IM-TWIN: from Intrinsic Motivations to Transitional Wearable INtelligent companions for autism spectrum disorder

a European funded project

Affect and emotional classification 1

Deliverable 3.1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952095.

Project duration 36 months (November 2020, October 2023). Consortium: Consiglio Nazionale dele Ricerche (ITA), Universiteit Utrecht (NLD), Centre de Recherches Interdisciplinaires (FRA), Università degli Studi di Roma La Sapienza (ITA), Plux-Wireless Biosignals S.A. (PRT).

Deliverable data

Work Package:	3 Affect classification, PlusMe AI, and IM-TWIN integration
Work Package leader:	CNR
Deliverable beneficiary:	UU
Dissemination level:	Public
Due date:	31th December (Month 26)
Type:	Report
Authors:	Lukas P.A. Arts, E.L. van den Broek

Acronyms of partners

CNR-ISTC	Consiglio Nazionale delle Ricerche, Istituto di Scienze e Tecnologie della Cognizione (Italy)
UU	Universiteit Utrecht (The Netherlands)
CRI	Centre de Recherches Interdisciplinaires (France)
LA SAPIENZA	Università degli Studi di Roma La Sapienza (Italy)
PLUX	Plux - Wireless Biosignals S.A. (Portugal)

Table of contents

1. Overview of the deliverable

This report describes the initial developments of a multi-faceted IM-TWIN affective classification processing pipeline designed to classify the emotional state of a patient using physiological biosignals. Signals are obtained using the custom designed T-Shirt with embedded textile-electrodes by PLUX. As the IM-TWIN system is expected to be primarily used in noisy everyday situations, the system has to be designed durable, robust and noise-resilient. In this report we describe the work on the specific processing pipelines for ElectroCardioGraphy (ECG) and ElectroDermal Activity (EDA) analysis. Both biosignals are known measures for arousal. Additionally, there are hints that ECG is also usable for valence estimation. Together they allow the estimation of emotions using the valence-arousal model.

In the next section, we will describe the theory behind emotional models including the valence-arousal model. Next, we zoom in on the ECG and EDA specific processing pipelines. We end with a concise discussion and conclusion describing future steps.

2. Emotional models

Emotion is a complex phenomenon that has been studied extensively by psychologists, neuroscientists, and philosophers for many years [1,2]. One of the challenges in studying emotions is developing a model that can accurately capture the range of experiences people have. There are several emotion models that have been proposed over the years, each with their own strengths and weaknesses.

One of the earliest models of emotions was the basic emotions model, which posits that there are a set of discrete, universal emotions that are biologically hardwired into humans. Basic emotions include but are not limited to happiness, sadness, anger, fear, disgust, and surprise [3]. While this model provides a useful framework for understanding emotions, it has been criticized for oversimplifying the complexity of human emotional experiences. As a consequence, the dimensional approach was invented which suggests that emotions can be described along three primary dimensions: valence, arousal, and dominance [1,4]. Valence refers to how positive or negative an emotion is, arousal to how intense the emotion is, and dominance to how much control a person has or to what extent an individual feels restricted in his behavior [1]. Often, the third dimension, dominance, is ignored for simplification [5]. Dominance is simply too difficult to measure using only biosignals.

The valence-arousal approach is popular in psychophysiological research because both map directly to specific biosignals [2]. Arousal has been shown to correlate strongly with EDA, EMG, and ECG. Valence, on the other hand, correlates with facial EMG and brain activity (EEG). As such, in the scope of the IM-TWIN project, we use the valence and arousal dimensions to map

the distinct emotional states of the child. Specifically, we identify three major states: i) low arousal, ii) high-arousal, negative-valence, and iii) high-arousal, positive-valence.

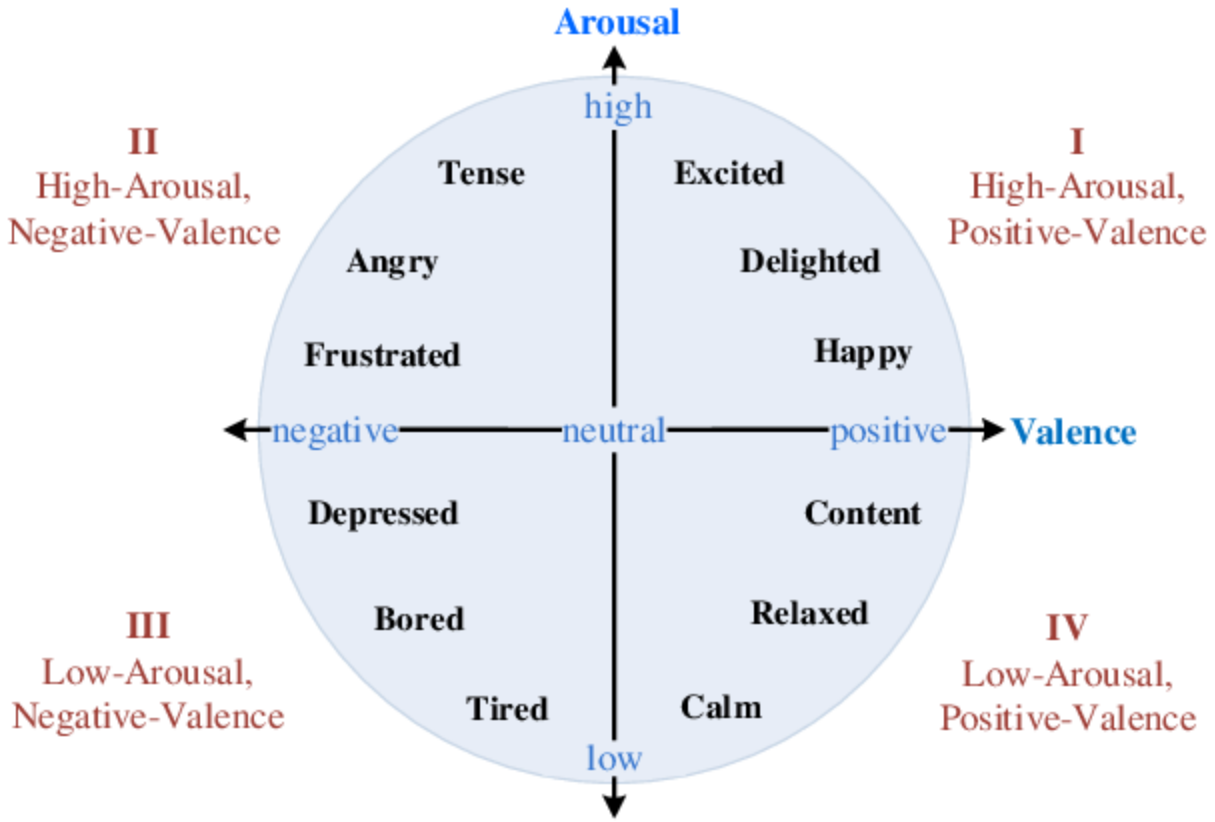

***Figure 1**: The valence-arousal model maps emotions to a 2D space where the x-axis represents the valence and the y-axis the level of arousal. Discrete emotions are shown to show the mapping between both the discrete and continuous emotional models.*

To categorize the child's emotions into these three major states, we use ECG and EDA measured via textile biosensors embedded into a T-Shirt developed by hardware manufacturer PLUX. Typically, the Skin-Conductance Level (SCL) and Skin-Conductance Responses (SCRs) of EDA are used to estimate the level of arousal physiologically. Heart Rate (HR) and Heart Rate Variability (HRV) are also known to be good estimators of arousal, but tend to be used for valence estimation as well. However, the quality of such valence estimators is questionable. Unfortunately, golden standard measures for valence such as facial EMG and brain activity (EEG) are not an option. Facial EMG would be too obtrusive for the child and EEG is unusable without clinically validated wet electrodes which are tedious to set up. Hence, EDA and ECG are IM-TWIN's two primary biosignals.

3. Physiological signal processing

The T-Shirt developed by PLUX incorporates several biosensors, three of which have been identified as indicators of emotional and/or mental state: i) ECG, ii) EDA and iii) temperature. It is important to note that these signals differ in their response time, with temperature being the slowest to respond, taking 15-20 seconds [5]. EDA reacts significantly faster with changes observable a few seconds after a stimulus. ECG has the fastest response time, taking only 1 to 2 seconds [5].

As real-time emotion estimation is crucial for the scope of IM-TWIN, temperature cannot be used as a direct measure. Still, temperature could serve as a correction measure for EDA [2,5]. Regarding IM-TWIN's real-time requirements, ECG seems to be the best fit. However, ECG cannot serve as a reliable valence-arousal estimator on its own due to its high sensitivity to noise, which distorts the signal. In contrast, EDA is less sensitive to noise, but slower to respond. Thus, a combination of ECG and EDA is proposed as the optimal solution. In the following sections, we will provide a detailed analysis of both signals and present a custom processing pipeline specifically designed to operate on real-world recordings, which are inherently noisy.

3.1 ElectroCardioGraphy (ECG)

ElectroCardioGraphy (ECG) is a medical diagnostic tool that records the electrical activity of the heart over a period of time using electrodes attached to the skin. It was first introduced in the late 1880s by Waller [6], who developed a system to record the electrical signals generated by the heart and display them as a graph known as an electrocardiogram. ECG has since become a widely used and essential tool for the diagnosis of various cardiovascular diseases and abnormalities. By analyzing the amplitude, duration, and shape of the waves created by the contractions of the different heart muscles, clinicians can diagnose various cardiac conditions [7] and monitor the effectiveness of treatments [8].

In recent years, ECG has also been used as a non-invasive method to estimate emotional and mental states [2], making it an increasingly valuable tool in the field of affective computing, especially in the scope of IM-TWIN.

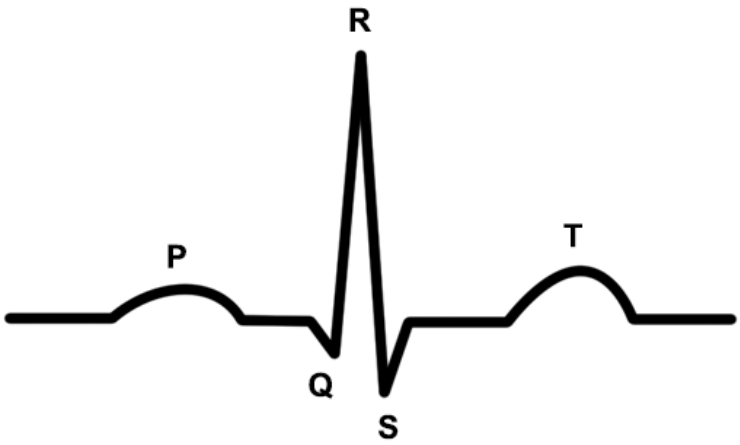

***Figure 2**: The ElectroCardioGram (ECG) waveform consists of several distinct components. The P wave, QRS complex and T wave. Generally, the QRS complex is used for HR and HRV extraction due to its striking morphology.*

ECG has a characteristic waveform that is made up of several distinct waves and components (see Figure 2). These waves are labeled P, Q, R, S, and T. The P wave represents the depolarization of the atria, which is the contraction of the atrial muscles as they push blood into the ventricles. The QRS complex represents the depolarization of the ventricles, which is the contraction of the ventricular muscles as they pump blood out of the heart. Finally, the T wave represents the repolarization of the ventricles as they prepare for another cycle of contraction and relaxation.

Of these waves, the QRS complex plays a crucial role in determining HR and HRV. In HR analysis, QRS complex detection serves as the primary indicator of heartbeats. By measuring the time between successive QRS complexes, clinicians can calculate the heart rate, which is a fundamental metric for assessing the physical state of the body. Similarly, in HRV analysis, QRS complex detection is used to determine the variability in the time intervals between successive heartbeats. This variability is a crucial parameter for assessing the health of the autonomic nervous system, which plays a significant role in regulating heart rate and blood pressure.

In both HR and HRV analysis, QRS complex detection acts as the first step in the processing pipeline. Without QRS complex detection, heart beats cannot be separated, and no HR or HRV can be extracted. When ECG is measured in controlled, clinical settings using wet, adhesive electrodes, QRS complex detection is a trivial problem that has already been solved decades ago [9]. Their odd morphology makes them easily detectable using simple slope-based detectors. However, when dry electrodes are used in noisy, every-day environments, peak detection becomes much more challenging. Sudden bursts of noise are easily mistaken for QRS complexes (see Figure 3).

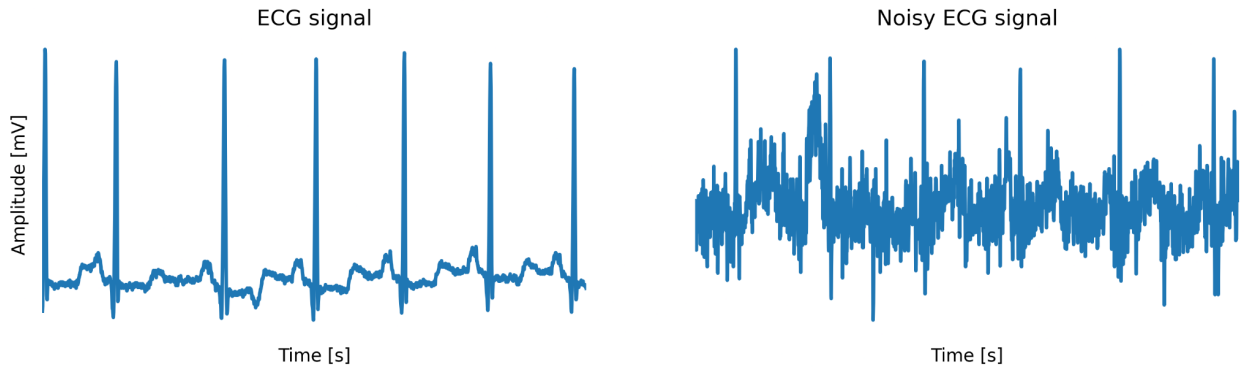

***Figure 3**: ECG recorded with wet, adhesive electrodes in controlled, clinical environments poses no challenge for QRS complex detection. However, when ECG is obtained using dry electrodes in noisy environments, QRS detection becomes much harder.*

Reliable detection of QRS complexes in clean and noisy ECG has already been a research topic for four decades. Initially, the development of automated QRS complex detection was difficult due to the availability of annotated datasets. This all changed when the first open-source datasets were released in the early ‘80s [10]. Suddenly, QRS complex detection became a problem that anyone could solve without the need for clinically validated equipment or patients. Consequently, the first detection algorithms began to appear short after. A very popular detector from that time was developed by Pan and Tompkins[9]. Their detector is still the most accessible and, hence, most used QRS detector today. Besides Pan and Tompkins’ algorithm, many techniques have been developed. Together they make up the vast algorithmic landscape we experience today.

In IM-TWIN, we need a QRS detector that is both fast and accurate. However, there are numerous QRS detection algorithms making it difficult to select one. Hence, we performed a concise benchmark of 4 algorithms [9,11,12,13] on 2 datasets. The algorithms were chosen to represent four different strategies towards QRS detection: derivative-based, decomposition-based, deep learning-based, and morphology-based detection. To test performance on clean ECG data, the MIT-BIH dataset [10] was selected; a clinical dataset recorded with wet electrodes. The Check-Your-Biosignals-Here initiative database [14] was chosen to represent the ECG we expect in IM-TWIN; noisy, but usable ECG recorded with dry electrodes. Benchmark results are shown in Figure 4.

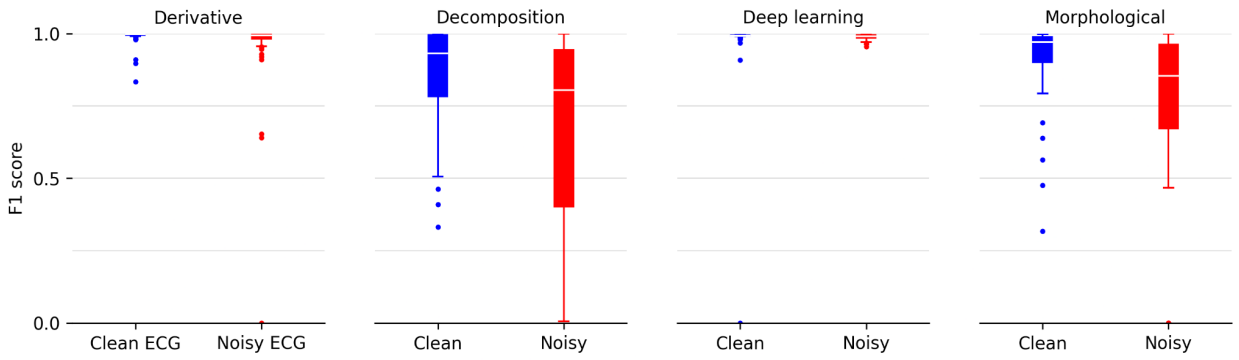

***Figure 4**: Accuracy in terms of F1-score of four QRS complex detection algorithms on a clean and noisy ECG dataset. Boxplots show the distribution of accuracies per dataset across all recordings.*

Pan-Tompkins' derivative-based algorithm and the deep learning Convolutional Neural Network (CNN) based algorithm [12] performed best on the clean and noisy datasets. Others suffered from large deviations in accuracy across records. Hence, they do not show reliability and cannot be used in the scope of the IM-TWIN project. When we compare Pan-Tompkins and the CNN-based algorithm, we see that both have very small deviations in accuracy. However, Pan-Tompkins still fails on two noisy ECG records. The QRS detection performance of both algorithms on one of them is shown in Figure 5.

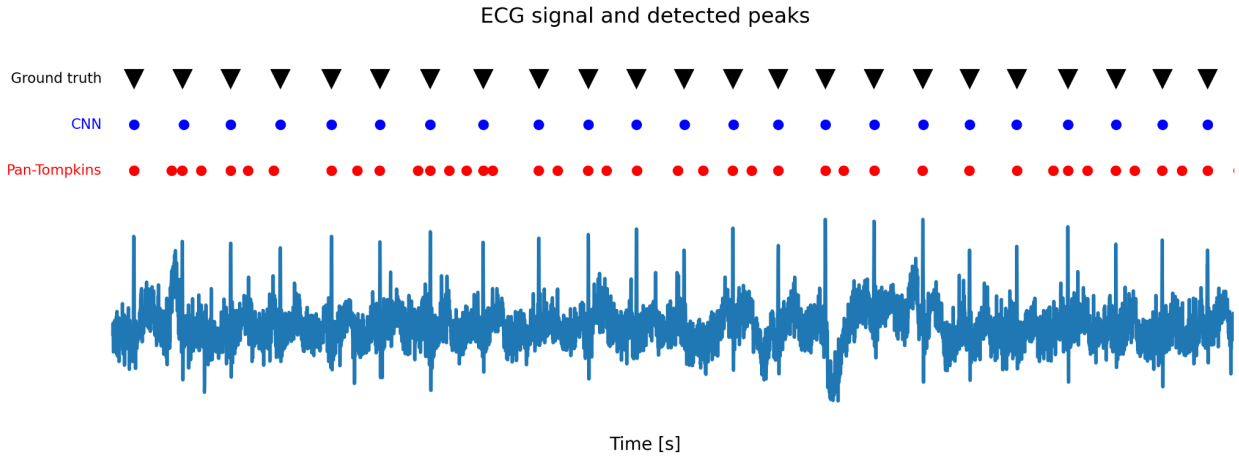

***Figure 5**: The 1985 Pan-Tompkins detector [9] and recent CNN-based deep learning detector [12] compared on a noisy record from the CYBHi dataset. Pan-Tompkins becomes unstable due to the high noise level. In contrast, the CNN-based detector makes no mistakes.*

Figure 5 provides a clear distinction between Pan-Tompkins and the CNN-based detector when faced with severe noise. In such situations, the Pan-Tompkins detector becomes unstable, generating many false positives. As further analysis uses the time between two consecutive beats, one can easily see that HR and HRV estimations become meaningless. Pan-Tompkins only shows this behavior on two records. However, the ECG quality of both records falls well

within the range of ECG quality we expect to see in IM-TWIN. Hence, Pan-Tompkins shows to be too unstable. The deep learning algorithm is, therefore, favored over the others. Unfortunately, deep learning does not come without limitations. In this case, computational cost. When comparing all algorithms in terms of computational cost, the results are clear. Figure 6 shows the number of calculations (additions and multiplications) needed to process one second of 1000 Hz ECG.

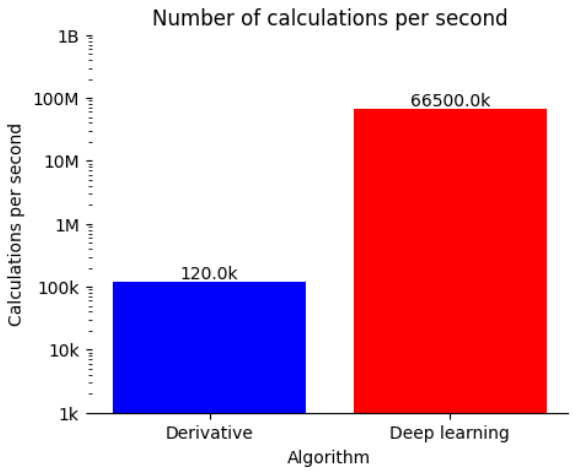

***Figure 6**: The Pan-Tompkins and CNN-based deep learning detector compared on computational complexity. The CNN-based detector needs 66 million additions and multiplications per second of ECG, whereas Pan-Tompkins only needs 120 thousand.*

The deep learning-based CNN detector is 550x more expensive to execute than the derivative based Pan-Tompkins. In practice, this translates to a 550x higher power usage or 550x slower calculation. Both are very unpleasant in the scope of IM-TWIN's requirements.

Processing noisy, real-world ECG in resource and/or time limited environments remains challenging. On the one hand, deep learning achieves superior detection accuracy at an astronomical cost. On the other hand, traditional derivative-based algorithms are efficient but become unstable in noisy conditions. Consequently, new approaches have to be explored to merge the best of both worlds: fast and accurate QRS detection.

3.2 ElectroDermal Activity (EDA)

Electrodermal activity (EDA) is a measure of the electrical conductance of the skin, which reflects changes in the activity of sweat glands in response to various stimuli. Also known as skin conductance or galvanic skin response, EDA was discovered by Vigouroux in the late 19th century to reflect one's psychological state [15]. Vigouroux observed that the skin's resistance to electrical current decreased when the subject was emotionally aroused. Since then, EDA has been widely used as a tool for psychophysiological research, particularly in the fields of emotion and stress, where it is used to measure changes in sympathetic nervous system activity [2,15].

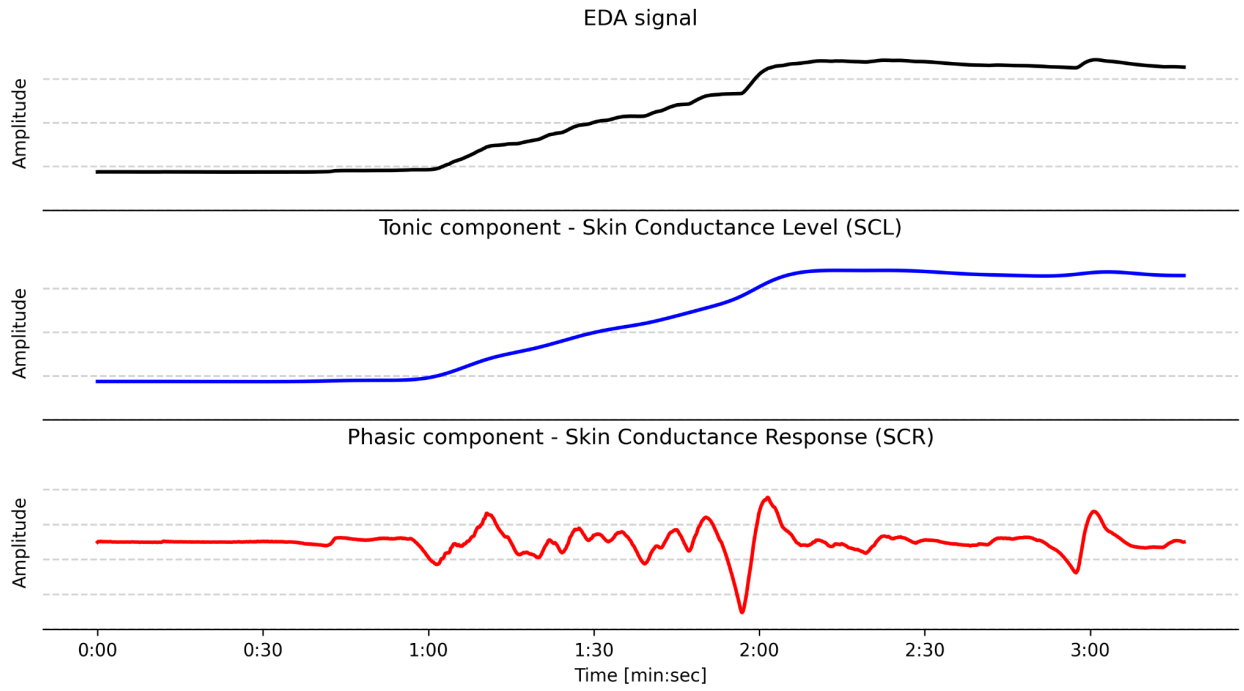

***Figure 7**: Top: Raw EDA signal. Middle: Tonic component reflecting the low-frequency skin conductance level (SCL). Bottom: Phasic component reflecting the higher-frequency skin conductance response (SCR).*

EDA is generally analysed in the time-domain. Typically, EDA is split into a tonic and phasic part [15]. The tonic component represents long-term Skin Conductance Level (SCL) and the phasic part the short-term Skin Conductance Responses (SCRs). The latter can be both specific or non-specific. Specific SCRs occur within a predefined interval after a stimulus. In other words, they can be directly related to a specific stimulus. Non-specific SCRs occur randomly. Researchers have hypothesized that the intensity and frequency of such NS-SCRs is related to mental state. In IM-TWIN, we only focus on NS-SCRs as there are, by definition, no specific SCRs in IM-TWIN's real-world environment.

EDA analysis in the context of IM-TWIN is restricted to the analysis of non-specific skin conductance responses (NS-SCRs). However, there is little research examining the direct relationship between NS-SCRs and arousal, as defined in the valence-arousal model. Most studies that investigate EDA tend to focus on physical and long-term stress estimation. It remains unclear whether NS-SCRs and emotional stress are linked to lower levels of arousal, or if they enable the continuous estimation of arousal, rather than only act as a discriminator between low and high levels of arousal. Consequently, we set out to compare two descriptive features of NS-SCRs to real-time self-reported arousal, using a unique dataset called the Continuously Annotated Signals of Emotion (CASE) dataset [16]. This dataset contains recordings of various biosignals, including EDA, along with continuous annotations of emotional state by participants. Participants were asked to annotate their feelings while they viewed

emotionally stimulating videos. As the annotations are continuous, it is possible to correlate EDA and emotional state in real-time.

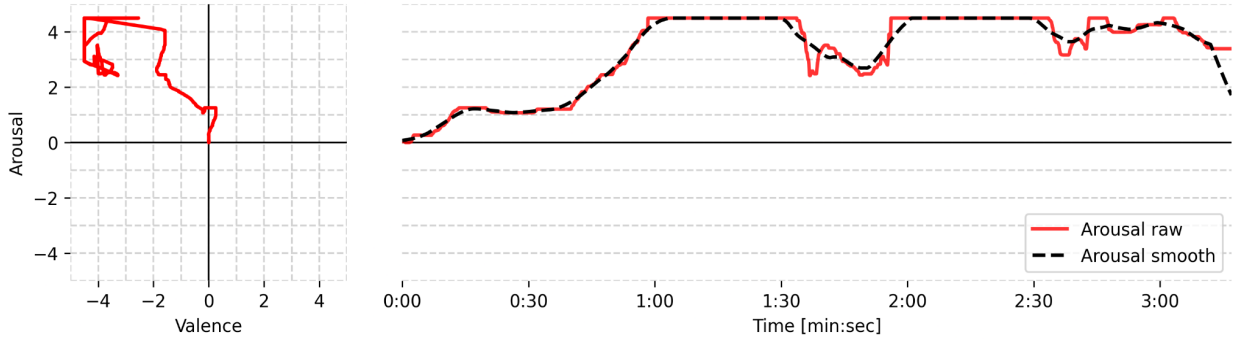

***Figure 8**: In the Continuously Annotated Signals of Emotion (CASE) dataset, participants annotate their feelings continuously on a 2D valence-arousal plane using a joystick. We only extracted the arousal axis to study its correlation with EDA.*

In CASE [16], the self-reported annotations are obtained with a joystick that the participant controls during the entire experiment. As such, annotations are represented as trajectories in Cartesian space where the x and y-axis represents valence and arousal, respectively (see Figure 8). Normalization was performed to correct for the personal bias across participants (i.e., one person might annotate more intensely in general, compared to others).

The Neurokit2 package in Python [17] is utilized to extract the NS-SCRs from the EDA signal. The signal is then partitioned into 30-second windows to compute the frequency and average amplitude of the NS-SCRs. Concurrently, a 30-second moving average filter is applied to the vertical axis of the joystick trajectory. The average arousal is then determined in the same 30-second windows as the EDA signal. Finally, the correlation between the NS-SCR frequency and average amplitude and average arousal is evaluated using a Pearson correlation test (see Figure 9). This process is repeated for every video and participant. Results are categorized per video as not every video was designed to elicit arousal. As such, the videos that were selected to be ‘boring’ or ‘relaxing’ were grouped as ‘low arousal’ videos. The ‘amusing’ videos were categorized as ‘medium arousal’ and ‘scary’ videos as ‘high arousal’ contexts. Results are shown in Figure 10

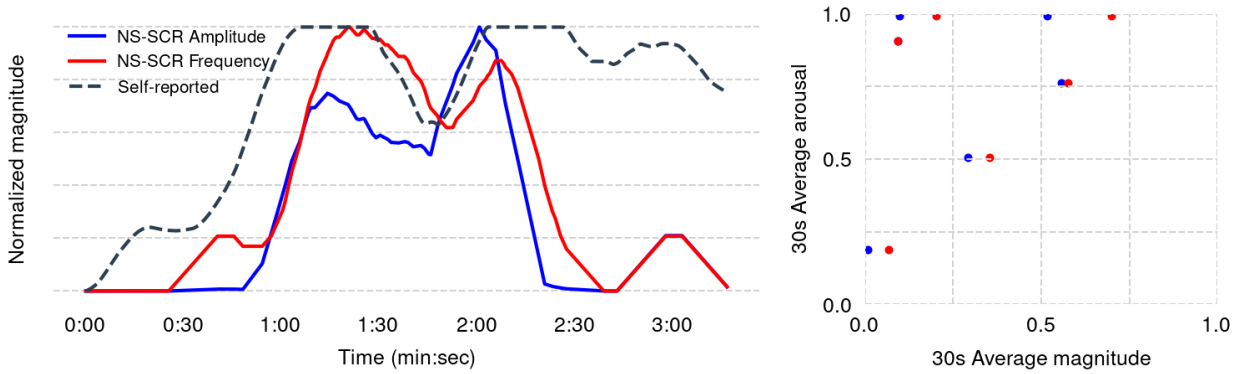

***Figure 9**: Left: The frequency and amplitude of Non-Specific Skin Conductance Responses (NS-SCRs) is plotted together with the arousal axis of the continuous annotations. Right: The average frequency and amplitude within 30s windows is calculated and plotted against average arousal. Based on this plot, a Pearson Correlation test is performed to calculate correlation.*

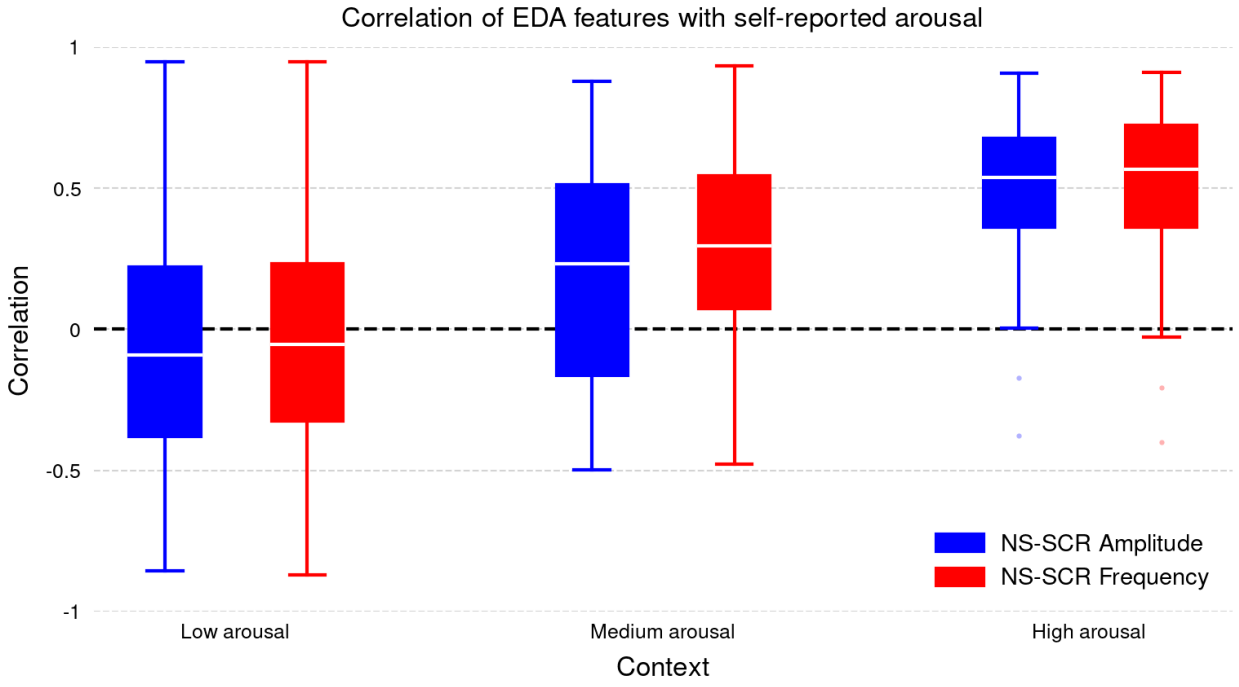

***Figure 10**: Distribution of correlation coefficients for frequency and amplitude of 'non-specific' Skin Conductance Responses (NS-SCRs) with self-reported arousal. Distribution is categorized into three arousal contexts based on the videos that participants were watching. Boring and relaxing videos were labeled as 'low arousal' contexts. Amusing as 'medium arousal' and scary as 'high arousal' contexts.*

In low arousal contexts, we observed no significant correlations. Conversely, when a context elicits a sufficiently strong stimulus, such as medium and high arousal contexts, significant and

strong correlations between self-perceived emotional arousal and EDA features were identified (corr.=0.5). This suggests the existence of a threshold, whereby arousal levels above a certain threshold are directly reflected in the physiology. Consequently, we conclude that emotional arousal cannot be estimated by EDA in isolation, and necessitates contextual or multimodal input. If one can determine the contextual state through other sources of information, EDA may serve as a dependable estimator for emotional arousal as per the valence-arousal framework. In the context of IM-TWIN, this implies that EDA analysis should be complemented with other biosignals such as ECG, along with other sources of information such as audio and/or video.

Currently, NS-SCRs are extracted using thresholds in the time-domain. Unfortunately, this methodology is very sensitive to noise. Noise, even when suppressed by low-pass filtering, makes NS-SCR detection in the time-domain challenging (see Figure 11). As such, more robust techniques are needed to extract SCRs reliably in noisy, real-world EDA.

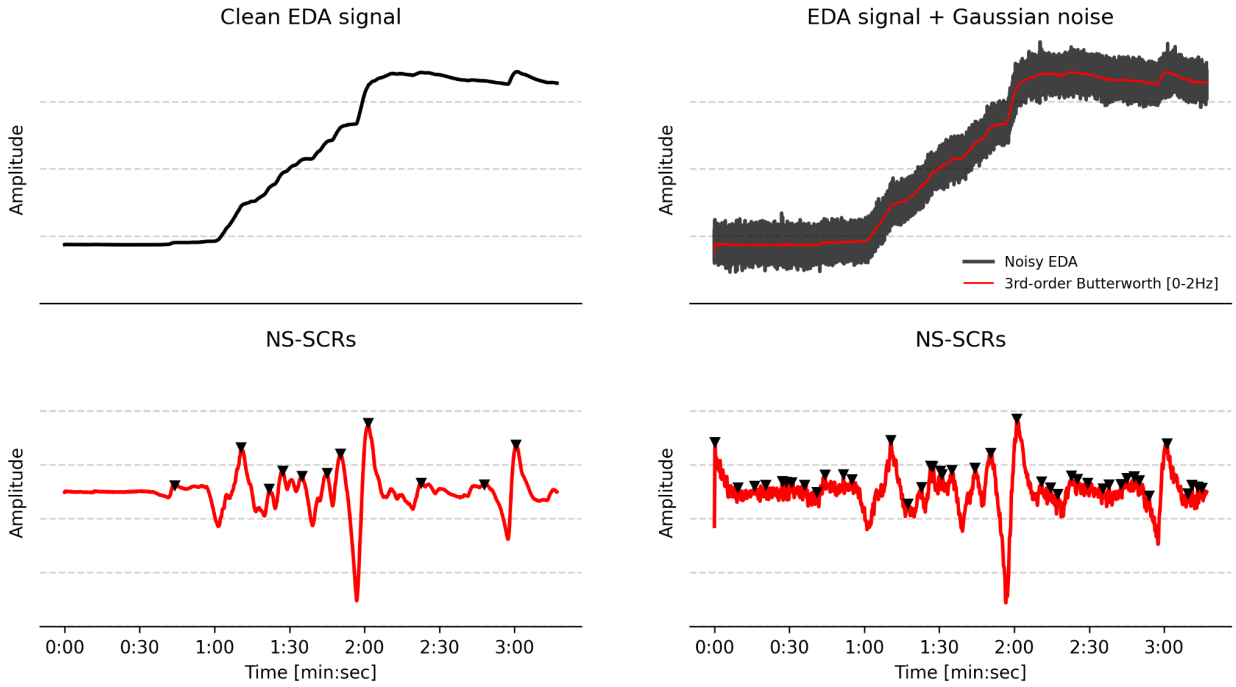

***Figure 11**: Extraction of peaks from EDA is highly noise-sensitive. Even noise suppressed by low-pass filtering will affect NS-SCR detection to the extent it becomes unusable.*

Analyzing and extracting SCRs in the frequency or time-frequency domains could offer new ways to handle real-world noise. Time-frequency techniques such as the Short-term Fourier Transform and the Continuous Wavelet Transform have already been applied to EDA successfully [18,19,20]. Future developments can test noise-resilience of such techniques and their value for real-world signal analysis.

4. Future Developments

In conclusion, processing noisy ECG and EDA signals in IM-TWIN's resource and time-limited environment presents significant challenges. In ECG analysis, deep learning algorithms offer superior detection accuracy but are computationally expensive. In contrast, traditional heuristic-based algorithms are efficient but become unstable in the presence of noise.

Conversely, in EDA analysis, the common technique of extracting NS-SCRs in the time-domain using thresholds is very sensitive to noise for EDA signals, making it challenging to apply in real-world scenarios. To address these challenges, future developments should focus on finding new approaches that combine the strengths of existing algorithms and achieve fast and accurate QRS detection for ECG signals, and more robust techniques to extract SCRs reliably in noisy, real-world environments. As signal quality is key in the scope of IM-TWIN's difficult challenges, such improvements are vital for IM-TWIN to succeed.

References

1. Bakker, I., Van Der Voordt, T., Vink, P., & De Boon, J. (2014). Pleasure, arousal, dominance: Mehrabian and Russell revisited. *Current Psychology*, *33*, 405-421.
2. Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). *Handbook of psychophysiology*. Cambridge university press.
3. Harmon-Jones, E., Harmon-Jones, C., & Summerell, E. (2017). On the importance of both dimensional and discrete models of emotion. *Behavioral sciences*, *7*(4), 66.
4. Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological bulletin*, *139*(4), 917.
5. van den Broek, E. (2011). *Affective Signal Processing (ASP): Unraveling the mystery of emotions*. PhD-thesis, Human Media Interaction (HMI), Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands. ISBN: 978-90-365-3243-3. DOI: https://doi.org/10.3990/1.9789036532433 .
6. Waller, A. D. (1887). A demonstration on man of electromotive changes accompanying the heart's beat. *The Journal of physiology*, *8*(5), 229.
7. Watanabe, T., Tomitani, N., Yasui, N., & Kario, K. (2022). Validation of an ambulatory blood pressure monitoring device employing a novel method to detect atrial fibrillation. *Hypertension Research*, *45*(8), 1345-1352.
8. Olson, J. A., Fouts, A. M., Padanilam, B. J., & Prystowsky, E. N. (2007). Utility of mobile cardiac outpatient telemetry for the diagnosis of palpitations, presyncope, syncope, and the assessment of therapy efficacy. *Journal of cardiovascular electrophysiology*, *18*(5), 473-477.

9. Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE transactions on biomedical engineering*, (3), 230-236.
10. Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE engineering in medicine and biology magazine*, *20*(3), 45-50.
11. Hossain, M. B., Bashar, S. K., Walkey, A. J., McManus, D. D., & Chon, K. H. (2019). An accurate QRS complex and P wave detection in ECG signals using complete ensemble empirical mode decomposition with adaptive noise approach. *IEEE Access*, *7*, 128869-128880.
12. Zahid, M. U., Kiranyaz, S., Ince, T., Devecioglu, O. C., Chowdhury, M. E., Khandakar, A., ... & Gabbouj, M. (2021). Robust R-peak detection in low-quality holter ECGs using 1D convolutional neural network. *IEEE Transactions on Biomedical Engineering*, *69*(1), 119-128.
13. Zhang, F., & Lian, Y. (2007, August). Electrocardiogram QRS detection using multiscale filtering based on mathematical morphology. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 3196-3199). IEEE.
14. Da Silva, H. P., Lourenço, A., Fred, A., Raposo, N., & Aires-de-Sousa, M. (2014). Check Your Biosignals Here: A new dataset for off-the-person ECG biometrics. *Computer methods and programs in biomedicine*, *113*(2), 503-514.
15. Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.
16. Sharma, K., Castellini, C., van den Broek, E. L., Albu-Schaeffer, A., & Schwenker, F. (2019). A dataset of continuous affect annotations and physiological signals for emotion analysis. *Scientific data*, *6*(1), 196.
17. Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., ... & Chen, S. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior research methods*, 1-8.
18. Ganapathy, N., Veeranki, Y. R., & Swaminathan, R. (2020). Convolutional neural network based emotion classification using electrodermal activity signals and time-frequency features. *Expert Systems with Applications*, *159*, 113571.
19. Rao Veeranki, Y., Ganapathy, N., & Swaminathan, R. (2021). Electrodermal activity based emotion recognition using time-frequency methods and machine learning algorithms. *Current Directions in Biomedical Engineering*, *7*(2), 863-866.
20. Feng, H., Golshan, H. M., & Mahoor, M. H. (2018). A wavelet-based approach to emotion classification using EDA signals. *Expert Systems with Applications*, *112*, 77-86.