A Concept for Integration of Voice Assistant and Modular Cyber-Physical Production System

Maxim Ya. Afanasev, Yuri V. Fedosov, Yuri S. Andreev, Anastasiya A. Krylova, Sergey A. Shorokhov, Kseniia V. Zimenko, Mikhail V. Kolesnikov
Faculty of Control Systems and Robotics ITMO University, St. Petersburg, Russia
amax@niuitmo.ru

***Abstract*—Voice interaction is helping to increase efficiency in many areas of human activity. However, most studies either research its general aspects or domestic applications. On the contrary, there is almost no research on using this technology in production field, although, the idea of retrieving data this way can be a successful solution for large production areas, where it is difficult to find the right display. The paper aims to develop the method of voice assistant's implementation as a part of Cyber-Physical Production System (CPPS) for production processes. The study includes creating a prototype of the voice assistant module based on Alice cloud voice service API from Yandex and Raspberry Pi 3 microcomputer. Main aspects of using voice assistants for accessing production data from a CPPS were described. A possible approach to the introduction of this technology is shown. And a working prototype has been successfully designed. The research shows the possibility of using voice control in production conditions and its potential effectiveness in data access automation.**

***Index Terms*—Voice assistant, cyber-physical production system, modular equipment**

I. Introduction

Historically, humanity in its activities seeks to make the work easier by delegating the processing operations to more and more improving tools up to complex automated machines. Modern cyber-physical systems are focused on minimizing the human factor in a production process. Generally, they include not only the equipment itself, but also the interaction channels with the virtual environment, where huge amounts of data can be collected and processed. Wide use of information technologies, local area networks and data aggregation systems are all aimed primarily at automating routine actions, reducing the number of errors and failures in production, and improving product life cycle support. For the same purpose a "digital twins" of production line are being developed. A digital twin is a virtual model of a physical device that simulates internal processes, technical characteristics and behavior of a real object under the influence of interference and environment.

Each product manufacturing company is faced with a large amount of data on the production process, such as: machine parameters, surface treatment parameters, sensor data and more. Different systems of data aggregation cope with the task of collecting and organizing information very well, but the task of timely delivery of necessary data to an operator is just as important. It is also necessary to emphasize the fact that data should be prompt, videlicet, the request should be processed as soon as possible, and data should be presented in a convenient for an operator form.

Although data is often displayed on monitor screens and machine displays, there are often situations where it is not enough. For example, to get the required number, a person has to first find the desired display, approach it, find the target number in the complex interface of the program and go back to their place. It is obvious that the larger the production area, the more time will be spent on such a trivial routine in terms of manufacturing process. To solve the problem of process optimization to simplify data reading, the so-called "augmented reality"—the integration of virtual objects into a real world image—is being used more and more often. For the first time, the experience of augmented reality in production was obtained in 1992 by Boeing Company: the image of wired circuits was displayed on an aircraft body, which significantly reduced the number of errors in production.

Currently, augmented reality systems are widely distributed in both entertainment and industry. Generally, such systems imply operator-wearable equipment in the form of glasses, but there are options that use the built-in camera of mobile devices like tablets. One of the most famous solutions is the German RE'FLECT company, which specializes in producing the appropriate equipment for leading German enterprises. Another brand, EON Reality, chose a simpler way and focused on developing software for smartphones and tablets. Among Russian developments Itorum MR is the most well-known.

Another way to get information quickly is to use voice assistants. According to literary sources, the visual channel gives 80–90% of all information [1], and the second most important, auditory, is practically not involved in production. For example, if a colleague stands in a workshop next to the desired screen, then it is much easier to ask him than to go and look for the needed parameter yourself. In addition, operators are often being deeply involved in working process so that voice control remains the only possible way of interacting with an information system.

The benefits of voice assistants were assessed back in the 70s, when the first prototypes of speech recognition devices appeared. Lately, almost all leading IT-companies have among their developments a voice assistant: a voice input by Google, Alexa by Amazon, Siri by Apple, Cortana by Microsoft, Alice from Yandex and others. At the moment, their capabilities have significantly expanded from a simple search for information to

978-1-7281-2927-3/19/$31.00 ©2019 IEEE

the full voice control of various household devices. Of course, the use of voice assistants in a production environment has its own nuances like noise pollution of manufacturing premises, simultaneous interaction with several operators, and security issues.

Another important task is the integration of a voice assistant system in a structure of a modular architecture. Modularity is a modern concept of organizing a decentralized control system for production equipment aimed at increasing flexibility and scalability of a system. Plainly, such architecture implies not only physical components, but also its own software, which simplifies the integration of new components into a control system. In the platform architecture described in previous papers, each piece of equipment can be represented as a set of modules, otherwise stated, managed by a dispatcher unit.

Most modules are closed autonomous entities working on their tasks, but dependent system components may also exist. Dependencies mainly include sensors that provide one-way data sending to a request. Naturally, the voice assistant module will be autonomous, since its task is only to communicate with CPPS directly, download data from a "digital twin" and issue it to an operator. Next the organization of the voice assistant in industrial equipment will be discussed.

The article is organized as follows. Section II is dedicated to an overview of related work. In Section III the reader can find a description of the prototype module of the voice assistant module. Section IV describes the limitations of the technology and how to overcome them at this stage of development. The final section is dedicated to an analysis of the results.

II. Related work

Despite the fact that in recent years, interest in voice assistants has been steadily growing, today there are not many implementations for its industrial and manufacturing use. The most interesting work in this area is the initiative of connyun GmbH companies to introduce Alexa voice assistant from Amazon to industry [2]. The platform I_Station Optimizer, developed by the company, will connect with Alexa voice assistant in production processes. This will allow workers and employees of enterprises to conduct a direct dialogue with the IIoT platform, for example, to request data, set tasks and solve problems in dialogue with the voice assistant.

There are also a sufficient number of theoretical studies related to the analysis of existing cloud services and prospects for the speech recognition technologies and voice assistants' introduction and use in various areas of production for solving the widest spectrum of problems.

In the study [3], in particular, a detailed analysis of the automatic voice recognition applicability in medicine, industrial robotics, forensic science, arms industry, and aviation was conducted. It has been shown that with the current level of technology development, automatic speech recognition can be a worthy alternative to the classical tools of human-machine interaction. Nevertheless, the authors come to the conclusion that a success of speech recognition applications requires the limitations elimination in modern technology, the use of new specialized hardware, in particular. However, it should be noted that this work is primarily aimed at implementing an autonomous speech recognition system that does not use cloud service resources.

Hoy [4] provides a comparative analysis of popular cloud services that provide voice assistants. The paper discusses basic principles of modern voice assistants' work and its common features, and also discusses some confidentiality and security issues, as well as some potential possibilities of using it in the future.

Bradley et al. in his work [5] addresses the applicability problems of interactive agents, made to simplify the work of software developers. As an example, a voice assistant prototype called Devy, based on the Amazon cloud service from Amazon, is described. The implemented prototype demonstrates that developers can successfully launch complex workflows without interrupting their current work. Based on the study, the authors conclude that in the future, voice assistants will be able to increase the productivity and/or efficiency of software developers, allowing them to focus on their core tasks.

In general, it can be noted that recently there has been a steady increase in the number of works devoted to voice assistants' introduction in various areas of human activity. However, most of studies are either of research nature and assess the potential use of voice assistants in the future, or are devoted to information security of this approach, or are focused on the consumer sector and are associated with home automation and mobile applications. There exists virtually no work on the integration of voice assistants to manufacturing production associated with the fourth industrial revolution. In this regard, it could be assume that the research in this direction is relevant and appropriate.

III. Proposed Approach

A. Digital twin of modular CPPS

A digital twin is the basis of the described modular CPPS. An important feature of a digital twin is that the information from sensors of a real device operating in parallel with a twin is used as input influences for it. Work is possible in both online and offline modes. After that, it is possible to compare the information of virtual sensors of a digital twin with sensors of a real device, identifying anomalies and their causes. Differently put, a real production process is combined with its digital model, and any impact on the model leads to a change in the physical process and vice versa.

The modular CPPS is based on a module. It can be a standalone device or a part of a production equipment unit. The concept of modular equipment is discussed in more detail in [6]. According to this concept, the core of a CPPS digital twin is a distributed registry, which is a decentralized database. From a technical point of view, this registry is a binary JSON tree that stores all data on a production process and also allows modules to register in the system and merge to solve common tasks (within an industrial equipment unit), in other words, it is an aggregator of data and services of CPPS [7]. The interaction

between modules is carried out through a mesh network based on the OpenThread protocol [8], [9].

A registry is a collection of nodes, each of which is either a unitary or a modular device. In both cases, all the functions of an equipment unit are determined by a dispatcher, in which modules can be registered. If only one module is registered in a dispatcher, this is a unitary device; if several, it is a modular unit of industrial equipment. Otherwise stated, a dispatcher can be configured to connect a certain number of modules, which can vary from 1 to 32, which is set at the design stage of new equipment integrated into CPPS.

A module is registered in a controller when it is physically connected to it. Each module is allocated to a separate slot. A slot is a structural information unit of a registry that stores information about services, received commands, and data that can be transmitted to a CPPS network. The slot includes the following fields:

- address;
- the name of the module/sensor;
- functions;
- return value;
- the limits of the return value.

A data set from all dispatcher modules determines the parameters of a piece of equipment, the data set from all dispatchers defines CPPS parameters in general.

Robotic systems (warehouses, automated guided vehicles, multi-axis industrial robots), conveyors and production lines, industrial equipment with computer numerical control can be modular. On the other hand, various sensors of a production process, as well as equipment for indication and notification, are unitary devices, i.e. from software implementation point of view, they combine the functions of a dispatcher and one module connected to it.

As an example, modular equipment with computer numerical control (CNC) is considered. Such equipment can be conditionally represented as a set of multi-axis chassis, moving in the space of any module that determines the purpose of an equipment unit, as well as a set of certain external units. For example, equipment for the selective photopolymers curing consists of three modules: a chassis, a laser head and an external source of laser radiation, as well as a set of sensors that determine the return value.

In addition, each of the modules registers its functions as available G-codes and M-functions according to standards ISO 6983-1 and ISO/TR 6983-2. The return value does not have to be just one. The description of each of them includes a type, limits and an array of values with timestamps. The dispatcher controls the output of each value beyond the permissible limits, and in the event of such situation, it analyzes the error that has occurred, and the decision is made whether it is possible or impossible to continue operating the equipment.

B. Voice assistant Integration into modular CPPS

According to the concept presented in the previous section, a voice assistant will be a unitary device operating at the CPPS level. Voice assistant's main function is to receive commands, transfer them to a cloud service for recognition, and also output the result in the form of an audio message. A general architecture of modular CPPS using voice assistants is shown in Figure 1.

As noted earlier, in the registry each module is associated with a set of sensors and the limits of acceptable values for each of them. Therefore, the first scenario of the voice assistant's work is an alert when a parameter is out of acceptable values. This task is multilevel, since the registry structure is flat, in other words, data from all sensors is always available at the CPPS level. Values that exceed permissible range are a violation of CPPS and spread throughout the network as soon as one of the controllers detects this event.

Therefore, the question at what point to carry out the notification of any parameter violation arises. The following solution is proposed. All sensors in the system are ranked according to a degree of their influence on the production process. According to these ratings, a separate custom alert policy has been implemented, which includes the following levels:

- General production level (critical error, all modules are synchronized in time and the message is transmitted via speakerphone).
- Local error of an equipment unit (the message through the closest module, the OpenThread protocol allows determining the neighboring nodes in the mesh network).
- Minor error (the message is not pronounced, but entered into the general log; a signal lamp is lit or a warning is displayed on a remote control).

The ranking is carried out by experts using the analysis of the subject area at the CPPS design stage. From the general registry point of view, an additional level field is added for each sensor with values ranging from 0 to 2 (low, medium, above average, high).

The second scenario of using voice assistant in the presented CPPS structure is to obtain data on the current equipment parameters. According to the accepted model of data representation in the registry, each unitary module and each equipment unit has a unique machine-readable identifier that allows you to search for it in the registry. For the voice assistant, an additional identifier has been added, which is displayed on the screen of the voice assistant module (see the next section). The identifier in this case is a four-digit number unique to the production facility.

Commands which help an operator to request from the assistant a list of all modules or equipment closest to him, if it is out of sight of the operator, are implemented in the system. Other sampling commands for any parameters can also be implemented. For each unitary module or equipment unit, the voice assistant can list all available parameters. As a result the following voice command is built {**keyword for accessing the voice assistant**} {**identifier**} {**parameter**}, for example "Alice, what's the temperature on the 4078 sensor?". Human memory works in such a way that after some time the operator will remember production units identifiers important for him

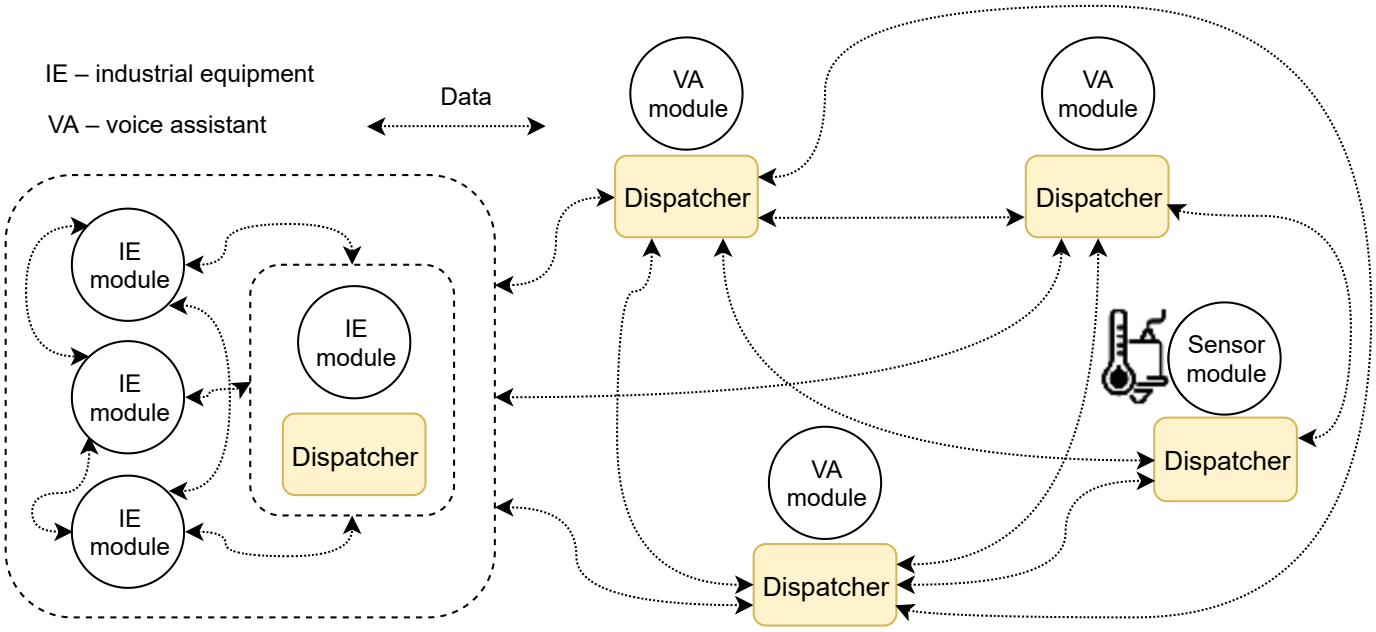

Fig. 1. General architecture of the CPPS with the voice assistant modules.

and the names of parameters that interest him most often and that will significantly increase the speed of data access.

C. Software and hardware implementation of the voice assistant module

A voice assistant module includes a control unit and peripheral devices: a microphone, a dynamic head (speaker) and a display. The module prototype is built on the basis of the Raspberry Pi 3 microcomputer and the nRF52840 Nordic Semiconductor module, which is responsible for the radio channel communication using the OpenThread protocol. An appearance of the described prototype is presented in Figure 2.

The prototype uses ReSpeaker Core v2.0 voice recognition board (https://respeaker.io/rk3229_core/) and the Alice cloud voice service API [10] from Yandex, which allows user to interact with the cloud service using a software library written in the Python programming language. The microcomputer is responsible for receiving data from the microphone, sending messages to the built-in speaker and integrating into the general structure of the CPPS, as well as accessing the Alice cloud service (Figure 3). Also in the structure there is a service for receiving commands from Alice and translating it into requests used in the modular CPPS under consideration (generating a message using the *nanomsg* protocol to fetch data from the registry).

IV. Discussion

Despite a number of obvious advantages of using voice assistants for the operator's operational interaction with CPPS, built based on a modular approach, a range of problems can be identified that need to be solved at this stage of development. Such problems may primarily include:

1) Noise pollution of manufacturing premises.
2) Interaction with several operators simultaneously.
3) Control or monitoring, security issues.

A. Noise pollution of manufacturing premises

It is not a secret for anyone that any industrial production is associated with an increased noise environment, the source of which is primarily industrial equipment. There are standards that regulate a permissible ambient noise in manufacturing. Different maximum allowable noise levels have been established in different countries, but mostly this parameter ranges from 70 to 90 dB [11].

Fig. 2. Appearance of the voice assistant module.

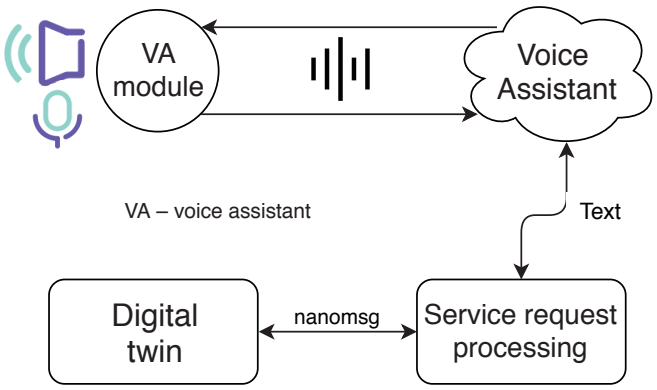

Fig. 3. Integration with Yandex cloud service Alice.

It is obvious that such noise values have a negative impact not only on the hearing organs of operators, but also on microphones used in modules of the considered voice guidance system. Certainly, the correct solution to this problem could be a complete rejection of the background analysis of incoming commands from an operator. Indeed, various technologies of active noise cancellation, filtering and signal amplification are implemented in modern acoustic headsets. The most illustrative in this regard are special solutions used for military purposes but household appliances are often able to provide acceptable performance as well, because the noise level of modern megalopolises often exceeds that in manufacturing.

However, this decision imposes a number of significant restrictions on voice assistants' usage in the workplace. Firstly, the use of headsets reduces the working convenience for an operator, because its constant wearing is associated with a number of disadvantages. People often do not feel comfortable wearing such devices for a long period of time; it reduces the ability to communicate with other operators, which, with industrial noise background added, can lead to emergencies or accidents. Apart from that headsets require constant maintenance, recharging etc. Lastly, headsets must be wireless and interact with CPPS via a separate radio channel (now, this is usually bluetooth). The developed CPPS modular architecture uses the OpenThread wireless protocol [12], which is not designed to transmit audio data, and it is highly undesirable to simultaneously deploy a bluetooth network operating in the same frequency band.

Thus, it becomes obvious that the use of acoustic headsets for interaction between an operator and CPPS is impractical. Instead, it is proposed to add a certain redundancy to the system, which can be implemented with the modular approach used.

As noted earlier, all CPPS modules are located in a common decentralized network, therefore they can communicate with each other and exchange information on commands received and the signal level at the receiving device (microphone) used. Accordingly, if one of the modules detects a voice command from an operator but cannot decipher (understand) the signal, it informs the nearest modules that a signal, different from the background noise, has been received but cannot be recognized. At the same time, it is possible to select one or several modules that will determine the parameters of the background noise of a production room as a whole or its individual parts and record this data into the CPPS database so that other modules can filter the background noise and extract voice commands.

Thus, if one of the modules recognizes a command and can process it, all its neighbors are informed that the command has been received and its processing begins now. Consequently, all voice accompaniment modules become a kind of omnidirectional microphone that perceives commands from the entire production room. At the same time, the quality of commands recognition can be significantly increased with the correct selection of the number of required modules and their location.

It should also be noted that the problem of increased noise exposure is receiving new solutions, and the overall noise level of production equipment is reduced due to improved noise insulation and the use of modern technical tools. All the mentioned tendencies should have a positive impact on the voice assistants implementation in various fields of industrial production in the future.

B. Simultaneous interaction with several operators

The considered approach to voice assistants' introduction in production suggests that such a system of interaction with an operator will not be individual. The voice assistant, in this case, is not an analogue of an operator's console, but a part of CPPS, which any of the operators can interact with by giving commands in the background mode. Clearly, such a system has to effectively recognize voice commands from several people, and it is necessary that the following requirements will be met:

1) The system must be self-taught, i.e. a preliminary recording voice samples from different operators for later recognition is unacceptable.
2) The system should be able to receive commands from different operators in an arbitrary order, otherwise stated, be able to quickly switch from one operator to another.

This problem is partially solved due to the modularity described above, when all voice tracking modules communicate with each other and can control the signal level of each neighboring module. In this case, all the assistants work within a single model (digital twin). In fact, it is difficult to imagine a situation in production process when several operators, being in close proximity to each other, try to interact with the voice assistant. A traditional production facility is divided into independent zones or areas, occupied by only one person.

Nevertheless, work on the recognition and separation of commands from several operators is still being conducted and there are certain positive results. In particular, Hershey et al. [13] describes an approach called deep clustering. The use of this method makes it possible to effectively separate and reconstruct the speech of two people speaking into one microphone, with an accuracy of up to 90%, three—up to 80%. At the same time, an approach similar to the one under consideration makes it possible to achieve an accuracy of 51%

with the use of two microphones connected to the same voice recognition system without measuring the signal level [14].

C. Control or monitoring, security issues

An important issue that needs to be addressed when using voice assistants in production is a safety issue. In this regard, it is necessary to immediately determine which CPPS functions can be implemented in the voice assistant interface, and which functions cannot. It is obvious that, like any other human-machine interface, the voice assistant can be used for both input and output of information. The implementation of the voice output system, in response to the command given by an operator, seems to be quite safe. For the vast majority of industries, the leakage of data related to the production process is not critical, especially if it is related to the current parameters of production facilities and equipment. Much more serious are commands for managing the production process, the parameters of production premises or equipment. Here safety issues come to the foreground. On the one hand, access to the voice assistant is collective and does not require authorization. That is what makes its use so convenient and allows one to reduce the time needed to obtain operational information about a production process. Any kind of authorization would again require the presence of wearable devices associated with an operator that will control access rights to execute certain commands. The obvious solution could be the storage of voice samples of each operator, in other words, the use of biometric parameters, but here rises the question about whether or not the system understands if a command is really being pronounced at the moment or the recording is being played, that is, authentication is needed then. Based on the analysis of the sources, it can be concluded that with the current quality level of the sound reproducing equipment, this problem cannot be solved by software. On the other hand, the issues concerning security and access control at any enterprise are addressed centrally and include, among other things, the control of personnel access to certain premises. In other words, security issues concerning the use of voice assistants could be resolved in a complex and have a multi-level structure. Physical access to equipment should be carried out only through the operator's console using the appropriate authorization mechanisms and commands to control certain non-responsible parameters. In this case an apropriate example is the scenario of using an automated guided vehicle (AGV), when an operator of a production site calls an AGV using a voice command to load finished products. Since each unit of the equipment is associated with its own voice tracking module, AGV automatically obtains the necessary coordinates and moves to a given point. After that, loading takes place and an operator gives a voice command to move the AGV to the next point, to an automatic warehouse for example.

V. Conclusion

Since the advent of studies on speech recognition technology, voice control has been one of the attributes of the life in the future, however, currently, these devices have already become part of the present time. The input audio information processing algorithms continue to improve its quality and recognition speed; there are already a large number of commercial products with extensive functionality, the ability to filter out ambient noise, tune in to the voice of an operator and execute fairly complex commands.

In this paper, aspects of the use of voice assistant were considered as a channel for accessing production data from a CPPS. The introduction of such technology can be justified in large production areas, where it is difficult to quickly find a display with necessary information and get to it. The approaches to the introduction of technology into modular equipment were analyzed, possible operational difficulties and tasks requiring further research were identified.

Acknowledgment

This work was financially supported by Government of Russian Federation, Grant 08-08.

References

[1] N. Lindgren, "Organy chuvstv zhivotnyh i ih ehlektronnye analogi [The organs of sense of animals and their electronic counterparts, in Russian]," *Electronika*, vol. 35, no. 7, pp. 22–27, 1962.

[2] K. Tubbesing. (2018) Alexa is engaged as voice assistant in industry. [Online]. Available: https://www.hannovermesse.de/en/news/alexa-is-engaged-as-voice-assistant-in-industry-90434.xhtml

[3] J. Vajpai and A. Bora, "Industrial applications of automatic speech recognition systems," *Int. Journal of Engineering Research and Applications*, vol. 6, no. 3 (Part 1), pp. 88–95, March 2016.

[4] M. Hoy, "Alexa, Siri, Cortana, and more: An introduction to voice assistants," *Medical Reference Services Quarterly*, vol. 37, pp. 81–88, 01 2018.

[5] N. C. Bradley, T. Fritz, and R. Holmes, "Context-aware conversational developer assistants," in *Proceedings of the 40th International Conference on Software Engineering*, ser. ICSE '18. New York, NY, USA: ACM, 2018, pp. 993–1003.

[6] M. Y. Afanasiev, Y. V. Fedosov, A. A. Krylova, and S. A. Shorokhov, "Modular industrial equipment in cyber-physical production system: Architecture and integration," in *Proceedings of the 21th Conference of Open Innovations Association FRUCT*, November 2017, pp. 3–9.

[7] ——, "An application of microservices architecture pattern to create a modular computer numerical control system," in *Proceedings of the 20th Conference of Open Innovations Association FRUCT*, April 2017, pp. 10–19.

[8] C. Garrido-Hidalgo, D. Hortelano, L. Roda-Sanchez, T. Olivares, M. C. Ruiz, and V. Lopez, "Iot heterogeneous mesh network deployment for human-in-the-loop challenges towards a social and sustainable industry 4.0," *IEEE Access*, pp. 1–1, 05 2018.

[9] M. Y. Afanasiev, Y. V. Fedosov, A. A. Krylova, S. A. Shorokhov, and Z. Ksenia V., "Mesh networking in cyber-physical production systems: Towards modular industrial equipment integration," in *Proceedings of the 23rd Conference of Open Innovations Association FRUCT*, April 2018, pp. 3–11.

[10] (2018) Yandex dialogs documentation. [Online]. Available: https://tech.yandex.ru/dialogs/alice/doc

[11] L. S. Goodfriend and F. M. Kessler, *Industrial Noise Pollution*. Boston, MA: Springer US, 1973, pp. 572–586.

[12] (2018) Openthread node roles and types. [Online]. Available: https://openthread.io/guides/thread-primer

[13] J. R. Hershey, Z. Chen, J. Le Roux, and S. Watanabe, "Deep clustering: Discriminative embeddings for segmentation and separation," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 31–35.

[14] J. C. Murray, S. Wermter, and H. R. Erwin, "Bioinspired auditory sound localisation for improving the signal to noise ratio of socially interactive robots," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2006, pp. 1206–1211.