EMPhASIS: An EMbedded Public Attention Stress Identification System

Jessica Leoni, Asia Ciallella, Luca Stornaiuolo, Marco D. Santambrogio, Donatella Sciuto
Politecnico di Milano, Dipartimento di Elettronica Informazione e Bioingegneria (DEIB), Milan, Italy
{jessica.leoni, asia.ciallella}@mail.polimi.it
{luca.stornaiuolo, marco.santambrogio, donatella.sciuto}@polimi.it

***Abstract*—Stress is a general adaptation syndrome, caused by a person's perception of danger about an event. While the perception of an event as stressful varies for each individual, several studies show that public speaking is a source of anxiety and panic for most people. Moreover, the inability to prevent or manage these conditions affects their overall performance negatively. The literature depicts different systems developed in order to recognize the subject's stress level by monitoring physiological signals. However, they are specifically designed for laboratory environments and clinical analysis and are often characterized by several limitations. Therefore, aiming at developing a daily-life application able to help people prevent and handle anxiety conditions by improving their self-perception, we propose an embedded system for real-time biometric feature analysis. This system is able to handle all the steps from the acquisition to the stress and anxiety status classification. The whole framework is implemented on a Xilinx PYNQ-Z1 board, leveraging both the ARM processor and the FPGA. Taking advantage of the programmable logic brings to a system able to process data with higher performance and energy efficiency, helping to face the embedded application constraints.**

***Index Terms*—embedded system, FPGA, stress detection, public speaking, PYNQ-Z1, biometric**

I. INTRODUCTION

As first defined by Hans Selye in 1936, stress is a general adaptation syndrome generated as a response to the state of alarm induced by the perception of a threat [1]. The interest that this topic has encountered in recent decades has revealed how stress impacts negatively on people's health status and well being, causing sleep disorders, difficulty concentrating, but also weakening of the immune system, digestive problems, and heart disease. These symptoms have a negative impact on the individual's performance in everyday life, especially considering work productivity and quality, as assessed by the results shown by Lazarus et al. in [2]. Also, the American Institute of Stress, in [3], reveals how more than 70% of Americans regularly experience physical and psychological symptoms caused by stress, and how more than 48% recognize that this harms both their personal and professional lives. And this trend is increasing.

As reported in [4], the professions that are more likely to generate a great amount of stress are the ones that pose a real threat to the safety of the subjects and the ones that require direct contact with the public. In this last scenario, the triggering factor turned out to be the presence of the audience, which leads the presenter to experience a well-known phenomenon, called *public speaking anxiety* [5].

An improved self-perception helps people preventing and managing stress and anxiety attacks, allowing them to optimize their performance when dealing with a public speaking-related task. Aiming at providing an effective solution, we propose an embedded and personalized system able to report real-time the level of anxiety and stress perceived by an individual. The proposed hardware architecture is based on PYNQ-Z1 and offloads part of the computation to the FPGA to best meet the performance requirements. .

The main contributions can be summarized as follows:

- An algorithm to determine the stress level, that can be customized for the physiology of each user;
- The implementation of a high-performance FPGA-based embedded system able to process biomedical signals in real-time;
- The capability of the device to be energy efficient and allow constant autonomy using lightweight batteries.

II. BACKGROUND

EMPhASIS stress detection logic is based on the analysis of the electrocardiograph (ECG) signal. As demonstrated by Goel et al. in [6], the combination of ECG time and frequency domain analysis is sufficient to accurately perform this classification task. ECG signal represents the trend of the cardiac potential over time. It is defined as pseudo-periodic, as it is characterized by a template, corresponding to a single heartbeat, which repeats over time with a variable frequency. To calculate the duration of a heartbeat, the distance between two consecutive R peaks is usually considered. R peak is chosen as it is the maximum deflection in the heartbeat template. Therefore, from the ECG is extracted a signal containing only the time instants of the occurrences of R peaks, defined *RR signal*. The time-domain parameters in the RR signal that are relevant to perform the classification task are:

- ***mRR***, that is the average duration of a heartbeat. Reductions of this value, compared to the baseline, indicate an increase in the level of anxiety and stress perceived by the subject;
- ***mHR***, i.e. the average heart rate.

Since the time-domain analysis provides only a partial insight into the whole information, the Fourier transform is leveraged to extract the signal's frequency content, i.e. its harmonics. As reported by Mohr et al. in [7], frequency domain directly reveals grasp from the *rr signal* information on the subject's stress and anxiety conditions. As demonstrated by Malliani et al. in [8], it is possible to find a correspondence between the normalized spectrum of the *RR signal* and the activity of the two branches of the autonomous nervous system: the sympathetic and the parasympathetic one. The first predominates in situations of stress and emergency, while the second concerns moments of relaxation. The area of the normalized spectrum underneath the low frequency (LF), identified between 0.04Hz and 0.15Hz, is directly proportional to the sympathetic activity, while the area underneath the high frequency (HF), identified between 0.15Hz and 0.4Hz, is an indication of the parasympathetic activity. Frequency domain features considered by the system presented are:

- ***nLF***, i.e. the area underlying the LF, normalized for the overall area underlying both the LF and the HF. The more this ratio tends to 1, the greater will be the state of stress and anxiety perceived by the individual;
- ***nHF***, i.e. the normalized area underlying HF. The more this contribution tends to 0, the higher the level of stress and anxiety;
- ***dLFHF***, i.e. the difference between the area underlying LF and HF. States of stress and anxiety correspond to values of this difference greater than 0;
- ***SVI***, i.e. the ratio between the area underlying LF and HF. The more the value of this ratio is greater than 1, the more the individual is perceiving a state of stress and anxiety.

III. Methodology

The hardware/software architecture of our system implements four phases: *calibration*, *acquisition*, *features extraction*, and *classification*.

Calibration: during this phase, the system records and analyzes the user's ECG signal to customize the classifier thresholds. The analysis is performed offline and requires ECG recording related to both physiological and stressed conditions. The analysis is conducted extracting time and frequency-domain features for each observation and then performing k-means in the PCA space. This procedure is completely unsupervised to prevent the user from providing the ground truth labeling the data. As an example, figure 1 shows the clusters identified using 5 minutes of recording from the MIT-BIH Noise Stress Test Dataset [9].

Acquisition: EMPhASIS supports two different acquisition modes: offline and online. The first one manages as an input a pre-recorded signal, while the second one acquires the ECG signal through a sensor connected to the programmable logic of the PYNQ-Z1 board. In this work, we target the DFRobot Heart Rate Monitor Sensor that allows acquiring the ECG through the electrodes, arranged to form a triangle of Einthoven, for the online acquisition. Both offline and online modes require to set the size for the sample window that the FPGA has to compute at each iteration. Increasing this size

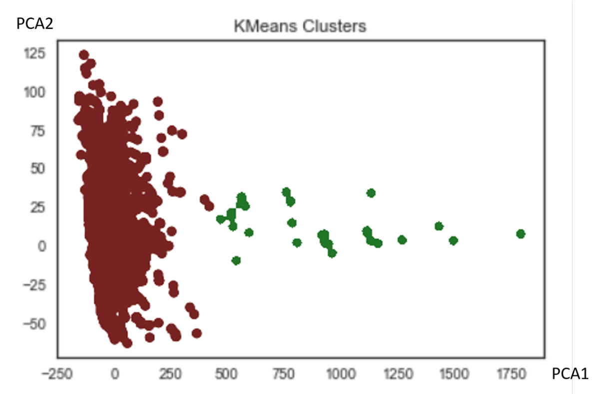

Fig. 1. K-Means Clusters Division. This figure shows the division in the two clusters of physiological state and stress state operated by k-means on the 5-minute recording logging to the system to allow calibration on the individual user.

helps the FFT function to extract more accurate frequency content, but affects the iteration execution time.

Feature Extraction: this phase takes as input the *RR signal* and its power spectrum, and provides the set of necessary features to perform the classification task. It is organized into two blocks working in parallel. The first block extracts time domain features from the *RR signal*, while the second block extracts frequency-domain features from its spectrum. More in detail, the time domain features extraction function works by applying a sliding window to the signal and then computing ***mRR*** and ***mHR***. The first feature consists of the series of the average duration of the beats contained in each window: $mRR_i = \frac{(\sum_{k=1}^{M} rr_k)}{M}$, where M is the number of RR intervals contained in the considered window. The second features consist of the series of the average frequency in the analyzed window. The average heart rate is defined as: $mHR_i = \frac{60000}{mRR_i}$, where the factor 60000 is introduced to bring the heart rate in beats per minute unit conventionally used, instead of in beats per millisecond, and it goes from 0 to the number of windows.

Even to perform frequency-domain feature extraction a sliding window is applied to the signal. For each window, the following features are extracted: ***nLF***, ***nHF***, ***dLFHF***, and ***SVI***. Each value in the nLF series corresponds to the normalized area subtended by the spectrum in the low frequencies in the window i: $nLF_i = \frac{LF_i}{LF_i+HF_i}$. For what concern nHF, each value corresponds to the normalized area under the spectrum in the high frequencies for each considered window i: $nHF_i = \frac{HF}{LF_i+HF_i}$. In $dLFHF$, each value corresponds to the difference between the value of the area underneath the low and the high frequencies for each considered window i: $dLFHF_i = |nLF_i - nHF_i|$. Finally, each value of SVI corresponds to the ratio between the value of the area underneath the low and the high frequencies for each considered window i: $SVI_i = \frac{LF}{HF_i}$. To reduce the latency, both time and frequency-domain blocks are designed to allow concurrent execution of the parallel operations, as shown in figure 2. This was possible since each window is completely data independent from the others.

Classification: The classification method implemented in hardware is composed of a PCA-based event detection algorithm. In particular, given the set of selected features, the value of the first PCA component is calculated and it constitutes the input of a threshold-based classification algorithm, properly tuned according to the parameters provided as the result of the calibration phase. Values that exceed the threshold concern physiological state, otherwise they are related to stress and anxiety conditions.

IV. Experimantal results

The whole system has been specially designed to best adapt to the public speaking scenario. For this purpose, we targeted the FPGA-based PYNQ-Z1 board to allow real-time acquisition and high-performance computation of the signal, while being extremely energy efficient. More in detail, EMPhASIS architecture leverages two different *overlays* to work. Overlays are hardware libraries containing both the bitstreams to configure the FPGA and the drivers to exploit the hardware functions from the software level. The first one, provided by Xilinx, manages the sensor during the acquisition phase. The second one, designed by us, takes as inputs the *RR signal* and its Fourier Transform, and performs features extraction and classification returning to the Zynq Processing System the final predictions. Both the overlays have been implemented with a clock frequency of 100MHz. We exploited Vivado HLS to design our custom computational kernels for the FPGA. The communication between the Programmable Logic and the Processing System leverages AXI stream interface and the library is offered at the software level as a Python library, in accordance with the PYNQ platform. In particular, at power-on, the first bitstream is downloaded and configures the FPGA. Then, the acquisition phase starts. As soon as the required amount of RR intervals has been recorded, the FPGA is dynamically reconfigured with the second bitstream to perform features extraction and classification tasks on the recorded data. Once the classification ends, the acquisition phase starts again and new data are recorded, while the classification outputs are notified to the user through RGB LEDs.

One of the fundamental requirements for EMPhASIS consists in being able to cover the entire duration of a presentation, without requiring excessively heavy batteries. From the results shown in table I, it is possible to observe that the total power consumption is about 1.757W. Most of this consumption, about 92%, depends on dynamic resources, in particular on the PS7, while only 8% depends on static ones. As a result, the power dissipated by the Joule effect is also reduced, which translates into a temperature of the junctions of only 45.3°C. Considering a normal power bank, which works at a voltage of 5V, it is possible to estimate that our system needs 351.4mA. Assuming reasonably that this power bank provides a current intensity of 2Ah, it will allow EMPhASIS to operate in full autonomy for more than 5 hours. The lower power consumption required by the system makes the choice of an embedded board far more efficient than using a powerful PC or a cloud infrastructure. Another fundamental requirement for our system was that it be able to provide real-time results. The current execution time for the classification of 40000 ECG-related raw data is 194.91s (each result is displayed to the end user each 10s by analyzing a subset of his measured heartbeat). Although the results obtained in this proof of concept are satisfactory, further improvements can be made in the implementation of the final system. A slight delay in the system is given by the time required by the FPGA to dynamically reconfigure itself. For the whole period of operation of the system, the two used overlays are mutually configured to the FPGA. However, the second overlay requires about half of the available resources. The more used resources are DSP, with a utilization percentage of 50.91%, LUT, with a utilization percentage of 22.66%, and flip flops, with a utilization percentage of 15.37%. Therefore, it would be interesting to be able to manage the reading from the ECG sensor and the leds in the same overlay, exploiting the available space. This way, the FPGA can manage all the phases in the system in just one overlay, without dynamically reconfiguring itself during the execution. Avoiding the reconfiguration runtime would give a further speed up to the system.

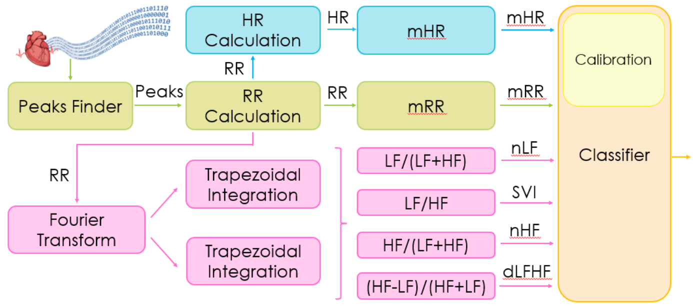

Fig. 2. High Level Block Design of EMPhASIS Overlay. This figure shows, at a high level, the block division of the EMPhASIS system.

TABLE I
Power Utilization On-Chip. This table reports the power utilization on-chip of the FPGA performing the computation required by the EMPhASIS system.

Total On-Chip Power	1.757W
Junction Temperature	45.3°C
Thermal Margin	39.7°C (3.3W)
Effective $\frac{\theta}{A}$	11.5° $\frac{C}{W}$
Power Supplied to Off-Chip Device	0W
Confidence Level	Medium

TABLE II
FPGA Resource Utilization. This figure shows the resource utilization for the implementation of the proposed overlay.

	LUT	LUTRAM	FF	BRAM	DSP	BUFG
Available	53200	17400	106400	140	220	32
Utilization(%)	22.66	5.60	15.37	7.50	50.91	3.13

V. Related Work and Evaluation

An accurate analysis of the literature has revealed several works aimed at automatically identifying the state of stress and anxiety in the subjects. As reported by Deng et al. [10] and by Keshan et al. [11], the signals used to classify this status are the most varied, such as ECG, EEG, GSR, respiration. However, this combination of sensors cannot be applied in the scenario of a public presentation, where the presenter must be free to move and the audience must not notice the presence of the sensors. At best of authors' knowledge the only work in literature that aims at estimating stress leveraging an FPGA architecture is the one in [12]. Again, the proposed setup is not designed to be applicable in a public speaking scenario.

Also, we investigate the devices currently on the market. Most of them deal with improving sleep quality. As an example, Spire's Health Tag sensor, presented in [13], that is designed to be worn on the underwear, in direct contact with the skin. Monitoring breathing frequency, this device is able to identify conditions of anxiety or stress, managed through a smartphone app that provides guided breathing exercises. To suit a daily-life scenario are proposed devices as The Pip [14], properly designed to recognize the stress level by monitoring the electrodermal activity (EDA) and to report the estimated results through the associated smartphone application. Since both of these types of devices require the users to constantly monitor the associated smartphone application to know their level of estimated anxiety, they are not feasible in a public speaking context. Therefore, we aim at providing EMPhASIS output leveraging a vibrating wristband, avoiding to involve any smartphone application. Another issue of analysis on the market devices is the lack of a calibration phase. In fact, since each subject's physiology differs from the others, a stress detection system should require a preliminary tuning phase to customize its parameters according to the specific individual.

Regarding the system autonomy, the minimum guaranteed by EMPhASIS system is 5 hours, so more than enough to cover the duration of a presentation. This result is comparable with those of The PiP and SpireHealth. The first has an autonomy of 8 hours, while the second, being a simple sensor, can last a whole day. However, since these two sensors require the use of specific smartphone applications, also the battery life of the phone has to be considered as a limiting factor.

Making an approximate estimation of the cost of the system, the Pip sensor cost ranges between 150$ and 200$ per person, while SpireHealth sensor is around 50$ per person. Besides, each individual must have a smartphone. The DFRobot sensor used for EMPhASIS costs 30$ per person, while a PYNQ Z1 costs about 200$. However, this board is able to compute data for a pool of people in a room. In fact, the final design can be repeated several times on the board, using all available resources. Reconfiguring each instance according to the parameters of the user to whom it is dedicated, the system would scale with a minimal added cost. This makes EMPhASIS much more affordable than other solutions on the market.

VI. Conclusion

In this paper, we presented EMPhASIS, an EMbedded Public Attention Stress Identification System. The system, entirely developed on a PYNQ-Z1 board, manages all the phases necessary to identify anxiety and stress status on an individual, overcoming the limitations of the existing solutions being as minimally invasive as possible, customizable and fast.

In fact, the ECG sensor is minimally invasive and easily concealed under clothing. Besides, given the PYNQ-Z1's low power consumption, the batteries required to ensure the device operates for the entire duration of a presentation are extremely light. Speed performances allow the system to provide real-time feedback of the user's status. Finally, EMPhASIS is made customizable thanks to a calibration system, which returns, for each new user, the parameters with which to tune the classifier.

Regarding future directions, the first improvement is to replace the LEDs with vibration sensors embedded in a wristband, that can be controlled bt the board without the need for a physical connection. In this way, the board can be placed in a turret in the middle of the room, from which it receives the signal acquired by the ECG sensors and communicates the prediction by modulating sensors' vibration. This improvement would make the choice of an FPGA extremely advantageous, leveraging its parallelization and scaling capabilities, without additional overhead. Besides, EMPhASIS is suitable for all those applications where immediate decision support is required regarding the physiological state of a subject, thus its performances in scenarios different from the one proposed will also be evaluated.

References

[1] H. Selye *et al.*, "A syndrome produced by several nocuous agents," *Nature*, vol. 138, no. 3479, p. 32, 1936.

[2] R. S. Lazarus, J. Deese, and S. F. Osler, "The effects of psychological stress upon performance." *Psychological bulletin*, vol. 49, no. 4, p. 293, 1952.

[3] T. A. I. of Stress, "Daily life - the american institute of stress," 2011. [Online]. Available: https://www.stress.org/daily-life

[4] K. A. Renzulli, "The most stressful job in america pays $26,802—here are the other 9," 2019. [Online]. Available: https://www.cnbc.com/2019/03/07/the-most-stressful-jobs-in-america.html

[5] J. A. Daly, A. L. Vangelisti, and S. G. Lawrence, "Self-focused attention and public speaking anxiety," *Personality and Individual Differences*, vol. 10, no. 8, pp. 903–913, 1989.

[6] S. Goel, G. Kau, and P. Toma, "A novel technique for stress recognition using ecg signal pattern." *Current Pediatric Research*, vol. 21, no. 4, 2017.

[7] E. Mohr, E. Witte, and B. Voss, "Heart rate variability as stress indicator," *Archiv fur Tierzucht*, vol. 43, pp. 171–176, 2000.

[8] A. Malliani, F. Lombardi, and M. Pagani, "Power spectrum analysis of heart rate variability: a tool to explore neural regulatory mechanisms." *British heart journal*, vol. 71, no. 1, p. 1, 1994.

[9] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "Ecg pattern analysis for emotion detection," *IEEE Transactions on affective computing*, vol. 3, no. 1, pp. 102–115, 2011.

[10] Y. Deng, Z. Wu, C.-H. Chu, Q. Zhang, and D. F. Hsu, "Sensor feature selection and combination for stress identification using combinatorial fusion," *International Journal of Advanced Robotic Systems*, vol. 10, no. 8, p. 306, 2013.

[11] N. Keshan, P. Parimi, and I. Bichindaritz, "Machine learning for stress detection from ecg signals in automobile drivers," in *2015 IEEE International Conference on Big Data (Big Data)*. IEEE, 2015, pp. 2661–2669.

[12] N. Attaran, A. Puranik, J. Brooks, and T. Mohsenin, "Embedded low-power processor for personalized stress detection," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 12, pp. 2032–2036, 2018.

[13] S. Health, "Spire health tag," 2019. [Online]. Available: https://spirehealth.com/products/health-tag-membership

[14] T. Pip, "Stress management device," 2015. [Online]. Available: https://thepip.com/