W!NCE: Unobtrusive Sensing of Upper Facial Action Units with EOG-based Eyewear

SOHA ROSTAMINIA, University of Massachusetts Amherst, USA
ALEXANDER LAMSON, University of Massachusetts Amherst, USA
SUBHRANSU MAJI, University of Massachusetts Amherst, USA
TAUHIDUR RAHMAN, University of Massachusetts Amherst, USA
DEEPAK GANESAN, University of Massachusetts Amherst, USA

The ability to unobtrusively and continuously monitor one's facial expressions has implications for a variety of application domains ranging from affective computing to health-care and the entertainment industry. The standard Facial Action Coding System (FACS) along with camera based methods have been shown to provide objective indicators of facial expressions; however, these approaches can also be fairly limited for mobile applications due to privacy concerns and awkward positioning of the camera. To bridge this gap, W!NCE re-purposes a commercially available Electrooculography-based eyeglass (J!NS MEME) for continuously and unobtrusively sensing of upper facial action units with high fidelity. W!NCE detects facial gestures using a two-stage processing pipeline involving motion artifact removal and facial action detection. We validate our system's applicability through extensive evaluation on data from 17 users under stationary and ambulatory settings, a pilot study for continuous pain monitoring and several performance benchmarks. Our results are very encouraging, showing that we can detect five distinct facial action units with a mean F1 score of 0.88 in stationary and 0.82 in ambulatory settings, and that we can accurately detect facial gestures that due to pain.

CCS Concepts: • **Human-centered computing → Mobile devices**; *User studies*; • **Computing methodologies → Neural networks**; • **Applied computing → Consumer health**;

Additional Key Words and Phrases: FACS, Facial actions, EOG, JINS MEME, Pain monitoring, Motion artifact, Eyeglasses, Wearable

ACM Reference Format:
Soha Rostaminia, Alexander Lamson, Subhransu Maji, Tauhidur Rahman, and Deepak Ganesan. 2019. W!NCE: Unobtrusive Sensing of Upper Facial Action Units with EOG-based Eyewear. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1, Article 23 (March 2019), 26 pages. https://doi.org/10.1145/3314410

1 INTRODUCTION

One of the crucial missing elements in wearable sensing is the ability to measure facial expressions in natural settings. While facial expressions can easily be monitored with cameras pointing at the face, this is impractical in ambulatory settings and raises privacy issues in personal spaces. But alternate solutions have been hard to

Authors' addresses: Soha Rostaminia, University of Massachusetts Amherst, Amherst, MA, 01002, USA, srostaminia@cs.umass.edu; Alexander Lamson, University of Massachusetts Amherst, Amherst, MA, 01002, USA, alamson@umass.edu; Subhransu Maji, University of Massachusetts Amherst, Amherst, MA, 01002, USA, smaji@cs.umass.edu; Tauhidur Rahman, University of Massachusetts Amherst, Amherst, MA, 01002, USA, trahman@cs.umass.edu; Deepak Ganesan, University of Massachusetts Amherst, Amherst, MA, 01002, USA, dganesan@cs.umass.edu.

2474-9567/2019/3-ART23 $15.00
https://doi.org/10.1145/3314410

find — while wearable EEG [12, 34], electrodermal activity [16, 38], and respiration sensors [32] have all explored related topics like emotion detection, these provide relatively crude measures compared to the richness of facial expression sensing.

The ability to monitor facial expressions has a variety of applications. In the healthcare context, there has been substantial interest in developing wearable solutions to objectively measure pain [35]. While pain measurements have been performed primarily using subjective measurements like the visual analog scale (VAS), self-reports are unreliable and depend on the user's context and behavior. Another major area of interest is emotion sensing. For example, entertainment companies use biosensors on participants to figure out what types of videos and sounds affect people's emotion and makes them more engaged and aroused [11, 39]. Facial expressions are also important for AR/VR systems that need better ways of sensing user attention, intent, and context [22]. The ability to sense facial expressions can enable a more rich sensing interface to make VR/AR systems more natural and immersive for the user, particularly in interactive situations.

A intriguing new alternative that has emerged is wearable electrooculography (EOG). EOG electrodes have long been known to capture electrical potentials that exist between the cornea and ocular fundus (also referred to as the cornea-retinal potential). This has been widely used by practitioners — EOG electrodes are typically placed under, over and to the side of the eye as well as on the forehead, and the signal analyzed for tracking eye and eyelid movements. While such a setup is clearly too intrusive for normal use, recently available EOG-enabled eyeglasses use these electrodes in a much more unobtrusive manner by hiding three electrodes on the nose bridge [2]. The eyeglass is virtually identical to a normal pair of glasses except for the additional electronics and battery, and most importantly requires no cameras or electronics in the visual path of the user.

The availability of EOG electrodes on wearable glasses opens up a new avenue for sensing facial actions. In this work, we re-think what is considered signal and what is noise for EOG-based sensing — unlike traditional uses of EOG where the signal is eye movement and noise is all other activity, the information we need is contained in the noise. Our intuition is that the "noise" in the wearable EOG signal may carry useful information about facial expressions that are typically discarded in EOG processing pipelines. If we can separate the various sources of noise, we can obtain valuable contextual information about facial expressions without requiring external cameras.

But the design of the eyeglass for comfortable daily use raises two challenges. First, the design for comfort and wearablility means that the electrodes are lightly placed on the nose bridge. This results in impedance changes during head movement actions and large body movements. These changes add unwanted motion artifacts that seep into the EOG signal and need to be removed if we want to capture facial expressions. Second, unlike laboratory grade EOG which places five electrodes at several positions on the forehead and around the eyes to have sufficient spatial dimensions, wearable EOG at the nose-bridge provides a very low-dimensional signal from one location on the face.

This paper is an exploration of our hypothesis that wearable EOG can provide sufficiently good measures of facial expressions despite motion artifacts and a low dimensional signal. Our work overcomes these challenges in two ways. First, we design a motion artifact removal pipeline that leverages accelerometer and gyroscope information to remove the artifacts caused by head movement and large body movement, thereby allowing us to zoom into the facial expression signal. Second, we overcome the limited dimensionality by leveraging unique differences between facial features in terms of signal amplitude, direction, and temporal patterns.

In summary, our main contributions are:

(1) We offer a detailed look at motion artifacts in wearable EOG and design a neural network based motion artifact removal pipeline that leverages IMU signals to effectively remove noise across multiple EOG channels and many different head movement patterns.
(2) We design a novel upper face engaged action recognition system with commercially available comfortable daily eyewear device, J!NS MEME [2].

Table 1. Comparing *W!NCE* with other wearable face-engaged activity tracking systems (both facial expressions and eye expressions).

Work	Sensor	Target Classes	Evaluation Scenarios	Intrusiveness	Participants
Hickson *et al.* [22]	camera	brow lower, upper lid raise, cheek raise, eyes closed, left & right brow raise, left & right wink, neutral	stationary	× VR headset	23
Rantanen *et al.* [43]	proximity sensor (capacitive sensor)	brow raise, brow lower, mouth corners raise and lower	stationary	× on-face sensors	10
Hamedi *et al.* [20]	EMG	neutral, right & left cheek raise, brow lower, brow raise, clenching molar teeth, mouth open	stationary	× on-forehead sensors	10
Gruebler *et al.* [19]	EMG	neutral, cheek raise, brow lower, bite	stationary	× on-temple sensors	10
Bulling *et al.* [9]	EOG	copying a text, reading, writing, watching a video, browsing the Web	stationary	× on-face sensors	8
Bulling *et al.* [8]	EOG	reading	stationary walking	× goggle	8
Ishimaru *et al.* [23]	EOG IMU	typing, reading, eating, talking	stationary	✓ J!NS MEME eyeglass	2
W!NCE	EOG IMU	neutral, cheek raise, brow lower, brow raise, nose wrinkler, blink	stationary walking head-movement	✓ J!NS MEME eyeglass	17

(3) We evaluate W!NCE in both stationary and ambulatory settings from 17 users, and show via a systematic and in-depth exploration of signal characteristics that wearable EOG can be very effective in detecting upper facial actions.
(4) We demonstrate the utility of W!NCE with a case study for continuous pain monitoring. W!NCE accurately detects the instances when pain is induced as well as the action units involved in the individual's reaction to pain.

2 RELATED WORK

Our core contribution in this work is the development and evaluation of a robust, unobtrusive, privacy-sensitive, on-body upper face action unit tracking system with commercially available EOG-based eyewear. We look at related work on face-engaged activity recognition and motion artifact removal in this section.

2.1 Face-engaged Activity Recognition

Face-engaged activity recognition is an area that has seen substantial amount of work, primarily with cameras as the sensing modality but also with other sensors. We analyze related work with respect to three questions: a) how comfortable and unobtrusive are different existing facial action unit tracking techniques?, b) how applicable are these techniques to mobile scenarios?, and c) how extensively have these methods been evaluated in realistic natural settings? Table 1 summarizes the related work.

Camera-based approaches. There is a vast body of literature on using computer vision based approaches to detect facial actions and expressions (surveyed in [6, 15, 37, 48]). While this is a natural fit when the user is interacting with a computing system, it is poor fit for ambulatory settings. The difficulty is both in how to mount a camera such that it can observe the face, and in the privacy concerns raised by using cameras. As a result, this is a poor fit for applications like pain sensing that occur in ambulatory settings rather than during computer use.

A few related efforts on facial expression sensing do explore the possibilities with cameras facing the eye. [22] classifies upper facial action units by leveraging an eye facing camera embedded into a VR headset and [27] uses a custom eyeglass frame with cameras turned inwards. One technical challenge that vision-based facial action sensing systems face is the confined field-of-view (FOV) that is typically restricted to the eye. This limits the system to only detect activities which affect the shape of the eye and/or the eye movement and closure patterns. Other than this, there are several issues including intrusiveness and lack of wearability, as well as high power consumption for continuously using a camera at high frame rates.

Proximity sensor. Recent work has also used proximity sensors for face-engaged activity detection. The idea is to continuously track skin to sensor distance and thus capture the skin deformation caused by facial activities. In [52] and [30], the authors embed photo reflective based proximity sensors into a virtual reality headset and an eyeglass frame respectively in order to classify different facial expressions such as sad, happy, and surprised. Apart from the fact that such sensors are highly sensitive to the ambient lighting, they too need to be pointing at the face and have restricted field of view at close proximity, hence can only detect the activities that cause skin deformation in the sensed area.

Electromyography (EMG). Electromyography (EMG) is a technique for evaluating and recording the electrical activity produced by the muscles. There are several studies that classify different face-engaged activities by leveraging EMG sensors. Scheirer et al. [49] design a wearable system with EMG sensors embedded in the forehead pad attached to an eyeglass to classify the confusion and interest facial expressions. Hamedi et al. [20] propose a similar system capable of classifying different facial action units such as eyebrows lower and eyebrows raiser. Gruebler et al. [19] also classify a few types of facial expressions by leveraging on-temple EMG sensors. Again, the practical challenge is that EMG electrodes were attached to various parts of the face which is intrusive and unsuitable for daily use.

Electrooculography (EOG). EOG-based techniques have been widely used in a variety of contexts. Bulling et al. have shown EOG metrics to be useful for recognizing different types of activities such as reading and writing. since each activity has its own unique eye movement pattern [8, 9]. Ishimaru et al. [23] also classify typing, reading, talking, and eating activities by leveraging the eye movement pattern and head motion features extracted from the J!NS MEME eyewear. The main difference between our approach and this body of work is that while existing work tries to separate different eye movement patterns, we separate out face-engaged activities from eye movement patterns and use this information for classification. This opens up new applications that can leverage facial expressions for classifying activities.

2.2 Motion Artifact Removal from Physiological Signals

Motion artifact removal is a persistent issue in physiological signal analysis. While there has been extensive work in motion artifact removal from physiological signals such as the Photoplethysmogram (PPG) [28, 42, 54], the Electroencephalogram (EEG) [7, 45], and Electrocardiogram (ECG) [55], the effect of body motion on EOG signals has not been studied as extensively since wearable EOG devices are relatively recent.

Filtering methods. Motion artifact removal approaches can be broadly categorized into two categories. The first category is methods that remove motion artifact from quasi-periodic signals such as ECG and PPG by using frequency-domain filters [54]. These are less relevant to facial expression detection which is a non-periodic signal. The second category is time-domain noise removal methods like adaptive filters that do not require a periodic signal [8, 14]. Many adaptive filtering methods have been explored for motion artifact removal from physiological signals including Recursive Least Squares (RLS), Least Mean Squares (LMS), Normalized Least Mean Squares (NLMS) and others. In addition to classical filters, nonlinear function approximators such as Neural Networks have also been used as adaptive filters for motion artifact removal [18, 40, 56]. In this work, we design and

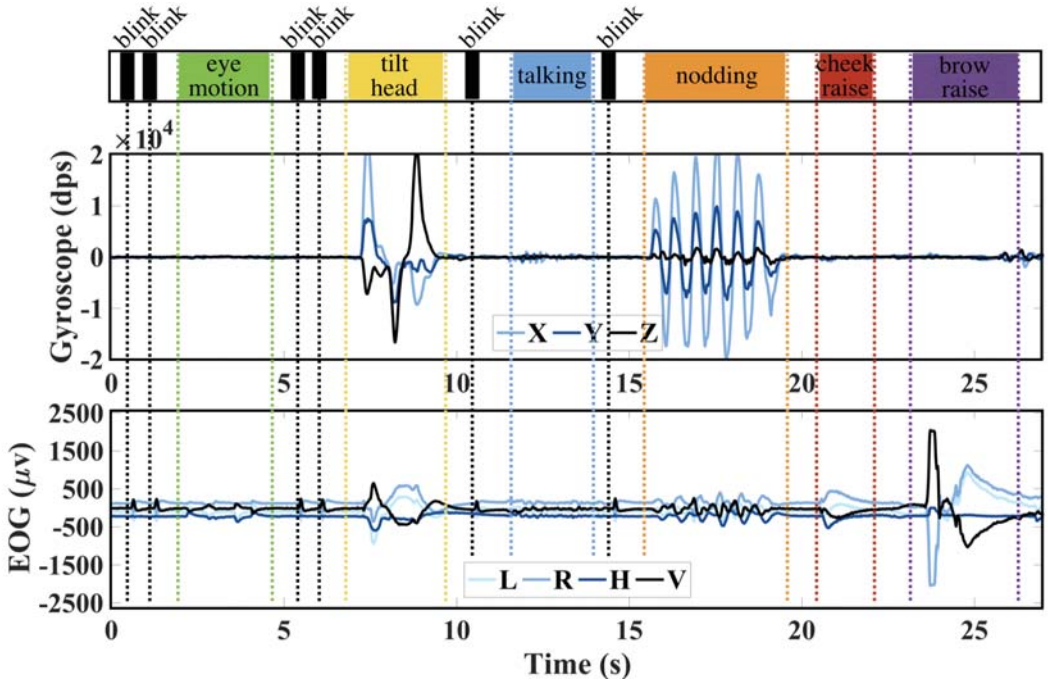


Fig. 1. The raw EOG and IMU signals measured in J!NS MEME corresponding to various eye-, upper face- and lower face-actions.

compare both an NLMS-based adaptive filter and an neural network-based filter for removing motion artifacts from the EOG signal.

Sensor choice. To effectively remove motion artifacts, we need a sensor that can accurately estimate the motions that cause the artifact in the biological signal of interest. The most common sensor for removing motion artifacts from physiological signals is the IMU, and more specifically, the accelerometer [8, 10, 44] but other specialized sensors such as piezo and stretch sensors have also been used [7, 21]. In our work, we focus on using the IMU sensor embedded in the J!NS MEME glasses to predict and remove motion artifacts.

3 CONTRASTING EOG VS IMU FOR MONITORING FACIAL ACTIONS

In this section, we provide an intuitive understanding of the signal characteristics from the eyeglass and contrast the EOG and IMU sensors in terms of their ability to capture key facial actions.

3.1 What does the EOG Measure about Facial Actions?

One question that might puzzle a reader is why EOG should carry information about facial actions. Although the traditional use of EOG is for measuring eye movement characteristics by sensing the changes in electric potential due to eye dipole movement, a rich variety of signals naturally seep through the three dry electrodes on the nose bridge. The most relevant signal to us is the generation of electric potential due to some external non-eye related muscle activation in the nearby facial regions, also referred to as electromyogram (EMG). Indeed, from the perspective of observing muscle activity on the face, the nose bridge is a good position since it is central and symmetric and capture EMG information from both sides of the face.

Fig. 2. J!NS MEME platform. It employs 3-point stainless steel EOG electrodes on the nose bridge and IMU, battery, and Bluetooth units on the eyeglasses temples.

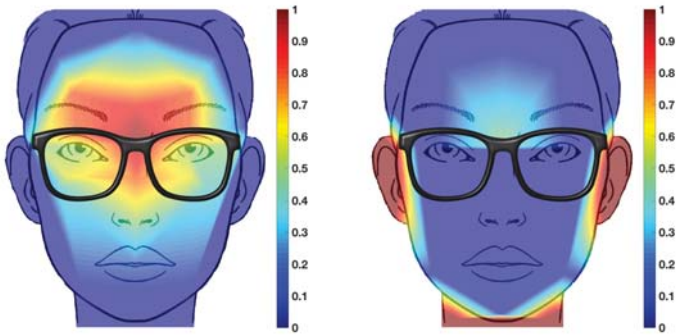
We collect our data with the J!NS MEME [2] eyewear, which is a light-weight (approx. 36g) and low-power (runs up to 16 hours) system. It employs three electrodes to sense eye movement through electrooculography (EOG): one electrode on the bridge of the nose and one on each side of nose. Figure 2 shows the external appearance of J!NS MEME. The signals it provides are the vertical difference (V), horizontal difference (H) and the raw left (L) and right (R) electrode readings. J!NS MEME also provides 3-axis accelerometer and 3-axis gyroscope signals.

Figure 1 illustrates an example – we show the raw EOG and gyroscope signal measured in J!NS MEME corresponding to a wide variety of eye-, upper face-, and lower face-engaged actions or activities. (While the IMU also contains an accelerometer, we only show the gyroscope signal for brevity.) The data was collected from a participant who remained in a seated position and performed a sequence of facial gestures in a natural manner. We can draw several interesting observations from this figure:

Upper face actions. We see that the upper face engaged actions including *cheek raise* and *brow raise* have unique EOG signatures. For example, cheek raise action leads to a sharp positive peak of the V-axis followed by a slower negative peak. The R-axis of the EOG shows a slower peak right after the sharp peak of the V-axis. The brow raise action shows a slightly different pattern and contains a much higher RMS energy compared to the cheek raise action. This is because the brow raise action activates muscles that are located near the electrodes resulting in the signal being captured with higher energy. Also, the brow raise action deforms the skin under the electrode on the nose bridge which results in a characteristic noise-like pattern due to skin-electrode friction.

Lower face actions. The lower face actions can also be captured by the nose bridge EOG, although only with much lower RMS energy. As the lower face actions happen due to the activation of distant muscles (located in the lower part of the face), the instantaneous electrical potential measured by the nose bridge EOG gets damped by the distance and is much lower than that of upper facial or eye-related actions. Because of the low energy of the lower face motion, it can easily be buried by other random noise present in the EOG signal.

Eye motions. The EOG signal also contains information about eye movements such as blinks and saccades. We see that *blinks*, a common eye activity, gives rise to a sharp peak in all the 4-dimensional EOG signals (except for the horizontal difference signal, since blink is a symmetrical action and has the same effect on both left and right EOG signals). The strongest signal can be seen in the vertical axis (V) of the EOG signal due to vertical nature of blink activity. The EOG signature corresponding to saccadic eye motion is, on the other hand, quite different from blinks. Unlike the sharp peak in the vertical axis (V) for blinks, we can observe activation of the horizontal, left and right axis (H, L, and R) for saccadic eye motion since it involves larger horizontal movement. The temporal spans of blinks and saccades also differ — blinks are temporally shorter than other eye movements and facial actions.



(a) EOG signal (b) IMU signal

Fig. 3. Heatmap of J!NS MEME's signal strength induced by facial muscles activities. a) The activities of muscles around the EOG sensors such as eyebrows would mostly affect the EOG readings. b) The jaw and neck movement would largely affect the IMU readings.

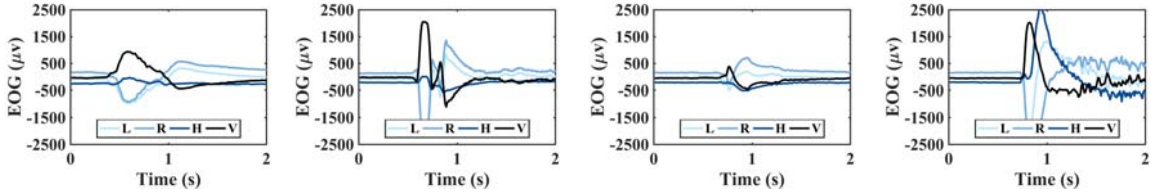
3.2 Contrasting EOG vs IMU for Facial Action Detection

The IMU is the single most commonly used sensor in wearable devices and therefore is often integrated into a wide range of smart wearable devices (including the J!INS eyeglass). Therefore, one natural question is what aspects of facial actions can be detected via an eyeglass-mounted IMU versus the EOG electrodes on the nose bridge. The interaction between IMU and EOG signals is more involved than we describe in this section since IMU also induces noise in EOG, but we defer the discussion of noise removal to §5.1. Our goal in this section is simply to contrast these two sensor modalities in terms of their ability to measure facial gestures under stationary conditions.

As can be seen in Figure 1, information about head movements like nodding and turning can be easily found in inertial sensor (i.e., gyroscope). Smaller head and mouth movement during talking can also be captured, but with lower energy. Prior work has suggested that IMUs inside the ear or behind the ear (attached to the skin) should contain such information as well [5] but the signal in our cases is weaker since an eyeglass only rests on the ear as opposed to being firmly attached.

The key difference between IMU and EOG is in their ability to detect upper facial versus lower facial gestures. In general, we find that the IMU has minimal information about upper facial actions whereas the EOG electrodes capture upper facial actions much more clearly. On the other hand, the IMU appears better than EOG in terms of measuring lower facial actions. We find that actions such as jaw movements exert large mechanical force on the temples of the eyeglass which can be sensed by the IMU. As a result, we find that actions like laughing and chewing that involve substantial jaw movement are prominently visible in the IMU signal but less visible in the EOG signal since the electrodes are relatively far from the mouth region.

Overall, the "facial coverage region" of EOG versus IMU looks as shown in Figure 3. In order to generate this figure, we asked a user to perform different facial action units that engage different facial muscles. Examples of these actions were lowering and raising eye brows, smiling, chewing, head movement, and etc. We calculate the signal power for the resultant EOG and IMU signals of each action and associate the value to the position of the engaged muscles. The normalized signal power values for all engaged muscles are used to generate a signal strength heatmap over all the face using linear interpolation. Given the fact that we are not stimulating a specific muscle and rather a group of muscles that are engaged in each action, the graph merely represents a



(a) Brow lower (b) Brow raiser (c) Cheek raiser (d) Nose wrinkler

Fig. 4. EOG signal corresponding to instances of brow lower (AU04), brow raiser (AU01), cheek raiser (AU06), and nose wrinkler (AU09) actions.

rough estimate of "facial coverage region" of EOG versus IMU. As can be seen in Figure 3, the muscles that are around the electrodes placed on the nose bridge of the glasses have higher signal strength whereas signal strength reduces as we go further from the electrodes. This is because of two reasons. The first is a bit more obvious — the EMG signal induced by the closer muscles are stronger at the EOG electrodes and therefore have higher signal strength. The second reason is more subtle — muscle movements in the proximity of the nose bridge, such as eyebrows, cause small changes in the electrode's positions thereby changing impedance and affecting the signal. In contrast, the IMU detects movements around the ear since that is where it is mounted. Thus, its coverage region is concentrated around large muscles such as the jaw that move the eyeglass position behind the ear. (Note that we have not discussed the effect of neck movements such as nodding or large bodily movements such as walking. These will also affect the signal but we defer this discussion to the next section on motion artifacts.)

4 CHALLENGES IN DETECTING UPPER FACIAL ACTIONS FOR PAIN MONITORING

In this paper, we focus on measuring signals corresponding to pain that is difficult to measure with other wearable modalities like IMU. Therefore, we restrict our attention to the detection of upper facial gestures that have been shown to be important for pain monitoring. Specifically, our focus is on classifying five different upper facial action units — raising eyebrows (AU01), lowering eyebrows (AU04), raising cheek (AU06), and nose wrinkler (AU09) which are known to be the most important facial actions for pain sensing. In addition to these, we also show that we can detect blinks (AU45) which provides useful information regarding fatigue, attention and even dopaminergic levels [25, 46, 47, 51]. Prior work has focused on detecting these changes using a camera but our goal is to detect these using EOG electrodes on an unobtrusive wearable eyeglass.

We begin by showing how the facial gestures look in the absence of any other noise. Figure 4 shows an instance of each of these four action units as observed by the EOG on the glasses. We can see that there are clear differences in the signal observed in the different cases in terms of the specific interleaving of H and V electrodes, as well as the overall amplitude, signal crossings, and temporal duration.

But the signal can change dramatically during typical head movement (e.g. nodding) or body movement (e.g. walking). In fact, these motion artifacts can be substantial and easily overwhelm the signal that we seek to detect. We describe why such motion artifacts occur in our EOG eyeglass and why they are difficult to remove.

4.1 Why do EOG Motion Artifacts Occur?

The J!NS MEME employs stainless steel electrodes that are in the category of stiff material dry electrodes [33]. Such electrodes are very popular in wearable computing because of their availability, price, and good electrical performance. Unlike gel-based electrodes they do not need to be replaced after several hours, do not cause any skin irritation or redness when used for prolonged monitoring, and they are unobtrusive. The main disadvantage of dry electrodes, however, is their susceptibility to motion artifacts due to the absence of conductive gel.

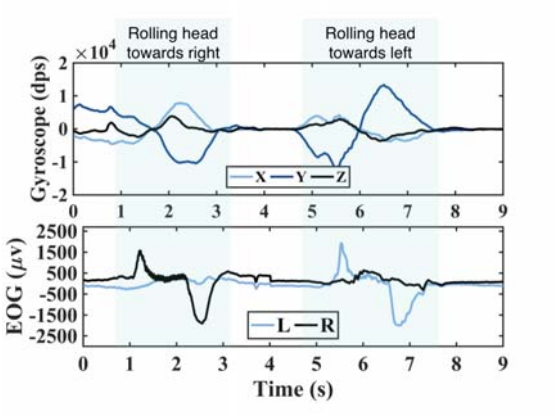


Fig. 5. EOG and IMU signals while the user tilts her head toward right and then towards left. As the user tilts her head toward right, the voltage of the right EOG signal increases and vice versa.

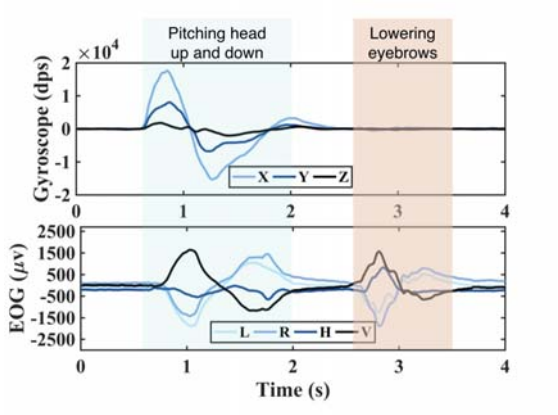


Fig. 6. EOG and IMU signals while the user nods and subsequently lowers her eyebrows. These actions have almost the same pattern on the EOG signal while the IMU signal differentiates them.

Figure 5 shows an example of head movement artifacts on J!NS MEME EOG signal. In this example, the participant tilts her head towards right (during the time interval $t = [1, 3]$) and left (during the time interval $t = [5, 7]$. As she tilts her head toward right ($t = [1, 2]$), the contact of skin and right electrode becomes loose while the other two electrodes remain connected. Therefore, as it can be seen in the figure, the electric potential between the ground electrode and right electrode increases. On the way back to the straight pose ($t = [2, 3]$), the initial head acceleration causes pressure on the contact point of skin and the right electrode that results in a sudden drop on the corresponding EOG signal due to the decrease in the connection impedance. During the time period $t = [5, 7]$, the same phenomena can be seen for when the user tilts her head toward left.

Thus, motions of the head induces force between the dry electrode and the skin, which in turn causes the skin to be compressed. This changes the coupling potential and thus the electrical potential measured by EOG also changes. This has been observed in other studies; for example, Taji et. al. [53] explored the effect of pressure on skin-electrode impedance and showed that the impedance decreases in the presence of the pressure.

In the case of more extreme head movements such as a sudden jerk, one can also observe complete loss of skin-electrode contact which results in random behavior of the EOG signal. However, in practice, such sudden movement is rare and most of the motion artifacts are due to changes in pressure.

4.2 Why is it Challenging to Remove Motion Artifacts?

While motion artifacts occur in virtually every physiological signal including PPG from wearable devices such as the Fitbit, the problem of EOG-based facial action detection is vastly different. First, facial gestures are not periodic signals unlike ECG/PPG/EEG, hence signal processing techniques that assume that the signal of interest is located in specific frequency bands do not apply. Second, head movements and facial action units occur over roughly the same timespan and have similar frequency content, so it is not possible to separate these signals using frequency-domain filtering. Third, the energy and amplitude of motion artifacts is quite high and often have similar (or higher) magnitude compared to the EOG signal due to the facial action. Taken together, this means that we cannot use frequency-domain separation techniques and cannot assume that the motion-induced noise is small compared to the desired signal.

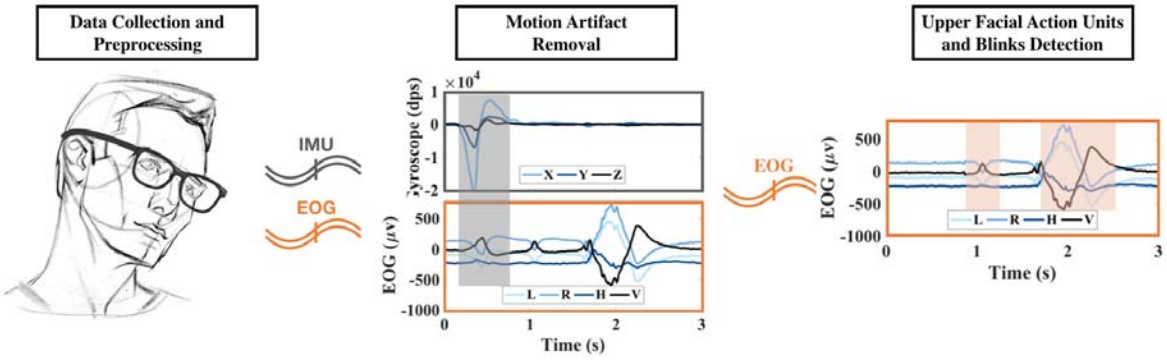


Fig. 7. Pipeline of W!NCE system design. After data collection and preprocessing, our proposed method has two main computational stages — the first stage is motion artifact removal by leveraging IMU information, and the second stage is facial action units detection.

Figure 6 demonstrates a situation where the user moves her head ($t = [1, 2]$) and then frowns ($t = [2.5, 3.5]$). The EOG signal caused by these two actions not only have a comparable amplitude and frequency, but also look similar and could be misinterpreted as each other. Therefore, filtering such motion artifacts is integral to robust detection and classification of facial action units.

5 W!NCE PROCESSING PIPELINE

Figure 7 shows an overview of the W!NCE processing pipeline. Our proposed method has two main computational stages — the first stage is motion artifact removal by leveraging IMU information, and the second stage is a classification stage where a Convolutional Neural Network (CNN) is used to detect facial action units. We now describe these two stages in detail.

5.1 Stage 1: Motion Artifact Removal

As we mentioned in §4, one of the most important challenges in our system and essentially any system working with dry electrodes is filtering out motion artifacts. This is a particularly significant concern in a loosely (and comfortably) worn wearable device such as an eyeglass where, unlike clinical settings, the electrodes are not fixed to the skin via adhesives.

Our approach is to learn a model that uses the readings of the IMU sensor embedded in the temples of the eyeglass to estimate the corresponding motion artifact on the EOG signals. Let $S(n)$ denote the raw EOG signal at time n and $M(n)$ be the estimated motion artifact on the EOG signal. Then with the assumption that the motion artifact-caused noise is additive [7, 50], we have

$$E(n) = S(n) - M(n) \tag{1}$$

where E(n) is the error signal (i.e. an estimate of the desired original EOG signal without motion artifact). The purpose is then to maximize EOG's signal-to-noise ratio (SNR) and minimize the power of motion artifact on the EOG signal by means of minimizing $E(n)$.

We model the effects of the motion artifact M(n) as a time-varying function of the IMU signal (3-axis Accel and 3-axis Gyro)

$$M(n) = h(n, \alpha, \omega) \tag{2}$$

where n denotes the time index, α is the linear acceleration and ω is the rotational angular velocity.

For learning the IMU to noise mapping function (i.e. h(n,α,ω)), we look at two different methods. The first method uses a combination linear adaptive Normalized Least Mean Squares (NLMS) filters, and the second uses a non-linear Neural Network model.

While there is extensive literature on motion artifact removal from various physiological signals like ECG and PPG [28, 44, 54], we are aware of little work on understanding the tradeoffs in the context of wearable EOG since such devices are quite recent [8]. Our exploration of this topic builds on the prior work in noise removal and identify tradeoffs and evaluate performance in the context of wearable EOG.

Multi-stage adaptive NLMS filter. Our first approach uses an adaptive filter to remove noise. A significant body of work on ECG and PPG (e.g. Schäck et al. [50]) has compared various types of adaptive filters such as the Kalman Filter, LMS, NLMS and others and generally found that NLMS offers the best performance for physiological signal filtering and has low computational complexity.

One issue that we face in using adaptive filters is that classical adaptive filters such as RLS and LMS are designed as single-input single-output (SISO) systems, i.e. they receive a one-dimensional signal as input and estimate the value of a one-dimensional desired output. In more complex scenarios, the SISO filters need to be extended and combined together to build more complex multiple-input single-output (MISO) or multiple-input multiple-output (MIMO) filters [24, 31, 50].

While typical motion artifact removal filters for ECG and PPG are relatively small since they use only the accelerometer as input and have only one output stream (e.g. PPG), our pipeline is larger. In our case, both the gyroscope and accelerometer are important since the gyroscope contains more information about head movements whereas the accelerometer contains information about large bodily movements (we show this empirically in §7.3). In addition, we have four output EOG axes over which the motion artifacts are present. Thus, we have six input streams (three accelerometer and three gyroscope axes) and four output streams (H, V, L, R axes of EOG). For each output stream, we need six NLMS filters per input stream and one adaptive filter to fuse these outputs. In other words, we need a total of 30 filters.

Figure 8(a) shows the overall structure of our model. For estimating each of the 4 axes of the EOG signal, a bank of filters comprised of 6 adaptive NLMS filters are used. Each of the 6 filters is dedicated to one of the 6 axes of the IMU signal and tries to estimate the motion artifact on the EOG signal given a history of IMU readings. The filters' weights are continuously updated during the online operation of the system to adapt to the EOG signal. The estimates made by the NLMS filters are then linearly combined with time-varying weights determined by a Kalman filter to make the final estimate. The final motion artifact estimate is removed from the corrupted EOG signal to recover a relatively clean signal.

We use NLMS filters of order 50, i.e. the filters require a history of 50 samples of IMU signals to estimate the current value the motion artifact on the EOG signal. The learning rate of the filters is also set to 0.05.

Neural network-based model. Given the high complexity of the adaptive filtering model, we also look at a neural network-based regression model that estimates the motion artifact by leveraging IMU information.

At a high level, this model takes all the six axes (accelerometer + gyroscope) as input and simultaneously predicts the four axes of the EOG signal resulting from the motion. The model structure is depicted in Figure 8(b). The proposed model contains one fully connected hidden layer (20 nodes) and a scaling post-processing module. The input is a vector of size 300 (window size of 50 x 6-axis IMU signal) and the output is a vector of size 4, which is the estimated motion artifact on the 4-axis EOG signal corresponding to the same time step as the last sample of the input window. In order to estimate all the motion artifact of all the samples, we shift the input window by one frame. ReLU activation functions are used to add some level of non-linearity to our model so it can capture non-linear relations between IMU readings and the EOG's motion artifacts.

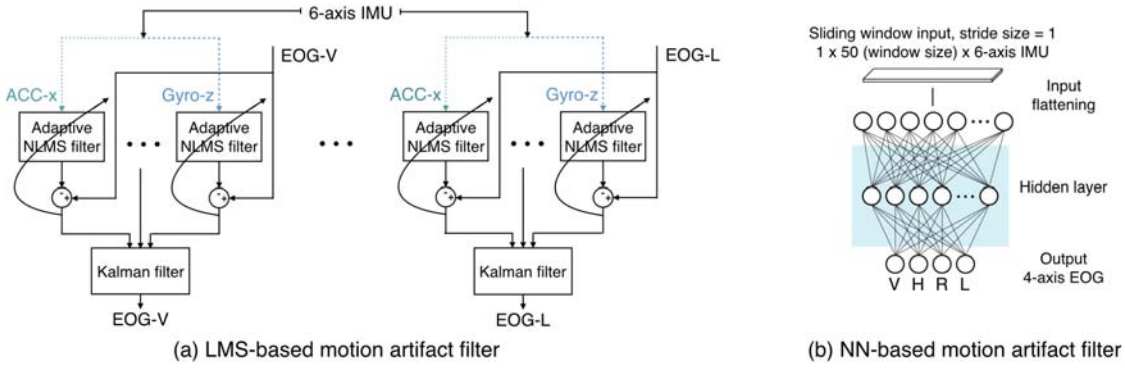


Fig. 8. Architectures of motion artifact multi-stage adaptive filter and neural network filter.

Similar to how adaptive filters can learn the pattern for each individual, we adapt our regression model by retraining our neural network model with a few minutes of individual's data. Also, we adaptively scale the amplitude of the estimated motion artifact signal so that it has the same standard deviation as that of the raw EOG signal calculated over a window of 50 data samples. This scaling helps the filter adapt to variations in signal strength due to eyeglass displacements over time.

5.2 Stage 2: Facial Action Detection

Once motion artifacts are removed and a significant EOG change is detected, we run our CNN model to detect instances of action units. Since blinks are frequently seen in the data, we also explicitly filter them out to avoid mis-classifying them as other action units. Thus, we use two different CNN models of the same architecture but different input size for detection of blink and other facial action units. We use a binary classifier CNN with the input window size of half a second (50 samples) for blink detection and a multi-class (5 classes) classifier with the input window size of one second (100 samples) for other facial action units. The reason for training a separate model for blink detection is that the duration of blinks are relatively much smaller than other action units and hence they require a shorter window size to be detected accurately. The mentioned window sizes were chosen experimentally.

CNN architecture. Our CNN model contains two convolutional layers and one fully connected layer (model structure and specifications are shown in Figure 9). We experimentally choose the hyperparameters by splitting our collected data to train (%80), validation (%10), and test dataset (%10). We implement our CNN model with the TensorFlow library [3].

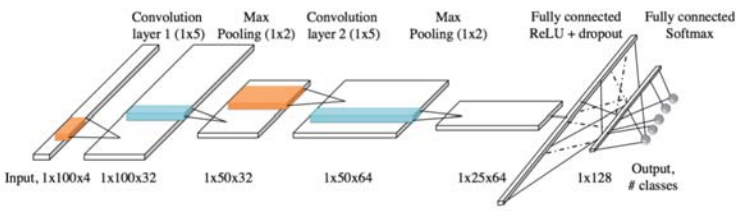
To produce a probabilistic class output, the softmax transformation is applied on the output layer.

$$y_i = \frac{exp(u_i)}{\Sigma_{j=1}^{k} exp(u_j)} \tag{3}$$

where u_i are the original outputs and y_i are the transformed outputs, and k is the number of outputs.

The network is trained using the Adam optimizer and with cross entropy chosen as the loss function. Furthermore, in order to control the complexity of the model and prevent co-adaptation of the features, we used the dropout technique on the fully connected layer.

One question that might arise is why we do not use a more complex and deep CNN structure. The reason is that we are limited by the number of participants we can recruit for a user study, and in particular by the number



(a) CNN architecture

Input size	height = 1 (time series data) width = window size dimension = 4-axis EOG
Output size	number of classes = 5
Convolution filter 1 size	1 x 5 x 32
Convolution filter 2 size	1 x 5 x 64
Fully connected layer	128
Learning rate	0.0001
Dropout probability	0.5

(b) CNN model specifications

Fig. 9. CNN architecture of facial action units classifier and its model specifications.

of times we can ask each participant to repeat an action unit. For example, we found out that it is exhausting for users to frown numerous times due to the resultant muscle fatigue. Therefore, we do not ask them to repeat each action more than 20 times, which in turn limits the total number of labeled data we have available. A deep CNN would be vulnerable to over-fitting to the data, hence we keep to the simplest model that is capable of discerning different facial action units given the EOG stream.

Personalizing with transfer learning. An additional issue that we had to consider is signal variation across individuals. While individual variability is normal when we deal with signals obtained from wearables, the eyeglass is particularly sensitive to such variability due to several factors. The first is the face shape and, in particular, the shape of the nose-bridge. This affects the contact between EOG electrodes and the nose which determines how much of the signal from muscle movements in the rest of the face can be observed. The second is the fit of the glasses behind the ear, which determines the lateral and vertical movement of the glasses during head turning and mobility. A snug fit means less movement of the glasses and more consistent signal. The third is variation in the way individuals use upper facial muscles which depends on the shape of the face, level of muscle control, and other factors.

To address this, we personalize the model to each new user. We obtain a few samples from each individual while performing facial gestures and use these labeled data for performing transfer learning on our trained CNN model. In order to do this, we just re-train the last layer (fully connected layer) of the CNN model with personalization data.

5.3 Reducing Computational Complexity

The above two-stage pipeline also helps us reduce the average computational requirement (and therefore power consumption). The motion artifact removal stage is considerably cheaper to execute compared to a CNN for action classification. Both the adaptive filter and neural network for motion artifact removal require far fewer computations compared to a full CNN which has two convolutional layers (see §7.3 for empirical estimates).

To reduce computational cost, we trigger the CNN model to run only when substantial EOG activity is detected after the motion artifact removal stage. Similarly, we trigger the motion artifact removal model to run only when significant variation is observed in the raw EOG signal. Thus, our multi-stage classification pipeline also acts as a triggering pipeline, allowing us to reduce power consumption in a real-time detection setting.

6 DATASET COLLECTION AND GROUND TRUTH LABELING

In this section we explain the details of our user study and ground truth labeling approach. We collect our data with J!NS MEME eyewear and sample the EOG sensor, accelerometer and gyroscope at 100Hz. All of these datasets were collected under Institutional Review Board approval.

6.1 Motion Artifact Dataset

In order to evaluate the performance of W!NCE motion artifact removal stage, we collected an "EOG noise" dataset comprising head-movement and walking data from 5 participants, 2 male and 3 female with average age of 23.

We asked each participant to perform three different head-movement patterns such as up-and-down (pitch), shoulder-to-shoulder (roll), and side-to-side (yaw) each for 12 minutes with different velocities (in total 36 minutes of head-movement data from each participant). For example, participants were asked to perform fast and slow nodding in a way that they would naturally do during a conversation. The participants were asked to do these actions with eyes closed to avoid eye movement signal cross-talk.

In order to collect data about walking-induced noise, we asked the participants to walk on a treadmill at a speed of 2 miles/hour for 10 minutes. We remove involuntary facial expressions (such as blinks and eye movements) from the data to ensure that the dataset only contains walking-induced EOG noise.

6.2 FACS Classification Dataset

In order to test the overall performance of W!NCE, we collected three datasets from 17 participants, 9 male and 8 female with average age of 26.

Stationary + facial actions dataset. We collected these data by asking all the 17 participants to perform the target upper facial action units while wearing the J!NS MEME eyeglasses. In order to provide a baseline reference for each participant, we first show them an example of each target upper facial action unit provided by a visual guidebook of Facial Action Coding System (FACS) [13]. Since, we need our participants to perform each target class multiple times, in order to ease the process, we make a looped video clip that contains 20 times of each action unit example. Then, during the collection process, the name of the action and the clip is provided to the participants and they are asked to perform the facial actions along with the video. Figure 10(a) shows one of the participants performing the task in the stationary setting.

In addition to the facial action target classes, which are brow raiser (AU01), brow lower (AU04), cheek raiser (AU06), nose wrinkler (AU09), and blink (AU45), we also add a "none" class in which we asked the participants to have a conversation that normally engages head movements. The participants are asked to avoid facial expressions during the conversation. Although they may have some involuntary facial expressions (such as blinks), we remove them from the data to provide the "none" label. In total, we collect 20 samples of each target class for every user, resulting in 1700 samples excluding blinks, for which we simply extract all involuntary blinks happened during the data collection process.

Walking + facial actions dataset. In order to understand our ability to extract upper facial expressions while the user is mobile, the same experiment is conducted on the same participants while walking on a treadmill at a speed of 2 miles/hour. Figure 10(b) shows the data collection setup for the mobile setting. This dataset also includes 20 samples of each target class for each user, resulting in 1700 samples excluding blinks, for which we again simply extract all involuntary blinks happened during the data collection process.

Head-movement + facial actions dataset. Another important scenario is facial expression with head movements since nodding and shaking of the head is common during conversation. However, we encountered significant difficulty in collecting a live dataset where participants performed both head movements and facial

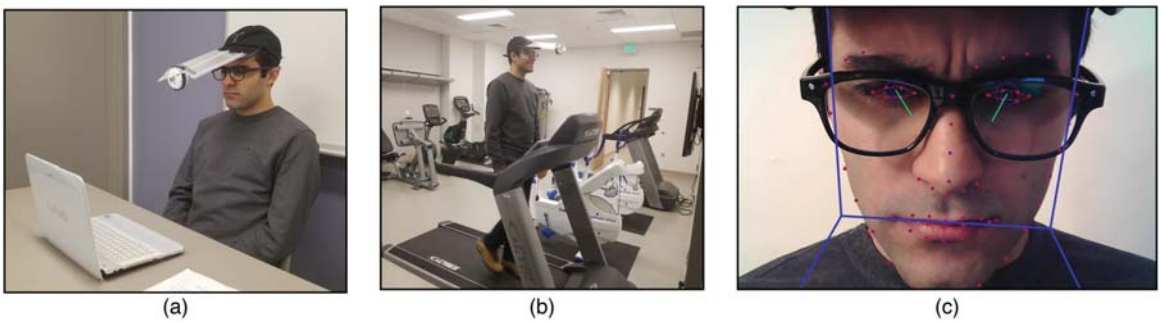


Fig. 10. Data collection setup: the participant is performing different facial action units in stationary (a) and mobile (b) settings. Figure (c) represents the output of OpenFace.

actions simultaneously. While people tend to unconsciously perform head movement with facial actions, we found that participants found it difficult to start and end facial actions while performing specific head gestures.

Instead, we use the "EOG noise" dataset described in §6.1 and add it to the EOG signal obtained in a stationary setting to synthetically create this dataset. The use of a noise model is common in physiological signal processing — for example, in the case of ECG signals, a noise model is often added to ECG data to simulate various noisy environments.

Ground truth labeling. In order to collect the ground truth measures for the given target classes, we use OpenFace, an open source facial behavior analysis toolkit [4]. OpenFace is capable of facial landmark detection, head pose estimation, facial action unit recognition, and eye-gaze estimation. In W!NCE, the OpenFace output values for intensity of the target classes brow raiser (AU01), brow lower (AU04), cheek raiser (AU06), nose wrinkler (AU09), and blink (AU45) are used to detect the onset and offset of each action. In order to avoid OpenFace's misannotation, we also leveraged the scripts and the time stamps of the facial actions video clip to validate OpenFace's output.

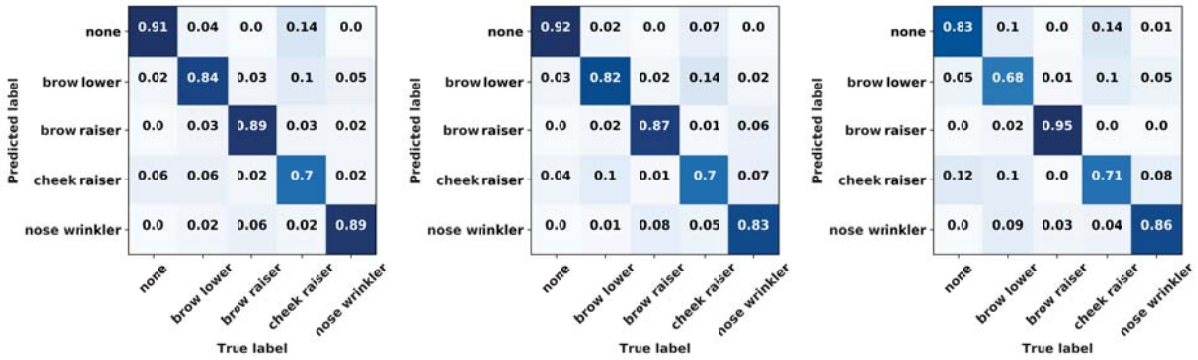
We record the participants' face video throughout the experiments with a wireless IP camera (sampling rate of 30 Hz) that is connected to the Image Acquisition MATLAB Toolbox and provides the time stamps of the video for the synchronization purpose. Figure 10(c) shows an example of OpenFace face annotation output. In order not to restrict the participant's head pose and head movement, the camera is attached to a cap that is worn by the participant (shown in Figure 10).

7 EVALUATION

We present the performance of W!NCE in three parts. To ensure focus on the key results, we start with the overall results in terms of facial action recognition and pain monitoring. We then describe several low-level benchmarks for individual stages of W!NCE.

7.1 Evaluation of Facial Action Unit (FACS) Detection

In this section, we summarize the overall performance of W!NCE for facial action detection in stationary and ambulatory settings. The results in this section use neural network-based motion artifact removal (see §7.3 for a comparison between noise-removal approaches).



(a) Stationary (b) Walking (c) Head-movement

Fig. 11. Classification results of FACS detection: the confusion matrices show the performance for the stationary dataset, walking dataset, and the head-movement dataset.

Table 2. W!NCE performance (over 1700 upper face action units and 8300 blinks). We see strong results across the board including scenarios that involve motion artifacts due to head movement and walking.

Facial action	Setting	Precision	Recall	F1 Score
FACS	**Stationary**	0.88 ± 0.06	0.88 ± 0.07	0.88 ± 0.06
	Walking	0.84 ± 0.08	0.83 ± 0.08	0.83 ± 0.08
	Head-movement	0.82 ± 0.07	0.80 ± 0.06	0.81 ± 0.07
Blinks	**Stationary**	0.96 ± 0.09	0.94 ± 0.08	0.95 ± 0.08
	Walking	0.90 ± 0.11	0.91 ± 0.1	0.90 ± 0.1
	Head-movement	0.91 ± 0.09	0.89 ± 0.08	0.90 ± 0.09

FACS detection performance. Table 2 shows the precision, recall and F1 score over all 17 participants for stationary, walking, and head-movement datasets. We show the results of blink classification separately from other facial actions since blinks use a shorter window size compared to the other classification tasks.

In a stationary scenario, we have an F1 score of 0.88 for FACS detection and this dips a bit to 0.83 and 0.81 for the more noisy walking and head movement scenarios. The F1 score for blink detection is higher — we get 0.95 for the stationary dataset and 0.9 for the more noisy walking and head movement datasets. Overall, these results are very promising and show that facial gestures can generally be accurately detected across a range of scenarios.

Figure 11 shows the confusion matrices of FACS detection for the different datasets. While the results are good for most of the FACS actions, "cheek raiser" seems to be confused with other classes more frequently. This is because we generally find that the signal is a bit weaker for the cheek raiser action compared to other actions due to the distance of the action from the electrodes. (This difference is also visible in Figure 4c where the cheek raiser signal has less amplitude compared to other actions.)

SNR improvement after filtering:. Figure 12 shows how much the signal to noise ratio (SNR) improves as a result of motion artifact removal (using the NN filter). The SNR increases significantly (about 6.5 dB) among all the facial action units. The benefit is particularly significant in the case of weak actions such as blink and cheek raiser since these actions involve small muscles that do not generate a strong EMG signal (blink) or occur farther

away from the electrodes (cheek raiser). The mean SNR is higher in case of brow lower, brow raiser, and nose wrinkler since these are relatively large actions and the muscles are closer to the electrodes.

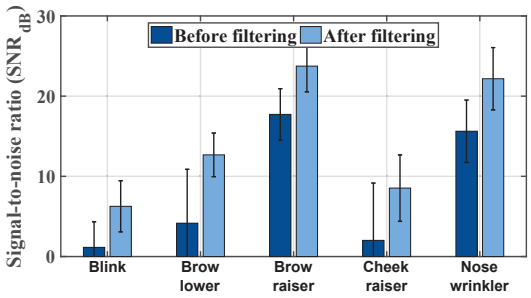


Fig. 12. Comparing the SNR_{dB} of Head-movement + Facial actions dataset before and after motion artifact removal.

Figure 13 illustrates the SNR improvement by visualizing the time-series before and after filtering. The participant is performing a nodding gesture from t = [0, 4], blinking from t = [4.5, 6] and from t = [12.5, 13.5], and moving her head slowly up and down from t = [7, 12] and t = [14, 16]. For brevity, we just plot the model's performance over vertical EOG signals rather than all four axes. We can clearly see that a significant fraction of the noise is removed by using the noise removal method while the actual facial expression signal (blinks) remains intact.

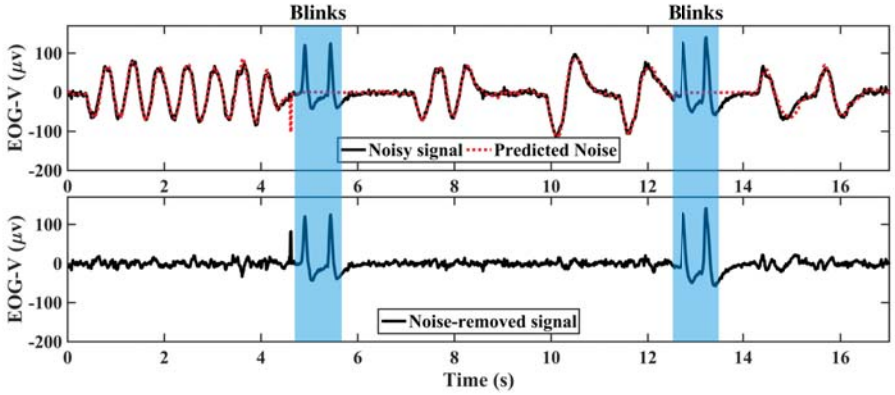


Fig. 13. Sample of motion artifact removal over EOG axes of vertical. The participant is performing nodding gesture t = [0, 4], blinking t = [4.5, 6] and t = [12.5, 13.5], and moving her head slowly up and down t = [7, 12] and t = [14, 16].

7.2 Case Study: Passive Pain Monitoring

A compelling and unique application for W!NCE is in the context of continuous pain monitoring. According to the National Health Interview Survey in 2012, 11% of the U.S. adult population experienced daily pain and 17.6% experienced intermittent pain. Chronic pain is the most common cause of long-term disability in the United States. Pain is also expensive, and studies have shown the national cost of pain to be roughly $600 billion annually [17]. Among the consequences of the inability to objectively monitor and manage pain is an increase in opioid

use which in turn has led to perhaps the worst addiction crisis the nation has faced. A device like W!NCE can help to make objective measurements during normal, ambulatory settings as opposed to relying on self-reports, thereby providing more information to health providers [35].

Previous research has studied the use of vision-based methods to determine facial action units, and map from these actions to pain. Specifically, Prkachin et al. [41] show that the action units brow lower (AU4), cheek raiser (A6), lid tightner (A7), nose wrinkler (A9), upper lip raiser (A10), and eye closure (A43) hold the majority of the information about pain. They also define the Prkachin and Solomon pain scale based on their observations:

$$Pain = AU4 + (AU6||AU7) + (AU9||AU10) + AU43 \quad (4)$$

where the sum of AU4, AU6 or AU7 (whichever is higher), AU9 or AU10 and AU43 provide a metric for pain degree. Lucay et al. [29] use the Prkachin and Solomon pain scale for pain detection by leveraging a video-based facial action detection system for estimating the different action unit intensities. In contrast to video-based approaches, W!NCE can be used in ambulatory settings for real-world measurement of symptoms of pain. Since we calculate a subset of the AUs in the Prkachin-Solomon pain scale, we calculate a limited version of their pain scale, where $pain_{W!NCE} = AU4 + AU6 + AU9$.

To examine W!NCE's potential for continuous and passive pain monitoring, we conduct a small case study. The dataset for this experiment is collected from 5 participants (2 male and 3 female), randomly chosen from the previous 17 participants group. Note that from the previous study we already have the participants' FACS dataset. In order to collect pain instances dataset, we need to induce pain on the participants. For that, we used algometers, which are widely used for identifying the pressure-pain threshold in clinical research [26]. The algometer has a small and pointed probe to apply pressure in a small probe area (approximately $1cm^2$). We use the algometer on the medial part of the participant's knee and the pressure is gradually increased until pain is felt by the participant. Each experiment is a 5-minute session, during which pain is induced on the participants' knee once every 30 seconds for 10 times while they are wearing the J!NS eyeglasses.

Pain detection results. Figure 15 illustrates the time series of the estimated pain level by W!NCE for one of the participants as a sample. The highlighted regions on the plot denote instances of pain. As it is shown in the figure, the estimated pain level successfully reveals actual instances of pain. It should also be noted that the probability varies a little since there are blinks and other actions in between instances when pain is induced but the large spikes are due to pain.

Once we calculate the pain scale probability, we use a very simple pain detection method that simply choose the peaks with probability higher than 0.5 as the pain instances. Figure 14 shows precision, recall and F1-score across five participants for pain detection using this method. Our results are uniformly high and we have an F1-score of 0.81 to 0.95 across the five participants. This shows that we can use W!NCE as an objective measure of facial grimacing due to pain.

Figures 16a and 16b provides a further breakdown across the different facial action units to show which of them were detected during instances of pain. Each row of the figure corresponds to one instance of pain being induced and each column shows the probability of a specific action unit. What we would like to see is that the action units that are related to pain, i.e. nose wrinkler, cheek raiser, and brow lower have higher scores compared to action units that are unrelated to pain such as brow raiser and none actions. We see that this is indeed the case. The results are very clean for participant B and a bit more uncertain for participant A but generally, we see that facial actions associated with pain have higher scores.

7.3 W!NCE Benchmarks

So far, we have highlighted the end-to-end benefits of using W!NCE but the underlying design choices in the W!NCE pipeline were driven by extensive benchmarking efforts. In the context of motion artifact removal, we looked at which method performed best, what IMU axes are most informative for filtering, and the contribution of

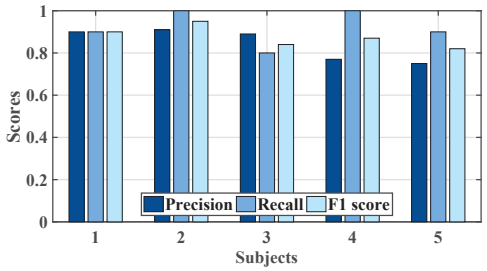


Fig. 14. Precision, recall, and F1 score of binary pain detection for 5 participants.

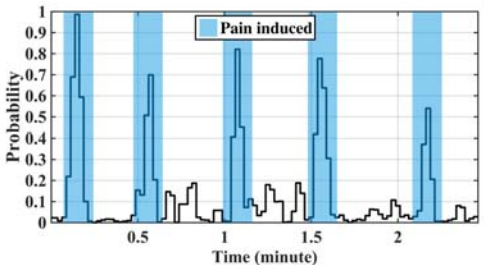
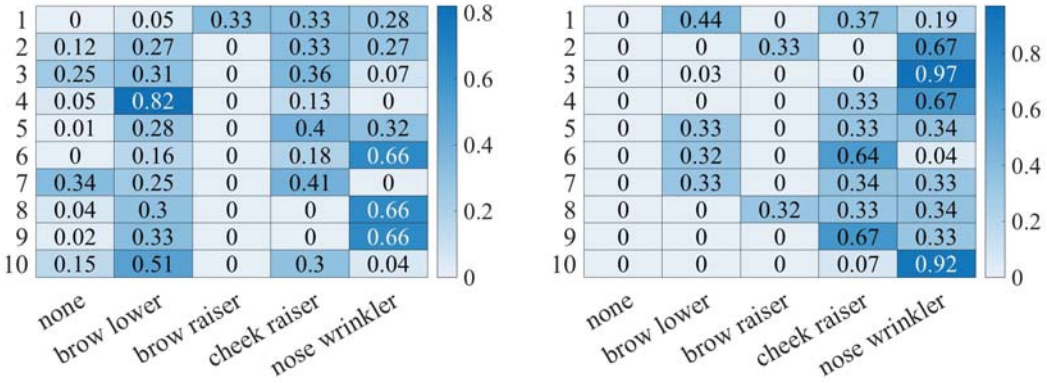


Fig. 15. Estimated pain scale probability during pain induction.



(a) Participant A

(b) Participant B

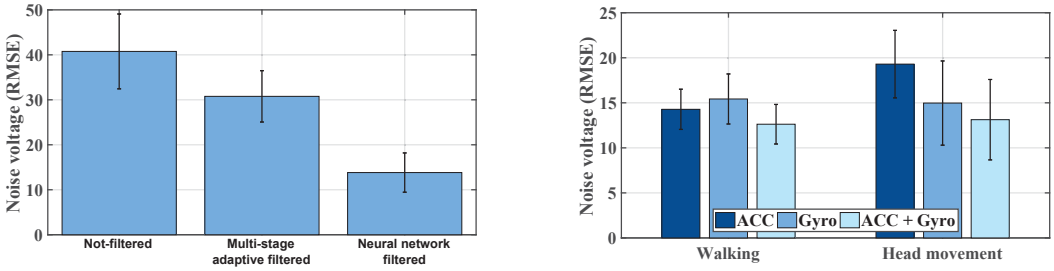
Fig. 16. Estimated score from our CNN-based upper facial action model for different upper face actions during 10 pain instances for participants A and B.

motion artifact removal stage on the overall W!NCE's performance. In the context of the facial action recognition module, we looked at different methods, robustness to various noise sources, and benefits of personalization. We also looked at various combinations of the filtering and classification modules including a combined CNN versus a two-stage pipeline. We present a salient subset of these benchmarks in this section.

Comparison of motion artifact removal methods. First, we compare the performance of the motion artifact removal methods described in §5.1. In order to evaluate our filtering methods, we use the "EOG noise" dataset described in §6.1. The evaluation is done in a leave-one-subject-out (LOSO) manner, where the model is trained on data from 4 participants and tested on the last participant with personalization.

Figure 17a shows the RMSE before filtering and after filtering with each of the two approaches. Note that lower noise level means that the filtering approach performs better. As can be seen, the neural network filter works significantly better and has roughly half the RMSE as the multi-stage adaptive filter.

We believe there are two reasons for this performance gap. First, as described in §5.1, we need to train filter parameters for each EOG output channel separately for the adaptive filter whereas the neural network is trained simultaneously across all four EOG outputs and can leverage correlations across them. Second, the nonlinear ReLU activation function of the neural network filter enables us to capture non-linear IMU-EOG correlations.



(a) Comparison of motion artifact removal methods.

(b) Importance of ACC and Gyro for filtering.

Fig. 17. Figure shows benchmarks for motion artifact removal module. (Left) RMSE after using the neural network filter is half as much as after using the multi-stage adaptive filter. (Right) accelerometer carries more information about walking whereas gyroscope has more information about head movements. Fusing both provides most information

The multi-stage adaptive filter, on the other hand, assumes linear relation between the IMU signal and the corresponding motion artifacts on the EOG.

Contribution of accelerometer and gyroscope to filtering performance. Our filtering methods used both the accelerometer and gyroscope as input, but one question is how much each of these contribute to filtering performance. To understand this, we train and compare three variants of the neural network filtering model, in which it receives as input 1) only Accelerometer signals (ACC), 2) only Gyroscope signals (Gyro), and 3) both Accelerometer and Gyroscope signals.

Figure 17b summarizes the results. We see that for the head-movement dataset, the gyroscope-only approach works significantly better than the accelerometer-only approach. This is intuitive because in this dataset the motion artifact is generated due to the head rotations in different directions which is best captured by the gyroscope. However, the contributions are reversed for the walking dataset where the accelerometer carries more information. This is because the motion artifacts largely come due to steps and large body motions that change electrode impedance. Fusing the accelerometer and gyroscope signals reduces motion artifacts the most. This is consistent with our observation that the two signals can provide complementary information — head rotation is captured via the gyroscope and whole-body movements via the accelerometer.

Analysis of motion artifact removal effectiveness on the overall FACS detection performance. Here, we look at three EOG-only classification methods to understand how much the motion artifact removal stage benefits the classification pipeline. We use the head movement dataset for this evaluation since we have control over how we add the noise to the signal which allows us to compare different ways of constructing the classification pipeline. As with the previous experiments, the motion artifact removal stage is trained on data that is separate from the noise that is injected into the signal.

We compare FACS classification result for three scenarios: 1) CNN classifier trained on clean stationary dataset and tested on the motion artifact-added dataset, 2) The full W!NCE pipeline with the motion artifact removal stage followed by the CNN classifier tested on the motion artifact-added dataset, and finally 3) CNN classifier trained and tested on the clean stationary dataset with no motion artifacts. The third scenario corresponds to the upper-bound of the performance of FACS classifier and corresponds to how well the classifier performs in the absence of any additional motion-induced noise.

Figure 18 compares the F1 score of the above mentioned scenarios over all the 17 participants. The difference between the first bar and the last bar is remarkable (roughly .07 in F1-score) which indicates that there can be a substantial performance dip if we do not use some method to remove motion-induced noise from the data. The

difference between the full W!NCE pipeline and the upper bound (when there is no noise) is relatively small (F1-score gap of .03) which shows that we can remove a significant fraction of noise via the filtering stage.

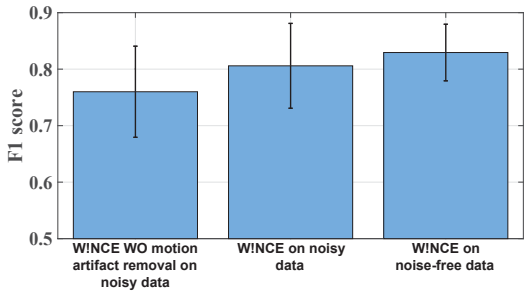


Fig. 18. Analysis of motion artifact removal effectiveness on the overall FACS detection performance: a) CNN classifier trained on clean stationary dataset and tested on the motion artifact-added dataset, b) The full W!NCE pipeline tested on the motion artifact-added dataset, and c) This is the upper-bound of the performance of FACS classifier in which CNN classifier trained and tested on the clean stationary dataset.

Comparing W!NCE against single-stage classification. While W!NCE uses a two-stage pipeline, a more direct approach would be to directly provide the EOG and IMU signals to the CNN and classify into facial actions. We now compare these two approaches.

Figure 19 shows the results in the stationary, walking, and head-movement settings. In a stationary setting, we see that the two-stage pipeline actually hurts performance a bit but this reverses in settings with more noise. This is because when the participant performs facial gestures while stationary, the motion artifact removal stage removes some useful information about head movement patterns. For example, facial expressions such as moving the brows cause small eyeglass shifts and slippage along the nose bridge that are captured by the IMU and help boost classification accuracy. But in situations with realistic head or body movement, the signal about facial actions in the IMU is tiny compared to the considerably higher amplitude signal about walking and head movements. Thus, even though the two-stage pipeline loses some performance in stationary settings, it is beneficial in the vast majority of settings where we would normally expect head movement together with facial gestures are performed (e.g. during a conversation).

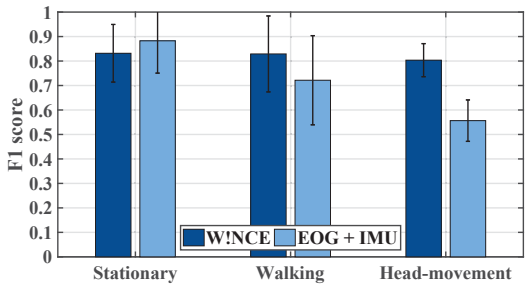


Fig. 19. Comparing W!NCE performance with the model that is trained on raw EOG and IMU signals (instead of motion artifact filtered EOG signal).

Benefits of personalization. So far, our results have assumed that the models were personalized to each user; we now look at how many samples per class we need for personalization to achieve high performance and how much benefit we obtain from such personalization. In order to do this, we assess W!NCE's performance with respect to the number of samples used for personalizing the model for each user (for test dataset, we set aside 10 out of 20 samples for each class). Here, the W!NCE's performance is evaluated for all three datasets (stationary, walking, and head-movement) over all participants and then the mean F1 score values are reported.

As it can be seen in Figure 20, as the number of samples used for personalization increases, so does the F1 score. However, the rate of increase gradually decreases especially after having 6 samples for personalization. Therefore, we chose 6 as the number of samples required per class for personalization training stage in upper face AU detection as it is also a reasonably small number to ask each new user to perform each facial action for model adaptation. Please note that the performance of the model with personalization has been improved by 28% comparing to the non-personalized model (0 personalization samples), which is quite impressive.

Figure 21 illustrates why personalization is so effective by looking at one of the classes, brow lower, for two participants. While the signals are visually quite similar, the specifics are different due to face shape, tightness of eyewear on the ear, and the facial muscle responses. The amplitude of the signal is clearly different but also the crossing between the different vertical and horizontal streams are different. As a result, personalization helps to improve results particularly since our data does not have hundreds of participants to overcome these variations.

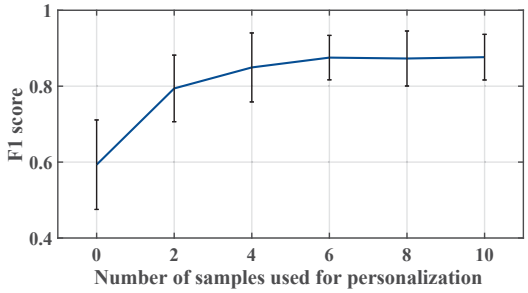
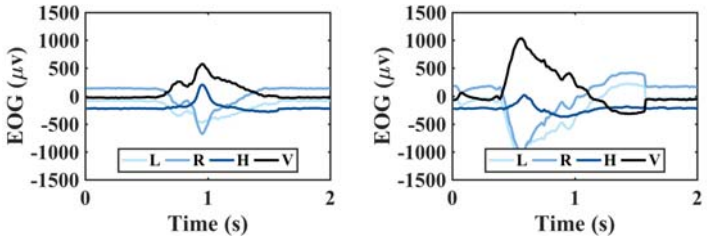


Fig. 20. W!NCE's performance with respect to number of samples per class used to personalize the model for each user.



(a) Participant A (b) Participant B

Fig. 21. Samples of brow lower action unit for two participants

Analysis of computation overhead of W!NCE's stages. We now provide some brief computation benchmarks for W!NCE. The J!NS MEME hardware is not modifiable, hence we cannot directly measure the computation overhead. Instead, we look at the computation cost if the wearable eyeglass were to be equipped with a low-power

Table 3. Computation overhead of W!NCE's stages.

Model	Exec. time (ms)	Avg. energy (μJ)
Multi-stage adaptive filter	0.038	1.16
Neural network filter	0.052	1.62
CNN FACS classifier	2.77	142.09

neural accelerator, specifically the GAP8 low-power neural accelerator [1]. To obtain these numbers, we use published benchmarking specifications in [36] and adapt them to the specifics of our network.

We see that both the adaptive filtering and neural network methods execute very quickly (30-50μs) and consume equivalent amount of energy ($1\mu J - 2\mu J$). The adaptive filter would normally be less expensive but the cost becomes equivalent to a neural network since we end up with thirty filters to deal with many input and output channels as described in §5.1. Both these methods consume more than two orders of magnitude lower energy than the CNN model for facial action unit detection. Even so, the CNN model can execute in less than three milliseconds. Since we expect the CNN to be triggered only when large EOG movements are detected after motion artifact filtering, the overall overhead of processing is not high.

8 CONCLUSIONS AND FUTURE WORK

In conclusion, our work explores how we can leverage the muscle activation signal that leaks into EOG electrodes to extract upper facial micro-expressions. Importantly, we show that this can be done with an easy to wear commercial eyeglass that has EOG electrodes embedded in the nose bridge and an IMU embedded in the temples of the eyeglass that rests behind the ear. Our results show that EOG carries much more information about upper facial actions than the IMU, and that key upper facial actions can be extracted with an F1 score of 0.88 in stationary settings and 0.82 in ambulatory settings. We also develop a motion artifact removal pipeline and show that a substantial fraction of artifacts due to head and body movement can be effectively removed from a multi-channel EOG signal. We demonstrate the utility of W!NCE in the context of pain sensing with a small user study where we demonstrate that several key upper facial action units are engaged upon inducing pain. Our results are promising and present a new way to use wearables for important applications like pain and emotion sensing without requiring intrusive cameras.

Our work opens up several directions that we plan to pursue. On the applications side, our preliminary results on pain monitoring using W!NCE are promising but will need to be validated in larger clinical trials. On the technical side, one direction that we are exploring is to further enhance our signal extraction pipeline and further leverage the complementary information of IMU signal to capture the small signals from faraway muscles to extract better measures of lower facial actions like talking, chewing, and laughing. By combining upper and lower face action detection, we can improve our ability to obtain a holistic sense of facial expressions to better understand the context surrounding upper facial expressions associated with pain or emotional expressions. We also plan to improve the motion artifact removal pipeline, so it can eliminate every possible motion artifact pattern as the neural network motion artifact filter is now confined to predict a pre-trained set of head-movements. We consider to make the training process of the motion artifact removal stage online, so the model can both personalize and enhance through time in an unsupervised manner.

ACKNOWLEDGMENTS

The authors would like to thank all of the reviewers for detailed feedback on the paper. This work was partially supported by grants and awards from the National Institutes of Health (1U54EB020404-01) and the National Science Foundation (1815347 and 1839999).

REFERENCES

[1] 2018. GAP8 Developer Boards. (14 Nov. 2018). https://greenwaves-technologies.com/en/developer-kits/.

[2] 2018. JINS MEME Eyewear. (14 Oct. 2018). https://jins-meme.com/en/.

[3] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Gregory S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian J. Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Józefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Gordon Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul A. Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda B. Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *CoRR* abs/1603.04467 (2016). arXiv:1603.04467 http://arxiv.org/abs/1603.04467

[4] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 2016. OpenFace: An open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 1–10. https://doi.org/10.1109/WACV.2016.7477553

[5] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: Using Wearable Sensors to Detect Eating Episodes in Unconstrained Environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 37 (Sept. 2017), 20 pages. https://doi.org/10.1145/3130902

[6] Vinay Bettadapura. 2012. Face Expression Recognition and Analysis: The State of the Art. *CoRR* abs/1203.6722 (2012). arXiv:1203.6722 http://arxiv.org/abs/1203.6722

[7] Giorgio Bonmassar, Patrick L. Purdon, Iiro P. Jääskeläinen, Keith Chiappa, Victor Solo, Emery N. Brown, and John W. Belliveau. 2002. Motion and Ballistocardiogram Artifact Removal for Interleaved Recording of EEG and EPs during MRI. *NeuroImage* 16, 4 (2002), 1127 – 1141. https://doi.org/10.1006/nimg.2002.1125

[8] Andreas Bulling, Daniel Roggen, and Gerhard Tröster. 2009. Wearable EOG Goggles: Seamless Sensing and Context-awareness in Everyday Environments. *J. Ambient Intell. Smart Environ.* 1, 2 (April 2009), 157–171. http://dl.acm.org/citation.cfm?id=2350315.2350320

[9] A. Bulling, J. A. Ward, H. Gellersen, and G. Troster. 2011. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 4 (April 2011), 741–753. https://doi.org/10.1109/TPAMI.2010.86

[10] Alexander J. Casson, Arturo Vazquez Galvez, and Delaram Jarchi. 2016. Gyroscope vs. accelerometer measurements of motion from wrist PPG during physical exercise. *ICT Express* 2, 4 (2016), 175 – 179. https://doi.org/10.1016/j.icte.2016.11.003 Special Issue on Emerging Technologies for Medical Diagnostics.

[11] Maurizio Codispoti, Paola Surcinelli, and Bruno Baldaro. 2008. Watching emotional movies: Affective reactions and gender differences. *International Journal of Psychophysiology* 69, 2 (2008), 90 – 95. https://doi.org/10.1016/j.ijpsycho.2008.03.004

[12] Emotiv. 2018. Emotiv: Mobile EEG Brainwear. (2018). https://www.emotiv.com/

[13] FACS. 2018. Facial Action Coding System (FACS). (2018). https://imotions.com/blog/facial-action-coding-system/

[14] B. Farhang-Boroujeny. 1998. *Adaptive Filters: Theory and Applications* (1st ed.). John Wiley & Sons, Inc., New York, NY, USA.

[15] B. Fasel and Juergen Luettin. 2003. Automatic facial expression analysis: a survey. *Pattern Recognition* 36, 1 (2003), 259 – 275. https://doi.org/10.1016/S0031-3203(02)00052-3

[16] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti. 2014. Empatica E3 — A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *2014 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*. 39–42. https://doi.org/10.1109/MOBIHEALTH.2014.7015904

[17] Darrell J. Gaskin and Patrick Richard. 2012. The Economic Costs of Pain in the United States. *The Journal of Pain* 13, 8 (2012), 715 – 724. https://doi.org/10.1016/j.jpain.2012.03.009

[18] Michael J. Grimble and Michael A. Johnson (Eds.). 2005. *Adaptive Filtering*. Springer London, London, 3–23. https://doi.org/10.1007/1-84628-121-0_1

[19] A. Gruebler and K. Suzuki. 2014. Design of a Wearable Device for Reading Positive Expressions from Facial EMG Signals. *IEEE Transactions on Affective Computing* 5, 3 (July 2014), 227–237. https://doi.org/10.1109/TAFFC.2014.2313557

[20] M. Hamedi, S. h. H. Salleh, M. Astaraki, and A. M. Noor. 2013. EMG-based facial gesture recognition through versatile elliptic basis function neural network. *Biomed Eng Online* 12 (Jul 2013), 73. https://doi.org/10.1186/1475-925X-12-73

[21] P. S. Hamilton and M. G. Curley. 1997. Adaptive Removal of Motion Artifact. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering' (Cat. No.97CH36136)*, Vol. 1. 297–299. https://doi.org/10.1109/IEMBS.1997.754531

[22] S. Hickson, N. Dufour, A. Sud, V. Kwatra, and I. Essa. 2017. Eyemotion: Classifying facial expressions in VR using eye-tracking cameras. *ArXiv e-prints* (July 2017). arXiv:cs.CV/1707.07204

[23] Shoya Ishimaru, Kai Kunze, Yuji Uema, Koichi Kise, Masahiko Inami, and Katsuma Tanaka. 2014. Smarter Eyewear: Using Commercial EOG Glasses for Activity Recognition. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 239–242. https://doi.org/10.1145/2638728.2638795

[24] D. Jarchi and A. J. Casson. 2016. Estimation of heart rate from foot worn photoplethysmography sensors during fast bike exercise. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 3155–2158. https://doi.org/10.1109/EMBC.2016.7591398

[25] CRAIG N. KARSON. 1983. SPONTANEOUS EYE-BLINK RATES AND DOPAMINERGIC SYSTEMS. *Brain* 106, 3 (1983), 643–653. https://doi.org/10.1093/brain/106.3.643 arXiv:/oup/backfile/content-public/journal/brain/106/3/10.1093/brain/106.3.643/2/106-3-643.pdf

[26] Ann M Kinser, William A Sands, and Michael H Stone. 2009. Reliability and validity of a pressure algometer. *The Journal of Strength & Conditioning Research* 23, 1 (2009), 312–314. https://doi.org/10.1519/JSC.0b013e31818f051c

[27] J. Kwon, D. H. Kim, W. Park, and L. Kim. 2016. A wearable device for emotional recognition using facial expression and physiological response. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 5765–5768. https://doi.org/10.1109/EMBC.2016.7592037

[28] Boreom Lee, Jonghee Han, Hyun Jae Baek, Jae Hyuk Shin, Kwang Suk Park, and Won Jin Yi. 2010. Improved elimination of motion artifacts from a photoplethysmographic signal using a Kalman smoother with simultaneous accelerometry. *Physiological Measurement* 31, 12 (2010), 1585. http://stacks.iop.org/0967-3334/31/i=12/a=003

[29] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, and K. M. Prkachin. 2011. Automatically Detecting Pain in Video Through Facial Action Units. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 41, 3 (June 2011), 664–674. https://doi.org/10.1109/TSMCB.2010.2082525

[30] 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 (IUI '16)*. ACM, New York, NY, USA, 317–326. https://doi.org/10.1145/2856767.2856770

[31] M. Boloursaz Mashhadi, E. Asadi, M. Eskandari, S. Kiani, and F. Marvasti. 2016. Heart Rate Tracking using Wrist-Type Photoplethysmographic (PPG) Signals during Physical Exercise with Simultaneous Accelerometry. *IEEE Signal Processing Letters* 23, 2 (Feb 2016), 227–231. https://doi.org/10.1109/LSP.2015.2509868

[32] MD2K. 2014. MD2K: NIH Center of Excellence on Mobile Sensor Data-to-Knowledge. (2014). https://md2k.org/

[33] N Meziane, J G Webster, M Attari, and A J Nimunkar. 2013. Dry electrodes for electrocardiography. *Physiological Measurement* 34, 9 (2013), R47. http://stacks.iop.org/0967-3334/34/i=9/a=R47

[34] MUSE. 2015. MUSE: The Brain Sensing Headband. (2015). http://www.choosemuse.com/

[35] NIH. 2017. NIH: Development of a Device to Objectively Measure Pain. (2017). https://grants.nih.gov/grants/guide/rfa-files/RFA-DA-18-012.html

[36] Daniele Palossi, Antonio Loquercio, Francesco Conti, Eric Flamand, Davide Scaramuzza, and Luca Benini. 2018. Ultra Low Power Deep-Learning-powered Autonomous Nano Drones. *CoRR* abs/1805.01831 (2018). arXiv:1805.01831 http://arxiv.org/abs/1805.01831

[37] M. Pantic and L. J. M. Rothkrantz. 2000. Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 12 (Dec 2000), 1424–1445. https://doi.org/10.1109/34.895976

[38] PIP. 2016. PIP: The Stress Management Device. (2016). https://thepip.com/

[39] Karolien Poels and Siegfried Dewitte. 2006. How to capture the heart? Reviewing 20 years of emotion measurement in advertising. *Journal of Advertising Research* 46, 1 (2006), 18–37. https://doi.org/10.2139/ssrn.944401

[40] Suranai Poungponsri and Xiao-Hua Yu. 2013. An adaptive filtering approach for electrocardiogram (ECG) signal noise reduction using neural networks. *Neurocomputing* 117 (2013), 206 – 213. https://doi.org/10.1016/j.neucom.2013.02.010

[41] Kenneth M Prkachin and Patricia E Solomon. 2008. The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. *Pain* 139, 2 (2008), 267–274. https://doi.org/10.1016/j.pain.2008.04.010

[42] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, and K. A. Reddy. 2012. A Novel Approach for Motion Artifact Reduction in PPG Signals Based on AS-LMS Adaptive Filter. *IEEE Transactions on Instrumentation and Measurement* 61, 5 (May 2012), 1445–1457. https://doi.org/10.1109/TIM.2011.2175832

[43] V. Rantanen, H. Venesvirta, O. Spakov, J. Verho, A. Vetek, V. Surakka, and J. Lekkala. 2013. Capacitive Measurement of Facial Activity Intensity. *IEEE Sensors Journal* 13, 11 (Nov 2013), 4329–4338. https://doi.org/10.1109/JSEN.2013.2269864

[44] M. A. D. Raya and L. G. Sison. 2002. Adaptive noise cancelling of motion artifact in stress ECG signals using accelerometer. In *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] [Engineering in Medicine and Biology*, Vol. 2. 1756–1757 vol.2. https://doi.org/10.1109/IEMBS.2002.1106637

[45] F. C. Robertson, T. S. Douglas, and E. M. Meintjes. 2010. Motion Artifact Removal for Functional Near Infrared Spectroscopy: A Comparison of Methods. *IEEE Transactions on Biomedical Engineering* 57, 6 (June 2010), 1377–1387. https://doi.org/10.1109/TBME.2009.2038667

[46] Soha Rostaminia, Addison Mayberry, Deepak Ganesan, Benjamin Marlin, and Jeremy Gummeson. 2017. iLid: Low-power Sensing of Fatigue and Drowsiness Measures on a Computational Eyeglass. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 23 (June 2017), 26 pages. https://doi.org/10.1145/3090088

[47] T. Sakai, R. Yoshida, H. Tamaki, T. Ogitsu, H. Takemura, H. Mizoguchi, M. Namatame, F. Kusunoki, E. Yamaguchi, S. Inagaki, Y. Takeda, M. Sugimoto, and R. Egusa. 2015. Electrodermal activity based study on the relationship between visual attention and eye blink. In *2015 9th International Conference on Sensing Technology (ICST)*. 596–599. https://doi.org/10.1109/ICSensT.2015.7438468

[48] E. Sariyanidi, H. Gunes, and A. Cavallaro. 2015. Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, 6 (June 2015), 1113–1133. https://doi.org/10.1109/TPAMI.2014.2366127

[49] Jocelyn Scheirer, Raul Fernandez, and Rosalind W. Picard. 1999. Expression Glasses: A Wearable Device for Facial Expression Recognition. (1999).

[50] T. Schäck, C. Sledz, M. Muma, and A. M. Zoubir. 2015. A new method for heart rate monitoring during physical exercise using photoplethysmographic signals. In *2015 23rd European Signal Processing Conference (EUSIPCO)*. 2666–2670. https://doi.org/10.1109/EUSIPCO.2015.7362868

[51] K. L. Shapiro, J. E. Raymond, and K. M. Arnell. 1994. Attention to visual pattern information produces the attentional blink in rapid serial visual presentation. In *Journal of Experimental Psychology: Human Perception and Performance*. 357–371. https://doi.org/10.1037/0096-1523.20.2.357

[52] K. Suzuki, F. Nakamura, J. Otsuka, K. Masai, Y. Itoh, Y. Sugiura, and M. Sugimoto. 2017. Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display. In *2017 IEEE Virtual Reality (VR)*. 177–185. https://doi.org/10.1109/VR.2017.7892245

[53] B. Taji, S. Shirmohammadi, V. Groza, and I. Batkin. 2014. Impact of Skin–Electrode Interface on Electrocardiogram Measurements Using Conductive Textile Electrodes. *IEEE Transactions on Instrumentation and Measurement* 63, 6 (June 2014), 1412–1422. https://doi.org/10.1109/TIM.2013.2289072

[54] Toshiyo Tamura, Yuka Maeda, Masaki Sekine, and Masaki Yoshida. 2014. Wearable Photoplethysmographic Sensors—Past and Present. *Electronics* 3, 2 (2014), 282–302. https://doi.org/10.3390/electronics3020282

[55] N. V. Thakor and Y. . Zhu. 1991. Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection. *IEEE Transactions on Biomedical Engineering* 38, 8 (Aug 1991), 785–794. https://doi.org/10.1109/10.83591

[56] Q. Xue, Y. H. Hu, and W. J. Tompkins. 1992. Neural-network-based adaptive matched filtering for QRS detection. *IEEE Transactions on Biomedical Engineering* 39, 4 (April 1992), 317–329. https://doi.org/10.1109/10.126604

Received November 2018; Accepted January 2019