Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 52 (2020) 4–8

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

5th International Conference on System-Integrated Intelligence

Interacting with a Digital Twin using Amazon Alexa

Stefan Wellsandt[a,*], Mina Foosherian[a], Klaus-Dieter Thoben[a,b]

[a]*BIBA - Bremer Institut für Produktion und Logistik GmbH, Hochschulring 20, 28359 Bremen, Germany*
[b]*University of Bremen, Faculty of Production Engineering, Badgasteiner Straße 1, 28359 Bremen, Germany*

* Corresponding author. Tel.: +49-0421-218-50166; fax: +49-421-218-50007. *E-mail address:* wel@biba.uni-bremen.de

Abstract

The Digital Twin is an evolving concept with many facets and applications in, for instance, engineering simulation, system control, and product-centric information management. This article focuses on the latter where literature uses the Product Avatar concept to refer to a product's digital counterpart. Such an avatar used to have one or more graphical interfaces to support user interactions with information about a product item. Over the last few years, voice user interfaces became more mature, and companies, such as Amazon and Google, used them to create digital assistants that support their users during tasks or by taking them over directly. This paper focuses on the hypothesis that a company could use a voice-enabled digital assistant to interact with item-level information. Our study used product tracking and tracing, and quality control in the production as a realistic application case. The design of the assistant bases on the information needs outlined in the Electronic Product Code Information Services (EPCIS) standard. We implemented this design in a small-scale demonstrator on an Echo Show 5 smart speaker with an integrated touch display and an embedded Amazon Alexa assistant. This paper concludes that significant technological barriers, such as low transcription accuracy for object identifier information and the handling of factory noise, remain. A significant non-technological barrier is the mistrust regarding the closed voice assistant technologies from companies, such as Amazon and Google. An approach to address the latter barrier is to use open technologies, such as the privacy-focused assistant Mycroft or Mozilla's transcription solution DeepSpeech. Further research and experiments with these technologies are useful to identify how they can support the interaction with Digital Twins.

Keywords: Digital Twin; Digital Assistant; Voice User Interface; Tracking and Tracing; Product Avatar.

1. Introduction

Miniaturization of microelectronics and lower deployment costs allow companies to use intelligent products along their supply chain [1]. An **intelligent product** possesses a unique computer-readable identifier, monitors its status and environment, stores data about itself, shares and receives information, and may make decisions. Its intelligence comes from an embedded computer or a remote computer with network access. When intelligent products move through the supply chain, business processes transform them into more valuable and often more complex products. During a business process, the product can incorporate additional information needed by its stakeholders. Stakeholders include the workers in the supply chain, product owners and users, and the staff that performs product-related services, such as maintenance, repair, and overhaul. The accumulated product information is the basis for the physical product's counterpart in the digital world. The so-called **Digital Twin** (DT) is an evolving concept to describe the digital representation of a product. Several research groups published articles about the application of a DT [2] Applications include the simulation of material stress [3], human-robot collaboration [4], and the stakeholder-specific provision of item-level product information [5]. The latter is the focus of this paper and typically refers to products as unique items.

10.1016/j.promfg.2020.11.002

The human interaction with a DT usually relies on one or more **graphical user interfaces** (GUI), such as simulation environments, three-dimensional virtual spaces, information dashboards, and faceted analytical displays [6].

GUIs support the human in creating, reading, updating, and deleting product information, and they convey more information at once compared to other digital media. They also require that information users learn how to use it. This expectation becomes a problem if a GUI changes regularly due to software improvements, and the employees have to relearn using it. A DT might also rely on multiple information systems, each one using a custom GUI. If workers must adapt to various visual interfaces continuously, they may lose confidence in working with a DT. The learning – especially professional courses – may also result in recurring costs for worker training.

Voice user interfaces (VUI) can provide a fast and intuitive method for human-computer interaction. An effective VUI allows its user to request information through natural language and without learning a specific query syntax. For a few years, consumers experience VUIs through digital assistant (DA) technologies, such as Siri, Alexa, or Google Assistant.

The **hypothesis** of this paper is that companies could use consumer-focused assistant technologies to interact with item-level information. Such an assistant would act as the digital representation of an item and could answer the user's questions about it. The **goal** of this paper is to test the hypothesis above.

The structure of this paper is as follows. Section 2 introduces the background of DAs, and the DT focused in this paper. Section 3 covers the application scenario used in the study and explains the relevant item-level information. Section 4 presents the design and implementation of the DA. Section 5 discusses the study results and concludes the hypothesis.

2. Background

The first part of this section introduces what DA means in this paper and how it relates to other research. The second part clarifies how this paper understands the DT concept and why we focus on tracking and tracing and quality control.

2.1. Digital assistant

An **assistant** supports or entirely takes over activities that are time-consuming, stressful, or otherwise not desirable for the client. A DA does the same for its user(s) through software and belongs to the group of so-called conversational agents. A **conversational agent** is an application that accepts user input in the form of voice or text and provides responses in natural language [7].

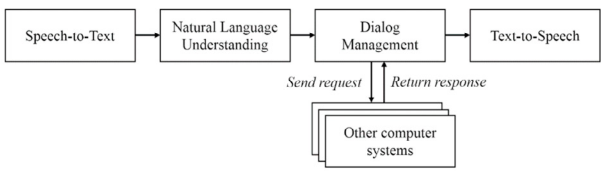

Figure 1: Voice-enabled digital assistant components

These agents include VUI agents, text-based agents (Chatbots), and embodied conversational agents [8]. This study focuses on the VUI agent that we call a voice-enabled DA to clarify that it is more than an interface. Figure 1 illustrates the components of such an assistant.

The *speech-to-text* (STT) component transcribes an audio input into a text output. *Natural language understanding* (NLU) extracts semantic elements from this text [9], typically through artificial intelligence. Important semantic elements are the user's intent and the related slots. *Intents* refer to what the user wants to do in a request. It can have pieces of information embedded that are necessary to know what or with what a user wants to perform an action. These information pieces are called *entities* [10]. The *dialog management* (DM) component contains the software that decides how to respond to the semantic elements. During the processing, the DM can interact with other computer systems to perform actions that are too complex for the DM. The *text-to-speech* (TTS) turns the generated text into a computer-generated voice.

The technological research behind voice-enabled DA's reaches back to the first interactive voice response (IVR) systems in the 1970s – some principles of dialog design are still valid since then [11]. Companies such as Amazon, Google, and Microsoft have opened their platforms for developers to create voice applications [10]. Considering that a *Skill* is a series of intents that a voice application can handle, by now, the underlying technology is so mature that even children can build simple skills [12]. Developing a skill with little programming knowledge is possible through web consoles such as Amazon Developer Console and Google Dialogflow. These web tools let the developer create new skills, and add intents with sample utterances and slots to it, using the components of a GUI.

The transfer of digital assistant technology and tools from the consumer domain to industrial applications is a new topic. There is almost no academic literature in this field despite challenges, such as:

- **Unclear benefits** of voice-enabled assistants in industrial tasks (e.g., product and process design, manufacturing, and maintenance and repair).
- **Uncertain personal data protection** (e.g., if service providers are large enough to ignore regulations).
- **Intransparent business knowledge protection** (e.g., if terms allow the service provider to process client data).
- **Unclear data security risks** (e.g., eavesdropping conversations and injecting false information).
- **Robustness under industrial conditions** (e.g., audio noise, time pressure, budget constraints, incomplete and inaccurate data).
- **Acceptance of assistants among employees** from different professions, education, and cultural backgrounds.
- **Accountability** when stakeholders associate an assistant's response with injuries, death, or damage to operations.
- **Lack of human-machine conversation experience for industrial applications** (e.g., how do employees talk to production processes or manufacturing machines).

Underlying technologies, such as artificial intelligence and cloud services, share many of the challenges above. Addressing the challenges in other areas will influence the applicability of digital assistants in the industry. The operationalization of the Ethics Guidelines for Trustworthy AI [13] is an example of recent activities in this domain.

2.2. Product Avatar and item-level information

Before the DT concept became popular, researchers developed similar ideas such as the **Product Avatar** (PA). Its vision is that of a "hybrid world" where a physical product has a digital counterpart [14]. This counterpart incorporates product-related information that resides in heterogeneous data sources and grows as the product progresses through its life cycle – i.e., from its design and production to its recycling or disposal. Humans access this information through the PA's user interfaces. Furthermore, they can use the avatar to perform actions on their behalf, which makes the PA a software agent.

An important aspect of the PA is that it adopts a **product-centric** perspective on information. This viewpoint means that the junction between the physical product and its digital counterpart is the unique identifier that allows computers to distinguish one item in the real world from another. This identifier connects the PA research to technologies, such as Radio Frequency Identification (RFID), and the concept of product tracking and tracing. The latter has a clear application domain with clear information needs, as described in the Electronic Product Code Information Services (EPCIS) standard [15].

EPCIS is a globally applied standard that aims to provide visibility for business processes, prescribing a common data model. The standard uses a Core Business Vocabulary (CBV) to fill the business context in this data model. Considering that a business process has a sequence of business steps, EPCIS data provides a detailed record of these steps. Information systems that use EPCIS create so-called **visibility events** that contain information to answer what-, where-, when-, and why-questions regarding a unique object identifier.

Future revisions of EPCIS will include information from the Internet of Things. This information, typically created during measurement processes, provides more details about items. It can support, for instance, the quality control in a factory by indicating when the production parameters exceeded quality-related thresholds.

3. Application case

This study uses an application case, which creates a realistic setting for a PA to test the hypothesis above. The **targeted product** is a bathroom cabin, which is an assembly of several components with a base structure made of composite material. Its production begins with the base structure, and several manufacturing and handling steps follow. Production plans for this product change frequently due to order changes. The ambient temperature and humidity inside the factory and the storage areas outside affect the product quality. The customer of the producer mounts the product into a module of a pre-fabricated building.

Tracking and tracing allow the producer to adjust the production plan faster. Measurements of the ambient environment can support the analysis of product quality issues (e.g., cracks in the composite structure). The technical infrastructure for tracking and measurements uses open-source software tools (e.g., Grafana, MongoDB, and InfluxDB). It adopts parts of the EPCIS standard to describe business process steps and visibility events. Desai et al. describe the technical concept for this application case in detail [16].

Figure 2 illustrates the relevant business steps, visibility events, and infrastructures of the application case. We created a **small-sized demonstrator** based on this case to experiment with item-level information in a realistic setup. The demonstrator focuses on Company 1 (producer) and the component installation, assembly area, and cold storage. These locations have sensors to detect air temperature and humidity.

Using a small-sized demonstrator is useful in this case because it provides us more control over the creation of visibility events and the measurements that build the basis for a PA. This way, we can iteratively design solutions to interact with the digital counterpart of an item without interfering with real production processes.

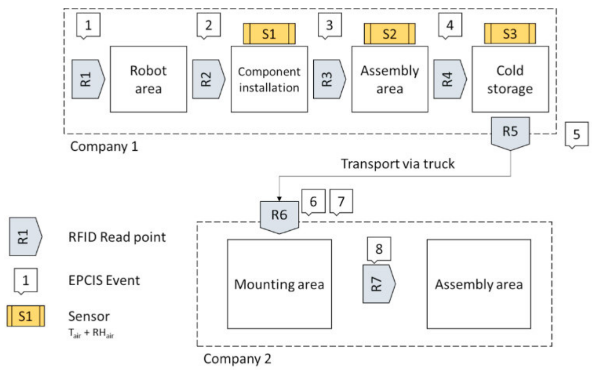

Figure 2: Application case [16]

4. Design and implementation

The first part of this section outlines the technical design of a DA that allows humans to interact with item-level product information. It bases on Amazon Alexa and uses the typical terminology used in its software documentation. The second part describes our implementation in a demonstration scenario that bases on the application case above.

4.1. Design

The first step of the design process was to create the interaction model. Alexa uses this model to interpret and translate the user utterances into a request [17]. The hand-crafted interaction model is a JSON file that contains, for instance, the sample utterances to activate the intents, and the slot names and types. We designed the intents based on the W-questions described in the EPCIS standard and the ambient temperature and humidity monitoring. Table 1 shows examples for user utterances, the related intents and slots, and the response with the values that answer the question.

Table 1. Example utterances, intents, slots, and responses of the assistant

Utterances	Intents	Slots	Responses
Tell me about the products in the robot area!	What	Business location	I can find LB1288 and LB5607 in the robot area.
Where is LB1288?	Where	Product identifier	Product LB1288 is currently in the robot area.
What is happening to LB1288?	Why	Product identifier	Product LB1288 is currently under installation.
Tell me all about the product LB1288	Product History	Product identifier	Product #LB1288is in cold storage before it was in the assembly area, and before that, <...>.
How cold is it in the storage area	Sensor Data	Business location	The temperature in the cold storage is 22 degrees.

4.2. Implementation

Amazon lets developers create custom skills for their voice assistant with the Alexa Skills Kit (ASK). We developed a custom skill to handle questions about items in a production setting. As an interaction device, we used the *Echo 5*, which is a smart speaker with an integrated touch display. The screen is useful because multi-dimensional information, such as graphs are challenging to describe with voice only.

The so-called *fulfillment code* runs in AWS Lambda, which is a cloud service from Amazon Web Services. The fulfillment code processes a user's request in voice applications [10]. When a user interacts with an Alexa, the fulfillment receives a request that contains the parameters to fulfill the intent and create a response. To build a response, we used API calls to get event and sensor data from other systems (refer to Section 3).

5. Conclusion

The implemented Alexa-based DA allows its users to interact with a product's digital counterpart. This proof-of-concept demonstrates that consumer-focused assistant technologies can act as an interface to item-level information. The hypothesis is, therefore, correct under the constraints of the lab demonstrator we used.

The assistant understands differently phrased questions related to EPCIS and helps the user to provide the right slots. The assistant remembers the information provided earlier, which makes the conversation more natural and faster.

Our prototype identified several barriers to implement a DA in an industrial context. The main **technological barrier** is the *low transcription accuracy* of the STT in some situations. This issue mainly concerns the beginning of an interaction where the human must provide the object identifier, e.g., alphanumeric product code. Similar sounding numbers and letters of one code result in an incorrect transcription. These errors are critical for this application because the product code is the primary reference for item-level information. Potential measures to solve this issue are an auto-correction of transcriptions errors, the use of other input formats (e.g., using a bar code taken with a smartphone), and the use of synonyms, such as human first names, location names. Another problem related to STT transcription is *factory noise*. We did not focus on this aspect of the presented study. Still, research in progress indicates that existing noise canceling and filtering solutions can mitigate this problem for some noise types.

The main **non-technological barrier** in our study is *intransparent technology*. Alexa is a free-to-use tool, but it is not clear how Amazon processes the recorded information for its business purposes. There is a mistrust regarding Amazon's intention and its technical and legal capabilities to protect the conversation data. Overcoming this barrier has at least two aspects to consider that describe a typical trade-off.

On the one hand, the *closed technologies* of companies, such as Amazon or Google, are most accurate because of the extensive audio data they collect and use for machine learning. They acquired these data because of their many consumers and benefit from them through the companies' liberal interpretation of data privacy. They also provide tools that support the design of DAs, which lowers the skill-floor to become a developer. On the other hand, *open technologies*, such as the privacy-focused voice assistant Mycroft and Mozilla's transcription solution DeepSpeech, are more transparent about how they operate. Their downside is that these community-driven technologies are not as accurate as closed technologies, and there are fewer tools to support their use. This statement, however, is a snapshot, and future improvements in these technologies are likely.

One of the follow-up hypotheses of this article is that open technologies can achieve similar results compared to the Alexa-based implementation. The follow-up research should also demonstrate that factory noise is manageable.

A second follow-up hypothesis is that users can benefit from a voice-enabled DA in situations where the digital counterpart has significantly more components and relations among them. This increased complexity includes the integration of item-level information from the entire product life cycle.

Acknowledgments

This project (DIAMOND) has received funding from the European Institute of Innovation and Technology (EIT), a body of the European Union, under the Horizon 2020, the EU Framework Programme for Research and Innovation.

References

[1] G.G. Meyer, K. Främling, J. Holmström, Intelligent Products: A survey, Computers in Industry 60 (2009) 137–148. https://doi.org/10.1016/j.compind.2008.12.005.

[2] J.-F. Uhlenkamp, K. Hribernik, S. Wellsandt, K.-D. Thoben, Digital Twin Applications A first systemization of their dimensions, IEEE International Conference on Engineering, Technologz and Innovation (ICE/ITMC) (2019) 1–8. https://doi.org/10.1109/ICE.2019.8792579.

[3] E.J. Tuegel, A.R. Ingraffea, T.G. Eason, S.M. Spottswood, Reengineering Aircraft Structural Life Prediction Using a Digital Twin, International Journal of Aerospace Engineering 2011 (2011) 1–14. https://doi.org/10.1155/2011/154798.

[4] A.A. Malik, A. Bilberg, Digital twins of human robot collaboration in a production setting, Procedia Manufacturing 17 (2018) 278–285. https://doi.org/10.1016/j.promfg.2018.10.047.

[5] K.A. Hribernik, L. Rabe, K.D. Thoben, J. Schumacher, The product avatar as a product-instance-centric information management concept, IJPLM 1 (2006) 367. https://doi.org/10.1504/IJPLM.2006.011055.

[6] S. Few, Dashboard Confusion Revisited, Perceptual Edge (2007).

[7] B.R. Cowan, N. Pantidi, D. Coyle, K. Morrissey, P. Clarke, S. Al-Shehri, D. Earley, N. Bandeira, "What can i help you with?" 1–12. https://doi.org/10.1145/3098279.3098539.

[8] J.-G. Harms, P. Kucherbaev, A. Bozzon, G.-J. Houben, Approaches for Dialog Management in Conversational Agents, IEEE Internet Comput. 23 (2019) 13–22. https://doi.org/10.1109/MIC.2018.2881519.

[9] A. Ram, R. Prasad, C. Khatri, A. Venkatesh, Conversational AI: The Science Behind the Alexa Prize (2017).

[10] D.A. Coates, M. Amordeluso, Voice applications for Alexa and Google Assistant, Manning Publications Co, Shelter Island NY, 2019.

[11] C. Pearl, Designing Voice User Interfaces: principles of conversational experiences, O'Reilly Media, 2016.

[12] Amazon Alexa Skills Challenge: Kids: Build an Alexa skill that educates, entertains, and engages kids under 13, 2018, https://alexakidskills.devpost.com/.

[13] High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, 2018, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419, accessed 21 April 2020.

[14] K.A. Hribernik, L. Rabe, K.D. Thoben, J. Schumacher, The product avatar as a product-instance-centric information management concept, IJPLM 1 (2006) 367. https://doi.org/10.1504/IJPLM.2006.011055.

[15] GS1, EPCIS and CBV Implementation Guideline: Using EPCIS and CBV standards to gain visibility of business processes, 1st ed., 2017.

[16] S. Desai, Q. Deng, S. Wellsandt, K.-D. Thoben, An Architecture of IoT-Based Product Tracking with Blockchain in Multi-sided B2B Platform 566 458–465. https://doi.org/10.1007/978-3-030-30000-5_57.

[17] Amazon, Understand How Users Interact with Skills, https://developer.amazon.com/en-US/docs/alexa/ask-overviews/understanding-how-users-interact-with-skills.html, accessed 28 February 2020.