Empowering Humans in a Cyber-Physical Production System: Human-in-the-loop Perspective

David Costa*, Flávia Pires*, Nelson Rodrigues*, José Barbosa*, Getúlio Igrejas*, Paulo Leitão*
* Research Center in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança,
Campus de Santa Apolónia, 5300-253 Bragança, Portugal
Email: {dcosta, fpires, nrodrigues, jbarbosa, igrejas, pleitao}@ipb.pt

Abstract—In the Industry 4.0 context, humans are taking up new roles in Cyber-Physical Systems (CPS) since it was realized that they constitute the most flexible part of an automated system. With this new reality, humans are becoming integral parts of CPS, entering in the realm of the Human-in-the-Loop of Cyber-Physical Systems (HiLCPS). In this paper, the challenges and technologies for the integration of humans in CPS environments are briefly discussed, and a practical application of the HiLCPS concept is presented. For this purpose, a human operation workbench is empowered with intelligent decision and assistant systems that use, among others, reasoning algorithms, image processing and speech recognition, to improve the productivity and quality of the tasks performed by humans, particularly focusing the assembly of highly customized products, as well as the training of operators to perform complex assembly, repairing or maintenance tasks.

Keyword: Cyber-physical systems, Human integration, Human-in-the-Loop, Image processing, Neural networks.

I. INTRODUCTION

The fourth industrial revolution, mostly known as Industry 4.0, is leading to a more intelligent manufacturing system, through intelligent products and production processes, being mainly based on Cyber-Physical Systems (CPS). CPS are the backbone of the Industry 4.0, being expected to shape the evolution towards this new future vision of production and affecting the interaction of humans with these key enabling technologies. CPS focuses on the coexistence of cyber and physical counterparts, which are organized in a distributed network and interact with each other to achieve a common goal. In such systems, the "cyber" part implies computation, communication and control [1]. This concept is the aggregation of several technological topics such as, robotics, wireless sensor networks, mobile computing and Internet-of-Things (IoT) [2].

Since the first industrial revolution, humans have always taken part of the manufacturing processes, but with the undergoing implementation of Industry 4.0, factories are changing the humans roles in production lines [3], [4]. These changes require that the skills and competencies of the humans must be enhanced. Some of these new capabilities are, e.g., safe human-robot collaboration and customer-operator collaboration [5]. In this new reality, humans are becoming an integral part of CPS, entering in the realm of the Human-in-the-Loop of Cyber-Physical Systems (HiLCPS) [2]. A typical HiLCPS consists of a loop involving a human, an embedded system and the physical environment. The embedded system, that represents the cyber component, augments the human interaction with the physical world [6].

Two different approaches of human integration arise in HiLCPS: Human-in-the-Loop (HitL) and Human-in-the-Mesh (HitM). In the case of HitL, it works at an operational level, involving different types of human activities that may influence the performance and the flexibility of the production processes. On other hand, the HitM is a type of intervention at strategic level that includes interactions with CPS network and applications, and also includes the interactions with other human roles [7]. The integration of humans in CPS is being studied from different points of view, such as, human-machine interfaces and human-in-the-loop control and coordination.

The integration of humans in CPS can enable them to achieve complex tasks with minimum specialization and skills [7]. This also allows them to be more efficient and productive, allowing to increase the employment opportunities [8]. Note that during their activities, usually if operators perform tasks quickly, they are more likely to make mistakes or if they don't make mistakes they take longer to perform their tasks. Additionally, in case of highly customized and complex tasks, operators have more difficulty to execute rapidly and efficiently these operations and require some external assistance.

This paper describes a human operation workbench, complemented with an intelligent assistance system, highlighting the integration of humans in a CPS context. The workbench, developed under the I4.0@TMAD project [9], considers emergent technologies associated to Industry 4.0 to develop an intelligent personal assistance system that improves the efficiency, quality and productivity of highly customized and complex operations performed by humans. For this purpose, artificial vision, speech recognition and augmented reality technologies, amongst others, are used to enhance and support operators during the execution of assembly, maintenance and training operations.

The rest of the paper is organized as follows: Section II overviews the existing technologies for integrating humans in CPS environments. Section III presents the developed experimental setup, in terms of hardware and software systems, as well as the working scenarios. Section IV presents the system architecture and Section V describes the developed solutions focusing on the image processing and the speech recognition solutions to support the intelligent assistance of operators.

Finally, Section VI rounds up the paper with the conclusions.

II. Technologies for Integrating Humans in Cyber-Physical Systems

Human resources are important pieces in CPS systems, since they introduce flexibility in the performed operations, requiring that human-centered technologies play a special role in the design of these systems [10].

The integration of the human in CPS is not an easy task. However, the technological advances, aligned with Industry 4.0, provide several solutions that allow an easy transition and integration. Before this new industrial revolution, the traditional industrial user interfaces only made one type of interaction. The operator would send through a mechanical input, a command to the system, and the interface would reply displaying an image or, when an error occurred or a warning was needed, would play an auditory signal. Now the interaction between human operator and CPS is made by natural forms combining touch-screens, voice recognition and gesture recognition, making it more intuitive [4].

There are several technologies that support the HiLCPS, facilitating the humans integration, such as, augmented reality (AR), virtual reality (VR) and intelligent personal assistant (IPA) [11]. All these technologies allow the augmentation of the original human capabilities. The AR is the direct or indirect view of a real physical environment, in real time, where those elements are augmented through computer-generated sensing, e.g. sound, video or graphics [11]. This technology improves the transfer of information from the digital to the real world. The VR is an immersive interactive computer simulation that digitally replicates the assembly or manufacturing environment, allowing for the operator to interact with real-time feedback [11].

IPA is a software system that supports the operator to interact with machines or computers, and to perform tasks or services [11]. According to [12], an IPA is an application that uses voice commands, images and contextual information to elaborate recommendations and actions to be taken. One of many applications of IPAs is in mobile health applications, where e.g., the AMBRO platform [13] supports the daily life aspects, with the users being able to access web services, consult information about body sensors under use and recommendations based on the learned user behaviour and routines (through the interactions with other objects in the surroundings and the IPA itself). Another example is the CORTEX aiASSIST, which offers assistance, e.g., in monitoring and diagnosis, process supervision and maintenance [14].

The presentation of data can use several types of devices, usually allowing bidirectional communication, namely:

- HMD (Head-mounted devices), e.g., Epson Moverio, Google Glasses and Microsoft Hololens, that allow the visualization of data while the hands are free to perform tasks, usually associated to VR/AR environments.
- Portable devices, e.g., smartphones and tablets, that require at least one of the hands to keep the device's display.
- Space screens, similar to portable devices but usually bigger and stationary, so it is not necessary to use the hands to hold them.
- Projectors: used for projection of augmented data about real objects that are perceived by the users.

Other interaction devices, that allow unidirectional communication human-machine, are gestures recognition devices (hands and fingers), e.g., the Leap Motion.

AR, VR and IPA applications can support operators during installation, operation and maintenance of automation systems. This can improve productivity and efficiency of the operator providing related information to the history and current state of the device/system and providing additional information, e.g. documents, websites or videos. As an example, the use of AR technologies allows for a reduction of 25 to 30% in the production time, which in the aeronautical industry can mean a production gain in the order of 4 M€ per product (airplane). Figure 1 illustrates an application of AR, running in 2D mode on a Android smartphone, to support the execution of maintenance tasks on an industrial manipulator robot.
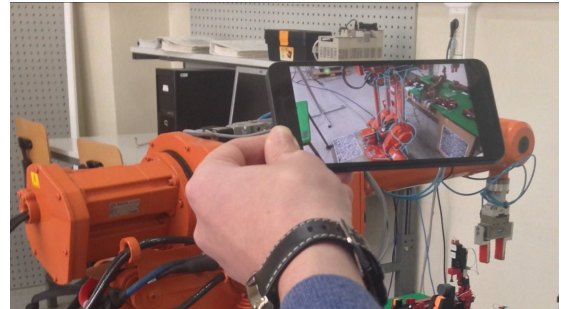

Figure 1: AR application running on a smartphone.

Although these technologies are very helpful for the operators, they also present disadvantages, e.g., HMDs are not comfortable in a long period of use and have a very short field of view, and gestures' devices require the use of the hands, which is an obstacle in cases of tasks that the operators need to utilize them. This means that the design of human-machine interfaces needs to consider several different aspects, e.g., functionality, ergonomics and work context.

III. Experimental Demonstration Setup

The human operation workbench developed in this work considers a kind of IPA system, where the tasks that are performed by the operator are supported by intelligent and interactive decision support applications developed for this purpose. The workbench shows the HiLCPS concept in an automated environment where operators or maintenance technicians are supported by an intelligent assistance system and collaborative technologies to support the execution of highly customized and complex assembly and maintenance tasks.

A. Structure of the Human Operation Workbench

The human operation workbench, illustrated in Figure 2, consists in a physical structure that provides a workspace for

the operator to perform manual operations, namely assembly and reparation. For this purpose, the workbench comprehends a set of hardware devices and software applications that constitute the intelligent personal assistance system to support the human during the execution of their tasks.
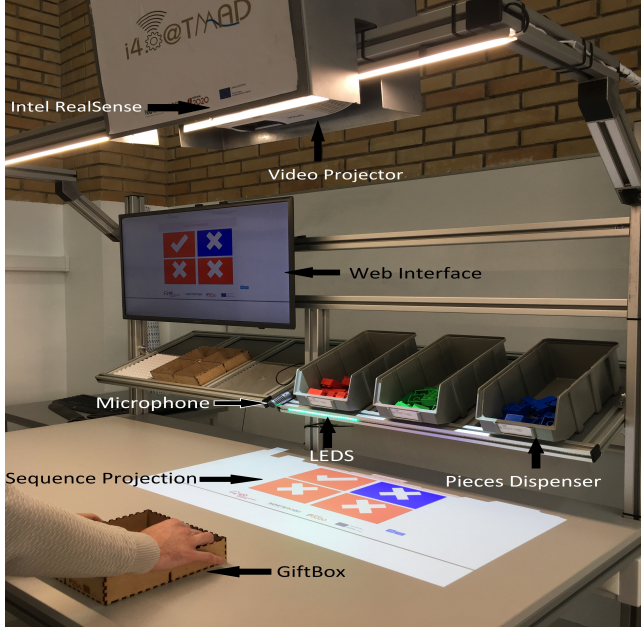

Figure 2: Human operation workbench in the HiLCPS concept.

In particular, the workbench comprehends the following hardware systems:

- A monitor to display the information related to the configuration or assembly steps, the results of correctness of the performed operations, and visual support to perform the required operations, e.g., images or tutorial videos.
- A video projector to project the same information as the monitor but directly in the working area.
- An Intel RealSense camera to acquire information of the workspace that will support the verification of the correctness of performed assembly operations.
- A microphone to acquire the voice commands from the operator and to support the flow management in the assistance system.
- A LED ribbon system placed under the pieces dispenser to help the user to visually know in which dispenser is the piece required to perform such operation.

These hardware systems are interconnected into the intelligent personal assistance software application.

B. Working Scenarios

Two working scenarios were created to illustrate the collaboration between the operator and the intelligent assistant system that guides the operator during the execution of highly customized and complex operations.

The first working scenario is related to the assembly of highly customized gift boxes, each one having four divisions where different shaped and colored objects can be placed,

constituting different configurations. Due to the characteristics of this highly customized assembly task, the operator can not performed their tasks in the same manner as performing repetitive tasks, and requires external support of the assistance system to instruct him to perform correctly the task and also to verify the correctness of the assembled configuration.

Initially, the target configuration for the assembled gift box is showed in the monitor and video projector. The operator picks the pieces from the proper dispensers according to the shape and color displayed in the monitor, and place them in the gift box. Aiming to support the operator, the intelligent assistance system turns on the lights of the LED ribbon to indicate in which dispenser is located the requested pieces. The assembly process is continuously monitored to verify if the pieces are correctly placed. If the configuration is correct, the system shows a check symbol and the operator can proceed to the next assembly of a gift box, e.g., using voice commands by saying "next" or "done". If the gift box is not correctly assembled the system shows a cross symbol in the divisions where the objects are wrongly placed, and the operator needs to correct them before to proceed to another gift box assembly.

The second working scenario is related to the assembly of complex products, during the execution of repairing or training interventions. For this purpose the intelligent assistance system provides the step-by-step instructions to perform the assembly and verifies if the assembly sequence is being performed correctly. In this way, the system displays a picture with the completed assembly product, a picture with the current assembly step, as well as a short explanation on how to proceed. If necessary, the operator can also request further assistance through the use of voice commands by saying "show video" to have access to a tutorial video that explains how to perform the current step of the assembly operation. The system is continuously checking the workspace to verify if the assembled piece is correctly assembled (checking the position, depth, shape and color). If the piece is assembled correctly, the system moves to the next step in the assembly sequence, but if not, it is requested to the operator to repeat the operation and to assemble the piece correctly.

IV. SYSTEM ARCHITECTURE

The structure of the intelligent assistance system for the human operation workbench, illustrated in Figure 3, comprehends the engine application, the Web-based application, the image processing and speech recognition modules and the context aware.

Briefly, an engine application, developed in Matlab®, is displaying, in the monitor and video projector, the information related to the assembly operation according to the predefined context of one working scenario (that defines the sequence of assembly operations to be performed by the operator). This information is related to photos of the final configuration, short explanation of the assembling procedure and tutorial videos to illustrate how the operation should be performed.

Interacting in a simple and intuitive way with humans is essential for a good perception of the system, in this
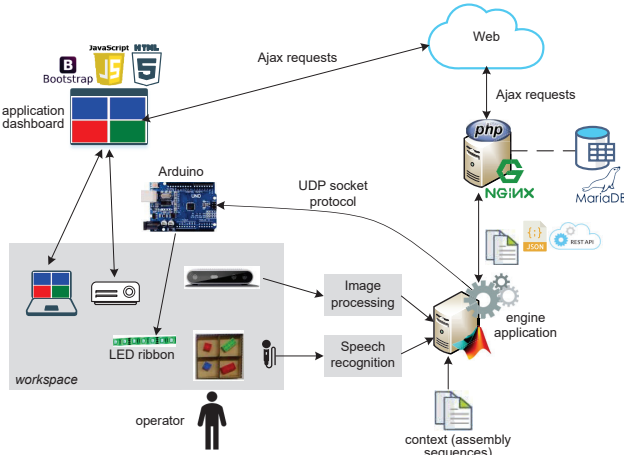
Figure 3: Structure of hardware devices and software applications for the intelligent assistant system.

case to provide a clear idea of how the assembly operations should be performed. In this context, and following the current technological trends, it is important to develop a device-agnostic application that is compatible with the most operating systems and may be accessible on any device, including notebooks, tablets, PCs, and smartphones. For this purpose, the Matlab®application exposes its outcomes through a Web Application, which API was developed in PHP, that works in real-time and was deployed on a Nginx Server using a typical table-based structured database for storing data.

In technical terms, a WebSocket Server allows JavaScript WebSocket Connections, e.g., to communicate with the Web Application, and a RESTful Web Service to communicate with the Matlab®application. On the Matlab®side, a client application is performing REST calls to the Web Application API to send and store the information to be displayed to the operator. On other hand, the front-end of the Web Application has a JavaScript worker that runs in the background, which fetches the information to be displayed from the database without interfering with the operator interface. In this context, the JavaScript worker, through Ajax calls, requests updated data from the database, which is sent as JSON (JavaScript Object Notation) format, to the front-end, while the animations are still running in the user interface. This data can then be displayed in the front-end using the jQuery library without refreshing the user interface. The Bootstrap was used as a responsive front-end framework to allow the adaptation of the user interface to a multitude of device's screens.

In parallel, the Matab®application is continuously acquiring the information about the workspace (image from the camera Intel RealSense and the voice commands from the microphone). The image processing module analyses the acquired image and execute proper algorithms to detect the objects placed in the workspace by shape and color, supporting the verification of the correctness of the task performed by the operator. The speech recognition module is responsible for the

processing of the voice signal to identify the voice commands that will allow to manage the workflow in the assistant application. The results of the verification of the correctness of the performed task are also provided by the means of a monitor and a video projector through the Web Application.

The system also interacts with the operator through a LED ribbon to indicate which dispenser contains the required object to perform the current assembly task. The connection between the Matlab®application and the LED ribbon is made through an Arduino using an UDP socket protocol, which lights green the desired dispenser position.

This system architecture is suitable for deployment and maintenance, once a new updated version is deployed only on the host server and became instantly available in all monitoring devices without activating any update process.

V. IMAGE AND SPEECH RECOGNITION ALGORITHMS

An important issue in the intelligent personal assistance system is the detection of the correctness of the assembly operations performed by the operator, as well as the management of the flow of the supporting application through voice commands. For this purpose, image processing and speech recognition techniques are used to implement these features.

A. Image Processing for the Verification of the Correctness of Product Configuration and Assembly Sequence

The detection of the objects that are being assembled is crucial to support the verification of the correctness of the placement of the object in the workspace during the assembly process by the operator. For this purpose, image processing techniques were used to detect the position, shape and color of the object placed in the workspace. In the first working scenario, the verification only uses the color detection and recognition but in the second scenario, the verification also considers the detection of depth and position.

The color recognition is performed by implementing the segmentation of the RGB components of the acquired image to identify and obtain them separately aiming to search and identify the object's color placed in each position of the image. Figure 4 illustrates the process of segmentation of the primary colors of the image acquired by the Intel RealSense camera, which is posteriorly compared with the target configuration to produce the results of correctness to be shown to the operator (in this case indicating that one of the objects is wrongly placed in the giftbox).

In this process, the acquired image is subdivided in RGB images and then converted to a gray scale in order to subtract them with each one of the images of the RGB components. In the Gray conversion process, each image (snapshot from the camera) is converted from a RGB to a two-dimensional gray scale image. The image is reduced from 24 bits to 8 bits per pixel, where the value of each pixel can vary from 0 to 256 depending on its intensity. The gray scale image is a matrix that holds the luminance (Y) values of the image. The luminance values of the image are obtained by combining the RGB values using the NTSC standard equation (1) that
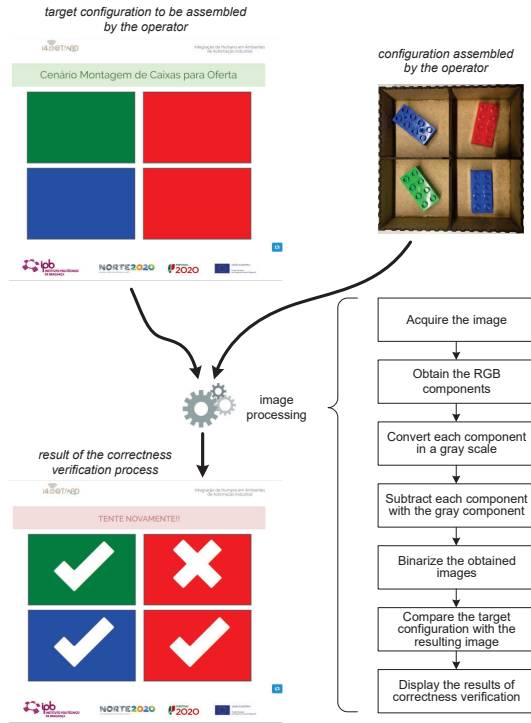
Figure 4: Verification of correctness of performed operations.

multiplies the primary colors (red, green and blue) with the coefficients based on the sensitivity of our eyes to these colors.

$$Y = 0.3 * R + .59 * G + 0.11 * B \qquad (1)$$

The obtained image is binarized with the values of the threshold of each RGB component with the Otsu's method [15], transforming it into an image that contains only 0's and 1's, where 0 represent the white pixels and 1 the black ones. This method is a type of global thresholding that only depends on gray values of the image [16]. This binarization is performed to count the black and white pixels allowing to calculate the area of the detected object.

In the second working scenario, the depth of each piece was obtained using the Intel RealSense's depth camera, allowing to verify if the pieces were correctly assembled.

According to the decision taken by the image processing algorithm, for the first working scenario, a check symbol will be placed in the compartments where the object is correctly placed and a cross symbol in the others; for the second scenario, the permission to move to the next assembly step is allowed only if the assemble piece is correctly assembled.

B. Speech Recognition to Manage the Assistance Flow

The speech recognition is currently a popular topic, making life more effective and easier, but also improving the efficiency and reducing costs in industry applications. In this human operation workbench, speech recognition is used to support the use of voice commands by the operators to manage the flow in the execution of the assistance application.

The developed speech recognition system for isolating words that will support the detection of voice commands uses neural networks. The first step was related to acquire several samples for training the Neural Network followed by the extraction of the coefficients to train the network into a M×1 matrix, where M is the number of extracted coefficients. For this purpose, the Mel frequency Cepstral coefficients (MFCC) technique, which allows to analyze and extract pitch vectors [17], was used because its filter behaves more like the human ear, its low complexity and its recognition accuracy is high and captures the main characteristics of the speech signal [18].

The MFCC procedure is illustrated in Figure 5. Initially, the signal is divided into small frames, in which, each frame overlaps others to avoid losing information. An Hamming Window is applied to the signal to minimize the disruptions at the start and end of each frame. In a third step, the signal is converted from spatial to frequency domain by using the Fast Fourier Transform (FFT), being the calculated spectrum mapped in a Mel scale to know the energy at each spot with the help of a triangular overlapping window, known as triangular filter banks. Later, a discrete cosine transform is applied to convert the Mel spectrum back into the spatial domain.
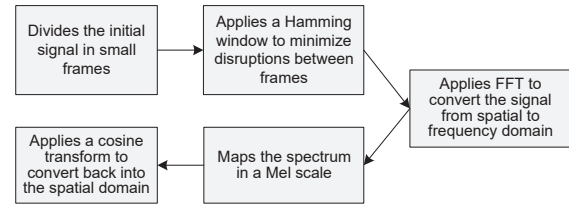

Figure 5: MFCC procedure to extract the coefficients.

After extracting the coefficients from all audio samples, they were concatenated into a matrix, composed of 440 samples (corresponding to 110 samples of each word) and 780 elements (coefficients) that is going to be the training matrix. It was also created another matrix of 440 samples and 4 elements (number of words) that is going to be the target matrix.

The samples of the training matrix were recorded by 55 women and 55 men from different nationalities, namely Portuguese, Polish, Brazilian, Spanish, Nepalese and Russian. The two created matrices were then used in the Matlab®'s Neural Net Pattern Recognition App, where the samples were divided in 70% for Training, 15% for Validation and 15% for Testing. The network is trained with 10 hidden neurons and using the scale conjugate gradient backpropagation (trainscg) function that stops the training when the generalization stops being improved, as indicated by an increase in the cross-entropy error of the validation samples following the structure of the Neural Network shown in Figure 6.

During the training phase, it was plotted a confusion matrix shown in Figure 7, which shows that the percentage of hit rate for the training phase is 100%, meaning no errors.
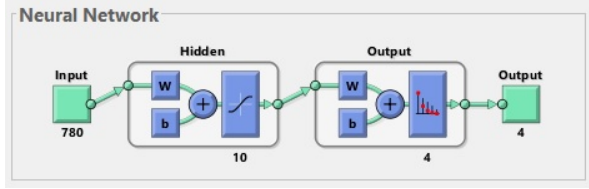
Figure 6: Structure of the Neural Network.

Figure 7: Confusion Matrix.

In the validation and testing matrices, the hit rate was 95.5%, leading to an overall hit rate of approximately 99% that means a good performance in detecting the voice commands.

VI. Conclusions

The factories of the future, aligned with the Industry 4.0 principles, considers the human as an important and flexible piece in automated CPS. The challenging is to integrate the human in CPS ecosystems by using key enabling ICT technologies such as AR/VR technologies and IPA systems. In particular, IPA systems allow the users to improve the efficiency and quality of performed operations, especially in cases of highly customized and complex tasks, such as assembly or maintenance operations.

The developed human operation workbench uses an IPA system to support, in a friendly manner, operators in the assembly of highly customized gift boxes (with different configurations) and in the execution of complex products, e.g., during maintenance or training interventions. The IPA application uses a set of emergent technologies, namely a monitor, a video projector and a web-based application to provide information about the configuration and sequence of the assembly procedure, the image processing to verify the correctness of the performed assembly operation and the

speech recognition to support voice commands to control the flow of the assistance system.

The workbench was tested by several operators executing their assembly tasks, showing the increase of flexibility and productivity of the system. The use of the workbench allowed the entrepreneurs of small and medium enterprises (SME) of the northeast region of Portugal to understand how ICT technologies can be used to integrate humans in new CPS systems, and improve the competitiveness of their businesses.

Future work is devoted to incorporate AR and other collaborative technologies to transform the workbench into a more immersive environment where humans are completely integrated in CPS systems, towards the HiLCPS concept.

References

[1] P. Leitão, A. W. Colombo, and S. Karnouskos, "Industrial automation based on cyber-physical systems technologies: Prototype implementations and challenges," *Computers in Industry*, vol. 81, pp. 11–25, Sep. 2016.

[2] D. S. Sousa Nunes, P. Zhang, and J. Sa Silva, "A Survey on human-in-The-loop applications towards an internet of all," *IEEE CST*, vol. 17, no. 2, pp. 944–965, 2015.

[3] H. Kagermann, W. Wahlster, and J. Helbig, "Securing the Future of German Manufacturing Industry: Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0," ACATECH – German National Academy of Science and Engineering, Tech. Rep., 2013.

[4] D. Gorecky, M. Schmitt, M. Loskyll, and D. Zühlke, "Human-machine-interaction in the industry 4.0 era," in *Proceedings of INDIN'14*, 2014, pp. 289–294.

[5] EFFRA, "Factories 4.0 and Beyond," p. 67, 2016. [Online]. Available: http://www.effra.eu/factories-future-roadmap

[6] G. Schirner, D. Erdogmus, K. Chowdhury, and T. Padir, "The Future of Human-in-the-Loop Cyber-Physical Systems," 2013.

[7] P. Fantini, G. Tavola, M. Taisch, J. Barbosa, P. Leitao *et al.*, "Exploring the integration of the human as a flexibility factor in CPS enabled manufacturing environments: Methodology and results," in *Proceedings of IECON'16*, 2016, pp. 5711–5716.

[8] NIST, "Foundations for Innovation in Cyber-Physical Systems," *Workshop Report. National Institute of Standards and Technology.*, 2013.

[9] "Promoção da Indústria 4.0 na Região de Trás-os-Montes e Alto Douro, I4.0@TMAD." [Online]. Available: http://i40tmad.ipb.pt/index.php/pt/

[10] G. Boy, *Orchestrating Human-centered Design*. Springer, 2012.

[11] D. Romero, "Towards an Operator 4.0 Typology: A Human- Centric Perspective on the Fourth Industrial Revolution Technologies," in *Proceedings of CIE'16*, no. April, 2017, pp. 0–11.

[12] J. Hauswald, M. A. Laurenzano, Y. Zhang, C. Li, A. Rovinski *et al.*, "Sirius: An Open End-to-End Voice and Vision Personal Assistant and Its Implications for Future Warehouse Scale Computers," in *Proceedings of ASPLOS'15*, 2015, pp. 223–238.

[13] J. Santos, J. Rodrigues, B. Silva, J. Casal, K. Saleem, and V. Denisov, "An IoT-based Mobile Gateway for Intelligent Personal Assistants on Mobile Health Environments," *JNCA*, March 2016.

[14] "CORTEX aiASSIST." [Online]. Available: http://cortexlogic.com/aiassist/

[15] D. Bradley and G. Roth, "Adaptive Thresholding using the Integral Image," *Journal of Graphics Tools*, vol. 12, no. 2, pp. 13–21, 2007.

[16] M. Vala and A. Baxi, "A review on Otsu image segmentation algorithm," *IJARCET*, vol. 2, no. 2, pp. 387–389, 2013.

[17] A. Saxena, A. K. Sinha, S. Chakrawarti, and S. Charu, "Speech recognition using matlab," *IJAICSCC*, vol. 1, no. 2, pp. 26–30, 2013.

[18] S. Naragan and M. D. Gupta, "Speech Feature Extraction Technique: A Review," *IJCSMC*, vol. 4, no. 3, pp. 107–114, 2015.