Procesamiento del Lenguaje Natural, Revista nº 65, septiembre de 2020, pp. 135-138 recibido 29-04-2020 revisado 26-05-2020 aceptado 07-06-2020

AREVA: Augmented Reality Voice Assistant for Industrial Maintenance

AREVA: Asistente por Voz Dotado de Realidad Aumentada para el Mantenimiento Industrial

Manex Serras[1], **Laura García-Sardiña**[1], **Bruno Simões**[1], **Hugo Álvarez** [1], **Jon Arambarri** [2]
[1]Vicomtech Foundation, Basque Research and Technology Alliance (BRTA)
[2]VirtualWare Labs Foundation
{mserras, lgarcias, bsimoes, halvarez}@vicomtech.org, jarambarri@virtuawareco.com

Abstract: Within the context of Industry 4.0, AREVA is presented: a Voice Assistant with Augmented Reality visualisations for the support and guidance of operators when carrying out tasks and processes in industrial environments. With the aim of validating its use for the training of new operators, first evaluations were performed by a group of non-expert users who were asked to carry out a maintenance task on a Universal Robot.
Keywords: Spoken Dialogue System, Augmented Reality, Industry 4.0

Resumen: Dentro del marco de la Industria 4.0 se presenta AREVA: un Asistente por Voz dotado con visualizaciones de Realidad Aumentada para el apoyo y guiado del operario a la hora de realizar tareas y procesos en entornos industriales. Con el objetivo de validar su uso para el entrenamiento y capacitación de nuevos operarios, se ha realizado una primera evaluación con un grupo de usuarios no-expertos a quienes se les pidió que realizasen una tarea de mantenimiento sobre un Robot Universal.
Palabras clave: Sistema de Diálogo Hablado, Realidad Aumentada, Industria 4.0

1 *Introduction*

Industry 4.0 is one of the main paradigms of the latest years, where the industrial factories are augmented with wireless connectivity, sensors, and AI mechanisms that can help with different processes and tasks. Industrial 4.0 revolution has shown how technology can alter the way work is performed, the structure of organisations, and the role that workers play in the manufacturing process (Simões et al., 2019; Posada et al., 2018; Serras et al., 2020).

This paper describes an Augmented REality Voice Assistant (AREVA) to assist operators during the maintenance of Universal Robot's (UR) Grippers. To this end, the system guides the operator combining voice interaction and real-time 3D visualisations. Augmented Reality (AR) glasses with a built-in microphone are used to provide hands-free interaction and as a solution to overcome practical limitations in tasks that rely on intense manual work.

This paper presents the AREVA system for industrial maintenance tasks. The presented prototype system was tested by a set of 10 participants (Gender: 5 male, 4 female, 1 would rather not say; Age ranges: 50% 25-29, 30% 35-39, 10% 30-34, and 10% 18-24) who had little-to-none technical expertise on the field and no prior knowledge of similar maintenance tasks, using the presented assistant to support and guide them through the process. After the experimental sessions, participants filled an assessment form to judge the system's validity.

Section 2 describes the architecture of the system and the use case, Section 3 details the experimental framework and presents the evaluation results. Finally 4 presents the conclusions and sets the future work.

2 *System Architecture*

The system architecture follows the classical SDS schema, but incorporating communication channels with the necessary hard-

ISSN 1135-5948. DOI 10.26342/2020-65-21

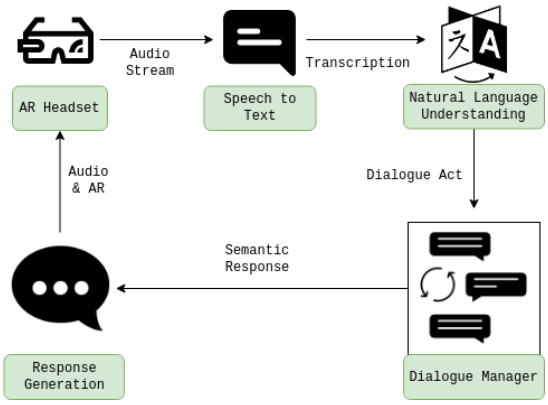

Figure 1: Architecture of the Augmented Reality enhanced Spoken Dialogue System

ware devices to allow for both visual and aural communication with operators, as shown in Figure 1. The core components are presented in more detail below.

- **AR Headset**: this device is responsible of capturing the users' speech, play AREVA's responses, and render AR animations. The selection of a head-mounted device is motivated so that operators have theirs hands free to perform the task while interacting with the system. For the proposed system, Microsoft HoloLens were used.
- **Speech to Text (STT)**: is the component in charge of converting the audio streams into texts. First, an energy-based automata is used as Voice Activity Detector (VAD) to segment audio streams into speech chunks, which are then sent to a speech transcription service to get the text utterance. Transcriptions were provided by the Google Speech API.
- **Natural Language Understanding (NLU)**: once the voice segments have been decoded into text, the NLU extracts the communicative intention of the transcribed text so it can be interpretable by the system. For the AREVA system, the grammar-based Phoenix parser was used (Ward, 1991).
- **Dialogue Manager (DM)**: the DM is the module that ensures the correct planning and interaction through the industrial process. The DM implemented for AREVA consists of an Attributed Probabilistic Finite State Bi-Automata (Torres, 2013; Serras, Torres, and Del Pozo, 2019) which uses a stack of expert-rules and simulated interactions to guide the operator through the task.
- **Response Generation**: this module receives the response selected by the DM to carry out with the interaction. It performs two actions: 1) it selects an adequate text template for the response which is then synthesised by a Text To Speech service (Hernaez et al., 2001), and 2) it checks if the action has an associated AR rendering. Then, these audio and visual responses are sent to the AR Headset to be conveyed to the operator.

2.1 Use Case

We considered as a use case the maintenance of the Universal Robot's gripper. The objective of the system is to assist untrained operators in real time with contextualised responses to the questions that may arise during the course of the task. AR is used to overlay animations that further enhance the experience.

The maintenance steps considered for the use case were retrieved from Subsections 7.1 and 7.3 of the official maintenance manual[1], which require the following actions:

1. Shutdown the UR, unplug the connected wires, and dismount the gripper.
2. Open the gripper to clean it and inspect it both visually and by moving its fingers to detect any wear, damages or incorrect joint articulations.
3. Close the gripper's fingers, mount it back to the UR, and attach the wires.

Once these four stages are completed, the monthly maintenance of the UR's gripper is finished. As it can be implied by the steps to perform the task, a heavy manual effort is required to perform the operation, so a hands-free system is critical to improve the User Experience (UX).

For the presented use case, the developed NLU module can understand a total of 10

[1] https://assets.robotiq.com/website-assets/support_documents/document/3-Finger_PDF_20190322.pdf?_ga=2.105997637.750885948.1563187207-1298495031.1563187207, last accessed 29/04/2020

Augmented Reality & Spoken Dialogue System	
Visualisations aside, how helpful has the voice interaction been in performing maintenance?	Mean score: 4,3/5
Voice interaction aside, how useful have the visualisations been to you?	Mean score: 3,7/5
Do you think you could have done the maintenance properly without the visualisations, just with the voice interaction?	Yes: 1, Yes, but it would have taken longer: 4, No, AR was a key feature: 4, I don't know: 1
Do you think you could have done the maintenance properly without the voice interaction, just with the visualisations?	Yes, but it would have taken longer: 4 No, voice interaction was a key feature: 6
If you had to say which technology was more relevant, which would it be?	AR: 1, SDS: 3, Both: 6
Do you think it was useful to combine both technologies?	Yes: 10, No: 0
Usability and perception	
Has the system made maintenance easier?	Mean score: 4/5
Have you felt frustrated at any point in the process?	Never: 1, Almost never: 4, Sometimes: 5
"The system has increased my confidence that I was performing the task correctly"	Mean agreement score: 3,8/5
How much has having your hands free made maintenance easier?	Mean score: 4,5/5
Task Completion and Recommendation	
Have you been able to complete the maintenance task?	Yes: 9, No: 1
Would you recommend using the assistant to someone who has never done the maintenance before?	Yes: 10, No: 0
Would you use the assistant to do maintenance a second time, the following month?	Yes: 8, No: 2

Table 1: Summary of the evaluation form filled by the participants

different intents, 20 entities, and more than 20 entity-values in total, using more than 2000 grammar rules to parse the operators' input into valid semantic actions.

The initial DM version was trained using 33 expert-rules. These DM rules map the dialogue state –which encodes the latest system's response, user action, and discrete memory attributes– with the possible actions to perform in the task, with physical objects (e.g. tools, wires, bolts), and with their properties, so that questions like "Which tool do I need?" and "Where are the bolts?" can be disambiguated according to the current step context. These rules also control the dialogue navigation aspects such as the welcoming prompt, repetitions, flow constraints (do not allow the operator to go to the next step if the current one is not satisfied), and so on.

The Response Generation had 48 possible response actions, amounting to 131 text templates (an action may have more than one associated template). A total of 8 possible AR animations were developed: 1) initial gripper position over the UR, 2) point shutdown button, 3) point the wires location, 4) point the bolts in the base of the gripper, 5) point the bolts under the gripper's fingers, 6) gripper's fingers opening, 7) gripper's fingers movement to check, and finally 8) gripper's fingers closing.

3 *Experimental Framework*

The objective of AREVA is to help and guide operators during maintenance tasks and make them more independent. Ten participants were recruited to validate the system and asked to perform a maintenance task in which they had no previous experience.

Figure 2: An operator using the AREVA system to carry out the gripper's maintenance

The robot arm was fixed on a pre-set position, with the gripper coupled and its fingers closed, as shown in Figure 2. The tools needed for the task were placed on the table. At the beginning of the session, while the participants were adjusting and getting familiarised with the AR Glasses, a researcher would present the UR, the system, and the task at hand that they should complete with no additional details.

The evaluation form had several questions regarding the technological components of the system, which were evaluated both separately and in conjunction. Besides,

additional questions about usability, participants' perception, and task completion were included. The results to these questions are summarised in Table 2.

The results reported in Table 2 show that the use of AR and SDS technologies combined in AREVA is useful for training and assisting applications for industrial processes.

The Spoken Dialogue System is perceived more positively by the participants, both in helpfulness and as a key feature, but, overall, the combination of both technologies is perceived as advantageous and both of them are relevant to assist and guide the operator through the gripper's maintenance task. Regarding the scores achieved in the system's usability and perception evaluation, all participants agreed that the system made the maintenance easier, especially by allowing them to have both hands free, and made them feel more confident that they were doing it correctly. Still, there were times when some users felt frustrated using the system, mostly due to STT or Voice Activity Detection errors.

Finally, 9 out of 10 users were able to complete the maintenance –note that no participant had made this process before and their knowledge of industrial processes was limited–. All of them would recommend the system to someone who has never done this task before, proving the usefulness of the system when training new operators. Moreover, 8 of them would use AREVA again for the following month's maintenance.

Additional observations made by the participants set the focus on improving the hardware ergonomics, as sometimes it would bother the participants when performing the maintenance. Also, the capability of adapting to each user profile (beginners/intermediate/advanced) to adjust the response verbosity was deemed as a useful feature.

4 Conclusions and Future Work

This paper describes an Augmented Reality Voice Assistant for natural and hands-free guiding for a Universal Robot Gripper's maintenance. The proposed solution facilitates the capture, distribution, and communication of domain specific knowledge to operators in training and production phases. The hands-free, technology-combining multimodal system was well accepted by a group of non-expert operators unfamiliar to the task. As future work, a wider sample of participants will participate as testers. In addition, UX factors will be further improved and evaluated, such as improvements in the STT and NLU modules and the use of Wake-up words to avoid ambient-noise activation of the VAD module. Also, the extrapolation to other industrial processes will be further investigated.

References

Hernaez, I., E. Navas, J. L. Murugarren, and B. Etxebarria. 2001. Description of the ahotts system for the basque language. In *4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis*.

Posada, J., M. Zorrilla, A. Dominguez, B. Simoes, P. Eisert, D. Stricker, J. Rambach, J. Döllner, and M. Guevara. 2018. Graphics and media technologies for operators in industry 4.0. *IEEE computer graphics and applications*, 38(5):119–132.

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.

Serras, M., M. I. Torres, and A. Del Pozo. 2019. User-aware dialogue management policies over attributed bi-automata. *Pattern Analysis and Applications*, 22(4):1319–1330.

Simões, B., R. De Amicis, I. Barandiaran, and J. Posada. 2019. Cross reality to enhance worker cognition in industrial assembly operations. *The International Journal of Advanced Manufacturing Technology*, pages 1–14.

Torres, M. I. 2013. Stochastic bi-languages to model dialogs. In *Proceedings of the 11th international conference on finite state methods and natural language processing*, pages 9–17.

Ward, W. 1991. Understanding spontaneous speech: The phoenix system. In *[Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing*, pages 365–367. IEEE.