ADVANCES IN MACHINE LEARNING AND INTEGRATED CIRCUITS FOR SMART ASSISTIVE
TECHNOLOGIES

By

Ehsan Ashoori

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Electrical and Computer Engineering - Doctor of Philosophy

2024

ABSTRACT

Assistive technologies have emerged as powerful tools for assessing physical health and wellness through monitoring physiological parameters such as movement and heart rate. However, our overall health is influenced not only by physiological parameters but also by mental health factors and environmental influences. Therefore, in the pursuit of holistic wellness, assistive technologies need to support multimodal sensing to monitor various aspects of individuals' health, including physiological health, mental wellness, and environmental parameters that influence personal health and wellness. The challenges arise when these technologies must be implemented in real-time and in miniaturized point-of-care platforms where multi-modal sensing algorithms must run efficiently, and resources, including power, are limited. Solving these challenges requires converging engineering practices with psychological and physiological principles. This work aims to implement resource-efficient algorithms to assess social interaction parameters as an important mental health factor and to enable high-performance point-of-care devices to monitor physiological and environmental parameters in a miniaturized and effective manner. In this work, an extensive dataset for human interaction in virtual settings was prepared. Efficient algorithms were developed to identify levels of two highly important social interaction parameters, 'affect' and 'rapport.' We analyzed affect in time intervals based on the conversation turns and analyzed rapport in 30-second time intervals, which is the highest temporal resolution reported in the literature. We achieved an affect prediction accuracy of 76.8% and a rapport prediction accuracy of 73.6%, which are the highest reported results for analyzing multi-person groups. Furthermore, to support monitoring physiological and environmental parameters, electrochemical solutions were identified as a

highly effective method. We introduced new architecture to overcome limited supply potentials in modern point-of-care devices. In our novel design, the potential window for electrochemical reactions doubles compared to the traditional designs. This, in return, facilitates a significantly wider range of target elements that can be monitored with this novel architecture. Overall, the enhanced algorithms and architecture introduced in this work enable multimodal sensing of important personal health and wellness parameters.

To researchers dedicated to making the world a better place

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

Chapter 1: Introduction

1.1. Applications and Significance of Assistive Technologies in Individuals' Wellness

Assistive technologies have emerged and gained popularity as powerful tools for tracking the physical health of individuals. However, our overall wellness is affected by factors such as emotional state and environmental factors that influence individuals' health. Therefore, in the pursuit of rounded and comprehensive wellness, it is paramount to develop assistive technologies for monitoring social, physiological, and environmental parameters to promote individuals' wellness.

1.1.1. Social interactions and individuals' wellness

Social connections and relationships are vital components of overall well-being, influencing mental health, emotional resilience, and a sense of belonging. Monitoring social parameters involves assessing the quality of interpersonal relationships, support networks, and community engagement. By tracking indicators such as social connections, loneliness levels, and participation in social activities, individuals and healthcare providers can identify areas for improvement and intervention. Therefore, developing technologies to monitor social interactions can provide insights into individuals' social behaviors, communication patterns, and social support systems, aiding in the identification of potential risks or opportunities for enhancing social wellness. These tools can be utilized in a wide range of settings, from helping healthcare professionals to improving the quality of interactions in workplaces. This can, in return, improve health services, and subsides anxiety, biases, and inequity for example in work places.

1.1.2. Role of physiological and environmental parameters on individuals' wellness

Physiological health is intricately linked to overall wellness, encompassing factors such as physical fitness, nutrition, sleep quality, and stress levels. Monitoring physiological parameters involves tracking key metrics such as heart rate, blood pressure, body composition, and biochemical markers. Assistive technologies that regularly assess these parameters could help individuals gain insights into their health status, identify potential health risks, and make informed lifestyle choices to optimize wellness.

Likewise, environmental factors play a significant role in shaping individual wellness, influencing physical health, mental well-being, and overall quality of life. Monitoring environmental parameters involves assessing factors such as air quality, noise levels, and temperature. By understanding the impact of the environment on wellness, individuals and communities can take steps to create healthier living environments and mitigate potential health hazards.

Developing assistive technologies for monitoring physiological and environmental parameters could improve individuals' wellness and quality of life. The emergence of wearable devices, smartphone apps, and point-of-care technologies that, for instance, monitor heart rate, air quality, or noise level, is enabling real-time tracking and analysis of health data. These assistive technologies empower individuals to take proactive control of their health, facilitating early detection of health issues and timely interventions to prevent or manage chronic conditions. Moreover, by leveraging these quantitative data, individuals can make informed decisions to optimize their living environments, reduce exposure to pollutants, and promote overall wellness.

1.2. Engineering challenges with assistive technologies

Assistive technologies have advanced in many areas. However, despite these advancements, several challenges and areas for further research remain.

- Multimodality and interoperability*:* many assistive technologies operate in isolation, lacking interoperability with other devices or platforms. Multifaceted assistive technologies and the integration of different data sources are needed. Therefore, multimodal assistive technologies that address different aspects of individual's health such as psychological and physiological are highly desirable. This allows for a holistic monitoring of individuals' wellness.

- Resource-efficient implementation: many advanced algorithms and devices in assistive technologies involve resource-heavy operations that prevent real-time execution. This also limits the range of target applications that a point-of-care device can support. Developing resource-efficient algorithms and devices remains an important area of research.

- Temporal resolution: many assistive technologies have been introduced, for instance, to analyze individuals' overall interaction and social engagement. However, the temporal resolution of analysis is often low in these assessments and they lack real-time analysis. Therefore, higher temporal resolution assessment solutions are necessary, especially for analyzing the dynamic of interactions over the course of an event. These fine-resolution analyses are paramount to devising individual plans for improving social interactions.

- Versatility of the solutions: translating from standard laboratory solutions to point-of-care assistive technologies often faces limitations such as reduced functionality. Wearable point-of-care devices must be implemented with a small form factor and low power consumption. These constraints often result in a limited range of operations, excluding many target parameters. Therefore, further research is needed to improve the performance of point-of-care devices within the limitations of wearable devices.

- Personalization and Accessibility: assistive technologies often adopt a one-size-fits-all approach, overlooking the diverse needs and preferences of users. Research is needed to develop personalized and customizable solutions that adapt to individual abilities, preferences, and contexts, thus enhancing user engagement and effectiveness.

1.3. Goals

Collaboration between researchers from different disciplines is essential for tackling these challenges and for the successful development and deployment of assistive technologies. The overall objective and vision of this work is to identify microsystems and algorithms to overcome the challenges specified in section 1.2. Discovering important parameters in social interactions and the avenues that technology can help is essential. This requires an understanding of individuals' psychology. Moreover, identifying the potential technologies for addressing these parameters requires a deep understanding of engineering solutions.

In this work, we bring expertise in machine learning and the extensive experience our lab has in developing efficient microsystems to tackle different aspects of challenges in assistive

technologies. We aim to take a holistic approach to improving individuals' wellness. Specifically, the following are the focus of this work.

1.3.1. Developing technologies that enable monitoring of important social interaction parameters with high temporal resolution.

There is limited literature on analyzing social interaction parameters in groups, especially with high temporal resolution. **In this work, the <u>goal</u> is to develop resource-efficient algorithms to monitor affect and rapport.** Engaging user interfaces that accommodate personal preferences and accessibility issues is of important consideration.

1.3.2. Applying microsystems techniques to bring laboratory utilities to assistive technologies.

Utilizing our lab's extensive experience in developing wearable technologies for point-of-care applications, this work focuses on developing assistive technologies for enhancing individuals' wellness by monitoring physiological and environmental parameters. More specifically, electrochemical solutions for detecting various physiological and environmental parameters that influence individuals' wellness are explored. Given the limited resources available for wearable devices, electrochemical solutions in these devices face serious limitations in the range of parameters that can be detected. Widening this range and targeting more diverse elements is the aim of this work. **The <u>goal</u> is to make CMOS potentiostat overcome the limitations of analyzing a wide range of elements.**

1.4. Outline

The following forms the content of this dissertation. Literature on employing assistive technologies for improved human interaction as well as monitoring of physiological and

environmental parameters is reviewed in Chapter 2. Chapter 3 presents the early work we did and the avenues we explored toward having a platform for improved virtual interactions. Chapter 4 describes the data collection and preparation for human trials along with the algorithms we developed for extracting social cues in virtual meetings. Chapter 5 presents the methods we employed for enhancing the efficacy of point-of-care electrochemical devices that are resource-efficient. Chapter 6 summarizes this dissertation and outlines potential paths for future works.

Chapter 2: Background and literature review

Literature has reported point-of-care devices and assistive technologies for improving different aspects of individuals' wellness. This includes technologies for monitoring psychological wellness such as individuals' emotions [1], [2], [3] as well as physiological parameters such as heart rate [4]. Some others focus on analyzing physical parameters such as skin conductivity using electrodermal activity sensors that can indicate stress levels and various health-related issues [3], [5]. Some others develop point-of-care devices to analyze human secretions, such as sweat [6], and monitor environmental parameters, such as particulate matter [7], [8], that can affect health. In this chapter, we explore the literature and identify the challenges and areas that require further research.

2.1. Employing technology for improved interaction

Among the solutions for monitoring individuals' emotional wellness, a body of work has recently gained attention that focuses on the interaction among people on different occasions, such as in a classroom, in a work meeting, in a clinical set, etc. [5], [9], [10]. After a recent shift in the trend that incorporated more and more virtual interactions, the need to improve online interactions has become more important than ever. That is especially because of the different nature of online interactions compared to in-person setups.

To have a productive meeting at work or to have an effective learning experience in a classroom, we benefit from recognizing non-verbal audial or visual cues in our audience. These cues help find out about the emotional states of people and the level of their engagement and, consequently, help establish more effective communication with our audience. For instance, a study [11] showed that social intelligence had a significant effect on the professional

performance of mathematics teachers. Thus, it is desirable to leverage technologies to help people perceive these cues in their audience.

The inability to detect important cues in an interaction is more pronounced in a virtual setting. Distance collaboration has become a common practice in recent years. Many companies and universities have opted to facilitate remote working and education. Even some companies went on to announce they would let their employees work remotely for the indefinite future. This trend shows distance collaboration will stay and only flourish in the coming years. Thanks to video conferencing technologies, we can now hold these virtual events that were not possible not too long ago. However, many elements of in-person interactions are missing in a virtual environment. For instance, lack of eye contact, noting body gestures, and other cues that are more easily assessable in an in-person meeting are missing in a virtual setup. This leads to less effective communication. Therefore, utilizing technologies to help people communicate better and have more effective interactions in this type of environment is highly desirable.

Recently, an increasing interest has been seen in the literature for developing technologies that are capable of detecting the emotional state of people [12]. Many of these methods rely on deep neural network implementation [13] which often is computationally heavy and not generally applicable to real-time implementation with limited computational resources. On the other hand, some of the reported works in literature employ machine learning algorithms that require less computation but often require hand-crafting features, which adds to the complexity of the problem [14]. Furthermore, these reported works are generally bound to the controlled lab environment [15], where the designed experiments induce desired emotions in the participants. These experiments, therefore, have a higher signal-to-noise ratio than a normal

interaction in a natural setup. Consequently, the developed algorithms might perform less effectively in a more natural setup. Moreover, some platforms were developed for detecting nonverbal cues from recorded videos [16] in a noncontrolled environment, but often focused on detecting very intense emotions, which are very different than the baseline emotion and hence easier to identify.

An example of using automated solutions to improve interactions is utilizing technologies to reduce unintended negative communications among participants. For instance, most unfavorable interactions in a workplace are being done unconsciously [17]. Many individuals may have unconscious bias against different groups of people. A common method that traditionally has been employed to overcome these challenges is through human experts. In this method, an expert analyzes the behavior of participants and provides constructive feedback to achieve higher-quality interactions. However, this method does not work well in real time. This means that this method is appropriate for an overall assessment of an interaction after it is over. Furthermore, since a human is involved in this type of assessment, some people may be uncomfortable with it and raise some privacy issues. Therefore, employing technology to enhance awareness of individuals helps to have more positive interactions.

The literature has investigated these technologies for various setups. The following are examples of literature that use technologies to improve virtual interactions in the most popular setups. The target applications were mostly for online settings such as online classrooms and online work meetings.

- Online learning environment

In a study [2] on virtual learning setups, researchers demonstrated that tutors who were provided with the emotional state of the learners in a virtual classroom used more affective elements in their report and wrote more formative and less summative feedback.

- Online work meeting

In another study [14], researchers developed a platform that processed audio and video data after a video conference session and extracted affective features such as smile and attention, as well as speech overlap and turn-taking. By providing feedback after finishing a session, participants demonstrated statistically significant improvements in balanced participation.

2.2. Types of cues extracted by algorithms

2.2.1. Emotional state

Emotions can be perceived as residing on two distinct dimensions: one concerning the degree of pleasure associated with the emotion and the other regarding the level of arousal or activation it entails [18]. Recently, literature has shown increasing interest in developing technologies capable of detecting people's emotional states [15], [19].

Emotions mirror responses of the sympathetic nervous system [20]. The Polyvagal Theory elucidates how emotional states influence both brain processes and bodily functions [21]. Moreover, this theory sheds light on the interplay between measurable physiological states tied to the autonomic and central nervous systems and resultant human behavior, proposing a mutual relationship between mind and body. It further suggests that environmental factors influence behaviors that subsequently impact physiological states. Thus, monitoring changes in

bodily physiological markers like respiration rate, heart rate, and perspiration rate can offer valuable insights into an individual's emotional state [22].

2.2.2. Engagement intensity

Social Cognitive Theory (SCT) asserts that individuals' interpretations of their surroundings can shape their emotional, physiological, and behavioral responses [23], thus impacting subsequent behaviors in a reciprocal manner. We define engagement by level of interest and cohesion shown by participants and their communication dynamics.

Multiple platforms were developed for detecting nonverbal cues from recorded videos [24], for finding the engagement intensity of people [9], [10], [25], [26]. Other cues such as head motion synchronization and empathy in face-to-face communications have been studied [27].

2.2.3. Rapport building

Rapport is defined as a friendly and harmonious relationship, especially, "a relationship characterized by agreement, mutual understanding, or empathy that makes communication possible or easy" [28]. Recent literature has explored monitoring rapport building between dyadic pairs. Studies have been done on both human-to-human and human-to-virtual agent interactions [29], [30], [31]. Studies in the literature focusing on analyzing rapport utilize various modalities such as audio [18], natural language [32] and video [33]. Machine learning approaches have been utilized in the literature [29], [33], [34] for analyzing rapport in various communication contexts. These algorithms focus on discerning the emotional valence of communication and identifying instances of agreement, disagreement, or conflict. Machine learning models can leverage audio and visual cues, such as tone of voice, intonation [18] and facial expressions [29], to infer underlying sentiments and attitudes.

2.3. Offline vs real-time feedback

A rich body of literature [9], [14], [15], [25], [26], [27], [35], [36], some of which were presented in the previous sections, has focused on detecting nonverbal cues in human interactions, though only provided off-line feedback to participants about their behavior, once the session is over. However, some other studies [37], [38] developed platforms for providing real-time feedback to the users using innovative visual representation, though limited to text-based communication in chatrooms. They analyzed the communication patterns as well as group dynamics using their platform. They also analyzed whether the feedback made any distractions for users.

The challenge of providing real-time analysis and feedback using technologies is that the required algorithms are extremely computationally demanding. Therefore, it makes it very difficult, if not impossible, to utilize common algorithms and methods for the real-time analysis of events. Extensive computational load manifests itself differently whether we are dealing with an in-person or a virtual setup. Since in an in-person/on-the-go situation, we would typically have limited hardware resources, we would like to increase both the computational and hardware efficiency to make the solutions viable on wearable devices. In virtual setups, however, access to powerful hardware (thorough computers for example) is not typically an issue, but we still need to increase the computational efficiency to speed up and facilitate the real-time processing of algorithms.

2.4. Sensor modalities for social cue extraction

Reported works in the literature use multiple modalities such as audio (tone, pitch, etc.), natural language, video, etc. [39], [40]. In [39], researchers used deep neural network to analyze

12

audiovisual data for affect recognition. [40] also uses deep neural networks to analyze the speech.

Other researchers used visual data to analyze the engagement intensity of people in different occasions such as a classrooms [9], [10], [25], [26]. They use facial expressions and physiological sensor data such as heart rate and employ various machine learning algorithms to identify students' engagement levels. Other cues, such as head motion synchronization and empathy in face-to-face communication, have been studied using the accelerometer in a lab environment [27], and it was shown that the level of empathy is mirrored in the frequency and phase of head motion synchronization.

2.5. Developing technologies to improve interactions in virtual meetings

We aim to develop efficient technologies to assist people in having more productive and positive meetings in the workplace, for example. We are also interested in improving the quality of virtual interactions in an online setting. To this end, we focus on developing algorithms to detect important cues from individuals, analyze them in tandem with the cues from other people, and feed the processed data back to the participants.

The feedback to the participants can be placed in offline or online modes. Although providing real-time feedback leads to having the most effective solution to increase the quality of interactions, offline assessment and feedback to the participants could also enhance awareness and improve the interactions. Another important aspect is the type and frequency of feedback data to the participants. We are interested in exploring different avenues for providing this information to the participants from both psychological and technical points of view.

To overcome these challenges, this work aims to develop methodologies for extracting social cues from participants in an online meeting in a natural setup without any artificial constraints. We aim to do this with the highest computational efficiency to enable future implementation of real-time analysis.

2.6. Effect of physiological and environmental parameters on individuals' wellness

The Polyvagal Theory highlights how emotional states affect both brain functions and bodily processes [21]. Additionally, this theory illuminates the dynamic interaction between measurable physiological states linked to the autonomic and central nervous systems and resulting human behaviors, proposing a bidirectional relationship between the mind and body. It also suggests that the environment influences behaviors that subsequently impact physiological states. Therefore, tracking changes in bodily indicators such as heart rate can yield valuable insights into an individual's emotional state. Similarly, monitoring environmental conditions can provide information on how the surroundings influence emotional states and other factors that directly affect the wellness of individuals. Therefore, a rich body of literature has studied these effects on the overall health and wellness of individuals and assistive technologies that have been developed for assessing these important parameters [41], [42].

2.7. Sensor modalities for monitoring physiological and environmental parameters

Among the sensors modalities, optical methods for measuring heart rate of individuals are widely popular [43], [44]. Among different methods, pulse oximetry is popular for measuring heart rate and hemoglobin oxygen saturation in a noninvasive manner. It also can be used for determining respiratory rates. Some recent literature [45] has suggested using camera feed for monitoring heart rate in an online session purely based on visual data.

14

Respiratory rates are also an indicator of different emotional states [46]. A sensor modality used for respiratory rate estimation is an inertial measurement unit (IMU) [47]. IMU detects chest movements and estimates the breathing rate based on the physical movements. Given how indicative of an emotional state a breathing rate can be, this method provides a noninvasive approach to the detection of the respiratory rate.

Another category of sensors is electrodermal activity (EDA) and galvanic skin response (GSR) sensors. These sensors have been reported in literature [48], [49], [50] to be used for monitoring emotional state of individuals. In other works, skin temperature [51] has also been investigated as an indicator of individuals' emotions. Furthermore, other approaches, such as utilizing electroencephalography (EEG) for monitoring the electrical activity of the brain, have been explored in the literature [52]. These brain waves can be indicative of the states an individual is in and, therefore, a valuable insight into human overall emotions.

Eye trackers use infrared and visual spectrum to monitor pupil diameter, gaze distance and coordinates and eye blinking. These parameters have been shown to be indicative of individuals' emotions as well as the level of engagement among people. Therefore, eye tracking is a viable approach for monitoring individuals' wellness [53].

Another popular category of sensor modalities in assistive technologies is electrochemical sensors. These sensors have a wide range of applications for monitoring physiological as well as environmental parameters. For instance, researchers in [54], [55] utilized electrochemical sensing to detect cortisol levels in sweat as an indication of stress level. Others utilized chemical sensing to analyze biosamples for detecting cancer precursors such as zinc ions [56]. Electrochemical methods also provide valuable insight about environmental parameters such as

air pollution and particulate matters [57]. Therefore, they facilitate a holistic approach for monitoring individuals wellness and how it is affected by various physiological and environmental parameters.

2.8. Electrochemical solutions for point-of-care devices

Since electrochemical sensors provide a rounded understanding of both physiological and environmental parameters and enable studying their effects on individuals' wellness, they provide a unique opportunity for integrating different aspects of health monitoring. Therefore, we examined this type of sensors in more detail.

Electrochemical measurements find extensive utility across scientific, technological, and everyday contexts, influencing various aspects of people's lives. They serve multiple purposes such as assessing food quality within supply chains [58], [59] evaluating human health through analysis of bodily secretions like salivary biomarkers [55], [60], identifying cancer precursors [56], monitoring air quality for toxic gases [61], as well as detecting heavy metals [62]. These applications empower individuals to make informed life style decisions, thereby enhancing their overall well-being.

For optimal utilization of electrochemical methods in diverse practical scenarios, it is crucial to employ them in compact, power-efficient, cost-effective, and preferably wearable devices. However, realizing these capabilities necessitates the development of miniaturized and economical electrochemical instruments as opposed to bulky and expensive laboratory equipment. In this pursuit, researchers have leveraged CMOS technology to craft small and wearable potentiostats [63], [64]. Significant strides have been made to broaden the current

readout range [65], reduce power consumption and device size [66], [67], and accommodate bidirectional current flow in electrochemical cells [66].

Despite the strides made in miniaturizing electrochemical systems, the reduction in feature size of modern CMOS technologies has led to diminished voltage supplies. For instance, while older 0.5 μm CMOS technology supported a 5 V supply, newer technologies like 180 nm only support a maximum of 1.8 V for regular transistors or 3.3 V for high-voltage transistors. Consequently, numerous electrochemical reactions cannot be sustained by contemporary integrated potentiostats. Moreover, since potentiostats must facilitate bidirectional current for redox reactions, only half of the supply voltage is available for each direction in an ideal rail-to-rail operation. With a 3.3 V supply, this translates to only 1.65 V for each reduction or oxidation reaction. Additionally, due to the necessity for the counter electrode in a standard three-electrode electrochemical cell to exceed the bias potential, only a fraction of this 1.65 V is usable as bias potential. However, many electrochemical reactions, such as those for detecting heavy metals like Mn, require bias potentials beyond the supported range [68]. Hence, conventional CMOS potentiostat designs implemented in newer technologies with lower supply voltages cannot support reactions for these elements. This opens a door for further research on empowering CMOS potentiostats with modern technologies to resolve the limited supply voltage issue. This advancement will allow a more versatile solution that can support a wider range of target elements in real-world applications.

2.9. Summary

Employing technology to improve individuals' wellness is of great interest. These technologies can be used to monitor different parameters that are indicative of individuals'

wellness. These parameters obtained from features extracted from audio data, visual data, physiological and environmental sensors, among others. These technologies help identifying parameters related to individuals' state of wellness including physiological and emotional wellness. Furthermore, they provide insight into interactions among people and how these interactions have mutual effects on participants' wellness. Moreover, these technologies can be deployed to monitor environmental parameters and study their effects on individuals' wellness, hence supporting more appropriate behavior to improve individuals' health.

The areas that require further research include efficient implementation of algorithms to enable resource-limited applications. Furthermore, providing higher temporal resolution, which enables the study of dynamic changes over time and allows real-time applications, is of high interest. Developing devices that are resource-efficient and tackling the limitations of translating laboratory instruments to wearable point-of-care devices is also of great importance for developing next-generation multifaceted assistive technologies.

Chapter 3: Methods and tools for analyzing social interactions

3.1. Introduction

Building rapport is an important element in having healthy and productive interactions in different situations, including in workplace environments or healthcare sets. This chapter presents the preliminary work on designing a framework for collecting data using sensors to infer human behavior and emotion and ultimately assess the rapport level in an interaction. The goal is to develop algorithms to process raw data and assess rapport building between dyads and leverage this information to enhance the quality of interactions in virtual meetings and overcome some of the shortcomings of virtual interactions compared to in-person setups. This chapter presents the work that has been conducted to converge knowledge across disciplines and identify suitable approaches and tools that can be utilized for analysis of important parameters in social ineractions. Different design iterations of the platform for conducting experiments and their design procedure are discussed. The analysis of tools and viable approaches for designing the aforementioned platform is reviewed. The algorithms that were developed, as well as methods to increase computational efficiency, are introduced. An analysis of the applicability of reinforcement learning (RL) for improving the platform is also presented. Finally, a discussion on how this preliminary work shaped the research path is discussed.

3.2. Sensor modality and data collection

To implement a platform for analyzing human behavior in a virtual environment, the first step is to collect raw data using sensors. This collected data will be processed down the line to infer human behavior and emotion. In this project, aiming at analyzing virtual meetings, camera data was employed for collecting visual data. Visual cues such as facial expression plays important

role in building rapport [69]. The goal of this project was to explore whether the visual data obtained by a camera can be utilized to extract information about affect and rapport. Camera provides rich data to work with for analyzing nonverbal cues. Moreover, in a virtual setup, a camera is often available, and thus, visual data can be obtained without the need for extra sensors, which makes this platform more widely accessible. In this work, different options for utilizing camera data were studied. Moreover, various features extracting from camera were studies to analyze human behavior.

3.3. Visual data for assessing affect and rapport

Literature suggests [69], [70] visual data such as direction of head and eye gaze as well as body pose including leg and arm posture are important elements in building rapport. Other nonverbal elements such as facial expressions are also important.

Among these visual cues, some elements like leg posture are not typically accessible in a virtual meeting. Some other features, however, can be collected using a camera in a virtual meeting which includes eye gaze, head movements and action units (AUs). AUs are the elements in the Facial Action Coding System (FACS) [71], [72] which is a system to taxonomize human facial expressions. This section introduces the analysis of tools for capturing visual data and presents an in-depth comparison of the options and how they can be integrated in a custom-built platform. This platform has been developed to collect data, process it and feed the processed data back to participants in a meeting.

3.3.1. Monitoring eye contact

Making eye contact is an important element of an effective communication [73]. It is an indication of engagement and attention levels of the audience. However, eye contact is missing

in a virtual environment, and therefore, participants miss an important cue in communication. Thus, one of the objectives of this research work was to utilize visual data to determine if participants in a virtual meeting were looking at each other and hence, they established "virtual eye contact" during the interaction. In this section, the options for monitoring eye gaze are analyzed and a comparison is presented. The methods that were developed for integrating eye gaze monitoring tools into our platform are explained.

The goal in this work was to implement a platform where participants in a virtual meeting can benefit without the need for an extensive setup on their end. For instance, utilizing special hardware/camera, which is often equipped with infrared detection and proprietary software, provides very accurate eye gaze data and improves the result of any analysis using this data. However, this special equipment is not typically readily available for users who use their laptops, for example, to attend a virtual meeting. Therefore, the objective in this research work was to limit the experiments to using hardware that is available to an average user, namely a webcam only. This constraint makes the platform usable without any special equipment and only require participants to install a piece of software that is developed for integrating different elements of the platform.

After analyzing different options, we chose GazePointer [74], which provides sufficient accuracy as an open-source software for detecting eye gaze on a regular 14" display. To facilitate this experiment, we developed a user interface in HTML as shown in Figure 3. 1.
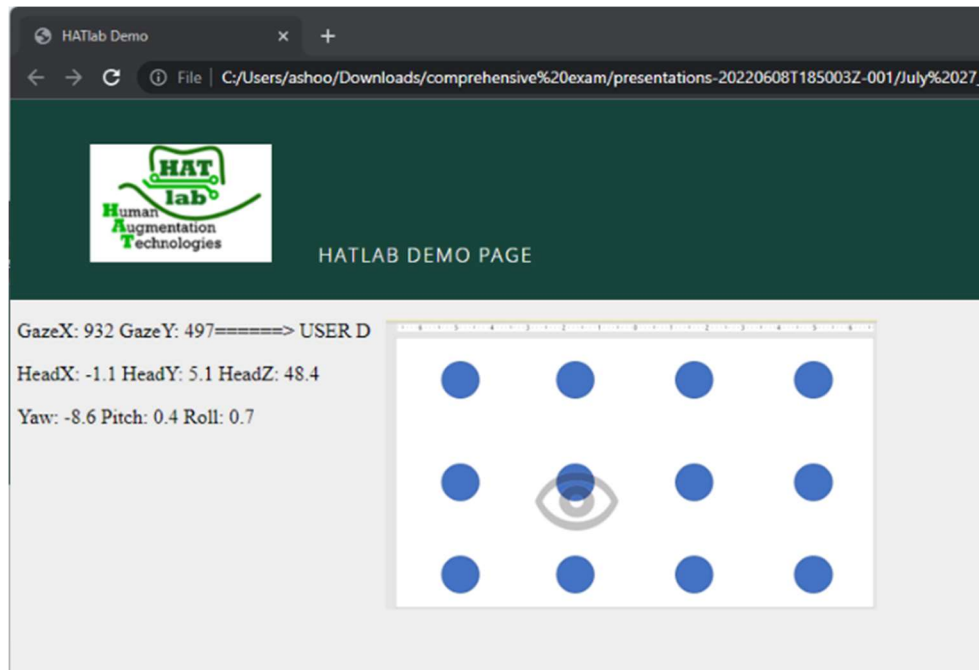
Figure 3. 1. Developed user interface in HTML that communicates with GazePointer and displays the coordinates of a user's gaze on the screen. The experiment was performed on a 14" display.

The HTML page communicates in the backend with GazePointer and displays the location at which a person is looking at on display. The HTML page also displays the coordinates of head location as well as yaw, pitch, and roll to monitor head movements. Despite the relatively good accuracy of GazePointer, a few disadvantages resulted in exploring other options to replace GazePointer. First, GazePointer requires a lengthy calibration process at the start of each session. Second, the result is very sensitive to the location of the head in front of the camera and might not be useful for a normal virtual session where participants move relative to the camera within the normal range of human movements. Third, it lacks support for analyzing prerecorded videos as well as videos that manifest multiple people. Finally, there is a lack of support for capturing action units (AUs), which are essential for detecting facial expressions [71]. It is worth mentioning for the case of real-time monitoring of participants in a virtual meeting, all different eye tracking

software require a dedicated camera because Windows settings do not allow two applications such as Zoom and GazePointer use one camera simultaneously.

3.3.2. Framework for processing action units and head movements using OpenFace

OpenFace [75] is open-source software that is slightly less accurate in eye gaze detection than GazePointer but without all the issues stated section 3.3.1, such as the need for a lengthy calibration process and the restriction of movements in front of the camera. This software provides an opportunity for seamless integration with the developed platform for analyzing the data using Python scripts. Besides eye gaze and head location/orientation, OpenFace provides information about action units. It also allows using recorded video and analyzes multiple people in one scene.

Since OpenFace allows the analysis of recorded videos, it facilitates analyzing the data from all participants on a single computer instead of analyzing data on each node. This method reduces the complexity of the experimental setup for each user as most of the heavy lifting is being done on a central node. Therefore, only a minimal software setup is required on each node. The approach to this method was recording the screen in small time frames and analyzing them immediately afterward. This method eliminates the need for a second camera for real-time analysis. The bottleneck, however, becomes the latency of processing the videos. Different methods for recording the screen and feeding the recorded data to OpenFace were explored. Namely, Python was employed to automatically use Camtasia [76], a third-party software, to record the screen and feed it back to Python. The Python script controls and sends commands to Camtasia using the command terminal. Another method that was used to achieve higher speed

was using Python directly for capturing screenshots and feeding the sequence of screenshots to OpenFace. The entire software backend was integrated and worked seamlessly and automatically. The Python script uses ZeroMQ [77] to communicate with OpenFace.

The rate of taking the screenshots and processing them through OpenFace was optimized to achieve the lowest latency. Figure 3. 2 shows the analysis for recording and processing the data for multiple subjects. The fastest solution we achieved was ~6 seconds latency for analyzing 4 people on the screen. This latency does not include the post-processing of our algorithms on the data obtained from OpenFace.
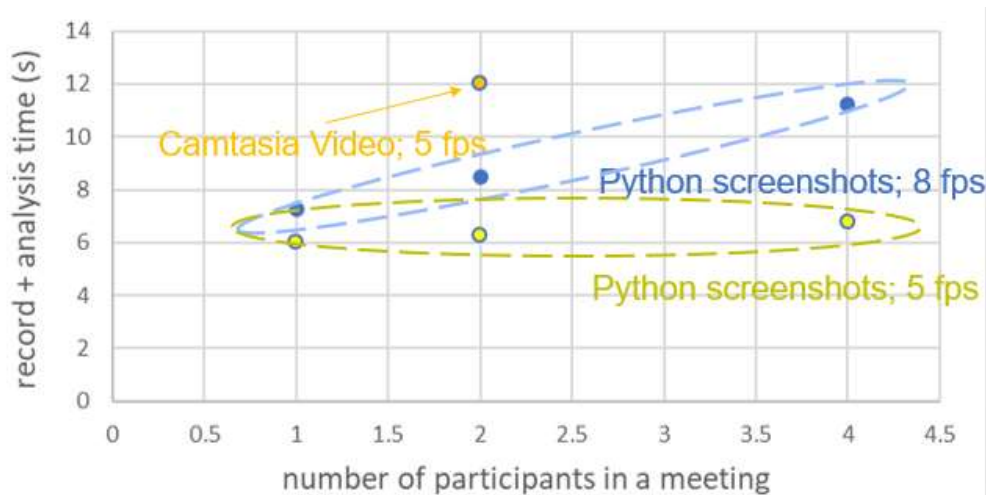

Figure 3. 2. Analyzing time delays on different methods of recording and processing the data using OpenFace. Our Python script uses ZeroMQ protocol to communicate with OpenFace back-end.

Besides relatively slow processing time, one downside of OpenFace is its low accuracy in eye gaze detection, especially in the vertical direction. To illustrate this limitation, Figure 3. 3 shows the result of the experiment where the eye gaze direction was assessed while looking at four corners of a 23-inch display. We were able to detect eye gaze in the horizontal direction with high accuracy, but the accuracy in the vertical direction was limited. This is a typical issue in eye

gaze detection systems as the movement of eyes in the y direction is more limited than in the x direction. Moreover, the vertical movement of the eyes is occluded by eyelids. This was a limitation that we observed in all the webcam-based eye gaze detection systems that we experimented with. It is worth mentioning that this experiment was focused on four extreme corners of a 23-inch display, and the result will be degraded when we want to use smaller screens or if we want to follow eye gaze in smaller range within a display. Therefore, without using specific hardware components for eye gaze estimation, we can only estimate the "virtual eye contact" for two people in a virtual meeting where the faces are displayed side by side on a screen.
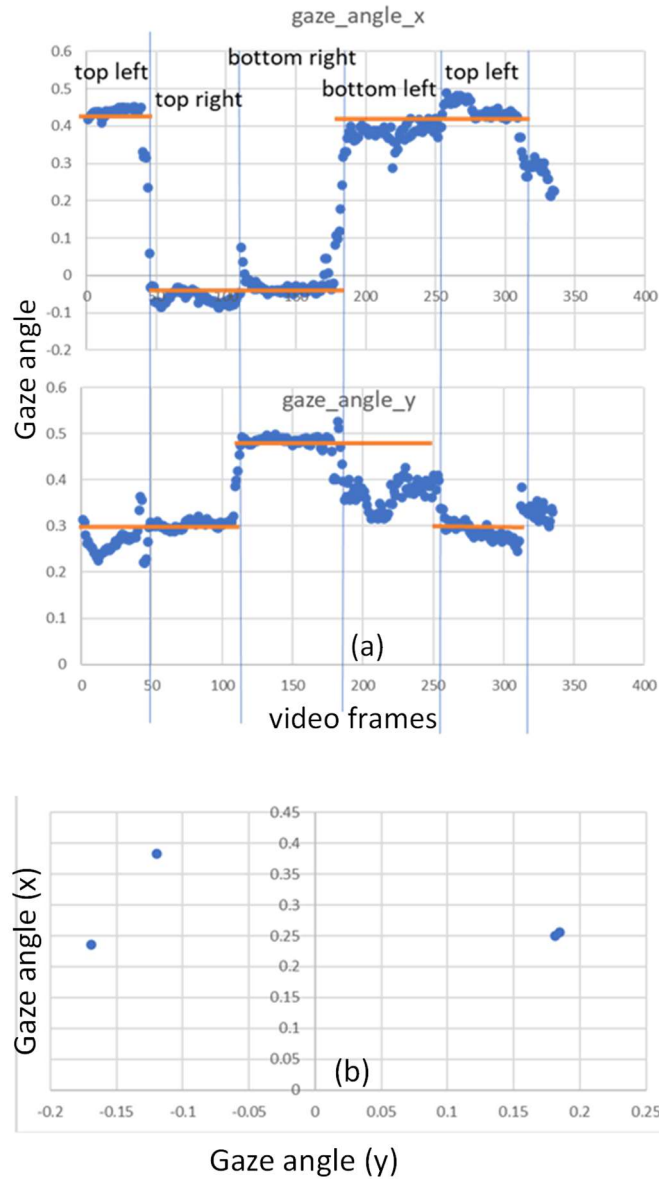
Figure 3. 3. Experimental results for estimating eye gaze when looking at four corners of a 23-inch display. (a) displays gaze angle in x and y direction vs video frames where the video was recorded at 30 fps. (b) shows XY coordinate of the estimated eye gaze on a 2D plane.

3.4. Preparing the platform for conducting experiments

In earlier sections, we discussed developing software for processing data and establishing communication between this software and other open-source software. We also briefly

26

discussed the web interface for running experiments for eye gaze detection. In this section, we present in more detail the considerations we had for developing the front-end and back-end for our experiments. The goal of this effort was to develop a dashboard where the data was effectively communicated to meeting participants. Efficient methods for data management and storing data in databases were also explored.

For visual display in the dashboard, the idea was to feed the processed data back to the users in an efficient and easy-to-understand way. The goal was for data not to distract participants, but rather let them take the key points away with a glance. Using Python, HTML and CSS, we developed the front-end of the display dashboard. We explored several options and ended up designing a split layout where the information is presented on both sides of the HTML page. The Zoom window was then laid out on top and in the middle of the page. We chose pie charts to include data about the participation of users in the conversation and bar charts for the dynamic of conversation between people. We also used Sankey diagrams to show the rapport building between each person and other participants as well as showing the level of affect for that participant. Figure 3. 4 shows an instance of the dashboard. The instances on the application, as well as the size of the text, are modifiable. Different options such as graphs with or without text and different layouts such as horizontal split of the display are selectable. Charts without text could convey the message with less distraction and are useful once the participant gets comfortable with the platform.

MySQL was used to manage the database running on a central node or server. The central node sends/receives data to/from all connected nodes through the local network. This

connection allows real-time communication between nodes and real-time updates of chart data on the dashboard display.

To improve the platform's accessibility, we also included color palettes suitable for color blindness cases. As for the networking, we used a central node to host the database and employed the MySQL protocol to connect different nodes. As of now, all the nodes need to be on a local network for the database to run effectively. Hosting the database on the web or a cloud to allow participants to connect from any location remains for future work.
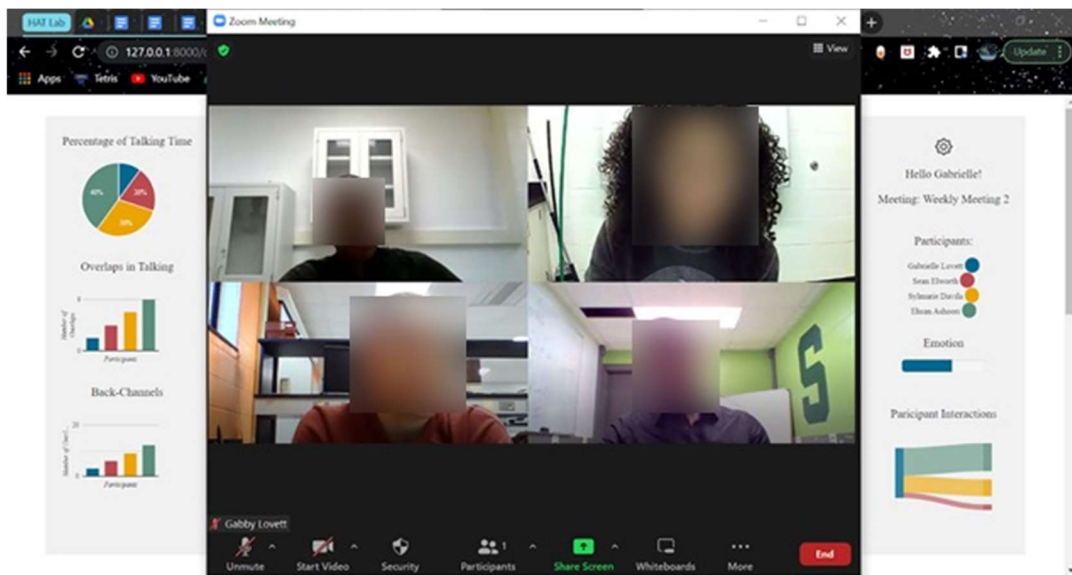

Figure 3. 4. Concept illustration of the split layout on the web interface with a Zoom window overlayed on top.

3.5. Exploring possibilities with reinforcement algorithms to enable person-specific recommendation

In the platform discussed so far, we ideally are interested in implementing a dynamic feedback system where the system analyzes the data and provides each participant with personalized recommendations to improve the quality of the interaction. To implement this

system, we intended to leverage reinforcement learning (RL) algorithms. Among all different RL methods, we are interested in methods that 1) do not require a model, 2) learn at each time step (as opposed to updating the parameters at the end of the experiment), and 3) provide control opportunity (as opposed to merely learning). After analyzing all the options with respect to these criteria, we created a short list of methods, namely SARSA, Expected SARSA, and Q-learning [78]. We analyzed the suitability of these options for this work. For example, Expected SARSA provides a more stable update target and lower variance, but it is more computationally expensive compared to SARSA as it requires the calculation of the expected estimate of the next action value for every state. Given the fact that we aimed for a real-time application where the speed of processing using a regular personal computer is a bottleneck, we chose the implementation of the SARSA method.

Another factor that is important is using approximate solution methods rather than tabular methods, as we typically do not have prior information about each possible state, and even if we did, constructing huge tables of states is not desirable in these applications. Moreover, using function approximation makes the learning methods applicable to partially observable problems which we expected to deal with when considering human involvement in our control system. Therefore, we chose to use approximation solution methods with neural networks (NN) as a function approximator.

To implement the quantized RL algorithm, we decided to apply the quantization technique to each main block of the algorithm and fine-tune it before implementing the complete solution. To this end, we started with the function approximator. To implement a quantized NN,

we took the standard NN problem of classifying hand-written digits and implemented our solution with the quantization technique.

Table 3.1. Classification accuracy of NN with quantized parameters.

Accuracy with:	Results
Original	~ 92%
Quantized weights	~ 88%
Quantized weights & activation functions (sigmoid(10x))	~ 85%
Activation function: sign Derivative: using sigmoid with a high coefficient	~50%

To begin with this phase, extreme cases of quantization which is binarization has been studies. Binarized version of NN was investigated first since its implementation is simpler than multilevel quantization and also supportive literature [79] exists on this topic. In this method, all the weights of the NN were replaced by -1 and 1. For the activation function, sign function is ideal but its derivative function causes problems in training a NN as it is equal to zero almost everywhere. Therefore, to mimic the sign function, we used sigmoid function with a large multiplier in its argument. This allows us to simulate the sign function behavior while still be able to take derivative of the activation function in backpropagation procedure. As shown in Table 3.1, the accuracy of 85% was obtained while all the weights have been binarized and the activation functions were nearly binarized as discussed above. This served as proof of concept that high precision is not always necessary for weights and activation functions in a NN. In the

next step, we replaced sigmoid functions with actual sign function to fully binarize the network. The accuracy of this design dropped to ~50%. To improve the results, we explored straight through estimator for implementing back propagation. In this method [79], the derivative of cost function ($\frac{\partial J}{\partial \theta}$) during backpropagation gets replaced by $\frac{\partial J}{\partial g} * 1_{|x| \leq 1}$ where J is the cost function, Θ is the weights and g is the activation function. What this means is that instead of calculating the derivative of the cost function with respect to the weights (Θ), the derivative is calculated with respect to the activation function which in this case is a sign function. Then the result is multiplied by $1_{|x| \leq 1}$ which equals to 1 in the vicinity of origin and 0 everywhere else. In other words, the backpropagation is calculated with this assumption that derivative of the sign function is 1 close to origin and 0 elsewhere.

To calculate the derivative of cost function with respect to the activation function, the following formula was derived. Given that the forward path is represented by (3.1),

$$out = g\left(a^{(1)}.\theta^{(1)}\right).\theta^{(2)} \ , \tag{3.1}$$

We assumed :

$$\frac{\partial out}{\partial g} = \theta \tag{3.2}$$

And we know:

$$(f^{-1})'_{(x)} = \frac{1}{f'(f^{-1}_{(x)})} \tag{3.3}$$

Applying the chain rule and (3.3), the following expression was obtained.

$$\frac{\partial J}{\partial g} = -\frac{1}{m}\sum\sum\left\{\left(\theta^{(2)}\log\left(h_{\theta(x)}\right) + h_{\theta(x)}\frac{\theta^{(2)}}{h_{\theta(x)}}\right) - \theta^{(2)}\log\left(1 - h_{\theta(x)}\right) + 1 - h_{\theta(x)}\frac{-\theta^{(2)}}{1-h_{\theta(x)}}\right\} +$$

$$\frac{\lambda}{2m}\sum\sum\sum\theta\frac{\partial h_{\theta(x)}}{\partial g} = -\frac{1}{m}\sum\sum\left[\theta\left(\log h_{\theta(x)}\right) - \log\left(1 - h_{\theta(x)}\right)\right] + \frac{\lambda}{m}\sum\sum\sum\theta \tag{3.4}$$

31

(3.4) was plugged into the algorithm. The accuracy of the results did not improve, but the computation overload was much less because the operations for calculating the derivative in backpropagation were just replaced with simpler mathematical equations as discussed above in (3.1) - (3.4).

Despite the potential we see and the progress we made with applying Reinforcement Learning to this problem, a major bottleneck remained the amount of data we needed for training the algorithms. Upon further investigation on implementing SARSA, we noticed with current human subjects and experiments, it was not feasible for us to follow this path for now. However, it remains a viable path to pursue in the future.

3.6. Conclusion and discussion

Given the findings in the preliminary work presented in this chapter, we organized the bulk of this thesis work on the following topics. The core of behavior monitoring in this framework is having a reliable assessment of the "affect" level of each individual. In psychology, affect is described as "the underlying experience of feeling, emotion, attachment, or mood" [80]. A valid assessment of affect in individuals may indicate the effect of the conversation on individuals over the course of a meeting. Such an assessment also gives clues to other participants, which may be utilized to improve the quality of interaction. Another important factor in assessing the dynamic of any conversation is assessing the rapport building between participants. Monitoring the rapport building between dyads in a conversation is a very important parameter that gives an understanding of the quality of interaction. Therefore, the next chapter is focused on assessing the effect and rapport in virtual conversations to foster higher quality interactions in virtual

meetings and improve the well-being of the participants. This, in return, facilitates the

productivity of meetings or online learning setups.

Chapter 4: Developing platforms for monitoring affect and rapport

4.1. Introduction

To ensure productive work meetings and effective learning environments, it's crucial to pay attention to non-verbal cues from our audience such as auditory or visual cues. These cues offer insights into people's emotional states and engagement levels, facilitating more impactful communication. However, individuals vary in their social intelligence, affecting their ability to interpret these cues accurately. This discrepancy directly influences the quality of interpersonal interactions.

This challenge is amplified in virtual settings, where remote collaboration has become increasingly prevalent. Despite advancements, virtual platforms often lack the richness of in-person interactions, such as eye contact and body language observation, making communication less effective. Consequently, there is a growing need for technologies to enhance communication and interaction effectiveness in virtual environments.

Recent literature reflects a surge in interest in developing technologies capable of discerning people's emotional states as well as rapport building among individuals. Many of these approaches rely on computationally intensive deep neural networks, limiting real-time implementation, especially with constrained computational resources. Alternatively, some studies utilize machine learning algorithms requiring less computation but necessitating manual feature engineering, adding complexity. Moreover, these methodologies often operate within controlled lab environments, where experiments induce specific emotions, resulting in a higher signal-to-noise ratio than natural interactions. As a result, algorithms developed under these conditions may exhibit reduced performance in more natural settings.

In this work, we developed a framework that utilizes neural networks to analyze the individuals' affect and rapport building in groups during virtual meetings. The contributions of this work are as follows:

- Analyzing affect and rapport where individuals are holding regular work meetings in a natural setup without acted sessions.

- Classification of subtle changes towards positive or negative affect as opposed to extreme cases.

- Analyzing affect and rapport with high temporal resolution, which enables providing real-time analysis and feedback.

- Analyzing affect and rapport in multiperson groups.

- Implementing neural network with minimum number of layers and input nodes by reducing the feature space and using raw features as opposed to hand crafting the features.

To the best of our knowledge, this work is the first to achieve high accuracy while satisfying all the requirements specified above. In this chapter, we present the details of our methods in dataset pteparation and data analysis. The results for analyzing rapport and affect are presented in different subsections. We conclude the chapter with summarizing the work and discussion on our findings.

4.2. Background and related works

4.2.1. Related work on affect monitoring

According to the American Psychology Association, together with 'cognition' and 'conation', affect is one of the three identified components of mind [81]. According to [81], affect

is defined as "any experience of feeling or emotion, ranging from suffering to elation, from the simplest to the most complex sensations of feeling, and from the most normal to the most pathological emotional reactions. Often described in terms of positive affect or negative affect, both mood and emotion are considered affective states." Reported works on affect recognition in the literature use multiple modalities of sensors and different types of machine learning algorithms some of which were explained in chapter 2. These modalities include audio, visual and natural language. In one study [39], researchers utilized deep neural networks to analyze audiovisual data for affect recognition, showcasing a significant improvement in emotion recognition performance compared to traditional methods reliant on handcrafted features. Similarly, another study [40] employs deep recurrent neural networks to analyze speech. Despite demonstrating promising results, these approaches are computationally intensive, which hinders their real-time application where computational resources are constrained.

Another work in [16] compares using logistic regression with linear support vector machine (SVM) to analyze videos. By extracting the action units (AUs) from the videos and analyzing them with these classifiers, the researchers were able to recognize disrespectful interactions with accuracy of ~62%. In [82], the researchers built on [16] and by adding audio features such as pitch and intensity, an accuracy of 79.86% was achieved in detecting disrespectful vs respectful interactions using logistic regression model.

Other researchers focus on visual data to gauge the engagement intensity of individuals in various settings, such as classrooms [9], [10], [25], [26]. Additionally, cues such as head motion synchronization and empathy in face-to-face communication are investigated using

accelerometers in a lab environment [27], revealing that empathy levels correlate with the frequency and phase of head motion synchronization.

This collective body of research motivates us to explore the potential of using visual features in natural settings, as opposed to controlled lab experiments, to analyze audience affect. To achieve this without relying on handcrafted features and to reduce computational complexity compared to deep neural network approaches, we opt to implement neural networks with only one hidden layer. Our objective is to classify affect in a natural meeting environment where extreme emotions are less prevalent, training the classifier to detect subtle shifts in participants' emotions. Throughout this dissertation, by "participants", we refer to people who were subject to monitoring their emotions and behavior and not the labelers and the study team who designed and conducted the experiments

4.2.2. Related work on rapport monitoring

Rapport, the establishment of a harmonious and empathetic connection between individuals, lies at the heart of effective communication and collaboration across diverse contexts from personal to professional interactions. Particularly, it is the cornerstone for building productive and impactful meetings across professional settings. Rapport encompasses the establishment of trust, understanding, and mutual respect among participants, fostering an environment conducive to open communication, collaboration, and creativity. The presence of rapport can greatly influence the dynamics of a meeting, shaping the level of engagement, the quality of discussions, and ultimately, the outcomes achieved. Research has consistently highlighted the significant impact of rapport on team performance, decision-making processes, and overall meeting effectiveness [83]. In this context, recognizing the importance of rapport and

its role in facilitating productive meetings is essential for organizations seeking to optimize their communication strategies and maximize team synergy.

Building rapport in virtual environments presents unique challenges compared to face-to-face interactions [84]. One of the primary obstacles is the lack of non-verbal cues, such as body language and eye contact, which are integral to establishing trust and connection. In virtual meetings, participants may find it challenging to interpret subtle cues or accurately gauge the emotions and intentions of others, leading to potential misunderstandings or miscommunication. As a result, individuals may struggle to develop the same level of rapport in virtual environments, requiring deliberate efforts and strategies to overcome these challenges effectively.

A body of work in this area focuses on recognizing rapport levels in the human interaction with a virtual agent [31], [32]. Others aim to analyze rapport levels in human-to-human interaction [29], [85] in dyadic pairs. Despite all the advances in this area, interpreting rapport in high temporal resolution and among multi-person groups is missing in the literature, leaving rapport analysis for dyadic conversations that is done for an entire session of interaction (as opposed to fine temporal resolutions). Therefore, granular information about the dynamics of conversations is absent in these studies. The focus in this work is analyzing rapport in multiperson groups and with high temporal resolution.

4.3. Dataset preparation

4.3.1. Collecting and preparing the data for analyzing affect

To perform this experiment, we recorded five work group meetings with the duration of approximately 40 minutes to 100 minutes with an average of ~62 minutes each. The first two meetings had 5 participants whereas the last three had 4 participants, both male and female.

Figure 4.1 shows a snapshot of one of these recording sessions in Zoom [86]. This study has been reviewed and approved by the Institutional Review Board (IRB) office at Michigan State University.

Figure 4. 1. An example of the setup for collecting data during a virtual meeting using Zoom. © 2023, IEEE.

The labels were assigned to each segment in which a participant was speaking. To simplify the labeling process, a Matlab script was developed to identify the conversation segments in each recording and generate time stamps accordingly. The recorded video files were cut into 2522 segments using the MATLAB script based on the generated time stamps. For labeling the cut segments, a graphical user interface (GUI) was developed using the App Designer tool of MATLAB 2019b. This GUI plays each segment one by one and lets the labeler assign the proper label from within the same GUI. Figure 4.2 shows the appearance of this GUI which greatly speeds up the labeling process [87].
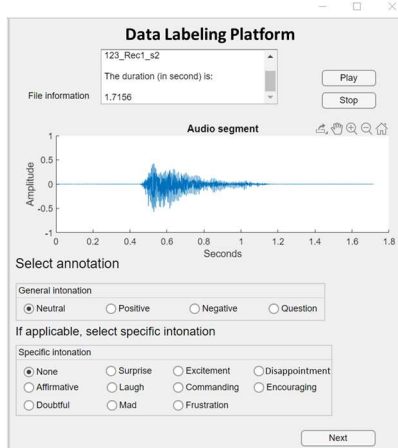
Figure 4. 2. Developed graphical user interface for labeling 'affect' [87]*.*

4.3.2. Labeling the affect dataset

Three labelers were trained to label the affect level of each participant for each segment of the recorded meeting. The GUI introduced in the prior section was used to assist with labeling. The labelers could play each segment and label it from within the app. The segments were played both in the order of occurrence and randomly and labeled separately. In this work, however, we focused on the labeling that was done in order and left the analysis of labeling the segments in a random fashion for the future work. The labelers were instructed to label the segments as positive, neutral, or negative. The app outputs a text file which contains the labels for each video segment. After completion of the labeling process, the labels that had a majority agreement among the labelers were kept, and the rest were disregarded. As seen in Figure 4.3, in 59.9% of datapoints, all three labelers assigned the same labels. For 36.8% of datapoints, only two labelers assigned the same label. And in 3.3% of datapoints, none of the labelers assigned the similar label. Therefore, we kept 96.7% of datapoints to which at least two labelers assigned a similar label and disregarded the remaining 3.3% of datapoints.
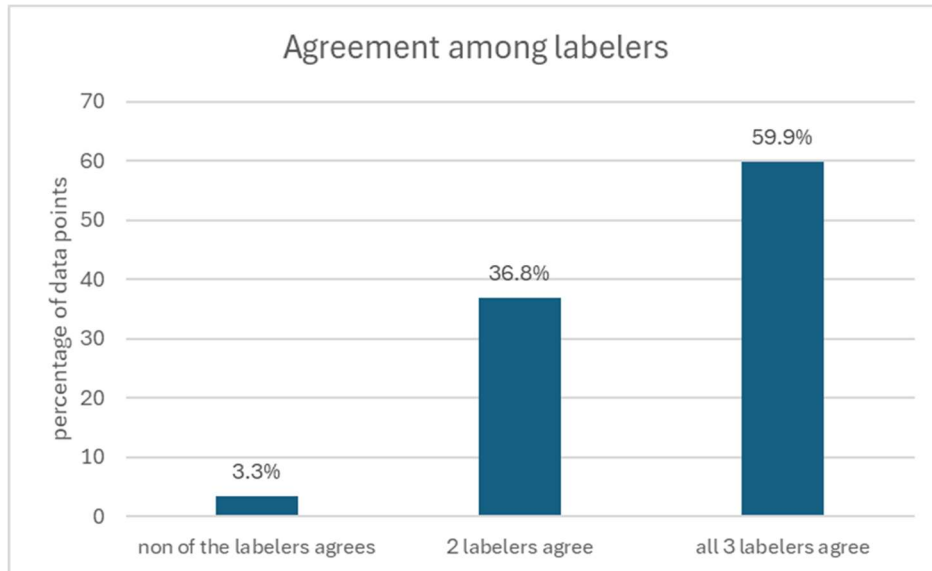
Figure 4. 3. percentage of datapoints that labelers agreed on a label.

A Python script then reads the labels and assigns them to the corresponding features and makes the dataset ready to be used with the classification algorithms. The details about features and algorithms will be presented in the methods section later in this chapter.

4.3.3. Collecting and preparing the data for rapport monitoring

For this experiment, we recorded twenty meeting sessions, of which eight meetings had three participants and twelve meetings had four participants. The duration of the sessions was between 18 minutes and 30 minutes, with an average of ~22 minutes. A total of 35 participants (people who were recorded and not the labelers and the study team who designed and conducted the experiments) were recruited for holding meetings. Tables 4.1 and 4.2 show the gender and age distribution of the participants. More than 90% of participants were in the age group of 18-34 years old.

Table 4.1. Gender distribution of participants.

Gender	Count
Male	21
Female	12
Other	2
Total	35

Table 4.2. Age distribution of participants.

Age group	Percentage
25 - 34	48.57%
18 - 24	42.86%
35 - 44	8.57%
Total	100.00%

Tables 4.3 and 4.4 show the distribution of the participants' education level and occupation status. More than 90% of participants had passed at least some college courses or had college or professional degrees.

Table 4.3 . Distribution of education level of participants.

Education level	Count
Some college, no degree	10
Bachelor's degree (e.g. BA, BS)	10
Master's degree (e.g. MA, MS, MEd)	10
Doctorate or professional degree (e.g. MD, DDS, PhD)	3
High school degree or equivalent (e.g. GED)	1
Do not wish to answer	1
Total	35

Table 4.4 . Distribution of occupation status of participants.

Occupation status	Count
Student	29
Employed full-time (40 or more hours per week)	3
Employed part-time (up to 39 hours per week)	3
Total	35

The 3-person groups formed three dyadic pairs and the 4-person groups formed six dyadic pairs for the purpose of analyzing rapport in these groups. Figure 4.4 shows a snapshot of one of these recording sessions in Zoom. This study has been reviewed and approved by the institutional review board (IRB) office at Michigan State University.
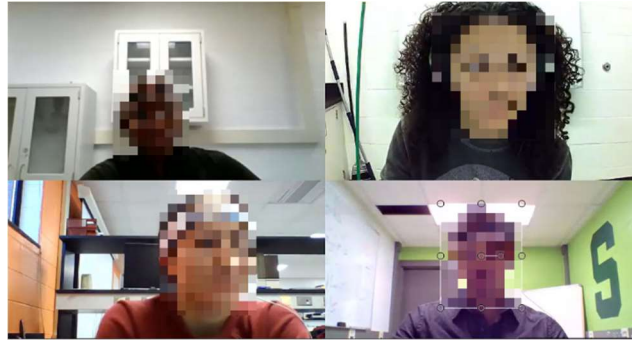
Figure 4. 4. An example of the recording session for collecting data during a virtual meeting.

To facilitate the labeling process, the Matlab script that was mentioned in section 4.3.1 was used to segment video files in 30-second windows as we were interested in analyzing the rapport in fine-grained time segments.This would allow us to analyze the dynamics of interactions during each session. For labeling these video segments, a graphical user interface (GUI) was developed using Python and PyQt. This GUI plays each segment of videos one by one and lets the labeler assign the proper label from within the same GUI. This GUI has features that facilitate faster and easier labeling processes, such as navigating between segments or skipping some segments. Figure 4.5 shows the appearance of this labeling assistant GUI which greatly smoothed the labeling process.
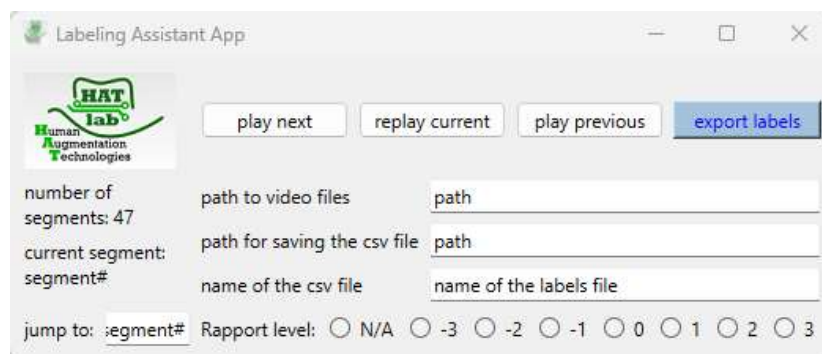

Figure 4. 5. Developed graphical user interface for labeling rapport.

4.3.4. Labeling the rapport dataset

Four labelers were recruited to label the 30-second segmented videos. The labelers were instructed to label each dyadic pair, three or six pairs for three-person and four-person groups, respectively. They were instructed only to label the dyadic pairs in which at least one person speaks for more than 10 seconds. The labelers were provided with the definition of rapport. To have consistency among the labelers, they were instructed to look for the parameters shown in Table 4.5. These parameters are derived based on the literature presented in [88], [89] and with the method introduced in [87]. However, they were instructed not to overemphasize these parameters and to rely on their first impressions and general intuition to gauge the rapport building in the groups.

Table 4.5. Parameters of interest in gauging rapport.

Well-coordinated	Boring	Cooperative	Harmonious	Unsatisfying	Uncomfortable
Cold	Awkward	Engrossing	Unfocused	Involving	Intense
Unfriendly	Active	Positive	Dull	Worthwhile	Slow

Given the subjective nature of these labelings, and based on our previous experience that many of the labelers tend to label instances as 'neutral', we labelers to label each segment on seven-point Likert scales where -3 represented extreme negativity and +3 represented extreme positivity as shown in the GUI in Figure 4.5. This was purely meant to have more labels other than 'neutral' or zero and then we binned these labels in only two classes of high and low rapport based on the statistical analysis presented in the Method section.

4.4. Method and results

4.4.1. Extracting Facial Action Units

In this work, we used facial action units (AUs) as features to analyze 'affect'. AUs are representative of movement of individual facial muscles and are commonly used as indicative of expression of emotions [90], [91]. Figure 4.6 shows some examples of action units [92]. To extract AUs, we used OpenFace [93], [94], [95], an open source software widely used by the community. OpenFace extracts a subset of AUs comprising intensity of 17 different AUs. These 17 features were used for classifying different affect levels in the various virtual meetings.

Figure 4. 6. Examples of facial action units [92].

4.4.2. Classification of affect

As described earlier, for classifying affect, the video segments were labeled as positive, neutral, and negative. We trained our classifier only on positive and negative labels as they are more reliably classified. In order to avoid developing the bias toward any of the classes during training, we balanced the dataset to have equal number of datapoints for positive and negative

46

labels. For the classification, to avoid manually crafting the features for the algorithms, we chose to use neural networks as opposed to other machine learning algorithms such as logistic regression or SVM. Moreover, to avoid heavy computational load as well as reducing the chance of overfitting, we chose to implement a neural network with only one hidden layer. We first implemented the neural network with all the 17 AUs as the input features. The videos were recorded at the rate of 30 fps. Although the change in the facial expression is not fast and our analysis does not need 30 fps, we did not downsample the recordings for now. Downsampling could be further investigated in the future. We just used the fully recorded data for this experiment. For each AU, the code takes the average of the values over the span of the start to stop time of each video segment. Therefore, for each AU, one value is assigned to each video segment. With all the 17 AUs used in the classifier, the design suffered from significant variance where despite using regularization, training accuracy of 92.9% was achieved while the testing accuracy was only 60.7%. To solve this problem, we employed principal component analysis (PCA). As shown in Figure 4.7, to retain at least 80% of variance, we projected the feature space to only 10 features and reformed the neural network with 10 input features and 10 nodes in the hidden layer.
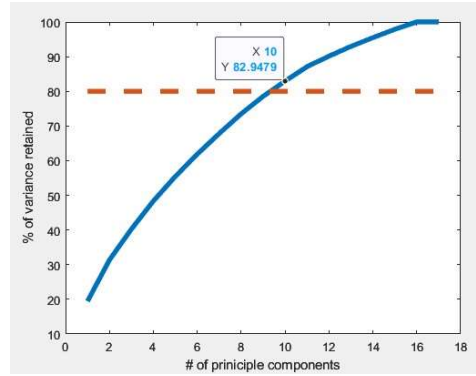
Figure 4. 7. Retention of variance vs number of principal components. By choosing 10 principal components, more than 80% of the variance has been retained © 2023, IEEE.

4.4.3. Results for classifying affect

The neural network was trained with 4-fold cross validation and the training accuracy and testing accuracy of 81.1% and 76.8% were obtained, respectively. Considering that we performed our experiments on natural setups without any constraints on participants, and considering that we did not aim at only classifying extreme cases such as disrespectful moments, 76.8% testing accuracy using a neural network with only one hidden layer is achieved. To the best of our knowledge, this result has been achieved for the first time in literature and paves the path toward real-time analysis of virtual meetings using local computational resources on a typical laptop for example. Table 4.6 summarizes the results of these experiments.

Table 4.6. Training and testing accuracy of affect with and without implementation of PCA © 2023, IEEE.

	Training accuracy	Testing accuracy
Full feature space (17 AUs)	92.9%	60.7%
Reduced feature space (PCA with 10 components)	81.1%	76.8%

4.4.4. Extracting gaze and head orientation

For the purpose of analyzing rapport, we extracted eye gaze and head orientation as well as head coordination in addition to AUs. These features were extracted using OpenFace as well. We were interested not only in these features, but more so on the synchrony of these features among dyadic pairs which is a more indicative of rapport building in groups. As a measure of synchrony, we used dynamic time warping (DTW). DTW is a measure of similarity between two temporal sequences. Similar to Euclidean distance, it measures the distance between two vectors. However, unlike Euclidean distance, it does not measure the distance between two vectors point by point. It takes into account the distance between neighboring points and chooses the minimum value for each point. Figure 4.8 shows the comparison between Euclidean distance and DTW. This method is widely used in applications such as language processing. It can be used for comparing two instances of data that are noisy or have different lengths. For instance, the similarity between one sentence pronounced by two people can be identified using DTW.
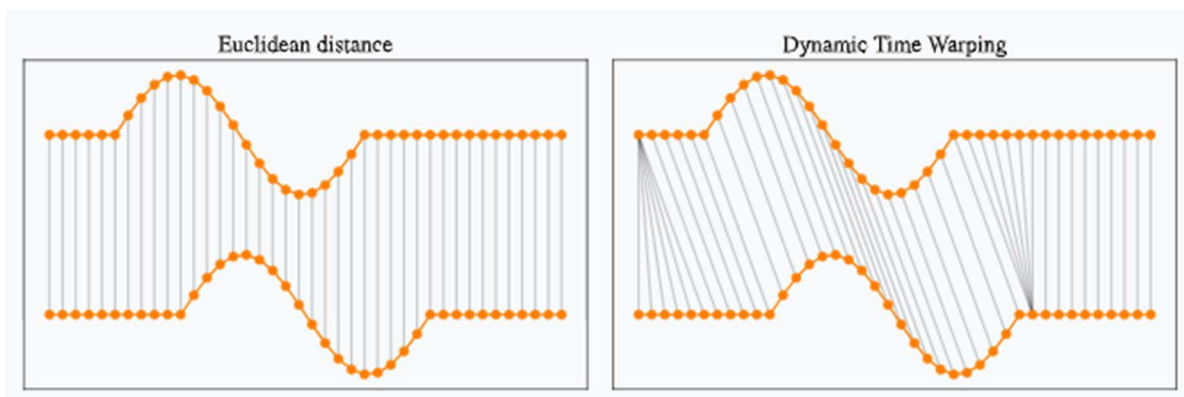

Figure 4. 8. Comparison between Euclidean distance and DTW [96] *.*

In this work, synchrony features were constructed for each of the base features, namely, eye gaze, head orientation, head coordination and action units. We used a Sakoe-Chiba band of

three seconds to calculate DTW and used it as a set of input features to our algorithm. The full

list of features are shown in Table 4.7.

Table 4.7. Full list of features for analyzing rapport building in groups.

Category	Comment	Number of components
Eye gaze	x, y, z coordinates for each eye angle (x, y) multiply by 2 (for each pair)	8 x 2 = 16
Head coordination	x, y, z coordinate multiply by 2 (for each pair)	3 x 2 = 6
Head orientation	Yaw, pitch, roll multiply by 2 (for each pair)	3 x 2 = 6
AUr	Intensity of AU multiply by 2 (for a pair)	17 x 2 = 34
AUc	Presence of AU multiply by 2 (for a pair)	18 x 2 = 36
DTW	Constructed between two participants (in each pair) for gaze, head orientation and coordination, AUr and AUc	5
Total		103

4.4.5. Classification of Rapport

The statistics of the rapport labels are presented in Figure 4.9. 2674 valid labels were generated. The main challenge in this dataset is the imbalance of data among classes. Therefore, we chose to bin high and low rapport in a way to have a more balanced dataset. This greatly helped train the algorithms. This graph also shows that the choice of having a seven-point scale for labeling helped the labelers to identify more instances of 'minor positivity' (indicated by scale '1'), which otherwise would be labeled as neutral and would skew the dataset massively.
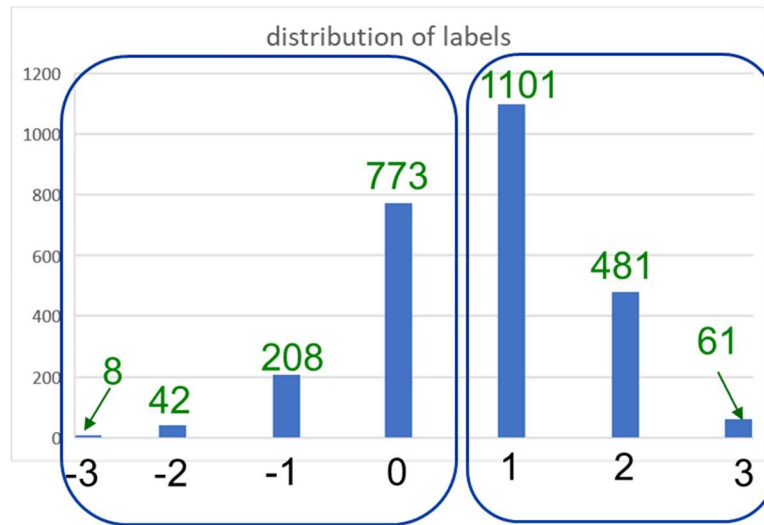
Figure 4. 9. Distribution of the dyadic rapport labels. The boxes show the group of labels used for high and low rapport.

4.4.6. Results for classifying rapport

A neural network with a single hidden layer and two output classes was implemented for classifying high and low rapport. The first experiment was done by the full feature set. Given a total of 103 features were used in this experiment, the rule of thumb for the required number of data points is:

$$N = 10 \times M/\alpha \qquad (4.1)$$

51

where N is the number of data points, M is the number of parameters and α is overparametrization ratio. We used 10 nodes in the hidden layer, therefore the number of parameters, M, is calculated as follows:

The number of parameters between the input and the hidden layer = 103 x 10 + 10 = 1040. Note that we added ten parameters for the bias nodes. Likewise, the number of parameters between the hidden layer and the output layer = 10 x 2 + 2 = 22. Therefore, a total of 1062 parameters should be trained. Let α = 5, from (4.1) we know at least N =10 x 1062/5 = 2124 data points are needed for training the network. Given the total number of 2674 data points we had, the concern was overfitting and our experiment results shown in Figure 4.10 confirms it as seen in the data for full set of features (the last two bars in the figure). Therefore, the feature space should be reduced to achieve more generalization. To this end, a study on each feature types were performed to identify the most relevant features. The idea was to keep the five DTW features as they are indicative of synchrony between participants. Then, each type of feature was added to the analysis to examine their impact on the results. The accuracy results for rapport classification of the dyads are shown in Fig. 4.9. The experiments were repeated 20 times, and each time, initialization of parameters was repeated in the training process. The same 80% of data chosen randomly was used for training and the remining 20% of the data for testing purposes in all these 20 runs. The average, along with the maximum and minimum error bars, are shown on the graph for different combinations of features as well as for the full set of features.
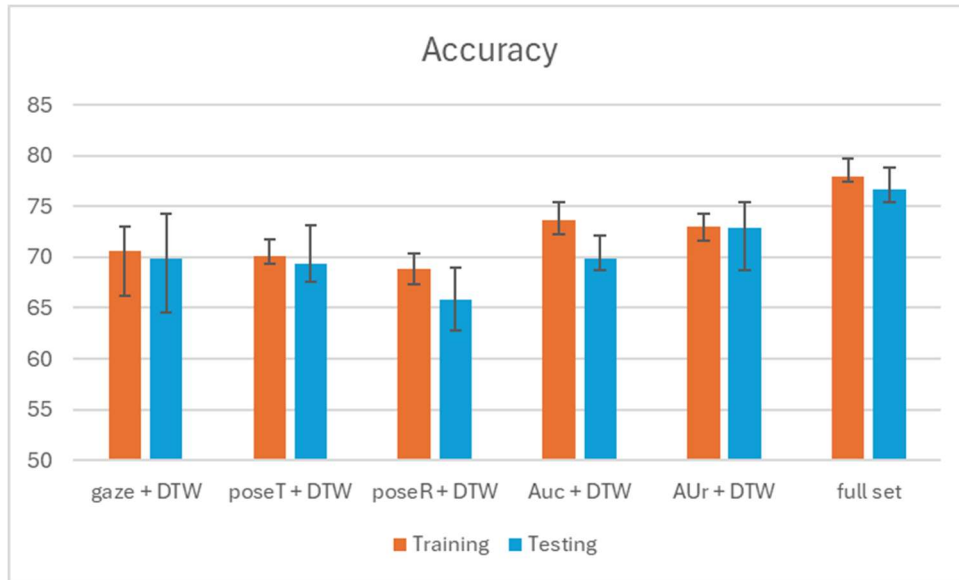
Figure 4. 10. Training and testing accuracy of rapport for different features. The right most column shows the result for the full set of features.

Since accuracy measures how often a classification model is correct in general, it is not a good metric when a dataset is imbalanced among different classes [97]. Since the dataset in this experiment was not fully balanced, precision and recall were calculated as well. Precision in this case is the measure of what portion of the items that have been detected as high rapport are correctly predicted [98]. In other words, it shows how often is the prediction correct when predicting a target class [97]. The formula for calculating precision was:

$$Precision = Tp/(Tp+Fp) \tag{4.2}$$

where Tp is true positive and Fp is false positive among the predicted labels. Recall measures what portion of all high rapport data points is correctly predicted [98]. In other words, it is a measure of how well all the instances of a target class is predicted [97]. It was calculated using the following formula:

$$Recall = Tp/(Tp+Fn) \tag{4.2}$$

53

where Fn represents false negative predictions. Figure 4.11 and Figure 4.12 represent the results
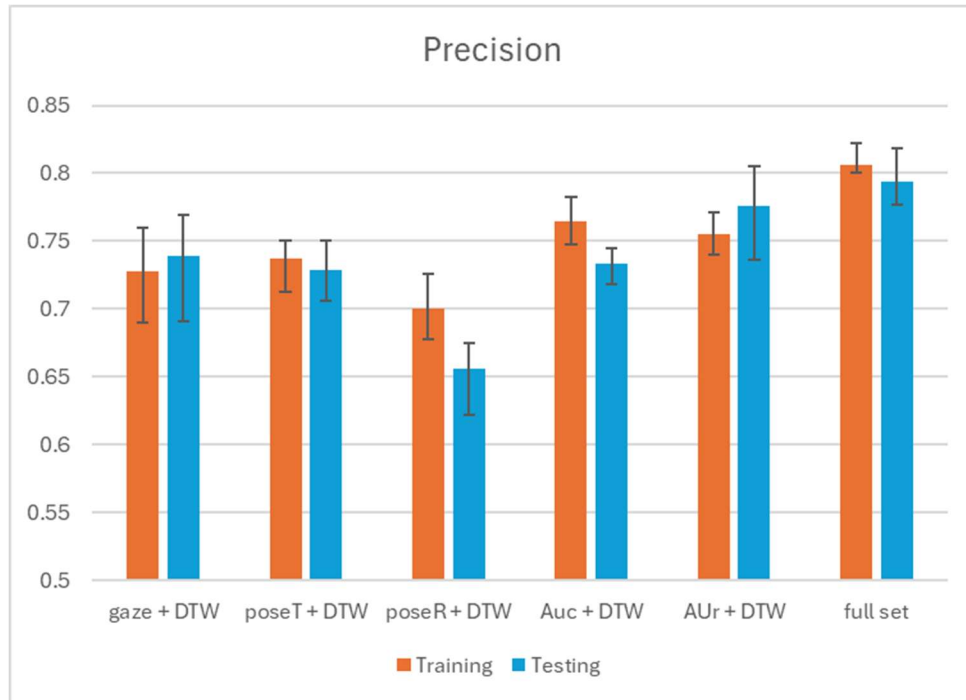
for precision and recall in these experiments

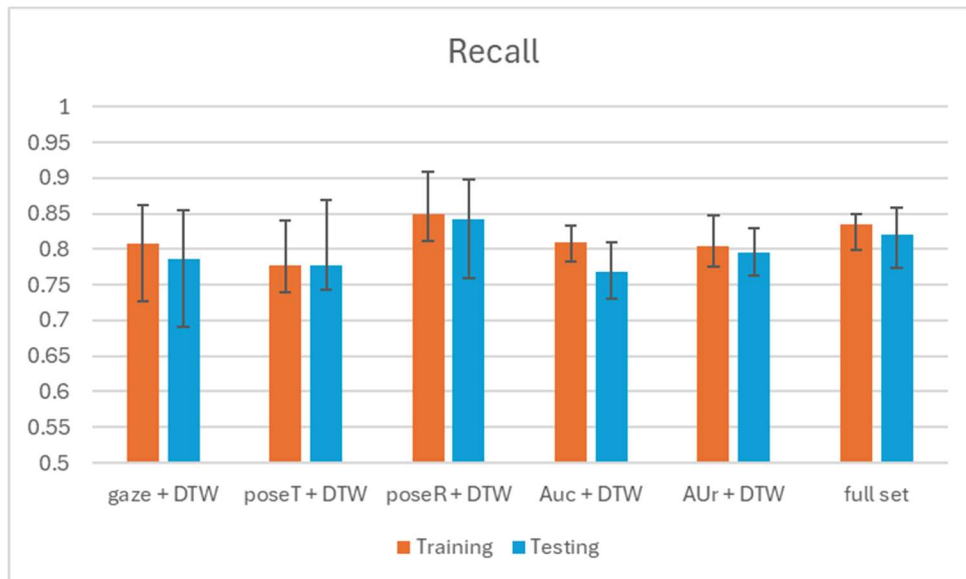*Figure 4. 11. Precision for the rapport classification.*

Figure 4. 12. Recall for the rapport classification.

Another commonly used metric that often is used to take into account both precision and recall is F1 score [99]. This metric was calculated using the following formula, and the results are shown in Figure 4.13.

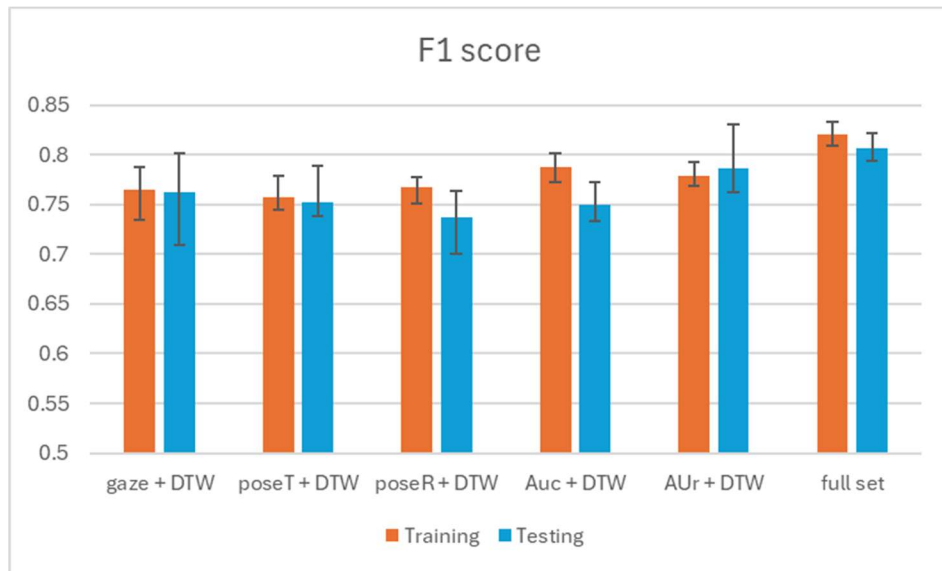$$F1 = 2*(precision*recall)/(precision+recall) \qquad (4.3)$$

Figure 4. 13. F1 score for the classification results.

According to the results of our experiments, we noticed that head coordination (poseT) and presence of action unita (AUc) are more significant than head orientation (poseR) and intensity of action units (AUr), respectively. Therefore, we did not include them directly as independent features. However, they still indirectly contribute to the classification because DTW features derived from those metrics have been utilized in the feature space. By eliminating these features, the number of features reduced from 103 to 63, which would help to the generalization of the classifier. These 63 features include eye gaze, head coordination and AUc, for both individuals constructing a dyadic pair. They also include five DTW features for each of eye gaze, head orientation, head coordination, AUc and AUr.

Using the newly constructed feature space, the classifier was trained on 80% of randomly selected data and was tested on the remaining 20% of data. This process was repeated 20 times, each time with a new subset of 80/20 data. The results for average and standard deviation of accuracy for both the full and reduced feature spaces are presented in Figure 4.14. As depicted in this figure, the difference between average accuracy of training and testing is smaller for the reduced feature space compared to that of the full set of features. The standard deviations of accuracy also follow the same pattern. This confirms the more generalized solution while achieving high accuracy of 73.6% for the testing experiment.
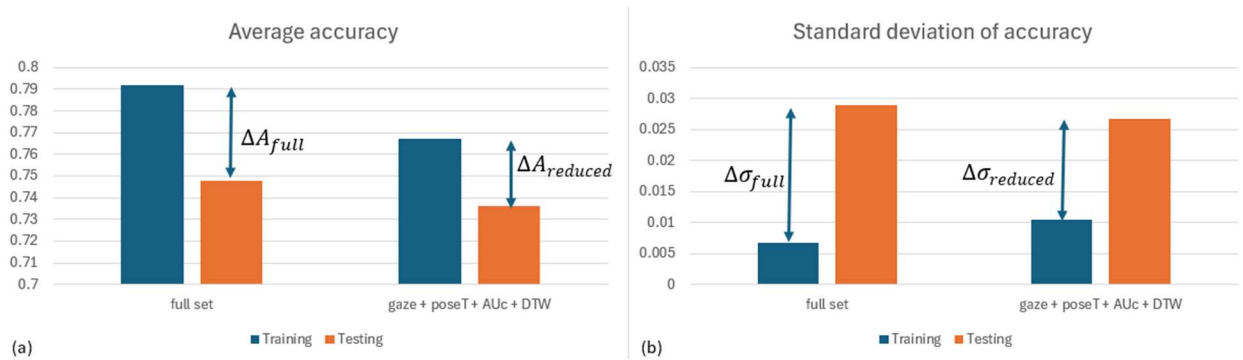

Figure 4. 14. (a) average accuracy and (b) standard deviation of accuracy across 20 experiments on training and testing data, for the full set of features and the reduced subset of features. (a) shows the difference in training and testing accuracy (ΔA) is smaller for the case of reduced features. (b) shows the standard deviation (σ) of the accuracy for the testing data is lower for the reduced feature space compared to the full feature space. Moreover, the difference of standard deviations ($\Delta \sigma$) between training and testing is smaller for reduced features compared to the full features.

We repeated these experiments to calculate precision and recall for reduced features and compared them with the case of using the full feature set. The results are depicted in Figure 4.15 and Figure 4.16.
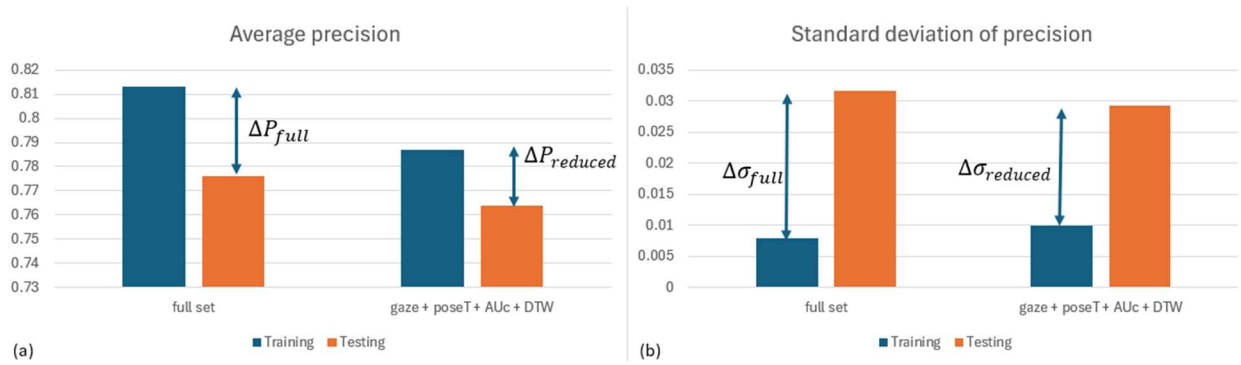
Figure 4. 15. (a) average precision and (b) standard deviation of precision across 20 experiments on training and testing data, for the full set of features as well as the reduced subset of features. (a) shows the difference in precision between training and testing (ΔP) is smaller in the case of reduced features. (b) shows the standard deviation (σ) of the precision of the testing data is lower for the reduced feature space compared to the full feature space. Moreover, the difference of standard deviations ($\Delta \sigma$) between training and testing is smaller for reduced features compared to the full features.
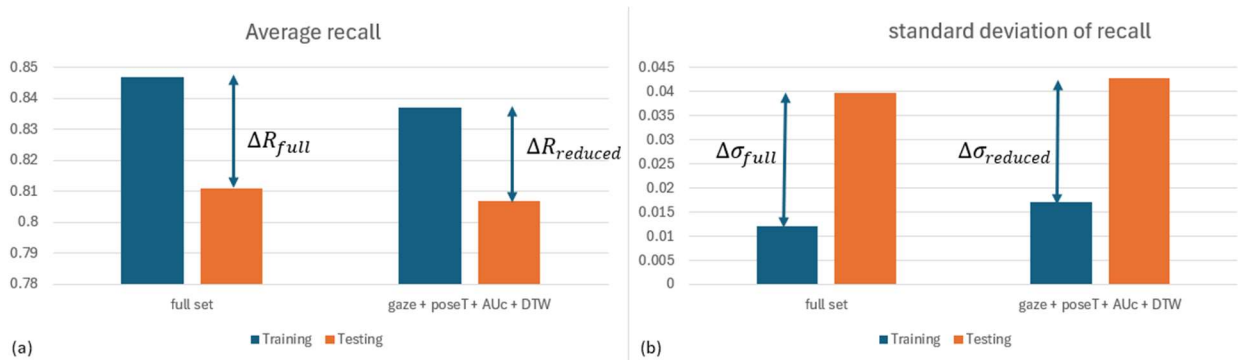

Figure 4. 16. (a) average recall and (b) standard deviation of recall across 20 experiments on training and testing data, for the full set of features and the reduced subset of features. (a) shows the difference in recall between training and testing (ΔR) is smaller in the case of reduced features. Although (b) shows the standard deviation (σ) of the recall for testing data is increased for the reduced features compared to the full features, the difference of standard deviations ($\Delta \sigma$) between training and testing is decreased for the reduced features compared to the full features. This result confirms a better generalization of the algorithm.

Moreover, the average and standard deviation of F1 score were calculated. The results are presented in Figure 4.17.
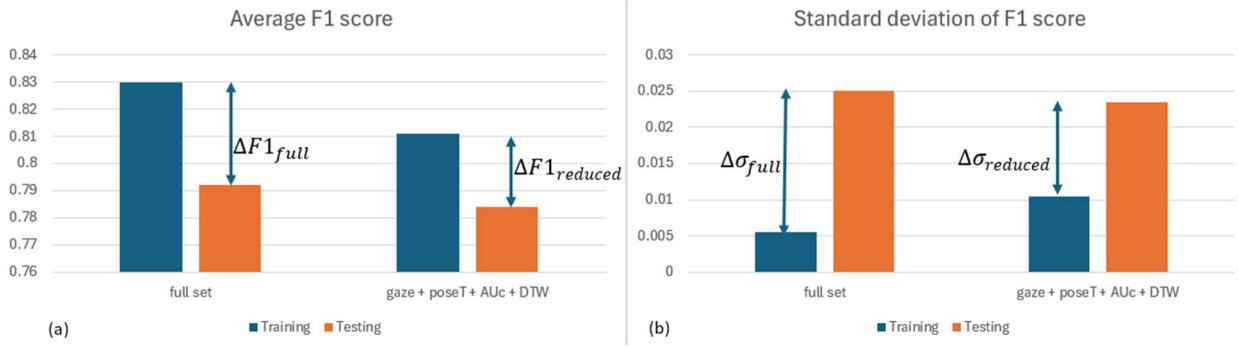
57

Figure 4. 17. (a) average F1 score and (b) standard deviation of F1 score across 20 experiments on training and testing data, for the full set of features and the reduced subset of features. (a) shows the difference in F1 score between training and testing ($\Delta F1$) is smaller in the case of reduced features. (b) shows the standard deviation (σ) of the F1 score for the testing data is lower for the reduced features compared to the full features. Moreover, the difference of standard deviations ($\Delta \sigma$) between training and testing is smaller for reduced features compared to the full features.

In all of these results, using the reduced features instead of full features decreased the difference between the average results between testing and training as seen in Figures 4.14 (a), 4.15(a), 4.16(a) and 4.17(a). Moreover, the standard deviation for the testing experiments on the reduced subset of features is smaller than that of the full set of features., except for the 'recall'. And in all cases, the difference between training and testing standard deviations of the reduced subset of features are much smaller than that of the full set of features as depicted in (4.4):

$$(\sigma_{test} - \sigma_{train})_{reduced\ features} < (\sigma_{test} - \sigma_{train})_{full\ features} \qquad (4.4)$$

where σ_{test} is the standard deviation of test results and σ_{train} is the standard deviation of the training results across all 20 runs of experiments. In other words, we have:

$$\Delta \sigma_{reduced\ features} < \Delta \sigma_{full\ features} \qquad (4.5)$$

where $\Delta \sigma_{reduced\ features}$ is the difference between σ_{test} and σ_{train} across 20 runs of experiments with reduced features and $\Delta \sigma_{full\ features}$ is the difference between σ_{test} and σ_{train} across 20 runs of experiments with full features. The fact that the standard deviation for test results is

58

lower in most cases (except for recall), and more importantly, $\Delta\sigma$ as in (4.5) is smaller for all cases (including recall) using reduced features shows that the classifier has achieved better generalization compared to the case of using the full feature set.

4.5. Summary and discussion

In this work, we focused on classifying subtle shifts in 'affect' in a completely natural setup without any constraints on the participants. We leveraged the power of neural networks but limited our design to the simplest neural network architecture with minimum number of nodes to help reduce the computational load. More investigation into the minimum number of frames per second for video recording could help to further decrease the time and resources needed for extracting the AUs from the video files. PCA was performed to reduce the dimension of features from 17 to 10 which helped reduce the variance in the results. Considering only positive and negative affect, testing accuracy of 76.8 % was achieved which is, to the best of our knowledge, the best results within the constraints discussed above. One observation in this study was reduction of accuracy while trying to classify the 'neutral' labels. We speculate that 'neutral' labels span more diverse characteristics compared to 'positive' or 'negative' labels; therefore, more training examples are likely necessary to train the algorithms to correctly classify 'neutral' instances. The bottleneck is increasing the number of 'neutral' datapoints alone is not helpful as it will result in skewed database which leads the classifier to massively develop bias towards identifying 'neutral' labels. In fact, according to our experience, most of the labels in a given dataset have been marked as 'neutral'. Therefore, for a given dataset, during the training phase, many of the 'neutral' labels were randomly removed to balance the dataset. This means even more data points have to be collected so that after balancing the dataset, remaining labels would

be enough for training the neural network to correctly detect 'neutral' labels. Tackling this challenge is a viable goal for future work.

As for rapport, we developed an architecture which leveraged DTW for gauging synchrony among participants. Five DTW features were constructed according to the gaze, head coordination, head orientation, AUc and AUr. Along these five DTW features, raw data for gaze, head coordination and AUc were used as input features. By leaving out head coordination and AUr data, a total of 63 features were used. 2674 data points were employed for training the neural network. An accuracy of 73.6% was achieved for testing over 20 experiments with a standard deviation of 2.68%. Precision, recall and F1 score for testing were achieved as 0.764, 0.807 and 0.784, respectively. To the best of our knowledge, these are the highest reported metrics for identifying rapport in multiperson groups and with highest temporal resolution of 30 seconds.

Further research could be done on the effect of more output classes such as high, neutral, and low rapport. The challenge lies in the number of additional datapoints needed for training the network. Moreover, balancing the dataset could get more challenging with higher number of classes which in return may require even more data point collection.

Another interesting path for research is incorporating the sequence of data in the analysis. As of now, the classifier does not consider the order of the datapoints. However, the labelers watched the video segments in order and that naturally affects the perception of 'rapport' and 'affect' by the labelers. Therefore, employing techniques such as recurrent neural networks and other methods for analyzing the sequence of data could further improve the

results. Our findings in this work pave the way and encourage the community to investigate

further the future work some of which were briefly mentioned here.

Disclaimer: A substantial portion of this chapter was published in [86] *© 2023, IEEE.*

Chapter 5: Advancing integrated electrochemical instruments for point-of-care devices

5.1 Introduction

As described in chapter 2, electrochemical sensing has been proven to be an effective approach for monitoring different physiological and environmental parameters. Therefore, implementing miniaturized electrochemical solutions could enhance assistive technologies for human health and wellness. To this end, researchers have utilized complementary metal-oxide semiconductor (CMOS) technology to develop small and wearable potentiostats [63], [64], [100], [101],and many advances have been made to develop potentiostats that increase the range of current readout [65], [66], decrease the power consumption and size [67], [102], [103], lower the noise [104] and widen the dynamic range [105], and support the bidirectional current of electro-chemical cells [66]. New processes have also been developed for implementing quasi-reference electrodes on the CMOS chip for a fully integrated electrochemical measurement [106].

Although these advances have enabled miniaturized electrochemical systems, as the modern CMOS technologies scale down in size, the voltage supply have become smaller [107]; for example, while an older 0.5 µm CMOS technology used to support a 5 V supply, newer technologies such as 180 nm support a maximum of 1.8 V for regular transistors or 3.3 V in the case of high-voltage transistors. As a result, many electrochemical reactions cannot be supported by modern integrated potentiostats, as illustrated in Figure 5.1.
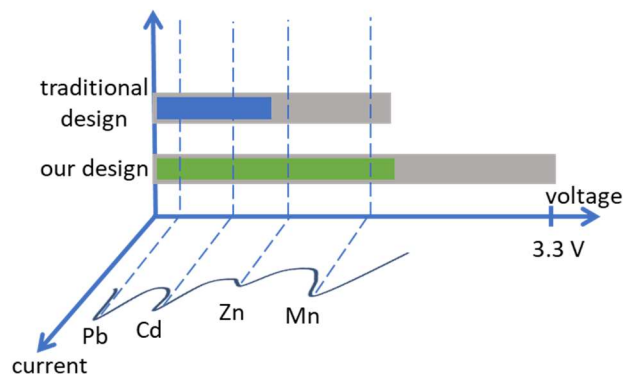
Figure 5. 1. The graph shows voltammetry of different heavy metals and indicates bias potentials for each target element to obtain peak current (data adapted from [68]). The blue and green bars show ideal ranges of bias potential that are supported with a traditional CMOS potentiostat and our novel potentiostat, respectively, both with a 3.3 V supply. In this example, the reactions for some elements such as Zn and Mn are not supported by a traditional CMOS potentiostat. Note that the gray bar represents $V_{CE\text{-}swing}$, the excess voltage beyond the bias potential required for an electrochemical cell [108].

Since a potentiostat needs to support bidirectional current for redox reactions, only half of the supply voltage is available to be used for each direction in an ideal rail-to-rail operation of the potentiostat. For a 3.3 V supply, this means only 1.65 V is available for each reduction or oxidation reaction. Furthermore, as detailed in section 5.2, because the counter electrode in a typical three-electrode electrochemical cell must be allowed to swing well beyond the bias potential, only a small portion of this 1.65 V is available to be used as bias potential, as illustrated in Figure 5.1. However, many electrochemical reactions, for example for detecting heavy metals such as manganese and zinc, require bias potentials of about 1.6 V and 1.2 V, respectively. As shown in Figure 5.1, these potentials fall outside the window that is supported by conventional CMOS potentiostats with power supplies of 3.3 V or lower. Therefore, conventional CMOS potentiostat designs implemented in newer technologies with lower supply voltages do not support voltammetry for detecting these elements. On the other hand, older CMOS process nodes, such as 0.5 μm that support supply voltages greater than 3.3 V, are not offered by

63

mainstream foundries as they are considered obsolete [109]. Therefore, it is inevitable to utilize these newer CMOS technologies for electrochemical measurements that come with the added benefits of smaller feature size, lower power consumption, and higher speed. Consequently, overcoming the issue of limited bias potential in CMOS potentiostats implemented in newer process nodes is crucial to accommodate a wide range of electrochemical reactions in wearable assistive technologies.

In this work, we introduce a novel potentiostat topology that addresses the limited supply voltage in newer CMOS technologies and supports bidirectional current measurement in a wide range of electrochemical reactions. For a given supply voltage, this new topology nearly doubles the voltage range for the electrochemical cell compared to conventional designs. Hence, it enables detecting a wider range of target elements than any previously reported integrated potentiostat. As desired with most integrated instrumentation circuits, this potentiostat also provides a small form factor and low power consumption for a compact system implementation which is necessary for wearable applications. We presents an in-depth analysis on voltage requirements of a three-electrode electrochemical cell as well as the challenges of Conventional CMOS potentiostats in section 5.2. Then we present the methodology and design for enhancing voltage range of the electrochemical cell along with the results of electrochemical experiments as well as simulation results of the implemented CMOS potentiostat.

5.2 Manifestation of Electrode Potentials and Challenges for Conventional CMOS Potentiostats

5.2.1 Electrochemical Cell Model and Manifestation of Potentials at Electrodes

As briefly asserted in in section 5.1, an important bottleneck in miniaturized CMOS potentiostats is their ability to support a wide bias potential window to extend the range of electrochemical targets that can be measured using CMOS instrumentation. To elaborate on this point, consider the electrochemical cell model shown in the circle at the center of Figure 5.2. A three-electrode electrochemical cell features a reference electrode (RE), a working electrode (WE) and a counter electrode (CE). The resistance between CE and RE is mainly attributed to the solution resistance. Similarly, the resistance between RE and WE is attributed to the solution resistance in series with a parallel capacitance and resistance that model the double-layer capacitance and charge transfer resistance at the WE surface.
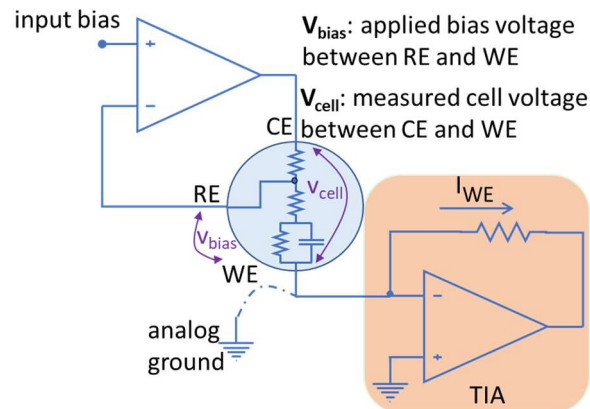

Figure 5. 2. Schematic of a traditional potentiostat with grounded working electrode. The electrochemical cell model is presented at the center of the figure with a circle symbol [108]*.*

In this three-electrode cell, a bias voltage is traditionally applied to the RE with respect to WE. In other words, V_{RE-WE} is applied to the electrochemical cell as shown in Figure 5.2. In this paper, we will refer to this applied voltage as V_{bias}. Note that V_{bias} is sometimes defined as V_{WE-RE}

[8], which is negative of V_{bias} as defined here. Both definitions are valid as long as one remains consistent. Therefore, throughout this paper, we define:

$$V_{bias} = V_{RE-WE} = V_{RE} - V_{WE} \qquad (5.1)$$

This definition facilitates a clearer discussion about the integrated CMOS potentiostat. While the V_{bias} is externally applied between RE and WE, the potential on CE can and will swing beyond V_{bias} in order to establish a desired electrochemical reaction. Let us define this CE swing voltage as:

$$V_{CE-swing} = V_{CE-RE} = V_{CE} - V_{RE} \qquad (5.2)$$

This $V_{CE-swing}$ depends on several factors such as electrolyte concentration and the geometry and material of electrodes, and it can be as large as V_{bias}, which extends the maximum potential the potentiostat must support to beyond two times V_{bias}. Finally, let us define the full cell potential, V_{cell} such that:

$$V_{cell} = V_{CE-WE} = V_{CE} - V_{WE} = V_{CE-swing} + V_{bias} \qquad (5.3)$$

Based on our extensive experience with integrated electrochemical platforms, we expect voltages at the cell electrodes to generally manifest similar to the graph in Figure 5.3. The absolute value of the cell potential is always more than that of the bias potential due to the existence of CE-RE resistance. Moreover, by lowering the electrolyte concentration, the CE-RE potential difference further increases due to the increase in the CE-RE resistance. Therefore, for a potentiostat with a limited voltage supply, the voltage swing on CE is the limiting factor.
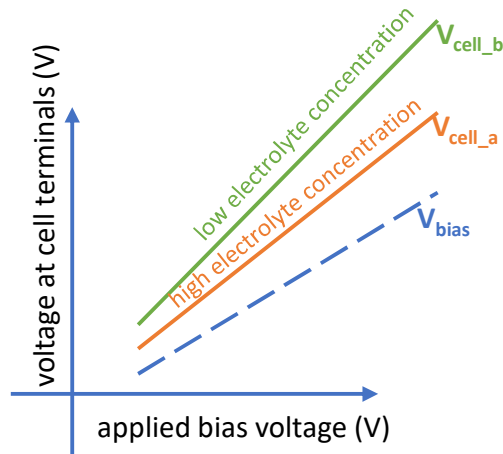
Figure 5. 3. Conceptual representation of V_{cell} and V_{bias}. V_{RE-WE} (V_{bias}) is always equivalent to the V_{bias} voltage applied to the electrochemical cell. V_{CE-WE} (V_{cell}), however, is more than V_{bias} and further increases if electrolyte concentration decreases [108].

5.2.2 Challenges of Conventional CMOS Potentiostats

A conventional CMOS potentiostat is shown in Figure 5.2. An operational amplifier is used to apply a bias voltage to an electrochemical cell. The current generated in the electrochemical cell is usually read using a transimpedance amplifier (TIA) as shown in the bottom right of Figure 5.2. The WE of the electrochemical cell in this design is tied to analog ground which is usually set to $V_{supply/2}$. This allows the potentiostat to support bidirectional current measurement and hence supports both reduction and oxidation reactions. For instance, in the old 0.5 µm CMOS technology with a 5 V supply, in an ideal rail-to-rail operation of the circuit, the analog ground is set to 2.5 V. Therefore, the available voltage for $|V_{cell}|$ is 2.5 V in either direction (negative or positive). Basically, the bottom half of the supply range (0 V to 2.5 V) is used to support negative V_{cell} (remember $V_{cell}=V_{CE}-V_{WE}$) and the top half (2.5 V to 5 V) is used to support positive V_{cell}. Only a portion of this 2.5 V in either direction can be assigned to V_{bias} because always $V_{bias}<V_{cell}$ (the exact ratio of V_{bias} to V_{cell} depends on the cell condition such as electrolyte concentration). This covers a relatively wide range of electrochemical experiments [110]. However, 0.5 µm CMOS

process node is not offered by major foundries anymore [109]. On the other hand, the supply voltage in newer CMOS technologies is drastically reduced compared to the older technologies. For example, going from 0.5 µm CMOS to newer 180 nm CMOS, the supply voltage drops from 5 V to 1.8 V (or 3.3 V in case of high-voltage transistors). This reduction in supply severely restricts the range of electrochemical experiments that can be conducted using a conventional CMOS potentiostat. In other words, this reduced supply voltage is not sufficient to support V_{bias} and V_{cell} in an electrochemical cell. In this case, for an ideal rail-to-rail operation of a CMOS potentiostat with 3.3 V supply, only $V_{supply/2}$=1.65 V is available for the electrochemical cell in either direction (negative or positive). Therefore, the absolute value for the maximum $|Vcell|=|V_{CE}-V_{WE}|$ in this case is 1.65 V (i.e. 'Vsupply - analog ground' or 'analog ground - gnd'). $|V_{bias}|=|V_{RE}-V_{WE}|$ in this case will be much lower than $|V_{cell}|$ as described in the previous section. The results of our experiments suggest $V_{bias}≈0.5V_{cell}$ as presented in section 4, but the exact ratio depends on the characteristics of the electrochemical cell. Consequently, only around 0.9 V is available as Vbias in this example with a conventional potentiostat. As shown in Figure 5.1, many of the electrochemical reactions happen in bias voltages outside this potential window [68], [110], [111] and hence are not supported by conventional methods. In this work, we present a novel circuit architecture for CMOS potentiostats that widens the supported windows for V_{bias} and V_{cell} to facilitate a wide range of electrochemical reactions and enable powerful miniaturized assistive technologies for point-of-care- devices.

5.3 Design Methodology for CMOS Potentiostats to Support High Voltage Requirements

To solve the problem of limited potential window, a novel architecture is introduced in this work to enable wider range of electrochemical experiments using cutting-edge CMOS technologies. The first step to widen the voltage swing is to allow the voltage on the WE to switch between high and low supply rails, instead of being tied to analog ground. This will allow V_{bias} to have a voltage swing of the full supply range in an ideal rail-to-rail operation of the circuit. The limitation, however, arises in reading the current. Traditionally, a TIA is used to read the current, whose reference point is tied to analog ground together with the WE as shown in Figure 5.2. By employing the proposed method, the reference point of the TIA should switch between high and low supply rails along with WE. However, this does not allow reading current in the original direction as it will push the output of the TIA beyond 3.3 V (the supply voltage) or less than 0 V (ground) which is not possible. Therefore, a current conveyor was employed to reverse the direction of the current and thus enable the TIA to read the current properly. The schematic of the current conveyor is seen in the middle of Figure 5.4.
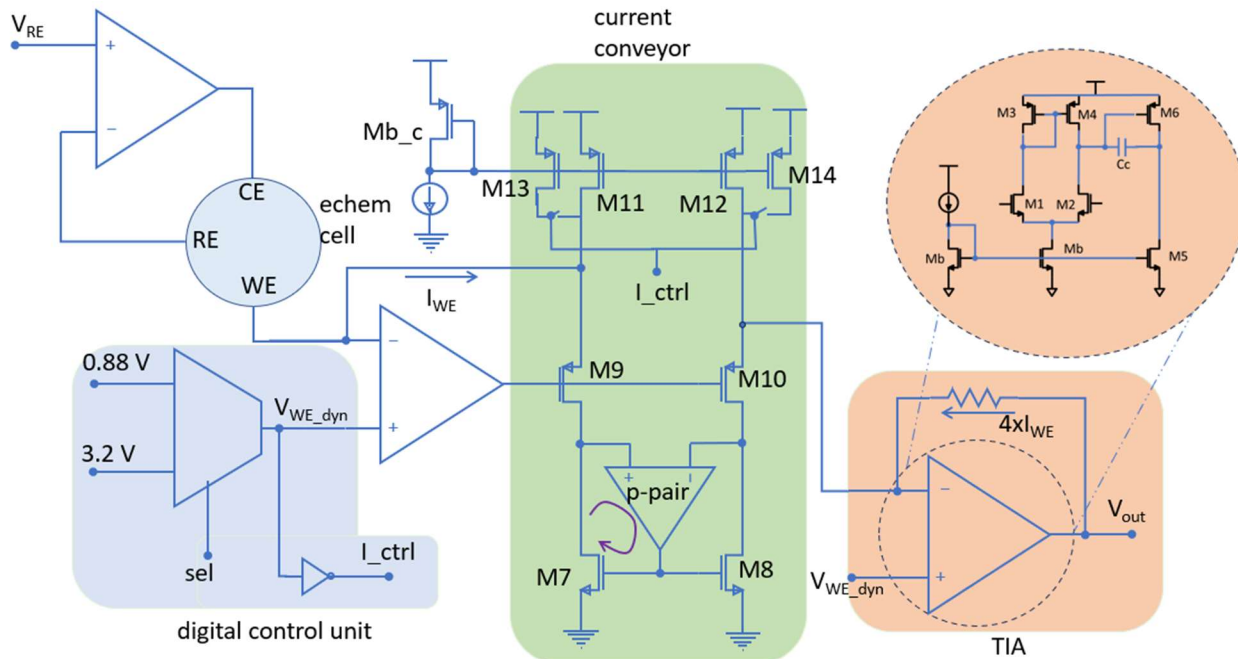
Figure 5. 4. Schematic of the implemented potentiostat. The current conveyor in the middle of the schematic is employed to reverse the direction of current to support bidirectional current measurement while allowing WE to switch between supply rails [108].

5.3.1 Current Conveyor

The current conveyor in this work was designed with the objective of enabling wide output voltage swing. In a typical current conveyor, a cascode current mirror is used to ensure the accuracy of copying current from the left leg to the right leg. However, to maximize the voltage swing at the output, single transistors (M7 and M8) were used in the current mirror to reduce the overhead voltage required for the circuitry and hence maximize the voltage swing for electrochemical reactions. These single transistors are shown in the bottom center of Figure 5.4. However, using single transistors results in mismatch between mirrored currents if the transistor's drain voltages do not follow each other. To ensure the matching of the current in both legs of the current mirror, an op amp was employed to match the drain voltages of the transistors. The input pair of this op amp was constructed of PMOS transistors to warrant that

70

the op amp remains in saturation mode even with low voltages at its input terminals. This op amp is placed in both positive and negative feedback loops. The impedance at the left leg of the current conveyor is higher than that of the right leg. To elaborate on this, note the simplified schematic of the circuit shown in Figure 5.5. In this structure, the mirrored current is multiplied from left leg to the right leg of the current conveyor. As it will be detailed in Table 5.1, the transistors in the right leg are four times wider than the ones in the left leg. Therefore, the output resistance of M7, M9, M11 and M13 are bigger than those of M8, M10, M12 and M14. Moreover, the cell resistance that we simplified in this schematic with R_{cell} is in the Mega Ohm range, as will be explained in more detail in section 5.4.2 while the feedback resistor on the TIA (R_f) is in kilo Ohm range. Therefore, the equivalent impedance at the drain of M7 is higher than the equivalent impedance at the drain of M8.
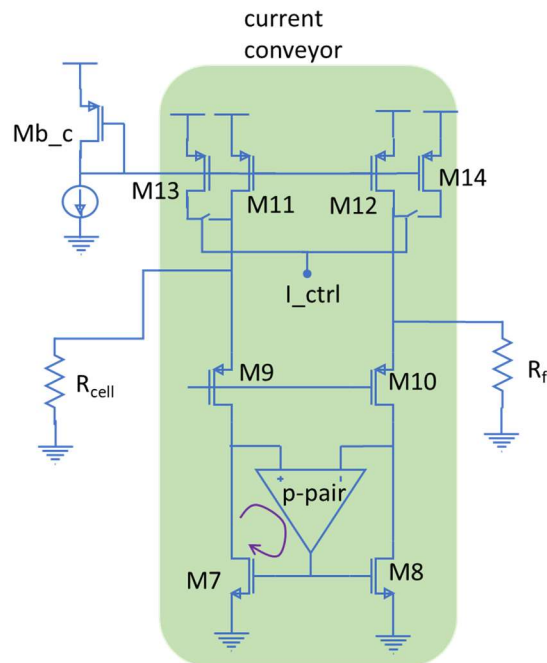

Figure 5. 5. Simplified schematic showing the impedance at input nodes of the p-pair op amp.

The positive terminal of the op amp is connected to the left side at the drain of M7 to form a negative feedback loop (as shown with the purple curved arrow in Figure 5.4) that is stronger than the positive feedback loop. Note that since M7 adds 180 degrees to the phase, the drain of M7 is tied to the positive terminal of the op amp to make a strong negative feedback loop. This guarantees the stability of the circuit, and it ensures the drain voltages of the two transistors match and the current is accurately mirrored.

To maximize the voltage range for electrochemical cells, the transistors at the output of the op amp in the current conveyor (M9 and M10) should be carefully designed. The first option considered was the NMOS-based design shown in Figure 5.6. However, the high threshold and overdrive voltages of the NMOS transistor at the output of the op amp were found to limit the available voltage for electrochemical reactions. This higher threshold voltage was due to the body effect of the NMOS transistor. To reduce the overdrive voltage that limits output swing, it was noted that a design based on PMOS transistors in isolated n-wells would eliminate the body effect and hence decrease the threshold and overdrive voltages of the transistors, given our design utilizes an n-well CMOS technology. Therefore, to decrease overdrive voltage and increase the range of voltage available for electrochemical reactions, the PMOS transistors with isolated wells were employed in the final design, as depicted in Figure 5.4. In addition, a PMOS transistor was added to the middle of the second leg of the current conveyor to balance the current in both legs.
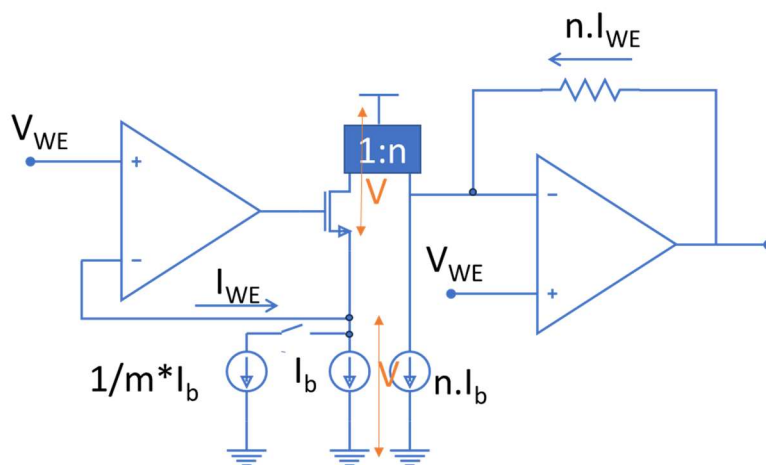
Figure 5. 6. Initial design of the current conveyor with NMOS transistors. This design suffered from limited voltage range available for the electrochemical cell that was mainly caused by body effect of the NMOS transistor .

5.3.2 Digital Control Unit

A digital control unit was employed to dynamically change the reference voltage that is applied to WE (bottom left of Figure 5.4) and the positive terminal of the TIA (right side of Figure 5.4). Also, a digital signal (I_ctrl) was created from the WE voltage, and this signal was used to control the current sources employed in the current conveyor, through the switches depicted in the top center of Figure 5.4. This control of the bias current is crucial to properly bias the current conveyor according to the voltage that is applied to the WE.

5.4 Results

5.4.1 Test Setup and Electrochemical Experiments

To assess the behavior of electrode voltages for varying electrochemical model parameters, chronoamperometry experiments were conducted using different electrolyte concentrations. Electrodes in these experiments were built in-house using standard microfabrication techniques, including photolithography and thermal evaporation. Interdigitated electrodes were made by depositing 10 nm of titanium and 100 nm of gold on a silicon wafer

73

containing a thin silicon dioxide layer. The titanium was used as an adhesion layer between the gold and the oxide substrate. For the electrolyte, phosphate buffer (PB) solution was used in low concentrations to increase the lifetime of thin-film gold electrodes. Phosphate buffer saline (PBS) solution was avoided because the chlorine (Cl) molecules released from saline were observed to dissolve the gold electrodes in previous experiments.

Experiments with our custom interdigitated gold electrodes were performed in a beaker using 0.05 M and 0.1 M PB solutions. A commercial electrochemistry instrument (CHI 760E) was used for chronoamperometry measurements. An illustration of the test setup and electrodes is shown in Figure 5.7, where a photo of the fabricated interdigitated electrode is provided as an inset. The experimental measurements shown in Figure 5.8 confirm the initial expectation that, while Vbias (V_{RE-WE}) stays at the applied bias potential, V_{cell} (V_{CE-WE}) is always greater than the applied bias potential. Notice also from Figure 5.8 that the CE potential further increases when the electrolyte concentration is decreased. For instance, for a V_{bias} of 1.4 V, a V_{cell} of 2.2 V and 2.9 V were measured for high and low electrolyte concentrations, respectively. This validates the importance of expanding the potential window that a potentiostat could support.
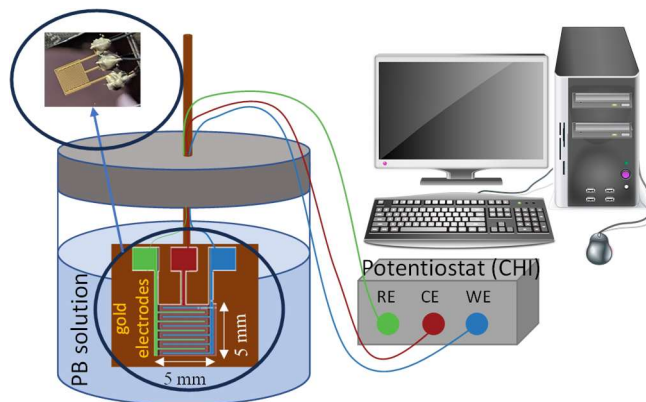

Figure 5. 7. Schematic of the experimental setup for concentration comparison using interdigitated electrodes and CHI potentiostat. A photo of the fabricated electrodes is shown inset. The size of electrode is 5 mm x 5 mm [108].
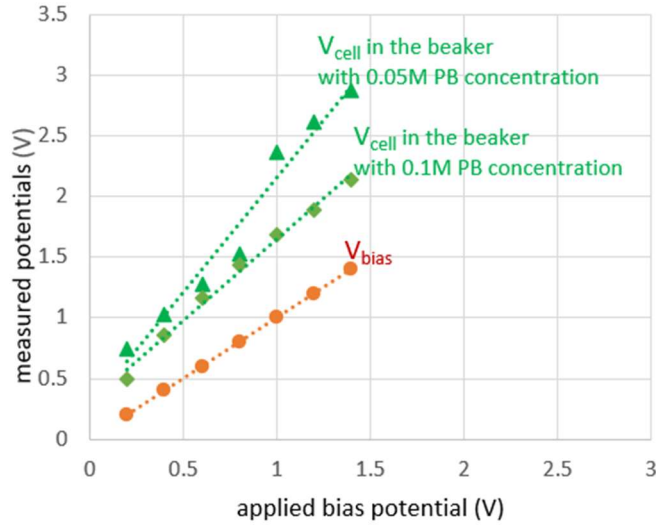
74

Figure 5. 8. Measured CE and RE voltages w.r.t. the WE voltage (i.e. V_{cell} and V_{bias}). V_{cell} is always higher than V_{bias} and this voltage difference increases as electrolyte concentration decreases [108].

5.4.2 Electrochemical Cell Model

As described in section 5.2.1. the parallel capacitance and resistance in Figure 5.9 model the double layer capacitance and charge transfer resistance at the WE surface. For all simulations of the novel wide-swing potentiostat, a typical value of 2.6 µF was chosen as the model capacitance and a value of 64 kΩ was used to model the charge transfer resistance based on the data presented in [112]. The solution resistance values of the electrochemical cell were empirically modeled from the experiments described in section 5.4.1 as follows: the measured steady state chronoamperometry current for a given bias voltage was used to calculate the RE-WE resistance; and the measured cell voltage for each chronoamperometry current was used to determine CE-RE resistance. For these calculations, measurements were performed at Vbias = 1 V where V_{cell} was measured as 2.44 V. Then, the RE-WE resistance was calculated as 10.2 MΩ and the CE-RE resistance was calculated as 14.7 MΩ. This calculation is based on the fact that the RE in a three-electrode electrochemical cell does not draw any current [113]. Therefore, the

75

resistances in the electrochemical cell can be considered to be in series and the voltage drop on

the resistances can be calculated by Kirchhoff's circuit laws. The voltage and current distribution

on the model electrochemical cell are depicted in Figure 5.9. This gives a reasonable

approximation of the resistance values that were used in simulations.
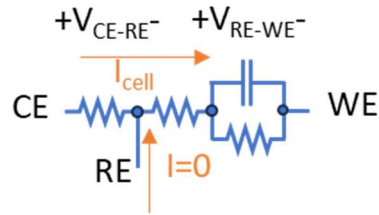
Figure 5. 9. The electrochemical cell characterization for simulation of the potentiostat. The characterization was performed using the measured voltage and current using the CHI 760E instrument [108]*.*

5.4.3 Simulation Results for the CMOS Potentiostat

Using the transistor sizes listed in Table 5.1, the new potentiostat design from Figure 5.4

was simulated in Cadence along with the electrochemical cell model described above.

Table 5.1. Transistor sizing of the readout portion of the potentiostat shown in Figure 5.4 [108].

device	W/L	fingers	device	W/L	fingers
M1,2	4u/500n	1	M8	4x2u/350n	1
M3,4	8.5u/500n	2	M9	8u/300n	1
M5	7.5u/500n	4	M10	4x8u/300n	1
M6	10u/500n	20	M11	2u/300n	2
Mb	6u/500n	1	M12	2u/300n	8
Mb_c	2u/300n	2	M13	2u/300n	4
M7	2u/350n	1	M14	2u/300n	16

The simulation results are shown in Figure 5.10. Figure 5.10 (a) demonstrates the voltage support for positive V_{bias} and V_{cell}. In this case, the digital control unit fixes the WE voltage at 0.88 V. By sweeping RE voltage from 0.88 V toward 3.3 V, V_{bias} increases until it saturates at V_{bias_max} = 1.1 V. Consequently, V_{cell} also increases until it saturates at V_{cell_max} = 2.41 V. Likewise, Figure 5.10 (b) explains voltage support for negative V_{bias} and V_{cell}. In this case, the WE voltage is fixed at 3.2 V and the RE voltage is swept from 3.2 V toward 0 V. As seen in Figure 5.10 (b), a V_{bias} of -2.12 V to 0 V and a V_{cell} of -3.11 V to 0 V are supported. This demonstrates that the new potentiostat enhances the potential window for oxidation and reduction measurements by supporting a maximum Vcell of 2.41 V and -3.11 V in positive and negative directions, respectively. In comparison, a conventional potentiostat, even with an ideal rail-to-rail operation at a 3.3 V supply supports a maximum Vcell of only ±1.65 V. Therefore, our new potentiostat architecture achieves 46% and 88% increase in the voltage range of V_{cell} for positive and negative voltages, respectively.
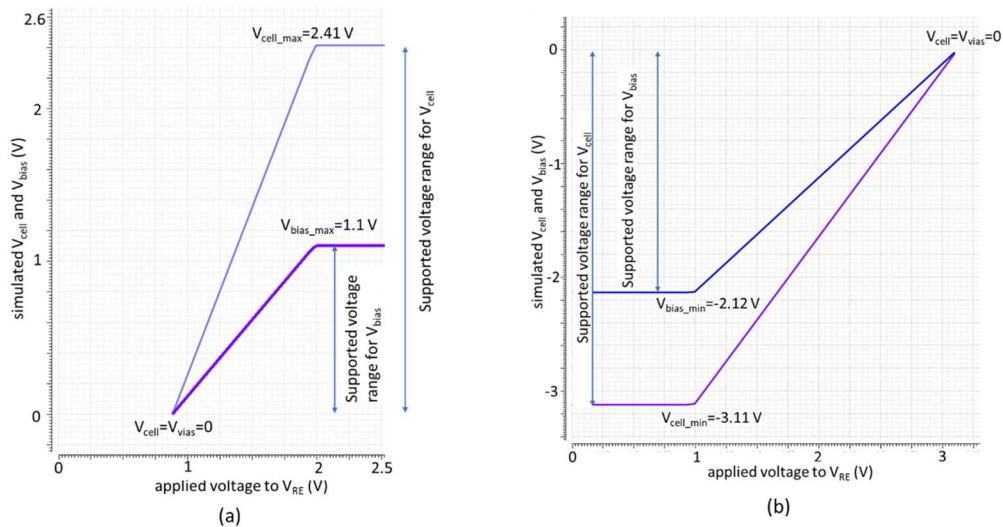

Figure 5. 10. Simulated voltage range for V_{bias} and v_{cell}. The graphs show the enhanced supported voltage range for (a) positive and (b) negative V_{bias} and V_{cell} [108].

The design operates from a 3.3 V supply and consumes only 2.047 mW of power. Figure 10 shows the layout of the new potentiostat designed in 180 nm CMOS technology, which occupies only 0.013 mm^2. The op amp employed in this design occupies 964 μm^2, consumes 435 μW, and demonstrates a bandwidth of 32.52 MHz with phase margin of 63 degrees. Figure 5.11 demonstrates the Bode plot for this op amp and Table 5.2 highlights the design and performance characteristics of the op amp designed for and employed in the potentiostat.

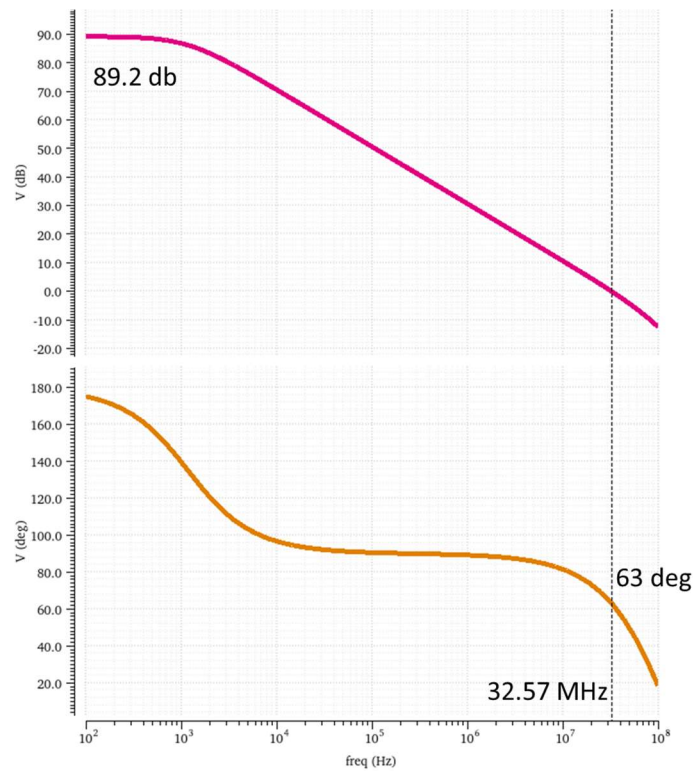*Figure 5. 11. Bode plot of the designed op amp used in the novel potentiostat design. It shows around 90 db of DC gain, bandwidth of 32.57 MHz and phase margin of 63 degree.*

Table 5.2. Characteristics of the op amp [108].

supply	technology	DC gain	area	power	bandwidth	slew rate
3.3 V	CMOS 180 nm	89.2 db	964 μm^2	435 μW	32.52 MHz	28.33 V/μs

Table 5.3 summarizes the characteristics of the whole novel potentiostat including the current conveyor, the digital control unit and the TIA. The 10% to 90% charge and discharge time of a typical 2.6 µF capacitor within the model electrochemical cell was less than 1 µs. Considering the reaction times of multiple seconds in a typical chronoamperometry experiment, the charge and discharge times are negligible and the potentiostat meets the speed requirements. Figure 5.12 demonstrates the layout of the entire potentiostat, implemented in CMOS 180 nm technology.

Table 5.3. Electrical Characteristics of the Potentiostat [108].

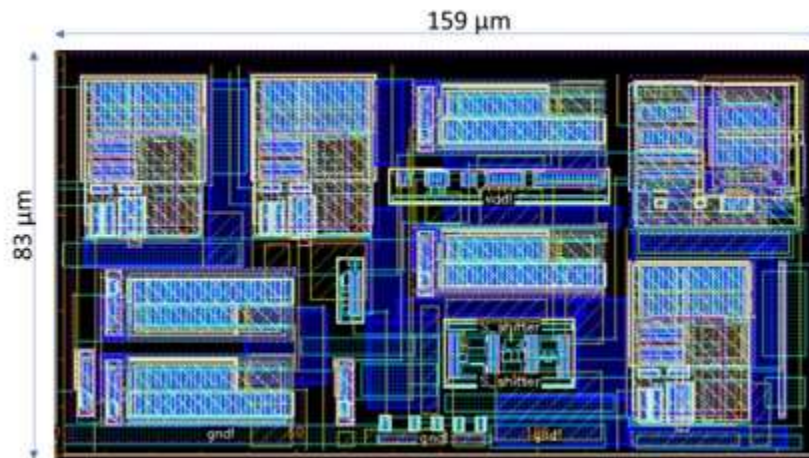| supply	area	max power	max cell voltage support	load capacitance
3.3 V	0.0132 mm^2	2.047 mW	3 V	2.6 µF

Figure 5. 12. Layout of the entire potentiostat designed in 0.18 µm CMOS technology. The total dimension is 159 µm x 83 µm [108].

5.5 Summary and discussion

5.5.1 Summary

As conventional integrated potentiostats support a limited cell voltage range, V_{cell}, they fail to accommodate many electrochemical reactions of interest. To resolve this challenge, we introduced a novel integrated potentiostat topology that was verified to support V_{cell} range between 2.41 V and -3.11 V (with 3.3 V supply). This increases the maximum supported V_{cell} by 46% and 88% for positive and negative voltages, respectively, compared to traditional CMOS potentiostat designs. This dramatic improvement in potential window permits the measurement of a much wider range of electrochemical targets, expanding applications for portable sensing and assistive technologies for point-of-care applications. The circuit was implemented in CMOS 180 nm technology and consumes only 2.047 mW of power. For a given electrochemical cell model, the maximum charge and discharge time was found to be under 1 µs, easily meeting the speed requirements for most electrochemical experiments. The greatly expanded potential window of this new potentiostat, along with its low power consumption and high slew rate, make this design well-suited for many current and future wearable electrochemical sensing and point-of-care applications.

5.5.2 Discussion and open challenges

Given the trend toward smaller integrated circuits and lower power supplies, the limited potential window for integrated CMOS potentiostats will be reduced as well. In this work, we addressed this challenge by leveraging existing technologies that support up to 3.3 V. In order to further increase the range of the potential window for such potentiostats or for implementing potentiostats with sub 1 V supplies, new methods should be explored. These methods could

80

utilize topologies such as charge pumps to increase the available voltages. The bottleneck is the current support, which also needs to be taken into account. Therefore, simply using voltage boosters alone cannot solve the issue, as they often do not support high-current applications. This opens the avenue for further research on this topic.

Disclaimer: A substantial portion of this chapter was published in [108].

Chapter 6: Conclusions and future works

6.1. Summary

In the pursuit of holistic wellness, monitoring various aspects of individuals' lives has become paramount. Assistive technologies have emerged as powerful tools for tracking physical health parameters. However, psychological wellness is another important factor of overall health. Therefore, it is crucial to develop assistive technologies to monitor social, physiological, and environmental parameters affecting health to promote individuals' wellness. In order to have multifaceted assistive technologies as point-of-care devices, these technologies need to support multimodal sensing and monitoring. The challenges arise when these technologies need to be implemented resource-efficiently and often in real-time. Moreover, transferring from standard laboratory equipment to personal miniaturized point-of-care devices limits their functionality. In this work, we introduced a framework and method for assessing social interaction parameters as important factors of emotional wellness. More specifically, we developed algorithms for monitoring 'affect' and 'rapport' as indications of the quality of interaction among individuals. Moreover, to facilitate multimodal sensing point-of-care devices, we developed novel electrochemical architecture to monitor important physiological and environmental parameters with highly efficient use of resources.

6.2. Contributions

The work presented in this dissertation converges engineering solutions to psychological and physiological aspects of individuals' wellness by contributing novel technologies and frameworks to facilitate multimodal sensing of important parameters. These contributions are:

- We developed a machine learning platform for monitoring 'affect' and 'rapport' as two important parameters indicative of the quality of interaction among individuals as well as the overall wellness of participants. Our analysis are done in **natural setups** as opposed to artificial and acting setups. We developed classifiers to identify **subtle changes** in affect and rapport as opposed to classifying only extreme cases. We introduced **the first** framework for analyzing these parameters in **groups** with a **high temporal resolution** of 30 seconds. The **accuracy, precision and recall** for assessing rapport are among the highest reported in the literature.

- Introduced novel architecture for enhancing electrochemical measurement of target elements. This novel architecture **doubles** the potential range of traditional solutions and hence facilitates **the most comprehensive** measurement of important health-related chemical parameters within the requirements of wearable point-of-care devices utilizing low voltage supply technologies. This architecture was implemented as an integrated circuit using CMOS 180 nm technology.

6.3. Future work

The work presented in this dissertation established a foundation for implementation of multifaceted assistive technologies that pursue a holistic wellness of individuals. The following are suggestions for future work:

- Developing personalized solutions for social interaction monitoring

Given personal differences, individuals may exhibit different baselines in emotions and rapport-building with others. In other words, a state of high affect and rapport for one individual may seem similar to the baseline of these parameters for another individual. Therefore, the

algorithms should take these differences into account. Developing solutions that can learn the baseline for each individual using techniques such as reinforcement learning can greatly improve the success of these methods. The key to this success is generating a rich dataset of interactions among individuals. In this work we recorded about 750 minutes total of interactions. The more the data collection process, the more success in getting results. Further statistical analysis is needed to design experiments for developing personalized affect and rapport monitoring and recommendation platforms.

- Context-aware approach for analyzing affect and rapport

In this work, our developed algorithms do not take into account the sequence of data points. As human labelers watched the video segments in order, they had an understanding of the context of the sessions. This undoubtedly influenced the labelers' perception about the events and their judgment in labeling the datapoints. However, the developed algorithms are agnostic to the contexts and assess the video segments by only looking at one datapoint at a time. Implementing classifiers that pay attention to the sequence of data could potentially improve the results. Methods such as recursive neural network could be explored for this purpose.

- Integration of more diverse modalities in assistive technologies

We showcased electrochemical methods as a viable solution for detecting physiological parameters. Incorporating other methods, such as optical solutions for monitoring health-related parameters, could improve the multimodal aspect of these devices and hence expand the efficacy of multifaceted assistive technologies. These methods could provide simple and yet effective methods for monitoring parameters such as heart rate. The challenge is effective

resource sharing among different modalities. Expanding the modalities of operation comes with more complexity and is resource-consuming. Integrating these modalities in efficient ways, such as smart multiplexing and innovative packaging, could be a future thrust for implementing multimodal assistive technologies.

BIBLIOGRAPHY

[1]   B. T. Nguyen, M. H. Trinh, T. V. Phan, and H. D. Nguyen, "An efficient real-Time emotion detection using camera and facial landmarks," in *7th International Conference on Information Science and Technology, ICIST 2017 - Proceedings,* Institute of Electrical and Electronics Engineers Inc., May 2017, pp. 251–255. doi: 10.1109/ICIST.2017.7926765.

[2]   M. Ez-zaouia, A. Tabard, and E. Lavoué, "EMODASH: A dashboard supporting retrospective awareness of emotions in online learning," *International Journal of Human Computer Studies*, vol. 139, no. October 2018, p. 102411, 2020, doi: 10.1016/j.ijhcs.2020.102411.

[3]   Y. R. Veeranki, N. Ganapathy, R. Swaminathan, and H. F. Posada-Quintero, "Comparison of Electrodermal Activity Signal Decomposition Techniques for Emotion Recognition," *IEEE Access*, vol. 12, pp. 19952–19966, 2024, doi: 10.1109/ACCESS.2024.3361832.

[4]   Y. Nam, Y. Kong, B. Reyes, N. Reljin, and K. H. Chon, "Monitoring of heart and breathing rates using dual cameras on a smartphone," *PLoS One*, vol. 11, no. 3, pp. 1–15, 2016, doi: 10.1371/journal.pone.0151013.

[5]   H. Yu, M. Xu, X. Xiao, F. Xu, and D. Ming, "Detection of dynamic changes of electrodermal activity to predict the classroom performance of college students," *Cogn Neurodyn*, vol. 18, no. 1, pp. 173–184, Feb. 2024, doi: 10.1007/s11571-023-09930-6.

[6]   F. Gao et al., "Wearable and flexible electrochemical sensors for sweat analysis: a review," *Microsystems and Nanoengineering,* 2023, Springer Nature. doi: 10.1038/s41378-022-00443-6.

[7]   N. M. Cusack, P. D. Venkatraman, U. Raza, and A. Faisal, "Review—Smart Wearable Sensors for Health and Lifestyle Monitoring: Commercial and Emerging Solutions," *ECS Sensors Plus,* vol. 3, no. 1, p. 017001, Mar. 2024, doi: 10.1149/2754-2726/ad3561.

[8]   E. Ashoori, S. Parsnejad, H. Yin, J. Figueroa, N. Sepúlveda, and A. J. Mason, "A Low-Cost Liquid-Based Capacitive Sensor for PM2.5 Monitoring," in *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS),* 2021, pp. 907–910. doi: 10.1109/MWSCAS47672.2021.9531893.

[9]   J. Whitehill, Z. Serpell, Y. C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: Automatic recognition of student engagement from facial expressions," *IEEE Trans Affect Comput*, vol. 5, no. 1, pp. 86–98, 2014, doi: 10.1109/TAFFC.2014.2316163.

[10]   A. Kaur, A. Mustafa, L. Mehta, and A. Dhall, "Prediction and Localization of Student Engagement in the Wild," *2018 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2018*, 2019, doi: 10.1109/DICTA.2018.8615851.

[11] W. Widodo, M. Suendarti, and H. Hasbullah, "Exploring The Effect of Knowledge Management and Social Intelligence on Professional Performance of Mathematics Teachers: A Mediating by Achievement Motivation," *Journal of Xidian University*, vol. 14, no. 6, pp. 749–757, 2020, doi: 10.37896/jxu14.6/085.

[12] M. Ez-Zaouia and E. Lavoué, "EMODA: A tutor oriented multimodal and contextual emotional dashboard," *ACM International Conference Proceeding Series*, pp. 429–438, 2017, doi: 10.1145/3027385.3027434.

[13] T. Fan et al., "A new deep convolutional neural network incorporating attentional mechanisms for ECG emotion recognition," *Comput Biol Med*, vol. 159, Jun. 2023, doi: 10.1016/j.compbiomed.2023.106938.

[14] S. Samrose et al., "CoCo: Collaboration Coach for Understanding Team Dynamics during Video Conferencing," *Proc ACM Interact Mob Wearable Ubiquitous Technol*, vol. 1, no. 4, pp. 1–24, 2018, doi: 10.1145/3161186.

[15] P. Schmidt, A. Reiss, R. Dürichen, and K. Van Laerhoven, "Wearable-Based Affect Recognition—A Review," *Sensors,* pp. 1–42, 2019, doi: 10.3390/s19194079.

[16] Samiha Samrose, "Visual Cues for Disrespectful Conversation Analysis," in *8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2019.

[17] V. G. Nelson, " Unconscious Bias in the Workplace." Accessed: Apr. 25, 2024. [Online]. Available: https://www.experthumanresources.com/unconscious-bias-in-the-workplace/

[18] S. Davila-Montero, J. A. Dana-Le, G. Bente, A. T. Hall, and A. J. Mason, "Review and Challenges of Technologies for Real-Time Human Behavior Monitoring," *IEEE Trans Biomed Circuits Syst*, vol. 15, no. 1, pp. 2–28, Feb. 2021, doi: 10.1109/TBCAS.2021.3060617.

[19] D. McDuff, K. Rowan, P. Choudhury, J. Wolk, T. Pham, and M. Czerwinski, "A Multimodal Emotion Sensing Platform for Building Emotion-Aware Applications," Mar. 2019, [Online]. Available: http://arxiv.org/abs/1903.12133

[20] W. B. Cannon, "The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory," *The American Journal of Psychology*, vol. 39, no. 4, pp. 106–124, 1927, [Online]. Available: https://www.jstor.org/stable/1415404

[21] S. W. Porges, "Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. A polyvagal theory," *Psychophysiology*, vol. 32, no. 4, pp. 301–318, 1995.

[22] G. Chanel and C. Mühl, "Connecting brains and bodies: applying physiological computing to support social interaction," Interact Comput, vol. 27, no. 5, pp. 534–550, 2015.

[23]   A. Bandura, "Human agency in social cognitive theory.," *American psychologist*, vol. 44, no. 9, p. 1175, 1989.

[24]   S. Samrose et al., "Visual Cues for Disrespectful Conversation Analysis," *2019 8th International Conference on Affective Computing and Intelligent Interaction, ACII 2019*, no. 2, pp. 580–586, 2019, doi: 10.1109/ACII.2019.8925440.

[25]   H. Monkaresi, N. Bosch, R. A. Calvo, and S. K. D'Mello, "Automated Detection of Engagement Using Video-Based Estimation of Facial Expressions and Heart Rate," *IEEE Trans Affect Comput*, vol. 8, no. 1, pp. 15–28, 2017, doi: 10.1109/TAFFC.2016.2515084.

[26]   V. T. Huynh, H. J. Yang, G. S. Lee, and S. H. Kim, "Engagement intensity prediction with facial behavior features," *ICMI 2019 - Proceedings of the 2019 International Conference on Multimodal Interaction,* no. June, pp. 567–571, 2019, doi: 10.1145/3340555.3355714.

[27]   T. Yokozuka, E. Ono, Y. Inoue, K. I. Ogawa, and Y. Miyake, "The relationship between head motion synchronization and empathy in unidirectional face-to-face communication," *Front Psychol*, vol. 9, no. SEP, pp. 1–10, 2018, doi: 10.3389/fpsyg.2018.01622.

[28]   "https://www.merriam-webster.com/dictionary/rapport."

[29]   S. Sharma, K. G. Gangadhara, F. Xu, A. S. Slowe, M. G. Frank, and I. Nwogu, "Coupled Systems for Modeling Rapport between Interlocutors," in *Proceedings - 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/FG52635.2021.9667067.

[30]   R. Zhao, T. Sinha, A. W. Black, and J. Cassell, "Socially-Aware Virtual Agents: Automatically Assessing Dyadic Rapport from Temporal Patterns of Behavior," *Intelligent Virtual Agents: 16th International Conference, IVA* 2016.

[31]   R. Zhao, A. Papangelis, and J. Cassell, "Towards a Dyadic Computational Model of Rapport Management for Human-Virtual Agent Interaction*," Intelligent Virtual Agents: 14th International Conference, IVA,* 2014.

[32]   R. Amini, M. Boustani, and C. Lisetti, "Modeling Rapport for Conversations About Health with Autonomous Avatars from Video Corpus of Clinician-Client Therapy Sessions," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* Springer Science and Business Media Deutschland GmbH, 2021, pp. 181–200. doi: 10.1007/978-3-030-77817-0_15.

[33]   P. Müller, M. X. Huang, and A. Bulling, "Detecting low rapport during natural interactions in small groups from non-verbal behaviour," in *International Conference on Intelligent User Interfaces, Proceedings IUI*, Association for Computing Machinery, Mar. 2018, pp. 153–164. doi: 10.1145/3172944.3172969.

[34] T.-H. Hwang, A. O. Effenberg, and H. Blume, "A Rapport and Gait Monitoring System Using a Single Head-Worn IMU during Walk and Talk," *2019 IEEE International Conference on Consumer Electronics (ICCE),* 2019.

[35] C. C. Yang, J. Yen, and J. Liu, "Social intelligence and technology," *IEEE Intell Syst*, vol. 29, no. 2, pp. 5–8, 2014, doi: 10.1109/MIS.2014.28.

[36] D. McDuff, K. Rowan, P. Choudhury, J. Wolk, T. Pham, and M. Czerwinski, "A Multimodal Emotion Sensing Platform for Building Emotion-Aware Applications," pp. 1–6, 2019, [Online]. Available: http://arxiv.org/abs/1903.12133

[37] G. Leshed et al., "Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups," C*onference on Human Factors in Computing Systems - Proceedings,* pp. 537–546, 2009, doi: 10.1145/1518701.1518784.

[38] Y. R. Tausczik and J. W. Pennebaker, "Improving teamwork using real-time language feedback," *Conference on Human Factors in Computing Systems - Proceedings*, pp. 459–468, 2013, doi: 10.1145/2470654.2470720.

[39] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-End Multimodal Emotion Recognition Using Deep Neural Networks," *IEEE Journal on Selected Topics in Signal Processing*, vol. 11, no. 8, pp. 1301–1309, Dec. 2017, doi: 10.1109/JSTSP.2017.2764438.

[40] S. Mirsamasdi, E. Barsoum, and C. Zhang, "Automatic Speech Emotion Recognition Using Recurrent Neuralnetworks With Local Attention," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2017, pp. 2227–2231.

[41] R. Haux, S. Koch, N. H. Lovell, M. Marschollek, N. Nakashima, and K. H. Wolf, "Health-Enabling and Ambient Assistive Technologies: Past, Present, Future," *Yearb Med Inform*, pp. S76–S91, Jun. 2016, doi: 10.15265/IYS-2016-s008.

[42] S. Abdi, I. Kitsara, M. S. Hawley, and L. P. de Witte, "Emerging technologies and their potential for generating new assistive technologies," *Assistive Technology*, vol. 33, no. sup1, pp. 17–26, 2021, doi: 10.1080/10400435.2021.1945704.

[43] R. He et al., "Optical fiber sensors for heart rate monitoring: A review of mechanisms and applications," *Results in Optics*, vol. 11, p. 100386, 2023.

[44] C. Nwibor et al., "Remote health monitoring system for the estimation of blood pressure, heart rate, and blood oxygen saturation level," *IEEE Sens J*, vol. 23, no. 5, pp. 5401–5411, 2023.

[45]   T. Curran, X. Liu, D. McDuff, S. Patel, and E. Yang, "Camera-based remote photoplethysmography to measure heart rate and blood pressure in ambulatory patients with cardiovascular disease: Preliminary analysis," *J Am Coll Cardiol*, vol. 81, no. 8_Supplement, p. 2301, 2023.

[46]   H. U. R. Siddiqui et al., "Emotion classification using temporal and spectral features from IR-UWB-based respiration data," *Multimed Tools Appl*, vol. 82, no. 12, pp. 18565–18583, 2023.

[47]   A. Angelucci and A. Aliverti, "An IMU-Based Wearable System for Respiratory Rate Estimation in Static and Dynamic Conditions," *Cardiovasc Eng Technol*, vol. 14, no. 3, pp. 351–363, 2023.

[48]   P. S. Kumar, P. K. Govarthan, N. GANAPATHY, and J. A. C. F. A. RONICKOM, "A comparative analysis of eda decomposition methods for improved emotion recognition," *J Mech Med Biol*, vol. 23, no. 06, p. 2340043, 2023.

[49]   W. Lin and C. Li, "Review of studies on emotion recognition and judgment based on physiological signals," *Applied Sciences*, vol. 13, no. 4, p. 2573, 2023.

[50]   M. Lee, S. Lee, S. Hwang, S. Lim, and J. H. Yang, "Effect of emotion on galvanic skin response and vehicle control data during simulated driving," *Transp Res Part F Traffic Psychol Behav*, vol. 93, pp. 90–105, 2023.

[51]   L. Crucianelli, D. Radziun, and H. H. Ehrsson, "Thermosensation and emotion: Thermosensory accuracy in a dynamic thermal matching task is linked to depression and anxiety symptomatology," *Physiol Behav*, vol. 273, p. 114407, 2024.

[52]   M. Jafari et al., "Emotion recognition in EEG signals using deep learning methods: A review," *Comput Biol Med*, p. 107450, 2023.

[53]   V. Skaramagkas et al., "eSEE-d: Emotional state estimation based on eye-tracking dataset," *Brain Sci*, vol. 13, no. 4, p. 589, 2023.

[54]   A. A. A. Zamkah, "Reusable electrochemical impedance spectroscopy biosensor for the detection of cortisol in sweat: Introducing novel techniques suitable for future affective wearable devices and emotional stress," University of Reading, 2023.

[55]   A. Zamkah, W. Lu, S. M. Mugo, and R. S. Sherratt, "Reusable Electrochemical Impedance Spectroscopy Cortisol Sweat Chemosensor as a Novel Emotional Stress Wearable Device," *Available at SSRN 4457316*.

[56] Y. Teng, C. K. Singh, O. Sadak, N. Ahmad, and S. Gunasekaran, "Electrochemical detection of mobile zinc ions for early diagnosis of prostate cancer," *Journal of Electroanalytical Chemistry*, vol. 833, pp. 269–274, Jan. 2019, doi: 10.1016/j.jelechem.2018.12.002.

[57] A. Barhoum, S. Hamimed, H. Slimi, A. Othmani, F. M. Abdel-Haleem, and M. Bechelany, "Modern designs of electrochemical sensor platforms for environmental analyses: Principles, nanofabrication opportunities, and challenges," *Trends in Environmental Analytical Chemistry*, vol. 38, p. e00199, 2023.

[58] L. Rotariu, F. Lagarde, N. Jaffrezic-Renault, and C. Bala, "Electrochemical biosensors for fast detection of food contaminants–trends and perspective," *TrAC Trends Anal. Chem.*, vol. 79, pp. 80–87, 2016.

[59] M. H. Abdalhai, A. M. Fernandes, X. Xia, A. Musa, J. Ji, and X. Sun, "Electrochemical Genosensor to Detect Pathogenic Bacteria (Escherichia coli O157:H7) As Applied in Real Food Samples (Fresh Beef) to Improve Food Safety and Quality Control," *J Agric Food Chem*, vol. 63, no. 20, pp. 5017–5025, May 2015, doi: 10.1021/ACS.JAFC.5B00675.

[60] P. Swetha, U. Balijapalli, and S. Feng, "Wireless accessing of salivary biomarkers based wearable electrochemical sensors: A mini-review," *Elsevier*, vol. 140, 2022.

[61] K. Ramaiyan and R. Mukundan, "Electrochemical sensors for air quality monitoring," *Electrochem Soc Interface*, 2019.

[62] R. Meng et al., "The innovative and accurate detection of heavy metals in foods: A critical review on electrochemical sensors," *Food Control,* p. 109743, 2023.

[63] H. Liu and C. Zhao, "Wearable electrochemical sensors for noninvasive monitoring of health—a perspective," *Curr. Opin. Electrochem*., vol. 23, pp. 42–46, 2020.

[64] M. D. Steinberg, P. Kassal, and I. M. Steinberg, "System Architectures in Wearable Electrochemical Sensors," *Electroanalysis*, vol. 28, no. 6, pp. 1149–1169, Jun. 2016, doi: 10.1002/ELAN.201600094.

[65] H. Li, X. Liu, L. Li, X. Mu, R. Genov, and A. J. Mason, "CMOS Electrochemical Instrumentation for Biosensor Microsystems: A Review," *Sensors*, vol. 17, no. 1, p. 74, Dec. 2017, doi: 10.3390/s17010074.

[66] J. Tsai, Y. Chen, and Y. Liao, "A power-efficient bidirectional potentiostat-based readout IC for wide-range electrochemical sensing," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS),* 2018.

[67] P. Li, T. Molderez, D. Villamor, A. Prévoteau, and M. Verhelst, "A 96-channel 40nm CMOS potentiostat for parallel experiments on microbial electrochemical systems," *IEEE*

Transactions on Circuits and Systems I: Regular Papers, vol. 70, no. 1, 2023, doi: 10.1109/TCSI.2022.3214470.

[68] P. Jothimuthu, R. A. Wilson, J. Herren, E. N. Haynes, W. R. Heineman, and I. Papautsky, "Lab-on-a-chip sensor for detection of highly electronegative heavy metals by anodic stripping voltammetry," *Biomed Microdevices*, vol. 13, no. 4, pp. 695–703, Aug. 2011, doi: 10.1007/s10544-011-9539-1.

[69] J. A. Harrigan, T. E. Oxman, and R. Rosenthal, "RAPPORT EXPRESSED THROUGH NONVERBAL BEHAVIOR," 1985.

[70] L. Tickle-Degnen and R. Rosenthal, "The Nature of Rapport and Its Nonverbal Correlates," 1990. [Online]. Available: http://www.jstor.org/about/terms.html.

[71] "Facial Action Coding System." Accessed: May 25, 2023. [Online]. Available: https://en.wikipedia.org/wiki/Facial_Action_Coding_System

[72] "https://imotions.com/blog/learning/research-fundamentals/facial-action-coding-system/."

[73] C. L. Kleinke, "Gaze and eye contact: a research review.," *Psychol Bull,* vol. 100, no. 1, p. 78, 1986.

[74] "GazePointer," https://gazerecorder.com/gazepointer/.

[75] "OpenFace," https://github.com/TadasBaltrusaitis/OpenFace.

[76] "https://www.techsmith.com/video-editor.html."

[77] "https://zeromq.org/."

[78] Sutton, R.S. and Barto, A.G., 2018. *Reinforcement learning: An introduction.* MIT press.

[79] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Quantized neural networks: Training neural networks with low precision weights and activations," *Journal of Machine Learning Research*, vol. 18, pp. 1–30, 2018.

[80] "https://en.wikipedia.org/wiki/Affect_(psychology)."

[81] "https://dictionary.apa.org/affect," APA Dictionary of Psychology.

[82] S. Samrose and E. Hoque, "Quantifying the Intensity of Toxicity for Discussions and Speakers," in *2021 9th International Conference on Affective Computing and Intelligent*

Interaction Workshops and Demos, ACIIW 2021, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/ACIIW52867.2021.9666258.

[83]   S. D. Tomasi, N. N. Parolia, C. Han, and T. Porterfield, "Exploring the impact of team rapport and empowerment on information processing and project performance in outsourced system development," *International Journal of Project Organisation and Management*, vol. 7, no. 3, pp. 284–305, 2015, doi: 10.1504/IJPOM.2015.070794.

[84]   S. Morrison-Smith and J. Ruiz, "Challenges and barriers in virtual teams: a literature review," *SN Appl Sci*, vol. 2, no. 6, Jun. 2020, doi: 10.1007/s42452-020-2801-5.

[85]   J. L. Hagad, R. Legaspi, M. Numao, and M. Suarez, "Predicting levels of rapport in dyadic interactions through automatic detection of posture and posture congruence," in *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, 2011, pp. 613–616. doi: 10.1109/PASSAT/SocialCom.2011.143.

[86]   E. Ashoori, S. Dávila-Montero, and A. J. Mason, "Analyzing Affect During Virtual Meetings to Improve Quality of Interaction," in *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC),* 2023, pp. 678–681. doi: 10.1109/CCWC57344.2023.10099285.

[87]   S. Dávila-Montero, "Real-Time Human/Group Interaction Monitoring Platform Integrating Sensor Fusion and Machine Learning Approaches," 2022.

[88]   L. Tickle-Degnen and R. Rosenthal, "The nature of rapport and its nonverbal correlates," *Psychol Inq,* vol. 1, no. 4, pp. 285–293, 1990.

[89]   F. J. Bernieri, J. S. Gillis, J. M. Davis, and J. E. Grahe, "Dyad rapport and the accuracy of its judgment across situations: a lens model analysis.," *J Pers Soc Psychol*, vol. 71, no. 1, p. 110, 1996.

[90]   J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, "Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders," *J Neurosci Methods*, vol. 200, no. 2, pp. 237–256, Sep. 2011, doi: 10.1016/j.jneumeth.2011.06.023.

[91]   T. Baltrusaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic Action Unit detection," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG),* IEEE, May 2015, pp. 1–6. doi: 10.1109/FG.2015.7284869.

[92]   "Action Units." Accessed: Apr. 26, 2024. [Online]. Available: https://github.com/TadasBaltrusaitis/OpenFace/wiki/Action-Units

[93] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "OpenFace 2.0: Facial Behavior Analysis Toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018),* IEEE, May 2018, pp. 59–66. doi: 10.1109/FG.2018.00019.

[94] T. Baltrusaitis, P. Robinson, and L.-P. Morency, "Constrained Local Neural Fields for Robust Facial Landmark Detection in the Wild," in *2013 IEEE International Conference on Computer Vision Workshops*, IEEE, Dec. 2013, pp. 354–361. doi: 10.1109/ICCVW.2013.54.

[95] A. Zadeh, Y. C. Lim, T. Baltrusaitis, and L.-P. Morency, "Convolutional Experts Constrained Local Model for 3D Facial Landmark Detection," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, IEEE, Oct. 2017, pp. 2519–2528. doi: 10.1109/ICCVW.2017.296.

[96] "An introduction to Dynamic Time Warping." Accessed: Apr. 26, 2024. [Online]. Available: https://rtavenar.github.io/blog/dtw.html

[97] "Accuracy vs Precision vs Recall." Accessed: May 07, 2024. [Online]. Available: https://www.evidentlyai.com/classification-metrics/accuracy-precision-recall

[98] "Precision and Recall." Accessed: May 07, 2024. [Online]. Available: https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall

[99] "F1 Score." Accessed: May 07, 2024. [Online]. Available: https://www.v7labs.com/blog/f1-score-guide

[100] M. Steinberg, P. Kassal, I. Kereković, and I. M. Steinberg, "A wireless potentiostat for mobile chemical sensing and biosensing," Talanta, vol. 143, pp. 178–183, 2015.

[101] E. Ashoori, S. Parsnejad, and A. J. Mason, "Neurochemical Sensing," in *Handbook of Neuroengineering*, Springer Nature Singapore, 2023, pp. 591–621. doi: 10.1007/978-981-16-5540-1_134.

[102] J. Tsai, C. Kuo, S. Lin, F. Lin, and Y. Liao, "A wirelessly powered CMOS electrochemical sensing interface with power-aware RF-DC power management," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 9, 2018, doi: 10.1109/TCSI.2018.2797238.

[103] Y. Chen, S. Lu, and Y. Liao, "A microwatt dual-mode electrochemical sensing current readout with current-reducer ramp waveform generation," *IEEE Trans Biomed Circuits Syst*, vol. 13, no. 6, 2019, doi: 10.1109/TBCAS.2019.2936373.

[104] K. Nam et al., "A low-noise and mismatch-tolerant current-mirror-based potentiostat circuit for glucose monitoring," *Microelectronics J*, vol. 132, Feb. 2023, doi: 10.1016/j.mejo.2023.105694.

[105] R. Ahmad, A. M. Joshi, and D. Boolchandani, "Programmable Readout With Integrated Bandgap Reference Potentiostat for Glucose Sensing," *IEEE Trans Instrum Meas*, vol. 73, pp. 1–11, 2024, doi: 10.1109/TIM.2024.3369142.

[106] L. Keeble, A. Jaccottet, D. Ma, J. Rodriguez-Manzano, and P. Georgiou, "An electroplated Ag/AgCl quasi-reference electrode based on CMOS top-metal for electrochemical sensing," *Electrochim Acta*, vol. 477, Feb. 2024, doi: 10.1016/j.electacta.2024.143780.

[107] M. Horowitz, E. Alon, D. Patil, S. Naffziger, R. Kumar, and K. Bernstein, "Scaling, power, and the future of CMOS," in *IEEE InternationalElectron Devices Meeting*, 2005. IEDM Technical Digest., 2005. doi: 10.1109/IEDM.2005.1609253.

[108] E. Ashoori, D. Goderis, A. Inohara, and A. J. Mason, "Wide Voltage Swing Potentiostat with Dynamic Analog Ground to Expand Electrochemical Potential Windows in Integrated Microsystems," *Sensors*, vol. 24, no. 9, 2024, doi: 10.3390/s24092902.

[109] "MOSIS Foundry Service." Accessed: Apr. 20, 2024. [Online]. Available: https://themosisservice.com/foundry-services

[110] J. Barón-Jaimez, M. R. Joya, and J. Barba-Ortega, "Bismuth electrodes, an alternative in stripping voltammetry," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, 2013. doi: 10.1088/1742-6596/466/1/012025.

[111] K. Xu and C. Wang, "Batteries: Widening voltage windows," *Nature Energy*, vol. 1, no. 10, Nature Publishing Group, Oct. 06, 2016. doi: 10.1038/nenergy.2016.161.

[112] S. Kumar, B. Tudu, R. Bandyopadhyay, and A. Ghosh, "An equivalent electrical network of an electronic tongue: A case study with tea samples," in *2017 ISOCS/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN)*, 2017. doi: 10.1109/ISOEN.2017.7968930.

[113] A. Bard, L. Faulkner, and H. White, *Electrochemical Methods: Fundamentals and Applications,* 3rd ed. John Wiley & Sons, 2022.