The Bigger the Better? Towards EMG-Based Single-Trial Action Unit Recognition of Subtle Expressions

Dennis Küster[1] [a], Rathi Adarshi Rammohan[1] [b], Hui Liu[1] [c], Tanja Schultz[1] [d] and Rainer Koschke[2] [e]

[1] *Cognitive Systems Lab, University of Bremen, Bremen, Germany*

[2] *AG Software Engineering, University of Bremen, Bremen, Germany*

Keywords: Action Units, Electromyography, Facial Action Coding System, EMG, sEMG, fEMG, Subtle Expressions, Pattern Recognition, Machine Learning.

Abstract: Facial expressions are at the heart of everyday social interaction and communication. Their absence, such as in Virtual Reality settings, or due to conditions like Parkinson's disease, can significantly impact communication. Electromyography (EMG)-based facial action unit recognition (AUR) offers a sensitive and privacy-preserving alternative to video-based methods. However, while prior research has focused on peak intensity action units (AUs), there has been a lack of research on EMG-based AURs for lightweight recording of subtle expressions at multiple muscle sites. This study evaluates EMG-based AUR for both low- and high-intensity expressions across eight AUs using two types of mobile electrodes connected to the Biosignal Plux system. The results of four subjects indicate that even limited data may be sufficient to train reasonably accurate AUR models. Larger snap-on electrodes performed better for peak-intensity AUs, but smaller electrodes resulted in higher performance for low-intensity expressions. These findings suggest that EMG-based AUR is viable for subtle expressions from short data segments and that smaller electrodes hold promise for future applications.

1 INTRODUCTION

Even if the face is not a proverbial "window to the soul", the notion that facial expressions play a key role in everyday nonverbal communication can be dated back to Charles Darwin's seminal work on "The Expression of Emotions in Man and Animal" (Darwin, 1872; Kappas et al., 2013). On the downside, however, this means that a lack of facial expressiveness can be a serious impediment to communication. For example, Parkinson's disease (PD) is characterized by hypomimia, and people with PD often experience reduced facial expressions (Sonawane and Sharma, 2021), as well as an impaired ability to recognize and discriminate between different facial expressions (Mattavelli et al., 2021). In fact, automated facial expression recognition may even be able to help diagnose PD (Jin et al., 2020).

However, we do not need to be afflicted by a condition such as PD to understand the negative impact of diminished or obscured facial expressions. In some situations, such as when talking on the phone, we may already be used to the absence of visual cues. In other instances, for example, when wearing a face mask, such as those widely used during the recent COVID-19 pandemic, listening to a speaker (Giovanelli et al., 2021) and recognizing their facial expressions may be impaired (Grahlow et al., 2022). Perhaps more importantly, even when the ability to discriminate between expressions remains, perceived interpersonal closeness and mimicry may be reduced (Kastendieck et al., 2022).

EMG-based AUR becomes particularly relevant when interacting through immersive devices, such as virtual reality (VR) headsets. Recent research on avatar-mediated virtual environments underscores that facial expressions may play a more critical role than bodily cues in fostering interpersonal attraction and liking (Oh Kruzic et al., 2020). However, VR headsets inherently obstruct half of the face, posing

[a] https://orcid.org/0000-0001-8992-5648
[b] https://orcid.org/0000-0002-8538-727X
[c] https://orcid.org/0000-0002-6850-9570
[d] https://orcid.org/0000-0002-9809-7028
[e] https://orcid.org/0000-0003-4094-3444

Küster, D., Rammohan, R. A., Liu, H., Schultz, T. and Koschke, R.
The Bigger the Better? Towards EMG-Based Single-Trial Action Unit Recognition of Subtle Expressions.
DOI: 10.5220/0013389300003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 1*, pages 100-110
ISBN: 978-989-758-731-3; ISSN: 2184-4305

a challenge for conventional video-based AUR systems (Wen et al., 2022). Potential approaches towards bridging this gap include the use of add-ons such as integrated eye-tracking and facial-tracking devices (Schuetz and Fiehler, 2022) or incorporating infrared light sources and cameras to develop visual databases for training visual AUR while users wear VR headsets (Chen and Chen, 2023). However, while these approaches may help to address the VR use case, EMG-based AUR could still outperform such approaches due to its superior sensitivity and time resolution (Veldanda et al., 2024).

Regardless of the camera type or additional tracking capabilities, vision-based AUR and facial expression analysis have relied on the Facial Action Coding System (FACS) since the 1970s (Ekman et al., 2002). FACS itself was upon earlier foundational work by (Hjortsjö, 1969), who cataloged facial configurations (Barrett et al., 2019) originally depicted in Duchenne's research (Duchenne and Cuthbertson, 1990). The system classifies facial expressions into 44 Action Units (AUs), each representing specific, independently controlled facial muscle movements. Unlike basic emotions (Ekman, 1999), AUs are purely descriptive and avoid interpretive labels (Zhi et al., 2020). FACS has therefore been the nearly universally accepted standard for behavioral research on facial expressions and 3D emotion modeling (van der Struijk et al., 2018). However, vision-based AUR using FACS faces several substantial challenges that could be addressed by EMG-based AUR.

1.1 EMG-Based Automatic Action Unit Recognition (AUR)

AUR aims to automatically identify the facial muscle movements associated with emotions, expressions, and communicative intentions (Crivelli and Fridlund, 2019) by analyzing action units such as nose wrinkling (AU9), eyebrow-raising (AU1, AU2), or lip corner pulling (AU12). Historically, AUR relied on labor-intensive manual annotation of videos by certified FACS experts, a process requiring over an hour to label just one minute of video (Bartlett et al., 2006; Zhi et al., 2020).

Today, advancements in automatic affect recognition have introduced tools ranging from early systems like the Computer Expression Recognition Toolbox (CERT) (Littlewort et al., 2011) to modern open-source software such as OpenFace (Baltrusaitis et al., 2018) and LibreFace (Chang et al., 2024), enabling cost-effective and efficient analysis of facial activity in research settings (Küster et al., 2020). While many tools have traditionally focused on detecting prototypical expressions tied to basic emotion theories (BETs) (Ortony, 2022; Crivelli and Fridlund, 2018), growing interest in facial AUR highlights its objectivity as a research tool independent of BET- or other theoretical frameworks (Küster et al., 2020).

However, efforts to evaluate and compare AUR platforms (Krumhuber et al., 2021) have often been constrained by a limited number of publicly available databases, such as those referenced in (Chang et al., 2024). As a result, performance estimates for these tools may be overly optimistic, particularly in spontaneous and noisy field recording conditions where accuracy tends to degrade (Krumhuber et al., 2021). A more in-depth analysis of raw movement data at the level of facial landmarks could help overcome these challenges and provide significant benefits for video-based AUR (Zinkernagel et al., 2019).

To date, AUR research has remained predominantly vision-based. Although camera-based AUR has demonstrated reliable accuracy under controlled recording conditions, advances in EMG-based methods for recording facial expressions have yet to be fully integrated into AUR research (Veldanda et al., 2024). This is despite a well-established body of emotion research utilizing facial EMG (fEMG) (Boxtel, 2001; Wingenbach, 2023; Tassinary et al., 2007) and the development of robust laboratory guidelines (Fridlund and Cacioppo, 1986; Tassinary et al., 2007) and placement schemes for high-resolution EMG (Guntinas-Lichius et al., 2023).

However, we argue that this state-of-the-art is beginning to change. Some recent work has examined the use of inertial measurement units (IMUs) for AUR, yielding promising early results (Verma et al., 2021). Other work has already integrated EMG electrodes into a VR-compatible device (Gjoreski et al., 2022). In our work, we have demonstrated encouraging pilot results, showing that EMG can provide reliable and real-time-capable data and models to classify four distinct AUs (Veldanda et al., 2024). In a similar approach, (Kołodziej et al., 2024) Similarly, (Kołodziej et al., 2024) used EMG to classify six discrete emotion categories, employing both a support vector machine (SVM) model and a *k*-nearest neighbor (KNN) classifier.

1.2 Methodological Challenges and Opportunities

EMG-based AUR offers a solution to several challenges that are difficult to overcome with camera-based AUR alone (Veldanda et al., 2024). On a technical level, camera-based AUR is influenced by factors such as the visibility of specific AUs, viewing angles, and the databases used for training and validation. Cross-database evaluations often rely on posed datasets, which may not reflect real-world conditions (Namba et al., 2021a; Namba et al., 2021b; Zhi et al., 2020).

Spontaneous facial expressions, while of greater interest to emotion researchers (Krumhuber et al., 2021), pose additional challenges. Spontaneous expressions are typically more subtle, dynamic, and complex, often involving co-occurring AUs (Veldanda et al., 2024). However, the greater variability inherent in spontaneous expressions makes it difficult for classifiers to accurately process less standardized data (Krumhuber et al., 2023; Zhi et al., 2020). On a conceptual level, facial expression research also deals with issues such as inconsistent emotion measurement and the interpretation of AUs within their physical and social contexts. In particular, there is often only poor agreement between physiological measures and self-reported emotional experiences (Kappas et al., 2013; Mauss and Robinson, 2009).

Advances in multimodal emotion recognition using machine learning appear promising but have rarely incorporated high-resolution facial EMG data, which could improve sensitivity compared to webcam-based methods (Schuller et al., 2012; Steinert et al., 2021). Here, EMG-based AUR could help to pave the way for a more robust and fine-grained study of facial expressions - in particular when studying facial expressions that are more spontaneous and subtle. However, facial electromyography as a method has always been limited with respect to the number of available electrodes as well as concerning the issue of cross-talk (van Boxtel et al., 1998; Tassinary et al., 2007). That is, when recording from only a small number of electrode positions, the source of the signal can be difficult to determine by conventional statistical measures, as neighboring muscle sites may produce a very similar, albeit weaker, signal than the targeted muscle site of interest. A simple and time-tested approach towards addressing this issue in the laboratory is to place electrodes on several different sites, and design experiments in such a way that there are clear predictions on which muscles should be activated - or to include an unobtrusive camera recording to exclude “noise” from unintended muscle activations. However, this latter approach effectively sacrifices much of the potential advantages of otherwise privacy-preserving EMG by introducing a camera for artifact checking. Furthermore, a camera-based correction is again limited to the visible signal, thus again voiding the inherent advantage of EMG to detect signals below the visible threshold.

One way to address this challenge is to increase the number of electrodes used. However, while camera technology has made significant strides in improving spatial resolution, high-density EMG recordings remain costly and constrained by the practical limits of electrode placement on the human face. In their recent effort to establish a high-resolution EMG recording scheme, (Guntinas-Lichius et al., 2023) utilized small, reusable pediatric surface electrodes with an Ag/AgCl disc diameter of just 4 mm. This enabled simultaneous bipolar recordings from 19 muscle positions to compare two different electrode placement schemes. However, while such a setup provides excellent coverage, it is likely to be impractical for most laboratories, which typically lack the resources for high-density EMG. Furthermore, the large number and sheer weight of the electrodes may hinder participants’ ability to perform facial expressions naturally. Therefore, a key goal for advancing EMG-based AUR is to harness machine learning to disambiguate signals using only a small number of electrodes. This would help identify the specific muscles responsible for a given AU while maintaining signal clarity. At the same time, we aim to build on the strengths of EMG to capture even subtle or invisible facial muscle activity.

1.3 The Present Work

In this paper, we aim to advance recent EMG-based AUR models to include automatic recognition of subtle facial expressions, which are characterized by a low intensity of the expression. As EMG has been the gold standard for the high-precision recording of facial expressions in the psychophysiological laboratory for decades (Fridlund and Cacioppo, 1986; Wingenbach, 2023), even a relatively small amount of data may be sufficient to train initial models. Additionally, we address the question of whether small and more lightweight electrodes may be more suitable for recording and building models for subtle expressions despite their smaller diameters. We, therefore, aim to examine a custom-built variant of the popular mobile Biosignal Plux EMG sensor to facilitate placement of electrodes at the distances that allow a closer and more accurate placement (Fridlund and Cacioppo, 1986) correspond to the requirements of established guidelines. While the vast majority of AUR research to date has been conducted on video data, our research aims to leverage EMG to pave the ground for a growing number of privacy-preserving AUR use cases.

To examine these questions, we use a newly recorded dataset of fEMG sensor data to predict a subset of eight AUs in both high and low expression intensity, as well as neutral, yielding a total of 17 distinct classes. The current work thus extends upon our recent work studying peak expression intensities of four AUs (Veldanda et al., 2024).

2 METHODOLOGY

To evaluate the performance of the two electrodes, we propose a framework that utilizes fEMG data synchronized with video recordings of facial expressions at both high and low intensities from four participants. We extract time-series features using the Time-Series Feature Extraction Library (TSFEL) (Barandas et al., 2020) to train a set of standard machine-learning models (RF, SVM, GNB, KNN). The best-performing model is then selected for further analysis.

2.1 Data Collection

The framework for fEMG dataset collection, including the synchronization with concurrent video recordings was adapted from the approach used in the study by Veldanda and colleagues (Veldanda et al., 2024). In this current study, data were recorded in single trials for each type and intensity of AU. Four participants (three female, one male) were recruited, with a mean age of 28.25 years (*SD* = 2.98). The task involved imitating facial expressions presented as stimulus videos via a customized graphical user interface (GUI) as in the Figure 1. The stimulus videos were sourced from the MPI Video Database (Kleiner et al., 2004), which provides accurate portrayals of AU activations.

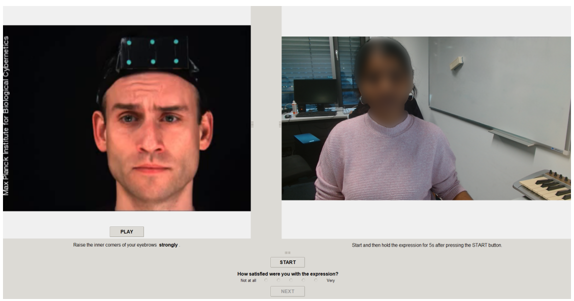

Figure 1: Graphical User Interface (GUI) for the data collection.

One of our primary objectives was to detect subtle facial expressions. To this end, participants were instructed to first hold the target facial expressions at a maximum (high) intensity and then repeat the same expression at a subtle (low) intensity. Both, the fEMG signals and corresponding video recordings were captured for a duration of 5 seconds for each expression to obtain short data segments featuring the same target expression and intensity.

The recording setup was adapted from the study by Veldanda and colleagues (Veldanda et al., 2024), with a desktop PC to display stimuli, a webcam, and an fEMG acquisition system. A bipolar recording configuration was used, comprising three channels covering upper facial regions (Lateral Frontalis, Corrugator Supercilii, Medial Frontalis) and three additional channels covering lower facial regions (Zygomaticus Major, Levator Labii Superioris, Mentalis). In total, we considered nine AUs, which additionally include neutral expressions, as listed in Table 1.

Table 1: Selected actions units for the pilot study.

Action Unit	Action
AU1	Inner Brow Raiser
AU2	Outer Brow Raiser
AU4	Brow Lowerer
AU9	Nose Wrinkler
AU12	Lip Corner Puller
AU17	Chin Raiser
AU20	Lip stretcher
AU24	Lip Pressor
AU0	Neutral Expression

Another important objective of this study was to compare the performance of two types of electrodes in recognizing the action units. The original EMG sensors (PLUX Biosignals[1]), with a diameter of 24 mm (in Figure 2a) and a hub, were used as part of our recording setup. In addition, a modified version of these sensors with lightweight adapters was employed, allowing the use of small Ag/AgCl electrodes with a diameter of only 5 mm (in Figure 2b). The smaller size allows electrode placement according to the guidelines of the Society for Psychophysiological Research (Fridlund and Cacioppo, 1986), which recommend maintaining the center-to-center distance between electrodes within 1 cm. Notably, this configuration ensured that both types of electrodes could be compared with the same settings, software, and amplifiers.

During data collection, one type of electrode (big, small) was placed in the upper region of the face and the other type on the lower region, respectively, as in Figure 3. At the end of a session, the electrode positions were swapped for the next session. To ensure a balanced design, the sequence of electrode placements was counterbalanced across the four partici-

[1] www.pluxbiosignals.com

(a) (b)

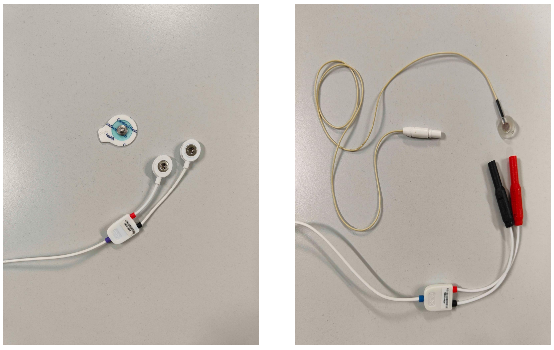

Figure 2: Original EMG sensor with 24mm diameter snap-on EMG electrodes (a), and modified EMG sensor with 5mm diameter Ag/AgCl EMG electrode (b).

Figure 3: Electrode placement.

pants.

The video recordings and EMG data were synchronized using the Lab Streaming Layer (LSL) protocol. The sampling frequency of the fEMG signals was set to 2,000 Hz. Facial expression segments from the EMG signals were extracted based on video timestamps, ensuring proper alignment between the modalities.

2.2 Feature Extraction

Traditionally, time-series data are filtered to remove noise before extracting domain-specific features for machine learning classification. This process can be both complex and time-consuming. However, the Time-Series Feature Extraction Library (TSFEL) (Barandas et al., 2020) provides a comprehensive, automated pipeline for efficient feature extraction across multiple domains, including temporal, statistical, and spectral feature sets. In our study, we segmented the raw EMG data into 100 ms windows with a 20% overlap and utilized all the feature sets provided by TSFEL.

2.3 Classification

We evaluated the performance of both electrode types for recognizing high and low-intensity AUs using the following popular machine learning classifiers:

1. Random Forest (RF)
2. Support Vector Machine (SVM)
3. Gaussian Naive Bayes (GNB)
4. *k*-Nearest Neighbors (KNN)

All models were obtained from the *scikit-learn* library (Pedregosa et al., 2011) and employed with their default hyperparameters. Training was carried out using the time-series features extracted from the TSEFL library.

3 RESULTS

3.1 Comparison of Classification Models

The machine learning models were trained to recognize all AUs for both types of electrodes. Their performance was evaluated based on overall accuracy, using 4-fold leave-one-out cross-validation (LOOCV) approach, in which data from three participants formed the training set in each fold. The mean accuracies are presented in Table 2.

Table 2: Comparison of the performance of machine learning models.

Model	Mean accuracy
RF	0.38
SVM	0.32
GNB	0.36
KNN	0.25

The overall decrease in performance across all models can be attributed to the short data segments represented by the small number of trials. Nevertheless, consistent with prior work (Veldanda et al., 2024), the Random Forest classifier performed best. As indicated by the confusion matrix of the RF in Figure 4, patterns of interest do emerge, prompting

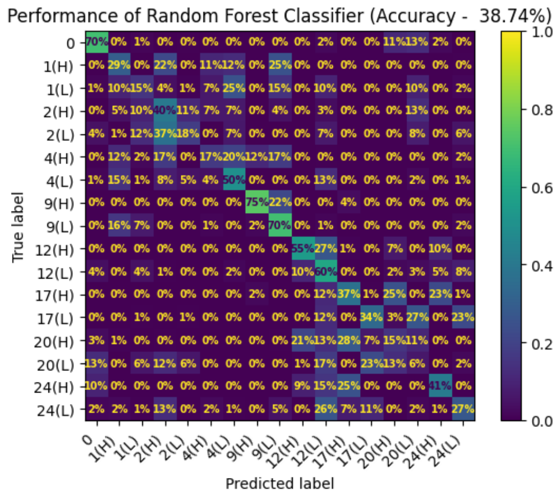

Figure 4: Confusion matrix illustrating classification performance for all action units across both electrode types. The labels in the confusion matrix indicate action units and their intensities: H = High, L = Low.

further investigation in subsequent analyses. As expected, this overall recognition performance is substantially lower than previous work examining five classes of AUs (Veldanda et al., 2024). However, the results are encouraging when considering the vastly greater number of 17 classes, the inclusion of subtle expressions, and the much smaller amount of training data per subject.

3.2 Impact of Electrode Size and Expression Intensity

To gain more insight into the observed differences, we performed a cell-wise, two-tailed two-proportion Z-test on the confusion matrices generated for different electrode sizes and intensities. Since multiple statistical comparisons were carried out, we applied a Bonferroni correction to each of the 81 individual comparisons to reduce the likelihood of Type I errors.

Figure 5 illustrates the differences in proportions associated with electrode size and expression intensity. Asterisks indicate significant differences at $p < .0001$. As illustrated by the results of the comparison between big and small electrodes (panel a), models based on the laboratory-grade 5 mm small electrodes overall performed dramatically better (76 %) than the big snap-on electrodes for correctly detecting the absence of an AU. Here, models run on the data for the big electrodes more often falsely predicted the presence of another expression, such as a movement of the lip stretcher (AU20). When considering the overall comparison between high- and low-intensity expressions (panel b), a complex pattern of confusions was observed, in particular with respect to different types of eyebrow movements. Here, e.g., a low intensity of lowering the eyebrows (AU4) was significantly better recognized than the high-intensity version of the same AU, whereas the opposite pattern was observed for raising the outer eyebrow. Considering the counteractive nature of both of these muscles, this pattern of results appears less surprising. Nevertheless, considering the complexity of these confusion patterns, we decided to further split the data by conditions to examine whether more systematic performance differences could be found.

Figure 6 shows the corresponding confusion matrices for the split of the data by electrode size and expression intensity of the AUs, which can be regarded as a 2x2 factorial design. Again, an RF classifier was trained and tested on these four conditions, and the resulting confusion matrices were analyzed using a chi-square test of independence. Before the analysis, each confusion matrix was treated as a contingency table, and any columns with zero totals were removed.

We found a significant performance advantage of 7.4% for using the bigger snap-on electrodes when classifying high-intensity expressions, $(\chi^2(61) = 1453.12, p < .0001)$, as well as a significant advantage of 5.81% for the smaller electrodes compared to big electrodes when classifying low-intensity expressions, $(\chi^2(76) = 2691.09, p < .0001)$. Simultaneously, models on the data from big electrodes performed significantly better for high vs. low-intensity expressions, yielding 8.66% better recognition performance for high-intensity expressions, $(\chi^2(67) = 2113.65, p < .0001)$. Finally, models on small electrodes performed significantly better on low-intensity expressions than high-intensity expressions, with a 4.55% increment for low-intensity AUs over high-intensity AUs, $(\chi^2(72) = 2364.60, p < .0001)$. This pattern of results appears to correspond to a disordinal (crossed) interaction effect, wherein both types of electrodes showed substantial performance gains for these two different types of expressions. These results suggest that small laboratory electrodes may be more suitable for subtle expressions, whereas the bigger snap-on electrodes may be able to more robustly detect peak intensity expressions.

4 DISCUSSION

The present results suggest that EMG-based AUR may be suitable for detecting a large number of different AUs - even with relatively little training data and a default RF baseline model. Notably, however, electrode positions in the lower and upper face

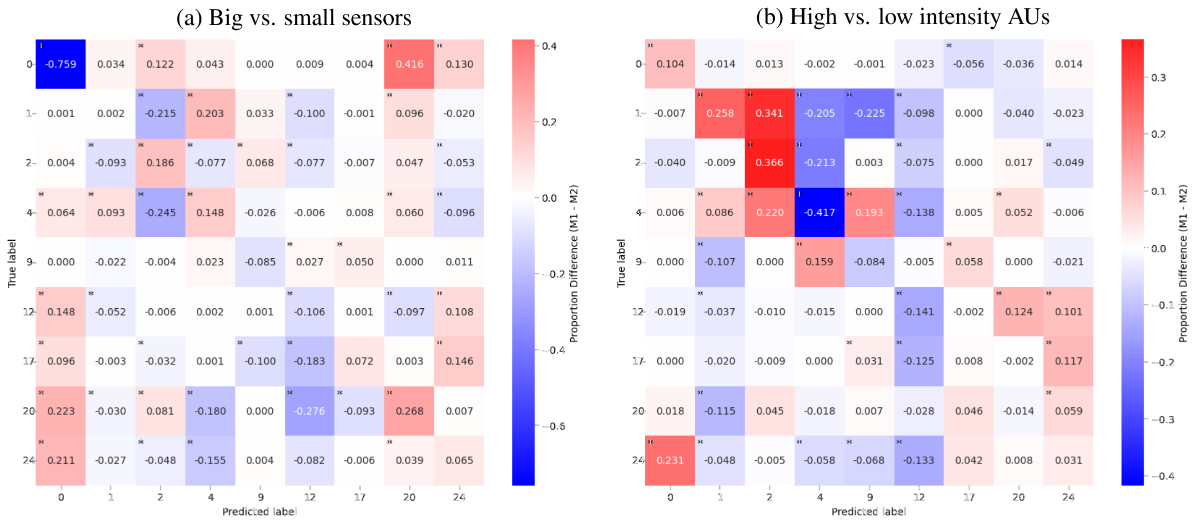

Figure 5: Comparison of differences in proportions for sensor size (a) and AU intensity (b). Red indicates greater proportions for big electrodes (a) or high intensity (b). Blue indicates greater proportions for small electrodes (a) or low intensity (b).

showed patterns of confusions suggesting that the models faced substantial challenges in distinguishing AUs that are physically close to each other. E.g., AU1, AU2, and AU4 all describe different types of eyebrow-related movements, whereas AU12, AU17, AU20, and AU24 all involve movements around the mouth region. Perhaps unsurprisingly, these two clusters of AUs showed a lot of confusions amongst the respective AUs, as these signals are likely to have involved substantial amounts of cross-talk. In contrast, the nose wrinkler (AU9), which is generally a relatively difficult AU to produce for laypeople, was recognized exceptionally well. Here, we speculate that this may have been the case because AU9 is sufficiently independent from both clusters, while still close enough to at least two of the electrode pairs to receive valid signals.

Another key finding of this work is that the two different electrode types appeared to suit different expression intensities. Here, the larger recording surface of the original single-use electrodes may be better able to differentiate the relative intensity of large muscle contractions at nearby sites. Conversely, the smaller electrodes may have allowed subjects to retain a better “feeling” for very fine-grained intensity differences, with less cross-talk - whereas moving muscles underneath the bigger electrodes could have required more effort and, possibly, more unwanted co-activation of neighboring muscle sites. This interpretation appears to be supported by the larger number of erroneous “neutral” labels for low-intensity expressions recorded by the big electrodes.

While the present results are encouraging, some limitations remain for the current pilot data set. First, the present study was still based on a very small number of participants, who performed the minimum number of expressions to train the present initial machine learning models. Here, we are presently collecting a more substantial data set with several repetitions of each of the 17 different AU classes examined in the present work. We expect that this expanded data set will provide a basis for better-performing models than the current baseline. Second, we have not yet conducted a formal statistical test of the apparent interaction between electrode type and AUR performance for low vs. high-intensity expressions. Here, we had expected a more clear-cut decision for one or the other type of electrodes, and we regard the apparent interaction between both factors as an exploratory finding at this stage. In our future work with a larger dataset, we plan to submit this hypothesis to a robust generalized linear mixed model test, with the subject as a random factor. Third, several different approaches could still be attempted to improve and further analyze the current model results. However, the RF classification has consistently emerged as the best model already in our previous work, and this study has only aimed to provide initial results for a proof of concept for EMG-based AUR for low-intensity expressions as well as the comparison of small laboratory and big snap-on electrodes with the same base system.

5 CONCLUSION

The present results are consistent with the notion that surface EMG is capable of detecting even very subtle muscle activity for EMG-based AUR - and that

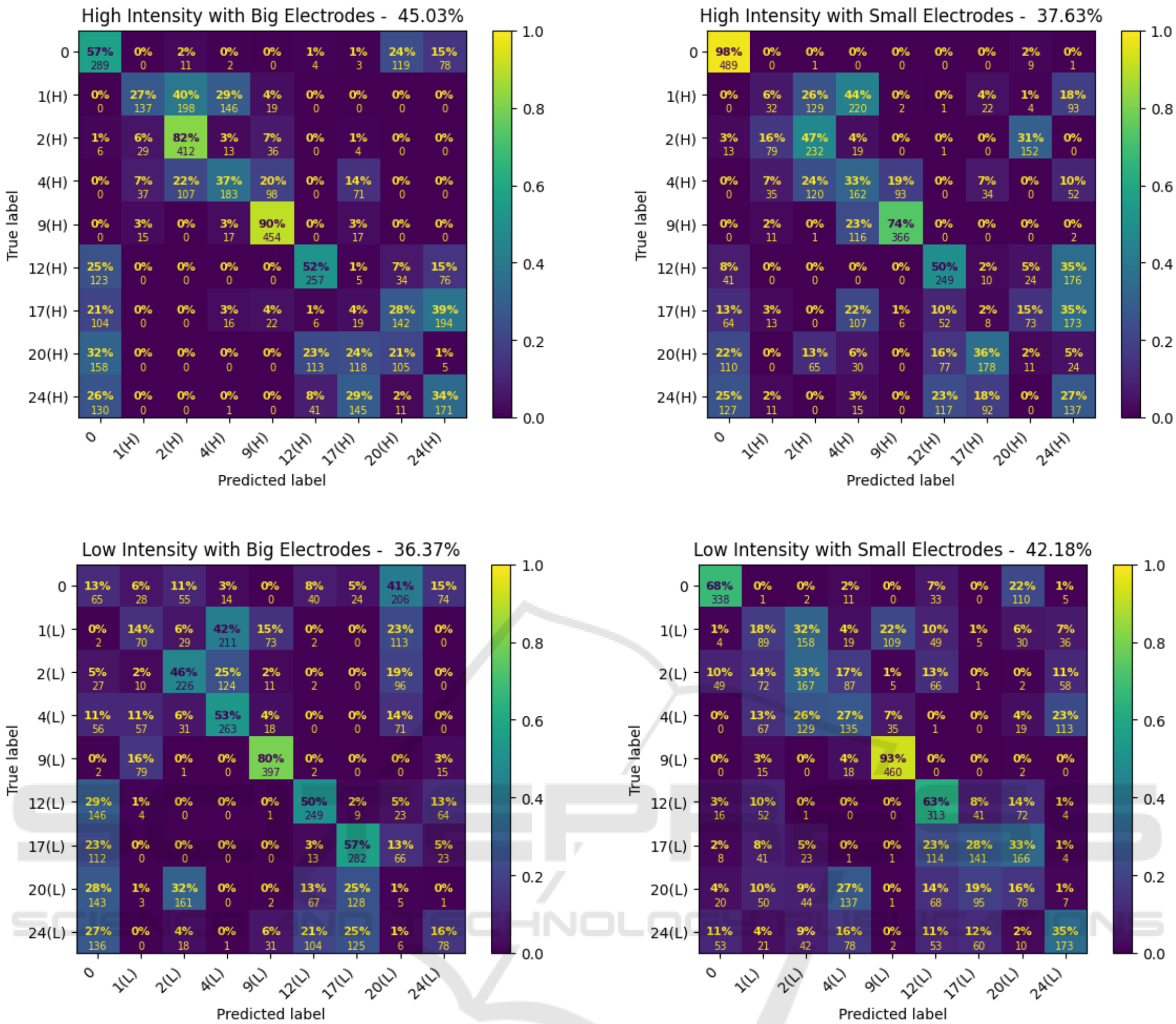

Figure 6: Confusion matrices under the four conditions.

with little relative degradation in performance compared to peak intensity expressions. To the best of our knowledge, this is the first study to have successfully detected low-intensity expressions via EMG-based AUR. Intriguingly, different types of electrodes may be more suitable for different use cases - even if they are attached to the same base amplifiers. This is consistent with findings from previous studies that have demonstrated that the control of human facial muscles is a complex process (Cattaneo and Pavesi, 2014), which is influenced by substantial anatomical variations (D'Andrea and Barbaix, 2006) as well as differences in signal strength across muscle regions (Schultz et al., 2019).

In future work, we aim to extend the current evaluation with further electrode types, while also varying the targeted electrode placement. Notably, the traditional placement guidelines (Fridlund and Cacioppo, 1986) were designed almost 40 years ago, with the purpose of better comparability of studies for statistical analyses across laboratories. When considering the relatively recent advent of advanced machine learning methods and current work involving high-resolution facial EMG (Guntinas-Lichius et al., 2023), this raises the question if there could be a more fine-grained adaptation of effective electrode placements for individual subjects. Indeed, while we had expected to see a more clearcut advantage of the more accurately placed smaller electrodes, our present results suggest that the optimal electrode type- and placement for the training EMG-based AUR systems may differ from the original guidelines that aimed to optimize comparability of mean activity between muscle recording sites. Considering the limited sample size of the present study, there is a clear need for further validation with a larger sample size and a substantially greater number of trials for each AU. This would allow conducting more robust hypothesis-guided statistical tests, in particular with regard to the present exploratory finding of an apparent interaction between sensor size and expression intensity on AUR performance.

Recording schemes for training machine learning models might benefit more from signals that are correlated with a particular AU, while simultaneously being as distinctive as possible from signals from other AUs. That is, instead of maximizing the mean signal strength at a recording site, EMG-based AUR may benefit from a somewhat more distal and individualized electrode placement. Here, another potential application on the horizon for real-time EMG-based AUR systems could be the development of automated placement guidance for subject-tailored optimal placement of recording electrodes. Finally, a more distal electrode placement would likewise be a requirement for the development of EMG-based AUR devices, e.g., for applications in VR, since current prototypes with inbuilt electrodes (Gjoreski et al., 2022) may still be too expensive and unwieldy for the majority of potential applications. Depending on the use case, the performance of AUR under laboratory conditions could just be a starting point. For instance, Ag/AgCl electrodes may oxidize over time, prompting considerations about whether electrodes in end-user devices should be cleaned or replaced. Together, these findings call for more research into EMG-based AUR, with the ultimate aim of building biosignals adaptive cognitive systems (Schultz and Maedche, 2023) that are designed to provide privacy-preserving AUR-capabilities across a broad range of fields for applications, from the diagnosis of Parkinson's disease to immersive avatar-mediated communication in VR.

ACKNOWLEDGEMENTS

This research work was funded by a grant of the Minds, Media, Machines High Profile Research Area at the University of Bremen (MMM-Seed Grant No.005). We furthermore gratefully acknowledge the contributions of our students, Romina Razeghi Oskouei and Ferdinand Rohlfing in conducting the study.

REFERENCES

Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L.-P. (2018). OpenFace 2.0: Facial Behavior Analysis Toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 59–66, Xi'an. IEEE.

Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., and Gamboa, H. (2020). TSFEL: Time Series Feature Extraction Library. *SoftwareX*, 11:100456.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., and Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 20(1):1–68.

Bartlett, M. S., Littlewort, G. C., Frank, M. G., Lainscsek, C., Fasel, I. R., and Movellan, J. R. (2006). Automatic Recognition of Facial Actions in Spontaneous Expressions. *Journal of Multimedia*, 1(6):22–35.

Boxtel, A. (2001). Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles. *Psychophysiology*, 38(1):22–34.

Cattaneo, L. and Pavesi, G. (2014). The facial motor system. *Neuroscience & Biobehavioral Reviews*, 38:135–159.

Chang, D., Yin, Y., Li, Z., Tran, M., and Soleymani, M. (2024). LibreFace: An Open-Source Toolkit for Deep Facial Expression Analysis. pages 8205–8215.

Chen, X. and Chen, H. (2023). Emotion recognition using facial expressions in an immersive virtual reality application. *Virtual Reality*, 27(3):1717–1732.

Crivelli, C. and Fridlund, A. J. (2018). Facial Displays Are Tools for Social Influence. *Trends in Cognitive Sciences*, 22(5):388–399. Publisher: Elsevier.

Crivelli, C. and Fridlund, A. J. (2019). Inside-Out: From Basic Emotions Theory to the Behavioral Ecology View. *Journal of Nonverbal Behavior*, 43(2):161–194.

Darwin, C. (1872). *The expression of the emotions in man and animals*. Oxford University Press, New York, 1998 ed. edition.

Duchenne, G.-B. and Cuthbertson, R. A. (1990). *The mechanism of human facial expression*. Studies in emotion and social interaction. Cambridge University Press ; Editions de la Maison des Sciences de l'Homme, Cambridge [England] ; New York : Paris.

D'Andrea, E. and Barbaix, E. (2006). Anatomic research on the perioral muscles, functional matrix of the maxillary and mandibular bones. *Surgical and Radiologic Anatomy*, 28(3):261–266.

Ekman, P. (1999). Basic emotions. In *Handbook of cognition and emotion*, pages 45–60. John Wiley & Sons Ltd, Hoboken, NJ, US.

Ekman, P., Friesen, W. V., and Hager, J. C. (2002). *Facial action coding system: the manual*. Research Nexus, Salt Lake City, Utah.

Fridlund, A. J. and Cacioppo, J. T. (1986). Guidelines for Human Electromyographic Research. *Psychophysiology*, 23(5):567–589.

Giovanelli, E., Valzolgher, C., Gessa, E., Todeschini, M., and Pavani, F. (2021). Unmasking the difficulty of listening to talkers with masks: lessons from the covid-19 pandemic. *i-Perception*, 12(2):2041669521998393.

Gjoreski, M., Kiprijanovska, I., Stankoski, S., Mavridou, I., Broulidakis, M. J., Gjoreski, H., and Nduka, C. (2022). Facial EMG sensing for monitoring affect using a wearable device. *Scientific Reports*, 12(1):16876. Publisher: Nature Publishing Group.

Grahlow, M., Rupp, C. I., and Derntl, B. (2022). The impact of face masks on emotion recognition performance and perception of threat. 17(2):e0262840. Publisher: Public Library of Science.

Guntinas-Lichius, O., Trentzsch, V., Mueller, N., Heinrich, M., Kuttenreich, A.-M., Dobel, C., Volk, G. F., Graßme, R., and Anders, C. (2023). High-resolution surface electromyographic activities of facial muscles during the six basic emotional expressions in healthy adults: a prospective observational study. *Scientific Reports*, 13(1):19214. Publisher: Nature Publishing Group.

Hjortsjö, C.-H. (1969). *Man's Face and Mimic Language*. Studentlitteratur, Lund, Sweden.

Jin, B., Qu, Y., Zhang, L., and Gao, Z. (2020). Diagnosing parkinson disease through facial expression recognition: Video analysis. 22(7):e18697. Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

Kappas, A., Krumhuber, E., and Küster, D. (2013). *Facial behavior*, pages 131–165.

Kastendieck, T., Zillmer, S., and Hess, U. (2022). (un)mask yourself! effects of face masks on facial mimicry and emotion perception during the COVID-19 pandemic. 36(1):59–69. Publisher: Routledge _eprint: https://doi.org/10.1080/02699931.2021.1950639.

Kleiner, M., Wallraven, C., Breidt, M., Cunningham, D. W., and Bülthoff, H. H. (2004). Multi-viewpoint video capture for facial perception research. In *Workshop on Modelling and Motion Capture Techniques for Virtual Environments (CAPTECH 2004)*, Geneva, Switzerland.

Kołodziej, M., Majkowski, A., and Jurczak, M. (2024). Acquisition and Analysis of Facial Electromyographic Signals for Emotion Recognition. *Sensors*, 24(15):4785. Number: 15 Publisher: Multidisciplinary Digital Publishing Institute.

Krumhuber, E. G., Küster, D., Namba, S., and Skora, L. (2021). Human and machine validation of 14 databases of dynamic facial expressions. *Behavior Research Methods*, 53(2):686–701.

Krumhuber, E. G., Skora, L. I., Hill, H. C. H., and Lander, K. (2023). The role of facial movements in emotion recognition. *Nature Reviews Psychology*, 2(5):283–296.

Küster, D., Krumhuber, E. G., Steinert, L., Ahuja, A., Baker, M., and Schultz, T. (2020). Opportunities and challenges for using automatic human affect analysis in consumer research. *Frontiers in neuroscience*, 14:400.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., and Bartlett, M. (2011). The computer expression recognition toolbox (CERT). In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pages 298–305.

Mattavelli, G., Barvas, E., Longo, C., Zappini, F., Ottaviani, D., Malaguti, M. C., Pellegrini, M., and Papagno, C. (2021). Facial expressions recognition and discrimination in parkinson's disease. 15(1):46–68.

Mauss, I. B. and Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & Emotion*, 23(2):209–237.

Namba, S., Sato, W., Osumi, M., and Shimokawa, K. (2021a). Assessing Automated Facial Action Unit Detection Systems for Analyzing Cross-Domain Facial Expression Databases. *Sensors*, 21(12):4222.

Namba, S., Sato, W., and Yoshikawa, S. (2021b). Viewpoint Robustness of Automated Facial Action Unit Detection Systems. *Applied Sciences*, 11(23):11171.

Oh Kruzic, C., Kruzic, D., Herrera, F., and Bailenson, J. (2020). Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments. *Scientific Reports*, 10(1):20626.

Ortony, A. (2022). Are All "Basic Emotions" Emotions? A Problem for the (Basic) Emotions Construct. *Perspectives on Psychological Science*, 17(1):41–61.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Schuetz, I. and Fiehler, K. (2022). Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability. *Journal of Eye Movement Research*, 15(3). Number: 3.

Schuller, B., Valster, M., Eyben, F., Cowie, R., and Pantic, M. (2012). AVEC 2012: the continuous audio/visual emotion challenge. In *Proceedings of the 14th ACM international conference on Multimodal interaction*, ICMI '12, pages 449–456, New York, NY, USA. Association for Computing Machinery.

Schultz, T., Angrick, M., Diener, L., Küster, D., Meier, M., Krusienski, D. J., Herff, C., and Brumberg, J. S. (2019). Towards restoration of articulatory movements: Functional electrical stimulation of orofacial muscles. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3111–3114.

Schultz, T. and Maedche, A. (2023). Biosignals meet Adaptive Systems. *SN Applied Sciences*, 5(9):234.

Sonawane, B. and Sharma, P. (2021). Review of automated emotion-based quantification of facial expression in parkinson's patients. 37(5):1151–1167.

Steinert, L., Putze, F., Küster, D., and Schultz, T. (2021). Audio-visual recognition of emotional engagement of people with dementia. In *Interspeech*, pages 1024–1028.

Tassinary, L. G., Cacioppo, J. T., and Vanman, E. J. (2007). The Skeletomotor System: Surface Electromyography. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G., editors, *Handbook of Psychophysiology*, pages 267–300. Cambridge University Press, Cambridge, 3 edition.

van Boxtel, A., Boelhouwer, A., and Bos, A. (1998). Optimal EMG signal bandwidth and interelectrode distance for the recording of acoustic, electrocutaneous, and photic blink reflexes. *Psychophysiology*, 35(6):690–697. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1469-8986.3560690.

van der Struijk, S., Huang, H.-H., Mirzaei, M. S., and Nishida, T. (2018). FACSvatar: An Open Source Modular Framework for Real-Time FACS based Facial Animation. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, IVA '18, pages 159–164, New York, NY, USA. Association for Computing Machinery.

Veldanda, A., Liu, H., Koschke, R., Schultz, T., and Küster, D. (2024). Can electromyography alone reveal facial action units? a pilot emg-based action unit recognition study with real-time validation:. In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies*, page 142–151, Rome, Italy. SCITEPRESS - Science and Technology Publications.

Verma, D., Bhalla, S., Sahnan, D., Shukla, J., and Parnami, A. (2021). ExpressEar: Sensing Fine-Grained Facial Expressions with Earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3):1–28.

Wen, L., Zhou, J., Huang, W., and Chen, F. (2022). A Survey of Facial Capture for Virtual Reality. *IEEE Access*, 10:6042–6052. Conference Name: IEEE Access.

Wingenbach, T. S. H. (2023). Facial EMG – Investigating the Interplay of Facial Muscles and Emotions. In Boggio, P. S., Wingenbach, T. S. H., da Silveira Coêlho, M. L., Comfort, W. E., Murrins Marques, L., and Alves, M. V. C., editors, *Social and Affective Neuroscience of Everyday Human Interaction: From Theory to Methodology*, pages 283–300. Springer International Publishing, Cham.

Zhi, R., Liu, M., and Zhang, D. (2020). A comprehensive survey on automatic facial action unit analysis. *The Visual Computer*, 36(5):1067–1093.

Zinkernagel, A., Alexandrowicz, R. W., Lischetzke, T., and Schmitt, M. (2019). The blenderFace method: video-based measurement of raw movement data during facial expressions of emotion using open-source software. *Behavior Research Methods*, 51(2):747–768.