EKIN: Towards Natural Language Interaction with Industrial Production Machines

EKIN: Hacia la Interacción en Lenguaje Natural con Máquinas de Producción Industrial

Arantza del Pozo[1], Laura García-Sardiña[1], Manex Serras[1], Ander González-Docasal[1], María Inés Torres[2], Eneko Ruiz[2], Izaskun Fernández[3], Cristina Aceta[3], Egoitz Konde[3], Daniel Aguinaga[4], Mikel de la Cruz[4], Iker Altuna[5], Joseba Agirre[5], Iker Etxebeste[6]
[1]Vicomtech Foundation, Basque Research and Technology Alliance (BRTA)
[2]University of the Basque Country UPV/EHU, Speech Interactive Research Group
[3]Tekniker Foundation, Basque Research and Technology Alliance (BRTA)
[4]Ikor Technology Center, [5]Machine Tool Institute, [6]UZEI
[1]{adelpozo, lgarcias, mserras, agonzalezd}@vicomtech.org, [2]{manes.torres, eneko.ruiz}@ehu.eus,
[3]{izaskun.fernandez, cristina.aceta, egoitz.konde}@tekniker.es, [4]{daguinaga, mcruz}@ikor.es,
[5]{altuna, agirre}@imh.eus, [6]{ietxebeste}@uzei.eus

Abstract: The industry and manufacturing sector could greatly benefit from 'hands-free' voice-based natural language interactions to assist operators across tasks requiring manual operations. However, the complexity of the industrial domain makes it very expensive to develop dialogue systems in this field. Also, the dominant cloud architectures for speech recognition and synthesis pose privacy, security and latency concerns. And for some languages with few resources such as Basque, there is a lack of formalised terminology and language resources for technology development. In this paper, we review the state of the art in this field and describe EKIN, a project which is being carried out to address some of the identified problems.
Keywords: Industry 4.0, Human-Machine Interaction, Basque.

Resumen: El sector de la fabricación industrial podría beneficiarse enormemente de las interacciones "manos libres" por voz, para ayudar a los operarios en tareas que requieren operaciones manuales. Sin embargo, la complejidad del dominio industrial hace que sea muy costoso desarrollar sistemas de diálogo en este campo. Además, las arquitecturas en la nube dominantes para el reconocimiento y la síntesis de voz plantean problemas de privacidad, seguridad y latencia. Y para algunas lenguas con pocos recursos como el euskera, se carece de terminología formalizada y de recursos lingüísticos para abordar los desarrollos tecnológicos necesarios. En este artículo, revisamos el estado del arte en este campo y describimos EKIN, un proyecto que se está llevando a cabo para abordar algunos de los problemas identificados.
Palabras clave: Industria 4.0, Interacción Persona-Máquina, euskera.

1 Project consortium and funding body

EKIN is a research project funded by the Basque Government through the Elkartek 2020 program from the Basque Agency for Business Development SPRI, under grant agreement KK-2020/00055.

The project has a total duration of 22 months, beginning on March 1, 2020 and ending on December 31, 2021.

EKIN is being carried out by the following consortium: Vicomtech[1], the Speech Interactive Research Group of the University of the Basque Country[2], Tekniker[3], Ikor Technology Center[4], the Machine Tool Institute[5] and UZEI[6].

[1]https://www.vicomtech.org
[2]https://www.ehu.eus/en/web/speech-interactive/about-us
[3]https://www.tekniker.es
[4]https://ikor.es
[5]https://www.imh.eus
[6]https://uzei.eus

2 Context and motivation

The use of voice and natural language is changing the way we relate to technology. As a result, conversational assistants have become one of the most innovative tools to simplify and make human-machine interactions more natural. Well-known examples of these interfaces are Apple's Siri, Google Now, Microsoft Cortana or Amazon Alexa.

Although these types of devices can work perfectly independently (e.g. to search the web, find songs on Spotify or report the weather forecast), they are increasingly integrating with the IoT of the home by enabling interactions with locks, light switches, heaters, air conditioners and/or kitchen appliances, becoming more and more indispensable devices.

Similarly, voice assistants could act as a central management element in IoT-enabled Industry 4.0 manufacturing plants. Some potential use cases in industrial manufacturing are: providing support in machine maintenance, repair and overhaul operations; facilitating the programming of manufacturing machines; or supporting with manufacturing and assembly tasks, among others. The operator is continuously involved in manual operations during these tasks, and searching through paper manuals or tablets for assistance slows down and considerably hinders their work.

Voice-based interfaces represent a relevant solution in this context. The obvious but important benefits of using voice to communicate with systems and machines in factories are as follows: (i) they are 'hands free' and 'eyes free', allowing operators to continue with physical tasks; (ii) they are natural for operators, requiring minimal training; and (iii) they are very flexible, allowing communication at different levels of detail and in contexts linked to multiple tasks.

Regarding industrial noise concerns for spoken interaction in manufacturing settings, recent studies have shown that the combination of existing hardware noise cancellation devices and speech recognition systems is robust enough for use in manufacturing environments (Gaizauskas, 2019).

Therefore, in theory, the development of voice interaction technology should allow industrial work to be carried out more effectively, as well as obtaining a positive response from the operators. However, its presence in industrial manufacturing environments is still rare because there are several challenges to overcome:

- Although much information is available in documents of a technical nature (e.g. manuals, manufacturing and assembly dossiers, maintenance notes), it is still very expensive to provide current dialogue systems with the knowledge necessary to implement meaningful manufacturing use cases and tasks. This is due to the specificity and complexity of the domain, compared to other more well-known areas of application such as the reservation of transport tickets, restaurants or hotels.
- The dominant cloud architectures for deploying speech recognition and synthesis technology, derived from the hardware requirements of neural paradigms, pose privacy, security and latency issues that concern the industry.
- In the particular case of the Basque industry, there is a tradition of oral communication in Basque in some factories that is not formalized. There is practically no specific terminology for the sector and the most common oral expressions used to interact with machines are not documented.

3 Technologies involved

3.1 Natural interaction in advanced manufacturing environments

Human-machine interfaces (HMI) in the industrial field have evolved rapidly in recent years with the development of new mobile technologies and new devices such as smartphones, tablets and/or augmented reality glasses. In the last decade, a considerable number of systems have been developed, mainly in the field of collaborative robotics, with the capacity for natural interaction between operators and machines, to varying degrees (Mavridis, 2015), (Serras et al., 2020). Despite these platforms integrate advanced interfaces, their ability to semantically understand human requests is still quite limited and the effort required for their implementation is very high in most cases.

Recent research that includes semantic technologies to improve industrial human-

machine interaction is mainly rule-based, which in terms of maintenance and/or extension requires high manual labor (Maurtua et al., 2017). There is little work focused on machine learning techniques for multimodal human-machine communication, on improving its adaptability to new scenarios (or even languages) or on improving its performance using as few resources (both linguistic or human) as possible. In line with reducing efforts, the authors in (Antonelli and Bruno, 2017) emphasize the role of ontologies, since they allow defining the domain understandable by humans and machines and contribute to reducing ambiguity across operators.

On the other hand, question-answering systems that allow obtaining information from a collection of unstructured documents have advanced considerably in recent years and are beginning to evolve towards conversational interfaces (Reddy, Chen, and Manning, 2019). Their adaptation to use cases in the industrial production domain would allow to automatically exploit the information contained in existing technical documents.

3.2 Embedded speech recognition and synthesis

With the introduction of deep neural architectures, the last few years have witnessed a significant leap in the performance of speech recognition and synthesis systems. However, until now, the high memory, processing and power consumption requirements of neural models and the computational and battery limitations of mobile devices have led to cloud integrations. In the industrial sector, the deployment of Wi-Fi facilities raises a series of difficulties linked to the extension of the facilities and the dispersion in different buildings. Also, there is a growing concern regarding privacy and security issues posed by the current cloud architectures of commercial systems.

On the other hand, important advances have been made in the development of specific hardware systems for embedded computing of neural models in recent years. Thanks to the advances in hardware obtained, compression techniques of neural models have also gained attention, obtaining important advances also at software level. Some of the techniques used to reduce the size of the models and the latency of the responses are precision reduction, data paralelization and model compression (He et al., 2019). The union between low-cost hardware and great computational power, together with the advances in neural model optimization, make it possible to explore the embedding of speech recognition and synthesis technology for industrial human-machine interaction applications.

3.3 Use of Basque in the industrial sector

The use of Basque in the industrial manufacturing sector is very scarce. Although Basque was used informally in the factories of some regions where the tradition of the machine tool industry is very powerful, its incorporation into the real industrial world has been closely linked to the plans for its promotion carried out by the companies in the sector. As a result, the social use of Basque in industry has increased, but there are still steps to take towards incorporating it into the actual manufacturing processes. In practice, both industrial machinery software and technical documentation are still made in the languages chosen by the machine suppliers, which usually do not formally include Basque.

Regarding availability of specialized language resources, some terminological dictionaries of the field exist such as the Numerical Control Dictionary and the Machine Tool Dictionary contained in the Euskalterm database[7], the LANEKI dictionary of vocational training[8], the Dictionary of the New Industry promoted by SPRI[9] or the DANOBAT dictionary[10]. Nevertheless, the specialized language resources available in Basque are too limited to be used for the development of natural language interaction interfaces with industrial machines in such language.

4 Project objectives and expected results

The EKIN project aims to advance the development of 'conversational interfaces' as a mechanism for interaction between operators and machines in industrial production plants in the Basque Country, with the aim of facilitating and improving the productivity of certain processes. Specifically, the following objectives are pursued:

[7] https://www.euskadi.eus/euskalterm/
[8] http://hiztegia.jakinbai.eus/
[9] https://www.spri.eus/hiztegia/
[10] https://hiztegia.danobatgroup.eus/

- To facilitate the development of natural language interaction interfaces between operators and industrial production machines, based on the information contained in technical documentation
- To optimize neural speech recognition and synthesis models, so that they can be embedded in electronic devices in the machines themselves to avoid the privacy, security and latency problems of their cloud deployments
- To formalize a terminology and a corpus of expressions of interaction with machines in Basque, that will serve as a reference for the development of conversational interfaces in the industrial sector in such language

The main expected result of the project is the explicit recognition by real operators that the use of natural language voice interfaces facilitates the maintenance, programming, manufacturing and assembly tasks of industrial production machines, increasing their productivity and satisfaction regarding the tasks performed. The technological stack under development will derive in specific interfaces for different use cases. As more specific results, we also expect to generate:

- Technological components that allow implementing operator-machine interaction systems faster and more efficiently than at present
- Speech recognition and synthesis models that can be embedded in low-performance hardware devices
- A terminology and corpus of reference expressions for operator-machine interaction in Basque

The first results of the project have been satisfactory. Considerable progress has been made in compiling a corpus of technical manuals and dossiers. Progress has also been made designing an ontology dealing with dialogue including domain aspects and investigating novel techniques for the development of question-answering systems, as well as systems for the automatic generation of dialogue acts and rules from technical documentation. Regarding the embedding of speech recognition and synthesis, an experimentation board has been designed and several neural model optimization frameworks have been explored. Finally, a significant effort has also been made to compile manuals and dossiers in Basque among project partners, from existing repositories in the field of vocational training and by contacting relevant stakeholders in the Basque industrial sector.

During the last phase of the project, efforts will be focused on finalising technological developments, implementing a prototype and having real operators evaluate it.

References

Antonelli, D. and G. Bruno. 2017. Human-Robot Collaboration using Industrial Robots. In *2017 2nd International Conference on Electrical, Automation and Mechanical Engineering (EAME 2017)*, pages 99–102. Atlantis Press.

Gaizauskas, R. 2019. Investigating spoken dialogue to support manufacturing processes. Technical report, The University of Sheffield, UK, April.

He, Y., T. N. Sainath, R. Prabhavalkar, I. McGraw, R. Alvarez, D. Zhao, D. Rybach, A. Kannan, Y. Wu, R. Pang, Q. Liang, D. Bhatia, Y. Shangguan, B. Li, G. Pundak, K. C. Sim, T. Bagby, S.-y. Chang, K. Rao, and A. Gruenstein. 2019. Streaming end-to-end speech recognition for mobile devices. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6381–6385.

Maurtua, I., I. Fernandez, A. Tellaeche, J. Kildal, J. Ibarguren, and B. Sierra. 2017. Natural multimodal communication for human-robot collaboration. *International Journal of Advanced Robotic Systems*, pages 1–12.

Mavridis, N. 2015. A review of verbal and non-verbal human-robot interactive communication. *Robotics and Autonomous Systems*, 63:22–35.

Reddy, S., D. Chen, and C. D. Manning. 2019. CoQA: A conversational question answering challenge. *Transactions of the Association for Computational Linguistics*, 7:249–266, March.

Serras, M., L. García-Sardiña, B. Simões, H. Álvarez, and J. Arambarri. 2020. Dialogue Enhanced Extended Reality: Interactive System for the Operator 4.0. *Applied Sciences*, 10(11):3960.