

# A Systematic View on Speech Assistants for Service Technicians\*

Joachim Baumeister<sup>1,2</sup> Veronika Sehne<sup>1,2</sup> Carolin Wienrich<sup>2</sup>

**Abstract.** The paper gives a systematic view on speech assistants in the field of technical service of industrial machines. We describe the results of a requirements analysis targeting companion technologies for service technicians and we report on a first reference implementation. The effectiveness of the approach is evaluated in a diagnosis task scenario and preliminary results of a user study are discussed.

knowledge-based diagnosis system intelligent personal assistant speech interaction dialog systems

## 1 Motivation

Intelligent personal assistants for private use are implemented in many devices nowadays. With their availability in smartphones, watches, and TVs, they diffused into daily life and support users when, for instance, sending text messages, setting reminders, and starting apps or media channels. They simplify the use of existing technology by making it more intuitive and quicker in execution.

In general, an intelligent personal assistant provides a natural language interface to take requests from the user and perform corresponding actions. Today, many assistants interact with the user by a speech interface. In research, the development of (smart) speech assistants was elaborated in many works, for instance see [1, 3, 9, 11, 12]. In this paper, we ask the research question, whether and how the obvious benefits of such assistants can be transferred from personal life to industrial use cases. Here, we especially look for applications of speech assistants in the context of Technical Service. In general, the domain of *Technical Service* considers the operation, the optimization and maintenance, and the repair of often very complex industrial machines.

---

\*Supported by German BMWi Project Grant ZF4172703BZ7 (MARS project)  
Copyright ©2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## 2 Background and Related Work

### 2.1 Assistance Systems for Service Technicians

Today's assistance systems for service technicians are often structured like search applications of the 2000er years: Providing simple textual search interfaces the technician need to manually formulate an appropriate query and click through the delivered results. Only a few number of systems provide a kind of semantic interface that is able to return relevant information bits not necessarily matching the original but the intended meaning of the query. Also, the technicians are required to actually hold and touch the assistance system's device. Speech assistants would enable the technicians to intuitively formulate the information request and also free them to touch a keyboard device while doing so. The aim of the present study is an user-centered development of a personal speech assistant for service technicians.

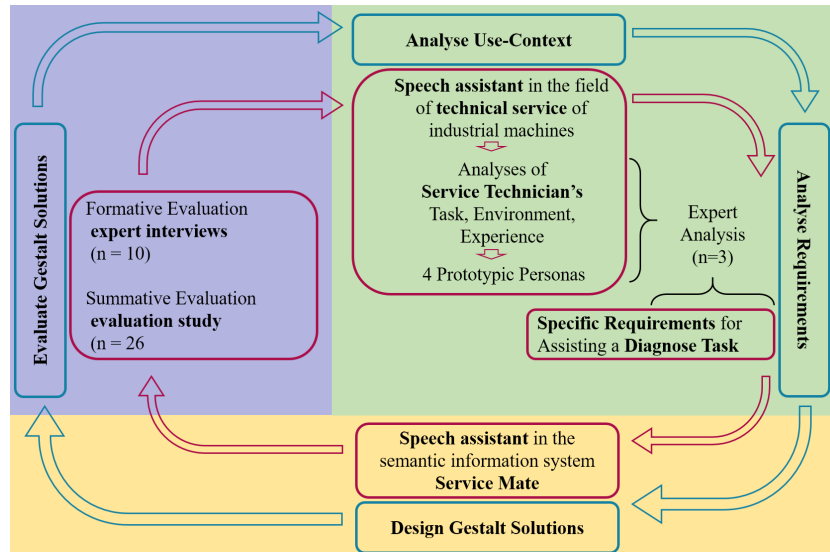
### 2.2 User-Centered Design Process

Norman and Draper [7] introduced the user-centered-design (UCD) process which condensed different approaches and methods known in the field of human-computer-interaction (HCI). Similar, the DIN EN ISO 13407 describes the basic steps of user-centered design processes consisting of the analysis of the context, the analysis of requirements as well as the iterative design and evaluation of gestalt solutions [5]. Integrating different perspectives, the approach of contextual design [4] involves users to analyze and evaluate systems (includes often summative evaluations) and the approach of usability engineering [6] involves experts (includes often formative evaluations) [8, 10]. The outer blue circle of Figure 1 demonstrates the common steps of the UCD process and how they are intertwined.

Even though, the UCD process is well established in the HCI, context and requirement analyses for speech assistance systems are rare. Dumas et al. [2] developed a guideline for multi-modal user interfaces including speech. However, the guideline seems rather universal and not specific enough to meet the requirements of a personal speech assistance for service technicians. Thus, the present article analyses the context and requirements of service technicians to design an user-centered speech assistance including an expert and user evaluation.

## 3 Analyzing Phase in the Present Study

Our use-context is the domain of the Technical Service, focusing particularly on the service technician. For this reason, we first introduce the domain of Technical Service of industrial machines and then domain specific tasks, environmental conditions, and experiences of a service technician. The use-context and first implications for the development of a speech assistant were analyzed iteratively by three experts from the domains human-computer-interaction and technical services. The analyses of context result in four prototypical personas. We conclude



**Fig. 1.** The outer blue circle shows a typical user-centered design process. The inner red circle shows the key points of the user-centered design process in the present article. Green, yellow, purple areas highlight the analyzing, designing, and evaluating phase.

the section with a description of the context-related requirements, that—in the view of experts—a targeted speech assistant need to meet. Figure 1 shows the key points of the analyzing phase in the present study (inner red circle, green area).

### 3.1 Analyzing the Use-Context: The Domain of Technical Service

Advanced industrial machinery is one of the most driving domains in today's life. Machinery produces almost all consumer goods with high efficiency. In this case, it touches all aspects of personal life, for instance harvesting machines feeding the food production of animals and humans, paper making and printing machines to produce newspapers, and automotive factories to produce cars and trucks.

However, machinery needs to be safely operated and maintained for its optimal performance. In case of malfunction it need to be brought back to work quickly. Often, the standstill of such machines yield exceptional costs in the area of hundreds of euros and more—per minute. The Technical Service, and namely the service technicians, are responsible for the maintenance, performance optimization, diagnosis and repair of industrial machinery.

### 3.2 Analyzing Work-related Characteristics of a Service Technician: Domain Tasks, Environmental Conditions, and Domain Experiences

**Domain Tasks** As an overall goal, the service technician should make sure to maintain the machine state and sometimes even to optimize the machine performance. To archive these goals we distinguish the following sub-tasks:

1. Operation and monitoring of the machine performance
2. Disassembly/assembly of machine components (optimization/repair)
3. Diagnosis of faulty machine behavior
4. Maintenance operations for a machine
5. Documentation of accomplished work for commercial and knowledge management reasons

The present paper focuses on the task of diagnosis including the use of an information system with a guided troubleshooting on the one hand and the work around/in/on/under the machine during the diagnosis on the other hand. The latter involves the use of special tools or working gloves while the former requests two free hands for navigate through the different diagnosis steps. As consequences for the development of the speech assistance result the implications of the integration of the speech assistant in a diagnosis interface of a semantic information system on the one hand and the possibility of a hands-free interaction on the other hand.

**Environmental Conditions** The service technician accomplishes the tasks within changing environmental parameters. Basically, the experts distinguish the following main parameters that affect the quality and speed of the technicians work:

1. *Location* The technician is working in a workshop or “in-field”. As consequences for the development of the speech assistant we see:
  - (Reduced) availability of required tools
  - (Reduced) support and knowledge provided by on-site colleagues
  - (Reduced) Internet access for remote communication
2. *Supporting Information and Infrastructure* considers the availability of information resources in breadth and depth, e.g., technical documentation, diagnostic systems, maintenance plans, communication channels, and mobile devices. As consequences for the development of the speech assistant result we see:
  - Required infrastructure for providing information resources (cloud systems and storage)
  - Different levels of information can be provided depending on the availability of mobile devices
3. Limited *Time* may stress the technician to complete the task. This primary limitation influences the use of tools, the quality of work, and the type of possible tasks. As consequences for the development of the speech assistant result we see:

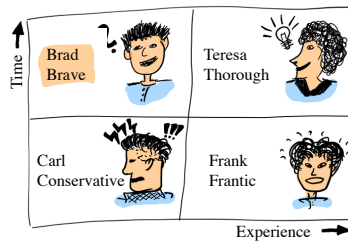
- Speech assistants need to be adaptable with respect to the available time and mental workload of the user. This affects the query patterns, the possible intents, and the (extensive) answers provided by the assistant.

**Domain Experience** The performance of the overall service task heavily depends on the experience of the technician. In general, it is difficult to classify the individual experience, but some indicators can help to distinguish different experience levels: a) The number and type of training sessions completed in recent years and b) The years of work in the particular domain.

Some companies define specific levels of experience based on the type of work the technician can complete individually, e.g., from *Level 1* (Technician can safely operate the machine) to *Level 4* (Technician can safely diagnose and repair faulty behaviour for which diagnostic knowledge is available) and up to *Level 5* (Technician can safely diagnose and repair a previously unknown issue). Consequentially, a speech assistant should support the individual experience level of a service technician.

### 3.3 Four Prototypical Personas of a Service Technician

Based on the analysis of basic work-related characteristics of a service technician’s tasks, environmental conditions, and domain experiences, the experts identified four typical personas, that represent different prototypes of service technicians. They focus in variations on domain experience and the available time to complete a given task. Thus, the four personas represent a typical cross-selection of the domain of Technical Service.



**Fig. 2.** Characterization of personas in the domain of Technical Service concerning the domain experience and the time for task completion.

In Figure 2, we see a classification of the identified personas: *Brad Brave* (“I am happy to learn new stuff every day!”) is innovation friendly, but a beginner of the domain. He has the motivation to use new things and invests time into the exploration and use of information systems. *Carl Conservative* (“I cannot spend extra time for this additional stuff.”) also has not much experience in the

field but has not motivation to invest time, e.g., in the exploration and use of an information system. *Frank Frantic* ("It is about the bolts not about the apps!") is very experienced in the domain, but shows no interest to explore new tools that possibly may improve his work. *Theresa Thorough* ("I love to improve every day!") has much experience in the domain and also is very innovation-friendly. She is interested to further improve her personal performance by trying and using advanced tools.

### 3.4 Analyzing Requirements

Following the context analysis, three experts analyzed specific requirements for developing a speech assistant for service technicians. In summary, 28 requirements were identified and clustered in five categories. Table 1 shows the five clusters and examples of the corresponding specific requirements.

**Table 1.** Analyzed requirements for a personal speech assistance supporting the diagnosis in Technical Service. Marked requirements show the implementations in the current system.

Requirement Cluster	Requirement for the Development
Symptom detection	Robust symptom detection (also various in one input), disambiguation*, recognition of technical & colloquial language
Orientation within dialog	Guided dialog* including pausing*, go backwards*, skipping & repeating* questions, multi-modal interaction*
Speech recognition	Online/Offline speech recognition, activation words recognition*
Explanations	Display further information, documents, explanation of the solution & differential diagnosis
Trust in the speech assistant	Using technical language*, transparency, personalization

## 4 Evaluation Phase 1: Expert Evaluation of Analyzed Requirements

In order to proof the relevance of the analyzed requirements, ten independent experts (age: 23 to 43 years, 9 male) from the field of diagnostic dialogues revisited the 28 requirements and rated the importance on a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree).

Table 2 shows the requirement clusters and ratings of the second expert group. *Symptom detection* was rated as most important, followed by *orientation within the dialog*, *speech recognition*, and *explanations*. *Trust in the speech assistant* was rated as less important.

In addition to the 28 requirements identified by the first expert group, the second expert group found 11 requirements. These 11 additional requirements fit

**Table 2.** Means and standard deviations of important ratings of the second expert group on the requirement clusters.

Requirement Clusters	Second Expert Group
Symptom detection	$M = 4.0, SD = 0.4$
Speech recognition	$M = 3.2, SD = 1.1$
Orientation within dialog	$M = 3.7, SD = 0.3$
Explanations	$M = 3.4, SD = 0.4$
Trust in the speech assistant	$M = 2.4, SD = 1.1$

in the above presented requirement clusters and did not changed the importance pattern of the clusters.

In sum, the two expert groups identified 40 requirements clustered in five groups with different importance for the development of a speech assistant for service technicians.

The first conceptional design of the speech assistant targets on the prototypical persona of *Brad Brave* (see Figure 2), because he is innovation-friendly and invests time into the exploration and use of information systems. Due to his limited working experience (Level 3, i.e., needs guidance), his working performance and experience can be improved by companion technologies. From the identified requirements, Brad needs the following:

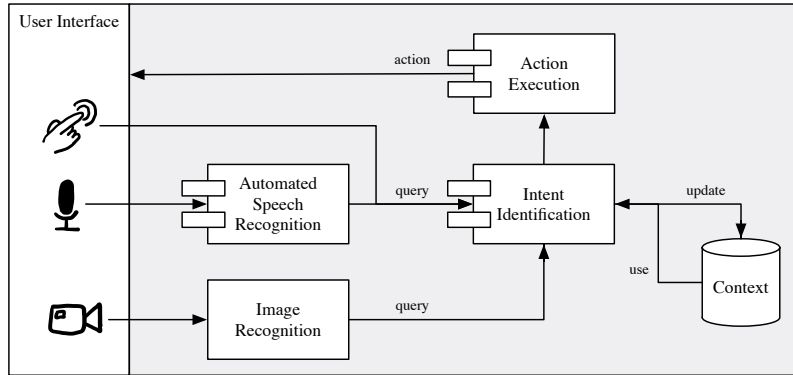
- The assistant works offline and online, because the availability is required not only in the workshop but also on-field with a probably worse internet connection.
- Brad often needs both hands to accomplish the service work, thus the assistant should be able to operate hands-free.
- The assistant needs to support the following tasks:
  - Basic information research for technical documentation
  - Support a diagnosis task by a guided diagnostic dialog

Following these requirements the assistant should be helpful for the remaining personas as well, at least in some situations. In the subsequent section, we introduce a concept of a speech assistant for the service technician Brad Brave, that takes the considerations from above into account.

## 5 Design Phase: Personal Assistant for Technicians

### 5.1 Conceptual View

The competence of a personal speech assistant can be characterized by the collection of *intents* it understands and can react on. Intents are formulated by the user in natural language and require one or more corresponding actions performed by the assistant. The requirements stated above yield the following main intents: a) Search for information in the available technical documentation. b) Provide diagnostic support for a specific problem description.



**Fig. 3.** Architecture of a multi-modal assistant including speech interaction.

Figure 3 shows the general architecture of the personal assistant. It consists of an automated speech recognition (ASR), intent identification, an action execution module, user context management, and a user interface. For the modules *Intent Identification* and *Action Execution*, we see adapter interfaces, so that the assistant can learn new intents/actions by plug-ins. In the following, we briefly describe each module in more detail.

### User Interface

*Query Formulation* An advanced user interface of an assistant is able to capture the user query in different modalities, e.g., speech utterances, video images and basic touch/text input. In the context of this paper, we focus on query formulation using speech utterances. The query needs to be transparently elicited by the system, i.e., the user sees the entered input instantly in the system while actually providing it. This is trivial for keyboard entries but also the text of recognized speech input and captured video images should be displayed to the user for transparent tracability.

*Deliver Action* The user interface outputs the results of the actions derived by the assistant. In the simplest case, the assistant simply displays the document the user asked for. In the case of a guided diagnosis system, the action module delivers an interactive dialog with the user.

**Automated Speech Recognition (ASR)** This module transforms spoken text (sound waves) into text input, so it can be later analyzed by the subsequent intent module.



**Intent Identification** The intent identification is responsible for the semantic interpretation of the recognized text. The module tries to identify the request of the user in order to find an appropriate action for the request. It includes techniques from natural language understanding and question answering. Classic approaches are based on rules and patterns, but recently also statistical learning approaches were introduced.

**Action Module** Based on the recognized intents an appropriate action is selected. In our scenario, we refer to the tasks that a speech assistant should support (see Section 3.4): Support the documentation research by providing useful information and facts for a given question and the support for diagnostic questionnaires.

**Context Update** The assistant is able to track the work of the technician and uses this work context to support the intent identification and the action selection.

Typical task information elements are the previously stated queries, the selected actions, and the environment of the current use.

## 5.2 Reference Implementation

A prototypical implementation of the speech assistant was developed for the existing information system Service Mate (<http://www.servicemate.de>). Originally, the application serves as an information system for the research and consumption of technical service documentation. For the implementation of the diagnosis capabilities, the existing speech assistant was extended by an interactive speech dialog. The diagnostic dialog now can be started by simply stating "start diagnosis" into the speech input.

Figure 4 depicts an example dialog of the diagnosis system. Here, the malfunction of a bicycle is analyzed by a question–answer dialog. The figure shows a question asking for the wear of the wheel rim with possible answers "visible" and "not visible". The multi-modal interface of the system allows for simply touching the buttons to answer the question, but also the answer can be given by speech input. After answering the question the next best relevant question is asked in order to derive a possible cause for the observed fault of the bike.

The next section outlines the second evaluation phase. Note, that the complete user evaluation is not described in this paper due to space constraints.

## 6 Evaluation Phase 2: User Evaluation of Speech Assistant

In order to investigate whether the speech assistant indeed support the service technician, we compared the user experience of the developed speech interaction with the established touch interaction of the system as well as expectations previous the usage.



**Fig. 4.** The diagnosis interface of the Service Mate showing a specific question, a helpful image and the possible answers for the question as touch buttons. It is also possible to answer the question by speech input.

## 6.1 Method

**Participants** The study was conducted by 26 participants having a mean age  $M = 31.23$  with standard deviation  $SD = 7.98$ , 4 females. The participants had medium experience with the information system.

**Material and Procedure** The empirical study use an exemplary but fully functional service system for a bicycle, i.e., the *dBike* system. Besides the technical documentation, a circuit diagram, and a 3D model the system also contains diagnosis routines for the most relevant functions of the bike. The experiments were conducted at a bike, a roughedized tablet computer was running the information system with the touch and speech interaction. Additionally, a separate speaker broadcasts the speech instructions. A further notebook was provided for answering the questionnaires of the experiment study. Participants had two tasks: (i) the diagnosis of faults with the gear shift and (ii) the diagnosis of a malfunction of the brake. The bike was specially prepared for each task. For the examination of the bike, the persons need to use the work gloves. Hence, for using the tablet computer, participants had to undress the gloves for touch interaction. In addition, some diagnostic steps request the usage of a special measuring tool.

Half of participants conducted the diagnosis using the speech assistant, the other half was using the touch interaction on the tablet (independent variable: mode of interaction). The groups only differ in the mode of interaction implying that participants using the touch interaction had to move between the bike and the laptop for each diagnostic step and had to undress the gloves for interaction on the one side. On the other side, participants using touch interaction

saw pictures with instruction. If participants of the speech interaction need a visualization, they had to move to the touch laptop, too.

Previous to the interaction, all participants provided demographical information. Participants using speech interaction rated the expected usefulness and expected problems of speech interaction for the diagnosis process on a 5-point Likert scale ranging from 1 (not useful) to 5 (very useful), or 1 (many problems) to 5 (no problems), respectively. After their interaction with the speech assistant, we assessed qualitatively positive and negative experiences. While the participants using speech interaction assessed their expectation prior the experiment, participants using touch interaction assessed the expected usefulness and problems of speech interaction after the experiment on a 7-point Likert scale ranging from 1 (not useful) to 7 (very useful), or 1 (many problems) to 7 (no problems), respectively. Other questionnaires were applied, but are not reported in the present paper.

## 6.2 Results

Prior the interaction with the speech assistant, participants rated speech assistance to support the diagnosis process as rather useful ( $M = 3.92$ ,  $SD = .64$ , range between 1 and 5). The auditive interaction mode and the corresponding free-hand interaction were mentioned as most frequent reasons. But also problems were expected ( $M = 3.62$ ,  $SD = .51$ , range between 1 and 5). Most problems were expected for the speech recognition. Further, some participants expected problems with the comprehension of diagnostic instructions. After the interaction with the speech assistant, most participants liked the hand-free interaction, the ease of use, and the intuitive control commands. On the other side, participants reported negative experience concerning the restricted use of language and the fact that they cannot interrupt the speech assistant when it was speaking.

After the interaction with the touch based assistant, participants also expected speech assistance as rather useful to support the diagnosis process ( $M = 5.58$ ,  $SD = 1.31$ , range between 1 and 7) and as medium problematic ( $M = 4.83$ ,  $SD = 1.40$ , range between 1 and 7). The reasons for their ratings were similar to those of the pre-experimental ratings. Free-hand interaction were mentioned most frequently for a useful application. Speech detection and the comprehension of diagnostic instructions were mentioned as most frequently problems. Further, participants using touch interaction reported positive experiences with pictures supporting the comprehensive of the corresponding diagnosis steps. Undressing the gloves for touch interaction was experienced as negatively.

## 6.3 Discussion

The results revealed the potentials of speech assistance in the domain of technical services. Particularly, the possibility of hand-free interaction is expected as useful. However, the results also revealed concerns. The use of auditive guidance might lead to comprehensive problems and pictures support the comprehension of corresponding diagnosis steps. Thus, a multi-modal interaction providing

hands-free speech interaction and the visualization of 3D-models might be the best support for service technicians.

## 7 Conclusions

In this paper we introduced speech assistants in the domain of Technical Service. Based on a user-centered requirements analysis for service technicians, we selected a number of relevant requirements to be implemented in an existing information system. We evaluated the research question, whether and how speech assistants can support a service technician. In the experiment we focused on the diagnosis task. In the future, we are planning to extend the reported user studies to a diverse range of specific research questions. Particularly, a comparison of speech, touch, and multi-modal interaction is planned for future studies. Further, a field study including *real* service technicians is planned.

## References

1. Christensen, H., Casanueva, I., Cunningham, S., Green, P., Hain, T.: homeService: Voice-enabled assistive technology in the home using cloud-based automatic speech recognition. In: 4th Workshop on Speech and Language Processing for Assistive Technologies. pp. 29–34 (2013)
2. Dumas, B., Lalanne, D., Oviatt, S.: Multimodal interfaces: A survey of principles, models and frameworks. In: Human machine interaction, pp. 3–26. Springer (2009)
3. Hamerich, S.W.: Benutzerfreundliche Sprachdialoge im Automobil. In: Sprachbedienung im Automobil, pp. 61–77. Springer (2009)
4. Holtzblatt, K., Beyer, H.: Contextual design: Design for life. Morgan Kaufmann (2016)
5. ISO, I.: 13407: Human-centred design processes for interactive systems. Geneva: ISO (1999)
6. Nielsen, J.: Heuristic evaluation. In: Usability inspection methods. pp. 25–62. John Wiley & Sons, Inc. (1994)
7. Norman, D.A., Draper, S.W.: User centered system design: New perspectives on human-computer interaction. CRC Press (1986)
8. Rubin, J., Chisnell, D.: Handbook of usability testing: how to plan, design and conduct effective tests. John Wiley & Sons (2008)
9. Santos, J., Rodrigues, J.J., Casal, J., Saleem, K., Denisov, V.: Intelligent personal assistants based on internet of things approaches. IEEE Systems Journal **12**(2), 1793–1802 (2018)
10. Sarodnick, F., Brau, H.: Methoden der usability evaluation. Wissenschaftliche Grundlagen und praktische Anwendung **1** (2006)
11. Sidner, C.L.: Building Spoken-Language Collaborative Interface Agents, pp. 197–226. Springer Netherlands, Dordrecht (2004)
12. Wobcke, W., Ho, V.H., Nguyen, A.T.L., Krzywicki, A.: A BDI agent architecture for dialogue modelling and coordination in a smart personal assistant. IEEE/WIC/ACM International Conference on Intelligent Agent Technology pp. 323–329 (2005)