Can Electromyography Alone Reveal Facial Action Units? A Pilot EMG-Based Action Unit Recognition Study with Real-Time Validation

Abhinav Veldanda[a], Hui Liu[b], Rainer Koschke[c], Tanja Schultz[d] and Dennis Küster[e]
Cognitive Systems Lab, University of Bremen, Germany

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

Abstract: Facial expressions play a crucial role in non-verbal and visual communication, often observed in everyday life. The facial action coding system (FACS) is a prominent framework for categorizing facial expressions as action units (AUs), which reflect the activity of facial muscles. This paper presents a proof-of-concept study for upper face action unit recognition (AUR) using electromyography (EMG) data. The study recorded facial EMG data of a subject over four sessions, who imitated facial expressions corresponding to four different AUs. The subject-dependent models that were trained achieved high accuracy in near-real time and were able to classify AUs not directly underneath the recording sites.

1 INTRODUCTION

A large part of human communication is believed to be nonverbal and visual in nature, with facial expressions playing a key role (Kappas et al., 2013). We may notice this in everyday life, when we cannot see someone's face (e.g., on the phone), or when facial expressions are partially obscured – for example, due to a face mask (Giovanelli et al., 2021), or when we interact with someone wearing a virtual reality (VR) headset (Oh Kruzic et al., 2020).

Considerable work has been done on facial expression analysis since the early 1970s. Perhaps most prominent among these is the Facial Action Coding System (FACS), which was developed by Paul Ekman and Wallace Friesen (Ekman et al., 2002), and based on prior work by (Hjortsjö, 1969), a Swedish anatomist who had catalogued the facial configurations (Barrett et al., 2019) depicted by Duchenne (Duchenne and Cuthbertson, 1990). FACS provides a framework for categorizing all possible facial expressions into constituent action units (AUs), which reflect the activity of facial muscles that can be controlled independently. In contrast to discrete or "basic emotions" (Ekman, 1999), AUs are purely descriptive for movements of certain muscles, and do not provide any inferential labels (Zhi et al., 2020). Therefore, accurate tracking of AUs provides an objective basis for behavioral research into facial emotional expressions, as well as for 3D-modelling of emotions (van der Struijk et al., 2018). In total, the FACS (Ekman et al., 2002) provides coding instructions for 44 AUs.

1.1 Automatic Action Unit Recogntion

Action unit recognition (AUR) is an important research direction within facial expression analysis, which aims to automatically identify the activation of AUs that correspond to specific emotions, expressions, and actions. This approach analyzes the dynamics of subtle changes in the face, such as wrinkling of the nose, raising of the eyebrows, or lip corner pulling. For decades, AUR had to rely exclusively on costly and time-consuming manual recognition by certified FACS experts, with a ratio of more than one hour to manually label one minute of video data (Bartlett et al., 2006; Zhi et al., 2020). Today, automatic affect recognition tools allow for a much more cost-effective consideration of facial activity in most experimental research paradigms (Küster et al., 2020). From early classifiers, such as the Computer

[a] https://orcid.org/0009-0007-3749-4971
[b] https://orcid.org/0000-0002-6850-9570
[c] https://orcid.org/0000-0003-4094-3444
[d] https://orcid.org/0000-0002-9809-7028
[e] https://orcid.org/0000-0001-8992-5648

Veldanda, A., Liu, H., Koschke, R., Schultz, T. and Küster, D.
Can Electromyography Alone Reveal Facial Action Units? A Pilot EMG-Based Action Unit Recognition Study with Real-Time Validation.
DOI: 10.5220/0012399100003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 1*, pages 142-151
ISBN: 978-989-758-688-0; ISSN: 2184-4305

Expression Recognition Toolbox (Littlewort et al., 2011)) to current open-source tools, e.g., OpenFace (Baltrusaitis et al., 2018), researchers can now rely on a wide range of out-of-the-box software for camera-based automatic affect recognition. Due to the popularity of basic emotion theories (BETs) (Ortony, 2022), many of these tools have traditionally aimed to distinguish prototypical patterns of expressions believed to reflect discrete emotional states such as happiness, anger, or sadness (Dupré et al., 2020). More recently, reliable assessment and validation of facial AUs has been gaining more attention because they can be measured objectively without requiring the lens of BET (Küster et al., 2020).

Although some work has previously tested their own database without comparative evaluations between different platforms (Krumhuber et al., 2021), machine learning (ML) models for camera-based AUR have been the focus of a number of recent challenges in facial expression recognition and analysis (Zhi et al., 2020). Additionally, a few works have studied the performance of freely available pre-trained AUR systems such as OpenFace (Namba et al., 2021a; Namba et al., 2021b; Lewinski et al., 2014). Overall, these works have demonstrated the usefulness and reliability of camera-based AUR. However, there still remain methodological and conceptual challenges, including the nearly exclusive reliance of facial AUR on visual data. Here, EMG research and other recent approaches such as the use of inertial measurement units (IMUs) (Verma et al., 2021) may contribute towards improving the convergent validity of ML-based AUR.

1.2 Methodological Challenges

Perhaps unsurprisingly, camera-based AUR performance still varies depending on factors such as the specific AU (Namba et al., 2021a), viewing angle (Namba et al., 2021b), and database (Zhi et al., 2020) in question. Cross-database evaluations and challenges for discrete and AU-based affect recognition have also generally been based on a limited number of well-known databases of mostly posed expressions (Küster et al., 2020; Zhi et al., 2020). Compared to well-controlled posed datasets, spontaneous facial expressions in the wild are likely to be more subtle (Zhi et al., 2020) and involve more complex dynamics (Krumhuber et al., 2023), as well as other cues such as head movements (e.g., nodding) (Zhi et al., 2020). Spontaneous facial behavior also includes the possibility of co-occurring AUs, e.g., smiling with the eyes and the mouth, which can potentially yield thousands of distinct classes (Zhi et al., 2020). Together, these considerations raise the question of how well the said classifiers will perform for completely new and less standardized data. Finally, including a camera may sometimes interfere with the phenomenon to be measured. For example, the feeling of being observed has been shown to eliminate facial feedback phenomena that were once believed to be robust and well-established (Noah et al., 2018). These factors still pose significant challenges to the vibrant field of camera-based AUR.

1.3 Conceptual Challenges

Conceptually, facial expression research still faces substantial challenges relating to a lack of cohesion between measures of emotion (Kappas et al., 2013), as well as the interpretation of AUs as part of their physical and social context (Kuester and Kappas, 2013).

As demonstrated by earlier reviews, agreement between physiological measures of emotion and subjective self-report has often been surprisingly low (Mauss and Robinson, 2009). Furthermore, while theories of emotion have generally assumed biosignals, cognitive, and behavioral components of emotions to be synchronized and/or coordinated, empirical data has repeatedly challenged notions of strong concordance (Hollenstein and Lanteigne, 2014). Here, novel approaches in ML combining different modalities may yield more stable predictions than previous psychological models, as well as eventually provide some further insights into the ways in which the different components of the emotional response may synchronize and relate to each other. In consequence, leveraging easily obtainable data, such as jointly recorded audio-visual emotional responses together has been a core aim of a series of multimodal emotion recognition challenges for over a decade (Schuller et al., 2012). More recently, such approaches have proven to be fruitful across a wide range of subject areas, e.g., recognition of emotional engagement of people suffering from dementia (Steinert et al., 2021). Perhaps surprisingly, however, fEMG has thus far rarely been included in such approaches.

Apart from the practical challenges of recording fEMG as a high-quality and high-resolution signal of facial activity, a second major conceptual challenge relates to the interpretation of facial muscle activity beyond a FACS-based categorization. Here, an increasing number of works have demonstrated that notions such as Ekman's "basic emotions" (Ekman, 1999) may no longer be tenable (Ortony, 2022; Crivelli and Fridlund, 2018; Crivelli and Fridlund, 2019).

However, while we are aware of this ongoing debate, the present work is focused on a methodological contribution. I.e., by demonstrating the possibility of a reliable automatic recognition of facial AUs from EMG, we aim to help pave the way towards providing a more sensitive and high-resolution measure of facial activity compared to the now commonly used webcam data.

1.4 Facial Electromyography for Automatic Action Unit Recognition

In this paper, we solely focus on the use of facial electromyography (fEMG) as our basis for AUR. While most research on facial expressions today has been conducted on the basis of video data or mainly video supplemented by electromyography (EMG) data. Nevertheless, the use of fEMG has been the true gold standard for the high-precision recording of facial expressions in the psychophysiological laboratory for decades (Fridlund and Cacioppo, 1986; Wingenbach, 2023). In particular, facial surface EMG is capable of detecting very subtle muscle activity, including muscle relaxation (e.g., of the eyebrows), below what would be observable with the naked eye (Kappas et al., 2013; Larsen et al., 2003). Thirdly, some previous studies further indicate that, beyond reliable detection of emotional facial expressions, these may be leveraged to substantially improve human-computer interaction (Gibert et al., 2009; Schultz, 2010). Last but not least, many up-to-date in-house and external research works have confirmed the practicality, convenience, and effectiveness of EMG in different areas of the human body and physiological exploration (Liu et al., 2023; Cai et al., 2023; Hartmann et al., 2023; Liu and Schultz, 2022).

Within the scope of this paper, we use our in-house recorded dataset of fEMG sensor data to predict a subset of AUs. To the best of our knowledge, no work has been published yet in AUR relying only on EMG data as its source.

2 METHODOLOGY

The proposed framework is based on a pilot dataset, which contains synchronised video modality data with fEMG recordings and output labels corresponding to appropriate AUs. Multiple widely-applied ML models with default hyperparameters will be trained using the acquired data and subsequently, classification metrics will be calculated for the same. The best-performing model will be chosen for further analysis.

2.1 Dataset Preparation

To construct a dataset, we recorded a proof-of-concept fEMG dataset of one subject across four sessions. Each session comprised of 25 recording trials, yielding a total of 100 trials. We have on average 2.12 ± 0.8 minutes of data per trial across all the sessions. Within each recording trial, the subject was asked to imitate facial expressions shown to them in the stimulus videos through a custom-made graphical user interface (GUI) 1(see Figure 1). These stimulus videos were taken from the MPI Video Database (Kleiner et al., 2004). These videos provide accurate portrayals of AU activation, which have been verified by FACs coders.

The subject was shown stimulus videos pertaining to four different AUs (see Table 1 and asked to imitate the AUs at maximum intensity and hold them for at least five seconds while fEMG data was being parallely recorded. In addition to the different AUs, the fEMG data corresponding to the neutral expression was also recorded, which we shall refer to as AU0 for representation purposes in the rest of the paper.

The recording setup consisted of a computer displaying the stimuli, a webcam, and an fEMG recording setup. The subject was seated in front of the display screen approximately 70 cm away. The fEMG setup was a bipolar recording setup consisting of 2 channels covering the Frontalis and Corrugator Supercili facial muscles. These positions are defined by the guidelines of the Society for psychophysiological research (Fridlund and Cacioppo, 1986), with slight deviations from the standard sensor positions (see Figure 2), based on extensive pre-testing to minimize the amount of crosstalk, and to account for the slightly larger size of our electrodes compared to the original guideline paper. These deviations were based on intensive pre-testing to optimize the quality of the recording. It is known that fEMG signals occur in the range of 15–500Hz (Boxtel, 2001), so the sampling frequency was chosen to be 2000 Hz as required by the limitations imposed by the Nquist theorem, which states that a periodic signal must be sampled at more than twice the highest frequency component of the signal. We employed the Biosignal Plux [1] sensors and hub as our acquisition system because its high-quality EMG acquisition was confirmed by many preliminary in-house works (Liu and Schultz, 2018; Hartmann et al., 2022; Liu et al., 2021a).

The synchronization between the imitated actions and the EMG sensor data is handled using the lab streaming layer (LSL) protocol. Along with the EMG data, we also record the video data of the participant

[1] www.pluxbiosignals.com

using the standard inbuilt webcam within the host computer. The timestamps associated with the video stream for the imitated actions are used to extract the relevant EMG signals and store them in a usable format for further processing.

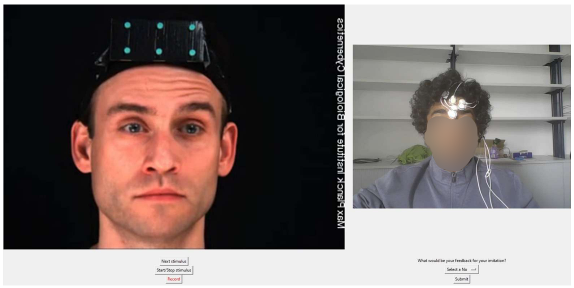

Figure 1: Graphical User Interface (GUI) implemented for recording trials.

Figure 2: Sensor Placement.

Table 1: Selected actions units for proof-of-concept study.

Action Unit	Action
AU1	Inner Brow Raiser
AU2	Outer Brow Raiser
AU4	Brow Lowerer
AU9	Nose Wrinkler
AU0	Neutral Expression

2.2 Pre-Processing and Feature Extraction

Raw EMG data typically includes a substantial amount of electrical noise, which should be removed before amplification (Tassinary et al., 2007). Traditionally, remaining noise (e.g., 50/60 Hz noise) and artifacts are then removed via filtering prior to any statistical analyses (Fridlund and Cacioppo, 1986; Tassinary et al., 2007). Within the field of biosignals-based ML, however, this latter type of noise may be better accounted for by the ML algorithm than by a filter, which might filter out some relevant data along with the noise. Therefore, some recent works in this field have suggested running their feature extractors directly on the raw fEMG (Liu and Schultz, 2019; Rodrigues et al., 2022; Liu, 2021), which we adopted in our work. The raw EMG data was first segmented into windows of a specified length with a pre-defined overlap percentage.

Training ML models require the proper set of features to be fed into the model. Considering our model orients a real-time application for interaction and control in the future, it is essential to consider the features that can be computed quickly. Therefore, following the windowing of signals, 16 temporal features were extracted using a time-series feature extraction library (TSFEL) (Barandas et al., 2020):

- Absolute energy
- Area under the curve (AUC)
- Autocorrelation
- Centroid
- Entropy
- Mean absolute difference (MAD)
- Median difference
- Negative turning points (NT)
- Neighbourhood peaks
- Peak to peak distance
- Positive turning points
- Signal distance
- Slope
- Sum absolute difference (SAD)
- Total energy
- Zero crossing rate

Most of the features listed above are frequently used in time-series ML, like AUC, MAD, and SAD; some features exist almost uniquely in TSFEL, such as NT, whose effectiveness and low computational cost have been validated in previous research (Liu et al., 2022). The feature extraction was followed by linear discriminant analysis (LDA) for dimensionality reduction (Hartmann et al., 2020; Liu et al., 2021b; Hartmann et al., 2021). Such an operation reduced the feature set and wad used for the training of different models.

2.3 Classification Models

We applied six widely-used ML and deep learning classification models to our pilot dataset.

1. Random forest RF
2. Support vector machine SVM
3. Gaussian naive Bayes GNB
4. K nearest neighbors *k*-NN
5. Artificial neural networks (ANN, deep)
6. Temporal convolutional networks (TCN, deep)

Five of these models were trained on features extracted using TSFEL library while the TCN model used the windowed data as its input. All non-deep models were trained on default hyperparameters provided by the *scikit learn* library (Pedregosa et al., 2011). The ANN model was designed as a vanilla seven layer network with leaky rectified linear unit (Leaky ReLU) and dropout layers stacked in between and ending with a softmax layer at the end. The TCN network was designed to takeinput as samples of defined windowed length. The output of the TCN module is of the same length as the input. This output is flattened out and fed to a single neural network with the number of output nodes same as the number of categories, followed by a softmax activation for prediction.

3 RESULTS

3.1 Cross-Validation Results

We conducted cross-validation studies using three of the acquired data sessions as our training set. The training data was first segmented into windows of 400 ms in length and 20% overlap, followed by temporal feature extraction and LDA. Subesequently, a five-fold cross-validation was performed using stratified sampling, and the mean accuracy results are showcased in Table 2.

Table 2: Accuracies for the results accumulated from five-fold cross-validation using combinations of different sessions of data.

Session(s):	1	1 and 2	1,2, and 3
RF	0.99	0.99	0.99
SVM	0.95	0.95	0.96
GNB	0.99	0.98	0.97
KNN	0.98	0.98	0.98
ANN	0.27	0.39	0.35
TCN	0.77	0.80	0.76

3.2 Leave-One-Channel-Out Testing

Considering the high performance of the models demonstrated in Table 2. We wanted to test how our models would perform when they were only trained on individual data channels. Cross-validation studies were done on the same training data but only using individual channels from our dataset. We analysed how changing the number of channels used for training the models would impact the performance. Table 3 indicates the results for the performances of the models on independent channels. Noticeably, there is an expected drop in the performance of the models but not much of a clear discernible pattern in terms of one channel performing better than the other. For example RF seems to be better trained with Frontalis data while SVM works better with the Corrugator data. TCN has the most significant drop in performance, suggesting that it may require a combination of features from multiple channels to recognize classes.

Table 3: Accuracies for the results accumulated from five-fold cross-validation using single channel data based on random forest.

	Frontalis	Corrugator
RF	0.97	0.91
SVM	0.75	0.83
GNB	0.88	0.84
KNN	0.89	0.89
TCN	0.28	0.37

3.3 Performance of Test Set

Although validation scores present one side of the argument on how the model performs on a similar distribution of data on which it was performed. We wanted to know how the model performs on an entirely unseen dataset, i.e session-independent data. We trained different models and calculated the accuracy metrics on the unseen test data set that we had kept apart from the start. The results are displayed in Figure 3. The trained models provide excellent test set scores. Based on the accuracy graphs obtained, it was apparent that all models except for ANN were showing promising results. We continue to see how the best-performing model performs in real time.

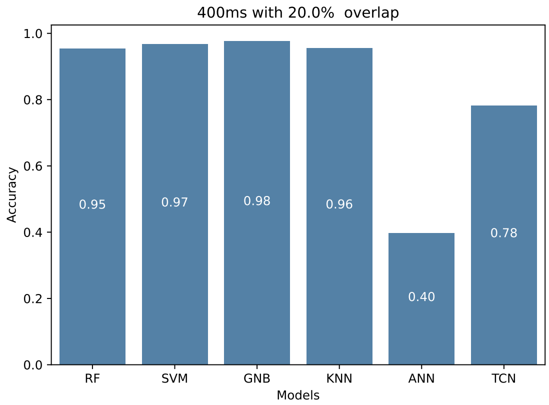

Figure 3: Accuracies on unseen test set.

3.4 Real-Time Analysis

RF was chosen as the base model for further analysis, as it outperformed the other models in our offline investigation. Satisfactorily, RF is up to the task in real time (see Figure 4) as it was able to recognize the participant's facial AU's in near-real time with a decent rate of correctness, which supports our hypothesis that EMG biosignals can be used to accurately predict AUs.

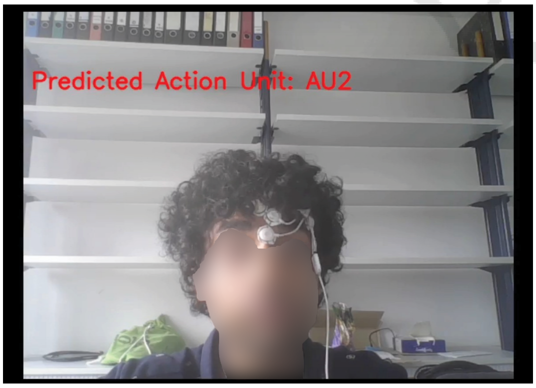

Figure 4: Real-time action unit recognition.

3.5 Joint Study of Window Length and Overlap Percentage

One helpful comparison to showcase is how the model performs when we change the window length and the overlap percentage value. We again chose RF as our basis and applied window lengths varying from 150ms to 500ms in increments of 50ms. The overlap percentages values varied from 10% to 50% in increments of 10% (see Table 4).

Table 4: Mean accuracy results with five-fold cross-validation for jointly studying window length and overlap percentage.

	10%	20%	30%	40%	50%
150 ms	0.97	0.97	0.97	0.97	0.97
200 ms	0.97	0.98	0.98	0.98	0.98
250 ms	0.98	0.98	0.98	0.98	0.98
300 ms	0.98	0.98	0.98	0.98	0.98
350 ms	0.98	0.98	0.98	0.98	0.99
400 ms	0.99	0.98	0.98	0.98	0.98
450 ms	0.98	0.98	0.98	0.98	0.98
500 ms	0.98	0.98	0.98	0.98	0.98

4 DISCUSSION

To the best of our knowledge, this paper is the first to provide a proof-of-concept for upper face AU recognition based solely on EMG data. The subject-dependent model achieved high accuracy in near-real time. Furthermore, we showed that our models could also classify AUs that were not directly underneath the respective recording sites, suggesting that future systems may achieve adequate results from conveniently placed electrodes that are even more distal from the respective facial muscles.

4.1 Evaluation and Analysis

We focused on recording EMG data from two muscle sites, Frontalis and Corrugator Supercili, to capture the activity of the eyebrows and distinguish between four different AUs (AU1, AU2, AU4, AU9). While we did not record above the levator labii (Nose Wrinkler), we assumed that sufficient signals could still be detected from these nearby sites also to allow a reliable classification of AU9. As demonstrated by our results, this classification was successful and robust, even when using only single channel data. Thus, while previous work has pointed out the often problematic effects of crosstalk phenomena of other muscles (Van Boxtel and Jessurun, 1993; van Boxtel et al., 1998), our models appear to have been able to successfully leverage these data to recognize the intended AUs.

As demonstrated by our validation results, all non-deep learning models obtained good results. As TCN is a deep learning-based model, we did not expect it to perform well given the relatively small amount of data. As suggested by our training runs, the TCN model was still unstable. Nevertheless, its results appeared promising, with performance at a level similar to that of an ANN. Future work could therefore investigate whether TCN-based models may achieve even better and more stable results once larger EMG-datasets become available. Contrarily, ANNs did not provide sufficiently strong results to be considered a good candidate compared to the other models. We hypothesize that this could be attributed to two factors. First, the vanilla neural networks utilized for our model training may require more extensive data. Second, ANNs may have performed worse than TCNs because they failed to construct adequate internal feature maps.

4.2 Comparison with State-of-the-Art Work

To the best of our knowledge, very few prior works have aimed to leverage EMG data for action unit recognition. (Perusquia-Hernandez et al., 2021) relied on the fusion of computer vision data along with EMG to train their models, while (Gibert et al., 2009) used EMG data only to predict prototypical facial expressions. Similarly, (Gruebler and Suzuki, 2014) developed a wearable device to detect smiling and frowning based on two electrode pairs. Some works have relied on independent video data (Baltrusaitis et al., 2018) or even using other modalities such as electroencephalography (Li et al., 2020). Finally a recent approach has utilised earbud IMUs (Verma et al., 2021) and TCNs to detect and classify a large range of AUs. While this approach has obtained promising results in a subject-dependent setting, ear-mounted IMU sensors may be at a disadvantage with respect to detecting more subtle naturalistic facial activity, the presence of movement artefacts (e.g., head movements) or interference due to the sound waves produced by the earbud when it is in use. Thus, the sound waves may themselves excite the IMUs as well as the earable device (Verma et al., 2021).

5 CONCLUSION

We propose a novel approach to upper-face AUR using EMG data, as demonstrated by the successful training of subject-dependent models in this initial case study. Our results furthermore show potential for classifying new AUs based on more distally placed electrodes in future applications, e.g., in VR. These results also suggest that deep learning models such as TCN can be considered for further research in this domain, while highlighting the limitations of using fewer channels. Overall, this work contributes to the emerging field of EMG-based AUR recognition and paves the way for future research.

We believe that this approach could be complementary to the development of IMU-based earable devices, as they are subject to different sources of noise and environmental as well as practical constraints. Thus, while EMG sensors require direct contact with the skin, they are likely to be more robust towards artefacts due to head movements or the sound waves produced by the earable itself. Conversely, earbuds are likely to be less susceptible to electrical noise from other devices, whereas EMG should outperform other sensor types for detecting subtle expressions. Considering the challenges of multimodal emotion recognition in the wild (Küster et al., 2020), we therefore envision a joint system comprising of both IMUs and a few EMG sensors to be able to provide the most robust and precise AUR performance. However, more work is still required to develop robust and versatile AUR from fEMG.

Despite the successful AU recognition in this pilot, our present models were still limited to training on subject-dependent data near the traditional recording sites for the respective action units in the upper face. As demonstrated by prior work, control of human facial muscles is complex (Cattaneo and Pavesi, 2014) and subject to significant anatomical differences (D'Andrea and Barbaix, 2006) as well as variability in signal power across muscle sites (Schultz et al., 2019). To address these challenges, we plan to build subject-independent models to examine whether the underlying muscle activity patterns are sufficiently reliable. In our future work, we therefore aim to to extend our EMG-based AUR approach also to lower face AUs, while further examining the viability of a distal electrode placement in multi-subject studies.

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.

Cai, L., Yan, S., Ouyang, C., Zhang, T., Zhu, J., Chen, L., Ma, X., and Liu, H. (2023). Muscle synergies in joystick manipulation. *Frontiers in Physiology*, 14:1282295.

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

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.

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.

Dupré, D., Krumhuber, E. G., Küster, D., and McKeown, G. J. (2020). A performance comparison of eight commercially available automatic classifiers for facial affect recognition. *PLOS ONE*, 15(4):e0231968. Publisher: Public Library of Science.

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.

Gibert, G., Pruzinec, M., Schultz, T., and Stevens, C. (2009). Enhancement of human computer interaction with facial electromyographic sensors. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, pages 421–424, Melbourne Australia. ACM.

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.

Gruebler, A. and Suzuki, K. (2014). Design of a Wearable Device for Reading Positive Expressions from Facial EMG Signals. *IEEE Transactions on Affective Computing*, 5(3):227–237. Conference Name: IEEE Transactions on Affective Computing.

Hartmann, Y., Liu, H., Lahrberg, S., and Schultz, T. (2022). Interpretable high-level features for human activity recognition. In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 4: BIOSIGNALS*, pages 40–49.

Hartmann, Y., Liu, H., and Schultz, T. (2020). Feature space reduction for multimodal human activity recognition. In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2020) - Volume 4: BIOSIGNALS*, pages 135–140.

Hartmann, Y., Liu, H., and Schultz, T. (2021). Feature space reduction for human activity recognition based on multi-channel biosignals. In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021)*, pages 215–222. INSTICC, SCITEPRESS - Science and Technology Publications.

Hartmann, Y., Liu, H., and Schultz, T. (2023). High-level features for human activity recognition and modeling. In Roque, A. C. A., Gracanin, D., Lorenz, R., Tsanas, A., Bier, N., Fred, A., and Gamboa, H., editors, *Biomedical Engineering Systems and Technologies*, pages 141–163, Cham. Springer Nature Switzerland.

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

Hollenstein, T. and Lanteigne, D. (2014). Models and methods of emotional concordance. *Biological Psychology*, 98:1–5.

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

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.

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.

Kuester, D. and Kappas, A. (2013). Measuring emotions in individuals and internet communities. In Benski, T. and Fisher, E., editors, *Internet and emotions*, pages 48–62. Routledge.

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.

Larsen, J. T., Norris, C. J., and Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, 40(5):776–785. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1469-8986.00078.

Lewinski, P., den Uyl, T. M., and Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, 7(4):227–236.

Li, X., Zhang, X., Yang, H., Duan, W., Dai, W., and Yin, L. (2020). An EEG-Based Multi-Modal Emotion Database with Both Posed and Authentic Facial Actions for Emotion Analysis. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 336–343, Buenos Aires, Argentina. IEEE.

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.

Liu, H. (2021). *Biosignal processing and activity modeling for multimodal human activity recognition*. PhD thesis, University of Bremen.

Liu, H., Hartmann, Y., and Schultz, T. (2021a). CSL-SHARE: A multimodal wearable sensor-based human activity dataset. *Frontiers in Computer Science*, 3:90.

Liu, H., Hartmann, Y., and Schultz, T. (2021b). Motion Units: Generalized sequence modeling of human activities for sensor-based activity recognition. In *29th European Signal Processing Conference (EUSIPCO 2021)*. IEEE.

Liu, H., Jiang, K., Gamboa, H., Xue, T., and Schultz, T. (2022). Bell shape embodying zhongyong: The pitch histogram of traditional chinese anhemitonic pentatonic folk songs. *Applied Sciences*, 12(16).

Liu, H. and Schultz, T. (2018). ASK: A framework for data acquisition and activity recognition. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018) - Volume 3: BIOSIGNALS*, pages 262–268.

Liu, H. and Schultz, T. (2019). A wearable real-time human activity recognition system using biosensors integrated into a knee bandage. In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2019) - Volume 1: BIODEVICES*, pages 47–55.

Liu, H. and Schultz, T. (2022). How long are various types of daily activities? statistical analysis of a multimodal wearable sensor-based human activity dataset. In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 5: HEALTHINF*, pages 680–688.

Liu, H., Xue, T., and Schultz, T. (2023). On a real real-time wearable human activity recognition system. In *Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2023) - WHC*, pages 711–720.

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.

Noah, T., Schul, Y., and Mayo, R. (2018). When both the original study and its failed replication are correct: Feeling observed eliminates the facial-feedback effect. *Journal of Personality and Social Psychology*, 114(5):657–664.

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.

Perusquia-Hernandez, M., Dollack, F., Tan, C. K., Namba, S., Ayabe-Kanamura, S., and Suzuki, K. (2021). Smile Action Unit detection from distal wearable Electromyography and Computer Vision. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–8, Jodhpur, India. IEEE.

Rodrigues, J., Liu, H., Folgado, D., Belo, D., Schultz, T., and Gamboa, H. (2022). Feature-based information retrieval of multimodal biosignals with a self-similarity matrix: Focus on automatic segmentation. *Biosensors*, 12(12).

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. (2010). Facial Expression Recognition using Surface Electromyography.

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.

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 Boxtel, A. and Jessurun, M. (1993). Amplitude and bilateral coherency of facial and jaw-elevator EMG activity as an index of effort during a two-choice serial reaction task. *Psychophysiology*, 30(6):589–604. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8986.1993.tb02085.x.

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.

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.

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.