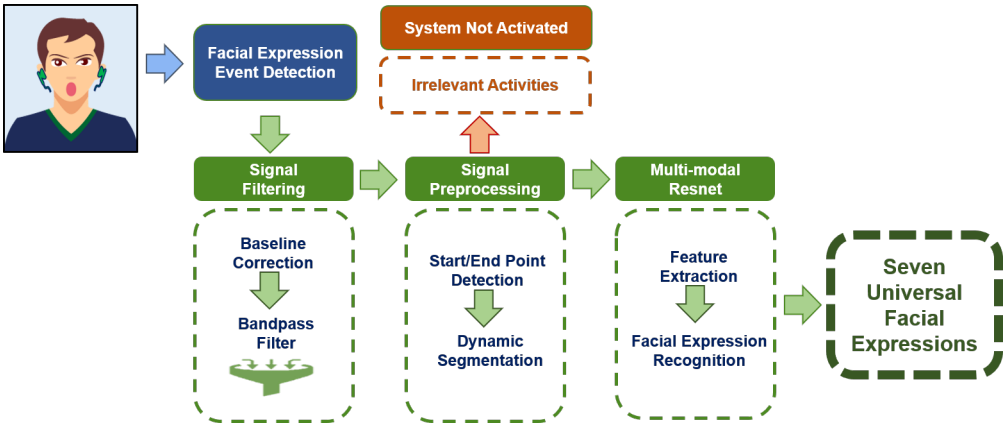


Fig. 9. PPGface system overview

5.2 Facial Expression Event Detection

Both facial and non-facial expression related activities can affect PPG signals. Irrelevant activities that cause facial muscles to contract, such as eating, yawning, or swallowing, should not activate PPGface. As seen from Fig. 10, happiness shows much higher amplitude than swallowing, and yawning has a longer duration and higher amplitude of both PPG and accelerometer signals. When people swallow, neck muscles affect the pharynx (known as the throat). The pharynx is connected to the Eustachian tube, which is a small passageway that connects the throat to the middle ear. As the Eustachian tube is connected to the ear canal, swallowing indirectly causes ear canal deformation. While spontaneous expressions normally last between 1 and 4 seconds [27], swallowing is a rapid process occurring within a second [63], with an insignificant signal fluctuation. Yawning takes 5.5 seconds to finish on average (ranging from 3 to 45 seconds [4]), and involves the posture of tilting one's head back and opening the mouth widely. Hence, non-facial expression related activities cause more transitory facial muscle movements resulting in sensor displacement, as discussed in Section 4. To filter out the irrelevant activities, we empirically chose the pre-defined threshold values by setting time duration and peak-to-peak values, which are shown as follows:

$$T_s(\text{duration}, \text{peakTOpeak}) = \begin{cases} 1, & \text{if (duration} < d_s + \alpha_s) \text{ and (peakTOpeak} < p_{s_{ppg}} \times \beta_s) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$T_y(\text{duration}, \text{peakTOpeak}) = \begin{cases} 1, & \text{if (duration} < d_y + \alpha_y) \text{ and (peakTOpeak} < p_{y_{acc}} \times \beta_y) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where d_s and d_y are the average duration time of swallowing and yawning. Moreover, $p_{s_{ppg}}$, and $p_{y_{acc}}$ are PPG and accelerometer signal's peak-to-peak values of the swallowing and yawning activities, respectively. When the T_s or T_y is equal to 1, the corresponding signals are filtered out; otherwise they will proceed to the next processing stage. As different users have different habits, parameters α_s and β_s are user specific regularization coefficients that need to be fine-tuned.

5.3 Signal Processing

PPGface aims to examine and utilize the PPG segments that result from and are aligned with the corresponding facial expressions. Therefore, it is critical to filter out unwanted noise arising from respiration and cardiac

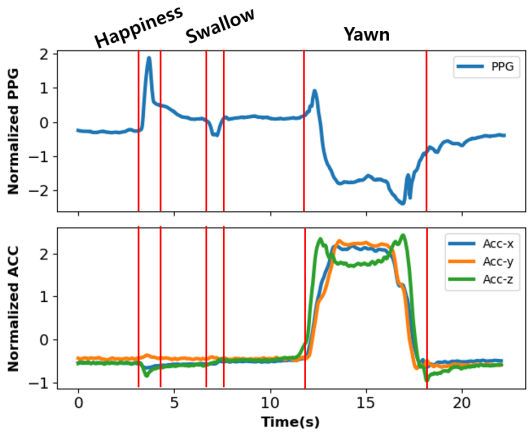


Fig. 10. Illustration of facial expression and irrelevant activities: happiness expression, swallowing, and yawning activity.

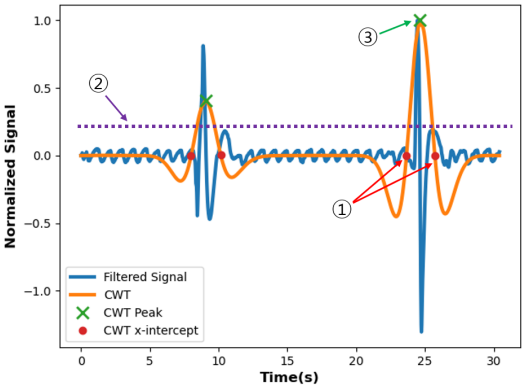


Fig. 11. Dynamic segmentation of PPGface

activities while preserving the facial expression and related body posture data. In addition, it is also crucial to localize accurate start and end points as people make facial expressions irregularly, resulting in different time intervals between two consecutive expressions and different signal lengths among users.

5.3.1 Signal filtering. To scale and restore the signals from distortion, signal filtering techniques are applied to the raw PPGface signals in this stage. First, the detrend technique using the convex hull approach is used to eliminate the DC components from the signals.

The *convex hull* of a point set $P \subseteq \mathbb{R}^d$, denoted conv(P), can be defined as the smallest convex set that contains P [5]:

$$\text{conv}(P) = \left\{ \sum_{i=1}^{n} \lambda_i p_i \mid n \in \mathbb{N} \wedge \forall i \in \{1, ..., n\} : \lambda_i \geq 0 \wedge p_i \in P \right\} \tag{3}$$

$\sum_{i=1}^{n} \lambda_i = 1$, and n is the number of data points. Here, the λ_i is the scalar coefficient that is used to construct from one point to another. After the signal is detrended, the 4th-order band-pass filter is used to get rid of the noise. we set the low cut-off frequency to 0.5 Hz as low frequencies (i.e., 0.01 − 0.5 Hz) are primarily caused by breathing, the parasympathetic and sympathetic nervous systems. In addition, the frequency range of the PPG signal is from 0.5 to 4 Hz, and the frequency of motion artifacts (e.g., moving head) is from 0.1 to 10 Hz [54, 55, 65]. Unlike previous PPG-related studies which eliminated the motion artifacts for a better signal quality, PPGface aims to capture the dynamic characteristics of PPG signals associated with facial expressions that are represented by the facial muscle movements and body postures. Hence, to preserve both characteristics, we set the high cut-off frequency to 10 Hz.

5.3.2 Dynamic Segmentation of Facial Expression. Contrary to previous papers that required the user to record themselves by pressing a key while making a facial expression, we designed to automatically detect the start and end points of PPGface signals [57, 93]. We proposed the dynamic segmentation by applying the continuous wavelet transform (CWT) and peak detection algorithm.

After the signals are conditioned and filtered from the previous steps, this stage seeks to segment all the possible candidates of start and end points using the CWT technique. The wavelet transform (WT) technique in

general, provides an optimal solution for most biosignals [76]. The WT analysis uses local wavelet transform to do the convolution operation with the signals using long windows at low frequencies and short windows at high frequencies.

Since the signals are time-series data and the length of the signals varies, CWT is an effective tool for detecting the motion signals especially in noisy data. Here, the noisy data refers to ordinary periodic PPG signals between PPGface segments. The rationale behind this is that the CWT has different window sizes that can provide optimal time resolutions for all frequency ranges, so it is capable of capturing abnormal signals better. Moreover, it is used to decompose time-domain signals into wavelets that are small oscillators localized in time, and then sum all of the signals. The WT of a continuous signal with the wavelet function is defined as [1]:

$$C_x(a,b) = \frac{1}{|s|^{1/2}} \int_{\infty}^{\infty} c(t)\bar{\psi}(\frac{t-b}{s})dt \tag{4}$$

where $c(t)$ is a time-domain signal, $\bar{\psi}$ is the complex conjugate mother wavelet divided by a scale factor s and minus a factor b where it can also be interpreted as a time location, and $\frac{1}{|s|^{1/2}}$ is an energy preservation. As can be seen from Equation 4, the CWT of a signal $c(t)$ is defined as the inner product in the Hilbert space of the L^2 norm. Then, the outcomes are the intersection points of two signals in the detection stage. We set the scale factor as 128 Hz and apply the Mexican hat wavelet which is the second derivative of the Gaussian function, defined by:

$$m(t) = \frac{1}{|e|^{-1/2t^2}} \tag{5}$$

Fig. 11 shows that the CWT can accurately detect the start and end points. The peak detection algorithm detects the peak by comparing the neighboring values with the moving window technique. Also, all the x-intercepts of the CWT are the possible start and end points of PPGface. After detecting the peaks and interception points, we calculate the average peak heights (i.e., y-axis value of the peak) to use as a threshold. That is to say, if the peak of the segmented signal is located in between two adjacent intercept points and above the threshold, the system detects an occurrence of facial expression activity.

5.4 Facial Expression Recognition Model

5.4.1 Data Augmentation. One of the main challenges of applying the deep learning model to a self-collected dataset is the limited amount of data; extensive amounts of data is needed to avoid the over-fitting problem and enable better generalization of deep learning models. Data augmentation is one of the most widely used techniques to mitigate the lack of data. Typical data augmentation methods rotate, scale, time-warp, and crop the original data to create new data [86], and we utilize these techniques for our self-collected PPG and accelerometer data. We first apply rotation by randomly rotating the original data to alleviate the individual's different sensor placement habits. Then the time-warping method is used to perturb the temporal location of the data in a single window by altering the time intervals between samples. Lastly, the scaling method, which changes the amplitude of the data by multiplying a random number, is adopted for data augmentation. In particular, a random number is selected from the Gaussian distribution with a mean of 1 and standard deviation (STD) of 0.1.

5.4.2 Multimodal ResNet. As the advancement of mobile and wearable devices which are equipped with heterogeneous sensors, researchers have explored multimodal sensing techniques in a wide range of real-life applications [6, 100]. The multimodal technique has been used to recognize facial expressions from eye movement and EEG [9], identify the user leveraging ECG and EEG [6], or quantify the anxiety level from EEG and PPG [100]. The basic idea of multimodal learning is to combine different modalities to understand and learn the data better as these modalities are correlated to each other, providing complementary information. Specifically, using both PPG and accelerometer modalities simultaneously can help detect and find the inter-correlation between facial

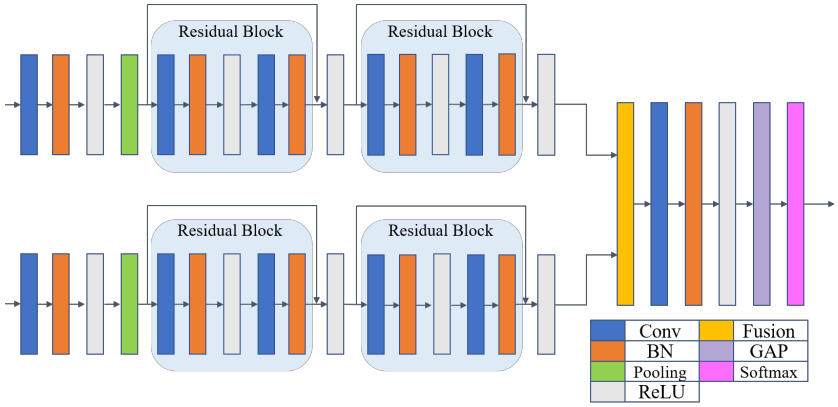


Fig. 12. Multimodal ResNet deep learning architecture proposed in this study (BN : Batch Normalization, GAP : Global Average Pooling)

muscle movement and spontaneous body posture. Moreover, multimodal learning is less prone to noisy data and overcomes the imperfect data acquisition conditions which can lead to false reject.

As can be seen from Fig. 12, this paper employs the Residual Neural Network (ResNet), which extends the CNN to a deeper structure with a skip connection in each residual block and mitigates the gradient vanishing problem. The input is added to link the output of a residual block and results the gradient to flow directly through the connections. Usually ResNet adopts the Global Average Pooling (GAP) layer instead of the Fully Connected (FC) layer. The GAP layer reduces several feature maps into one feature map for each class category for the classification task. Compared to the FC layer, the GAP layer is more robust to the overfitting problem since it has no trainable parameters to optimize. In addition, since it sums out the spatial information from different feature maps, it is also robust to spatial translations of the input [58]. Adopting ResNet is suitable for this study, which can also mitigate the data insufficiency problem along with data augmentation.

Therefore, in this study, we propose the 1D multimodal ResNet by combining the multimodal learning approach and applying feature-level fusion technique to learn correlated features between two modalities better. First, two different modalities are separately passed through the convolution layers consisting of 32 filters with the stride of 5 with the batch normalization (BN) technique. Then the ReLU activation function is applied and the features go through the Max-pooling layer. For the ResNet architecture, we design four residual blocks, where each block contains two convolution layers which consist of 64 filters with the stride of 3, the BN technique and ReLU activation function. Features from the two modalities are concatenated to be an input to the last convolution layer with 128 filters and stride of 3 followed by the BN technique and another ReLU activation function. This is lastly linked to a GAP and softmax layer; the latter classifies the seven universal facial expressions.

6 SYSTEM EVALUATION

6.1 Performance Evaluation

6.1.1 Procedure & Data Collection. In this study, we recruited 20 participants; 5 females (median age : 27) and 15 (median age : 22) males with ages ranging from 20 to 32 years old. Before the experiment, we sufficiently explained the objective of this experiment and potential hazards to participants. During the data collection process, they were asked to wear the PPGface device on the left ear while sitting comfortably on a chair. To elicit their emotional states through facial expressions, we asked them to recall the memory of certain emotions, which is a common way for the actors to evoke emotions. However, as none of the participants were professional actors

Table 2. Description of body posture for different facial expressions. [25]

Facial Expression	Description
Anger	Glaring others by leaning forward with chin jutting forward and wrinkling nose.
Contempt	Having upright posture and roll one's eyes.
Disgust	Turning the body away from the source of disgust.
Happiness	No certain posture (can be either upright or still).
Fear	Immobilizing by freezing one's body.
Sadness	Looking downwards with a hunched posture.
Surprise	Moving the head towards backwards away from surprising object.

or acting students, and to collect the data more similar to the real world, we also played short movie clip videos, which were carefully selected from the internet in random orders. The average and standard deviation of the length of video clips are 77.73 and 49.56 seconds, respectively. Please refer to Appendix A.1 for more details. Furthermore, before data collection, we showed them different AUs and facial expression images to help them better elicit their emotions and practice facial expressions. Then we asked them to follow the set of AUs for each facial expression from Table 1 with the proper body posture described in Paul Ekman's group [25], which is shown in Table 2. We collected each facial expression 50 times, allowing for a one minute break in between whenever the users felt fatigue. For each of the facial expressions, users were allowed to repeat the expression whenever they were ready, since the proposed system is designed to dynamically segment the irregular signals. The experiments were approved by the Internal Review Board (IRB) of the University at Buffalo, State University of New York for human subjects.

6.1.2 Evaluation Metrics. For performance evaluation, we adopted the most commonly used measurements in existing work [14], which are *accuracy*, *precision*, *recall*, and *F1-score*. The *accuracy* demonstrates the percentage of correctly verified data among the total number of data. As shown from Eq. 6 to Eq. 8, *precision* is the ratio of correctly verified positive data over the total number of positive data. The *recall* is the correctly verified positive samples over the total number of observations, and the *F1-score* is the weighted average of *precision* and *recall*.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Postivie} + \text{False Positive}} \tag{6}$$

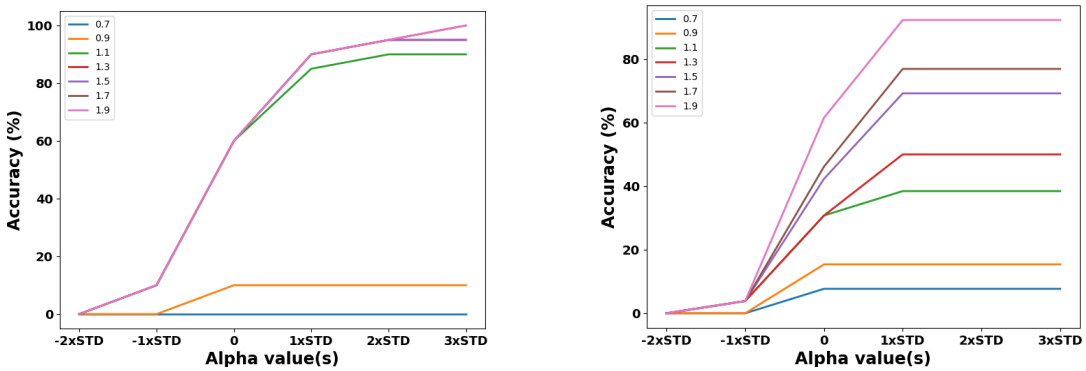
$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{7}$$

$$\text{F1}-\text{score} = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

6.2 Performance Evaluation of Event Detection Module

As can be seen from Fig. 9, non-speaking activities (or NULL class samples) are processed and filtered out in the signal preprocessing module before fully segmented facial expression signals are sent to the classifier. It will be problematic if the system is frequently turned on by mistake when the user performs non-speaking activities, such as yawning, head moving, swallowing, or masticating. Therefore, it is critical to distinguish the PPGface signals resulting from valid facial expressions from both the resting PPG and those caused by irrelevant activities. To this aim, we conducted an experiment to evaluate how well the PPGface can filter out the NULL class samples in the signal preprocessing module. In particular, we were interested in exploring how α and β values from Eq. 1 and Eq. 2 can affect performance. Hence, we asked two participants to yawn and swallow their saliva 20 times each. Since it is unnatural to swallow every few seconds, participants repeatedly performed swallowing activities

a few times between facial expressions. The average duration time of swallowing and yawning (i.e., d_s and d_y) are 0.9 seconds (SD = 0.12) and 8.03 seconds (SD = 1.71) respectively, whereas the peak-to-peak values of swallowing and yawning (i.e., $p_{s_{ppg}}$ and $p_{y_{acc}}$) are set to 0.87 (SD=0.12) and 2.67 (SD=0.21) respectively. As seen from Fig. 13, we set the α_y and α_s values from -2 × SD to 3 × SD and β_y and β_s values from 0.7 to 1.9. It can be seen from Fig. 13a that, when β_s is lower than 1, accuracy can't be improved beyond a zero α_s value. However, if the α_s value is properly set where the β_s value is greater or equal than 1, there is in upward climb in accuracy because swallowing activities show insignificant patterns compared to facial expressions, which can be easily discarded. This demonstrates that a signal duration time can affect the accuracy when filtering out the swallowing activity. On the other hand, from Fig. 13b, it is observed that the yawning activity is affected more by the β_y than the α_y value since accuracy doesn't increase when the $_y$ values are greater or equal than $1 \times STD$, suggesting that the peak-to-peak value should be carefully chosen to accurately eliminate the yawning activity.



(a) Swallowing activity (b) Yawning activity

Fig. 13. Event detection performance with different α and β values for different activities

6.3 PPGface Performance Evaluation

We quantify the PPGface results in terms of classification accuracy in this study with multimodal ResNet. Among the collected dataset, 80% was used to train the model and 10% was used for the validation and testing purposes respectively, with a batch size of 32 (i.e., 5 samples from each class per subject were used to evaluate the performance). Here, the sample corresponds to the fully segmented facial expression signals from the signal preprocessing module. We then apply the 5-fold cross-validation (CV) protocol to optimize the model and make it less biased. Fig. 14a and Fig. 14b show the confusion matrices of user-dependent and user-independent cases respectively from the 5-fold CV protocol testing samples. Moreover, we adopted the Adam optimizer to optimize the PPGface with setting the learning rate to 0.002 and weight decay to 0.001. As shown in Table 3, PPGface achieves 0.935 accuracy (SD = 0.086), 0.951 precision (SD = 0.086), 0.934 recall (SD = 0.074), and 0.934 f1-score (SD = 0.094). It is observed from the confusion matrix in Fig. 14a that fear and surprise expressions are relatively more confusing compared with other expressions. This is because facial muscles used to express surprise and fear overlap a lot. As aforementioned, target users of this paper are normal people who are not professional in acting, thus, it is difficult for them to precisely control their facial muscles.

We wanted to dig deeper into whether the proposed system can also perform well under the user-independent case. Different from the user-dependent case, we merged each facial expression without considering the user

Table 3. Classification accuracy for user-dependent and user-independent cases (SD = Standard Deviation).

	User-Dependent		User-Independent	
	Mean	SD	Mean	SD
Accuracy	0.935	0.086	0.848	-
Recall	0.934	0.074	0.854	0.074
Precision	0.951	0.086	0.854	0.048
F1 Score	0.934	0.094	0.842	0.054

label. From Table 3, accuracy is 0.848, precision is 0.854 (SD = 0.048), recall is 0.854 (SD = 0.074), and f1-score is 0.842 (SD = 0.054). Please note that accuracy has no standard deviation value as it is calculated by dividing the correctly classified number of samples by total number of samples. In addition, different from the user-dependent case, SD values show the deviation between seven different facial expressions, not between different users. It is understandable that performance becomes worse than user-dependent case because each participant has different expression habits. We believe that the accuracy is high enough and far beyond a random guess (i.e., randomly guess among seven emotions, i.e., expected accuracy is $\frac{1}{7} \approx 0.14$). Due to the slight difference in facial expressions among each individual, there are more misclassified data shown in Fig. 14b. Interestingly, observation from the earlier section (Section 3.1), expressions that share similar facial muscles are confusing the system more than other expressions.

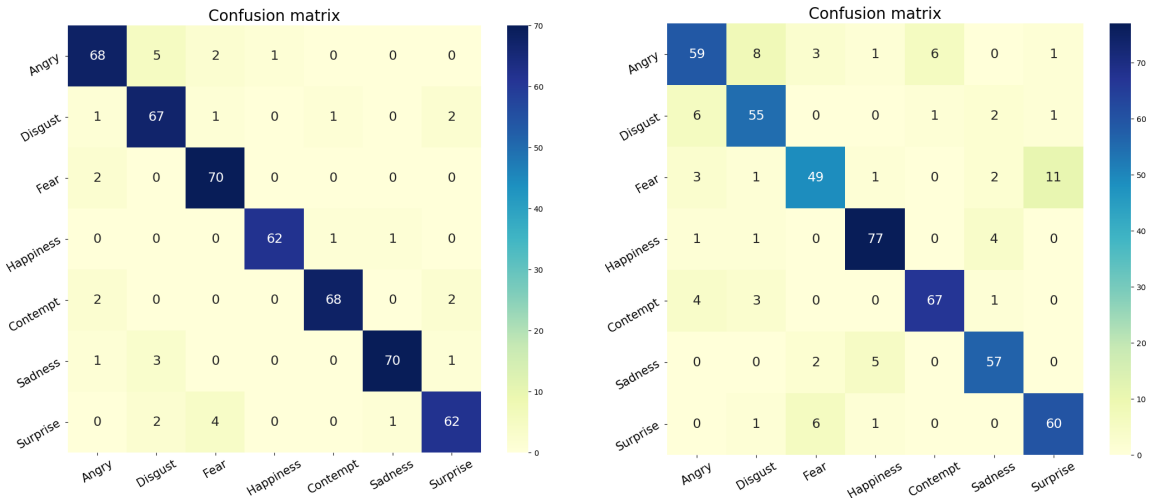
Although these two cases achieve acceptable performance, we delve into a more challenging scenario which is leave-one-user-out cross validation (LOOCV). LOOCV trains with all the collected data but one user (i.e., 19 out of 20 in this paper), and tests the user who is not seen during the training stage. This tends to have a lower bias and therefore we don't need to re-train the model when a new user is enrolled. However it is more challenging as the model has no prior of the new user. The results show 0.53 accuracy, 0.54 precision, 0.54 recall and 0.52 f1-score, which is far behind the performance of the previous two cases. We presume the performance drop is due to the uniqueness of the user's PPG signals and muscle movement from their own biological traits. Moreover, this is due to the small likelihood of a generalization model to fit a new user.

6.4 Performance Comparison Between Using PPG and Accelerometer Sensors

In addition to the importance of utilizing both sensors successfully to detect facial expressions, it is also important to provide quantitative analysis. As can be seen from Fig. 15, we plotted the average recall and precision values for seven different facial expressions with three different combinations using: 1) only the PPG signals, 2) only the accelerometer signals, and 3) both signals together. It is observed that all three cases show above 0.8 precision and 0.77 recall, with the lowest precision and recall rates for when using only the accelerometer sensor. However, as can be seen from Fig. 15b, which uses the IMU sensor alone, it is observed that the accelerometer is not enough to capture the micro-movement of facial muscles, which is aligned with our explanation from the previous section. From Fig. 15b, facial expressions that involve distinctive spontaneous body postures (i.e., anger, disgust, sadness, or surprise) show slightly better performance compared to those that do not, but overall lower than when using only PPG signals (Fig. 15a). On the other hand, as shown in Fig. 15c, using both sensing modalities shows promising results, outperforming the other two cases that use only one sensor. Thus, the results address that deploying both sensors can enhance the performance to recognize different facial expressions precisely.

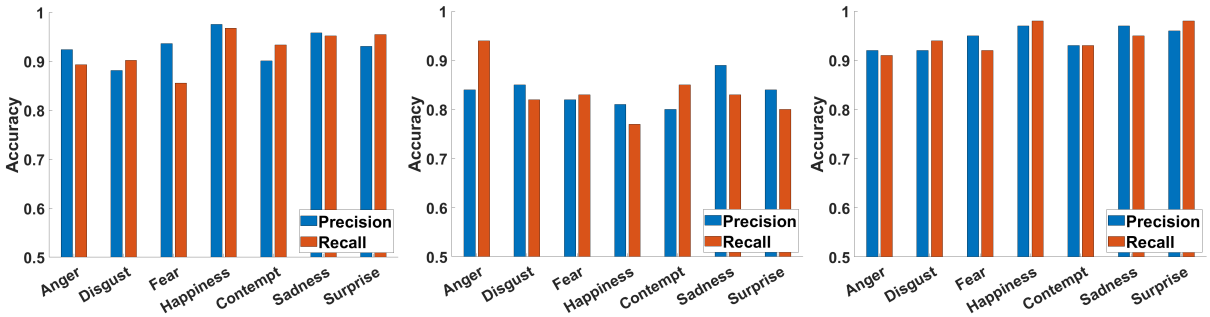
6.5 Performance Comparison Between Using Different Light Wavelengths

Normally commodity earable devices are equipped with one type of wavelength of light. Therefore, we explore whether PPGface is suitable to be adopted by the market products by evaluating the performance with using one



(a) User-dependent case (b) User-independent case

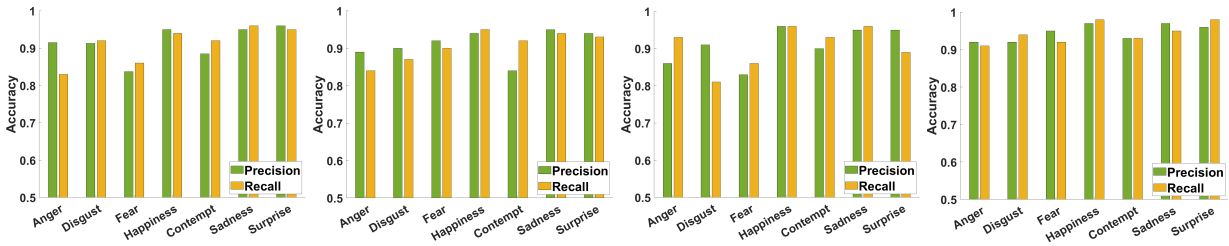
Fig. 14. Confusion matrix of seven universal facial expressions. (a) user-dependent, (b) user-independent case.



(a) PPG signals only (b) Accelerometer signals only (c) PPG + Accelerometer signals

Fig. 15. Average precision and recall values of different signal combination cases.

type of wavelength of light. Fig. 16 shows the overall result of average recall and precision values for different lights. Instead of using three lights altogether, we used one type of light with accelerometer sensors in this experiment. When using the green light only, due to its low signal-to-ratio (SNR) and better absorbability rate to the hemoglobin, it achieves similar performance to the IR light and slightly better performance than the red light. On the contrary, the red light shows higher SNR than the green, and can only penetrate up to the dermis skin layer, showing lower accuracy. The IR also shows higher SNR, but can penetrate deeper than the other two lights, allowing it to capture muscle movement more efficiently and leads to better results than the red light. Although using all three lights achieves the best performance, which is shown in Fig. 16d, employing either the green or IR light is also acceptable for integration into commodity devices as there is only a slight performance



(a) Green light only (b) Infrared light only (c) Red light only (d) Green + IR + Red lights

Fig. 16. Average precision and recall values of three different wavelengths of the light.

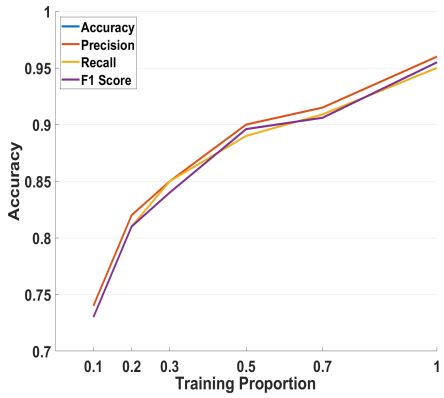


Fig. 17. Accuracy, precision, recall and F1-score of different training size ratio.

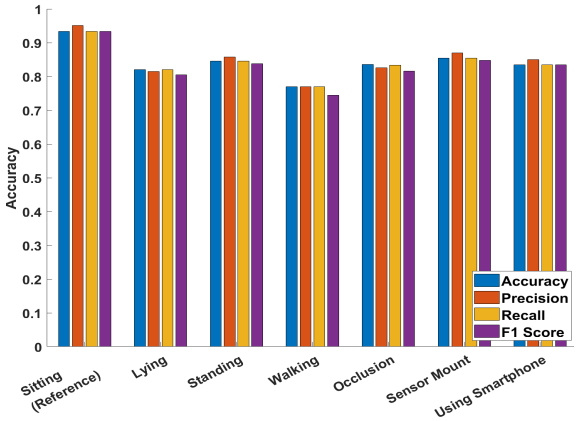


Fig. 18. Robustness Test Performance

drop. Therefore, light selection can be determined by the purpose of the system, whether the system focuses more on the performance or the weight.

6.6 Impact of Training Dataset Sizes

The size of training data can affect the user's experience as most users prefer a simpler sampling and calibration process. With this in mind, this section explores the performance with respect to the training dataset size. Fig. 17 illustrates the average accuracy for four activation phrases when the training dataset portion changes. We conducted an experiment by changing the proportions of training samples, P_{train}, from 10% to 90% where the rest of the samples are used for testing purposes. It is observed that PPGface is robust against the size of the training dataset as it can achieve an accuracy of 0.81 even when only 30% of the total data is used for training. However, the accuracy drops to 0.73 when P_{train} is 10% which was expected, because the model cannot learn with insufficient data samples.

7 SYSTEM ROBUSTNESS EVALUATION

Although PPGface shows convincing results, we designed and examined different experiments which emulated real-world usage scenarios to strengthen our hypothesis by conducting several robustness evaluations. We

recruited five participants ages between 27 and 32 to take part in the user study (three males and two females). The participants were asked to first sit on a chair comfortably and wear the PPGface device to perform certain tasks 20 times, and go through making seven facial expressions one by one.

7.1 Impact of Body Motions

Many people watch videos standing while waiting for a bus or train, or before going to bed lying on one's back. Hence, understanding how different body motions (e.g., walking or hand-using) can affect performance is important since the user's body motions could cause undesired noise. To evaluate the robustness of the proposed system, we conducted an experiment with multiple users' body motions: sitting, standing, lying, walking, and using hands freely. We included the walking scenario as some people listen to music or comedy shows without looking at their smartphone screens. Furthermore, as people aren't stagnant; constantly moving their bodies especially their hands and upper torsos, we assumed a more realistic scenario to validate the PPGface's robustness against body motions. We asked the participants to freely use their phones with both hands (or either hand) without being asked to perform any specific task. Some participants sent text message while others read an article or did web-browsing. Then, the model predicted their facial expressions using the users' previously collected data from Section 6.3. We evaluated the performance using the same evaluation metrics from previous studies. Compared with the sitting case accuracy from Fig. 18, standing achieves 0.85 accuracy, 0.86 precision, 0.85 recall, and 0.84 f1-score, lying down achieves 0.82 accuracy, 0.82 precision, 0.82 recall, and 0.81 f1-score, walking achieves 0.77 accuracy, 0.77 precision, 0.77 recall, and 0.75 f1-score, and using smartphones achieves 0.84 accuracy, 0.85 precision, 0.84 recall, and 0.84 f1-score. There is a slight performance drop for lying down, walking, and using a smartphone because it is hard to make specific facial expressions under these scenarios. For example, it is hard to sway backwards when making a "surprise" expression in a sleeping position. Walking shows lowest performance among other body motions because the IMU and PPG signals are corrupted due to the vibrations from the walking motion. Moreover, for the using smartphone case, it was observed during data collection that participants were slightly distracted and unconsciously moving (e.g., shaking legs or moving head constantly), and the leading cause for the slight performance drop. However, PPGface signals induced by the facial expression are less affected by motion artifacts because they show more distinctive patterns, which may offset those noises. These results seem much more promising than existing solutions, such as WiFi-based facial expression systems because WiFi-based systems require an extra sensor at a specific location, near the user, which could restrict the body positions or motions when collecting data.

7.2 Impact of Occlusions

Unlike PPGface, which deploys in-ear PPG signals as a means of classification, other facial expression recognition systems utilize a camera, WiFi, or physiological-based sensors. However, a critical limitation of these modalities is that they cannot accurately measure the signal in the presence of occlusions, especially when the user is wearing a mask. In particular, after the COVID-19 breakout, most people wear masks on a daily basis due to government regulations and overall health concerns surrounding the current climate. This makes the aforementioned modalities ineffective or even incapable of recognizing facial expressions. Hence, to evaluate the effectiveness of PPGface under the special circumstance with mask occlusions, we asked the participants to wear masks and collected the data when they make facial expressions (under the mask). As shown in Fig. 18, the average performance shows an accuracy of 0.84, precision of 0.833, recall of 0.83, and f1-score of 0.82. Compared with other robustness study scenarios, there is a slight performance drop when wearing masks. This is because the mask straps are worn around the outer ear and hold tension force, which can possibly affect the earphone displacement and cause ear canal deformation. This issue could be resolved with a more compact, well-designed

prototype within the ear canal, lessening the influence from outer ear stretching forces. Moreover, we can include the part of the data when the user is wearing a mask during the training stage to enhance performance.

7.3 Impact of Sensor Mount Positions

Users have personal preferences on how to wear their earable devices and don't usually wear them all day. Most use earable devices whenever they listen to music, make a phone call, or watch videos. System performance shouldn't be degraded due to the changing sensor mount positions. Therefore, we asked the participants to wear and take out the device based on their preferences before and after collecting the data. From the Fig. 18, the result shows 0.88 accuracy, 0.9 precision, 0.88 recall, and 0.87 f1-score. It is observed that performance was similar to Section 6.3, which is a reference accuracy of the PPGface. Unlike the IMU signals which are significantly affected by sensor mount positions due to the angle rotation from the gyroscope and directional movement from the accelerometer, PPG signals are not greatly affected by mount positions since they are captured inward from the ear canal caused by muscle movement and blood vessel deformation. Therefore, it is concluded that the proposed approach shows robustness to sensor mount positions.

8 USABILITY STUDY

8.1 Longer-term Case Study

Although the set of robustness studies from the previous section shows reasonable performance, data was collected from a shorter period of time repeatedly. Hence, we delved into more challenging scenarios by evaluating when a user is watching a video for a longer period of time, to show whether PPGface performs well under uncontrolled and unrestricted environments. We recruited two participants for this sub-study and asked them to watch the famous TV show "Friends," season 2 episode 14 for 10 minutes. Participants were allowed to watch a TV show using any device (e.g., TV, mobile phone, or computer) in any body position (e.g., lying on sofa, standing, or sitting on a chair) and conduct any activity (moving one's head, eating a snack, or even yawning if the participant found it boring). Description of the scene was annotated by two human subjects who didn't participate in this study. Like previous studies, the model was trained to predict the facial expressions from the collected data. Then it detects and predicts which facial expression the user had made during watching the TV show. Fig. 19 shows the time segments of the TV show with the description of the scene and the predicted facial expressions from the participants. As people make facial expressions occasionally, we only color those segments for the expressions. The empty segments indicate either a neutral expression or irrelevant body/head motions. The overall accuracy is 78% which shows that the user's predicted facial expression showed a high similarity with the labels. It is observed from the results that the fear and surprise expressions are two commonly confused cases. This is because the surprise expression is associated with similar key facial muscle movements with that of fear as discussed earlier, and the surprise expression is brief and may merge into a fear expression.

8.2 Subjective Ratings of Different Facial Expression Recognition System

Although we demonstrated PPGface's robustness by conducting several experiments, it is important to assess user experience, including usability, privacy, and user acceptance. Therefore, we recruited 11 subjects who participated in this study and conducted a user survey using 5-point Likert scales (from 1–strongly disagree to 5–strongly agree). All users who participated in this survey replied they have at least one earable device in their possessions, and 9 out of 11 subjects considered themselves very active users who wear earable devices frequently in daily life. As can be seen from Fig. 20, the consensus was that they would rather use PPGFace over other systems; WiFi, sEMG, and camera-based systems. In regards to usability, we asked *"Do you feel comfortable to wear the sensor and interact with the system?"*, and the scores were 4.54 for PPGface (SD = 0.52), 3.18 for WiFi (SD = 1.17), 1.72 for sEMG (SD = 1.10), and 3.36 for the camera (SD = 0.92). About 63% of the subjects answered negatively for the

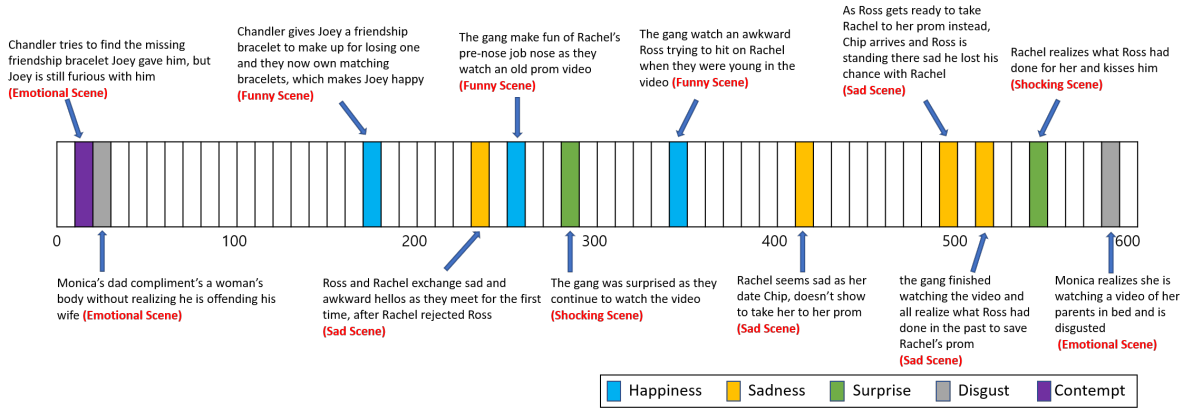


Fig. 19. Longer-term case usability study.

sEMG system, which we assume is because wearing lots of wires on the face is obtrusive. For privacy, we asked *"Do you feel safe from the privacy perspective such as personal data leakage?"*, and average scores were 4.72 for the PPGface (SD = 0.47), 3.54 for the WiFi (SD = 0.82), 3.81 for the sEMG (SD = 0.87), and 1.72 for the camera (SD = 0.9). It is observed that more than half of the subjects felt unsafe when using the camera-based system, although they were comfortable with the system itself. Last, for user acceptance, we asked *"Are you willing to use the system in daily life?"*. Average scores for this question were 4.72 for the PPG face (SD = 0.46), 2.90 for the WiFi (SD = 1.13), 1.72 for the sEMG (SD = 0.78), and 2.18 for the camera-based system (SD = 1.17). As expected, people were not interested in using the sEMG-based system due to the inconvenience of sensor wearing, although they had replied they feel safe from using it. Moreover, people were reluctant to use the camera-based system, although more than 80% answered neutral or positive regarding usability. From this survey, it can be concluded that when it comes to willingness to employ the system in daily life, people consider not only usability but also the privacy issue seriously.

9 DISCUSSIONS

9.1 Comparison with Existing Work

In Table 4, we compare the proposed PPGface with other representative facial expression recognition systems, including ExpressEar [88], Wi-Face [17], Amesaka *et al.* [2], and Hamedi *et al.* [45] from different perspectives. The reason for the larger number of expressions in ExpressEar and Amesaks's work is because ExpressEar actually focused on 32 different AUs (rather than more complicated facial expressions) and Amesaka's work involved 21 facial gestures (e.g., protrude tongue, face down, or tilt head). In addition, as shown in the On-the-fly Recognition column, excluding Wi-Face and PPGface, the rest of the studies either manually segmented the signals or didn't explicitly mention segmentation strategy, which is essential in determining the start and end points of each expression. This suggests that these systems may not be able to meet the real-time requirements due to lack of an automatic segmentation module, requiring manual annotation per facial expression. As demonstrated in the longer-term case study in Section 8, PPGface can successfully detect and recognize expressions in a real-world setting. Moreover, compared with other works, PPGface is robust towards motion artifacts and can be utilized regardless of body motions (i.e., walking or using hands), environments (i.e., indoors or outdoors) and occlusions (i.e., wearing mask). Moreover, unlike some prior works that may require the user to record themselves before and after each facial expression, PPGface can detect and segment facial expressions in an unsupervised, intuitive

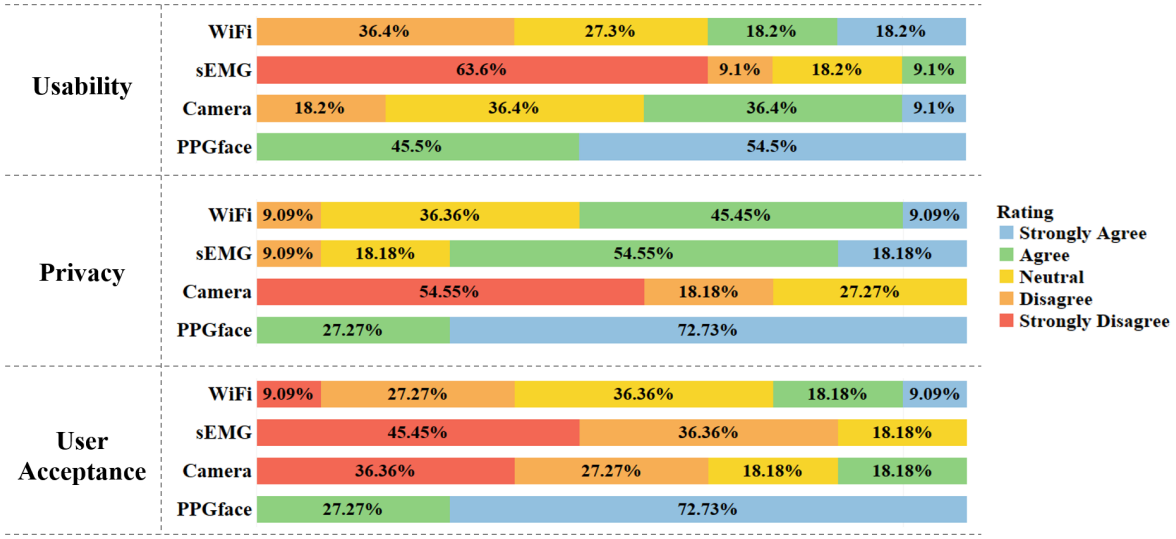


Fig. 20. User assessment using 5-point Likert scales for different systems using WiFi, sEMG, and camera-based compared with the PPGface. The x-axis represents the percentage.

Table 4. Comparison with existing facial expression recognition solutions (ACC : Accelerometer, GYR : Gyroscope).

	ExpressEar [88]	Wi-Face [17]	Amesaka [2]	Hamedi [45]	PPGface
Modality	ACC + GYR	WiFi	Acoustic	Surface EMG	PPG + ACC
Accuracy (%)	89.9	94.8	64.7	94.48	93.5
Occlusion Free	Yes	No	Yes	Yes	Yes
Num. Expressions	32 (AUs)	6	21 (Gestures)	8	7
Num. Participants	12	20	11	10	20
Num. Samples Each Participant	2,240 (32×70)	300 (6×50)	1,763 (21×84)	20 seconds (200 ms windows)	1,000 (20×50)
Environment Free	Yes	No	Yes	No	Yes
Motion Resistant	No	No	No	No	Yes
On-the-fly Recognition	No	Yes	No	No	Yes

way without needing user involvement. In addition, the experiments were conducted under different real-world settings, encompassing seven universal facial expressions which is a big plus in terms of ubiquitousness, usability, and robustness.

9.2 Limitations and Future Works

PPGface introduces a novel approach in facial expression recognition; in an unparalleled secure, obscure manner. However, some limitations need to be addressed before PPGface can be used as an independent embedded system for earable devices.

(1) **Customized Prototype Optimization.** The proposed design of PPGface applies commercial silicone type earbuds to secure a solid fixed position between the ear and the sensor. The device used in this study was not equipped with a real earphone; just the PPG and accelerometer sensors, which is not a viable wearable facial expression recognition system. For future work, we will design a customized prototype by embedding these sensors to the earable device to make the system more practicable. In addition, due to the I2C address interruption issue in this study, we were only able to utilize one sensor at a time. In reality, people wear a pair of earbuds, not just one. Hence, we plan to design a way that allows data collection to be synchronized from two sensors simultaneously, which would further improve system performance.
(2) **Practicability and Comprehensive Evaluation.** Although the COVID-19 pandemic presented difficulties in overall data collection, we were able to collect data from 20 users to validate the possibility of PPGface in this pilot study. However, a larger subject pool is needed to assess the system's performance when it comes to designing a delicately engineered product. Moreover, owing to different skin elasticity levels among participants due to ethnicity, body fat, and age, more data is needed. Hence, we plan to delve into the impact of demographics from a more diverse and larger subject pool to examine the robustness and permanence of the PPGface system.
(3) **Versatility** In this study, we only focus on the seven universal facial expressions. However, Chowdhury *et al.* [18] mentions that facial expressions can be used for other purposes. Therefore, we can develop a single system, which not only detects facial expression, but also authenticates the user based on facial expressions. For future work, we will enhance the algorithm and deep learning architecture by applying multi-task learning to improve the versatility of PPGface.

9.3 Application Scenario

As PPGface only needs an accelerometer and PPG signals, which are normally equipped in recent earable devices, PPGface is suitable to be used in a wide range of user engagement scenarios without the user's involvement. The most noteworthy part of the proposed system is that it is highly usable; anywhere, anytime without the need for extra device setup or distance restraints, to and from the sensor. This section briefly describes the several application scenarios of PPGface.

9.3.1 User Engagement. We built an earable device application that can both benefit the user and content creators alike. From the user's perspective, existing content recommendation systems generally require users to click the like or dislike buttons to collect preference data. However, PPGface is able to automatically detect the user preference through PPGface. Content can be segmented, and PPGface can construct a micro-level recommendation system from segmented parts where the user focuses on and reacts. In turn, content creators can provide better quality service by accommodating micro-level precise user preference data.

9.3.2 Accessibility. We believe that this study can be applied in accessibility scenarios for people with disabilities, such as mute persons. Because PPGface doesn't require speaking, it is able to detect the emotions and preferences of these people without having to converse. For instance, user involvement is automatic as long as the user wears PPGface, and facial expressions are aptly recorded in a simple manner. Moreover, mute persons usually communicate their feelings and opinions through sign language in a physically visible environment, but PPGface overcomes these environmental challenges significantly.

9.3.3 Applicability of User Command. We believe that this study can be applied to user commands. For example, PPGface will allow users to easily decide on checking incoming messages or declining phone calls during meetings, navigate exercise videos in between instructions, when the body is restricted, or play the next video or continue watching current content by simply making a facial expression.

10 CONCLUSION

This paper proposes PPGface, a novel facial expression recognition system, which leverages unique in-ear PPG variations aided by an accelerometer, from the user's facial expressions. We validated that PPGface can sufficiently serve as an alternative to existing facial expression recognition systems such as WiFi, camera or physiological sensor (EMG or EEG) based modalities. In particular, we successfully present dynamic signal segmentation and detection, based on drastically changing signal patterns affected by the facial muscle movements. After the signal preprocessing stage, we further extract the features and classify seven universal facial expressions by adopting multimodal ResNet. Based on a group of 20 participants from the first user study, we demonstrated that the classification performance of our proposed system can achieve an accuracy of 93.5%, a recall of 0.93, a precision of 0.95, and f1-score of 0.93. To explore the robustness and practicability of PPGface in real-world applications, we conducted several experiments under different settings in addition to a daily life longer-term usability case study and user assessment. It is believed that this work has great potential to be employed in future smart earable devices as a convenient, ubiquitous facial expression recognition solution and can provide insight into non-intrusive in-ear sensing research.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. CNS-1840790 and CNS-2050910. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Paul S. Addison. 2017. *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance.* CRC press.

[2] Takashi Amesaka, Hiroki Watanabe, and Masanori Sugimoto. 2019. Facial expression recognition using ear canal transfer function. In *Proceedings of the 23rd International Symposium on Wearable Computers*. 1–9.

[3] R. R. Anderson, J. Hu, and J. A. Parrish. 1981. Optical radiation transfer in the human skin and applications in in vivo remittance spectroscopy. In *Proceedings of the Bioengineering and the Skin*. Springer, 253–265.

[4] Jean J. M. Askenasy. 1989. Is Yawning an Arousal Defense Reflex? *The Journal of Psychology* 123, 6 (1989), 609–621.

[5] C. Bradford Barber, David P. Dobkin, and Hannu Huhdanpaa. 1996. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)* 22, 4 (1996), 469–483.

[6] Khayrul Bashar. 2018. ECG and EEG based multimodal biometrics for human identification. In *Proceedings of the 2018 IEEE International conference on systems, man, and cybernetics (SMC)*. IEEE, 4345–4350.

[7] F. Berzin and C. R. H. Fortinguerra. 1993. EMG study of the anterior, superior and posterior auricular muscles in man. *Annals of Anatomy-Anatomischer Anzeiger* 175, 2 (1993), 195–197.

[8] Shengjie Bi, Tao Wang, Nicole Tobias, Josephine Nordrum, Shang Wang, George Halvorsen, Sougata Sen, Ronald Peterson, Kofi Odame, Kelly Caine, et al. 2018. Auracle: Detecting eating episodes with an ear-mounted sensor. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–27.

[9] Maneesh Bilalpur, Seyed Mostafa Kia, Manisha Chawla, Tat-Seng Chua, and Ramanathan Subramanian. 2017. Gender and emotion recognition with implicit user signals. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. 379–387.

[10] Mehdi Boukhechba, Lihua Cai, Congyu Wu, and Laura E. Barnes. 2019. ActiPPG: using deep neural networks for activity recognition from wrist-worn photoplethysmography (PPG) sensors. *Smart Health* 14 (2019), 100082.

[11] Brian Bradke and Bradford Everman. 2020. Investigation of Photoplethysmography Behind the Ear for Pulse Oximetry in Hypoxic Conditions with a Novel Device (SPYDR). 10, 4 (2020).

[12] K. Budidha and P. A. Kyriacou. 2018. In Vivo Investigation of Ear Canal Pulse Oximetry during Hypothermia. *Journal of Clinical Monitoring and Computing* 32 (2018), 97–107.

[13] Nam Bui, Nhat Pham, Jessica Jacqueline Barnitz, Zhanan Zou, Phuc Nguyen, Hoang Truong, Taeho Kim, Nicholas Farrow, Anh Nguyen, Jianliang Xiao, Robin Deterding, Thang Dinh, and Tam Vu. 2019. *EBP: A Wearable System For Frequent and Comfortable Blood Pressure Monitoring From User's Ear.* Association for Computing Machinery, New York, NY, USA.

[14] Yetong Cao, Qian Zhang, Fan Li, Song Yang, and Yu Wang. 2020. PPGPass: Nonintrusive and Secure Mobile Two-Factor Authentication via Wearables. IEEE Press.

[15] Tuochao Chen, Yaxuan Li, Songyun Tao, Hyunchul Lim, Mose Sakashita, Ruidong Zhang, Francois Guimbretiere, and Cheng Zhang. 2021. NeckFace: Continuously Tracking Full Facial Expressions on Neck-mounted Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–31.

[16] Tuochao Chen, Benjamin Steeper, Kinan Alsheikh, Songyun Tao, François Guimbretière, and Cheng Zhang. 2020. C-Face: Continuously Reconstructing Facial Expressions by Deep Learning Contours of the Face with Ear-Mounted Miniature Cameras. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 112–125.

[17] Yanjiao Chen, Runmin Ou, Zhiyang Li, and Kaishun Wu. 2020. WiFace: facial expression recognition using Wi-Fi signals. *IEEE Transactions on Mobile Computing* (2020).

[18] Romit Roy Choudhury. 2021. Earable computing: A new area to think about. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*. 147–153.

[19] Sebastian Cotofana, Natalia Lowry, Aditya Devineni, Gianna Rosamilia, Thilo L. Schenck, Konstantin Frank, Sana A. Bautista, Jeremy B. Green, Hassan Hamade, and Robert H Gotkin. 2020. Can smiling influence the blood flow in the facial vein?—An experimental study. *Journal of Cosmetic Dermatology* 19, 2 (2020), 321–327.

[20] Sebastian Cotofana, Hanno Steinke, Alexander Schlattau, Markus Schlager, Jonathan M. Sykes, Malcolm Z. Roth, Alexander Gaggl, Riccardo E. Giunta, Robert H. Gotkin, and Thilo L. Schenck. 2017. The anatomy of the facial vein: implications for plastic, reconstructive, and aesthetic procedures. *Plastic and Reconstructive Surgery* 139, 6 (2017), 1346–1353.

[21] Antonios Danelakis, Theoharis Theoharis, and Ioannis Pratikakis. 2018. Action unit detection in 3 D facial videos with application in facial expression retrieval and recognition. *Multimedia Tools and Applications* 77, 19 (2018), 24813–24841.

[22] John Delaney. 2016. You Can Hear the Future Calling. *Communications of the ACM* (May 2016). https://cacm.acm.org/news/202492-you-can-hear-the-future-calling/

[23] Shichuan Du, Yong Tao, and Aleix M. Martinez. 2014. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences* 111, 15 (2014), E1454–E1462.

[24] Xujun Duan, Qian Dai, Qiyong Gong, and Huafu Chen. 2010. Neural mechanism of unconscious perception of surprised facial expression. *Neuroimage* 52, 1 (2010), 401–407.

[25] Paul Ekman. [n.d.]. *Paul Ekman Group*. https://www.paulekman.com/universal-emotions Last accessed: 2021-08-25.

[26] Paul Ekman. 1989. The argument and evidence about universals in facial expressions. *Handbook of Social Psychophysiology* 143 (1989), 164.

[27] Paul Ekman. 2003. Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences* 1000, 1 (2003), 205–221.

[28] Paul Ekman and Dacher Keltner. 1970. Universal facial expressions of emotion. *California Mental Health Research Digest* 8, 4 (1970), 151–158.

[29] Luke Everson, Dwaipayan Biswas, Madhuri Panwar, Dimitrios Rodopoulos, Amit Acharyya, Chris H Kim, Chris Van Hoof, Mario Konijnenburg, and Nick Van Helleputte. 2018. Biometricnet: Deep learning based biometric identification using wrist-worn PPG. In *Proceedings of the 2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 1–5.

[30] Jesse Fine, Kimberly L. Branan, Andres J. Rodriguez, Tananant Boonya-ananta, Ajmal, Jessica C. Ramella-Roman, Michael J. McShane, and Gerard L. Coté. 2021. Sources of Inaccuracy in Photoplethysmography for Continuous Cardiovascular Monitoring. *Biosensors* 11, 4 (2021), 126:1–36.

[31] E. Friesen and Paul Ekman. 1978. Facial action coding system: a technique for the measurement of facial movement. *Palo Alto* 3, 2 (1978), 5.

[32] Crystal A. Gabert-Quillen, Ellen E. Bartolini, Benjamin T. Abravanel, and Charles A. Sanislow. 2015. Ratings for emotion film clips. *Behavior Research Methods* 47, 3 (2015), 773–787.

[33] Yang Gao, Wei Wang, Vir V Phoha, Wei Sun, and Zhanpeng Jin. 2019. EarEcho: Using Ear Canal Echo for Wearable Authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.

[34] Mohammad Ghamari, Denisse Castaneda, Aibhlin Esparza, Cinna Soltanpur, and Homer Nazeran. 2018. A Review on Wearable Photoplethysmography Sensors and Their Potential Future Applications in Health Care. 4, 4 (2018), 195–202.

[35] T. Lee Gilman, Razan Shaheen, K. Maria Nylocks, Danielle Halachoff, Jessica Chapman, Jessica J. Flynn, Lindsey M. Matt, and Karin G. Coifman. 2017. A film set for the elicitation of emotion in research: A comprehensive catalog derived from four decades of investigation. *Behavior Research Methods* 49, 6 (2017), 2061–2082.

[36] Michael Goodman. 2021. *U.S. SVOD Forecast (2010 - 2026).* Technical Report. Strategy Analytics, Newton, Massachusetts USA. https://www.strategyanalytics.com/ Last accessed: 2021-08-25.

[37] Valentin Goverdovsky, David Looney, Preben Kidmose, and Danilo P. Mandic. 2016. In-Ear EEG From Viscoelastic Generic Earpieces: Robust and Unobtrusive 24/7 Monitoring. 16, 1 (2016), 271–277.

[38] Valentin Goverdovsky, Wilhelm von Rosenberg, Takashi Nakamura, David Looney, David J. Sharp, Christos Papavassiliou, Mary J. Morrell, and Danilo P. Mandic. 2017. Hearables: Multimodal physiological in-ear sensing. *Scientific reports* 7, 1 (2017), 1–10.

[39] Grand View Research. 2022. *Video Streaming Market Size, Share & Trends Analysis Report, 2022 - 2030.* Technical Report GVR-2-68038-629-5. San Francisco, CA, USA.

[40] Malcolm J. Grenness, Jon Osborn, and W. Lee Weller. 2002. Mapping ear canal movement using area-based surface matching. *The Journal of the Acoustical Society of America* 111, 2 (2002), 960–971.
[41] James J. Gross and Robert W. Levenson. 1995. Emotion elicitation using films. *Cognition & Emotion* 9, 1 (1995), 87–108.
[42] Anna Gruebler and Kenji Suzuki. 2010. Measurement of distal EMG signals using a wearable device for reading facial expressions. In *Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 4594–4597.
[43] Yu Gu, Xiang Zhang, Zhi Liu, and Fuji Ren. 2020. WiFE: WiFi and Vision based Intelligent Facial-Gesture Emotion Recognition. *arXiv preprint arXiv:2004.09889* (2020).
[44] Y.Y. Gu and Y.T. Zhang. 2003. Photoplethysmographic authentication through fuzzy logic. In *Proceedings of the IEEE EMBS Asian-Pacific Conference on Biomedical Engineering, 2003.* IEEE, 136–137.
[45] Mahyar Hamedi, Sh-Hussain Salleh, Tan T. Swee, et al. 2011. Surface electromyography-based facial expression recognition in Bi-polar configuration. *Journal of Computer Science* 7, 9 (2011), 1407.
[46] Masaki Hasegawa, Kotaro Hayashi, and Jun Miura. 2019. Fatigue Estimation using Facial Expression features and Remote-PPG Signal. In *Proceedings of the 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1–6.
[47] Dustin A. Hatfield, Eugene A. Whang, Robert A. Boyd, Duy P. Le, Yi-Fang D. Tsai, David J. Feathers, Shota Aoyagi, and Sean S. Corbin. 2020. Earphones. Apple Inc.. https://patents.justia.com/patent/20200314518 US Patent Application Number: 16/564,804.
[48] Jiesheng He and Wei Wu. 2021. Deformation of Heartbeat Pulse Waveform Caused by Sensor Binding Force. arXiv:2108.10014 [physics.med-ph]
[49] Earnest Paul Ijjina and C. Krishna Mohan. 2014. Facial expression recognition using kinect depth sensor and convolutional neural networks. In *Proceedings of the 2014 13th International Conference on Machine Learning and Applications*. IEEE, 392–396.
[50] Maxim Integrated. 2021. MAXM86161 Single-Supply Integrated Optical Module for HR and SpO2 Measurement. Data Sheet.
[51] Masood Mehmod Khan, Robert D. Ward, and Michael Ingleby. 2004. Automated classification and recognition of facial expressions using infrared thermal imaging. In *Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems, 2004.*, Vol. 1. IEEE, 202–206.
[52] Byung S. Kim and Sun K. Yoo. 2006. Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE Transactions on Biomedical Engineering* 53, 3 (2006), 566–568.
[53] Jangho Kwon, Da-Hye Kim, Wanjoo Park, and Laehyun Kim. 2016. A wearable device for emotional recognition using facial expression and physiological response. In *Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 5765–5768.
[54] Han-Wook Lee, Ju-Won Lee, Won-Geun Jung, and Gun-Ki Lee. 2007. The periodic moving average filter for removing motion artifacts from PPG signals. *International Journal of Control, Automation, and Systems* 5, 6 (2007), 701–706.
[55] Jongshill Lee, Minseong Kim, Hoon-Ki Park, and In Young Kim. 2020. Motion artifact reduction in wearable photoplethysmography based on multi-channel sensors with multiple wavelengths. *Sensors* 20, 5 (2020), 1493.
[56] Min Seop Lee, Yun Kyu Lee, Dong Sung Pae, Myo Taeg Lim, Dong Won Kim, and Tae Koo Kang. 2019. Fast Emotion Recognition Based on Single Pulse PPG Signal with Convolutional Neural Network. *Applied Sciences* 9, 16 (2019), 3355.
[57] Seungchul Lee, Chulhong Min, Alessandro Montanari, Akhil Mathur, Youngjae Chang, Junehwa Song, and Fahim Kawsar. 2019. Automatic Smile and Frown Recognition with Kinetic Earables. In *Proceedings of the 10th Augmented Human International Conference 2019*. 1–4.
[58] Min Lin, Qiang Chen, and Shuicheng Yan. 2013. Network in network. *arXiv preprint arXiv:1312.4400* (2013).
[59] Yong-Jin Liu, Minjing Yu, Guozhen Zhao, Jinjing Song, Yan Ge, and Yuanchun Shi. 2017. Real-time movie-induced discrete emotion recognition from EEG signals. *IEEE Transactions on Affective Computing* 9, 4 (2017), 550–562.
[60] Katsutoshi Masai, Kai Kunze, and Maki Sugimoto. 2020. Eye-based interaction using embedded optical sensors on an eyewear device for facial expression recognition. In *Proceedings of the Augmented Humans International Conference*. 1–10.
[61] Katsutoshi Masai, Yuta Sugiura, Masa Ogata, Kai Kunze, Masahiko Inami, and Maki Sugimoto. 2016. Facial expression recognition in daily life by embedded photo reflective sensors on smart eyewear. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. 317–326.
[62] David Matsumoto. 1992. More evidence for the universality of a contempt expression. *Motivation and Emotion* 16, 4 (1992), 363–368.
[63] Koichiro Matsuo and Jeffrey B. Palmer. 2008. Anatomy and physiology of feeding and swallowing: normal and abnormal. *Physical Mdicine and Rehabilitation Clinics of North America* 19, 4 (2008), 691–707.
[64] Johannes Michalak, Judith Mischnat, and Tobias Teismann. 2014. Sitting posture makes a difference—embodiment effects on depressive memory bias. *Clinical Psychology & Psychotherapy* 21, 6 (2014), 519–524.
[65] Jermana L. Moraes, Matheus X. Rocha, Glauber G. Vasconcelos, José E. Vasconcelos Filho, Victor Hugo C. De Albuquerque, and Auzuir R. Alexandria. 2018. Advances in photopletysmography signal analysis for biomedical applications. *Sensors* 18, 6 (2018), 1894.
[66] Shohreh Nafisi and Howard I. Maibach. 2018. Skin Penetration of nanoparticles. In *Proceedings of the Emerging Nanotechnologies in Immunology*. Elsevier, 47–88.
[67] Takashi Nakamura, Yousef D. Alqurashi, Mary J. Morrell, and Danilo P. Mandic. 2020. Hearables: Automatic Overnight Sleep Monitoring With Standardized In-Ear EEG Sensor. 67, 1 (2020), 203–212.

[68] Takashi Nakamura, Valentin Goverdovsky, and Danilo P. Mandic. 2018. In-Ear EEG Biometrics for Feasible and Readily Collectable Real-World Person Authentication. 13, 3 (2018), 648–661.

[69] Maja Pantic. 2009. Machine analysis of facial behaviour: Naturalistic and dynamic behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3505–3513.

[70] Vasileios Papapanagiotou, Christos Diou, Lingchuan Zhou, Janet van den Boer, Monica Mars, and Anastasios Delopoulos. 2016. A novel chewing detection system based on PPG, audio, and accelerometry. *IEEE Journal of Biomedical and Health Informatics* 21, 3 (2016), 607–618.

[71] Stefanie Passier, Niklas Müller, and Veit Senner. 2019. In-Ear Pulse Rate Measurement: A Valid Alternative to Heart Rate Derived from Electrocardiography? 19, 17 (2019).

[72] Andrea Pedrana, Daniele Comotti, Valerio Re, and Gianluca Traversi. 2020. Development of a Wearable In-Ear PPG System for Continuous Monitoring. 20, 23 (2020), 14482–14490.

[73] Y. A. Pinar and F. Govsa. 2006. Anatomy of the superficial temporal artery and its branches: Its importance for surgery. *Surgical and Radiologic Anatomy* 28, 3 (2006), 248–253.

[74] Phillip Qian, Edward Siahaan, Scott C. Grinker, and Jason J. LeBlanc. 2017. Earbuds with biometric sensing. https://patents.google.com/patent/US9838775B2 US Patent 9,838,775.

[75] Phillip Qian, Edward Siahaan, Erik L. Wang, Christopher J. Stringer, Matthew Dean Rohrbach, Daniel Max Strongwater, and Jason L. LeBlanc. 2015. Earbuds with compliant member. Apple Inc.. https://patents.google.com/patent/US20180063621A1/ US Patent: 10,856,068.

[76] M. Raghuram, K. Venu Madhav, E. Hari Krishna, Nagarjuna Reddy Komalla, Kosaraju Sivani, and K. Ashoka Reddy. 2012. Dual-tree complex wavelet transform for motion artifact reduction of PPG signals. In *Proceedings of the 2012 IEEE International Symposium on Medical Measurements and Applications*. IEEE, 1–4.

[77] Raj Rakshit, V Ramu Reddy, and Parijat Deshpande. 2016. Emotion detection and recognition using HRV features derived from photoplethysmogram signals. In *Proceedings of the 2nd workshop on Emotion Representations and Modelling for Companion Systems*. 1–6.

[78] K. Ashoka Reddy, Boby George, and V. Jagadeesh Kumar. 2008. Use of fourier series analysis for motion artifact reduction and data compression of photoplethysmographic signals. *IEEE Transactions on Instrumentation and Measurement* 58, 5 (2008), 1706–1711.

[79] K. Ashoka Reddy and V. Jagadeesh Kumar. 2007. Motion artifact reduction in photoplethysmographic signals using singular value decomposition. In *Proceedings of the 2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007*. IEEE, 1–4.

[80] Catherine L. Reed, Eric J. Moody, Kathryn Mgrublian, Sarah Assaad, Alexis Schey, and Daniel N. McIntosh. 2020. Body matters in emotion: Restricted body movement and posture affect expression and recognition of status-related emotions. *Frontiers in Psychology* 11 (2020), 1961.

[81] Salah Rifai, Yoshua Bengio, Aaron Courville, Pascal Vincent, and Mehdi Mirza. 2012. Disentangling factors of variation for facial expression recognition. In *Proceedings of the European Conference on Computer Vision*. Springer, 808–822.

[82] Pau Rodriguez, Guillem Cucurull, Jordi Gonzàlez, Josep M. Gonfaus, Kamal Nasrollahi, Thomas B. Moeslund, and F. Xavier Roca. 2017. Deep pain: Exploiting long short-term memory networks for facial expression classification. *IEEE transactions on cybernetics* (2017).

[83] Jocelyn Scheirer, Raul Fernandez, and Rosalind W. Picard. 1999. Expression glasses: a wearable device for facial expression recognition. In *Proceedings of the CHI'99 Extended Abstracts on Human Factors in Computing Systems*. 262–263.

[84] Jiacheng Shang and Jie Wu. 2019. A Usable Authentication System Using Wrist-worn Photoplethysmography Sensors on Smartwatches. In *Proceedings of the 2019 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 1–9.

[85] Meike K. Uhrig, Nadine Trautmann, Ulf Baumgärtner, Rolf-Detlef Treede, Florian Henrich, Wolfgang Hiller, and Susanne Marschall. 2016. Emotion elicitation: A comparison of pictures and films. *Frontiers in Psychology* 7 (2016), 180.

[86] Terry T. Um, Franz M.J. Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. 2017. Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. 216–220.

[87] Michel F. Valstar, Maja Pantic, Zara Ambadar, and Jeffrey F. Cohn. 2006. Spontaneous vs. posed facial behavior: automatic analysis of brow actions. In *Proceedings of the 8th International Conference on Multimodal Interfaces*. 162–170.

[88] Dhruv Verma, Sejal Bhalla, Dhruv Sahnan, Jainendra Shukla, and Aman Parnami. 2021. ExpressEar: Sensing Fine-Grained Facial Expressions with Earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–28.

[89] Shangfei Wang, Zhilei Liu, Siliang Lv, Yanpeng Lv, Guobing Wu, Peng Peng, Fei Chen, and Xufa Wang. 2010. A natural visible and infrared facial expression database for expression recognition and emotion inference. *IEEE Transactions on Multimedia* 12, 7 (2010), 682–691.

[90] Zi Wang, Sheng Tan, Linghan Zhang, Yili Ren, Zhi Wang, and Jie Yang. 2021. EarDynamic: An Ear Canal Deformation Based Continuous User Authentication Using In-Ear Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–27.

[91] T. P. Whetzel and S. J. Mathes. 1992. Arterial anatomy of the face: An analysis of vascular territories and perforating cutaneous vessels. *Plastic and Reconstructive Surgery* 89, 4 (1992), 591–603.

[92] Wingclips, LLC. [n.d.]. https://www.wingclips.com/ Last accessed: 2021-10-01.
[93] Wentao Xie, Qian Zhang, and Jin Zhang. 2021. Acoustic-based Upper Facial Action Recognition for Smart Eyewear. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–28.
[94] Uldis Zarins. 2017. *Anatomy of Facial Expression.* Anatomy Next Inc.
[95] Nianyin Zeng, Hong Zhang, Baoye Song, Weibo Liu, Yurong Li, and Abdullah M Dobaie. 2018. Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing* 273 (2018), 643–649.
[96] Guoying Zhao, Xiaohua Huang, Matti Taini, Stan Z. Li, and Matti PietikäInen. 2011. Facial expression recognition from near-infrared videos. *Image and Vision Computing* 29, 9 (2011), 607–619.
[97] Ke Zhao, Jia Zhao, Ming Zhang, Qian Cui, and Xiaolan Fu. 2017. Neural responses to rapid facial expressions of fear and surprise. *Frontiers in Psychology* 8 (2017), 761.
[98] Lei Zhao, Zengcai Wang, Xiaojin Wang, Yazhou Qi, Qing Liu, and Guoxin Zhang. 2016. Human fatigue expression recognition through image-based dynamic multi-information and bimodal deep learning. *Journal of Electronic Imaging* 25, 5 (2016), 053024.
[99] Tianming Zhao, Jian Liu, Yan Wang, Hongbo Liu, and Yingying Chen. 2019. Towards Low-cost Sign Language Gesture Recognition Leveraging Wearables. *IEEE Transactions on Mobile Computing* 20, 4 (2019), 1685–1701.
[100] Yali Zheng, Tracy C.H. Wong, Billy H.K. Leung, and Carmen C.Y. Poon. 2016. Unobtrusive and multimodal wearable sensing to quantify anxiety. *IEEE Sensors Journal* 16, 10 (2016), 3689–3696.

A APPENDIX

A.1 Description of the movie clips [32, 35, 41, 59, 85, 92]

No.	Emotion	Source Film & Duration	Description
1	Fear	Kill Bill II (1:10) (Quentin Tarantino 2004)	A woman is lying in a dark wooden box, only a flashlight is lighting her desperate face. Only her breath can be heard and the noise of the earth that is falling on the box in which she is trapped.
2	Fear	Elizabeth: The Golden Age (0:54) (Shekhar Kapur 2007)	Elizabeth refuses to punish the Catholics because of their beliefs, even though they pose a threat to her.
3	Fear	Monster (0:52) (Patty Jenkins 2003)	Extreme close-up of bound hands, then a shot of a woman lying face down in a car. A man is kicking her, cursing, and demanding that she screams. The woman screams and moans, contorted in pain.
4	Disgust	Joan of Arc (0:44) (Christian Duguay 1999)	Extreme close-up of an arrow sticking into an abdominal wound. Then a close-up of the pain-contorted face of a female knight, followed by another close-up of the bleeding wound. Someone is trying to pull the arrow out of the wound. In the background, the loud moaning of the injured woman can be heard.
5	Disgust	Blade (0:41) (Stephen Norrington 1998)	In a crowded room that is completely smothered in red paint, a darkly dressed man shoots at a vampire, who then crumbles with a gurgling noise.
6	Happy	Remember the Titans (3:00) (Boaz Yakin 2000)	Scene starts with coach saying "listen up, this is out time" A team wins its final football game and celebrates. End right before the music changes and the voiceover begins.
7	Happy	Wall-E (2:24) (Andrew Stanton 2008)	Starts as a white robot flies forward. Two robots dance in outer space and fail in love as people in the spaceship watch and music plays. Ends when the two robots fly away together (before shot of big spaceship).
8	Sad	Coach Carter (0:33) (Thomas Carter 2005)	Close-up of a basketball-player's sad face. Another close-up shows his disappointed team-mates. Blower's music accompanied the scene.
9	Sad	Bodyguards and Assassins (0:44) (Teddy Chan 2009)	A father finds his son is dead
10	Sad	Moulin Rouge (1:54) (Baz Luhrmann 2001)	A top shot shows a loudly weeping young man who kneels on a bed of flowers and holds his dead lover in his arms. Slow orchestral music accompanies the scene.
11	Sad	Do The Right Thing (2:00) (Spike Lee 1989)	A homeless man gets ridiculed by local teens who have no sympathy for his plight.
12	Amusement	Just Another Pandora's Box (2:00) (Jeffrey Lau 2010)	Humorous battle scenes
13	Amusement	What Women Want (0:45) (Nancy Meyers 2000)	Medium shot of a man. He is trying to get his legs into a woman's stockings. He has an anti-pimple plaster on his nose.
14	Amusement	Finding Neverland (1:53) (Marc Forster 2004)	A man in a suit and bow tie sits at a dinner table in fine company and shows the children who are present a magic-trick.
15	Amusement	Love on delivery (1:24) (Lee Lik-chi 1994)	The dog poop is patted to a bad man's face
16	Surprise	D.O.A (1:50) (Rudolph Mate 1950)	Scene opens on a man walking towards a building. He walks down a hallway and turns into the homicide division room, saying he is there to report a murder. Scene ends when he says HE was murdered and the man across the table reacts with surprise (along with the music).

17	Surprise	The Departed (0:28) (Martin Scorsese 2006))	Starts as the camera pans over a scene on a rooftop. Scene shows a confrontation between two men. Ends after DiCaprio has been suddenly shot, on the image of Matt Damon standing in an elevator looking shocked.
18	Surprise	Capricorn (0:49) (Peter Hyams 1978)	A man is sitting on a bed in an apartment. Begin recording at the first frame after the camera switches from a close-up of the man's face to a shot from down the hall. Men have just burst through the door.
19	Surprise	Sea of Love (0:09) (Harold Becker 1989)	A man has gotten out of an elevator and begun walking down the hall toward an exit door. He turns his back completely to the wall and is looking toward the left.
20	Anger	Gentleman S Agreement (2:40) (Elia Kazan 1947)	A man drives up to and enters a hotel. He tries to get a hotel room, but cannot because he is Jewish. Scene ends when he walks out while people start at him.
21	Anger	The Apostle (1:10) (Robert Duvall 1998)	After finding out his wife was cheating on him, Sonny cries out to the Lord in anger and frustration.
22	Contempt	A Monster Calls (0:30) (Juan Antonio Bayona 2017)	Conor gets threatened in class and physically intimated after class by school bullies.
23	Anger	A Question Of Faith (0:39) (Kevin Otto 2017)	After the tragic death of his son, David's father tells him that he is not in the right place to lead a church.
24	Anger	Crash (3:00) (Paul Haggis 2004)	Scene starts in a diner with a man talking on the phone
25	Contempt	How The Grinch Stole Christmas (0:44) (Ron Howard 2000)	After making a special ornament for his love, the young Grinch ends up getting ridiculed by his classmates, which sparks a life full of hate.
26	Contempt	Mr. Peabody And Sherman (0:44) (Rob Minkoff 2014)	Sherman gets teased by a bully at school for having a dog for a Dad.