

# Increasing the acceptance of speech assistants

## *How can dialogues with machines be improved?*

Monday, 1 March 2021

*Exposé Bachelor Thesis Anais Siebers - Hochschule RheinMain*

For humans, it is easiest to communicate via language - they learn it as children and gain experience their whole life [MSHH19]. That is why speech assistants spread worldwide. Nevertheless, it is interesting that the amount of users (about 31% of the German population) does not increase since 2018<sup>1</sup>. There are multiple reasons for this phenomenon. Most prominent is that the usage feels limited<sup>2</sup>.

### **GOAL**

The goal of this thesis is to increase the acceptance of speech assistants and improve the motivation to use them by improving the communication between humans and speech assistants. Therefore, problems in the communication will be researched and an approach to minimise these is offered. The focus will be the dialogue management assuming that other components of the speech assistant pipeline are working reliable and their performance is sufficient to be able to focus on the dialogue.

### **WHAT IS DIALOGUE MANAGEMENT?**

Speech assistants consist of different components which each have their own task [SRM<sup>+</sup>16]. First, there is automatic speech recognition (ASR) recognising if a person started to talk to the assistant. The following speech is converted to text by speech-2-text conversion (S2T). The next component is natural language processing (NLP) or rather natural language understanding (NLU). This means the text is being “understood” - meaning text is transferred to knowledge. NLU is preferred in speech assistants because it retrieves the user’s intent from a given text<sup>3</sup>. Now a dialogue manager (DM) forwards the retrieved information to intelligent algorithms (AI or similar) which decide which task is executed. In

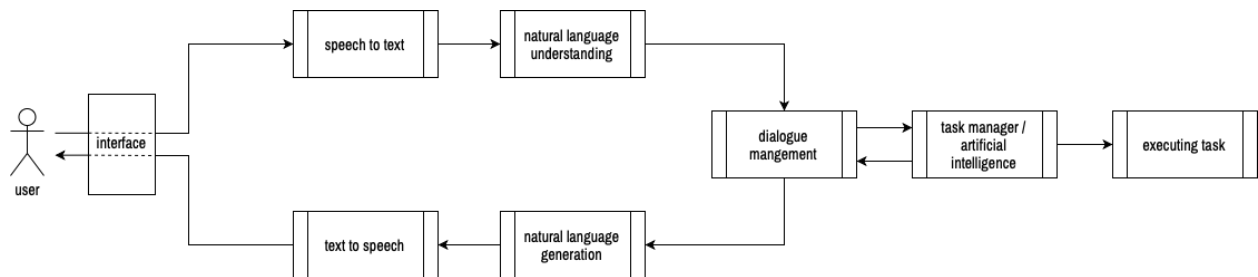
---

1 <https://de.statista.com/infografik/23155/umfrage-zur-nutzung-von-sprachassistenten-in-deutschland/> [last access: 22<sup>nd</sup> January 2021, 00:41 am]

2 <https://www.theverge.com/2019/11/6/20951178/amazon-alexa-echo-launch-anniversary-age-funtionality-not-changed-use-cases> [last access: 23<sup>rd</sup> January 2021, 11:26 am]

3 <https://kauz.net/chatbots/blog/was-ist-der-unterschied-zwischen-nlp-und-nlu/> [last access: 27<sup>th</sup> February 2021, 13:38 am]

the end the DM determines the feedback the user is given. This is synthetically generated speech using text-2-speech conversion (T2S). This is the complete pipeline<sup>4</sup>:



## APPROACHES TO DIALOGUE MANAGEMENT

The thesis will focus on dialogue management. There are different approaches to deal with dialogues. The most spread approaches being so called command-systems or menu-dialog systems [Sie19]. They are very common on telephone hotlines. Other approaches to deal with dialogue management in chatbots or speech assistants are graph-based, frame-based or a statistical approach [ARA13]. Furthermore, there are the tree-based, finite-state, plan-based, agent-based approaches<sup>5</sup>. They all have different focusses and the suiting approach is chosen depending on the use case, available data, scale and available resources [Bur17].

## WHAT ARE OPEN PROBLEMS?

**Creating Sympathy towards the speech assistant** Classical speech assistants create dialogues with the help of structured sentences [Sie19]. The sentences are build with gaps. If a user says exactly the same sentence, information is retrieved by looking at the words in the gap. This kind of interaction leads to a need for the user to learn the assistant's language [Pin94]. Since the motivation for learning new languages is very low<sup>6</sup>, it is an obstacle in interaction. Furthermore, from the perspective of the user, the assistant is the one who has to come to meet the user not vice versa. This contradicts learning the assistant's language. Since the assistant is used out of comfort, it cannot be difficult to interact with it [Udl18]. Another important aspect is that humans adapt to one another's language if they find one another sympathetic. This very natural and subconscious process does not take place with current speech assistants.

4 <https://blog.vsoftconsulting.com/blog/chatbot-development-designing-dialog-management> [last access: 27<sup>th</sup> February 2021, 13:17 am]

5 <https://people.ict.usc.edu/~traum/ESSLLI08/Lecture1.pdf> [last access: 27<sup>th</sup> February 2021, 17:16 am]

6 <https://de.statista.com/statistik/daten/studie/290926/umfrage/umfrage-zur-haeufigkeit-der-beschaefigung-mit-anderen-kulturen-in-der-freizeit/> [last access: 23<sup>rd</sup> January 2021, 11:40 am]

**Multi-staged dialogues / Topic-jumping during conversation** The currently and prospectively imaginable use cases for speech assistants<sup>7</sup> are mainly exercises which can be initiated with one command and do not need further specification. Just like e.g. calling a person: “Call Miss Meier”. But at place three, there already is an exercise requiring multi-stage interaction: searching which is a process often requiring multi-stage interaction. Even calling can turn into a multi-staged dialogue if multiple persons share the same name. Another problem dialogue management faces is that many people jump during conversations or start another questions before they answer an asked question.

**Dealing with inability and incomprehension** The user creates a long-term relationship to the devices in his surrounding [Reb17]. Assistants are laid out for long-term use. That is why mistakes have to be seen as a chance to improvement and this needs to be communicated to the user to sustain his motivation to use the assistant. Studies showed that people are more positive after a failure if they see the possibility that there will be improvement in the future [WR9].

**Adapting to cultural habits** Current assistants also lead to frustration because they do not maintain cultural habits. For example, they do not react if a user wants to interrupt an answer because it is not what he wanted. They also talk stiffly and often unfamiliar due to the use of different vocabulary [KOY7].

## FOCUS OF THE THESIS

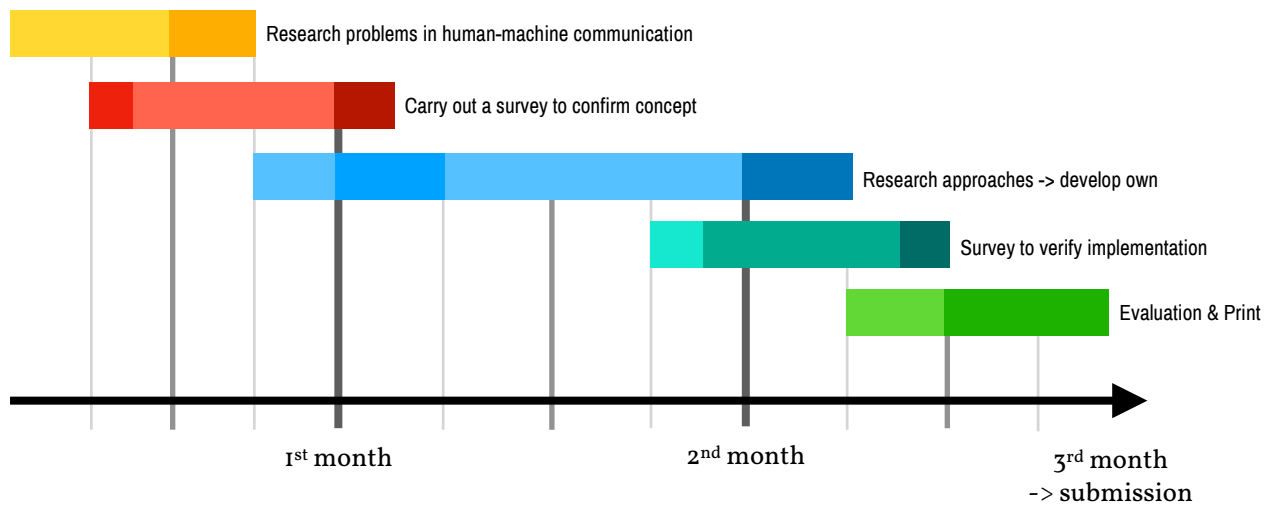
The thesis will focus on the main problem that speech assistants often lack sympathy or there is annoyance towards speech assistants. Furthermore, there is a lack of trust that speech assistants misuse data and listen to everything. Since trust is essential for technical devices of everyday life [Nor4], the goal is to research problems in the communication and provide an offline-capable transparent solution - focusing on dialogue management. The main problems which will be discussed are:

- creating sympathy towards the speech assistant
- multi-staged dialogues / topic-jumping during conversation
- dealing with inability and incomprehension
- adapting to cultural habits

---

<sup>7</sup> <https://de.statista.com/infografik/4928/anwendungsbereiche-von-digitalen-sprachassistenten/> [last access: 22<sup>nd</sup> January 2021, 00:57 am]

## COURSE OF ACTION / MILESTONES



- 1. Research problems in human-machine communication** The first phase researches problems in the interaction between humans and speech assistants.
- 2. Carry out survey to confirm concept** The second phase consists of a survey to find out whether the problems determined in phase 1 really are problems and if problems are neglected.
- 3. Research approaches in dialogue management** During this phase, approaches for dialogue management are researched and an own approach for a speech assistant to solve the problems discovered in phase 1 and 2 is developed.
- 4. Survey to verify implementation** To see whether the implementation could solve the problems, another survey is held.
- 5. Evaluation** The last phase evaluates found results and gives an outlook for future improvement.

The lighter parts of the beams are for development or creation. For the surveys it means: creating the surveys on a profound basis. The darker parts are for documentation and summarising the found results. The rest represents research, implementation or conduction.

# Literature

- [MSHH19] MERDIVAN, ERINC ; SINGH, DEEPIKA ; HANKE, STEN ; HOLZIGER, ANDREAS: Dialogue Systems for Intelligent Human Computer Interactions. In: *Electronic Notes in Theoretical Computer Science - May 2019* (2019)
- [SRM<sup>+</sup>16] SCHNELLE-WALKA, DIRK ; RADOMSKI, STEFAN ; MILDE, BENJAMIN ; BIEMANN, CHRIS ; MÜHLHÄUSER, MAX: NLU vs. Dialog Management: To Whom am I Speaking. From: *Workshop Interacting with Smart Objects, March 10th, 2016, Sonoma, CA, USA* (2016)
- [Sie19] SIEBER, A: Dialogsysteme in der Praxis. In: *Dialogroboter* (2019 - Springer Fachmedien), S. 49-127
- [ARA13] AHMED, MANZOOR ; RIYAZ, ROMANA ; AFZAL, SADUF: A comparative study of various approaches for dialogue management. In *COMPUSOFT - An International Journal of Advanced Computer Technology 2* (4), April, 2013 (2013 - COMPUSOFT, Volume-II, Issue-IV)
- [Buri7] BURGAN, DEENO: Choosing an Action Selection Method. In: *Dialogue Systems & Dialogue Management - January 2017* (2017 - Australian Government - National Security & ISR Division), S. 42-43
- [Pin94] PINKER, S.: *The Language Instinct* (1994 - Penguin), S. 193
- [Udl18] DR. UDLUFT, HEIKO: Schach, Suchmaschine, Sprachassistent - unser Verhältnis zur Künstlichen Intelligenz. In: *Digitale Welt* (4) (2018)
- [Reb17] REBAUDENGO, SIMONE: *Design for Living with Smart Products, The Intelligent Home* (2017 - O'Reilly Media, Inc.)
- [WR9] WHITE, RYEN W. ; ROTH, RESA A.: Defining Exploratory Search. In: *Exploratory Search: Beyond the Query-Response Paradigm (Synthesis Lectures on Information Concepts, Retrieval & Services)* (2009 - Morgan and Claypool Publishers), S. 9-23
- [KOY7] KIRAKOWSKI, JUREK ; O'DONNELL, PATRICK ; YIU, ANTHONY: The Perception of Artificial Intelligence as "Human" by Computer Users. In: *Human-Computer Interaction, Part III, HCII 2007 (Ed. J. Jacko)* (2007 - Springer Verlag)

[Nor4] NORMANN, DON; Emotional Design: Why We Love (Or Hate) Everyday Things. (2004 - Basic Books)