Concept of Implementing Computer Voice Control for CNC machines Using Natural Language Processing

Yury Fedosov, Aleksei Katridi
ITMO University
Saint Petersburg
{yf01, katridi}@yandex.ru

Abstract—**Currently, rapidly developing software and hardware technologies in the machine learning field boost natural language processing in many applications: from interactive voice response systems to fully automated customer support with chat-bots and even conversational bots for the entertainment industry. However, there are still fields that do not use machine learning to solve existing tasks and challenges. This paper introduces a high-level architecture of voice user interface, describes its requirements and use-cases in the context of using computer numeric control machines. Moreover, while other researchers suggest trying to automate the manufacturing process as much as possible to exclude humans from the manufacturing process, this article suggests choosing a slightly different approach and taking a fresh look at how the user interacts with the machine. The study proposes a new interaction scheme using computer voice control as an analog to computer numeric control term. Such an approach allows simplifying the human-machine interfaces on a machine and improving the safety of use and efficiency of interaction.**

I. Introduction

To provide an easy and convenient human-machine interface, in particular voice control, many IT companies address the following tasks: human-related (language processing, emotion recognition, intent recognition, speech to text, text to speech, etc.), as well as technical problems (latency, performance, design) to form a proprietary system for the customer and sell products using various ways (subscription, advertisement, and others). The area of Natural Language Processing (NLP) grows rapidly. Although there are many issues and challenges connected with retention [1] and monetization, this field constantly changes, and the quality of virtual assistance also improves. Interestingly, some unforeseen events, such as the COVID-19 pandemic, may increase the number of users [2]. However, almost all use-cases are oriented to a broad audience, while some niches could benefit from the same technologies but do not apply them.

II. State of the art

The Fourth Industrial Revolution is a complex term that could be defined through a semi- or fully automated decision-making process and overall production cycle improvement using new technologies. The group of authors made a review and suggestion about development in industrial engineering [3]. They claim people will make production systems as much as “30 percent faster and 25 percent more efficient” and elevate mass customization to new levels [Ibid.]. Although numbers may vary, the review captures the general trend that using state-of-the-art technologies from different agendas could significantly improve production.

The researchers highlight nine pillars: Big Data and Analytics, Autonomous Robots, Simulation, Horizontal and Vertical System Integration, Industrial Internet of Things, Cybersecurity, Cloud, Additive Manufacturing, Augmented Reality [Ibid.].

Most of the proposes on improving the production process generally fall into two categories: solutions either exclude the need for human interference or improve interaction interfaces between machines and humans.

There is research on human-free tasks or where little interference is needed [4, 5]. The suggested approaches are designed to increase the level of autonomous functioning. However, this approach cannot be applied to every case in terms of cost and limitations; thus, a worker still needs to have a simple and efficient way to interact with machines and make decisions based on data and forecasts provided.

An industrial equipment operator faces new challenges, such as increasing demand for customization and a constantly changing environment. Therefore, the new type of personnel has to be able to operate quickly and precisely to become "operator 4.0".

Manufacturing is yet human-centric [6]. Researchers still attempt to find how human-machine interaction will be performed in the new context and how “operator 4.0” will work [7].

Some researchers proposed to adapt computer vision approaches to take advantage of visual data processing [8], but due to limitations and a non-intuitive way of interacting with machines, it could be applicable only in limited cases.

Others proposed VR or AR and digital twins, in general, to complement the usual layout with some visual information to bridge a gap between cyber- and physical space and improve diagnosis and decision-making process [9–11], as a predictive analysis does to improve decision making and make diagnosing easier [12–14].

Some researchers claim [15] that usage of voice assistants in e-commerce is growing rapidly, yet they are not considered an interaction channel in an industrial environment, although voice technologies are no longer just a chat-bot to answer frequently asked questions. Such technologies could perform complex tasks, e. g., reserving a table in a restaurant [16] or helping tenants in hotels [17], and slow but sure become a part of our environment.

Still, these examples are goal-oriented and domain-specific. Researchers reported that their model needs no tuning to professional language meaning [18]. However, even though the ability to precisely recognize sentence and intention associated with it is decent, flow and scenarios still need to be implemented considering the risks associated with industrial specifics. Safety concerns impact the adoption speed of such technologies and the use of robots and cobots [19].

To date, there is no universal interface that would be widely used by different manufacturers and was standardized. It is hard to measure the economic impact of improving interfaces, as opposed to predictive algorithms or autonomous robot lines that operate without human interference [20]. Moreover, a clear and unified standard interface across all industries may contribute to working faster, less error-prone, and more effectively. This advantage should not be underestimated. The speed of adoption in the consumer domain is also likely to impact the working environment since the human expects devices to communicate through a voice channel. Thus, the last pillar of technological advancement is improved human-machine interaction (HMI). Fig. 1 shows a basic scheme of interaction.

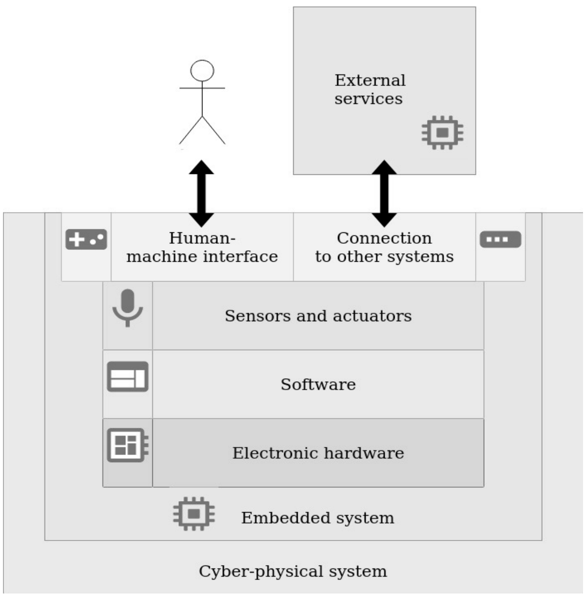

Fig. 1. Interaction between humans and machines via cyber-physical systems

Voice technologies could provide a more comfortable user interface and more safety for the user since they do not require additional interfaces and additional control for manual operations. To effectively adopt such technologies, one needs to understand which of their components are crucial and what problems exist in adopting such technologies.

III. Key components: purpose and main challenge

Mobile applications such as Siri, Google Assistant, Alexa, and others, which combine visual and auditory information, and voice-only devices, such as Amazon Echo and Google Home, are considered mainstream. Google reports that 20 % of its searches are now done via voice [21].

Voice has some decent features:

- saying command could be faster. The study showed speaking (dictating) text messages was faster than typing, even for expert texters [22];
- some cases require hands to be free, such as driving or cooking; speaking rather than typing or tapping is much more practical (and safer);
- this interface comes almost naturally because, from early ages, people can talk;
- voice, which includes tone, volume, intonation, and speech rate, conveys a great deal of information and emotions.

However, there are situations when voice could be less appropriate to use:

- a crowded workspace could be a potential source of artificial sounds that could corrupt interaction with the machine;
- although virtual assistants are becoming more popular, not everybody is comfortable speaking to a device;
- texting could be a casual way of interaction, and people usually do not appreciate changes;
- besides some discomfort about reading messages out loud, this could also violate the user's privacy and expose sensitive data [23].

When designing a VUI, people consider one-off tasks, such as answering a search query, setting up a calendar appointment, making a phone call, playing a song, etc. Sometimes, these tasks can be accomplished instantly. The best VUI designs also consider the next step since it could be a tree-like scenario considering the previous context [Ibid.].

The ability of modern applications to understand the user's speech and take appropriate actions is provided by a combination of three essential technologies: automatic speech recognition (ASR), text-to-speech (TTS), and natural language understanding (NLU).

The first component is the ASR component. The main problem of its use is the correct perception of sound and its translation into the text and dealing with various cases with a high percentage of correct recognition. Besides the linguistic complexities, there are also problems related to the environment in which the text is pronounced. A recognition model needs to be sufficiently robust to ambient noise.

The second component is the NLU component. NLU is an extensive-term. In our case, we focus on several aspects in this area:

- intent recognition:

by the incoming phrase (already in text form), the model tries to understand what exactly the user wants to get from the system—this is a non-trivial task when the system has a wide range of choices, and you need to choose the right choice between the options;

- entities extraction:

can be named NER (named entity recognition), for example: “Send this to John Smith."

Or it can contain some necessary data (slot filling) to transfer this data for the selected scenario, for instance: "Move to coordinates 20 20".

Finally, the last of the components is text-to-speech (TTS). TTS is required to respond to the user. These tasks are similar to ASR. It may be worth neglecting the liveliness and naturalness of speech in the industrial field in favor of the clarity and correctness of the information spoken since the error cost is high.

A natural language is difficult to formalize as the longer the list of tasks to be solved, the faster the complexity of this component's implementation grows.

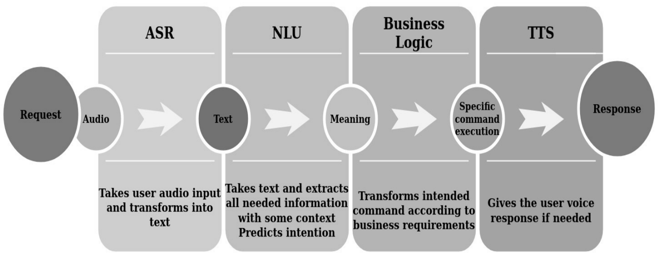

Fig. 2. Basic VUI pipeline

According to the pipeline represented in Fig. 2, the voice receiver grabs user input, sound waves are processed to an audio format, and the audio is processed to text format. The text is then sent to the NLU module with some context data: date, previous command stack, and all potentially useful information to improve the quality of the decision-making process for the system. The NLU module is based on overall input data producing predictions and entity extraction and sending them to the business logic component, where the domain decisions are made. Then, considering the chosen action, an operation is performed, and the user gets information about the result.

Although for the user verbal communication could seem an easy effort, for such system the task is quite complicated as besides dialogue strategies and the way data is exposed to the user and overall design, the inner machinery has to meet some critical requirements. One of them is latency. All data processing—from translation waves to audio format, audio to text, text to entities, and basic reasoning with those entities and way back to the user—comes at a price.

In graphical applications, when an application is loading, the user is waiting. Nothing can be done until an application is ready. The equivalent wait time in VUIs is waiting for the application to respond, recognizing what the user said, and understanding what they meant. The processes are time-consuming, and the latency time must be as short as possible.

Latencies are generally caused by:

- poor connectivity;
- data processing;
- request to external services.

Thus, the required operations could be either CPU-bounded or IO-bounded in different stages.

IV. Computer voice control

Given the same input conditions, computer numerical control (CNC) machines allow a more reliable, safer, and more predictable production process. The control system of machine tools has changed significantly, going from punched cards and punched tapes, which control program execution, to microcontrollers, which allow to perform a wide range of tasks and provide a more convenient interface for machine control. Today, modern machines enable to write a rather complex program consisting of different stages, run a simulation of the machining process, etc. Therefore, they are used intensively in manufacturing processes.

CNC machines have rich functionality and a wide range of capabilities. However, when it comes to the learning curve and successful mastering of interaction interfaces, there appear some difficulties. In addition to the standard of G-codes [24] itself, understanding the physical and mechanical processes that take place during machining, understanding the engineering task, that is, drawings and compliance of the manufactured part with the required technical conditions, there is a problem of the complexity and specificity of control of specific machines with a specific system CNC.

Fig. 3. CNC control panel SINUMERIK 808d [25]

Despite the significant efforts of manufacturers to unify the panel interface with which the user interacts, the efficiency of its use and ease of management is a challenge for panel creators. The extreme reluctance of operators to learn how to operate another machine with a different interface indicates that learning a specific interface is time-consuming.

Although Scientists have researched the topic of machine tool control using a voice interface [26-31], there is no universal standard for interaction with the machine through this channel. This is most likely because end-users do not see the practical use and scenarios for implementing interaction using such an interface. For successful implementation, the voice interface must be not only a way to close/open a door or move the machine to the desired coordinates [Ibid] but be a full-fledged interface for controlling the CNC machine, if not completely duplicating or expanding its functionality, then at least supplementing it. The practical value lies not in the execution of a specific team, which is not an unattainable task with the current development of technology, but in building up whole scenarios of interaction and the acquisition of practical use-value.

Usually, such studies confirm the hypothesis about using such an interface for specific equipment as proof of concept [Ibid].

It follows that such studies do not raise the issue of unifying groups and commands for different equipment but use ad-hoc solutions for specific equipment. That fact runs counter to the idea that, in order for the interface to be used, in addition to the value it carries, there must be portability and versatility for different groups of machines and from different manufacturers. So that not only one machine supports it, but any that have a CNC and a bus for data exchange to receive commands from it, thus reducing not only the speed and safety of using the machine but also the complexity of the transition from one equipment and the ability to use something else. Thus, it affects the speed of adaptation of such inventions. As a result, it is possible to reduce the cost of recruiting machine tool maintenance personnel by unifying the interaction interfaces.

In this study, to designate an additional interface that can be used to control a CNC machine and obtain reference information, the term computer voice control (CVC) is used similarly to CNC. As shown below, such interface can improve safety when working with CNC equipment, reduce the time spent on control and greatly simplify the interaction interface. Individually or collectively, these benefits can significantly impact lower manufacturing costs and increased productivity in instrumentation plants.

The enlarged scheme of interaction is shown in Fig. 4.

The key to this scheme is the ControlBox. It receives a command from the operator, sends it to the server for processing, receives a response from the server and responds to the user, or issues a command to execute the machine.

The installer has a headset and a plug-and-play [32] device (ControlBox), which they connect to the Ethernet port of the machine as, from the headset, voice commands are transmitted to the ControlBox via Bluetooth, the ControlBox via Ethernet sends an audio stream to the server, in the same place the command is defined and converted for a specific version of the rack program and is sent back to the ControlBox to issue appropriate commands to the machine's actuators by the machine mechanics.

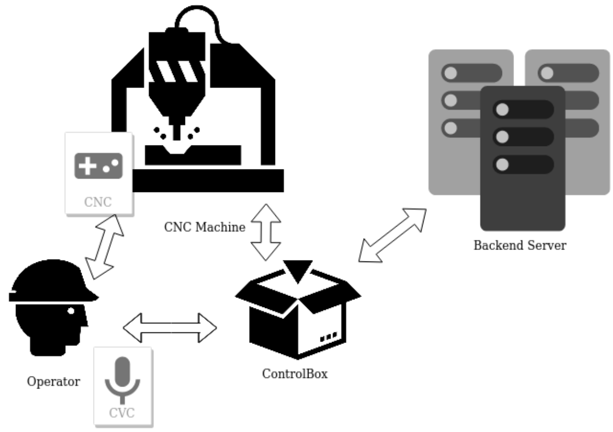

Fig. 4. Scheme of interaction with CNC machine tools using CVC interface and traditional

The ControlBox must have several key components to complete tasks:

- USB port (for charging and data exchange);
- Ethernet port (for interaction with the server);
- Ethernet port (for interaction with the machine);
- Bluetooth (for user interaction);
- CPU (to control the entire process);
- RAM (for short-term storage of information);
- Memory (for long-term storage of information).

Most modern CNC machines have Ethernet ports through which programs can be downloaded to execute them. One can use this functional feature and take this fact into account to adapt the ControlBox, which, when connected to the CNC machine, requests its parameters and gain access to the machine control. Obviously, in addition to the technical issues (without the necessary access rights, the command cannot be executed), there is also a legal aspect. The manufacturer can accuse such control method as external interference, and in the event of equipment failure, a machine will not be held liable and warranty serviced. However, if the manufacture approves such control method, the necessary audit and testing are done, then this dramatically simplifies the implementation and use of an interface in actual production.

The command comes from an operator or service technician through a headset connected via Bluetooth to the ControlBox, which in turn sends information to the server, where computational operations are performed. As mentioned above, it can be either a cloud (but with limited access to ensure data privacy) or some kind of internal server of the enterprise. After processing the results on the server, the data goes back to the ControlBox, which either gives a command to the machine or gives an answer to the user on the requested command—it depends on how the interaction of voice input and command execution is programmed.

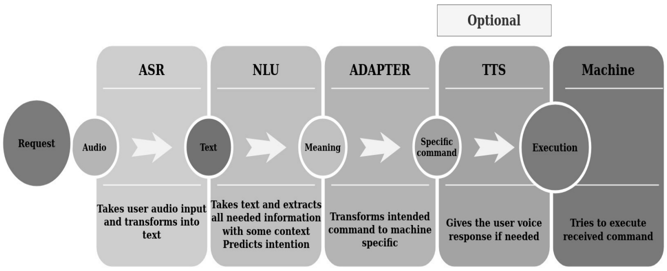

Fig. 5. Pipeline of proposed integration from request to execution

Technical details and implementation lie out of the scope of this paper and would be introduced and compared side by side in further works. Particular algorithms for automatic speech recognition or text processing are less important than high-level overview because they could change over time, and new methods could be introduced.

Several key components ensure the work of such bundle:

- ASR;
- NLU;
- ADAPTER;
- TTS.

ADAPTER is a specific component in this chain. Usually, at this stage, there is some kind of business logic needed to build the process of interaction with the user. This is a special component that, based on the previous steps, decides which command should be given to the machine for execution. However, the most important is the command itself. It is initially stored in an internal representation and then transformed to satisfy the protocol of interaction with a specific machine, allowing a universal interface for interaction with different manufacturers at the software level. We can take advantage of a component, such as Adapter, to translate our script commands into specific commands for a specific machine. Adapter works by analogy with post-processing in CAM systems: the system uses an internal representation converted into commands for the equipment to which the ControlBox is connected.

Nevertheless, it is imperative not only to consider the security of new implementations themselves but also how these implementations can affect the system's security as a whole. An example would be the "red button" interface, which is an emergency stop. Different machines can have different sized work areas, such as large open-type milling machines. When setting up a machine, the operator can be in the working area and, even if there is no remote control with the same interface at hand, they can stop the machine with a voice command in case of unforeseen circumstances. Thus, the voice command interface can save the operator from a serious injury or death and the equipment from damage.

V. COMMAND GROUPS AND CVC MODES

Commands for machine tools can be roughly divided into operational and diagnostic. For operational commands, one can use the principle of explicit confirmation when the system tells the user what will be done, and the user confirms the command. Thus, the command is executed only if there is explicit confirmation. Diagnostic (help) commands require no explicit confirmation, but implicit confirmation can be used, for example, to check whether the classifier assures that this specific help is being requested.

There are different approaches to verifying commands. They can be divided into explicit and implicit. The explicit approach is a way of communicating with the machine, in which, before executing a command, the machine announces the action to be performed and waits for confirmation. The user speaks a certain command, but it does not start executing yet because it requires explicit confirmation of this command. In the case of implicit confirmation, the command is executed without additional validation if the confidence of classification is higher than a threshold for execution.

In situations where the cost of error is high, no doubt an explicit confirmation approach should be taken. Despite the fact that this interface is more formal and less user-friendly, such approach can avoid unwanted consequences and make the system behavior more predictable for an operator.

To balance these two approaches could be applied hybrid method of confirmation: usage of three-tiered confidence [23]. In this case, the system will explicitly confirm information between a certain threshold for predicting probability, reject anything with lower confidence, and implicitly confirm anything above some high probability threshold. It is especially important to explicitly confirm information if the cost of a misrecognition is high.

A factory is a place where there is a potential danger to employees' life and health and where expensive equipment is used, the damage of which also carries great risks. Thus, special attention should be paid to safety considerations. When working with complex equipment, such as CNC machines, all standard safety precautions must be followed as improper use is a potential source of injury and equipment damage. Thus, these factors should influence the design of the dialogue system that the operator will use.

There is a variety of ways VUIs can make mistakes:

- no speech detected;
- speech detected, but nothing recognized;
- something was recognized correctly, but the system processed it wrong;
- something was recognized incorrectly.

Moreover, it is necessary to duplicate the possibility of a voice emergency shutdown of equipment. When there is a command responsible for the emergency shutdown of the machine and which must be processed locally on the device itself, an architectural scheme is worth designing. Because time is critical in emergency cases, such scheme reduces the waiting time for a response from the server, as some inaccurate commands can lead to severe damage or injuries. It is required to select the correct voice command, convenient enough to use it if necessary, with the possibility of false triggering minimized.

In order to implement the system, it is necessary to take into account the specifics and technical capabilities of different equipment (lathes, milling, machining centers), which have different functionalities. In order to standardize the set of voice commands for working with different groups of machines, it is necessary to make universal commands.

While the classification of commands and the need for them require additional research in order to determine the groups of commands that would be necessary but not included in the group of G-codes [33], it should be noted that whatever they may be beside the groups, it is also necessary to restrict the set of commands depending on the mode in which the machine is currently set.

Intuitively, not all commands need to be available at different times. For example, a tool change command during surface machining should not be available for safety reasons. Therefore, this understanding needs to be formalized. The modes can be divided into the following.

- Initial calibration mode:

Calibration and initial setup could be significantly accelerated by duplicating the interface for voice input and making some scripts or sets of commands that would allow for a quick search of the desired settings and check the necessary parameters. The setup procedure is described quite fully, and the implementation of such commands should not cause great difficulties in formalizing and designing the necessary sequence.

- Machine setting mode:

A mode where most of the commands that exist in the machine are available, except for commands that are only initial setup commands.

- Program execution mode:

In this mode, only the necessary set of functional commands should be available, which affects the machine's behavior but is relatively safe (turning on / off the coolant, turning on / off the chip conveyor).

Obviously, an emergency stop command must be available in all modes. The classification of operating modes and their characteristics require additional studies to identify such conceptual model that would be most effective, but at the same time suitable for a wide range of machines and systems.

As mentioned above, combining different interfaces should be the desired option. One can use the presence of the display and duplicate on it separately the answers from the dialogue system, help information, and other necessary data. This approach can help duplicate the screen interface into a voice interface and help use the screen more quickly and efficiently (simplify navigation commands, typing text, and combining a set of commands for the machine into a single voice macro).

VI. Conclusion

The rapid development of Industry 4.0 with its Internet of Things and cloud computing is unlikely to eliminate human workplaces (at early stages at least). However, demand for convenient ways of interaction between humans and machines will definitely grow, and it is voice control that is likely to provide a more natural and simple way of interaction for humans and machines.

The difficulties of implementing computer voice control are associated with designing a system that would meet the technical requirements. The system has to work with sufficiently high accuracy of speech recognition, intent, and entities. Moreover, the system should be implemented cooperating with manufacturers of machines and control panels, which would support a standardized communication protocol. For the supported protocol to control the machine through the ContolBox, it is necessary to develop and improve the existing groups of commands, classify and implement different modes: initial calibration mode, machine setting mode, program execution mode. This approach should be considered because it could contribute to the efficiency and safety of using this type of machine and make interfaces convenient and straightforward.

The use of voice technologies in enterprises in general and CNC machines in particular, can lead to increased safety, reduced time costs, and a simplified interface. Computer voice control is a promising area of study. During the production and operation of CNC machines, enough knowledge and experience have been accumulated to apply them to create a new interface for interacting with the machine user. This opens up new possibilities that were unavailable as the area of natural language processing was less developed in the past.

This paper demonstrates the main principles and approaches to design multimodal systems with voice interface. Technical details and challenges to implement and connect with specific equipment lie out of the scope of this paper and presented in the future. However, it would be based on this work and will include:

- connection scheme
- way of interaction
- concept of complementary interface e.g. multimodal interface with voice and graphical interface
- different groups of commands
- different modes of work
- concept of implicit and explicit confirmation for different groups

Moreover, besides machine control, such voice interface could be an entry point to connect with a digital factory to get some reference information from a database or PLM-system.

References

[1] Web: https://www.vox.com/2017/1/23/14340966/voicelabs-report-alexa-google-assistant-echo-apps-discovery-problem.

[2] Web: https://www.geekwire.com/2020/amazon-alexa-leader-covid-19-sparked-huge-increase-use-voice-home/.

[3] M. Rüßmann, M. Lorenz, P. Gerbert, and M. Waldner, *Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries*. Boston: The Boston Consulting Group, 2015.

[4] O.I. Borisov, V.S. Gromov, S.A. Kolyubin, A.A. Pyrkin, A.A. Bobtsov, V.I. Salikhov, and I.V. Petranevsky, “Human-free robotic automation of industrial operations”, *IECON 2016—42nd Annual Conference of the IEEE Industrial Electronics Society*, 2016, pp. 6867-6872.

[5] O.I. Borisov, V.S. Gromov, S.A. Kolyubin, A.A. Pyrkin, N.Y. Dema, V.I. Salikhov, and A.A. Bobtsov, “Case study on human-free water heaters production for Industry 4.0”, *2018 IEEE Industrial Cyber-Physical Systems (ICPS)*, 2018., pp. 369-374.

[6] J. Nelles, S. Kuz, A. Mertens, and C.M. Schlick, “Human-centered design of assistance systems for production planning and control: the role of the human in Industry 4.0”, *2016 IEEE International Conference on Industrial Technology (ICIT)*, 2016, pp. 2099-2104.

[7] D. Romero, P. Bernus, O. Noran, J. Stahre, and Å. Fast-Berglund, “The Operator 4.0: human cyber-physical systems and adaptive automation towards human-automation symbiosis work systems”, *Advances in Production Management Systems. Initiatives for a Sustainable World*, 2016, pp. 677-686.

[8] E. Magrini, F. Ferraguti, A.J. Ronga, F. Pini, A. De Luca, and F. Leali, “Human-robot coexistence and interaction in open industrial cells”, *Robotics and Computer-Integrated Manufacturing*, 61, 101846, 2020, pp. 2-19.

[9] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, “Digital twin-driven product design, manufacturing and service with Big Data”, *The International Journal of Advanced Manufacturing*, 94, 2018, pp. 3563-3576.

[10] L. Damiani, M. Demartini, G. Guizzi, R. Revetria, and F. Tonelli, “Augmented and virtual reality applications in industrial systems: a qualitative review towards the Industry 4.0 era”, *IFAC-PapersOnLine*, 51(11), 2018, pp. 624-630.

[11] M. Lorenz, M. Spranger, T. Riedel, F. Pürzel, V. Wittstock, and P. Klimant, “CAD to VR—a methodology for the automated conversion of kinematic cad models to virtual reality”, *Procedia CIRP*, 41, 2016, pp. 358-363.

[12] J. Yan, Y. Meng, L. Lu, and L. Li, “Industrial Big Data in an Industry 4.0 environment: challenges, schemes, and applications for predictive maintenance”, *IEEE Access*, 5, 2017, pp. 23484-23491.

[13] I. Sittón and S. Rodríguez, “Pattern extraction for the design of predictive models in Industry 4.0”, *Trends in Cyber-Physical Multi-Agent Systems*. The PAAMS Collection—15th International Conference, PAAMS, 2017, pp. 258-261.

[14] Z. Li, Y. Wang, and K.-S. Wang, “Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario”, *Advances in Manufacturing*, 5(4), 2017, pp. 377-387.

[15] Web: https://www.emarketer.com/content/us-voice-assistant-users-2019.

[16] Web: https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.

[17] Web: https://developer.amazon.com/en-US/alexa/alexa-for-hospitality.

[18] T.B. Brown, B. Mann, N. Ryder, M. Subbiah et al., *Language Models are Few-Shot Learners. OpenAI*. 2020. Web: https://arxiv.org/pdf/2005.14165.pdf.

[19] J. Kildal, A. Tellaeche, I. Fernández, and I. Maurtua, “Potential users’ key concerns and expectations for the adoption of cobots”, *Procedia CIRP*, 72, 2018, pp. 21-26.

[20] O. Gomes and S. Pereira, “On the economic consequences of automation and robotics”, *Journal of Economic and Administrative Sciences*, 36, 2019, pp. 134-153.

[21] Web: https://www.forbes.com/sites/miguelhelft/2016/05/18/inside-sundar-pichais-plan-to-put-ai-everywhere/.

[22] Web: https://www.intercom.com/blog/principles-bot-design/.

[23] C. Pearl. *Designing Voice User Interfaces*. California: O’Reilly Media, 2016.

[24] ISO 6983:2009, Automation systems and integration—Numerical control of machines—Program format and definitions of address words.

[25] Web: https://assets.new.siemens.com/siemens/assets/api/uuid:d6d415fe-4790-4827-919b-597d5a4f6c0c/width:1125/quality:high/version:1537527978/sinumerik-808-1280x720-ohne-hg.png.

[26] J. Norberto Pires, “Robot-by-voice: experiments on commanding an industrial robot using the human voice”, *Industrial Robot: An International Journal*, 32(6), 2005, pp. 505-511.

[27] A. Rogowski, “Robotized cell remote control using voice commands in natural language”, *15th International Conference on Methods and Models in Automation and Robotics*, 2010, pp. 383-386.

[28] A. Rogowski, “Industrially oriented voice control system”, *Robotics and Computer-Integrated Manufacturing*, 28(3), 2012, pp. 303-315.

[29] A. Rogowski, “Web-based remote voice control of robotized cells”, *Robotics and Computer-Integrated Manufacturing*, 29(4), 2013, pp. 77-89.

[30] L. Jing, X. Ting, and S. Nanyan, “Research on the voice control and its audio signal processing in flexible manufacturing cell”, *Sensors and Transducers*, 25, 2013, pp. 91-97.

[31] M.Ya. Afanasev, Yu.V. Fedosov, Yu.S. Andreev, A.A. Krylova, S.A. Shorokhov, K.V. Zimenko, and M.V. Kolesnikov, “A concept for integration of voice assistant and modular cyber-physical production system”, *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, 2019, pp. 27-32.

[32] Web: https://en.wikipedia.org/wiki/Plug_and_play.

[33] ISO 14649:2003 (en) Industrial automation systems and integration—Physical device control—Data model for computerized numerical controllers.