Prototyping Smart Eyewear with Capacitive Sensing for Facial and Head Gesture Detection

Denys J.C. Matthies
Technical University of Applied Sciences Lübeck, GER
denys.matthies@th-luebeck.de

Alex Woodall
Augmented Human Lab, The University of Auckland, NZ
woodall@ahlab.org

Bodo Urban
Fraunhofer IGD Rostock, University of Rostock, GER
bodo.urban@igd-r.fraunhofer.de

ABSTRACT

The prevalence of smart eyewear, particularly in the form of eye glasses, has continuously increased. In this paper, we explore instrumenting a pair of glasses with Capacitive Sensing (CapSense) technology. We conducted three studies, in which we (1) explore a suitable CapSense setup for rapid prototyping; (2) reveal the potential of detecting facial and head gestures with a CapSense glasses prototype; and (3) investigate the technical feasibility of using a passive electric field sensing. Instead of training a low-dimensional gesture set that works across different users (e.g., eye winks), we selected a set of 12 facial and head related gestures, which we classified by relying on user-dependent machine learning models.

CCS CONCEPTS

• **Human-centered computing** → **Interaction devices**; **Ubiquitous and mobile computing systems and tools**; **Ubiquitous and mobile devices**.

KEYWORDS

Capacitive Sensing; Electric Field Sensing; Facial Expressions; Smart Glasses; Eyewear; Prototyping; Machine Learning; Data Mining

ACM Reference Format:
Denys J.C. Matthies, Alex Woodall, and Bodo Urban. 2021. Prototyping Smart Eyewear with Capacitive Sensing for Facial and Head Gesture Detection. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3460418.3479361

1 INTRODUCTION

Facial expressions naturally occur and currently exist as grounds for research in the area of affective computing [12]. The detection of facial expressions has been of particular interest for collaborative Mixed Reality applications [10]. Circumventing the problem of expensive classification algorithms, recently, Facebook-Research [16] demonstrated a complex facial-gesture-mapping to a virtual avatar using a GAN Deep-Learning approach based on 9 cameras attached to a VR-HMD. Other optical approaches have been demonstrated by Masai et al. [6, 7] and Ishimaru et al. [3]. Another typical sensing approach is using an IMU, particularly in conjunction with EOG, such as showcased in KissGlass [5], W!NCE [15], Smarter Eyewear [4], and EOG Glasses [1]. Utilizing a different technology being inexpensive could be a Capacitive Sensing (CapSense) as demonstrated in literature [13, 14]. A 'face-hugging device' [14], nevertheless, may be too obtrusive to be socially acceptable. The first mobile approach was demonstrated in 2017 in EarFieldSensing [8], using a simple earplug, enabling the recognition of 5 gestures with reasonable accuracy of 90%. Wearable CapSense approaches that in particular utilize a proximity sensing are not yet explored widely.

In this paper, we explore the recognition of facial and head gestures using glasses prototypes. Our results underpin CapSense to be powerful enough to detect activities arising from our face.

2 STUDY 1: RAPID PAPER PROTOTYPE

The setup of CapSense, only requiring a single electrode in "Loading Mode", is simple and enables fairly flexible usage (e.g., detecting touch and hovering). Inspired by the idea of rapid paper prototyping, which is also common for hardware prototyping [11], we cut out a few pairs of glasses from cardboard and attached copper electrodes.

2.1 Apparatus

The initial glasses prototype consists of a cardboard, at which four copper tape electrodes (2 cm × 2 cm and 2 cm × 5 cm) are attached to the frame facing the user (see Figure 1).

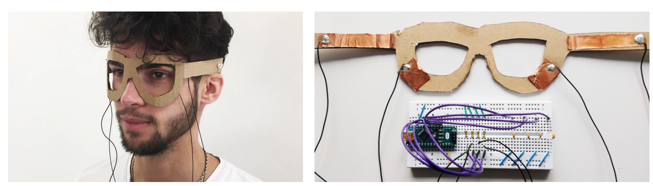

Figure 1: We built several rapid paper prototypes in shape of a pair of glasses to explore technical feasibility. A USB-wired Arduino streams the capacitive measurements picked up by the copper tape.

Our prototype relies on the Arduino CapSense library from Paul Badger. [1] The CapSense library takes a measurement for each of these electrodes in turn by setting the corresponding digital output pin (e.g., D2, D4, D6 or D8) to HIGH and measuring the latency until the shared digital input pin (e.g., D10) exhibits the same voltage. The higher the latency, the higher the capacity. To increase the stability of the sensor readings, a small capacitor is placed between each of the digital output pins and the ground. This way, we slightly decrease sensitivity but reduce judder and increases SNRs.

[1]**Paul Badger's CapSense Library:** http://playground.arduino.cc/Main/CapacitiveSensor

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3479361

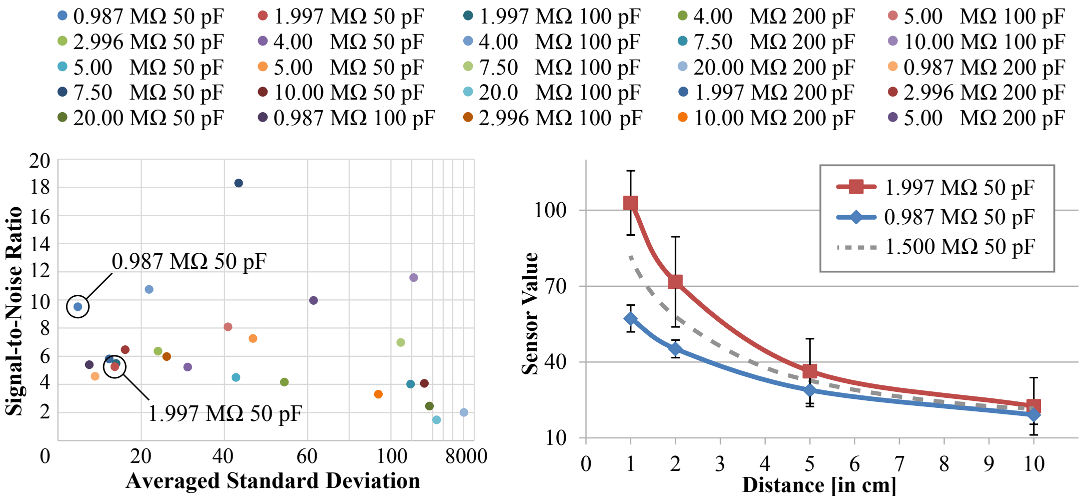

Figure 2: Showing the averaged Standard Deviation of all distances in comparison to their Signal-to-Noise Ratio. The setup of 1.5 MΩ and 50 pF seems to provide the best trade-off.

2.2 Study Design

To determine a suitable setup, we examined reasonable combinations of resistors (1, 2, 3, 4, 5, 7.5, 10 & 20 MΩ) and capacitors (50, 100, 200 pF). We carried out a study, in which we measured 10 values for 4 distances (0, 2, 5, 10 cm to skin) for all combinations. We were looking for a high SNR, which reflects a relatively low overlapping of the standard deviation ranges for 1, 2 and 5 cm distances.

2.3 Results

As indicated in Figure 2, the 0.987 MΩ 50 pF configuration yields relatively low standard deviation ranges and a comparably high SNR. However, sensing range reduces with small resistors. The curve of the 1.997 MΩ 50 pF configuration is steep enough, but exhibited standard-deviation ranges that are too large, as the sensed values cannot adequately be assigned to distances with the required resolution and confidence. For our continuing investigation, we determined the combination of a 1.5 MΩ resistor and a 50 pF capacitor to provide sufficient resolution for detecting facial expressions.

3 STUDY 2: LOW FIDELITY PROTOTYPE

As we previously determined a suitable hardware setup, we then continued our exploration investigating the spectrum of facial and head gestures in more detail with an improved prototype. We developed another prototype, a purchasable pair of glasses with some attached electronics.

3.1 Apparatus

As we already determined the best hardware configuration for a capacitive sensing glasses frame (Resistor: 1.5 MΩ, Capacitor: 50 pF), we implemented a three channel CapSense using this setup. Since the facial gestures we examine are symmetrical in nature, positioning the electrodes on one side of the glasses is sufficient for an initial investigation. Each copper electrode was attached to a 3D-printed layer, which was glued to a standard glasses frame (Nerd Clear^{TM}) – see Figure 3. We used an Arduino Mini Pro 3.3 V 328PU and a HC-05 Bluetooth modem, to process and transmit the data to a computer. In order to improve sensor data, we still connected a wire to the ground of a Power Supply Unit (PSU). Figure 4 depicts the streamed raw data gathered by the three electrodes.

3.2 Study Design

With our established hardware setup, we focused on recognizing facial expressions. In a pilot study, we recorded a gesture set of 25 (as demonstrated in Reference [8]) and determined 12 gestures to be detectable. Then, the main study was conducted with 10 participants, students, and research fellows aged between 25 and 51 years. Two of the subjects were female. The users' task was to perform the gesture set, consisting of 12 facial expressions, including the default class (relaxed face), with each gesture repeated 10 times. The participants were instructed, but did not undergo a training phase. Therefore, execution styles also slightly varied across users. The data was gathered with 100 Hz. The length of a gesture was restricted to 1.28 s due to the window size of 128 (Raw data: see Figure 4). Gestures were obtained by manually triggering the recording of a window just before performing the gesture. We computed 46 state-of-the-art features on the collected raw data. Since classifiers were trained on a per user basis, we refrained from using techniques aimed at obtaining a measure for time-independent between-gesture similarity, such as dynamic time warping.

3.3 Results

In a pilot study, we determined a suitable classifier and a useful gesture set using a *stratified 10-fold-cross-validation* on a set of 25 gestures [8]. We evaluated 5 state-of-the-art classifiers, which

Figure 3: The 2nd glasses prototype consists of a purchasable glasses frame with three copper electrodes, connected to an Arduino Pro Mini, which is driven by a 240 mAh battery. A Bluetooth modem and a charging unit is implemented.

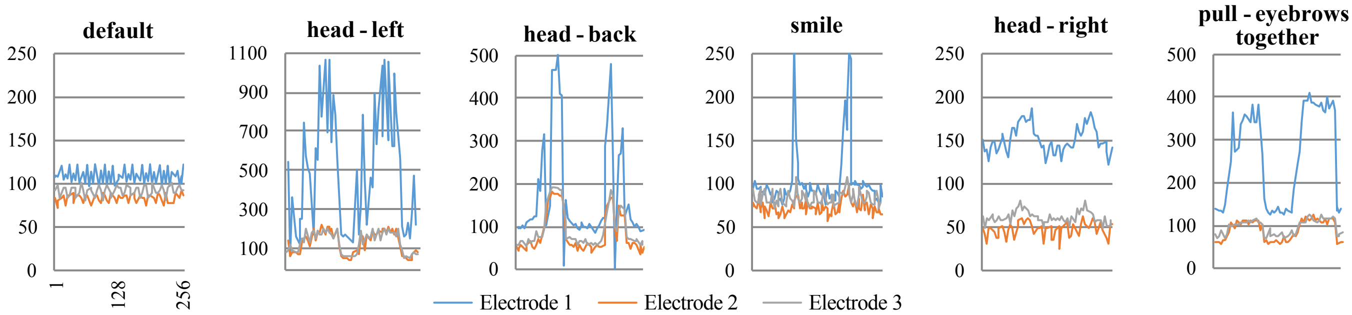

Figure 4: Displaying the raw data signals of six gestures from a random user. The graphs include two repetitions of the same gesture. As the sensing electrodes are only attached to a single side, certain gestures may be better detected, such as head moves to the left in contrast to the right.

performed significantly different from each other. Based on its significantly higher mean performance, we selected the *Random Forest* for our following analysis. In terms of facial gestures, vocalizations were hard to distinguish. Except for the eye-wink, muscle activity around the Orbicularis Oculi was hardly detectable. Head movements and broader gestures were detected comparably well. Therefore, we reduced the gesture set to the 12 gestures that are shown by Matthies et al. [8].

a	b	c	d	e	f	g	h	i	j	k	l	<
70	3	2	0	7	4	0	4	1	4	4	1	a
4	82	4	1	1	7	0	0	0	1	0	0	b
2	3	76	2	0	1	4	3	5	2	1	1	c
6	1	0	73	4	4	0	0	2	0	8	2	d
4	5	0	2	74	8	0	1	1	1	4	0	e
6	5	1	3	17	65	2	0	1	0	0	0	f
1	1	6	0	2	2	73	4	3	4	4	0	g
4	1	2	1	0	0	5	66	6	10	5	0	h
2	2	6	0	0	2	6	12	56	12	2	0	i
5	1	5	0	1	0	6	7	16	57	1	1	j
4	3	2	5	2	2	2	5	1	3	70	1	k
1	0	0	1	0	0	0	0	1	0	1	96	l

Figure 5: Confusion Matrix showing true-positive rates and confusions [in %] using a *Random Forest* (Average accuracy: 71.5%). The top 5 gestures include: (l) smile, (b) default, (e) head-left, (d) head-back, and (g) lift-eyebrows.

The reduced gesture set achieved a true-positive recognition of 71.5% across all users (mean *F1* = 0.7083, σ = 0.2135)—see Figure 5. However, the recorded data also presents two users as statistical outliers (P2: 38.33%, and P6: 27.5%), while the other eight users scored an average of 81.15% accuracy (*mean F1* = 0.80675, σ = 0.0485). Although Weka's *stratified cross-validation* showed accuracy rates around 71.5%, the system would not result in satisfying usability. Having three erroneously detected gestures out of ten is unacceptable. Furthermore, error rates will rise at a realistic interaction scenario. Therefore, we further reduced the gesture set and removed all gestures with an accuracy lower than 70%, which included all mouth-gestures, and the head-right movement. To simulate a more realistic scenario, we conducted an additional method; the *leave-*$k_{Instances}$*-out* method (k = 5), in which we compared five unknown gestures to five gestures the system has already learned. Being aware that poor performance can occur in reality, we did not declare further outliers. The results are summarised in Table 1.

Table 1: Accumulated accuracy rates across all 10 users comparing sizes of gesture sets with a *stratified 10-fold cross-validation*, and a *leave-k-out* method (*Random Forest*).

2	3	4	5	6	7	*n*-Gestures
99.5%	95%	91.7%	89.6%	86.5%	84.1%	**10-fold**
98%	95.3%	86%	85.2%	82.3%	66%	**leave-k$_5$-out**

As indicated by a *one-way ANOVA* for correlated samples ($F_{4,36}$ = 80.01; $p < 0.0001$, smaller gesture sets (using a *leave-k-out* method) enable significantly higher recognition rates. A *Tukey HSD* proofed a gesture set with $n = 2$ to provide significantly higher results than a set with $n = 7$ ($p < 0.01$). No further statistical differences were found. We see this study as an early stage to gain first insights on how a future pair of capacitive glasses could possibly perform. More professional CapSense devices, such as the *FDC2214 by Texas Instruments* [2] may provide higher accuracy.

4 STUDY 3: REFINED PROTOTYPE

In this study, we introduce a refined 3D-printed prototype comparing a passive electric field sensing that relies on a differential amplification against an active capacitive sensing in loading mode.

4.1 Apparatus

We enlarged the frame width to allocate for a maximised surface area of electrodes, potentially increasing detection accuracy. Eight electrodes on the glasses were connected to electric field sensing shields [8] (see Figure 7), commonly described as a passive capacitive sensing [2]. In active capacitive sensing, such as in loading mode, typically at least one electrode is frequently being charged, while the time between the signal being sent and received is measured. The reference electrode is located at a relatively far distance from the signal electrodes, such as on the user's neck (see Figure 8). The sensor data is streamed via USB and thus the prototype is grounded to the computer.

[2]**Texas Instruments FDC2214:** http://www.ti.com/lit/ds/symlink/fdc2214.pdf

Passive Sensing (EarFieldSensing)

a	b	c	d	e	f	g	h	i	j	k	l	<
80	0	0	0	0	0	20	0	0	0	0	0	a
0	100	0	0	5	0	0	0	0	0	0	0	b
0	0	75	0	20	5	0	0	0	0	0	0	c
0	0	0	80	20	0	0	0	0	0	0	0	d
0	0	0	30	70	0	0	0	0	0	0	0	e
0	0	0	0	10	90	0	0	0	0	0	0	f
0	0	0	0	0	0	100	0	0	0	0	0	g
0	0	0	0	0	0	0	100	0	0	0	0	h
0	0	0	0	0	0	0	0	100	0	0	0	i
0	0	0	0	0	0	0	0	0	100	0	0	j
0	0	0	0	0	0	0	0	0	0	100	0	k
0	0	0	0	0	0	0	0	0	0	0	100	l

Active Sensing (Arduino CapSense Library)

a	b	c	d	e	f	g	h	i	j	k	l	<
15	0	0	0	10	0	0	40	0	0	0	35	a
5	60	5	0	0	0	5	5	5	0	0	15	b
0	20	5	5	5	0	0	0	0	0	35	30	c
0	5	5	15	10	5	0	0	0	10	40	10	d
25	0	0	15	15	5	0	0	5	0	25	10	e
0	0	5	10	5	5	0	0	35	5	25	10	f
0	0	0	0	0	0	80	10	5	0	0	5	g
0	0	0	0	0	0	25	45	25	0	0	5	h
5	0	0	0	5	5	0	30	45	0	0	10	i
0	0	0	25	5	5	0	0	0	35	30	0	j
0	0	0	25	0	10	0	0	0	20	45	0	k
15	15	10	0	15	5	5	10	25	0	0	0	l

Figure 6: Head movements are prone to confusion for both technologies. The eye-wink and chin-on-chest was also not always clear. Still, most gestures were easily identified using the EarFieldSensing [8] Technology. In contrast, the Arduino CapSense Library proved to be less reliable, resulting in massive confusions [values in %].

4.2 Study Design

To gain first impressions of the performance, we ran a single subject pilot study. The procedure was similar to that of study 2, as we recorded the same gesture set of 12 gestures with 10 repetitions. First we recorded a training set containing all gestures. After some time, we recorded another set, which we used as a test set. We ensured that both sets were timely separated from each other. The gestures were executed in a window of 256 values, giving the user 2.5 s to perform them (at a 100 Hz sampling rate). The collected recordings of all gestures were computed, to which 46 features were calculated as optimally usable for conventional machine learning algorithms. Using the 3d-printed glasses, we evaluated both the passive (EarFieldSensing [8]) and active (Arduino Library) CapSense. The active sensing is a loading mode using the previously determined resistor & capacitor configuration.

4.3 Results

In study 2, we already determined the *Random Forest* as a suitable classifier for our type of data. Using the passive electric sensing method (EarFieldSensing), the glasses prototype was able to predict all 12 gestures with reasonably high accuracy (see Table 2 and Figure 6). Based on our previous experience with study 2, we do not anticipate a substantial raise or drop of accuracy with more extensive user studies. Overall, the accuracy produced a score of 91.25% (*mean F1* = 0.9105, σ = 0.0487). Using the active sensing technique, the Arduino Library demonstrated a lower performance level of 63.3% accuracy (*mean F1* = 0.605, σ = 0.07).

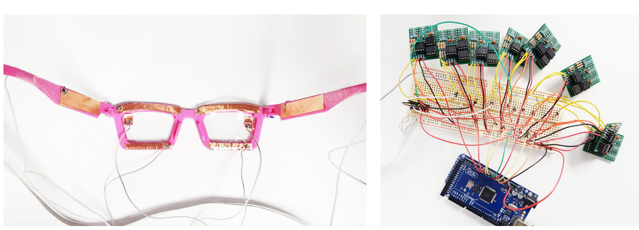

Figure 7: Our third prototype consists of a 3D printed glasses frame with six copper electrodes and two Indium Tin Oxide (ITO) electrodes. All eight electrodes are connected to an electric field sensing shield each being driven by an Arduino Mega.

Table 2: Showing True-Positive (TP), False-Positive (FP) rates and F1 score for both setups: Active Sensing: Paul Badger's CapSense and the passive sensing: EarFieldSensing.

TP	FP	F1	Sensing Technology
.9125	.008	.9105	**EarFieldSensing**
.633	.0335	.605	**Arduino CapSense**

These results indicate the performance differences between both setups. It is striking that the Arduino CapSense implementation demonstrated comparably low performance. The EarFS technology (passive sensing) seems to have the potential to outperform the Arduino CapSense Library (active sensing). This result may appear surprising, given that active sensing should exceed the reliability of passive sensing. The fact that we specifically tweaked the passive sensing for our purpose explains our a result.

5 CONCLUSION

In our investigation, we explored different types of CapSense with several glasses prototypes. We could demonstrate that training a machine learning model works with reasonably high accuracy for a rather great gesture set. During all investigations the participants were always in a seated position and not moving greatly around. This benefited the high accuracy. However, in truly mobile scenarios, we envision a different solution; one that generates an own small local EM-field, which can override environmental noise [9].

Figure 8: The prototype when worn. The reference electrode is attached to the user's neck, but could be placed at any other position, such as the ear lobe.

ACKNOWLEDGMENTS

This research was funded in part by the Federal Republic of Germany (BMWi 16KN049102 & 16KN049131), Fraunhofer Gesellschaft, Tertiary Education Commission New Zealand (3716751), and by *Assistive Augmentation* research grant under the Entrepreneurial Universities (EU) initiative of New Zealand. We acknowledge that some parts of study 1 has been adapted from the Master's Thesis entitled: *"A Wearable Facial Expression Control Interface to accomplish Microinteractions in Mobile Scenarios"* by Bernhard Arthur Strecker. Further we declare to have all rights to publish these parts following the §6 of the agreement *"Vertrag mit Masterstudenten"* signed on the 21.08.2015.

REFERENCES

[1] George Chernyshov, Kirill Ragozin, Benjamin Tag, and Kai Kunze. 2019. EOG Glasses: an Eyewear Platform for Cognitive and Social Interaction Assessments in the Wild. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–5.

[2] Tobias Grosse-Puppendahl, Yannick Berghoefer, Andreas Braun, Raphael Wimmer, and Arjan Kuijper. 2013. OpenCapSense: A rapid prototyping toolkit for pervasive interaction using capacitive sensing. In *Proceedings of the 2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 152–159. https://doi.org/10.1109/PerCom.2013.6526726

[3] Shoya Ishimaru, Kai Kunze, Koichi Kise, Jens Weppner, Andreas Dengel, Paul Lukowicz, and Andreas Bulling. 2014. In the blink of an eye: combining head motion and eye blink frequency for activity recognition with google glass. In *Proceedings of the 5th augmented human international conference*. ACM, 1–4. https://doi.org/10.1145/2582051.2582066

[4] 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*. 239–242.

[5] Richard Li, Juyoung Lee, Woontack Woo, and Thad Starner. 2020. KissGlass: Greeting Gesture Recognition using Smart Glasses. In *Proceedings of the AHs '20: Augmented Humans International Conference (AHs '20)*. ACM. https://doi.org/10.1145/3384657.3384801

[6] 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 AHs '20: Augmented Humans International Conference (AHs '20)*. ACM. https://doi.org/10.1145/3384657.3384787

[7] Katsutoshi Masai, Yuta Sugiura, Katsuhiro Suzuki, Sho Shimamura, Kai Kunze, Masa Ogata, Masahiko Inami, and Maki Sugimoto. 2015. AffectiveWear: towards recognizing affect in real life. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, 357–360. https://doi.org/10.1145/2800835.2800898

[8] Denys JC Matthies, Bernhard A Strecker, and Bodo Urban. 2017. Earfieldsensing: a novel in-ear electric field sensing to enrich wearable gesture input through facial expressions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1911–1922. https://doi.org/10.1145/3025453.3025692

[9] Denys JC Matthies, Chamod Weerasinghe, Bodo Urban, and Suranga Nanayakkara. 2021. CapGlasses: Untethered Capacitive Sensing with Smart Glasses. In *Proceedings of the Augmented Humans International Conference 2021 (AHs '21)*. ACM. https://doi.org/10.1145/3458709.3458945

[10] Dhwani Mehta, Mohammad Siddiqui, and Ahmad Javaid. 2018. Facial emotion recognition: A survey and real-world user experiences in mixed reality. *Sensors* 18, 2 (2018), 416. https://doi.org/10.3390/s18020416

[11] Katherine A Mirica, Joseph M Azzarelli, Jonathan G Weis, Jan M Schnorr, and Timothy M Swager. 2013. Rapid prototyping of carbon-based chemiresistive gas sensors on paper. *Proceedings of the National Academy of Sciences* 110, 35 (2013), E3265–E3270. https://doi.org/10.1073/pnas.1307251110

[12] Rosalind W Picard and Roalind Picard. 1997. Affective computing. Vol. 252. *MIT press Cambridge."EEG-detected olfactory imagery to reveal covert consciousness in minimally conscious state." Brain injury* 29, 13-14 (1997), 1729–1735.

[13] Ville Rantanen, Pekka-Henrik Niemenlehto, Jarmo Verho, and Jukka Lekkala. 2010. Capacitive facial movement detection for human–computer interaction to click by frowning and lifting eyebrows. *Medical & biological engineering & computing* 48, 1 (2010), 39–47. https://doi.org/10.1007/s11517-009-0565-6

[14] Ville Rantanen, Hanna Venesvirta, Oleg Spakov, Jarmo Verho, Akos Vetek, Veikko Surakka, and Jukka Lekkala. 2013. Capacitive measurement of facial activity intensity. *IEEE Sensors journal* 13, 11 (2013), 4329–4338. https://doi.org/10.1109/JSEN.2013.2269864

[15] 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. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–26. https://doi.org/10.1145/3314410

[16] Shih-en Wei, Jason Saragih, Tomas Simon, Adam W. Harley, Stephen Lombardi, Michal Perdoch, Alexander Hypes, Dawei Wang, Hernan Badino, and Yaser Sheikh. 2019. VR Facial Animation via Multiview Image Translation. *ACM Transactions on Graphics* 38, 4 (2019), 67. https://doi.org/10.1145/3306346.3323030