Development and evaluation of a speech interaction prototype for elderly people

Master Thesis

For attainment of the academic degree of
Master of Science in Engineering (MSc)

in the Master Programme Digital Healthcare
at St. Pölten University of Applied Sciences

by
Max Jesenko, BSc
1710756501

First advisor: Andreas Jakl, MSc
Second advisor: Dipl.-Sporting. Dr. Mario Heller

[Vienna, 14.05.2019]

Declaration

I declare that I have developed and written the enclosed Master Thesis completely by myself, and have not used sources or means without declaration in the text. Any thoughts from others or literal quotations are clearly marked. This work was not used in the same or in a similar version to achieve an academic grading or is being published elsewhere.

Vienna, 14.05.2019

....

Place, Date Signature

Abstract

Due to population aging, enhancing the methods of caring for elderly persons gets more important every day. Research in the field of ambient assisted living shows the potential of using digital technologies to facilitate caring work by providing smart assistance systems. Smart devices could offer a great benefit for elderly. However, these intelligent devices such as voice assistants and robotics are particularly challenging for older generations with regards to interaction and communication.

The “Smart Companion” concept, which was developed at UoAS St. Pölten, aims at tackling those challenges by combining low-cost robot sensors (vacuum robots) with smart speech interaction to simplify communication between human beings and machines. The purpose of this thesis is to explore the benefits, acceptance, possibilities and limitations of a speech interaction system and contribute to the development of the “Smart Companion”. This should pave the way for a concrete realization of the concept.

The main method to answer the research questions is the development and evaluation of speech interaction prototypes via measuring predefined scenarios in Wizard of Oz (WOz) tests and structured surveys with the study participants. The WOz tests were conducted with elderly people from a nursing home.

All in all, the developed prototypes received mostly positive feedback from the participants. Valuable insights into the communication of elderly people with a speech interaction prototype could be gathered. The participants consider it possible to gain benefits in their daily lives by communicating with an assistive speech interaction system.

Further development of the system and further studies with more participants are needed to determine how elderly people evaluate a fully functional speech interaction system.

Kurzfassung

Aufgrund der Alterung der Bevölkerung wird die Verbesserung der Betreuungsmethoden für ältere Menschen von Tag zu Tag wichtiger. Die Forschung im Bereich des Ambient Assisted Living zeigt das Potenzial der Nutzung digitaler Technologien zur Erleichterung der Pflegearbeit durch intelligente Assistenzsysteme. Intelligente Geräte könnten einen großen Nutzen für ältere Menschen bieten. Diese intelligenten Geräte wie Sprachassistenten und Robotik stellen jedoch für ältere Generationen eine besondere Herausforderung in Bezug auf Interaktion und Kommunikation dar.

Das an der FH St. Pölten entwickelte Konzept "Smart Companion" zielt darauf ab, diese Herausforderungen anzugehen, indem es kostengünstige Robotersensoren (Vakuumroboter) mit intelligenter Sprachinteraktion kombiniert, um die Kommunikation zwischen Mensch und Maschine zu vereinfachen. Das Ziel dieser Thesis ist es, die Vorteile, Akzeptanz, Möglichkeiten und Grenzen eines Sprachinteraktionssystems zu erforschen und zur Entwicklung des "Smart Companion" beizutragen. Dies soll den Weg für eine konkrete Umsetzung des Konzepts ebnen.

Die Hauptmethode zur Beantwortung der Forschungsfragen ist die Entwicklung und Bewertung von Sprachinteraktionsprototypen durch Messung vordefinierter Szenarien in Wizard of Oz (WOz)-Tests und strukturierten Befragungen mit den Studienteilnehmern. Die WOz-Tests wurden mit älteren Menschen aus einem Pflegeheim durchgeführt.

Insgesamt erhielten die entwickelten Prototypen von den Teilnehmern überwiegend positives Feedback. Wertvolle Erkenntnisse über die Kommunikation älterer Menschen mit einem Sprachinteraktionsprototyp konnten gewonnen werden. Die Teilnehmer halten es für möglich, durch die Kommunikation mit einem assistiven Sprachinteraktionssystem Vorteile in ihrem täglichen Leben zu erzielen.

Eine Weiterentwicklung des Systems und weitere Studien mit mehr Teilnehmern sind notwendig, um festzustellen, wie ältere Menschen ein voll funktionsfähiges Sprachinteraktionssystem bewerten.

Table of Content

1 Introduction

Due to population aging, enhancing the methods of caring for elderly persons gets more important every day. Research in the field of ambient assisted living shows the potential of using digital technologies to facilitate caring work by providing smart assistance systems [1], [2].

Intelligent technical systems like voice assistants, smart TVs, household robotics or wearables can help us with achieving tasks faster and more conveniently. For this reason, they are increasingly spreading in our living environment. For example, around 40% of Americans already possess a smart speaker [3].

1.1 Problem

Smart devices could offer a great benefit for elderly. However, these intelligent devices such as voice assistants and robotics are particularly challenging for older generations with regards to interaction and communication [1], [4].

If a system uses robust speech detection technology and is designed in an appealing way, though, elderly people accept this system for controlling ambient assisted living environments, as a study dealing with a personal life assistant has shown [5].

Still, difficulties and challenges remain. The “Smart Companion” concept, which was developed at UoAS St. Pölten, aims at tackling those challenges by utilizing robot vacuum cleaners which are proven to have a high level of acceptance among the older generation [6] and by implementing bi-directional speech interaction to simplify communication between human beings and machines.

The purpose of my thesis is to explore the benefits, acceptance, possibilities and limitations of assistive speech interaction systems and contribute to the development of the “Smart Companion”, a mobile smart speech interaction system for elderly people. The thesis will explore the reasonableness of such a system and its acceptance among elderly alongside perceived benefits. This should pave the way for a concrete realisation of the concept.

1.2 Research questions

Based on the problems stated above, following research questions can be identified:

Main question:

- Do elderly people living in a nursing home consider it possible to gain benefits in their daily lives by communicating with an assistive speech interaction system based on testing a speech interaction prototype?

Sub questions:

- Which functionalities of a speech interaction prototype do elderly people consider as most useful?
- Do elderly people consider it possible to have an improved sense of independence through usage of an assistive speech interaction system?
- Do elderly people consider it possible to engage in more activities because of suggestions from an assistive speech interaction system?

To accomplish the goals in this thesis, an extensive literature research as well as using a combination of quantitative and qualitative research methods is done. Literature research of relevant topics like ambient assisted living and speech recognition is done in both medical and technical databases like PubMed and IEEE Xplore.

1.3 Method

The main method to answer the research questions is the development and evaluation of speech interaction prototypes via measuring predefined scenarios in Wizard of Oz (WOz) tests and structured surveys with the study participants. Although the “Smart Companion” concept includes the combination of a speech interaction system with a robot vacuum cleaner, only the speech interaction system has been prototyped for reasons of cost and time efficiency.

The prototypes were created in a user centred design process with the study participants. Additionally, the study participants answered a questionnaire about technology attitude before and after dealing with the prototypes.

1.4 Goals

The findings of this thesis will provide information on how elderly people interact with a smart voice assistant and how useful they find it to be. The results shall serve as a foundation for the realization of the “Smart Companion” concept by identifying problems, possibilities and limitations. If fully developed, the “Smart Companion” should offer elderly people a great benefit.

It is not a goal of this thesis to create a fully functional speech interaction system.

1.5 Structure

The further structure of this thesis is structured as follows:

- Chapter 2 deals with the theoretical background of all necessary areas, such as Ambient Assisted Living and speech recognition.
- Chapter 3 comprehensively describes the chosen methodology used to answer the questions.
- Chapter 4 presents all the developments, implementations, results, and evaluations that have emerged in the course of the study.
- Chapter 5 analyses the most important results of the study and draws conclusions.
- Chapter 6 presents a brief summary of the core results of the study and deals with further new questions.

2 Theoretical Background

This chapter aims at giving a theoretical introduction to the topics which are dealt with in this thesis. Central terms, which are necessary for the understanding of the thesis, are explained with the help of technical literature. The following subchapters deal with population ageing and explain the term Ambient Assisted Living. There is also an explanation of speech recognition in the context of AAL and a presentation of the “Smart Companion” concept.

2.1 Population ageing

Already in 2002, demographic change has been identified by the World Health Organization (WHO) as one of the greatest challenges facing society in Western countries. [7] This development is also subject of discussion at the European Commission. The commission has reported that in EU member states the ageing population will have an enormous impact on future economic growth and productivity. [8]

The part of the population of EU member states aged 65 years or older is expected to increase significantly across the whole EU in the coming decades. In 2010, the population part aged 65 or older was about 25%. In 2016, this number has increased to 29,6%. Up to 2050, the number is expected to rise further, until it eventually reaches 51,2% in 2070. This means that there will only be around two working-age people for every person aged over 65 in contrast to four working-age people in 2010. [9]

The ageing population will put pressure on health care systems and social care. This will lead to a reduction in the availability of care staff in the respective areas. To cope with these challenges, it is important that people stay healthy and live independently at home over a longer period of time. To help people achieving this, information and communication technologies (ICT) and assistive technologies will play an important role. A concept which aims at accomplishing this goal is ambient assisted living. [4]

2.2 Ambient Assisted Living

To approach the challenges of an ageing population in an innovative way, Ambient Assisted Living (AAL) technologies have developed in recent decades. There are many areas in which AAL technology could help older people to live and age well.

It has the potential to help individuals maintain and continue their current activities. Another area is the facilitation of continued participation in activities in the community and at home. AAL technology could also improve the quality and cost-effectiveness of social and health services [10].

When speaking of AAL, it is generally referred to the use of information and communication technologies (ICT), smart home technologies and standalone tools in a person's daily life and work environment with the goal that an individual using these technologies can live independently into old age through remaining active longer and remaining socially connected [11]. Empowerment of older people in the health care system is also increased by the growing trend of personal health monitoring. This is due to the growing understanding and awareness of health concepts and systems. Examples are mobile health applications, health monitoring systems and electronic health records [12]. AAL aims at providing a supportive home environment, which is realized through integrating intelligent interfaces and artificial intelligence, among other technologies [13].

AAL systems can also consist of computer hardware, software applications, medical sensors, wireless sensor and actuator networks (WSANs), computer networks, and databases. Ideally, all elements are connected to each other to exchange information and provide environmental assistance in an AAL setting. To transmit medical data to health monitoring systems, medical sensors and actuators are connected to the AAL applications and home gateways [14].

2.3 AAL and speech recognition

Speech is mankind's most intuitive and important method of communication. Meanwhile speech is increasingly perceived as a very effective possibility of human computer interaction. This means that spoken dialogue systems are conceived as more natural than graphic-based interfaces [15]. Human computer interaction based on speech is also one of the most convenient ways of communication. This convenience can minimize the communication gap between machine and human [16].

In recent years, large technology companies have developed smart voice assistants that already work very well. Voice assistants are essentially software programs that are able not only to capture and interpret human speech, but also to respond with synthesized voices. The most well-known voice assistants are currently Amazon's Alexa, Microsoft's Cortana, Apple's Siri, and Google's Assistant. These assistants are either integrated into computers, smartphones or specially designed speakers. These voice assistants can be used to perform simple tasks such as retrieving emails, creating lists or reminders of appointments. Users can also play music, ask questions on many topics and control home

automation devices [17]. All these things are done via voice commands. This type of interaction with technology would therefore be particularly interesting for older people, as they would neither have to look at a screen nor control anything with their hands.

Human voices change with age, however. Automatic speech recognition (ASR) devices currently available do not usually work well with older people. This is due to the fact they have been optimized for the speech characteristics of young adults to meet mainstream business needs. For example, the deterioration in automatic speech recognition (ASR) performance in older people has been illustrated by Vipperla et al [18] and Wilpon & Jacobsen [19]. Performance improves significantly, though, when automatic speech recognition is optimized specifically for older people, as the same authors showed in their studies. It can be said that there is a need to adapt ASR to the speech characteristics of older people.

Hämäläinen et al. [5] have shown in a study dealing with a personal life assistant that if a system uses robust speech detection technology and is designed in an appealing way elderly people accept this system for controlling ambient assisted living environments.

2.4 The „Smart Companion“ concept

"Smart Companion" is an exploratory project of the University of Applied Sciences St. Pölten. The "Smart Companion" concept aims to develop a platform for older people that combines everyday low-cost robot sensors with methods of smart speech recognition and speech actuators for assistance services. The innovation is generally to use and combine the data from known smart everyday technology to reduce both the inhibition threshold for personal use and the costs.

The first innovation is bi-directional, contextually relevant speech assistance. Current speech assistance systems like Amazon Alexa are purely passive and respond to questions. For this it is necessary that users know in advance what possibilities the system offers and have to think about actually questioning it. The innovation goal of "Smart Companion" extends the system and actively speaks to users. The concept described in the project proposal already provides for a concrete plan of scenarios to be created. The sensor technology of the vacuum cleaner robot collects and incorporates additional information about the context. The environment map created by the robot can be used, for example, to identify unusual obstacles that may indicate a fallen person. In "Smart Companion", this is a trigger for active contact with the user - if help is requested or no answer is given, an automatic alarm is started. An important differentiation is that "Smart Companion" is an always available addition to the nursing staff. [20]

The second innovation is acceptance through everyday technology. Technology must offer people concrete added value, but at the same time it must be accepted and realistically applicable. Experts should therefore evaluate whether the respective products are available on the market at a favorable price and are adaptable to the scenarios. This forms the basis for a realistic commercial model in follow-up projects. [20]

One of the factors why social acceptance of technology solutions is reduced among older people is if they are marketed as products for needy senior citizens and thereby convey an image of disability [21]. This is avoided by the use of everyday technology such as vacuum cleaner robots. There is even evidence that vacuum cleaner robots are "enthusiastically" accepted by seniors as they have positive effects on prioritization and other activities in daily life [6].

Vacuum cleaning robots usually integrate a wide range of sensors. Among other things, digital building maps can be created with it. "Smart Companion" controls the robot and scans the maps in real time. These are combined with the rest of the contextual knowledge to provide clues for actively triggered communication (e.g. unusual contamination from buried objects or an unforeseen obstacle that could be a fallen person). [20]

The third innovation is the emphasis on data protection and ethics. To realize the concept, a wide range of information about users and their context must be integrated. This raises questions, especially with regard to the EU General Data Protection Regulation (GDPR) [22] and the dominance of American companies in language assistance, with often intransparent data processing and storage for users (Amazon Alexa, Google Home, Apple HomePod/Siri, Microsoft Cortana). In addition, there are ethical concerns, especially in the social sciences. An integral part of the concept is therefore the evaluation of the relevant framework conditions and their classification into three categories: Harmless, questionable and blocker. This is to be investigated by an ethics board from several directions. The innovation goal is to prepare this information in a way that does not patronize senior citizens, but allows for a well-informed selection. [20]

The results of the study presented in this thesis shall serve as a foundation for the realization of the “Smart Companion” concept by identifying problems, possibilities and limitations. If fully developed, the “Smart Companion” should offer elderly people a great benefit.

3 Methodology

This chapter deals with the methodological approach used. It deals with the selection of the research method and of the target group. It also describes the study period and explains the structure of the questionnaire and surveys.

3.1 Selection of the research method

As shown in Chapter 1.2, a main research question along with three sub questions have been defined for this thesis. In order to answer the questions in the best possible way, a descriptive, explorative study design was chosen.

Speech interaction prototypes were developed and evaluated for this purpose. The development took place in a user centered design process. User centered design is a method in which later users are included from the start. This ensures that the structure, content and design of the end product are significantly influenced by the needs, expectations and understanding of the users. The product is usually conceived from the user side and visualized very early on. This means that very concrete specifications are possible for the implementation, because it is clear how the applications, processes and products should look and work [23]–[25].

The prototypes were tested in Wizard of Oz (WOz) tests. A WOz test is a method for designing user interfaces in which a system is simulated by a human being. Functions and their usability can thus be tested without the functions having already been implemented. Therefore, it is possible to test the usability of later functions at an early stage of development and to optimize them if necessary [26], [27].

This allowed scenarios to be rated on the one hand and direct feedback from participants to be obtained through structured surveys on the other. This data was then analyzed in an explorative-descriptive statistic. The strength of this method lies in its ability to graphically represent data and to create a deeper understanding of the connections [28].

In addition, a short questionnaire was created with questions about technology attitude, which was answered by the participants once before the first WOz test and once after the second WOz test. The results of the questionnaire should offer valuable insights into the usefulness and acceptance of speech assistance

technologies in the target group. Thus, they form a good basis for future projects in the AAL environment.

3.2 Target group selection

The study participants were recruited from a nursing home in January 2019.

The inclusion criteria for participating in the study were a minimum age of 60 and personal consent ability. Exclusion criteria were mental, speech and language disorders. Based on these criteria, fitting people from the nursing home were asked for their participation in the study. The persons who agreed to participate were given a detailed explanation of the subject and objectives of the study as well as of data protection and anonymization. They were then required to sign a declaration of consent in duplicate. One declaration was kept by the participants themselves, one by the study director. The ideal minimum number of participants was defined beforehand as 5, the maximum number 10. A total of 8 persons could be acquired as study participants.

Of these 8 study participants, 6 are female and 2 male. The age span is between 72 and 93.

3.3 Study period

Between 8 January and 17 January 2019, persons meeting the inclusion criteria were invited to participate in the study. From then on, the first prototype was developed.

Between 5 February and 8 February, the first Wizard of Oz Test was conducted with the study participants. After that, the development of the second prototype began.

Between 19 March and 22 March, the second Wizard of Oz Test was conducted with the study participants.

3.4 Structure of questionnaire and surveys

The short questionnaire about technology attitude consists of five closed questions. Only closed questions have been used for several reasons: they are faster to complete, they give clear opinions and they increase comparability by providing uniform answers [28].

Participants were asked the following questions:

- Are you interested in technology?
- Do you use technical devices yourself?
- Do you have confidence in technology?
- Do you find devices such as lawn mower robots or vacuuming robots useful?
- Could you imagine being supported by a robot in everyday life?

Every question could be answered with yes or no. The original German version of the questionnaire can be found in the appendix on page 57.

The structured survey for the first WOz test was designed to provide the most meaningful feedback possible for the further development of the prototype. For this purpose, a table was prepared in which the opinion of the participants on each individual scenario was asked. In order to be able to compare the rating of the scenarios by the participants, the rating was based on a simple three-step scale. A scenario could be rated as “good”, “neutral” or “bad”. In addition, there was the option of commenting on each scenario.

After evaluating the scenarios, the participants were asked four closed questions and one open question.

The closed questions were:

- Do you think that communication with a voice computer could be useful for elderly people?
- Do you think that elderly people could feel more independent by using a voice computer?
- Do you think that elderly people could participate in more activities based on suggestions from a voice computer?
- Do you think that elderly people might feel more secure using a voice computer?

These questions could be answered with yes or no. It was also possible to add a comment to each question.

The open question was:

- Do you have any other ideas what the device could say or what you could ask it?

In this way, suggestions for new scenarios should be collected. The structured survey can be found in the appendix on page 62.

After the first WOz test the structured survey was slightly changed. The rating of the scenarios was still based on a three-step scale, but the designations for the rating were changed to “useful”, “neutral” and “useless”. In addition, one scenario was omitted and three new ones added. The updated structured survey is to be found in the appendix on page 64.

4 Development and Testing

Between 8 January and 17 January 2019, persons meeting the inclusion criteria were invited to participate in the study. The persons who agreed to participate were given a detailed explanation of the subject and objectives of the study as well as of data protection and anonymization. They were then required to sign a declaration of consent in duplicate. One declaration was kept by the participants themselves, one by the study director. A total of 8 persons were acquired as study participants.

Between January and March 2019 the development and testing of the speech interaction prototype took place. This was done in a User Centered Design Process with two iterations.

In the following subchapters the development of the first prototype, the first WOz test with the participants and the development of the second prototype as well as the second WOz test are described in detail.

4.1 Development of the first prototype

Due to time and cost efficiency, it was decided that the first prototype would not yet consist of hardware or software, but of a structure for a voice user interface for different scenarios. This voice user interface should then be tested in the context of a WOz test with the study director as "Wizard". Realistic task scenarios had to be created for the first WOz experiment.

The TAALXONOMY [29], a taxonomy for effective classification of AAL products, was used to assist in the creation of the scenarios. In TAALXONOMY there are the following main areas: Health & Care, Living & Buildings, Safety & Security, Mobility & Transport, Work & Training, Vitality & Abilities, Leisure & Culture and Information & Communication. These main areas are in turn subdivided into subcategories. In many subcategories use cases are also described. Each main area, subcategory and use case is assigned a specific ID (starting with T01).

For the purposes of the prototype, subcategories and use cases from the main areas Health & Care (T01), Safety & Security (T02), Work & Training (T05), Vitality & Abilities (T06) and Information & Communication (T08) were used.

Table 1 provides a clear summary of the selected areas of TAALXONOMY and the scenarios created. After each description of a scenario the number of the scenario is shown in brackets, in the same order as it was in the following WOz test. The

“a” behind the number represents the first iteration, since the prototype will be developed further in the process.

TAALXONOMY area	Scenario
Risk prevention and disease prevention (T01-01-01)	The device asks the user if he is dizzy. (1a)
	The device asks the user if he is nauseous. (2a)
Fall detection in buildings (T03-03-01)	The device asks the user if he has fallen. (3a)
Learning and training (T05-02)	The user can ask the device if it gives him some quiz questions. (4a)
	The user can ask the device how to detect a stroke. (5a)
	The user can ask the device how to detect a heart attack. (6a)
Memory (T06-02-01)	The device asks the user about the day of the week. (7a)
	The device asks the user on which station he lives. (8a)
	The device asks the user what he ate for breakfast today. (9a)
Remembering (T06-02-02)	The device asks the user when he was born. (10a)
	The device asks the user where he was born. (11a)
	The device asks the user what his last profession was. (12a)
Information and knowledge (T08-01)	The user can ask the device what is available for lunch the next day. (13a)
	The user can ask the device what the next event in the house is. (14a)

Appointment and task management (T08-04-02)	The user can ask the device if he has any appointments on that day. (15a)

Table 1: Scenarios for first prototype

Based on these scenarios, a script was created with which the head of the study carried out the first WOz test as a "wizard". The script can be found in the appendix on page 58. The script plays through one scenario after the other. Using simple decision trees, the "Wizard" is exactly “programmed” how to react within a scenario in order to simulate a speech interaction system as realistically as possible.

For scenarios in which the users should first address the device, some exemplary statements were defined which should be recognized as the activation of the scenarios. In order to avoid frustration on the part of the participants and to document many different statements of the participants for further development, it was decided to be very tolerant regarding the "recognition rate" of the statements.

Furthermore, a structured survey was created in order to collect feedback from the participants as efficiently as possible during the first WOz test. The first step in the survey was to find out which scenarios the participants found to be useful or less useful. For this purpose, the scenarios were titled as "functions" of the device and the participants were given the opportunity to rate each scenario as good, neutral or bad. In addition, a free comment could be submitted for each individual scenario. Subsequently, some questions were asked regarding the connection of a voice interaction system with perceived usefulness, sense of independence, participation in activities and sense of security. Participants were also asked to contribute their own ideas for improving the system.

4.2 First WOz test

Between 5 February and 8 February, the first Wizard of Oz Test was conducted with the study participants.

The concrete procedure was as follows:

- The participants were alone in a quiet room with the study director.

- The head of the study explained the purpose of the test to them once again. The study director explained that they should first complete a short questionnaire and then try out the " speech computer " for the first time.

- Then the participants received the short questionnaire on their attitude to technology.

- The study leader then explained to them that, as described in the participant information, no actual computer existed at that time. Instead, the study leader himself would pretend to be the computer. This would be necessary in order to find out whether it could work as intended, without much time being spent on development.
The head of the study made sure that he was allowed to make an audio recording. The recordings were made with a Samsung Galaxy S9+ smartphone.

- Then the participants received a note with the procedure. There were alternating scenarios in which the "computer" speaks to the people and scenarios in which the people should speak to the "computer". No exact phrases were given, but rather requests like "Ask the computer what's for lunch the next day". This was done to document various statements about activating a particular scenario. The participants were allowed to read the note beforehand.

- Then the scenarios were played through one after the other. Delayed answers or difficulties with individual scenarios were either documented immediately or afterwards based on the audio recordings.

- After the Wizard of Oz test the feedback of the participants was collected by means of a structured survey.

4.2.1 Results of the questionnaire

The results of the questionnaire can be seen in Figure 1. Six out of eight participants said that they were generally interested in technology. Three participants said they used technical equipment themselves. According to the questionnaire, six of the eight participants have confidence in technology. Six participants also found equipment such as lawnmower robots or vacuuming robots useful. Only one participant can imagine being supported by a robot in everyday life.

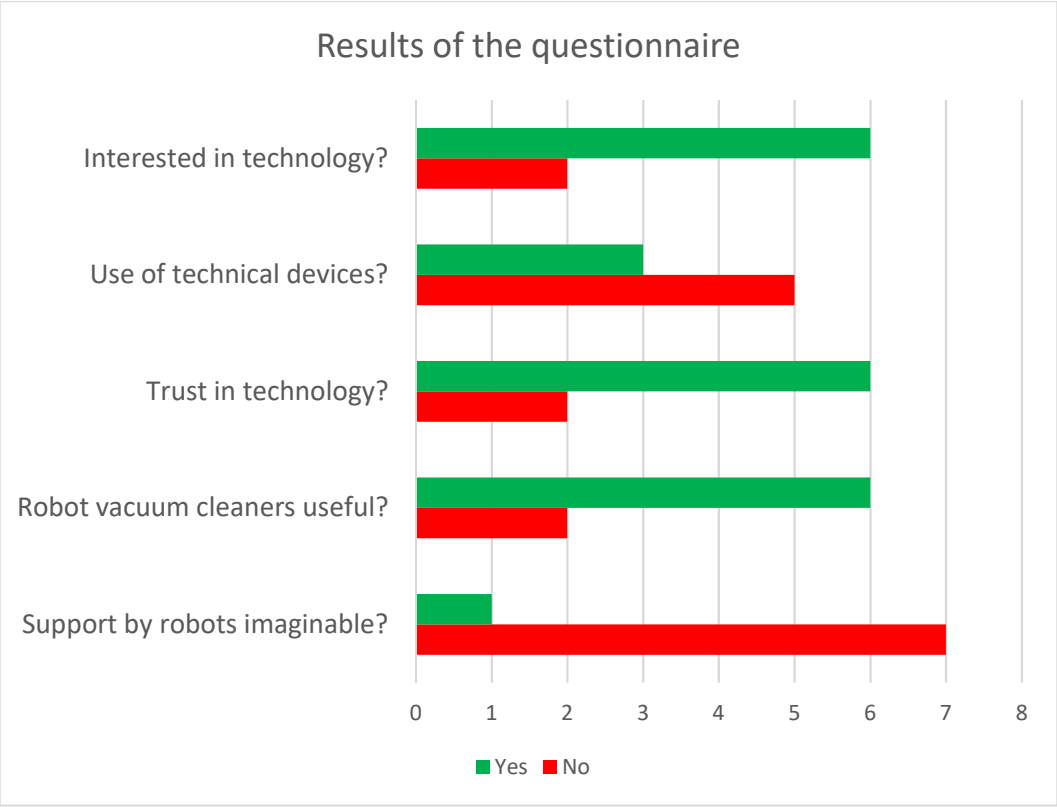

Figure 1: Results of the questionnaire

4.2.2 Feedback regarding the first prototype

Figure 2 gives an overview of the participants' feedback regarding every scenario. Green bars represent the number of participants who rated a scenario with "good", while orange bars stand for "neutral" and red ones for "bad". It can be seen that no scenario was rated "good" less than 4 times. Seven scenarios received "bad" ratings. Scenarios 1a, 2a, 4a and 9a received one "bad" rating each. The scenarios with the least positive ratings were 10a, 11a and 12a.

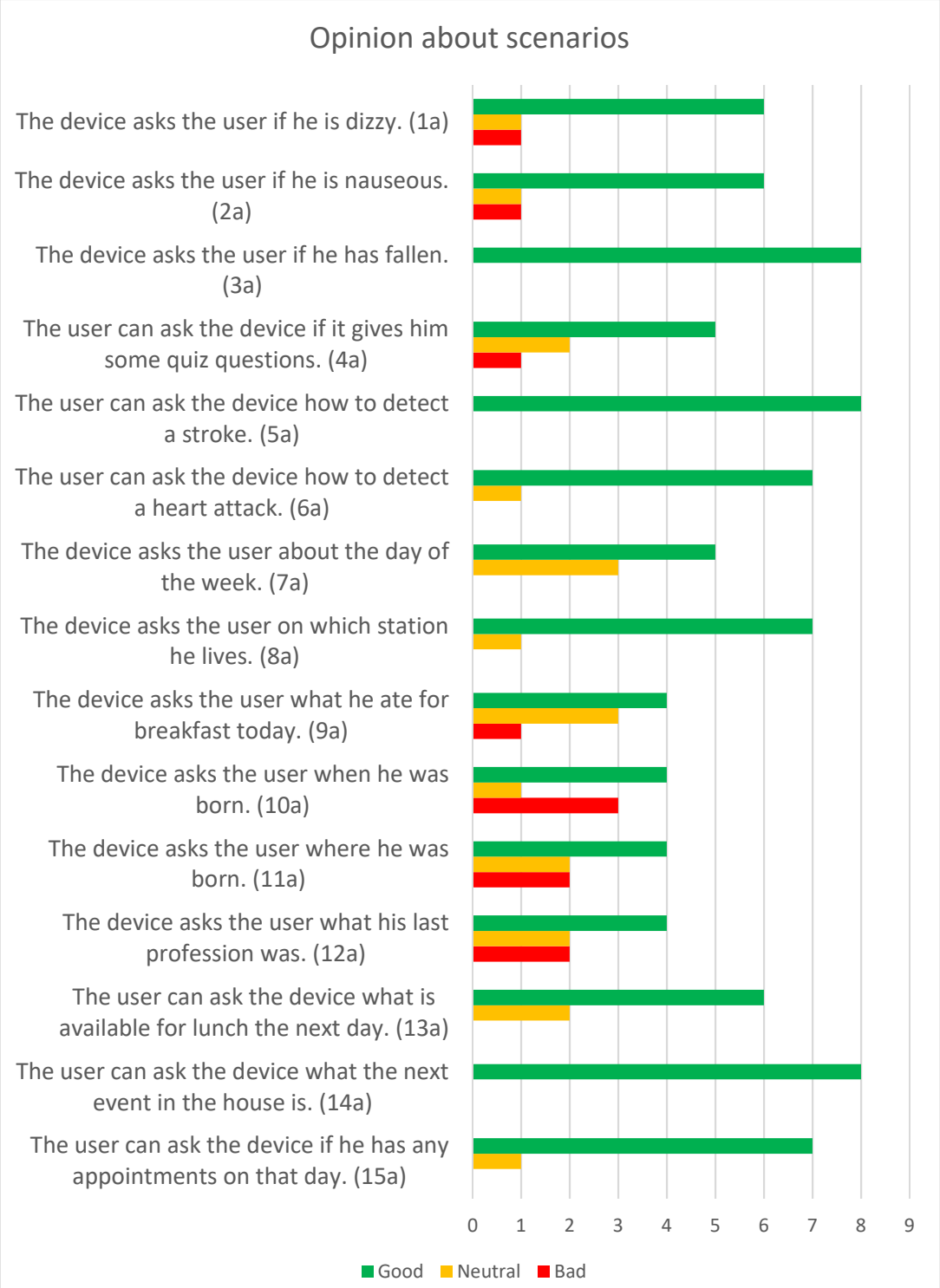

Figure 2: Feedback scenarios first prototype

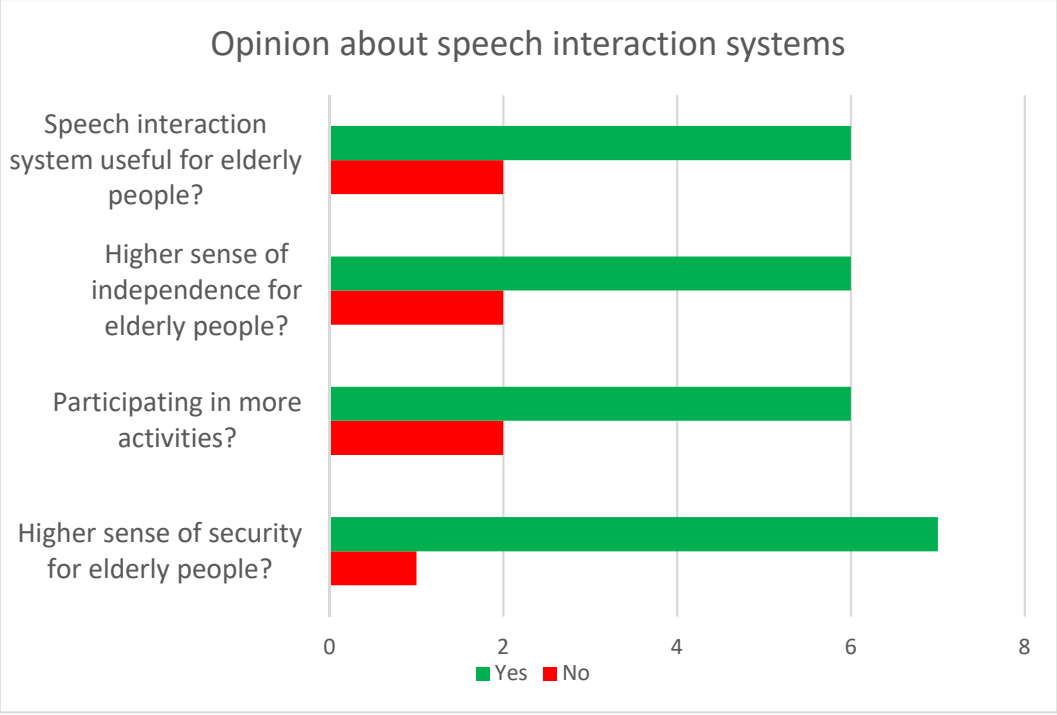

Figure 3: Opinion about speech interaction systems

The participants' opinion about speech interaction systems after the first WOz test is visualized in Figure 3. Six out of eight participants think that communication with a speech interaction system could be useful for elderly people. Also, when asked "Do you think that elderly people could feel more independent by using a voice computer?" six respondents said yes. Furthermore, six participants think that older people could participate in more activities based on suggestions from a speech interaction system. All but one participant felt that older people could have a higher sense of security by using a speech interaction system.

A total of five suggestions for new scenarios could be collected from the participants:

- You can ask the device about the weather.
- You can ask the device for the time.
- You can ask the device what day it is.
- The device informs you 10 minutes before lunch that the meal is about to arrive.
- The device informs you if a therapy is cancelled.

4.2.3 Problems and conclusions

Based on the audio recordings, some problems which occurred during the WOz Test could be identified. The frequency of problems encountered is shown in Figure 4.

The most common problem was that more than half of the participants tried to ask the study leader unrelated questions during the execution of the individual scenarios. They did this even though it was explained clearly beforehand that they should treat the study leader as a “computer”. One participant lost track of the current scenario on the note with the procedure. Some participants read requests for questions to the "computer" - which were found on the note with the procedure - directly, instead of formulating questions themselves. Another problem was that participants talked too much after a question from the “computer”, telling unrelated stories.

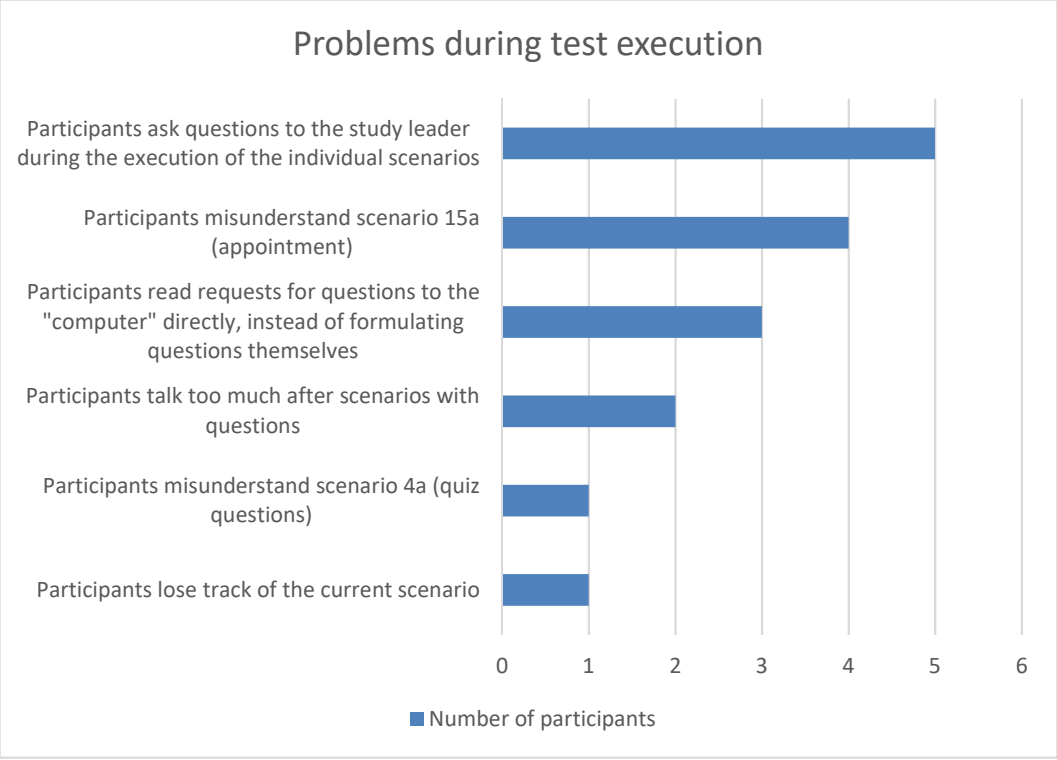

Figure 4: Problems during first WOz test

There also occurred problems with understanding certain scenarios. One participant misunderstood scenario 4a (quiz questions). It was not clear to the participant that he/she should answer the first quiz question. Half of the participants misunderstood scenario 15a (appointment). Three participants were under the impression that the scenario was about asking the “computer” if he had some appointments on this day. One participant did not understand that it was about

entering an appointment but wanted to query an appointment. The participant thought the “computer” would assign him/her an appointment.

Problems also arose during the following structured survey. Some participants thought they had to answer the written statements again (instead of evaluating the functions). Apparently, the three ways to evaluate the functions "good", "neutral" and "bad" were titled in a misleading way. Some participants perceived these terms to express their physical condition in some cases. For example, the first statement of the structured survey was "The computer asks you if you are dizzy". Some participants then looked at the three possible answers and said things like, "No, I'm not dizzy now, I'm fine. So I tick off Good". After a detailed explanation it was clear to the participants that it was about the evaluation of the functions itself, but it led to a loss of time.

As a conclusion the following things were considered for the development of the Alexa prototype:

- It would be useful to be able to start each scenario individually. Otherwise it could happen that in case of an unforeseen error the whole process does not work anymore.
- There should be the possibility to get Alexa to listen to input manually. Otherwise it could happen that Alexa is in Waiting for user input mode, while a participant says something regarding the previous scenario. For example: Alexa is waiting for input to scenario 6a (heart attack), while the participant still says "Thanks for the info about the stroke" (scenario 5a). Of course, Alexa wouldn't recognize this correctly.
- The procedure for the participants should not be on a single sheet of paper. Instead, each scenario in which the participants have to ask something actively should be on a separate card, which the participants get one after the other.
- In the scenarios where the participants have to ask something from the computer: Next time, after three failed attempts, a fixed question should be given for reading. This would at least allow the participants to end the scenario themselves, although it would not be counted as successful completion.
- In the structured questioning that follows, the terms "Useful", "Neutral" and "Useless" should be used instead of "Good", "Neutral" and "Bad". This should make it clearer to the participants that it is a question of evaluating the scenarios or functions of the device.

- Scenarios 10a, 11a and 12a (reminder questions) should be reformulated. They should be less personal. A different introduction could be chosen for this purpose. For example, "Good day, let's do a little memory training. When were you born?
- Scenario 15a (Date) should be removed due to the many difficulties encountered.
- New scenarios should be added based on feedback from participants. It should be possible to ask the device for weather, day and time.

4.3 Development of the second prototype

In order to create a prototype of a speech interaction system that was as authentic as possible, it was decided to use the automatic speech recognition system Amazon Alexa. For this purpose an Echo Dot of the 3rd generation, a smart speaker with the ability to recognize human speech and answer with synthesized voice, was acquired. For reasons of time efficiency, it was not considered necessary to program an Alexa skill with custom code. Instead, the Voiceflow service in combination with an Amazon Developer account was used to develop a prototype.

4.3.1 Voiceflow

The explanations in the following paragraph follow the explanations from the official Voiceflow website:[1]

Voiceflow is a service that lets users design and build voice apps for Alexa and Google Assistant. For the purpose of creating the prototype, the free version of Voiceflow was used. It is limited to three projects, but apart from that, the user has full access to all features.

The Voiceflow service links to the user's Amazon developer account. When a user creates a Voiceflow project and uploads it to Alexa, an interaction model is generated which acts as "front-end" of the created voice app. The voice application has its backend point on Voiceflow, which generates a logic system specific to the created app to power the interface. Figure 5 illustrates this structure in a clear way.

[1] https://university.getvoiceflow.com/what-is-voiceflow/about-voiceflow/how-voiceflow-works (accessed 24.04.2019)

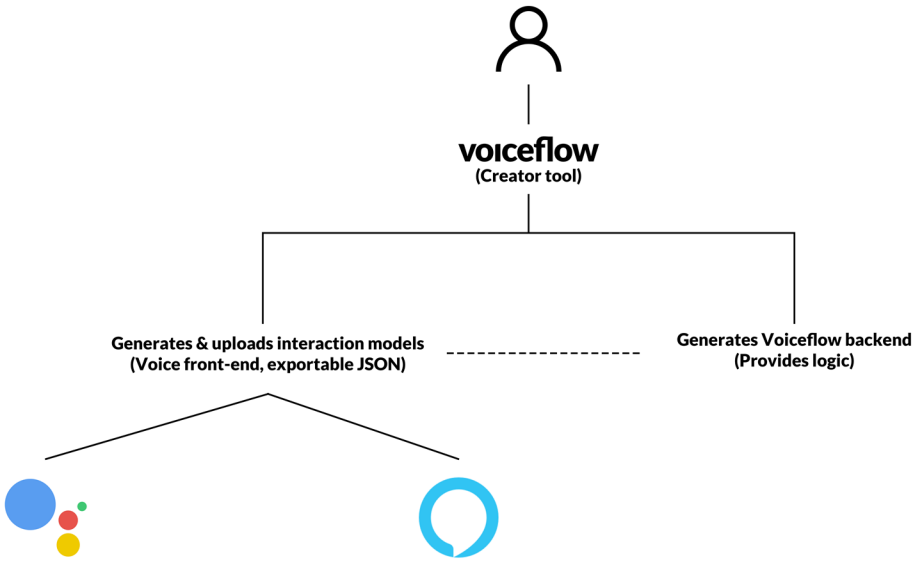

Figure 5: Explanation of Voiceflow | Source: https://university.getvoiceflow.com/what-is-voiceflow/about-voiceflow/how-voiceflow-works (accessed 24.04.2019)

Structure of Voiceflow:[2]

In Voiceflow, building a voice app is done in a visual interface. There is a so-called canvas, which is the building area where one can create a project. A project is made up of flows. Flows in turn consist of blocks. This is pictured in Figure 6.

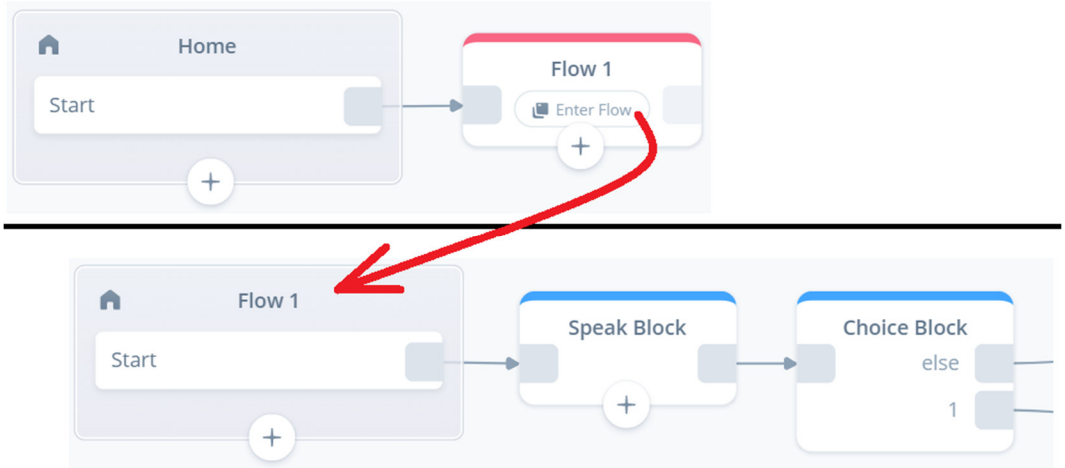

Figure 6: Flows and Blocks

[2] https://university.getvoiceflow.com/learn-the-basics-of-voiceflow/the-voiceflow-platform/voiceflow-terminology (accessed 24.04.2019)

Voiceflow also allows to add variables to the voice application, which can store data (numbers and words). The value of variables can be set using various methods. One can drag different blocks with different functionalities on the canvas and connect them to create a specific course of events in the voice app. Blocks have block ports, which are the spaces where one can drag out a connector to connect two blocks together. Most blocks have both an inbound and outbound port, but some blocks only have an outbound port. The aforementioned connectors are the lines that connect the blocks on the canvas. To successfully run a project, the blocks must be connected.

There are several types of blocks. Block types that were used for creating the voice app are:

- *Start*
  This is always the first block in every flow. It cannot be removed.
- *Speak*
  With this block, one can tell Alexa what to say or play audio clips.
- *Choice*
  A choice block is used when Alexa should listen for the user to make a choice from a list of options one can set.
- *Set*
  The Set block is used to set the value of a variable or many variables at once.
- *If*
  With this block, conditions can be set that activate paths only when true.
- *Intent*
  This block can handle intents from within the skill and upon skill launch.
- *API*
  By creating an API block, external APIs can be used and responses can be stored into variables.
- *Flow*
  This creates a new flow to organize the project into manageable sections.
- *Exit*
  This block ends the skill on the current flow.

Based on the feedback of the participants from the first WOz test, changes were made to the scenarios. Scenarios 1a – 14a were kept, with scenarios 10a -12a slightly changed. Scenario 15a was deleted. Three new scenarios were added (15b -17b). The “b” behind the scenario number stands for the second iteration. The new scenarios were about being able to query the day, time and weather from

the prototype. However, these new scenarios were not created by the study leader himself, as these functions are already very well integrated in Alexa.

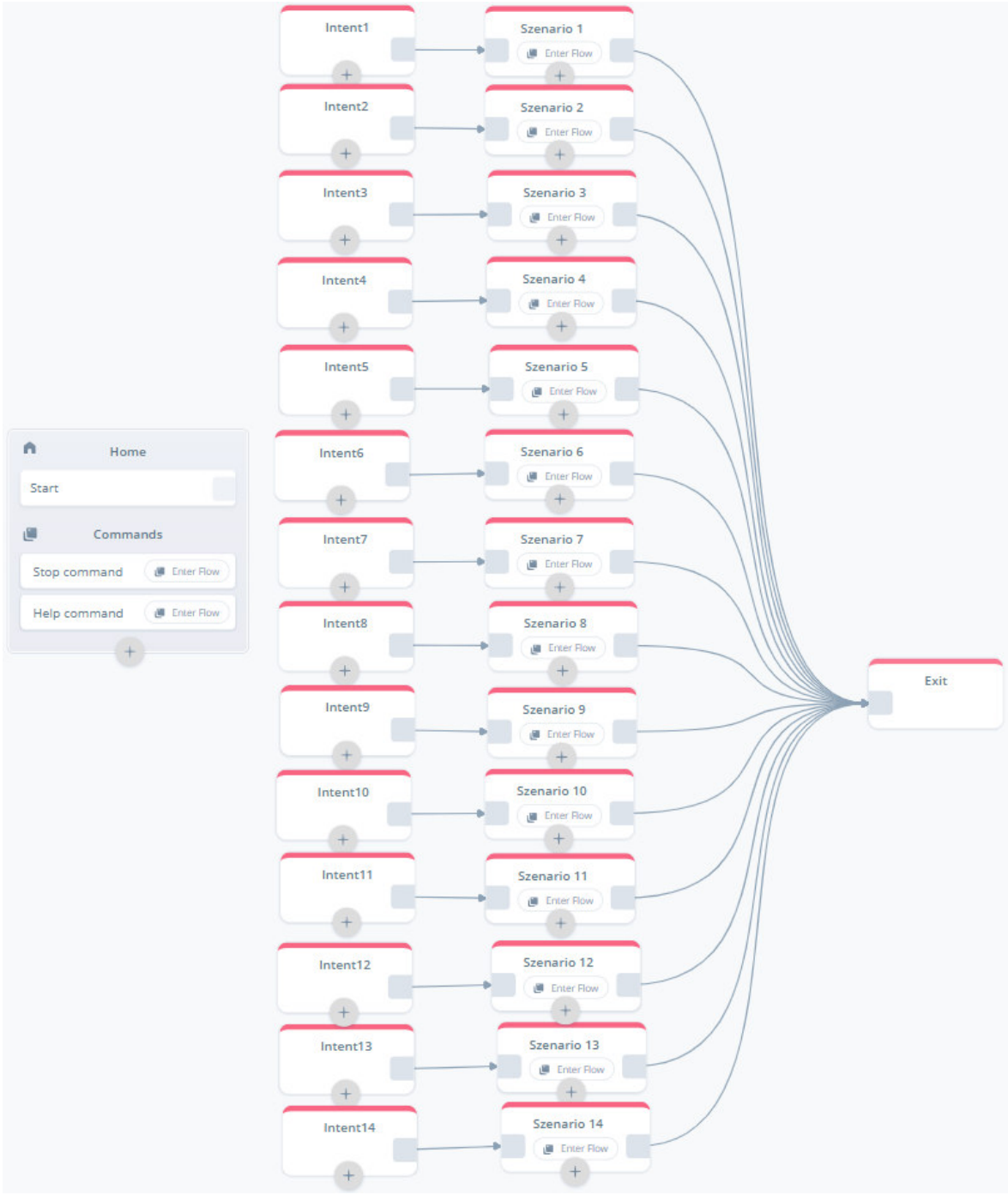

Figure 7: Scenarios in Voiceflow

It was decided to design the prototype in such a way that it can be controlled via smartphone when which scenario starts. This decision was made because not all functions of a prototype have to be fully functional yet and mainly the participants'

handling of the system needed to be documented. The possibility of manual control also ensures that the interaction with the system does not break off prematurely due to an error.

In order to accomplish this, the so-called “One shot invocation” was used. This Alexa feature allows a user to start a skill and a certain function in the skill with a single command. The input "Alexa, start Smart Companion and scenario one / two /...fifteen" jumps directly to the desired scenario. Each scenario ends the skill automatically at the end.

As shown in Figure 7, the skill was created by using 14 flows, each one representing a separate scenario. Every flows’ outbound port is connected with an Exit block, which ends the skill. Every flows’ inbound port is connected with an own Intent block. The Intent blocks serve the purpose of starting the respective flow / scenario when saying the defined trigger word. At first, trigger words were defined as “Scenario one” and “one”, “Scenario two” and “two”, etcetera. However, while testing the prototype before the second WOz test, this solution sometimes resulted in a scenario being started in the middle of another scenario due to an answer in which a number occurs. For example, when answering the question of scenario 10b, the study leader answered with “1994”. The “four” at the end of the answer triggered scenario 4b to be started. Therefore, the Austrian spelling alphabet was used for distinctive keywords that are certainly not mentioned during a scenario. The trigger word for the first Intent block was defined as “Anton”, the one for the second Intent block as “Berta”, etcetera.

Every scenario accepts certain user input, which was defined before. For every accepted input, several synonyms were defined. These synonyms were mostly chosen because of the participants’ answers in the respective scenario in the first WOz Test. The aim was to achieve an optimal understanding of the participants' inputs. This is also the reason why the number of synonyms varies from scenario to scenario - in some scenarios at the first WOz test the answers of the participants were often the same or very similar. Therefore, for such scenarios in Voiceflow not as many synonyms were created as for other scenarios.

The following subchapters contain detailed descriptions of the creation of the scenarios. The scenarios are arranged according to the corresponding TAALXONOMY areas, including the corresponding ID.

4.3.2 Scenarios regarding risk prevention and disease prevention (T01-01-01)

Scenario 1b was created using a combination of *Speak* blocks and a *Choice* block, as shown in Figure 8. With the first *Speak* block, Alexa asks the user “Do you feel dizzy?” (“Fühlen Sie sich schwindlig?”). After the question, the *Choice* block is activated, with Alexa waiting for user input. Choice 1 is “yes” with six synonyms. It is connected with a *Speak* block which makes Alexa say “It's best to call a nurse right away. Pay more attention than usual, especially when walking.” (“Rufen Sie am besten gleich eine Pflegekraft. Passen Sie besonders beim Gehen mehr auf als sonst.“) Choice 2 is “no”, also with six synonyms. It leads to another *Speak* block, which contains the text “That's nice! Enjoy your day.” (“Das ist schön! Genießen Sie Ihren Tag.”) The else option is triggered when neither Choice 1 nor 2 are activated by the user’s answers. It is connected with a *Speak* block that makes Alexa say “I'm afraid I didn't get that. Do you feel dizzy?“ (“Das habe ich leider nicht verstanden. Fühlen Sie sich schwindlig?“) This *Speak* block leads back to the *Choice* block.

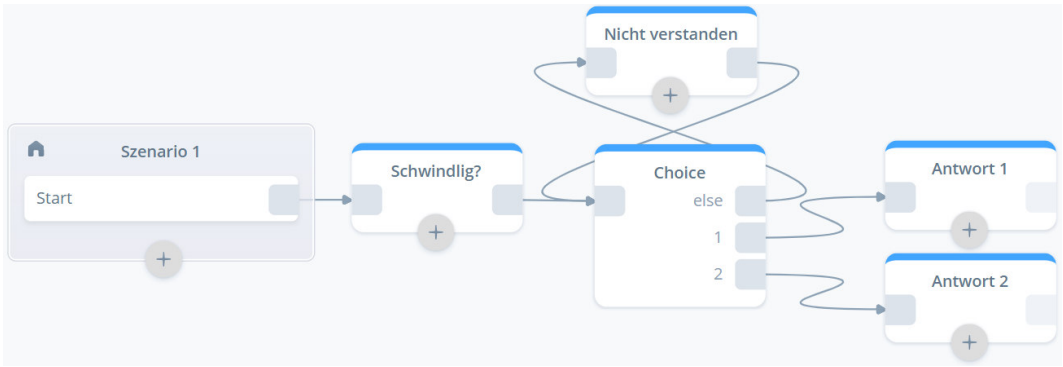

Figure 8: Voiceflow scenario 1

Scenario 2b was created similar to scenario 1b. *Speak* and *Choice* blocks were likewise used, as illustrated in Figure 9. The first *Speak* block contains the text “Do you feel sick?” (“Ist Ihnen übel?”) The flow continues with a *Choice* block with two answer options and an else option. Choice 1 is triggered by “yes” or one of the synonyms and leads into a *Speak* block which makes Alexa say “It is best to contact a nurse right away. You may need a medicine.“ (“Melden Sie sich am besten gleich bei einer Pflegekraft. Eventuell brauchen Sie ein Medikament.“) Choice 2 is activated when the user says „no“ or one of the 4 synonyms. It continues with a *Speak* block that contains the text “That's nice! Enjoy your day.” (“Das ist schön! Genießen Sie Ihren Tag.“) The else option is triggered when neither Choice 1 nor 2 are activated by the user’s answers. It is connected with a *Speak* block that makes Alexa say “I'm afraid I didn't get that. Do you feel sick?“

(“Das habe ich leider nicht verstanden. Ist Ihnen übel?“) The *Speak* block’s outbound port is connected with the inbound port of the *Choice* block.

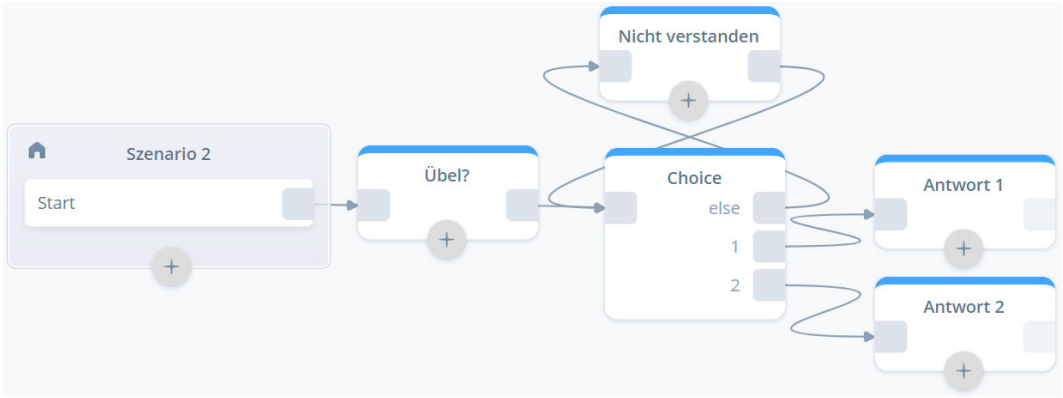

Figure 9: Voiceflow scenario 2

4.3.3 Scenarios regarding fall detection in buildings (T03-03-01)

For scenario 3b, a combination of *Speak*, *Choice*, *Set* and *If* blocks as well as a variable was used. Figure 10 illustrates this in a comprehensible way. The flow starts with a *Speak* block that asks the user “I detected an object on the ground. Did you fall?“ (“Ich habe ein Objekt am Boden erkannt. Sind Sie gestürzt?“) It continues with a *Choice* block. There are three possible options. When the user answers with “yes” or one the four synonyms, a *Speak* block is activated which makes Alexa say “All right. I'll set off an alarm. Someone will come to you in a moment.“ (“Alles klar. Ich löse einen Alarm aus. Es kommt gleich jemand zu Ihnen.“) When the user says „no“ or one of the three synonyms, the flow continues with a *Speak* block containing the text “All right. Thanks for the answer.” (“Alles klar. Danke für die Antwort.“)

If the user gives an incomprehensible answer, the else option is activated, leading to a *Set* block. The *Set* block adds 1 to the value of the variable *time*, which was created beforehand with a starting value of 0. The flow continues with an *If* block which checks if the value of the variable *time* equals 3. If it does not, it leads to a *Speak* block containing the text “I'm afraid I didn't get that. Did you fall?” (“Das habe ich leider nicht verstanden. Sind Sie gestürzt?“), which in turn leads into the *Choice* block again. The user gets two more chances of answering the question with synonyms of “yes” or “no”. If the user gives two more incomprehensible answers, the value of the variable *time* equals 3, which causes the *If* block to lead into a *Speak* block that says “I'm going to set off an alarm now.” (“Ich löse jetzt einen Alarm aus.”)

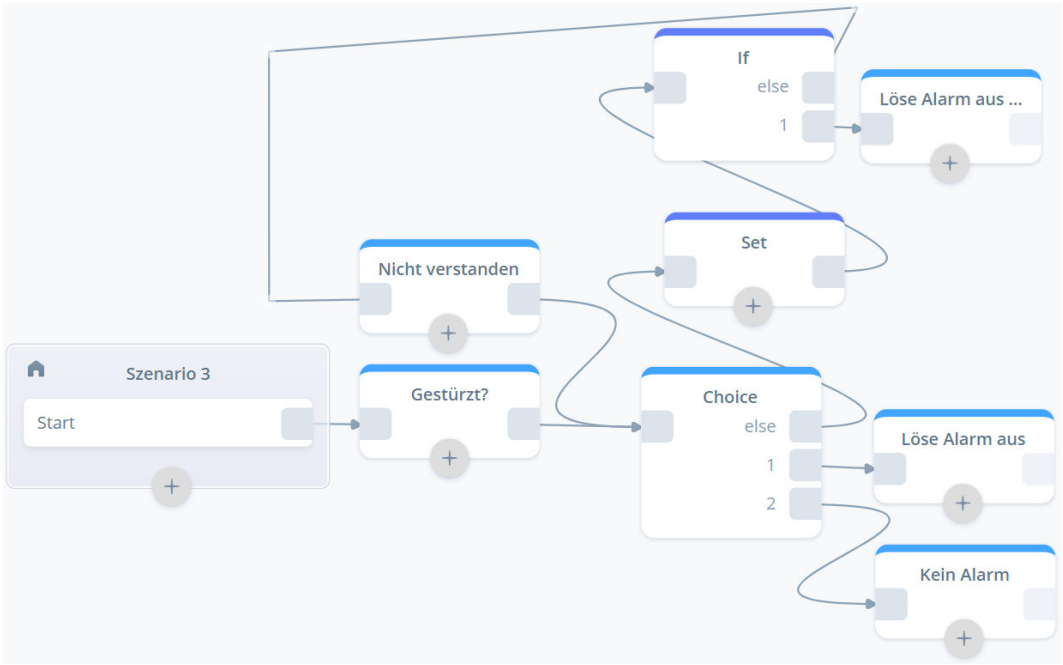

Figure 10: Voiceflow scenario 3

4.3.4 Scenarios regarding learning and training (T05-02)

Scenario 4b utilizes *Speak* and *Choice* blocks to ask the user two quiz questions. The combination of blocks is illustrated in Figure 11. The flow begins with a *Speak* block greeting the user with the text “Hello, how can I help you?” (“Hallo, wie kann ich helfen?”) and continues with a *Choice* block. The user has the possibility to ask for quiz questions with “I'd like to have some quiz questions.” (“Ich möchte ein paar Quizfragen”) or twelve synonyms. If he / she does so unsuccessfully due to an incomprehensible question, the else option of the *Choice* block leads into a *Speak* block which tells the user “I'm afraid I didn't get that. What would you like?“ (“Das habe ich leider nicht verstanden. Was möchten Sie?“) The *Speak* block connects with the inbound port of the *Choice* block, giving the user another chance of activating the scenario. If the user succeeds to trigger Choice 1, the flow continues with a *Speak* block which asks him or her the first quiz question: “Do you think London has more inhabitants than Vienna?” (“Glauben Sie, dass London mehr Einwohner hat als Wien?”)

The *Speak* block leads into a *Choice* block. The user can answer with variations of “yes” and “no”, each having four synonyms. If Alexa does not receive a variation of “yes” or “no” from the user, the else option is activated which leads into a *Speak* block saying “I'm afraid I didn't get that. Do you think London has more inhabitants than Vienna?” (“Das habe ich leider nicht verstanden. Glauben Sie, dass London mehr Einwohner hat als Wien?“) and leading into the *Choice* block again. Choice 1 continues with a *Speak* block that congratulates the user with the text “That's right! London has four times as many inhabitants as Vienna.” (“Richtig! London hat

viermal so viele Einwohner wie Wien.“) Choice 2 leads into a *Speak* block that tells the user “I'm afraid that's not true. London has four times as many inhabitants as Vienna.” (“Das stimmt leider nicht. London hat viermal so viele Einwohner wie Wien.“) Both *Speak* blocks connect with the inbound port of another *Speak* block containing the next quiz question.

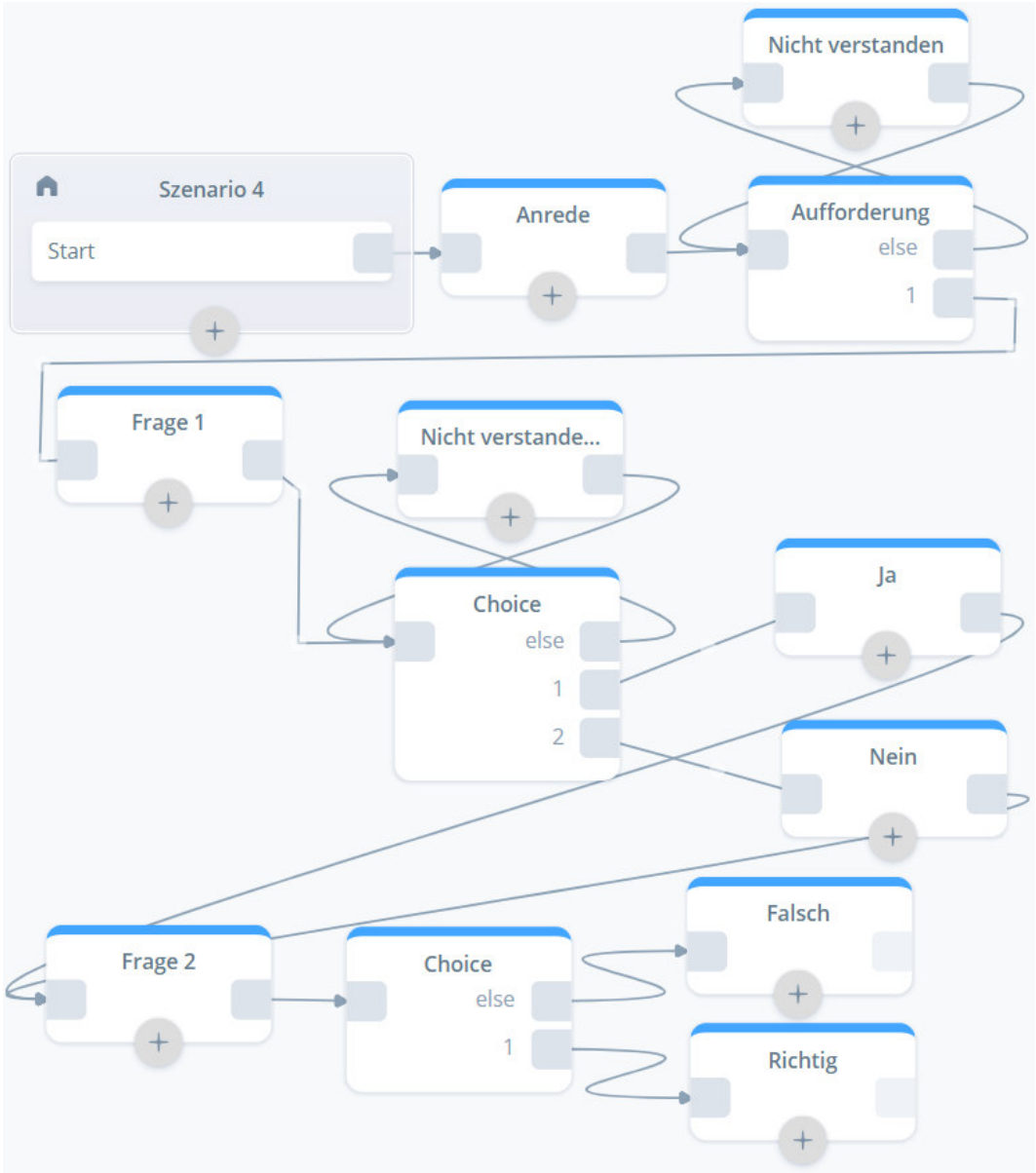

Figure 11: Voiceflow scenario 4

The *Speak* block with the second quiz question contains the text “<break time="2s"/> Next question. In which year did the First World War end?” (“Nächste Frage. In welchem Jahr endete der Erste Weltkrieg?“) The phrase “<break time="2s"/> is an SSML tag which represents a pause of two seconds in the

speech. Without this SSML tag, Alexa would read the text from the previous *Speak* block(s) and the text from this *Speak* block without any break.[4]

After the *Speak* block, the flow continues with a *Choice* block. Option 1 is triggered when the user says “nineteenhundredandeighteen” (“neunzehnhundertachtzehn”) or one of the four synonyms and leads into a *Speak* block containing the text “That’s right!” (“Das stimmt!”) The synonyms do not allow much tolerance in terms of the year. Synonyms are “das war neunzehnhundertachtzehn”, “das war neunzehnachtzehn”, “achtzehn” and “neunzehnachtzehn”. This means that users can formulate the sentence in various ways and say the year in different ways, but the year itself has to be correct. If the user answers with anything incorrect, the else option is activated and leads into a *Speak* block that tells the user “I'm afraid that's not true. Nineteen hundred and eighteen would have been right.” (“Das stimmt leider nicht. Neunzehnhundertachtzehn wäre richtig gewesen.“)

The flow for scenario 5b contains *Speak* and *Choice* blocks, as can be seen in Figure 12. The first *Speak* block asks the user "Hello, how can I help?” (“Hallo, wie kann ich helfen?“) The user has the possibility to ask for information about a stroke with “How can I recognize a stroke?” (“Wie kann ich einen Schlaganfall erkennen?”) or seven synonyms. If successful, option 1 is triggered which leads into a *Speak* block containing the text „A stroke can cause various symptoms, such as impaired vision and speech, headaches, dizziness, tingling in the arms and legs, numb fingers or lips, swallowing difficulties and facial paralysis. If you experience any of these symptoms suddenly, contact a doctor or nurse immediately.” (“Bei einem Schlaganfall können verschiedene Symptome auftreten, zum Beispiel Seh- und Sprachstörungen, Kopfschmerzen, Schwindel, Kribbeln in Armen und Beinen, taube Finger oder Lippen, Schluckbeschwerden und Gesichtslähmungen. Sollten einige dieser Symptome plötzlich bei Ihnen auftreten, sollten Sie sich sofort an eine Ärztin oder Pflegeperson wenden.“) If the user’s answer is incomprehensible, the else option is activated, which leads to a *Speak* block saying “I'm afraid I didn't get that. What would you like to know?” (“Das habe ich leider nicht verstanden. Was möchten Sie wissen?“), which in turn leads back to the inbound port of the *Choice* block.

[4] https://developer.amazon.com/docs/custom-skills/speech-synthesis-markup-language-ssml-reference.html#break (accessed 24.04.2019)

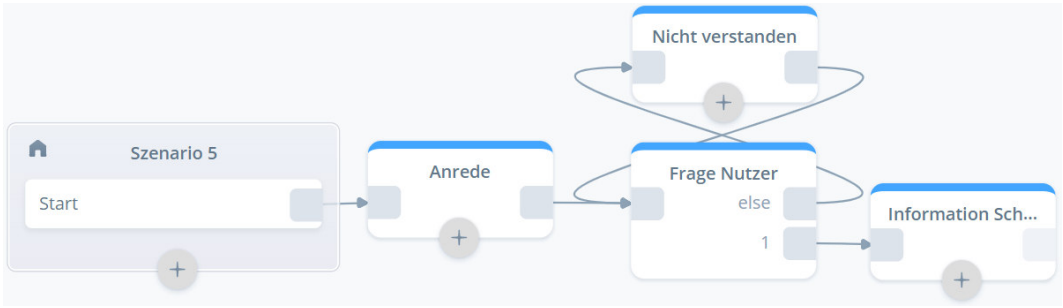

Figure 12: Voiceflow scenario 5

The flow for scenario 6b contains *Speak* and *Choice* blocks, as illustrated in Figure 13. The first *Speak* block asks the user “Hello, how can I help you?” (“Hallo, wie kann ich helfen?“) The user can ask for information about a heart attack with “How can I detect a heart attack?” (“Wie kann ich einen Herzinfarkt erkennen?”) or five synonyms. If successful, option 1 is triggered which leads into a *Speak* block containing the text „A heart attack can cause various symptoms, such as tightness, intense pressure or burning sensation in the chest. A pale facial colour and cold sweat on the forehead and upper lip can also be warning signs. If you experience any of these symptoms suddenly, you should consult a doctor or nurse immediately.” (“Bei einem Herzinfarkt können verschiedene Symptome auftreten, zum Beispiel Engegefühl, heftiger Druck oder Brennen im Brustkorb. Auch eine blasse Gesichtsfarbe und kalter Schweiß auf Stirn und Oberlippe können Warnzeichen sein. Sollten einige dieser Symptome plötzlich bei Ihnen auftreten, sollten Sie sich sofort an eine Ärztin oder Pflegeperson wenden.“)

If the user’s answer is incomprehensible, the else option is activated, which leads to a *Speak* block saying “I'm sorry, I didn't get that. What would you like to know?“ (“Tut mir leid, das habe ich leider nicht verstanden. Was möchten Sie wissen?“), which in turn leads back to the inbound port of the *Choice* block.

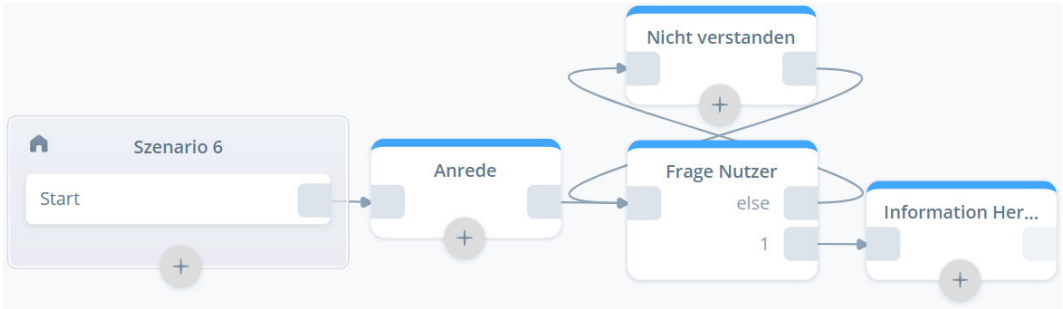

Figure 13: Voiceflow scenario 6

4.3.5 Scenarios regarding Memory (T06-02-01)

Scenario 7b utilizes *Speak*, *Choice* and *API* blocks as well as a variable. The arrangement of the blocks can be seen in Figure 14. The scenario starts with a *Speak* block asking the user “What day of the week is today?” (“Welcher Wochentag ist heute?”) The *Speak* block connects with an *API* block. The *API* block makes an API call to a Google Sheet.

The Google Sheet was created beforehand. It contains two cells. The first cell (A1) contains the excel function =TODAY(), which returns the current date. The second cell (A2) contains the function =TEXT(A1; "dddd"), which converts the date value from the first cell into a text value and displays the day name. In the *API* block, it was defined to get the text value from the A2 cell in the Google Sheet and store the value in a variable. For this purpose, the variable *day* was created. By utilizing the combination of an *API* block, a Google Sheet and a variable, it is possible to make Alexa say the current day name by using the variable *day*.

The flow continues with a *Choice* block with two options. If the user says “I don’t know” (“Weiß ich nicht”) or one of three synonyms, option 2 is triggered and leads into a *Speak* block containing the text “It doesn't matter. Today is {day}.” (“Macht nichts. Heute ist {day}.”) By placing the variable in the text, Alexa says the text value stored inside the variable, which is the current day name in this case. If the user says the correct day, option 1 is activated. In this case, the flow leads into a *Speak* block which confirms “That's right! You have a good memory.” (“Das stimmt! Sie haben ein gutes Gedächtnis.”) If the user says an incorrect day, the else option is triggered and leads into a *Speak* block saying “I'm afraid that's not true. Today is {day}.” (“Das stimmt leider nicht. Heute ist {day}.“)

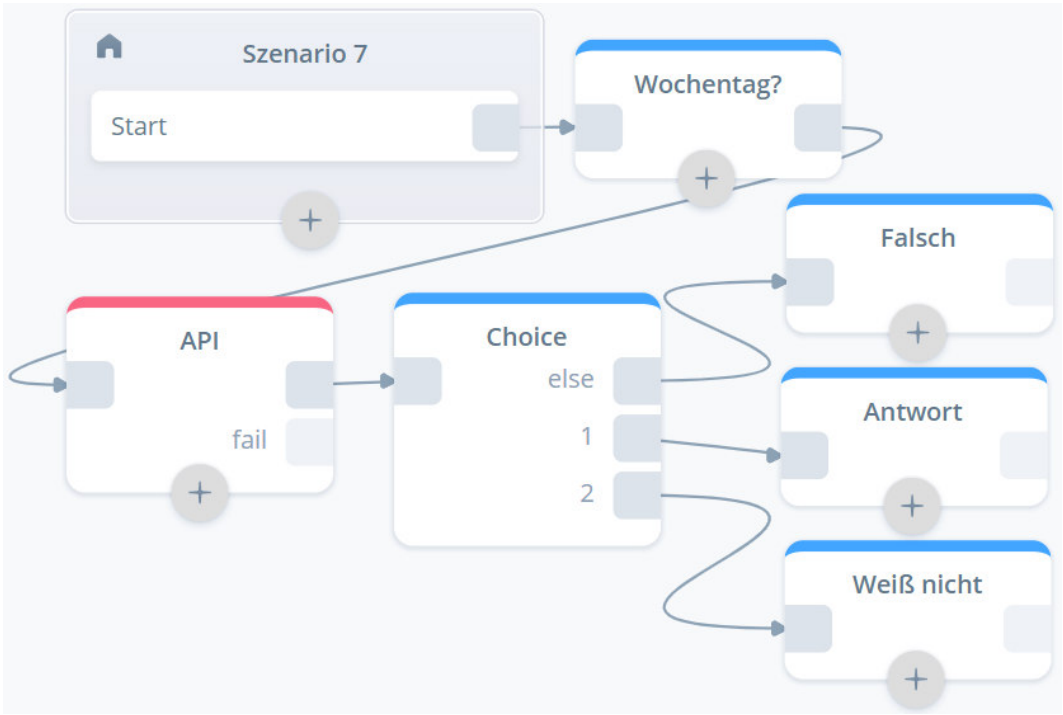

Figure 14: Voiceflow scenario 7

In scenario 8b, *Speak* and *Choice* blocks are used. Figure 15 illustrates the arrangement of the blocks. The flow starts with a *Speak* block. Alexa asks the user “What station do you live on?” (“Auf welcher Station wohnen Sie?“) The flow continues with a *Choice* block. To activate option 1, the user has to answer with “I live on…” (“Ich wohne auf…”) and the name of a station. For the purpose of the prototype, every station name was defined as synonym, since a personalized query is not possible with this setup. Upon activation of option 1, it continues with a *Speak* block saying “Great! You know your stuff.” (“Super! Sie kennen sich aus.“) If the user says “I don’t know” („Ich weiß nicht“) or one of the two synonyms, it leads to a *Speak* block saying “It doesn't matter. Maybe it will come back to you later.” („Das macht nichts. Vielleicht fällt es Ihnen später wieder ein.“) If the user gives an incomprehensible answer, the flow leads into a *Speak* block containing the text “I'm afraid I didn't get that. On which station do you live?” (“Das habe ich leider nicht verstanden. Auf welcher Station wohnen Sie?“), which in turn leads back into the *Choice* block.

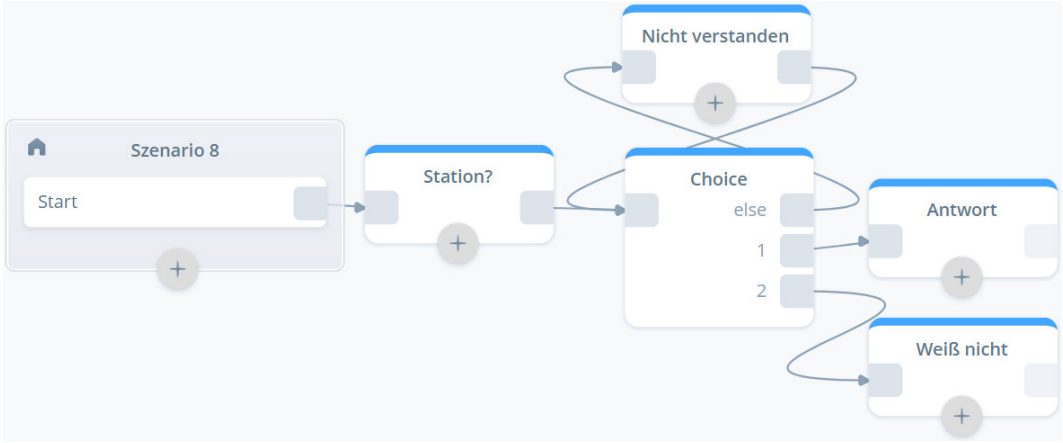

Figure 15: Voiceflow scenario 8

As can be seen in Figure 16, the flow for scenario 9b contains *Speak* and *Choice* blocks. The first *Speak* block asks the user „What did you have for breakfast today?“ („Was haben Sie heute zum Frühstück gegessen?“) The flow continues with a *Choice* block with one option and an else option. If the user says “I don’t know” (“Weiß ich nicht”) or one of three synonyms, option 1 is triggered and a *Speak* block saying “It doesn't matter. Maybe you'll remember it later.” (“Macht nichts. Vielleicht fällt es Ihnen später ein“) is activated. If the user says anything else, the flow leads into a *Speak* block that makes Alexa say “That sounds good! I hope it tasted good.” (“Das klingt gut! Ich hoffe, es hat geschmeckt.“)

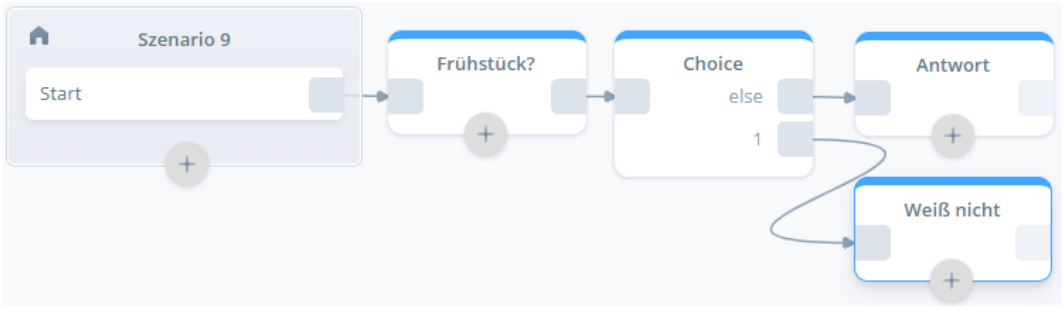

Figure 16: Voiceflow scenario 9

4.3.6 Scenarios regarding remembering (T06-02-02)

As illustrated in Figure 17, the flow for scenario 10b contains *Choice* and *Speak* blocks. The first *Speak* block asks the user „Let's do a little remembering training. In which year were you born?“ (“Machen wir ein kurzes Erinnerungstraining. In welchem Jahr sind Sie geboren?”) The flow continues with a *Choice* block with one option and an else option. If the user says “I don’t know” (“Weiß ich nicht”) or one

of two synonyms, option 1 is triggered and a *Speak* block saying “It doesn't matter. Maybe you'll remember it later.” (“Macht nichts. Vielleicht fällt es Ihnen später ein.“) is activated. If the user says anything else, the flow leads into a *Speak* block that makes Alexa say „Not bad, so you have some life experience.” (“Nicht schlecht, Sie haben also einiges an Lebenserfahrung.“)

Figure 17: Voiceflow scenario 10

Scenario 11b uses a combination of *Speak* and *Choice* blocks. The arrangement of the blocks can be seen in Figure 18. The first *Speak* block contains the question „Let's do a little remembering training. In which place were you born?“ (“Machen wir ein kurzes Erinnerungstraining. In welchem Ort sind Sie geboren?“) The flow leads into a *Choice* block with two options. If the user answers with “I don’t know” (“Weiß ich nicht”) or one of two synonyms, option 1 is activated which leads into a *Speak* block saying “It doesn't matter. Maybe it will come back to you later.” (“Macht nichts. Vielleicht fällt es Ihnen später ein.“) If the user answers with combinations of “Wien”, “Graz”, “Brünn”, “bin geboren”, “in”, “ich komme aus”, option 2 is activated. These trigger words were chosen because of the participants’ answers in this scenario in the first WOz Test. The activation of option 2 leads into a *Speak* block that congratulates the user with “All right. You have a good memory.” (“Alles klar. Sie haben ein gutes Erinnerungsvermögen.“)

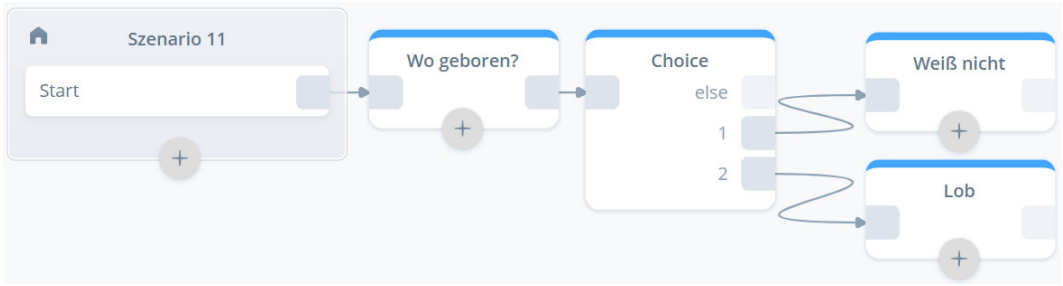

Figure 18: Voiceflow scenario 11

As illustrated in Figure 19, the flow for scenario 12b contains *Choice* and *Speak* blocks. The first *Speak* block asks the user „Let's do a little remembering training.

What was your last profession?“ (“Machen wir ein kurzes Erinnerungstraining. Welchen Beruf haben Sie zuletzt ausgeübt?”) The flow continues with a *Choice* block with one option and an else option. If the user says “I don’t know” (“Weiß ich nicht”) or one of two synonyms, option 1 is triggered and a *Speak* block saying “It doesn't matter. Maybe you'll remember it later.” (“Macht nichts. Vielleicht fällt es Ihnen später ein“) is activated. If the user says anything else, the flow leads into a *Speak* block that makes Alexa tell the user “I understand. You have a good memory.” (“Ich 36erstehen. Sie haben ein gutes Erinnerungsvermögen.“)

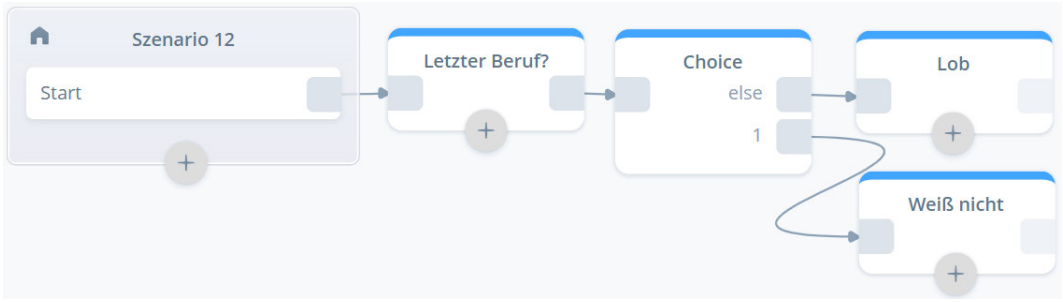

Figure 19: Voiceflow scenario 12

4.3.7 Scenarios regarding information and knowledge (T08-01)

Scenario 13b utilizes a combination of *Speak* and *Choice* blocks. Figure 20 shows how the blocks are connected with each other. The flow begins with a *Speak* block greeting the user with the text “Hello, how can I help you?” (“Hallo, wie kann ich helfen?”) and continues with a *Choice* block. The user has the possibility to ask for the meal of the next day with “What's for lunch tomorrow?” (“Was gibt es morgen zum Mittagessen?”) or six synonyms. If he / she does so unsuccessfully due to an incomprehensible question, the else option of the *Choice* block leads into a *Speak* block which tells the user “I'm afraid I didn't get that. What would you like to know?” (“Das habe ich leider nicht verstanden. Was möchten Sie wissen?“) The *Speak* block connects with the inbound port of the *Choice* block, giving the user another chance of activating the scenario. If the user succeeds to trigger option 1, the flow continues with a *Speak* block which tells him / her “Tomorrow we have schnitzel with fried vegetables and mashed potatoes for lunch.” (“Morgen gibt es zum Mittagessen Schnitzel mit Bratgemüse und Kartoffelpürree.”)

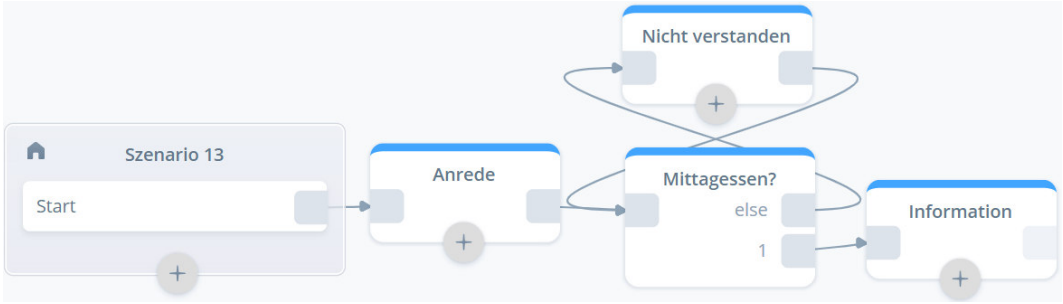

Figure 20: Voiceflow scenario 13

In Scenario 14b, *Speak* and *Choice* blocks are combined. Figure 21 shows the way the blocks are connected. The scenario starts with a *Speak* block greeting the user with the text “Hello, how can I help you?” (“Hallo, wie kann ich helfen?”) The *Speak* block leads into a *Choice* block. The user can ask for the next event in the house. Variations of the phrase “What's the next event here at the house?” (“Was ist die nächste Veranstaltung hier im Haus?”) and of six synonyms are accepted. If the user is not successful, the else option of the *Choice* block leads into a *Speak* block which tells the user “I'm afraid I didn't get that. What would you like to know?” (“Das habe ich leider nicht verstanden. Was möchten Sie wissen?“)

The *Speak* block leads back into the *Choice* block, which gives the user another chance for receiving the desired information. If the user succeeds to trigger option 1, the flow continues with a *Speak* block which tells him / her “The next event here in the house is the Mother's Day concert on the ninth of May.” (“Die nächste Veranstaltung hier im Haus ist das Muttertagskonzert am neunten Mai.”)

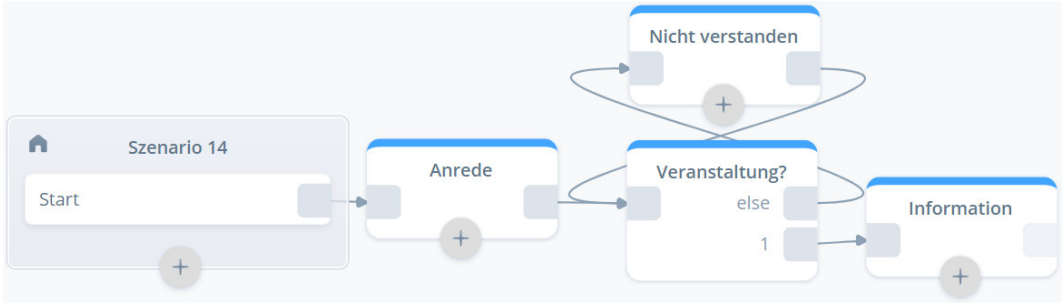

Figure 21: Voiceflow scenario 14

4.4 Hardware

On the hardware side, the following setup was used:

A smartphone is wirelessly connected to Bluetooth headphones. One of the ear pieces of the headphones is attached directly to one of the microphone openings of the Echo Dot used. With the smartphone, voice commands can be played via the Bluetooth headphones. These voice commands are almost inaudible to people without hearing impairment at a distance of 1 meter from the Echo Dot, but the Echo Dot can recognize the command. The voice commands trigger various scenarios manually.

Figure 22: Alexa protoype - Echo Dot with Bluetooth headphones

4.5 Second WOz-Test

Between 19 March and 22 March, the second Wizard of Oz Test was conducted with the same study participants.

The concrete procedure was as follows:

- The participants were alone in a quiet room with the study director.

- The head of the study explained the purpose of the test to them once again. The study director explained that they would first try out the speech computer prototype and complete a short questionnaire afterwards.

- The study leader showed the participants the Echo Dot prototype and explained that this time they would only speak with the device during the test, while the study leader would sit in the background and observe or take notes. The head of the study made sure again that he could make an audio recording. Figure 23 shows the setup for the second Woz test.

Figure 23: Setup for second WOz test

- The study director explained to the participants that they should answer the device in the most natural way possible if it spoke to them. If they had to ask the device a question, the study director would give them a card with a note telling them what to ask directly before. The cards did not give an exact phrase, but rather a prompt like "Ask the computer what's for lunch the next day". This was intended to check if a particular scenario could be activated by multiple, unforeseen phrases.

In the last three scenarios (day, time, weather), the cards indicated that the device was to be addressed with "Computer". This had to do with the fact that these three scenarios were not included in the actual skill but are functions that are already very well integrated in Alexa itself. "Computer" was used as trigger instead of "Alexa".

The reason for the change regarding single cards is that some participants were slightly overwhelmed at the beginning of the first WOz test, when they received the piece of paper with the whole procedure on it.

- Then the scenarios were played through one after the other. Successful completion of the scenario, required attempts, difficulties with individual scenarios, etc. were either documented immediately or afterwards based on the audio recordings.

- After the WOz test the feedback of the participants was collected via a structured survey.

- Finally, they got the short questionnaire about the attitude to technology for the second time.

4.5.1 Results of the questionnaire

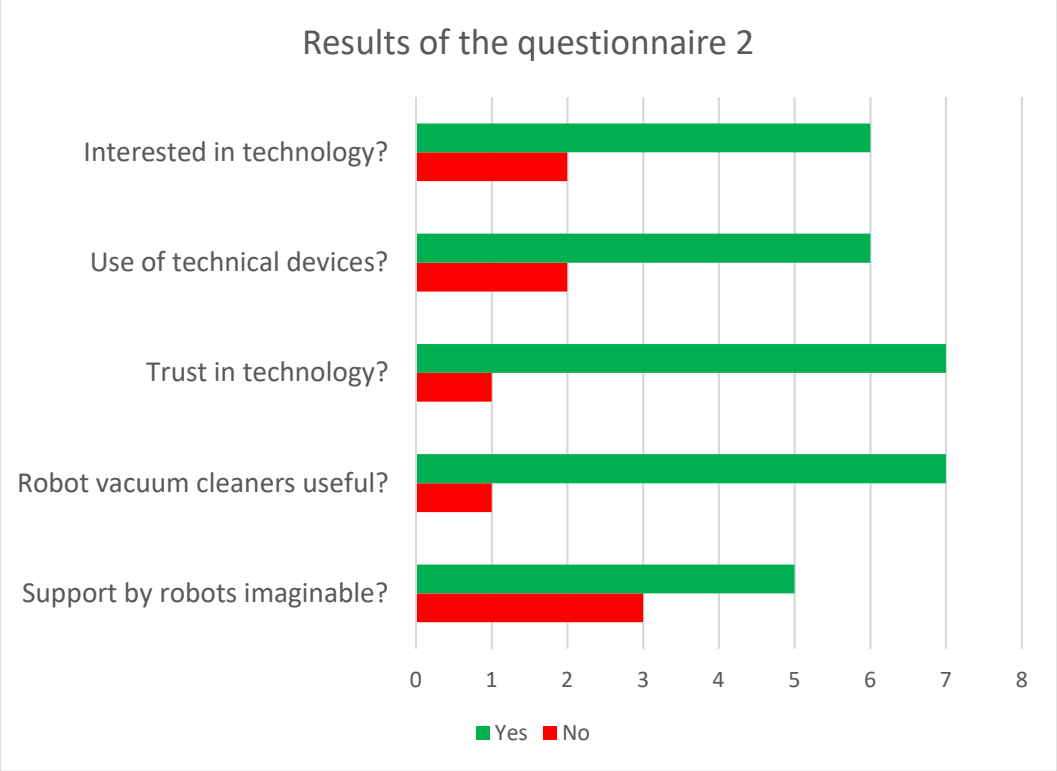

Figure 24: Results of the questionnaire 2

The results of the questionnaire are shown in Figure 24. Six out of eight participants said that they were generally interested in technology. Six participants stated that they used technical equipment themselves. According to the questionnaire, seven of eight participants have confidence in technology. Also seven participants found lawnmower robots or vacuuming robots to be useful. Five participants can imagine being supported by a robot in everyday life.

4.5.2 Feedback regarding the second prototype

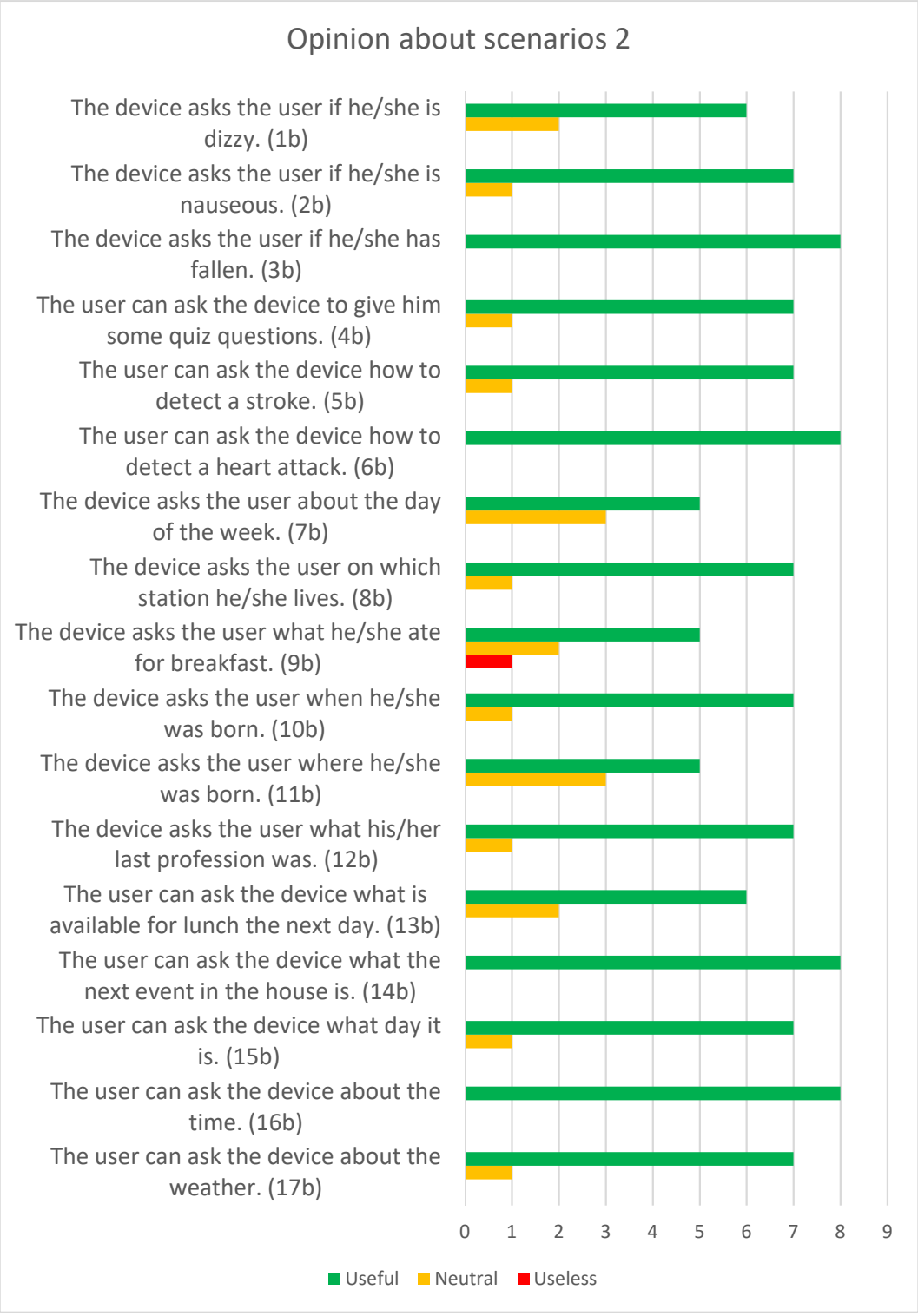

Figure 25: Feedback scenarios second prototype

Figure 25 gives an overview of the participants' feedback regarding every scenario. Green bars represent the number of participants who rated a scenario with "useful",

while orange bars stand for “neutral” and red ones for “useless”. It can be seen that no scenario was rated “useful” less than 5 times. Only one scenario got a single “useless” rating.

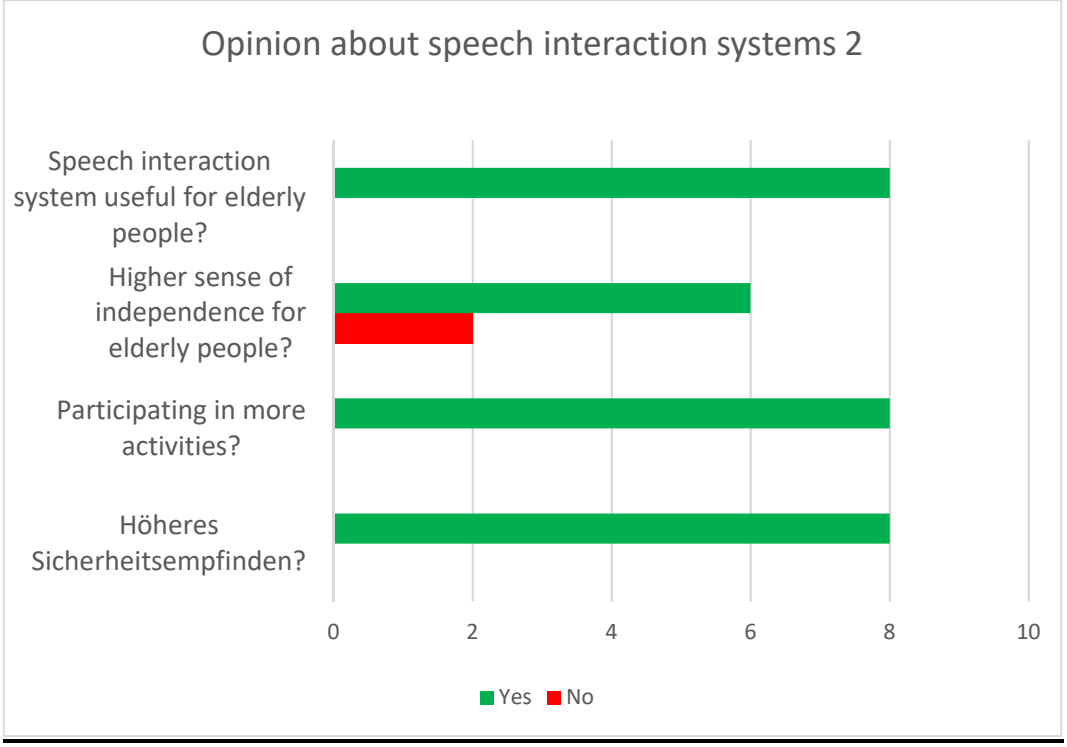

Figure 26: Opinion about speech interaction systems 2

The participants’ opinion about speech interaction systems after the second WOz test is visualized in Figure 26. Eight out of eight participants think that communication with a speech interaction system could be useful for elderly people. When asked "Do you think that elderly people could feel more independent by using a voice computer?" six respondents said yes. Furthermore, all eight participants think that older people could participate in more activities based on suggestions from a speech interaction system. All eight participants felt that older people could have a higher sense of security by using a speech interaction system.

A total of three suggestions for new scenarios could be collected from the participants:

- You could give the device instructions (e.g. "I want to go into the garden"). This instruction could then be given to a caregiver to comply with the request.
- The device could say something fun or nice.
- The device could teach you a foreign language.

4.5.3 Problems and conclusions

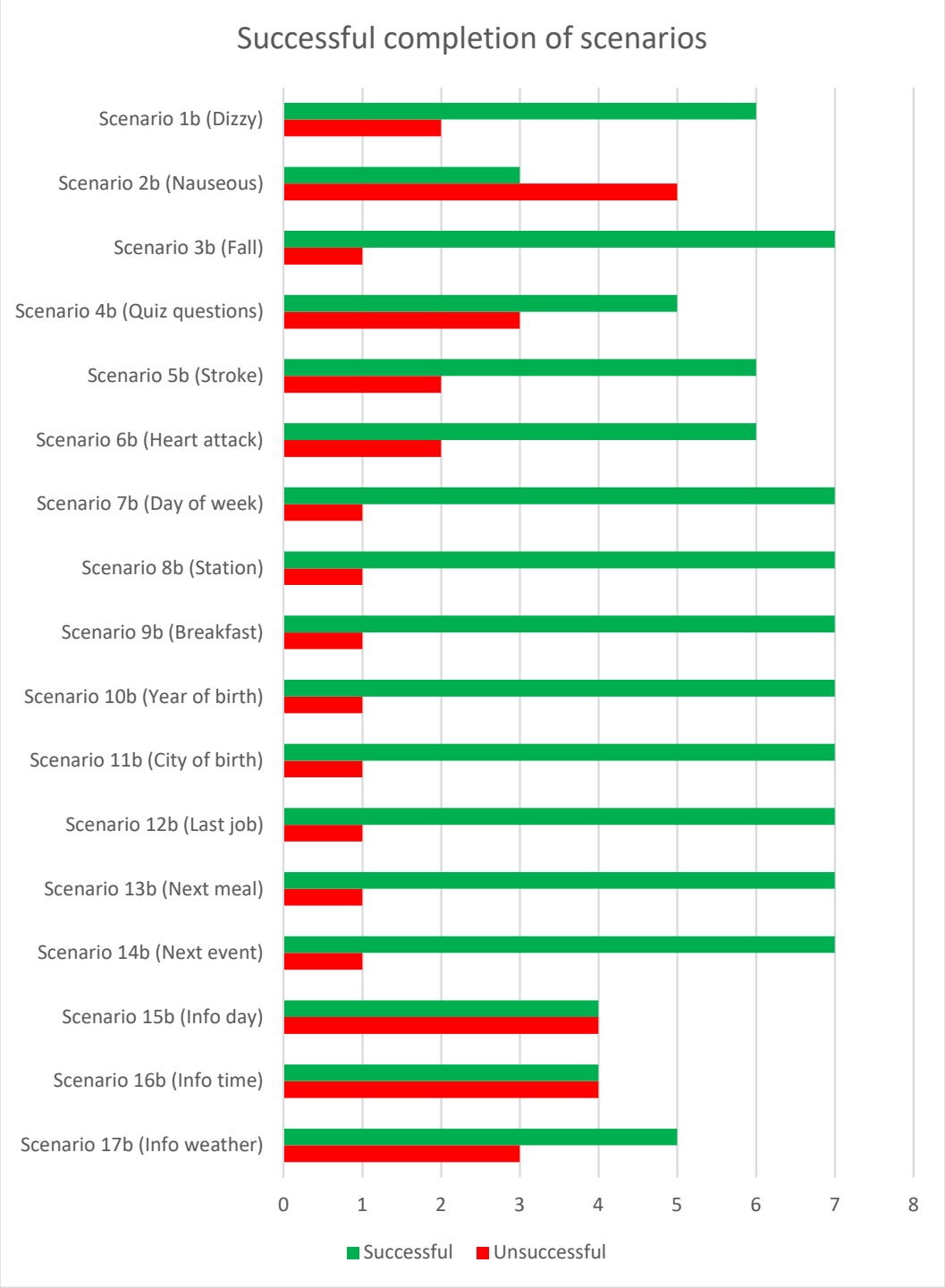

Figure 27: Successful completion of scenarios

Figure 27 visualizes the participants' success of completing specific scenarios. Completing a scenario successfully means a participant can finish a scenario without help. It can be seen that scenario 2b is the one with the least successful completions. Only 3 people managed to complete it themselves successfully. Not a single scenario could be completed by all eight participants. This is due to the fact that one of the participants has a hearing impairment which allows her to

recognize human voice, but the participant was not able to understand the synthesized voice of Alexa, not even at full volume. Therefore some scenarios, which could be completed by the other seven participants without great effort, could not be completed by the hearing impaired participant.

Scenarios which had the highest rate of successful completion were 3b and 7b-14b. These are scenarios in which the device asks the user instead of vice versa.

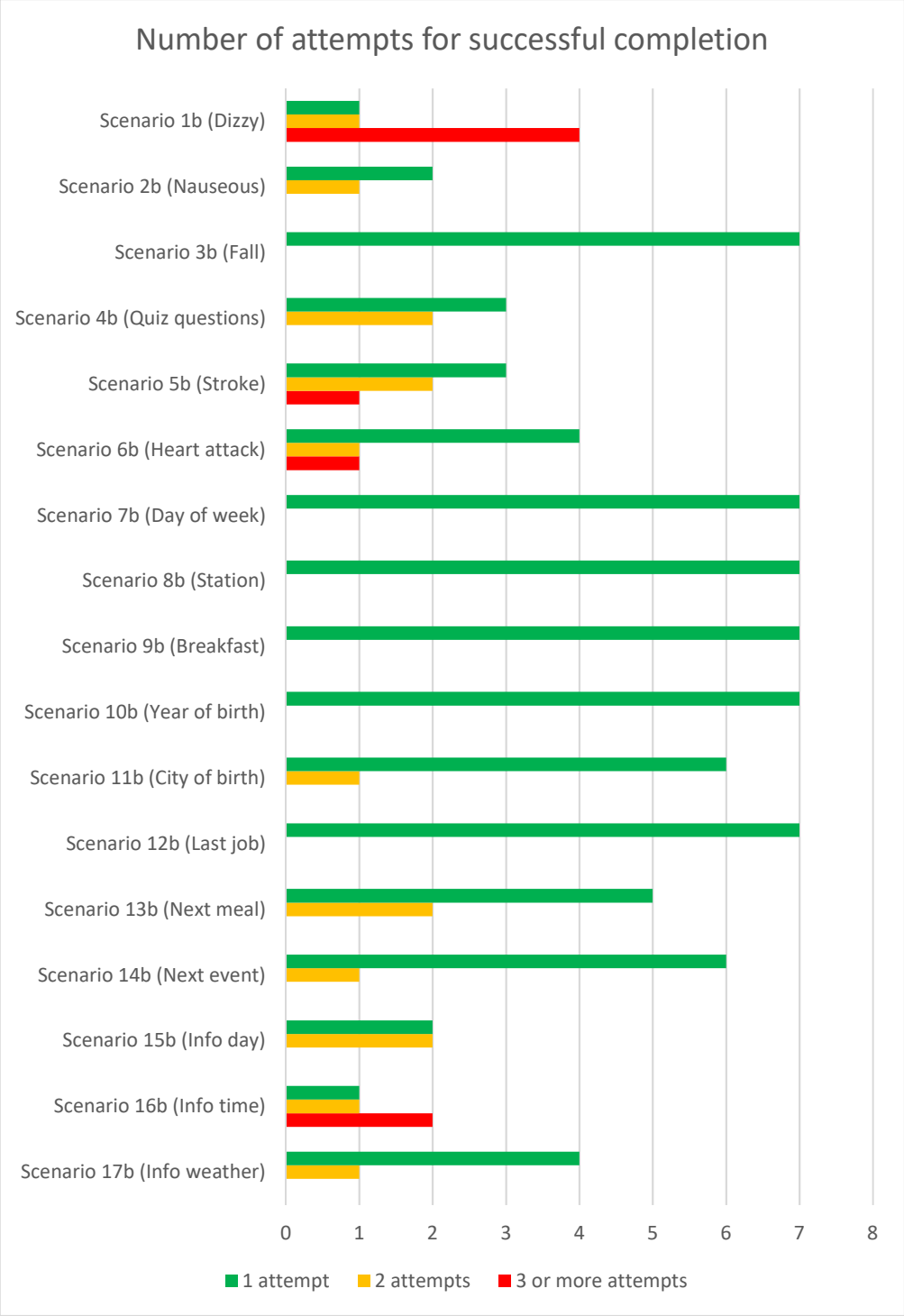

Figure 28: Number of attempts for successful completion

Figure 28 shows the number of attempts the participants needed to successfully complete a scenario. The green bars represent the number of participants who needed one attempt for successful scenario completion. Likewise, orange bars stand for two attempts and red bars for three or more attempts. It can be seen that scenario 1b had the most “3 or more” attempts for successful completion. This is most likely due to the fact this was the first scenario the participants encountered.

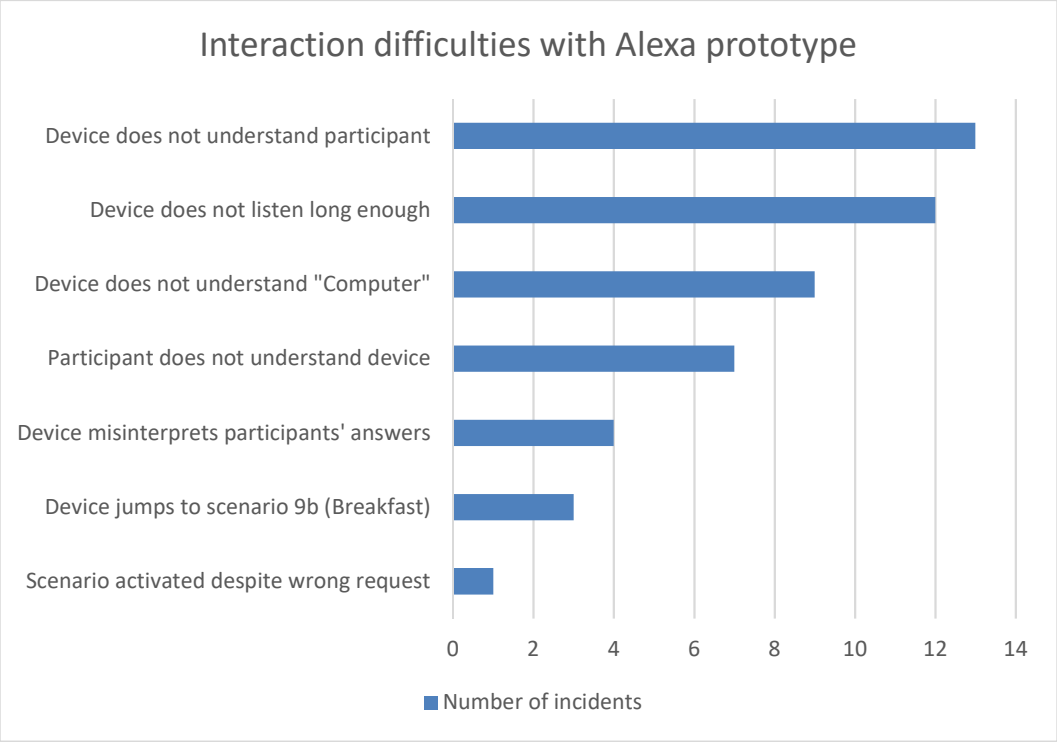

Figure 29: Interaction difficulties with Alexa prototype

Although the second WOz test went well overall, there were some difficulties regarding the interaction with the prototype. Figure 29 provides an overview of all interaction difficulties encountered. The size of the bars represent the number of incidents for each difficulty.

The most frequently encountered difficulty was that the device did not understand a participant. In the whole second WOz test, this happened 13 times. Close behind with 12 incidents was the difficulty “device does not listen long enough”. This means that participants were searching for an answer or question, but before they could say it, the device stopped listening. Another problem was that the Echo Dot did not understand the trigger word “computer” when said by participants. All in all, this happened 9 times. Other problems included participants not understanding the device (7 incidents), the device misinterpreting participants’ answers (4 incidents), the device jumping to scenario 9b while in the middle of another scenario (3 incidents) and the activation of a scenario despite a wrong request (1 incident).

Some reasons for these difficulties can be identified:

- The device did not understand participants, because they either spoke too much or they answered with phrases that could not be interpreted because of too few previously defined synonyms.
- The short active listening time of the Echo Dot is a problem that was already known before the second WOz test. The time in which the device actively listens to user input is set by Amazon to about 7 seconds and cannot be changed.
- The fact that the Echo Dot did not understand “computer” when said by participants is an indication for a poor optimization of Alexa’s speech recognition capability for older people’s voices.
- The participants did mostly not understand the device when it said certain words. The most common word not understood by participants was “übel” in Scenario 2b.
- The misinterpretation of participants’ answers by the device is due to too few synonyms.
- The jumps to scenario 9b can be explained with a suboptimal choice of scenario trigger words. As explained in chapter 4.3.1, the Austrian spelling alphabet was used for indistinguishable scenario trigger words. It seems that the trigger word for scenario 9b (“Ida”) was a poor choice, as the device interpreted similar words to be “Ida” which led to starting scenario 9b.
- The scenario which was activated by a participant despite he / she did not use the correct request is also due to an insufficient number of intent synonyms, as it seems.

As a conclusion the following things could be considered for a further development of the Alexa prototype:

- Scenarios should have more synonyms for their intents. This would lead to a much higher recognition rate.
- The active listening time of Alexa needs to be prolonged for older people, as it seems. Since Amazon has set a fixed time, a makeshift solution is needed.
- It seems that Alexa’s automatic speech recognition technology needs to be optimized for elderly speech.
- Apparently, Alexa’s german accent is not fully understandable for elderly people in Austria speaking Austrian german.

5 Discussion

In this chapter, the most important findings of the study are summarized and interpreted. The following subchapters contain interpretations of the results of the questionnaire about technology attitude, of the feedback to scenarios and of the opinions about speech interaction systems. Furthermore interaction difficulties with the Alexa prototype are interpreted and limitations of the study are presented.

5.1 Interpretation of the questionnaire results

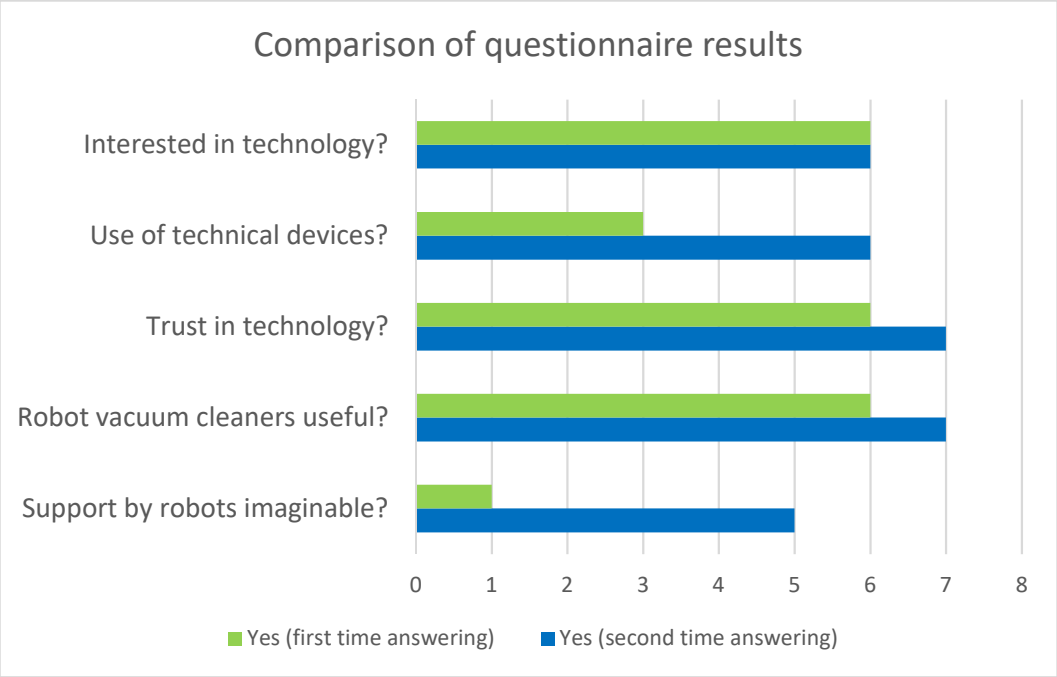

Figure 30: Comparison of questionnaire results

As can be seen in Figure 30, results from the questionnaire about technology attitude differ depending on when the participants received the questions. The same questionnaire was answered by the participants once before the first WOz test and once after the second WOz test. This means that when answering the questionnaire for the first time, the participants had not dealt with the speech interaction prototypes at all. When answering the questions for the second time, they had already experienced interactions with the prototypes. The light green bars represent the number of participants answering with “yes” when answering the questionnaire for the first time, while the dark blue bars represent the number of participants answering with “yes” the second time the questionnaire was handed out.

While the general interest in technology as well as confidence in technology and perceived usefulness of lawn mower robots or vacuum cleaning robots has not changed much, two topics are more or less different in terms of a before-after-comparison.

In the first answer to the questionnaire, only three out of eight participants stated that they used technical equipment themselves. In the second answer, six people answered yes to the question of whether they used technical equipment. This could be explained by the fact that, after dealing with the prototypes, a greater awareness among participants has developed regarding actual technical devices used in everyday life.

The biggest difference was found in the answers to the question of whether support in everyday life by robots was conceivable. In the first answer, only one participant was in favor, the remaining seven clearly rejected. In the second answer to the questionnaire, five participants stated that they could imagine everyday support from robots. This question is provocative in its nature. However, the increased encouragement for robot support after the second response could be an indication that dealing with the prototypes has triggered a thought process in some participants. It is a possibility that some participants have changed their perception of the term “robot” after seeing the Echo Dot speaker and interacting with it.

However, in order to obtain significant results on such issues as presented, it is necessary to conduct a large-scale study with a significantly larger sample.

5.2 Interpretation of feedback to scenarios

When looking at the feedback to the scenarios from the first and second WOz test, there are similarities. However, a side by side comparison should be done with caution, for two reasons: Firstly, the number of scenarios as well as some scenarios itself have changed. Secondly, there may be a bias because the naming for the three-step rating scale has been changed from the first feedback to the second. This was done because of understanding problems during feedback on the part of the participants after the first WOz test, as can be read in chapter 4.2.3.

For this reason, mostly the second scenario feedback is of importance, since after dealing with the Alexa prototype during the second WOz test, the participants had a better understanding of what a finalized version of a speech interaction system could look like.

The feedback regarding the scenarios after the second WOz test was positive overall, as can be seen in Figure 25 in chapter 4.5.2. The most positively rated scenarios were 3b (the device asks the user if a fall occurred), 6b (the device gives the user information about heart attacks), 14b (the device informs the user about

the next event in the house) and 16b (the device informs the user about the time of day). Each of these scenarios was rated as “useful” by all eight participants. The least positively rated scenarios were 7b (the device asks the user about the day of week), 9b (the device asks the user about his/her breakfast) and 11b (the device asks the user about the city of birth). These scenarios got rated with 5 “useful” each.

These findings indicate that scenarios which are about security and where the device provides practical information to the user are received very well. On the other side, scenarios where the device asks questions which require the user to actively think about an answer are not accepted equally well. This should be considered for further development.

An interesting detail: Scenarios 10a, 11a and 12a received the most negative ratings in the first WOz (only four times rated "good"). These scenarios are about remembering. The participants were therefore asked about their year of birth, place of birth and last profession. The scenarios were slightly changed for the Alexa prototype. It was pointed out before each of the three scenarios before the actual question that it is about a remembering training. This was positively reflected in the evaluations after the second WOz test: Scenarios 10b and 12b each received seven "useful" evaluations, scenario 11b five. A possible conclusion is that older users of a speech interaction system do not want to feel that they are being interrogated. If a question is clearly explained what it is about, it will be received more positively. So for more personal questions it is important to prevent misunderstandings.

5.3 Interpretation of opinions about speech interaction systems

Directly after each WOz test, the participants’ opinion about speech interaction systems was conducted. Figure 31 gives an overview of the compared results. The light green bars represent the number of participants answering with “yes” when being asked for feedback after the first WOz test. The dark blue bars represent the number of participants answering with “yes” after the second WOz test.

After dealing with the Alexa prototype, all the areas surveyed except one received a full “yes” feedback from the participants: Eight of eight participants stated that communication with a speech interaction system could be useful for elderly people, in contrast to six people after the first survey. All eight participants also thought that elderly people could participate in more activities based on suggestions from a speech interaction system, while this number was six after the first WOz test. They are also of the opinion that elderly people might feel more secure using a speech interaction system.

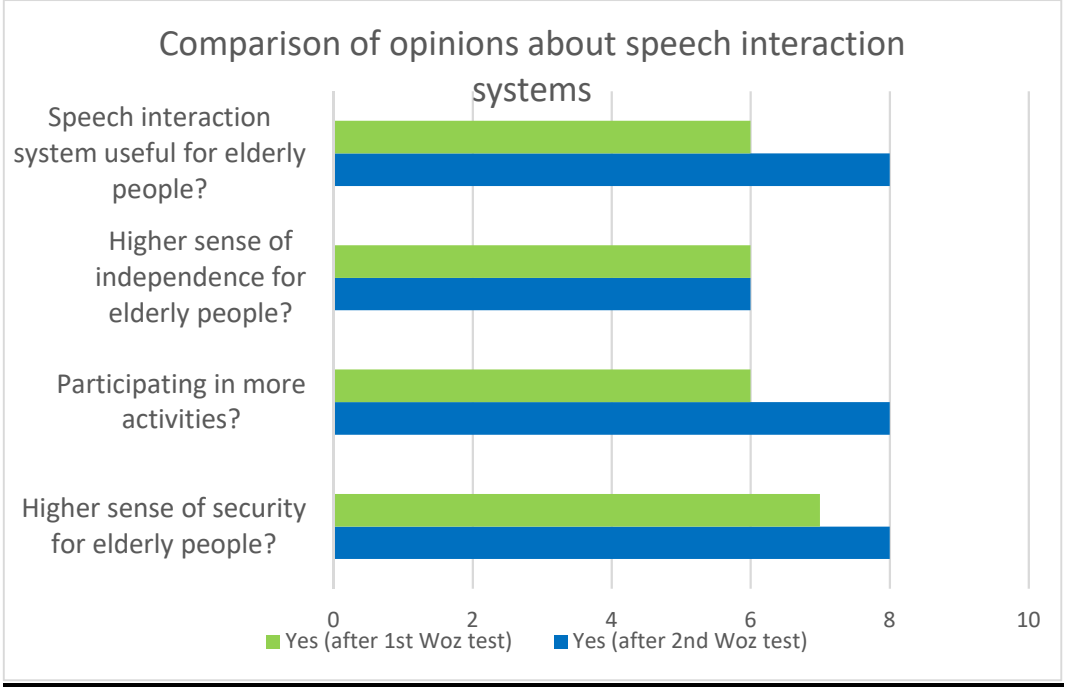

Figure 31: Comparison of opinions about speech interaction systems

The opinion about independence did not change in comparison to the first WOz test: Six of eight participants said they thought elderly people could feel more independent by using a speech interaction system.

All in all, it can be noted that most participants already had a positive opinion about speech interaction systems after dealing with the first prototype, although the first prototype did not yet consist of hard- or software. Dealing with a “touchable” Alexa prototype improved the participants’ positive opinion further. This could be an indication that there is no considerable reluctance to interact with smart assistants in the target group. However, in order to obtain significant results on such issues as presented, it is necessary to conduct a large-scale study with a significantly larger sample.

5.4 Interaction difficulties with Alexa prototype and conclusions

The most frequently encountered difficulty during the second WOz test was that the device did not understand a participant. In some cases, this occurred because participants used too complex phrases while speaking with the device. On the other hand, there where possibly too few previously defined synonyms for the device to understand everything. To resolve this issue, it is proposed for further

development that users need to have a training phase and that a large number of synonyms is used.

Another difficulty was that the device did not listen for user input long enough. It is strongly recommended for future implementations that either a makeshift solution is used to prolong the listening time for Amazon Alexa or that another system is used. Elderly people need more time for answering a speech interaction system, as it seems according to the findings.

A certain trigger word (“computer”) could not be correctly understood by the device in some cases when said by participants. It could be perfectly understood by the study leader, who is a trained speech-language pathologist, though. This indicates that Alexa’s speech recognition capability is poorly optimized for older people’s voices. This finding matches with study results from Wilpon & Jacobsen [19] and Vipperla et al [18], who stated that automatic speech recognition systems are optimized for younger adults’ voices. They also showed that the performance of understanding elderly voice of an automatic speech recognition system improves when it is adapted for understanding specifically that. For future implementations, this fact needs to be kept in mind.

5.5 Limitations

The study presented in this thesis is limited by some factors:

- The study was based on the development of prototypes in a user centered design process. For this purpose, the sample of eight participants was enough. However, for an evaluation with the goal of generating significant statistical results, a much larger sample is needed.
- The study leader has a background in health studies. With a more technical background, a more sophisticated prototype could have been developed. For follow-up projects it is proposed that a team of several persons from different backgrounds work on the project.
- The time of the study was limited given the fact that the study was conducted in the context of a thesis. It is recommended that future implementations allow the project plenty of time. This would allow for more iterations in the user centered design process.

6 Conclusion

For answering the research questions in this thesis, a speech interaction prototype has been developed. Two WOz tests with the same eight participants within the context of a user centered design process enabled data to be collected. Feedback from the participants on the functions of a developed speech interaction prototype was collected. In addition, the opinion of the participants on speech interaction systems was collected, which was positive overall. A separate questionnaire was also used to determine the participants' attitudes to technology.

Based on all data conducted during the study, the research questions can be answered in the following way:

- Elderly people who dealt with a speech interaction prototype consider it possible to gain benefits in their daily lives by communicating with an assistive speech interaction system. This answer is based on the overall positive feedback of the study participants testing the prototype.
- Elderly people living in a nursing home consider functionalities of a speech interaction prototype, which offer practical information and are about security, more useful than others.
- Elderly people who dealt with a speech interaction prototype consider it possible to have an improved sense of independence through usage of an assistive speech interaction system.
- Elderly people who dealt with a speech interaction prototype consider it possible to engage in more activities because of suggestions from an assistive speech interaction system.

In the context of the study presented in this thesis, valuable insights into the communication of elderly people with a speech interaction prototype could already be gathered. The next question that arises is: How do elderly people evaluate a fully functional speech interaction system in terms of usefulness and increased independence and security?

To answer this question, further development of the system and further studies with more participants are needed.

Literature

[1] P. Rashidi and A. Mihailidis, “A survey on ambient-assisted living tools for older adults,” *IEEE J Biomed Health Inform*, vol. 17, no. 3, pp. 579–590, May 2013.

[2] C. Siegel, A. Hochgatterer, and T. E. Dorner, “Contributions of ambient assisted living for health and quality of life in the elderly and care services - a qualitative analysis from the experts’ perspective of care service professionals,” *BMC Geriatr*, vol. 14, Oct. 2014.

[3] “U.S. Smart Speaker Ownership Rises 40% in 2018 to 66.4 Million and Amazon Echo Maintains Market Share Lead Says New Report from Voicebot,” *Voicebot*, 07-Mar-2019. [Online]. Available: https://voicebot.ai/2019/03/07/u-s-smart-speaker-ownership-rises-40-in-2018-to-66-4-million-and-amazon-echo-maintains-market-share-lead-says-new-report-from-voicebot/. [Accessed: 13-May-2019].

[4] C. Siegel and T. E. Dorner, “Information technologies for active and assisted living-Influences to the quality of life of an ageing society,” *Int J Med Inform*, vol. 100, pp. 32–45, 2017.

[5] A. Hämäläinen, A. Teixeira, N. Almeida, H. Meinedo, T. Fegyó, and M. S. Dias, “Multilingual Speech Recognition for the Elderly: The AALFred Personal Life Assistant,” *Procedia Computer Science*, vol. 67, pp. 283–292, 2015.

[6] S. Frennert and B. Östlund, “The domestication of robotic vacuum cleaners among seniors,” *Gerontechnology*, vol. 12, no. 3, pp. 159–168, 2014.

[7] World Health Organization, *Active ageing: a policy framework*. Geneva: Switzerland, 2002.

[8] K. Linz, S. Stula, “Demographic change in Europe—an Overview,” *Observatory for Sociopolitical Developements in Europe 2010*, vol. 4, pp. 2–10, 2010.

[9] European Commission, Directorate-General for Economic and Financial Affairs, and Economic Policy Committee of the European Communities, *The 2018 ageing report: underlying assumptions and projection methodologies.* 2017.

[10] S. Blackman *et al.*, “Ambient Assisted Living Technologies for Aging Well: A Scoping Review,” *Journal of Intelligent Systems*, vol. 25, no. 1, pp. 55–69, 2016.

[11] A. Muñoz, J. Augusto Wrede, A. Villa, and J. Botía, “Design and evaluation of an ambient assisted living system based on an argumentative multi-agent system,” *Personal and Ubiquitous Computing*, vol. 15, pp. 377–387, Apr. 2011.

[12] D. Rhode, “Personal(ized) Health Monitoring, Personalization, and Personality,” *Journal of Technology in Human Services*, vol. 29, pp. 83–100, Apr. 2011.

[13] A. Sixsmith *et al.*, “A User-Driven Approach to Developing Ambient Assisted Living Systems for Older People: The SOPRANO Project,” *Intelligent Technologies for Bridging the Grey Digital Divide*, pp. 30–45, 2011.

[14] M. Memon, S. R. Wagner, C. F. Pedersen, F. H. A. Beevi, and F. O. Hansen, “Ambient Assisted Living Healthcare Frameworks, Platforms, Standards, and Quality Attributes,” *Sensors (Basel)*, vol. 14, no. 3, pp. 4312–4341, Mar. 2014.

[15] A. Spournias *et al.*, “Experimental Evaluation of a Novel Personal Assistant in Greek Language for Ambient Assisted Living Environments employing home robots,” in *2018 South-Eastern European Design Automation, Computer Engineering, Computer Networks and Society Media Conference (SEEDA_CECNSM)*, 2018, pp. 1–9.

[16] K. A. S. V. Rathnayake, W. K. I. L. Wanniarachchi, and W. H. K. P. Nanavakkara, “Human Computer Interaction System for Impaired People by using Kinect Motion Sensor: Voice and Gesture Integrated Smart Home,” in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2018, pp. 531–536.
[17] M. B. Hoy, “Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants,” *Med Ref Serv Q*, vol. 37, no. 1, pp. 81–88, Mar. 2018.
[18] R. Vipperla, S. Renals, and J. Frankel, “Longitudinal study of ASR performance on ageing voices,” 2008, pp. 2550–2553.
[19] J. G. Wilpon and C. N. Jacobsen, “A study of speech recognition for children and the elderly,” in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, 1996, vol. 1, pp. 349–352 vol. 1.
[20] A. Jakl, “Smart Companion for the Best Generation,” *UoAS St. Pölten*, 2016.
[21] S. T. M. Peek, E. J. M. Wouters, J. van Hoof, K. G. Luijkx, H. R. Boeije, and H. J. M. Vrijhoef, “Factors influencing acceptance of technology for aging in place: A systematic review,” *International Journal of Medical Informatics*, vol. 83, no. 4, pp. 235–248, Apr. 2014.
[22] “„Reform der EU-Datenschutzvorschriften 2018“,” *EU-Kommission - European Commission*.
[23] A. De Vito Dabbs *et al.*, “User-Centered Design and Interactive Health Technologies for Patients,” *Comput Inform Nurs*, vol. 27, no. 3, p. 175, 2009.
[24] Z. J. Nagykaldi, M. Jordan, J. Quitoriano, C. A. Ciro, and J. W. Mold, “User-Centered Design and Usability Testing of an Innovative Health-Related Quality of Life Module,” *Appl Clin Inform*, vol. 5, no. 4, pp. 958–970, Dec. 2014.
[25] M. Garschall, K. Neureiter, J. H. Haslwanter, D. Bertel, D. Krainer, and C. Moser, “Investigating user-centered design practices in Austrian AAL projects,” p. 13.
[26] S. Dow, B. MacIntyre, J. Lee, C. Oezbek, J. D. Bolter, and M. Gandy, “Wizard of Oz support throughout an iterative design process,” *IEEE Pervasive Computing*, vol. 4, no. 4, pp. 18–26, Oct. 2005.
[27] A. Steinfeld, O. C. Jenkins, and B. Scassellati, “The Oz of Wizard: Simulating the human for interaction research,” in *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2009, pp. 101–107.
[28] A. Steland, Ed., “Deskriptive und explorative Statistik,” in *Basiswissen Statistik: Kompaktkurs für Anwender aus Wirtschaft, Informatik und Technik*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 1–67.
[29] P. Leitner, J. Neuschmid, S. Ruscher, “TAALXONOMY - Entwicklung einer praktikablen Taxonomie zur effektiven Klassifizierung von AAL-Produkten und Dienstleistungen,” 2015.

List of Figures

Appendix

A. Questionnaire

Sind Sie prinzipiell an Technik interessiert?

- Ja
- Nein

Nutzen Sie selbst technische Geräte?

- Ja
- Nein

Haben Sie Vertrauen in Technik?

- Ja
- Nein

Finden Sie Geräte wie Rasenmäherroboter oder Staubsaugroboter sinnvoll?

- Ja
- Nein

Könnten Sie sich vorstellen, von einem Roboter im Alltag unterstützt zu werden?

- Ja
- Nein

B. Script for “Wizard”

1. **Fühlen Sie sich schwindlig?**

   - Ja.
     - Rufen Sie am besten gleich eine Pflegekraft. Passen Sie besonders beim Gehen mehr auf als sonst.
   - Nein.
     - Das ist schön! Genießen Sie Ihren Tag.

2. **Ist Ihnen übel?**

   - Ja.
     - Melden Sie sich am besten gleich bei einer Pflegekraft. Eventuell brauchen Sie ein Medikament.
   - Nein.
     - Das ist schön! Genießen Sie Ihren Tag.

3. **Ich habe ein Objekt am Boden erkannt. Sind Sie gestürzt?**

   - Ja.
     - Alles klar. Ich löse einen Alarm aus. Es kommt gleich jemand zu Ihnen.
   - Nein.
     - Alles klar. Danke für die Antwort.
   - Keine Reaktion: Frage wird noch zweimal wiederholt. Kommt dann noch immer keine Antwort, wird automatisch ein Alarm ausgelöst.

4. **Nutzer fragt Computer nach Quizfragen.**

Variationen folgender Sätze werden akzeptiert: Ich möchte ein paar Quizfragen. Ich möchte ein paar Rätsel. Ich hätte gerne Rätsel. Stell mir Quizfragen. Ich will Quiz spielen.

- Gerne!

- Glauben Sie, dass Wien mehr Einwohner hat als Berlin?
    - Ja: Das stimmt leider nicht. Berlin hat fast doppelt so viele Einwohner wie Wien.
    - Nein: Richtig! Berlin hat fast doppelt so viele Einwohner wie Wien.
- In welchem Jahr endete der Zweite Weltkrieg?
    - 1945: Das stimmt!
    - Anderes Jahr: Das stimmt leider nicht. 1945 wäre richtig gewesen.

5. Nutzer fragt Computer nach Schlaganfall.

Variationen folgender Sätze werden akzeptiert: Woran erkennt man einen Schlaganfall? Wie kann man einen Schlaganfall erkennen? Wie weiß ich, ob jemand einen Schlaganfall hat? Was passiert bei einem Schlaganfall?

- Bei einem Schlaganfall können verschiedene Symptome auftreten, zum Beispiel Seh- und Sprachstörungen, Kopfschmerzen, Schwindel, Kribbeln in Armen und Beinen, taube Finger oder Lippen, Schluckbeschwerden und Gesichtslähmungen.

Sollten einige dieser Symptome plötzlich bei Ihnen auftreten, sollten Sie sich sofort an eine Ärztin oder Pflegeperson wenden.

6. Nutzer fragt Computer nach Herzinfarkt.

Variationen folgender Sätze werden akzeptiert: Woran erkennt man einen Herzinfarkt? Wie kann man einen Herzinfarkt erkennen? Wie weiß ich, ob jemand einen Herzinfarkt hat? Was passiert bei einem Herzinfarkt?

- Bei einem Herzinfarkt können verschiedene Symptome auftreten, zum Beispiel Engegefühl, heftiger Druck oder Brennen im Brustkorb. Auch eine blasse Gesichtsfarbe und kalter Schweiß auf Stirn und Oberlippe können Warnzeichen sein.

Sollten einige dieser Symptome plötzlich bei Ihnen auftreten, sollten Sie sich sofort an eine Ärztin oder Pflegeperson wenden.

7. Welcher Wochentag ist heute?

- Antwort richtig: Das stimmt! Sie haben ein gutes Gedächtnis.

- Folgende Variationen werden akzeptiert: Montag. Heute ist Montag. Montag, glaube ich. Ich denke, heute ist Montag.

- Antwort falsch: Das stimmt leider nicht. Heute ist…
- Antwort „Ich weiß nicht“ oder Variationen davon: Heute ist…

8. <u>Auf welcher Station wohnen Sie?</u>

- Antwort richtig: Stimmt genau! Sie kennen sich aus.
    - Folgende Variationen werden akzeptiert: Auf der Station St. Katharina. Auf Katharina. Katharina. Ich wohne auf St. Katharina.
- Antwort falsch: Das stimmt leider nicht. Sie wohnen auf der Station…
- Antwort „Ich weiß nicht“ oder Variationen davon: Sie wohnen auf der Station…

9. <u>Was haben Sie heute zum Frühstück gegessen?</u>

- Antwort des Nutzers.
- Darauf Aussage: Das klingt gut, ich hoffe, es hat geschmeckt.
- Antwort „Ich weiß nicht“ oder Variationen davon: Macht nichts. Vielleicht fällt es Ihnen später ein.

10. <u>In welchem Jahr sind Sie geboren?</u>

- Antwort des Nutzers. Darauf Aussage: Nicht schlecht, Sie haben also einiges an Lebenserfahrung.
- Antwort „Ich weiß nicht“ oder Variationen davon: Macht nichts. Vielleicht fällt es Ihnen später ein.

11. <u>In welchem Ort sind Sie geboren?</u>

- Antwort des Nutzers. Darauf Aussage: Alles klar. Sie haben ein gutes Erinnerungsvermögen.

- Antwort „Ich weiß nicht“ oder Variationen davon: Macht nichts. Vielleicht fällt es Ihnen später ein.

12. <u>Welchen Beruf haben Sie zuletzt ausgeübt?</u>

- Antwort des Nutzers. Darauf Aussage: Ich verstehe. Sie haben ein gutes Erinnerungsvermögen.
- Antwort „Ich weiß nicht“ oder Variationen davon: Macht nichts. Vielleicht fällt es Ihnen später ein.

13. <u>Nutzer fragt Computer nach dem Mittagessen.</u>

Variationen folgender Sätze werden akzeptiert: Was gibt es morgen zum Mittagessen? Ich möchte das Mittagessen für morgen wissen. Was gibt es morgen zu Mittag?

- Antwort: Morgen gibt es zum Mittagessen…

14. <u>Nutzer fragt Computer nach nächster Veranstaltung im Haus.</u>

Variationen folgender Sätze werden akzeptiert: Was gibt es demnächst für Veranstaltungen? Was ist die nächste Veranstaltung hier im Haus? Ich möchte die nächste Veranstaltung wissen. Was findet demnächst statt?

- Antwort: Die nächste Veranstaltung ist…

15. <u>Nutzer fragt Computer nach Terminen.</u>

Variationen folgender Sätze werden akzeptiert: Habe ich Termine? Habe ich heute einen Termin?

- Für heute haben Sie keine Termine eingetragen. Wollen Sie einen Termin eintragen?
    - Ja: Wann ist Ihr Termin? Blablabla… Und welche Art des Termins ist es? Blablabla… Danke, der Termin wurde eingetragen!
    - Nein: Alles klar.

C. Structured survey 1

Welche Funktionen finden Sie gut und welche weniger?

Funktionen	Gut	Neutral	Schlecht	Kommentar?
1. Der Computer fragt Sie, ob Sie sich schwindlig fühlen.				
2. Der Computer fragt Sie, ob Ihnen heute übel ist.				
3. Stellen Sie sich vor, Sie sind gestürzt und liegen am Boden. Der Computer fragt Sie, ob Sie gestürzt sind.				
4. Der Computer gibt Ihnen ein paar Quizfragen.				
5. Der Computer sagt Ihnen, wie Sie einen Schlaganfall erkennen können.				
6. Der Computer sagt Ihnen, wie Sie einem Herzinfarkt vorbeugen können.				
7. Der Computer fragt Sie nach dem Wochentag.				
8. Der Computer fragt Sie nach Ihrer Station.				
9. Der Computer fragt Sie nach dem heutigen Frühstück.				
10. Der Computer fragt Sie nach Ihrem Geburtsjahr.				
11. Der Computer fragt Sie nach Ihrem Geburtsort.				
12. Der Computer fragt Sie nach Ihrem letzten Beruf.				
13. Der Computer sagt Ihnen, was es am nächsten Tag zum Mittagessen gibt.				
14. Der Computer sagt Ihnen, was die nächste Veranstaltung hier im Haus ist.				

15. Der Computer sagt Ihnen, ob Sie heute irgendwelche Termine haben.				

Denken Sie, dass die Kommunikation mit einem Sprachcomputer für ältere Personen nützlich sein könnte?

- Ja
- Nein
- Kommentar?

Denken Sie, dass ältere Personen durch die Nutzung eines Sprachcomputers ein höheres Unabhängigkeitsgefühl haben könnten?

- Ja
- Nein
- Kommentar?

Denken Sie, dass ältere Personen aufgrund von Vorschlägen eines Sprachcomputers an mehr Aktivitäten teilnehmen könnten?

- Ja
- Nein
- Kommentar?

Denken Sie, dass ältere Personen durch die Nutzung eines Sprachcomputers ein höheres Sicherheitsempfinden haben könnten?

- Ja
- Nein
- Kommentar?

Haben Sie noch Ideen, was der Computer sagen könnte oder was man ihn fragen könnte?

D. Structured survey 2

Welche Funktionen finden Sie sinnvoll und welche weniger?

Funktionen	Sinnvoll	Neutral	Sinnlos	Kommentar?
1. Der Computer fragt Sie, ob Sie sich schwindlig fühlen.				
2. Der Computer fragt Sie, ob Ihnen heute übel ist.				
3. Stellen Sie sich vor, Sie sind gestürzt und liegen am Boden. Der Computer fragt Sie, ob Sie gestürzt sind.				
4. Der Computer gibt Ihnen ein paar Quizfragen.				
5. Der Computer sagt Ihnen, wie Sie einen Schlaganfall erkennen können.				
6. Der Computer sagt Ihnen, wie Sie einen Herzinfarkt erkennen können.				
7. Der Computer fragt Sie nach dem Wochentag.				
8. Der Computer fragt Sie nach Ihrer Station.				
9. Der Computer fragt Sie nach dem heutigen Frühstück.				
10. Der Computer fragt Sie nach Ihrem Geburtsjahr.				
11. Der Computer fragt Sie nach Ihrem Geburtsort.				
12. Der Computer fragt Sie nach Ihrem letzten Beruf.				
13. Der Computer sagt Ihnen, was es am nächsten Tag zum Mittagessen gibt.				
14. Der Computer sagt Ihnen, was die nächste Veranstaltung hier im Haus ist.				

15. Der Computer sagt Ihnen, welcher Tag heute ist.				
16. Der Computer sagt Ihnen, wie spät es ist.				
17. Der Computer sagt Ihnen, wie das Wetter heute ist.				

Denken Sie, dass die Kommunikation mit einem Sprachcomputer für ältere Personen nützlich sein könnte?

- Ja
- Nein
- Kommentar?

Denken Sie, dass ältere Personen durch die Nutzung eines Sprachcomputers ein höheres Unabhängigkeitsgefühl haben könnten?

- Ja
- Nein
- Kommentar?

Denken Sie, dass ältere Personen aufgrund von Vorschlägen eines Sprachcomputers an mehr Aktivitäten teilnehmen könnten?

- Ja
- Nein
- Kommentar?

Denken Sie, dass ältere Personen durch die Nutzung eines Sprachcomputers ein höheres Sicherheitsempfinden haben könnten?

- Ja
- Nein
- Kommentar?

Haben Sie noch Ideen, was der Computer sagen könnte oder was man ihn fragen könnte?