Master Thesis
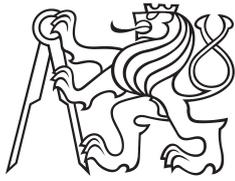
Czech Technical University in Prague

F3

**Faculty of Electrical Engineering
Department of Computer Graphics and Interaction**

Emotional module for virtual voice assistant

Bc. Marek Klement

**Supervisor: Ing. Rudolf Talácko
Field of study: Open Informatics (OI)
Subfield: Human-Computer Interaction (HCI)
May 2022**

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Klement**   Jméno: **Marek**   Osobní číslo: **465928**

Fakulta/ústav: **Fakulta elektrotechnická**

Zadávající katedra/ústav: **Katedra počítačové grafiky a interakce**

Studijní program: **Otevřená informatika**

Specializace: **Interakce člověka s počítačem**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Emoční modul pro virtuálního hlasového asistenta

Název diplomové práce anglicky:

Emotional module for virtual voice assistant

Pokyny pro vypracování:

Analyzujte vědeckou literaturu zabývající se vývojem hlasových asistentů, tvorby adaptivních uživatelských rozhraní a možností analýzy emocí uživatele. Na základě získaných znalostí navrhněte a implementujte rozšíření hlasového asistenta o modul identifikace emocí. Tento modul bude schopen z hlasu uživatele rozpoznat vybrané druhy emocí, na základě kterých umožní adaptaci dialogu hlasového asistenta během komunikace s uživatelem.
Pomocí principů 'User-Centered Design' vytvořte funkční prototyp adaptivního hlasového asistenta a proveďte alespoň dvě iterace experimentů s cílovou skupinou uživatelů, pomocí nichž ověříte správnost vašeho návrhu a zhodnotíte vliv emočního modulu na uživatelskou přívětivost.

Seznam doporučené literatury:

[1] Bertero, D.; Siddique, F. B.; Wu, C.-S.; aj.: Real-Time Speech Emotion and Sentiment Recognition for Interactive Dialogue Systems. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas: Association for Computational Linguistics, November 2016, s. 1042–1047, doi:10.18653/v1/D16-1110. Available from: https://aclanthology.org/D16-1110
[2] J. A. Galindo, S. Dupuy-Chessa, N. Mandran, and E. Céret, "Using user emotions to trigger ui adaptation," in 2018 12th International Conference on Research Challenges in Information Science (RCIS). IEEE, 2018, pp. 1–11.
[3] TALÁCKO, Rudolf: Conversational assistant for smart home. September 2021. Faculty of Electrical Engineering, Czech Technical University in Prague. Available from: https://dspace.cvut.cz/handle/10467/89932.

Jméno a pracoviště vedoucí(ho) diplomové práce:

Ing. Rudolf Talácko   katedra počítačové grafiky a interakce   FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **04.02.2022**   Termín odevzdání diplomové práce: **20.05.2022**

Platnost zadání diplomové práce: **30.09.2023**

_____
Ing. Rudolf Talácko
podpis vedoucí(ho) práce

_____
podpis vedoucí(ho) ústavu/katedry

_____
prof. Mgr. Petr Páta, Ph.D.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací.
Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

.

Datum převzetí zadání   Podpis studenta

Acknowledgements

First of all, I would like to thank my supervisor, Ing. Rudolf Talácko, for his valuable insights and guidance throughout this thesis.

I also would like to thank doc. Ing. Zdeněk Míkovec, Ph.D. for his help with the original idea of this thesis and consultations.

I am really grateful and I would like to thank my wife and parents for supporting me in achieving my study goals.

Finally, I would like to thank my assistants and participants for their help with testing.

Declaration

I hereby declare that I have submitted the work elaborated separately and that I stated all information sources used in accordance with the Methodological Guideline on Ethical Principles for College Final Work Preparation.

In Prague, 20. May 2022

Abstract

This work deals with the design of an emotional module for voice assistants to solve problems in the household. This emotional module ensures that the conversation is adapted to the user's emotions. Part of the research is the analysis of available tools for detecting emotions and then deals with the possibilities of voice assistants. The first part implements a prototype of a virtual voice assistant for maintenance and repair of washing machines. The main part of the work is devoted to the design of an emotional module for a voice assistant based on User-Centered Design. The module's task is to analyze the user's emotions and adjust the conversation accordingly. This module is then tested in an experiment together with a prototype voice assistant for repair and maintenance of the washing machine on a group of senior participants.

Keywords: virtual voice assistant, smart home, emotions

Supervisor: Ing. Rudolf Talácko

Abstrakt

Tato práce se zabývá návrhem emočního modulu pro hlasové asistenty na řešení problémů v domácnosti. Tento emoční modul zajišťuje adaptaci konverzace na prožívané emoce uživatele. Součástí výzkumu je analýza dostupných prostředků pro detekci emocí a následně se věnuje možnostem hlasových asistentů. V první části je implementován prototyp virtuálního hlasového asistenta pro údržbu a opravu praček. Hlavní část práce se věnuje návrhu emočního modulu pro hlasového asistenta na základě User-Centered Designu. Modul má za úkol analyzovat emoce uživatele a na základě toho přizpůsobit konverzaci. Tento modul je následně v rámci experimentu otestován společně s prototypem hlasového asistenta pro opravu a údržbu pračky na skupině seniorních participantů.

Klíčová slova: virtuální hlasový asistent, chytrá domácnost, emoce

Překlad názvu: Emoční modul pro virtuálního hlasového asistenta

Contents

Figures

Tables

Chapter 1

Introduction

1.1 Motivation

We live in a decade where modern technologies are on the rise. There are many devices that have the nickname "smart": phones, televisions, cars, and many more. Smart houses can push living standards to an incredible level. Today, all of these technologies can also use voice instructions. In the world of technology giants, there are already several forms of so-called voice assistants: Google Assistant, Siri, Alexa, or Cortana [19].

All of these assistants can do great things. Just give the order, and they will turn off the light in the living room [17], play soothing music before sleep [3], turn on the automatic vacuum cleaner [18][14], or even turn off the washing machine miles away [16]. Although the range of existing assistants is large, their contextual memory and perception of emotions are minimal. However, in some situations, it is advantage to perceive emotions and adjust the approach accordingly.

What if each of these assistants had something like an emotional module, which would remember the context and be able to respond promptly to every emotion of the user? Perhaps by sensing emotions at any point in the conversation, the assistant can become more emphatic and can adapt to ones mode, which can be beneficial for user experience. Such assistant behavior could better keep the user in a good flow state. The user would not have to worry so much about maintaining attention.

Everyone wants to be more independent in their household. For example, it is very important for older people to be independent in their household. However, it is also necessary to pay attention to their safety. A voice assistant with emotion detection could help by assisting these people with various tasks and keeping them safe at the same time.

■ 1.2   Goals

The first goal of this research project is to explore the possibilities of capturing and responding to the emotions of existing voice assistants. An important part of this goal is to explore methods that can capture emotions from the user's voice or, for example, from their facial expressions.

There is currently a prototype voice assistant for washing machines. One of the important goals of this project is to transform this prototype into a virtual form and analyze its usability in real life by testing it with users.

The main goal of this work, based on previous research and implementation of a virtual assistant, is to design an emotional module for the prototype voice assistant. This module is designed to respond to the individual emotions of the user. These emotions will be anger, happiness, sadness, fear, and surprise. New paths that will respond to these emotions will be implemented into the prepared scenarios. The prototype will be improved so that the moderator is able to capture and record individual emotions.

The last and most important goal is to test the designed prototype with an emotional module on a group of users. In this section, it is considerable to prepare the test, analyze the test results and compare the benefits of the emotional module in the voice assistant with the results of testing without an emotional module.

Chapter 2

Research

This chapter focuses primarily on describing the problems associated with a user's emotions when using the voice assistant. First of all, we define the different types of problems that can occur during communication and then analyze the existing possibilities of detecting emotions.

2.1 Voice assistant

Intelligent personal assistant, virtual voice assistant, or smart assistant - these are some of the names for this technology. It is a type of product marketed mostly by big corporations such as Apple, Amazon, and Google. This technology is based on natural language speech recognition and allows users to do some useful things such as switching the lights on or off or even turn the TV on by voice commands. [17] Basic functions of the voice assistant:

- Remotely control devices by voice - give commands

- Information retrieval - hear the news, joke, or tales

Each user can also set different voices for the assistants and can use them on different devices - Phone, TV, car, washing machine, speakers, and many more. Voice assistants are the most common feature on the smartphone.
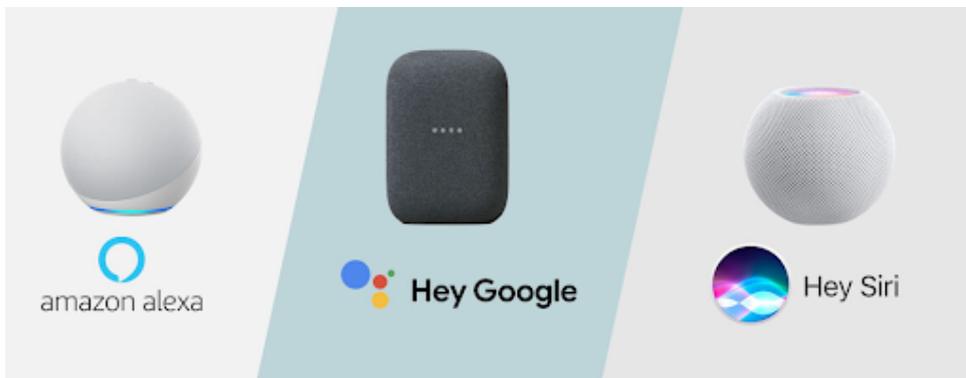
Figure 2.1: Different types of assistants and their starting phrase [34]

2.1.1   Usage

Every day, new possibilities for usage of the voice assistant are created. It can currently be used for a very wide range of functions. Here is a short list of the most well-known usages of voice assistants:

- Ask a question

- Listen to streaming music service

- Check the weather

- Set an alarm

- Control the TV remotely

- Controlling the lights in the house [17]

- Call someone

- Control an automatic vacuum cleaner - for example, a Roomba [14]

- Using cat feeders and fountains - Tesla pet care [33]

- Control other smart devices - washing machine, fridge, sound system, phone, cars, and many more

2.1.2   Google assistant

Google Assistant is an AI (Artificial Intelligence [28]) based virtual assistant developed by Google. Google Assistant is available on each Android device such as phones, TV and other smart home devices. Users are able to communicate with an assistant through voice or by typing text. Before Google assistant, Google also developed Google Now. However, Google Assistant, compared to Google Now, is able to engage in two-way conversations. [7] [1]

2.1.3   Alexa

Amazon Alexa virtual assistant is based on a previous Polish speech synthesizer named Ivona. Ivona was bought by Amazon in 2013 and is the forerunner of the world-famous Alexa. Nowadays, Alexa is able to control several smart devices, thanks to which it has become a home automation system. Users are able to extend Alexa's skills, allowing them and developers to create and use so-called apps such as weather programs and audio features. It uses automatic speech recognition, natural language processing, and other forms of AI to perform these tasks. [20]

---

[1]**What is Artificial Intelligence (AI)?** AI is the scientific intent of creating intelligent technologies (programs, systems, and software) to solve tasks that require human thinking. AI processes big data (pictures, text, videos, etc.) in order to learn about the problem.[28]

■ 2.1.4 Siri

Siri is a virtual assistant that was developed by Apple and is part of their family of devices (iPhone, Macbook, iPad, and others). It uses voice commands, gesture-based control, focus-tracking, and a natural-language user interface to answer questions and perform actions on other devices in its network. Siri is able to adapt itself to the user's individual needs - search preferences and individualized results. [4]

■ 2.1.5 Cortana

Also, another of the technology giant, Microsoft, has its own virtual assistant. It is called Cortana and uses the Bing search engine to perform the assigned tasks. Cortana's functions are similar to those of other assistants - setting reminders, answering questions, and more similar tasks. [25]

■ 2.1.6 Alquist AI and Flowstorm

Alquist is a conversation AI [28] developed by a group of students from CTU (Czech Technical University in Prague). The main goal of the developers is to create an AI, which will be able to have a fluent conversation with a person about all kinds of popular topics, such as movies, sports, music, etc. The name of the chatbot was inspired by a popular Czech drama by Karel Čapek - R.U.R. The word "Robot" was first used in this drama. [9]

Flowstorm in voice assistant development tool. The developer can use the tool to create complex and individual conversations - scenarios. The toolkit offers a simple tree structure that includes, for example, user input, speeches, or variable-based decision-making. The tool is based on Kotlin and can be easily controlled by a drag-and-drop approach. Flowstorm has great documentation and can be tried for free. Flowstorm is developed by the same group of students as Alquist. [29]

■ 2.2 Emotions and sentiment

Emotions are part of a person's character that consists of their feelings, as opposed to their thoughts. [12] It influences how a person may act in any situation. There are some types of emotions - happiness, sadness, disgust, fear, surprise, anger along with a few others.

Sentiment is a thought, opinion, or idea based on a feeling about a situation, or a way of thinking about something. [32] For the sentiment, we distinguish mainly basic types - negative, neutral, and positive.

Emotions and sentiment are really closely related and in the scope of this

project we are going to consider them as each emotion is part of the positive, negative or neutral sentiment. For example, happiness will be positive sentiment and anger a negative sentiment.

2.3 Recognition of Emotions and sentiment

2.3.1 Voice

Most available tools can only recognize sentiment from a voice or text. This is recognized mainly by the words used and the tone of the sentence. The whole sentence thus gains positive, negative, or neutral sentiment. If we wanted to recognize individual emotions (fear, anger, happiness, etc.) from the voice, then we would need a longer part of the speech, otherwise, the recognition of emotions could be distorted and inaccurate. In the same way, each person has a different voice, so it would be necessary to adapt to the user. Most of the tools are able to recognize these emotions - happiness, sadness, disgust, fear, surprise, and anger.

There are some tools for recognizing emotions in voice using the Convolutional Neural Network (CNN). CNN takes a large number of known noises of emotions in speech and learns from them. Then it is prepared to find emotions in each conversation it gets. [5] [2]

CNN is one type of NN that is based on the mathematical linear operation. This operation is called convolution, hence the name Convolutional Neural Network. [2]

2.3.2 Face

Recognition from the face is similar. Again, most of the tools use recognition by Convolutional Neural Network (CNN) or its modifications. CNN gets a great number of pictures with different types of emotions presented in them. Then the learning process begins, the CNN finds similar spots in the face and adds these spots to particular emotions. After the learning process is done, the CNN is able to recognize some emotions. Usually, these emotions are perceived from the face by most of the tools - anger, sadness, fear, happiness, surprise, and confusion. [15]

---

[2]**What is Convolutional Neural Network (CNN)?** Neural networks (NN) are one approach to solve the problems. NN analyzes the available data, which learns the behavior of the participant. This behavior is further able to apply and recognize other data. The model for the LV human nervous system and human thinking. [6]
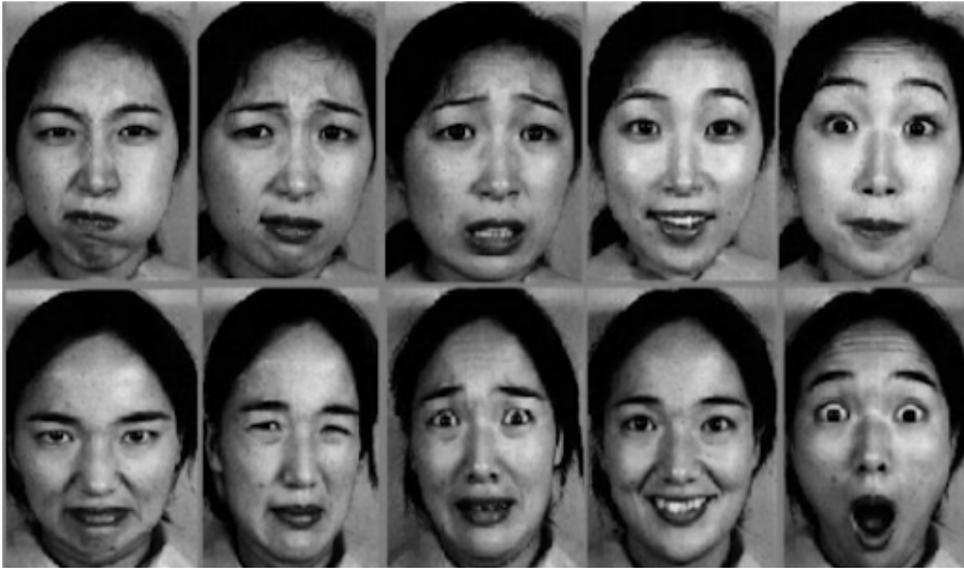
Figure 2.2: Facial expression during study [15]

■ 2.3.3 Combination

As we can see, both approaches recognize similar emotions and none of them is accurate. Thanks to the combination of these approaches, we can better recognize individual emotions, or we can verify whether this emotion is available in the voice and face at the same time - for more accurate adaptation/reaction. It is also possible to find out the given emotion from the face and confirm its occurrence by the observed sentiment in the user's voice.

■ 2.3.4 Mood detection in different assistants

Each of the virtual assistants mentioned previously is able to use AI to find out the user's mood and react to it differently. There are only a bad and good mood. This approach brings in a wide variety of possibilities. [24] Imagine that you are in a bad mood and when you come home, the assistant dims the lights so you don't get irritated by the brightness of the room, on the other hand, when you are in a good mood, your assistant could read you a joke or start some happy songs.

■ 2.4   Related work

■ 2.4.1 Voice assistant for washing machine

At the Department of Computer Graphics and Interaction of the Czech Technical University in Prague, Ing. Rudolf Talacko was dedicated to create a voice control for a smart home in his thesis [30]. The work first focuses on creating a voice assistant for laundry. This assistant helps the visually

impaired people and seniors go through the recommended laundry process. The benefits of such an assistant could not be confirmed. However, the voice control of the washing machine got good results.

Another part of this work is devoted to the voice assistant for maintenance and repair of washing machines. The assistant was created as a prototype, which helped the participants of the experiment to solve three different scenarios with maintenance or repair of the washing machine. They selected a total of 12 participants for the experiment. 8 participants were seniors (over 65 years of age) and 4 participants were visually impaired. The experiment confirmed the usefulness of such an assistant. In addition, when implementing the prototype of this assistant, a universal tool was created to test new possible scenarios.

2.4.2 Emoji study

Figure 2.3: Study used application with emoji [21]

In this work, a team of researchers from the University of Munich researched human emotions when speaking with a voice assistant [21]. During the experiment, they used an emoticon on the phone, which the participant had the task of either making, laughing, or annoying. Participants were to speak on emoji for 2 seconds. This speech was recorded and the whole thing was repeated 5 times with different emoji for each emotion. After each answer, the emoji became happier.

The study included a total of 52 participants and most of the participants' responses were in neutral emotion. The experiment found that emotions in the voice are language and culture-dependent. Gender differences, although found, could not be explained.

2.4.3 Voice emotions detection

In another study, a researcher from the Nirma Institute of Technology Ahmedabad focused on detecting emotions based on several fundamental attributes of a voice such as pitch, tint, volume, and vocal tone [10]. Three test cases were examined, each focused on a different emotion. One for anger, the other for panic, and the third for normal emotions (neutral).

The cornerstone for the other two emotions was neutral emotion. All attributes set in the test were set to mean value. Anger was characterized by increased loudness and pitch. Panicked emotions were explored when participants were in a panic or overwhelmed state. It was characterized by a different intonation.

Detecting emotions from a user's voice has been found to be potentially beneficial for improving conversation.

2.4.4 Text emotion detection for virtual agents

In this work, Xianda Zhou and William Yang Wang focused on recognizing emotions from written text [35]. The study chose a large corpus of data (more than half a million) from Twitter conversations that contained emoji as inside the text and assumed that emoji represented the feelings the text expressed. Based on these data, a deep generative model was trained.

Quantitative and qualitative tests were subsequently performed on this model. One of the tests was also the Turing test, where participants had to choose whether the answer was written in human. In this test, 18% of participants chose the answer that was generated by the system and 27% said they could not distinguish the two answers. It was found that the model is able to generate quality responses to emotions.

Chapter 3

Virtual voice assistant

With regard to the identified opportunities, in this work, we decided to create a virtual voice assistant for a smart home, which is based on previous research in 2.4.1. This work dealt with the creation of a prototype voice assistant to help with the repair and maintenance of washing machines.

3.1 Virtual prototype vs. basic prototype

For future parts of this work, it is good to define in advance the difference between the virtual form of the prototype and the basic prototype. The virtual form of the voice assistant is automated. The virtual prototype is not controlled by anyone and responds to the user independently. In the context of this work, we will call such a prototype **virtual**.

On the contrary, the basic prototype works on the principle of the Wizard of Oz. Throughout the experiment, a moderator must control the conversation and start individual speeches. We will call this type a **prototype**.

3.2 Implementation

In this part of the work, we deal with the design and implementation of a virtual voice assistant. In order to be able to compare the results of the subsequent experiment, it will be good to choose the created prototype in 2.4.1 as a template. We want to create a fully virtual form of each scenario, which will not require human assistance. Each of the scenarios solves a slightly different problem with the maintenance or repair of the washing machine. We had the opportunity to use the Flowstorm tool [29] since we have contact with the developers and it is possible to ask them for help.

Flowstorm is an online tool that provides an environment for developing voice assistant scenarios. This tool is largely based on assembling the individual graphical components of the interview into a diagram. However, the tool is based on Kotlin (Programming language developed by Jetbrains [31]) and also allows developers to access and modify the code.

Although Flowstorm currently has Czech language support, it does not work quite well. For example, recognizing this language has occasional outages. Therefore, the scenarios will be translated into English.

■ 3.2.1 Scenario

In the context of this project, we understand that the term voice assistant scenario is just one conversation between the assistant and the user. It is a directed cyclic graph of recurring assistant speeches and user responses (intents), which is finite in most of the cases.

Under the terms of the scenario, we can imagine, for example, a conversation about a user's day or the steps of following a recipe with the help of a voice assistant.

■ 3.2.2 Path

The word path here means a certain sequence of several speeches/points in a scenario. This is one subgraph of the scenario. The path does not contain cycles, it always has a start and end node. The path is final.

■ 3.2.3 Flowstorm

In the following paragraphs, different types of possible nodes of conversation are described.

■ Enter

Each scenario has to begin somewhere. No matter if this is the main dialogue or sub dialog, enter represents the beginning of it. Enter is a node in the scenario which precedes one other node (Speech or User intent) and does not have any input.
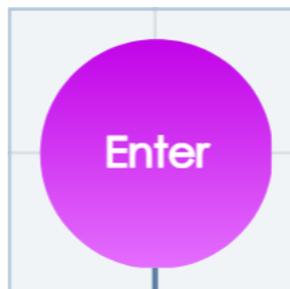
Figure 3.1: Enter node

Speech

The assistant needs to be able to say something. For this purpose, speech is the node which represents the saying of the assistant. In this node lies the setting of the text to be read, which can actually be more than one variant. For example, if we want to say "yes", then we can use two variants of the answer: "yes" or "ok".

There must be a path to the speech node coming from the enter node. The Speech node is then followed by one other node (see following nodes).
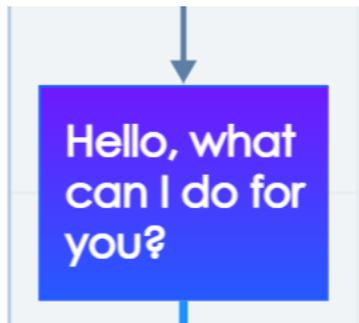
Figure 3.2: Speech node

User input

Arguably, the most important node is the user input node. It is a node that is triggered by user input and is followed by intents based on the input. It can be followed by more than one intent node. Before each user input node is usually some speech - ask.

Multiple variants can be set in the intent again. As in the previous example, the user can answer differently to approve the request - "yes" or "ok". The developer's task is to cover as many variants as possible that can occur for each path and thus avoid unknown answers. However, this is a very challenging task, Flowstorm is equipped with several automatic messages that indicate that the assistant does not understand the request.
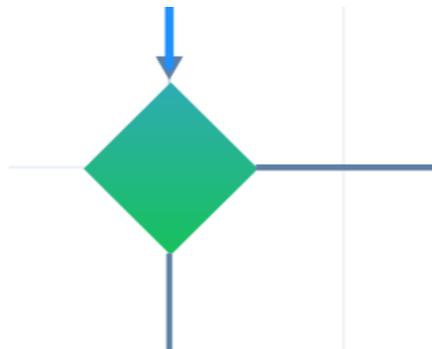
Figure 3.3: User input node

■ **Intent**

The Intent node was mentioned in the previous paragraph. Intent node offers developers the ability to anticipate user input. Flowstorm offers several ready-made Intents: "yes", "no", and more. Several expected phrases must be defined in the intent, and the AI-based Flowstorm will include the remaining options in an intent. Before an intent node, there must be a user input node, and intent is usually followed by speech.
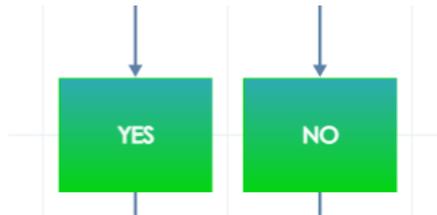
Figure 3.4: Intent nodes - positive and negative

■ **Global intent**

The global intent node is similar to normal intent nodes, except it can be triggered at any point of the conversation. It has no previous node and usually is followed by speech. For example, it can be used to trigger a joke while repairing something in your smart washing machine.

Figure 3.5: Global intent node

■ **Function**

The Function node is one of the approaches to the code itself. It allows developers to create conditions, write variables, or perform various procedures. The Function node is able to continue on one of the possible paths, or the following nodes, based on the evaluation of conditions.

Example. We will approach the work of The Function node with a simple example. We have a scenario that reveals which parts of the washing machine are broken. In this example, there are a total of four parts that can be broken. In the conversation, we go through the individual options and set a variable that records whether a fault has been found or not. This variable is set by The Function node. At the end of the conversation, there is another Function node that points to different ends depending on which fault was found (or which variables come out as true or false).

14
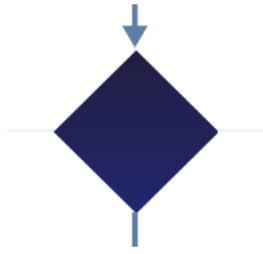
Figure 3.6: Function node - decision point

End/Exit/Go back

Each conversation has to end somewhere. The End, Exit, and Go Back nodes are used to end a given part of the conversation. The End node is the terminal station; there is no other node after ending conversation. The Exit node does not necessarily mean terminal node. Rather, it leads to the exit of a subdialog. Go back is used to end global intent and come back to the spot where the global intent was triggered so the conversation can follow.
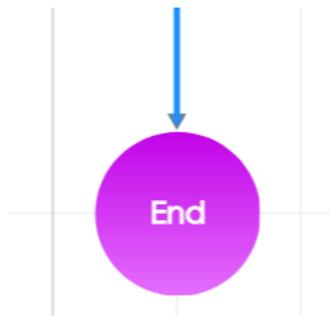
Figure 3.7: End node

3.2.4   Washing machine repair

To use the voice assistant, several scenarios involving repair and maintenance of the washing machine are introduced in this work. The choice of washing machine repair is not random, it was used in 2.4.1.

Compare the results

The main reason for the washing machine repair and maintenance scenarios is to compare the results of the previous study with our results. The individual scenarios are strongly based on previous work so that it is possible to compare the results and find out the results as best as possible.

Good source of emotions

Individual washing machine repair scenarios are a great way to evoke and observe emotions in the user. The user is almost never happy when his

15

washing machine breaks, he may be irritated when he has to perform its routine maintenance or, on the other hand, he may enjoy the repair.

■ Length of scenarios

Most of the voice assistants used are short-term activities. For example, turning on the television, listening to the weather, or turning off the light are just one-off activities in which we would not have room to observe individual emotions. Scenarios with washing machine repairs are not a simple task and are rather a longer guide. There is a sufficient amount of data for observing and reacting to individual emotions.

■ 3.2.5 Scenarios

We currently have 4 scenarios for repairing the washing machine. Each scenario solves a slightly different problem, which is triggered when the washing machine breaks or when a maintenance action is required. The links lead to a site where the scenario can be tried. [1] [2] [3] [4]

- S00 - Introduction

- S01 - Error E20 and cleaning the filter

- S02 - Cleaning the filter of natural hose

- S03 - Flood

■ Introduction of scenarios

These scenarios were part of a study 2.4.1 to design an emotional assistant model for washing machine disorders. Each scenario is designed to be used for the experiment - testing on a circle of users. The following paragraphs briefly describe each scenario:

S00 - Introduction. The first of all scenarios is a short introduction. This is mainly a test of communication with the assistant, getting used to the speaker's voice and adjusting the required volume. The first scenario is intended mainly for users who have no experience with voice assistants. Thanks to this scenario, they can easily try communicating with it.

Expectations

The main reason is to set the appropriate volume and show the user with the form of communication. It's important to make sure everything works for him and that he wants to continue the experiment.

---

[1]Link to S00 - https://bot.flowstorm.ai/61bd247c5955211cf43e34b6
[2]Link to S01 - https://bot.flowstorm.ai/61b91617ebb78f6a632d7a94
[3]Link to S02 - https://bot.flowstorm.ai/61bd200b5955211cf43e2c0d
[4]Link to S03 - https://bot.flowstorm.ai/61be57355955211cf43e9301

S01 - Error E20. This scenario is the first non-simple. Participants of the experiment [30] should imagine that it is Friday afternoon and they want to go on a family vacation on weekends. They decide they want to start the washing machine with a delayed start before the vacation so the laundry is ready to be hung in the morning. But when they came to the washing machine first thing in the morning, Error E20 occurred. Participants have to figure out what needs to be done with the help of a conversation assistant.

Ideal course of the scenario

In this scenario, the user first checks that the drain hose is at the correct height and is not kinked anywhere. He then opens the drain filter door and prepares a drain tank, preferably lower. He then slowly releases the filter and lets the container fill. When the container is full, the filter retightens, pours the water into a larger container, and continues to drain. When all the water is out, he finds that the filter is really dirty and cleans it, including its hole. Then it just guides the filter correctly into the hole, tightens it, and continues washing.

S02 - Cleaning the filter. The participant of the experiment was drawn into a situation where, after one of the washing machine failures, they read that these situations could be prevented by regular cleaning of the filter. In that situation, the participant decided to ask the voice assistant for help with cleaning. Then they could go through the whole process together.

Ideal course of the scenario

In this scenario, the user first prepares a rag and an old toothbrush. Subsequently, it closes the water supply so that the bathroom does not flush out. He first unscrews the supply hose and then cleans it. After cleaning, the hose is screwed back on. The water opens again and everything is done.

S03 - Flood. In this case, the participant in the experiment witnessed a very unpleasant situation, and the bathroom was flooded. Their task, together with the voice assistant, is to find out the cause of the flooding of the bathroom and solve the situation.

Ideal course of the scenario

The main defect that the user is looking for is a disconnected supply hose in the washing machine drum. The user assistant will gradually check the following possible defects:

- Check that no water leaks around the washing machine drum. It is important not to open the washing machine door.

- Check that the drain filter is not enabled. The filter must not be unscrewed.

17

- Check supply hose - must not be broken or damaged.

- Check drain hose.

Since the drain hose can be moved in the direction from the washing machine, there is a fault. The user should then call the service and describe the whole process to a specialist.

▪ Simple dialog with recognition of sentiment

A sentiment recognition scenario was created to represent emotions in the virtual voice assistant. The current possibilities of the Flowstorm development environment cannot recognize individual emotions [29]. However, sentiment is very close to emotion.

This scenario prompted the user for any narration when started. Subsequently, the sentence said by the user is automatically analyzed and the sentiment is recognized. The script thus demonstrates the reaction to the recognized sentiment.

- **Negative sentiment** - In this case, it ends the conversation. Before the end of the conversation, we ask the participant if he wants to call someone for help or to cheer him up.

- **Positive sentiment** - Participant is happy or in a good mood. In this case, the assistant decides to ask the participant if he wants to hear a joke. Then, depending on the response of the participant, either tell a joke or end the conversation.

- **Neutral and no sentiment** - In this case, the assistant asks for more words or sentences so he can better decide on sentiment. And the process is repeated.

18
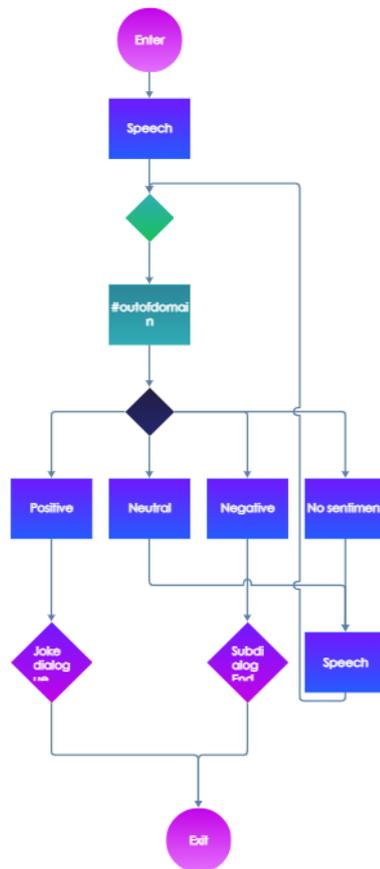
Figure 3.8: Emotion scenario

This dialog is available here: Recognition of sentiment from conversation

Other scenarios

For the purpose of the big scenarios, a couple of simple ones were created. They are used as sub dialogues in main dialog. See the link and try these basic scenarios for yourself. [5] [6] [7] [8]

- End of conversation

- Joke

- Fancy ending

---

[5]Link to End of conversation - https://bot.flowstorm.ai/61bca0a65955211cf43d169f

[6]Link to Joke - https://bot.flowstorm.ai/61bd2065ebb78f6a63a35325

[7]Link to Fancy ending - https://bot.flowstorm.ai/61bd20f15955211cf43e2d40

[8]Link to sentiment dialog - https://bot.flowstorm.ai/61be577c5955211cf43e948a

▪ 3.3 Experiment

In order to verify that the virtual voice assistant works properly, an experiment is necessary to verify that it can be used and under which conditions. One focus of this section is also to compare the results with a previous experiment.

▪ 3.3.1 Research questions

Before we embark on a proposal, it is appropriate to prepare research questions. We want to get an approximate answer to them after the experiment.

- **P1** - Are people able to ask for the help of a virtual assistant in repairing/maintaining a washing machine?

- **P2** - Is there some benefit of the virtual assistant opposed to the previous prototype in 2.4.1?

- **P3** - How do people behave if their assistant has difficulty understanding?

▪ 3.3.2 Participants

In this part of the experiment, we will focus on the selection of the experiment participants and the recruitment process.

▪ Target group

Development of virtual voice assistants can help a wide amount of people. When it comes to the user's age, then we would like to focus mainly on older people. The main purpose of selecting older people for the experiment was to closer compare the results from the previous experiment 2.4.1. This group of people may not have any experience with virtual voice assistants and also requires assistance with repairs the most. On the other hand, it is difficult to find entities in Prague, where the experiment takes place, that are seniors and are able to communicate with the assistant in English. Therefore, at least 45% of seniors will be enough for us. The groups were divided as follows (approximate percentage of the required representation in brackets):

- Under 18 [0%]

- 18 - 60 [55%]

- 60 and more [45%]

As already mentioned, knowledge of the English language is very important. We therefore want to select participants who are able to speak at least at level B2. Another attribute that needed to be monitored was gender. We made the number of women similar to the number of men. However, this attribute is not entirely essential.

■ Number of participants

It is necessary to ensure a sufficient number of participants to ensure that each area is properly tested. In the previous experiment, the choice of the number of participants was based on a relationship defining dependence between the number of user tests and the number and nature of errors found, which Jakob Nielsen described in his article. [27]

To be able to achieve at least 95% chance of capturing most of the problems we need at least 5 participants. In research by Tom Landauer and Jakob Nielsen is shown that the number of usability problems found in a usability test with n users is:

$$N(1 - (1 - L)^n) \tag{3.1}$$

- ■ N is the total number of usability problems

- ■ L is the proportion of usability problems discovered while testing a single user.

- ■ The typical value of L is 31
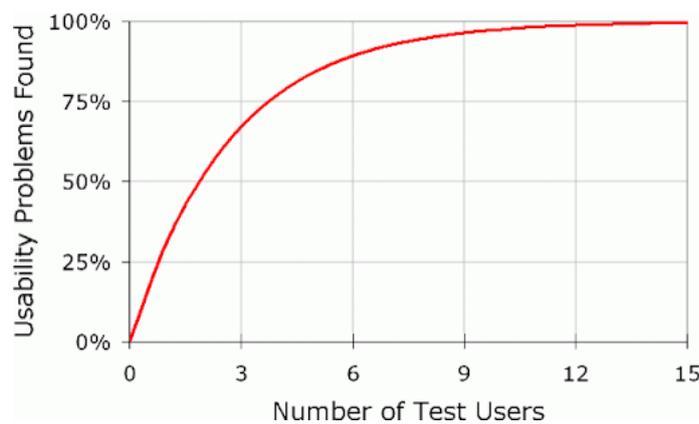
Figure 3.9: Relationship between users and problems where L = 31%. [23]

Therefore we need at least 5 participants.

■ Recruitment

Recruitment of experiment participants was done using experiment-friendly groups on social networks and contacts from our friends. In this way, we were able to get the necessary participants.

- ■ Recruitment was done using a pre-test form in Facebook group Active seniors from Prague and the surrounding areas. This form is not only used for getting needed information about each user, but also as a screening tool. We chose participants as needed by their English level and age (possibly also gender).

- We tried to contact the retirement home and primarily made contact with a retirement home in Kladno (CZE).

- We tried to get workers from the office in Děčín (CZE).

Recruitment was done using the pre-test form.

3.3.3 Environment

Testing will be performed at the Department of Computer Graphics and Interaction CTU in a prepared classroom. The virtual voice assistant created in the Flowstorm will be used for our experiment. The participant must be able to communicate with the assistant on a laptop that simulates an assistant. Therefore, this assistant can be used from different devices (washing machine, phone, speaker). The experiment environment must therefore be equipped with a laptop on which this assistant will be.

Figure 3.10: Environment of testing virtual voice assistant

Technical specification

Although the test participant was only to communicate with the virtual voice assistant, an experiment moderator was in the room. Due to the presence of the moderator, no cameras were needed. There was also another external microphone in the room, for recording all conversations.

■ Room preparation

The virtual assistant testing room was equipped with several tables. The first table, shown in the figure 3.11 on the left, is ready for the test participant. There is a washing machine with surrounding equipment. The supply hose is attached to the wall to resemble the connection to the water supply system. The drain hose drains into the container.

A laptop with a virtual assistant is ready on the table shown in the figure 3.11 on the right. A moderator will be ready there to avoid any problems.
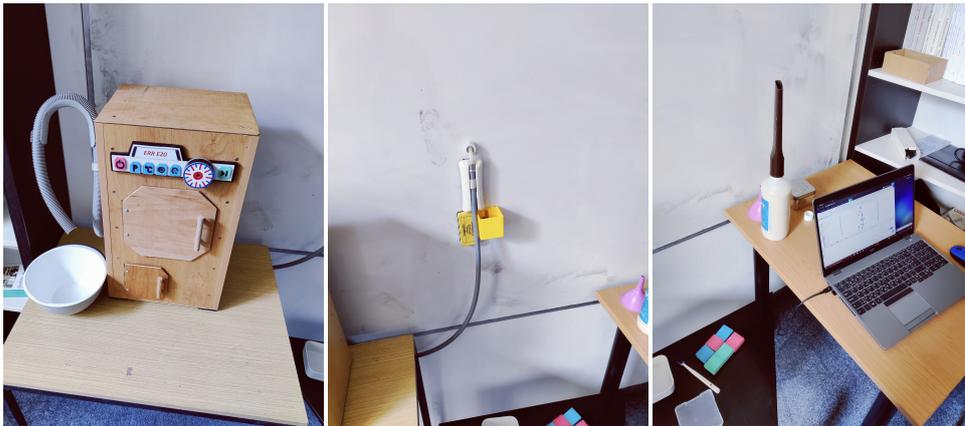
Figure 3.11: Setting of the assistant and props

■ Props

For the authenticity and relevance of the experiment, it was necessary to provide several items that were mostly also used to repair the washing machine in the previous experiment mentioned in 2.4.1. These props are, for example, water containers, drying cloth, and a toothbrush. A virtual assistant was also available to the participant in the room.

■ 3.4 Analyses of pilot testing results

■ 3.4.1 Participants profiles

Each participant went through an entry form (online or on-site) and answered the following questions. Each of the questions is marked differently and is then referenced under this designation in table 3.1.

- **A1** - Would you say that you are a technical type?

- **A2** - How would you rate your manual skill level?

- **A3** - Do you have any experience with repairing the washing machine?

- **A4** - Do you think you are communicative?

- **A5** - Do you have any experience with virtual conversation assistants?

In total, 5 users were selected. The following table lists their profiles. For question A3, the supplementary question was what experience the given participant has with the repair of the washing machine. The answers to this question were for information only and are not listed in the table 3.1. One of the users had little experience with repairing the washing machine, although not much. He only helped his father in reparation. You can find the whole example of the questionnaire in the appendix.

Participant	Gender	Age	A1	A2	A3	A4	A5
SU1	Male	25	a	c	b	c	c
SU2	Female	25	b	c	a	d	c
SU3	Male	23	b	b	b	a	c
SU4	Male	23	a	b	a	b	c
SU5	Female	22	b	d	a	a	c

Table 3.1: Pilot testing profiles

3.4.2 Analyses of testing with each participant

Participant 1 (SU1)

The first of the pilot tests were marked by major outages on the Flowstorm side. The system was probably very overwhelmed and the response to individual user sentences was very long. The participant was initially patient but only ended with two scenarios. For the other two, the system got stuck in the middle and didn't respond.

Improvements. This initial Flowstorm outage was considered unique and it was hoped that the system would be stable in most cases. Several conversation points have been fixed for further testing. The errors in the assistant's sentences were primarily fixed and the passages of the scenario were slightly modified.

Participant 2 (SU2)

The second test was already significantly better than the first. As a result, the problem in the first test seemed rare. The participant went through most of the scenarios. He finished 3 out of 4. In one case, the assistant was in loop and it was not possible to get out of the situation. After a short time, the moderator intervened and ended the scenario.

Improvements. Following this test, more extensive adjustments were made to the scenarios to avoid looping. During the test, several minor errors in the texts appeared, which were subsequently corrected.

■ Participant 3 (SU3)

Improvements before test.   After consultations, which resulted in several suggestions for editing conversations, several modifications were added before this test. The first was to add global intents that allow the user to end the conversation or skip steps. The possibility of returning has also been added. For the purposes of the experiment, a telephone directory, a description of the washing machine and a description of important assistant instructions were added as props.

Test.   The testing itself went smoothly and for the first time the test participant was able to succeed in the end in all scenarios. In the first scenario, he had a problem removing the filter too soon, but that was only one minor problem. The participant was enthusiastic about testing and in the ensuing discussion we came up with some findings and possible improvements. When calling for service, the participant requested a call, even earlier than the assistant expected. Therefore, the participant suggested that it would be good to add this path to the previous point. He also suggested that it would be a good idea to start a conversation from a certain point.

There was only one outage in the fluency of the assistant, which was resolved by restarting the experiment.

Improvements after test.   Following the need to restart the entire scenario, a meeting with Flowstorm developers was arranged. Thanks to this meeting, it was possible to improve the assistant by the possibility of starting a conversation from a certain point. This eliminates the need to repeat the whole scenario. Subsequently, several intents were modified. Response options have been added or linked so that the assistant can respond to them correctly.

■ Participants 4 & 5 (SU4/5)

Unfortunately, after a successful test and subsequent extensive adjustments, there was another service outage. Automatic speech recognition (ASR) did not work this time. These two tests were scheduled in one day. The first test was to take place in the morning. The participant had great patience and tried to wait until the outage disappeared. In the end, however, the moderator had to cancel this testing because the ASR did not respond.

Before the afternoon test, we were advised to use the written answer form. Although it was not user-friendly at all, the participant was willing to try it. In the end, however, the written form failed and the test was canceled again.

Improvements.   This test was followed by a discussion of the stability of the Flowstorm system. Although the system is very good and has a great future, the momentary instability was critical for our testing purposes. The discussion showed that it would be better to fail the virtual assistant test and focus on modifying the prototype to support emotions. This also created a

requirement for the prototype to contain an unexpected misunderstandings - random error.

■ 3.4.3   Partial answers to research questions

Although we did not fully test, we tried to answer predefined questions so that we have at least a partial insight into the problem.

■ P1 - Are people able to ask for the help of a virtual assistant in repairing/maintaining a washing machine?

All test participants, despite outages and unfinished scenarios, reported that such a virtual repair or maintenance assistant was a great idea. Three participants said it was better to ask a virtual assistant than to call someone on the phone. Thanks to that, they are not so ashamed that they cannot do it without help. The assistant will not tell it anywhere.

■ P2 - Is there some benefit of the virtual assistant opposed to the previous prototype in 2.4.1?

The benefits of the virtual assistant could not be evaluated because we were unable to fully perform the experiment. Therefore, we cannot confirm the presence of benefits.

■ P3 - How do people behave if their assistant has difficulty understanding?

Experiment shows that people start to get irritated. They raise their voices until they almost scream if their virtual voice assistant doesn't understand them. The loop of the experiment (the participant refused to say the question differently and the assistant still did not understand) even became so frustrating that the scenario had to end unsuccessfully.

■ 3.4.4   Discussion

The word "pilot" in the title of this part of the thesis is not accidental. During testing, we encountered major outages of the Flowstorm online development environment. So after pilot testing, it was evaluated that this path is currently closed.

Although this problem has not been fixed, it is very important to state in this section the testing procedure and especially the reason why it did not work. Subsequently, we will also present the possibilities of correcting unstable tools during the testing.

3.4.5 Known problems

On-going development of Flowstorm and stability

During testing, we encountered errors with the unstable Flowstorm several times. This was mainly due to the fact that the product is still in development. New versions are deployed once in a while, so it is possible that a bug will enter production. Unfortunately, this is at odds with our testing, where it needs a stable, error-free version.

Possible solution. One possible solution is to create a separate version on a private server to which the moderator will have access. It can thus lock the versioning on the day of the test, or test its stability before the test itself.

Offline backup

One of the interesting problems that arose during the development was the absence of an offline backup. At one point, the development of the individual scenarios took place during the deployment of the new version of the Flowstorm product. An unknown error caused the scenarios to be unavailable and possibly a minor heart attack by the author. Fortunately, in collaboration with the Flowstorm development team, the scenarios were made available from a new account. However, the old account is now useless. It would therefore be a good idea to have an offline backup of the scenarios.

Chapter 4

Prototype voice assistant with emotion module

In the previous chapter, we unsuccessfully tested a virtual voice assistant, which was based on scenarios in 2.4.1. We need to take a different direction in order to deal with the emotional aspect of the conversation and design an emotional module.

We therefore decided to make full use of the prototype from 2.4.1 and try to design and implement an emotional module to it so that we could experiment with the module's functionality.

4.1 Used prototype

First, we need to get a little closer to the prototype of 2.4.1, which we are going to expand. This part of the chapter deals mainly with the used prototype and its possibilities so that the reader has a better understanding of the following design.

4.1.1 Type of communication

The prototype voice assistant works on the principle of the Wizard of Oz [13]. This method is used for testing with users and the prototype, although it looks completely automated for the test participant, and is controlled by moderators. The name Wizard of Oz refers to a book by L. Frank Baum. In this book, there was a wizard who created an improved image of himself by manipulating controls. In prototype meaning, it creates an illusion of a fully automatic version, thanks to which we can observe relevant test results.

4.1.2 Possibilities of prototype

The prototype has several blocks that the moderator controls. This section lists the most important sections.

Figure 4.1: Prototype from previous study [30]

■ Scenarios

In the middle of the prototype is XML, which shows the conversation diagram. A total of four scenarios (S00, S01, S02, and S03) are available, each addressing a slightly different issue. The scenario is always selected when the page is loaded.

Figure 4.2: XML with conversation

Each scenario contains two basic types of nodes.

Assistant's speech. This is the assistant's speech to the user, in which he tells him information about the given point in the scenario. Only the assistant always uses this speech. The assistant's speech is rectangular.

User's response. A moment ago, we showed the assistant's speech. In order for the moderator to be able to follow the script, potential user responses have also been added. Depending on how the user answers, the assistant chooses the next speech. The user response has an oval shape. The prototype was designed so that each moderator could create their own scenarios. The scenario consists of several parts. For us, two main ones will be relevant - the SVG image, which is displayed after the prototype is loaded, and the clickable XML, which is created in order to trigger the voice of the speech. Each speech refers to the soundtrack with that speech. This clickable XML can be selected in the settings, which we will cover next.

31

■ Top right panel

The right panel serves as the prototype settings. The most important functions here are the **Start** - start testing, **Stop** - end testing, and **Upload XML** buttons. The last button allows you to put the sound overlay to the scenario, so that you can play the sound after clicking on the individual nodes. Other settings on this panel are not relevant for our testing.

Figure 4.3: Right panel for settings

■ General speeches

General speeches give the moderator the option to use general messages. Speeches are the same across all scenarios and can be used by the moderator if appropriate during the testing. As a result, the scenarios are smaller, as we do not have to give these general speeches everywhere.

They are also available in several colors. Each color represents a different type of speech.

- ■ **Blue** - Basic neutral speeches to support conversation.

- ■ **Black** - Slow down the conversation types of speeches.

- ■ **Red** - Error speeches.

- ■ **Green** - Praise speeches.

Figure 4.4: Different general speeches

4.1.3 Controls

When loading the page, the moderator first selects the required scenario by typing S00, S01, S02, or S03 into the input. Based on the selection, the given XML is loaded in the middle of the screen. Subsequently, it is necessary to put in the sound layer of the scenario (general speeches already contain the sound layer). This is done by clicking the load XML button and selecting a clickable XML layer (sXY_click.xml - XY represents the number of chosen scenario).

Once we have the scenario ready, we need to press the start button to measure our time and create a log. The moderator then clicks on the individual speeches (squares) and monitors the user's answers (ovals). At the end of the scenario, it is needed to press stop. As a result, the end of the conversation is recorded in the log.

The log is available in the browser console. It is most often opened with the F12 key.

4.2 Design and implementation - Emotion module

In this section, we'll take a whole prototype voice assistant from 2.4.1 and try to extend it with an emotional module. This module will consist of several newly designed paths in each scenario so that we can respond appropriately to emotions during some of the user's responses. These new paths will be designed based on what is important at the time to execute the scenario and how it may be affected by the user's emotions. An integral part of the

33

implementation will also be the creation of a tool that can capture emotions to get rid of the moderator's bias.

◼ 4.2.1 Scenario-specific emotions

The main goal of this part is to find places in conversations where the emotion aspect should be triggered. Part of this goal is also to decide what to do on each type of emotion. We will start with different types of possible conversations.
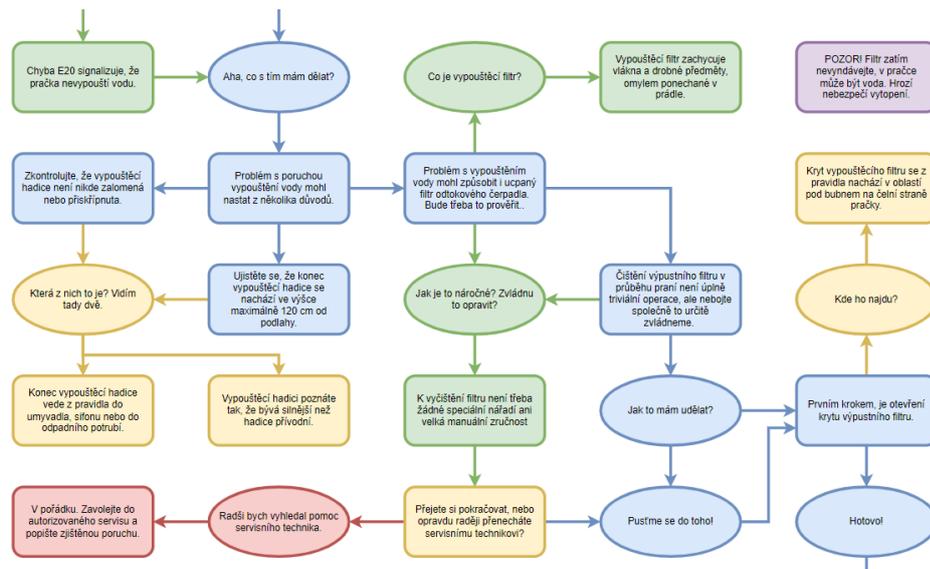
Figure 4.5: Different types of speeches in scenario

The original prototype of 2.4.1 contains several types of different paths the user can take. He is on the **main path** most of the time. This path is marked in blue and is the most used passage through the scenario. There are various turns from this main path. One of them is a **detailed path**. This path is marked in yellow and is used when the user wants to know more information. In addition to the detailed path, there is also a **bad path**, which is marked in red. This path is undesirable because it usually means that the user did something wrong or ends the scenario unsuccessfully.

Additional information path also exist in scenarios. It answers additional questions when the user asks them. There are also **critical points** in the scenarios. These are speeches that are important and the user must be very careful when carrying out the activity. You can see the part of the scenario that contains all the mentioned paths in the picture 4.5.

In the following points, we will supplement the existing scenarios with new types of paths that will depend on the user's emotions. We will generally call it the emotion path and mark them in orange. These paths will respond

to the emotions experienced in user responses. Answers that will be able to respond to emotions will be marked with an orange outline.

■ Uncertain path

At some points of conversation, assistant can detect that the user is not certain about what he should do. The user usually experiences emotions such as uncertainty, fear, or surprise. After this situation, we should trigger an uncertain path.

Proposed design. At this point, the assistant needs to slow down, pause over what the assistant has already said, and repeat it. There is also an option for the assistant to go into more detail. The main goal of this journey is to provide certainty to the user. For example, if a user is unsure, they may have missed a step or performed incorrectly.

- **Emotions detected** - Repeat last few steps or go to more detailed version of these steps.

- **Emotions not detected** - Continue with the main path.

Fear / Sad / Surprise

Are you sure you have the right cover? Located in the front area under the washing machine drum. The door is easy to open.

It is done.
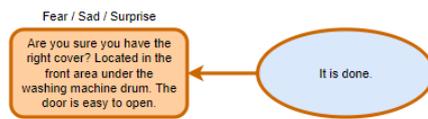
Figure 4.6: Uncertain path

■ Happy path

The user is generally happy and surprised and feels good. Our task will be to support him in keeping him in the same mood or even improving it to enthusiasm.

Proposed design. In user's response will detect happiness and surprises. In that case, the user will be offered a message that will further encourage him.

- **Emotions detected** - An encouraging speech is used by the assistant.

- **Emotions not detected** - Conversation stays on the main path.

Happiness / Surprise

Good work. The hardest part of cleaning the filter is done. Also try to check that the water is not rusty. This could mean rust on the washing machine drum.

All the water is out.
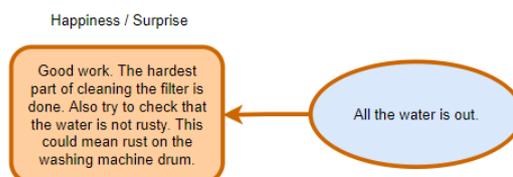
Figure 4.7: Happy path

■ Dangerous path

These places in conversation are really dangerous and critical. These are usually places when there is no going back and it is not possible to restore the washing machine to its original condition in the event of an error. In these cases it is really important to make sure that the participant is ready to do that and is really prepared for a possibly dangerous activity.

Proposed design. It is really important to watch for negative emotions such as anger, fear, and confusion. User has to be sure that he wants to do the next steps. Depending on the situation, it is necessary that the assistant does not let the user go on immediately unless he is absolutely sure. There are several options that can occur. **Relaxed approach** We are mostly listening for negative emotions. When negative emotions are detected then we may ask him if he is sure or not to continue with the other steps immediately and end the conversation.

- ▪ **Negative emotions detected** - Ask again if the user is sure or end the conversation.

- ▪ **Negative emotions not detected** - Continue or ask again if the user is sure.

Strict approach We do not allow any uncertainties. In this approach, it is important to not allow the user to do anything unless he is absolutely sure that he wants to continue. Strict approach is designed for the most critical places in the conversation.

- ▪ **Positive emotions detected** - Allow user to continue or ask again if he is sure.

- ▪ **Positive emotions not detected** - Strictly end the conversation.
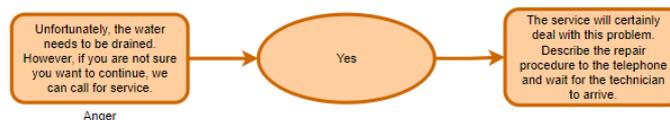
Figure 4.8: Dangerous path

■ Appeasement path

It often happens during a conversation with an assistant that the participant is scared, a little upset, surprised and sad. At this point, he has no idea what awaits him, or what will happen. We will call this state Appeasement state. Instead of continuing at this point, there should be some reassurance - Appeasement path.

Proposed design.   The user needs to be reassured that there is no need to worry. Tell him that the repair is not difficult and that if you proceed together step by step, you will manage everything. It gives the user the impression that he is not alone and that he could handle it. Alternatively, the user can be offered a service call if repairs are feared.

It would not be good to calm down users at this time. For example, tell him that he seems scared or upset. This is definitely out of place and should be avoided. This can upset the user even more or confirm that he should definitely not continue.

- **Emotions detected** - We'll add a node to the conversation to reassure users that they don't have to worry about continuing.

- **Emotions not detected** - Continue with the main path.
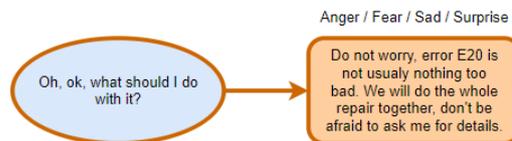
Figure 4.9: Appeasement path

■ Decision point

There are also points in the conversation where you need to respond differently to different types of emotions. This point is called the Combined point. In most cases, this is a combination of several approaches outlined earlier in this chapter.

Proposed design.   The difference from previous paths is that we do not follow only one emotion, but more at once. For example, if the user has the task of cleaning the filter and says he is done. Two emotions at once can be important at this point.

- **Anger** - The user is exhausted from repairs or maintenance, he wants to quit as soon as possible. It is also in the assistant's interest to come to an end quickly, as the user may be less careful in anger.

- **Fear, sad, or surprise** - The user is not sure if it is cleaned well. It helps him when the assistant says he should do it properly so that he doesn't have to do it again soon.

- **Other emotions or no emotion** - Continue with themain path.

Figure 4.10: Decision point and it's paths

■ 4.2.2 Global emotions

So far, we've only focused on emotions that were specific to one scenario. However, just like general speeches (Yes, Great, I don't understand), there are general reactions to emotions. These responses can be used across all scenarios at any point in the conversation.

Unlike scenario-specific emotions, these reactions cannot be used to only by detecting the emotion.

Example. The user's emotion "happiness" was detected. In general speeches, there is a sentence that it is surprising how easy repair can be. However, if we look at the overall representation of the emotion of happiness, among other emotions, it is unique. For example, anger prevails. The user is annoyed that it should be easy and the consequences will be such that the user is likely to stop listening to the assistant.

■ Emotion level

In order to use general announcements, the prototype needs to be improved by the level of emotions. This setting is available on the left side of the prototype. It is thus possible to set a separate level of representation for each emotion, among other emotions, for which the message is available. The default setting is 50% for all emotions. You can check the setting in the figure 4.11.

Figure 4.11: Emotion level setting

Although the individual general emotional speeches are hidden and revealed automatically, according to the representation among all emotions, the prototype was equipped with a histogram that graphically shows the representation. The histogram can be used by the moderator to monitor the levels at which individual emotions reach, or to decide on scenario-specific speeches based on the levels. The histogram was added to the upper right panel of the settings, as it was not used during testing at all. You can see the histogram in the figure 4.12.
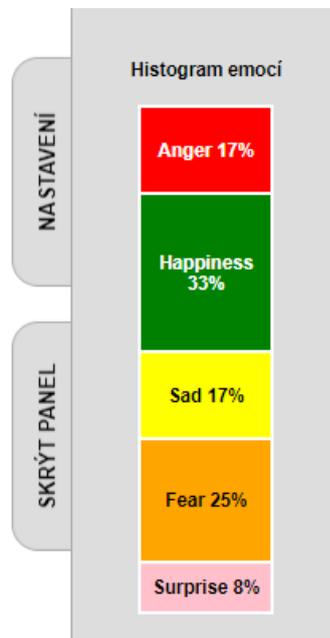
Figure 4.12: Histogram of emotions

Block with emotion level collapses when the start is pressed. Setting should not be changed during testing.

39

■ General emotion speeches

As already mentioned, several general speeches have been created in the prototype, which is tied to emotions. This section lists the individual emotions and related speeches.

Figure 4.13: General emotion speeches

Anger. In case of predominance of anger among other emotions, we can offer the user alternative ways. If the user is angry, his actions can be dangerous. He can damage some components irreversibly in anger. The following speeches can be used:

- This activity is usually challenging. We can call the service if you want.

- If you want, we can take a break. The work doesn't have to be done immediately.

The first option is an alternative ending. We give the user space to decide if he wants to continue at all. The second option is less radical. We offer the user a break to gain strength and continue in a slightly better mood. Both messages should be used later in the scenario - when we have a better overview of the representation of anger among other emotions.

Happiness. Same as in scenario-specific speeches, here happiness means that the user is on the right track. Maybe he's even experiencing "Flow" [26]. To keep him on this path, we want to support him with compliment.

- Awesome. You're doing great.

- You deserve a big dinner tonight!

The first message is a simple compliment, thanks to which the user will feel good. The second is a smaller, funny speech. Although it may seem

unnecessarily complicated, it is surprising how positive an effect it has on the conversation with the user.

Happiness and surprise. The combination of these two emotions suggests nothing more than that the user is pleasantly surprised. At this point, the assistant has the opportunity to use the message, which confirms to the user that it is really simple and does not have to worry about anything.

- It is surprising how easy repair can be.

- It is surprising how easy maintenance can be.

These messages are semi-specific, as they depend on whether the scenario is undergoing maintenance or repair.

Sad, fear, and surprise. As we know from previous chapters, the combination of these emotions is actually uncertainty. At this point, we should be able to provide some insight into the problem again, or repeat the speeches. In the general case, we will offer repetitions.

- If you are not sure if you have understood correctly, we can repeat the step.

- Let's rather repeat this step.

In the first speech, we offer the user the option of repetition. On the other hand, the second speech immediately notifies the user that it will repeat the previous speech. This speech can be used even in case of long user inactivity.

■ 4.2.3 Emotions counters

In order to be able to record individual emotions, the prototype was equipped with emotions counters. There is a separate counter for each emotion, which represents the number of measured points of that emotion on the user. As the moderator may click, there is also the option to remove the emotion point or delete all counters altogether.

■ Connection with emotion level and histogram

The histogram reacts to the change of individual counters. The ratio of individual emotions changes with the press of counters. For each emotion, the percentage is calculated as follows:

$$\textbf{Emotion level = Emotion counter / sum of all counters}$$

After each counter change, general messages are hidden or uncovered, depending on whether they meet the set level of emotions. During testing, only speeches that can be used are available in the left column of the emotion block.

Figure 4.14: Emotion counters

4.2.4   Emotion detection application

Everyone perceives emotions differently, and has different qualities of their perception and is not always able to determine them. To this end, there has been a demand for a general tool that will not depend on humans, but will capture emotions automatically.

In the analysis, we went through several options for capturing emotions. There are basically two main ways to perceive emotions - voice and face. The voice seems to be very obvious and unambiguous, but it has several fundamental problems. We have therefore considered the pros and cons of using different approaches.

▪ Voice

Pros.

- Voice assistant is using voice perception anyway.

- Less invasive than capturing face from a camera.

Cons.

- We cannot recognize emotions immediately. Longer text is required.

- We can only know sentiment immediately.

- There are few frameworks that can be used.

▪ Face

Pros.

- We recognize emotions from points in the face very well and immediately.

- There are several frameworks available that can be used.

Cons.

- Very invasive method. People don't like being watched.

- The need for a camera.

Decision

In this part of the thesis, we work only with a prototype of a voice assistant, which also does not capture the voice automatically. Therefore, voice sensing loses one advantage.

The availability of possible frameworks for capturing emotions also played a big role in the selection. After analyzing possible frameworks for capturing emotions from the face, we found that a very expensive camera is not needed. For example, the AffDex SDK [22] can detect emotions even from lower video quality. This allows you to use a cheaper HD webcam. This eliminates one point from the cons.

Finally, we focused on the user aspect. We know that it's certainly more user-friendly when the cameras aren't pointed at them. Voice recording is definitely better. Unfortunately, we do not always have only the pros themselves, so we decided to sacrifice this point and go with benefits of using the camera.

Frameworks

The basic requirement for the framework that we will use in the application is the ability to detect these emotions - anger, happiness, sadness, fear, and surprise. In this chapter, we will briefly describe the two most well-known emotion capture frameworks.

Affdex SDK by Affectiva. This tool can analyze and capture emotions in real time. It is also a cross-platform tool (Android, iOS, Windows) that is available for Java, JavaScript, Python, and more. Affdex SDK can capture six or seven emotions - anger, happiness, sadness, fear, surprise, disgust (+ neutral expression) and is free to use. So it contains all the necessary emotions and is also the most widespread framework available. It also brought the possibility of using it in JavaScript. [1]

FaceReader by Noldus. This software can recognize and analyze emotions - happy, sad, angry, surprised, scared, disgusted, and neutral. Unlike the first variant, however, it is not free. [11]

Implementation

The need for an emotion capture tool has resulted in a minimalist application that provides a simple interface with real-time emotion determination.

The application is written in the JavaScript programming language, as the prototype and its side tools are written in HTML, CSS, and JavaScript.

When the application starts, the first tool available is the detection control panel. It contains a total of 3 buttons. Start starts detecting emotions from the default camera. This requires that you enable the camera for the page before you can use it. The stop button, on the other hand, stops detection. The reset button resets the detection tool to its initial state. You can see this panel in the figure 4.15.

Figure 4.15: Control panel of detection tool

When the detection is turned on, a preview from the camera is displayed and a list of individual emotions with the current percentage level is displayed below it. The level changes instantly, so the moderator can monitor the changes smoothed as the user answers the assistant's questions.
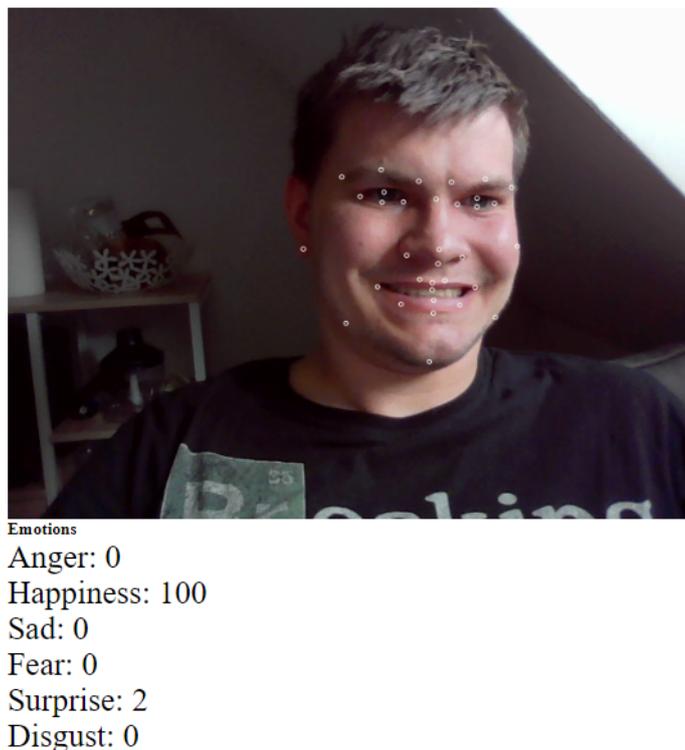
Figure 4.16: Detection of emotions and camera preview

The application is also equipped with a log of events that took place. This log is for information only and has no deeper meaning.

Figure 4.17: Log of actions

The whole application was inspired by the implementation of a similar tool on the Github platform. [8]

4.2.5 Unexpected misunderstanding

Based on testing of the virtual voice assistant, it was found that the virtual form of the assistant in contrast to the prototype is very erroneous. It often happened that the assistant did not understand what the user was saying.

Example. The user wants to know how to open the washing machine door. That's why he says, *"How do I open the washing machine door?"* The assistant does not understand him, so he says *"Can you repeat that, please?"*.

Such misunderstandings can occur for a variety of reasons. The assistant does not understand a word in a sentence or there is an unrelated sound around the conversation. It is also possible that only an error occurs in the assistant.

Therefore, the prototype has to be extended by a set of random errors (unexpected misunderstandings).
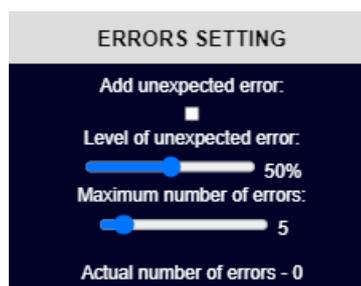
Figure 4.18: Random errors settings

The picture shows a total of three setting options and one informative text. We will go through the each individual one.

Add random error. If this checkbox is checked, then the unexpected misunderstanding is turned on. Otherwise, it is not used at all, and other settings in this section are unnecessary.

Level of random error.   This is the level of probability at which an unexpected misunderstanding occurs. If this setting was 100%, then the error would be reported over and over again, if it was 0%, then the error would not occur at all. The optimal number can be set by each moderator.

Maximum number of errors.   This number shows the maximum number of errors that can occur from the start of the scenario to its end. Below this setting is also the current number of errors.

▪ Speeches used when error occurs

If such an error occurs, then an error speech appears instead of the speech the moderator wants to run. In total, two different error speeches are used. The moderator must not repeat the original speech at that moment and must wait for the user to repeat it. Speeches used:

- Can you repeat that?

- What do you mean?

▪ 4.3   Experiment

In this section, we will focus on the experiment itself, which will monitor the contribution of the emotional module in the prototype compared to its version without the module. The experiment setup is largely based on working in 2.4.1.

▪ 4.3.1   Research questions

As part of the experiment, we will try to answer the following research questions.

- **Q1** - What are the most common emotions in washing machine repair / maintenance scenarios?

- **Q2** - What effect do emotions have on the course of the experiment?

- **Q3** - What is the option of users on sensing emotions?

▪ 4.3.2   Participants

First, it was necessary to specify which participants we will need and what the recruitment process will look like.

▪ Target group

As the whole setup of the testing apparatus was complicated, we also had to perform several pilot tests to find the ideal setup. In total, we need **12 participants** in the experiment. Thanks to the same number of participants, there is a better opportunity to compare the results and benefits of the emotional module.

▪ Recruitment

As we did not use all available resources to test the virtual voice assistant, we were able to contact these participants for a prototype experiment. Older people became the most important contacts for us in this experiment.

- We contacted the retirement home in Kladno.

- We contacted workers from the office in Děčín

- Contacts obtained from Facebook group or from the previous form for the first experiment was included as possible candidates.

▪ 4.3.3 Environment

Before the experiment itself, several pilot tests took place, which helped us prepare for testing, find out the ideal room settings, and set up a prototype.

Since the tested target group are mainly seniors, it is necessary to adapt to them. Therefore, we prepared for the possibility of moving to their environment. The test environment is again richly equipped with props and will also be adapted to capture emotions.
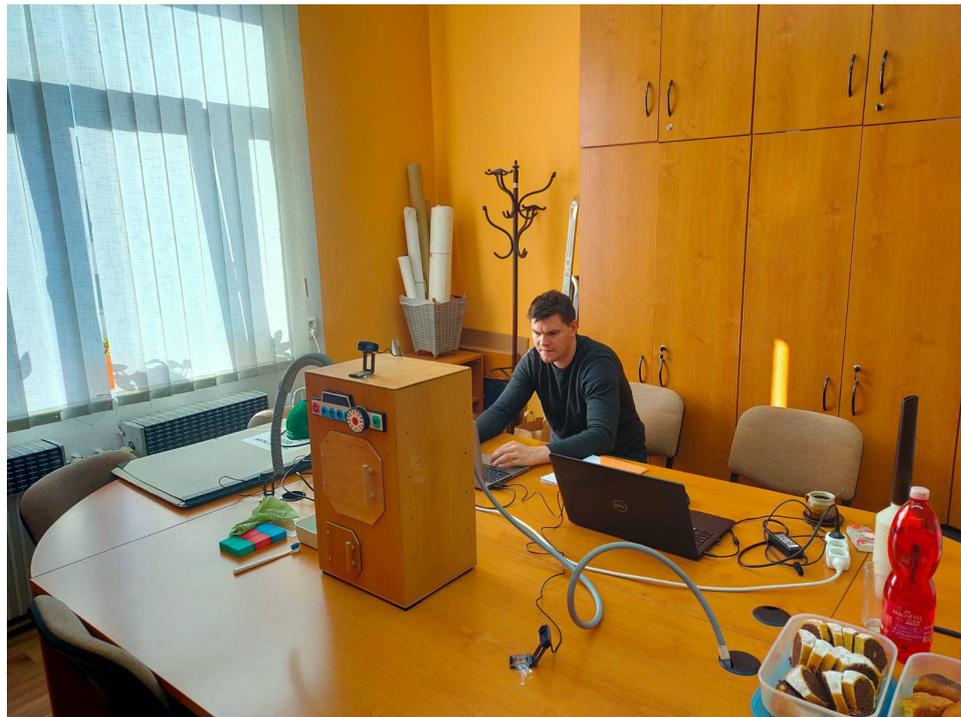
Figure 4.19: Environment adapted to capture emotions

■ Technical specification

Cameras. The biggest difference compared to testing with a virtual voice assistant is the presence of cameras. These are used precisely so that the implemented application can capture emotions. A total of 3 cameras are mounted around the washing machine, each with a different view of the tested user so that we can monitor his emotions in different positions.

Another problem occurred when exposing the cameras to the washing machine. The moderator must be within range of the cameras to be able to plug them into his laptop. The test environment has been modified based on this issue.

Second moderator. Unlike the first experiment with a virtual voice assistant, there was also a need for a second moderator. One of the moderators manages one instance of the prototype and focuses mainly on going through the scenario, the other with another instance of the prototype monitors the script, emotions and adds emotional counters based on that.

Devices. The equipment of the room has thus been expanded to include two laptops containing extended prototypes and one monitor, on which there

are three instances of the emotion monitoring application. Each instance is linked to a different camera.

■ Room preparation

The first fundamental change is the fact that the washing machine is no longer on the table by the wall, but on the table in the middle. The table is wide enough to fit a washing machine, a container in front of the washing machine, and one of the moderators behind the washing machine with his laptop and screen. This table is also long enough to accommodate the drain hose container (if available) next to the washing machine and the props on the other side.

The second moderator is placed with his laptop on his own table in the immediate surroundings of the user so that he can interact with him. You can see the setting in the figure 4.20.

Figure 4.20: Setting of the table

■ Props

For the authenticity and relevance of the experiment, it was necessary to provide several items that were mostly also used to repair the washing machine in previous experiment by R. Talácko [30] and the experiment with virtual voice assistant. These items are for example toothbrush, containers, foam cubes or a rag.

■ **Settings of the prototype**

Emotion level. This level was set at 50% for all emotions. If the emotions in the histogram exceed this value, possible messages in general speeches are displayed.

Unexpected misunderstanding. The unexpected misunderstanding level is set to a maximum of 3 errors and it occurs with a probability of 8%.

■ **4.3.4 Process**

The aim of the experiment is, as already mentioned, to compare the results with the experiment on which ours is based. Thanks to this, we can summarize the benefits of the existence of an emotional module in the voice assistant. Due to the use of a prototype and not a virtual version of the assistant, the Wizard of Oz method will be used for research. This is an alternative solution that was chosen based on the unfavorable results of the virtual version of the assistant. The whole testing process is described in the session guide for the tested users.

■ **Introduction**

First of all, it is necessary to say a few words about yourself and our assistant - name, school, faculty, field, and a short reason for the meeting. This is followed by an introduction to the informed consent, its completion, and signing. In this section, we emphasize that the user does not worry about his privacy. It is therefore necessary to say that the recording is for personal purposes only and the cameras do not shoot, they only capture emotions in real-time.

■ **Experiment introduction**

In the first phase, it is necessary to present the prototype that the experiment will cover. Describe to the participant that we are currently developing a voice assistant at the department. It is necessary to briefly outline the principle on which the assistant works and that in order to create such an assistant, several prototypes are created. The prototype that users test focuses on the emotional aspect of the conversation.

This is followed by a request to try to shield users from the fact that it is only a prototype and to try to communicate with the assistant in the same way as with a virtual assistant. We can also give examples of other voice assistants for a better idea (if they know any).

In the next section, it will be necessary to say that users will perform repairs or maintenance of the washing machine with the help of the assistant. Each of these actions is presented in a slightly different scenario, which

will be presented before its implementation. This can be followed immediately by a washing machine and props. It is also necessary to say why there are cameras and again assure the user that there is no recording of them.

Finally, in this section, it is difficult to introduce a form of communication with the assistant. A comparison with, for example, a remote repairman is suitable. It should be noted that the assistant does not see what they see, so things need to be described.

■ Scenarios

In this phase of the experiment, we try to bring the individual scenarios closer to the experiment participant. In this experiment, we use a total of four scenarios, all of their introductions can be found in this 3.2.5 section. These scenarios are adjusted according to the prepared emotional module. This part only describes expectations in terms of emotions.

S00 - Introduction. Example of a conversation assistant.

S11 - Error E20 - Expectations. This scenario is already modified by several paths with emotions. Since this scenario is the longest, we expect the greatest involvement of emotions of all scenarios. The scenario contains several difficult points where emotions will occur and need to be monitored. The first such point is to open the drain filter cover, where the user may feel unsure whether he has the correct hole. In this case, the assistant can help prevent confusion.

The next point is to unscrew the filter. At this point, we are again trying to monitor the user's uncertainty, which tells us that he is afraid it will overflow. In this case, the assistant will assure the user that he can screw in the filter again and pour out the container. This will prevent water overflow.

There are a few other points in the scenario that react to happiness, for example, and in that case, praise the user. This will give them confidence in their work. It is then difficult to react to anger. If the user is notified that a lot of water is flowing from the washing machine, we will offer to end the scenario and call for help.

S12 - Cleaning the supply hose filter - Expectations. Right at the beginning of a conversation, users should be commended if we detect happiness. Thanks to this, we can motivate him to take further preventive actions that will extend the life of the washing machine.

Here are some points where the user may feel insecure. In that case, we will again choose a more detailed path - for example when choosing the right hose or screwing the hose back on.

Particularly specific is the part where the user closes the water and fails. He's getting upset. At this point, the user needs to be guided correctly to close the cock, but at the same time tell him to be careful that the cock can break. This will prevent reckless dealing and breaking of the washing machine.

S13 - Flood - Expectations. Uncertainty can again arise at each point of control. In this case, we expect the assistant to prevent the user from taking the wrong step - opening the washing machine drum or unscrewing the filter or supply hose.

It can also happen that the user does not find the fault correctly. In that case, he may be unsure of his answer. The assistant should recognize this behavior and describe to the user how to check the fault in great detail.

In case of possible anger when finding a fault, we expect the assistant to offer a quick call to the service, instead of further speeches.

▊ Discussion

At the end of all scenarios, we need to have a short conversation that leads to a discussion with the user. The interview has several prepared questions that are open and can be discussed.

- How do you fell?

- How natural was the communication with the assistant?

- Have you noticed the reactions to current emotions? Where and how do you feel about it in general?

- Did you notice the assistant's misunderstanding? Where and do you think it is relevant?

- How would you rate the overall usability of such an assistant with an emotional module?

From the discussion, we are most interested in reactions to emotions. Try to understand whether it is perceived as a benefit or as a burden. Find out under what conditions it could be beneficial. In addition to emotions, we are also very interested in the question of misunderstanding.

▊ 4.4 Analyses of results

▊ 4.4.1 Scenarios

▊ S11 - Error E20 with emotion paths

The scenario was the longest of all and therefore there are several problems. The first of these was to remove the filter. Participants often needed to

overtake the assistant and do things first. 7 participants leaked a small amount of water on a rag (or floor) and one even leaked completely.

The subsequent return of the filter caused a problem for a total of 6 people and one failed to return it at all. The others, with the help of an assistant, finally succeeded.

None of the participants terminated the scenario before the end, although they were offered it 3 times based on emotions. One participant reported that the water drain hose was too high.

Emotion paths. Uncertain path emerged the most. Exactly 6 people used it, and they did it several times in a row in this scenario. As a result, they were able to complete some tasks faster.

Three participants were offered a dangerous path as anger was detected. However, none of the participants accepted the offered end of conversation.

In addition, the assistant also used a happy path, an appeasement path, and even a global response. However, the assistant used each of these paths only once.

■ S12 - Cleaning of the filter with emotion paths

This scenario went mostly smoothly without much difficulty. One of the participants wasn't sure what he would need a toothbrush for. Furthermore, the participants were not sure where to find the tap (5 times) or the supply hose (4 times). Two participants pushed the washing machine away closer to them.

Cleaning was not done twice thoroughly, but it was no problem.

One participant had a problem screwing in the hose, but in the end, he managed it. Opening the water to participants was also not a problem. They just weren't sure of the tap's position from time to time.

Emotion paths. Most of the participants felt happy during the scenario. Based on this, the happy path was run 10 times. One of participants was also praised for preventive cleaning. However, it had no effect.

Furthermore, fear and surprise reappeared in the scenario. Thanks to this, the uncertain path was used 3 times. At one point, there was also a decision point at which a dangerous path was subsequently used based on the detected anger.

■ S13 - Flood with emotion paths

In most cases, this scenario was fine. During the initial inspections, the washing machine drum was opened twice and one participant reported a defect in the supply hose.

At the drain hose, exactly half of the participants were not determined that the defect was right there. That's why their assistant had to help them. Unfortunately, two participants did not find a defect at all.

During the subsequent call, 8 participants described the fault, 2 participants described the entire course of the inspection and 2 unfortunately only said that they had a flooded bathroom.

Emotion paths. This scenario is very critical. The flood is not something to cause happiness. Therefore, there are only 2 different paths (dangerous and uncertain) and global speeches to use in this scenario.

Uncertain path was used a total of 10 times. Of these, two participants used it 4 times in different places during the scenario.

The Dangerous path was used only once, when the assistant made the call quickly so as not to deepen the participant's anger. However, this does not mean a failed end, only the assistant did not say the last few speeches, only important ones.

■ 4.4.2 Participants

■ General

Before each experiment, each of the participants received a questionnaire that showed us their profile. Individual questions of the questionnaire:

- **Gender**

- **Age**

- **A1**: Would you say that you are a technical type?

- **A2**: How would you rate your manual skill level?

- **A3**: Do you have any experience with repairing the washing machine? What experience?

- **A4**: Do you think you are communicative?

- **A5**: Do you have any experience with virtual conversation assistants?

You can see the possible answers to these questions in the attached form at the end of this work. The answers recorded in the table 4.1 are marked with

letters of the alphabet. These letters refer to the individual answer as they are listed sequentially on the form. For example, for question A1, the answers are:

- a) surely yes

- b) probably yes

- c) not sure

- d) probably not

- e) surely not

Participant	Gender	Age	A1	A2	A3	A4	A5
B1	Female	22	b	d	b	c	a
B2	Male	22	d	c	b	b	a
B3	Female	57	b	c	a	c	a
B4	Female	29	c	c	b	a	a
B5	Female	77	d	c	b	a	a
B6	Female	66	b	c	b	a	b
B7	Female	67	b	c	b	a	a
B8	Female	65	c	d	b	b	a
B9	Female	65	b	d	b	d	c
B10	Male	71	a	b	a	b	a
B11	Female	74	d	c	b	a	a
B12	Male	65	b	c	b	c	a

Table 4.1: Testing profiles

The first four participants (B1-B4) are under 65 years old. The others (B5-B12) are over 65 years old. We have a total of 3 men and 9 women. Their profiles are in table 4.1

Participants' responses to whether they were technical types varied widely. However, none of the participants stated that it was not a technical type at all. Most participants manage to repair minor damage on their own, and no one said they were a handyman.

Only two participants stated that they had experience with repairing the washing machine. However, during the experiment, several participants found out that they had done this before.

Participants usually have no experience with voice assistants. One participant said he used them occasionally and one said he tried.

■ **Participant B1**

Emotion	S11	S12	S13
Anger	0	0	2
Happiness	5	4	2
Sad	1	0	0
Fear	3	2	1
Surprise	4	1	2

Table 4.2: Emotions counters of participant B1

Scenario S11. The participant proceeded very carelessly and tried to go ahead in the scenario. Unfortunately, this caused the water to drain to the floor. The other steps were already going smoothly. Participant even managed to screw the filter back on the first try.

It is even more necessary to discuss which **emotions** were recognized by our application. The table 4.2 shows the representation of individual emotions.

Emotions in S11. The participant entered the performance of the scenario in a good mood. Although the scenario should create a feeling of panic, the participant was rather looking forward to solving the problem. The change in emotions occurred later at a very rapid pace. Since then, the participant has been insecure. However, it was the detailed path associated with the feeling of insecurity that helped the user to correctly guide the filter lock into the hole.

Number of unexpected misunderstanding speeches: 0

Scenario S12. In this scenario, the participant proceeded exactly as given. First, the participant himself asked for help with cleaning the filter. Furthermore, the user was not sure where the water tap was located but was not afraid to ask. Participant cleaned the supply hose filter and screwed it back on quickly. In the end, he just turned on the water automatically and it was all done.

Emotions in S12. There was no significant emotional point in this scenario. The user was only unsure about the tap closure. However, before the assistant was able to react, the participant asked himself how to close the tap.

The assistant responded only to the emotions of happiness, when he woke the participant by saying that the hardest work was done. This reaction brought a smile in the next steps.

Number of unexpected misunderstanding speeches: 0

Scenario S13. The participant followed the instructions. He responded confidently to the checks of the individual options and, as a result, passed

them quickly. He also correctly identified the cause of the problem and called for help with his assistant. He only said in the description that the drain hose was disconnected and that there would probably be a problem.

Emotions in S13. The participant got into the script better this time, trying to be upset that his washing machine had leaked. This anger was revealed by an application. Later, the assistant detected only minor uncertainty at the disconnected hose so uncertain path was used.

Number of unexpected misunderstanding speeches: 1

Reaction to unexpected misunderstanding. The unexpected misunderstanding speech occurred only once. The user simply repeated what he said and went through the whole situation.

Discussion. The participant felt fine. The naturalness of the communication with his assistant seemed fine to him but was aware that he often acted a bit more. He also mentioned that he goes to these repairs with a cold head.

The participant did not notice the reactions to his emotions, although there were some reactions. The assistant was not annoying and invasive, which participant takes as a big bonus. The participant did not notice the unexpected misunderstanding at all and does not remember that it would be in the scenarios.

The participant considers the overall use of the assistant to be very beneficial. He could imagine it at home and sees emotion sensing as a bonus that can help in tense situations where the assistant can prevent damage.

Participant B2

Emotion	S11	S12	S13
Anger	8	7	1
Happiness	0	1	0
Sad	0	0	1
Fear	4	0	1
Surprise	7	3	3

Table 4.3: Emotions counters of participant B2

Scenario S11. The participant went through the first part scenario without any problems, he asked about the height of the hose.

The first frustration occurred when the assistant answered slowly. This was followed by a series of questions about the drain filter - can I open the filter when the washing machine is on, how high must the container be, will

this rag be enough? The speech that it was necessary to drain the water slightly irritated the participant, as he did not know if he should shut the water tap to the washing machine. Unfortunately, the speech about screwing filter back when needed came late, as the filter was already out. The participant then failed to return the filter and the water leaked out.

Subsequently, the participant cleaned the filter and tried to return it. The filter has already been returned without any problems. Eventually, the participant requested that an assistant could call for a staff to clean the water.

Emotions in S11. The participant got into a situation greatly and in some situations proved to be very uncomfortable. Great frustration arose, as already mentioned, after the need to drain the water. The overflow of the container itself only escalated the situation. Upon anger, the participant was offered to call for help (dangerous path), but he vehemently refused.

In addition to the permanent anger, surprise and fear also appeared during the scenario testing. Taken together, this represents uncertainty. The assistant used the uncertain path several times to help the participant better screw in the filter or find the filter cover.

Number of unexpected misunderstanding speeches: 0

Scenario S12. The participant approached the assistant himself, then with his help he pushed the washing machine away and unscrewed the hose. In doing so, he asked about the water tap and the correctness of the hose. The hose was cleaned very thoroughly. Then there was a problem where the participant thought that something else needed to be cleaned up as well and the assistant was unable to cope with this request. Subsequent screwing and tightening took place very quickly. The participant eventually opened the water.

Emotions in S12. At the beginning of the scenario, the participant did express happiness. The turning point came when the assistant was unable to cope with his request. Since then, the participant has been very frustrated and insane. He also exchanged several swearing words.

The assistant first responded to the initial happiness. After a complete change of mood, he tried to react to the anger and when cleaning the hose, he tried to speed up the process so that the anger did not develop further.

Number of unexpected misunderstanding speeches: 0

Scenario S13. The participant again contacted the assistant himself. Subsequently, he went through the individual options confidently. Only at the supply hose he did stop a little. He immediately revealed that the hose on

the back was not holding tight. In the end, the participant was not sure which defect it was. He told the technician on the phone that the problem is probably in the drain hose, that it was probably disconnected.

Emotions in S13. In this scenario, the participant did not show as much emotions as in the previous ones. The assistant reacted only once, when the participant did not react confidently (sadness, fear, surprise) and recommended not to unscrew the hose.

Number of unexpected misunderstanding speeches: 0

Reaction to unexpected misunderstanding. No unexpected misunderstandings occurred.

Discussion. The participant evaluates the communication with the assistant rather naturally. He mentioned that he was very upset when assistant was unable to respond properly. The participant did not notice the reaction to the emotions. He would expect the assistant to reassure the participant more when he is very upset. Or use motivational speeches.

The unexpected misunderstanding was not used, so he has no opinion on that. He evaluates the overall usability very positively. He can imagine it instead of various manuals and emotion module is bonus. But cameras might be invasive, voice emotion perception would be better.

Participant B3

Emotion	S11	S12	S13
Anger	2	1	5
Happiness	2	2	1
Sad	0	1	3
Fear	0	3	4
Surprise	0	4	4

Table 4.4: Emotions counters of participant B3

Scenario S11. The course of this test was very straightforward. The participant already had previous experience with cleaning the filter, so it was not difficult for him. At the same time, he was also very attentive. When checking the height of the drain hose and its breakage, participant was very quick with solving these tasks.

Before the assistant could react, he prepared a smaller container and a rag under the drain filter. He dropped the water slowly and tightened the filter quickly when necessary. Everything went without water. Minor frustration occurred when the filter could not be returned to the hole. But it

succeeded after a short effort. The participant completed the repair and even pressed the button on the washing machine himself to complete the wash.

Emotions in S11.   From the beginning, the user was very happy and happy. Anger was detected during the frustration. For the rest of the scenario, the emotions were neutral. Due to the low number of detected emotions, the assistant did not respond with any emotion speech.

During the conversation, there were a total of two speeches about the assistant not understanding. The participant considered this to be normal and repeated his questions.

Number of unexpected misunderstanding speeches: 2

Scenario S12.   The participant no longer had any experience with this operation on the washing machine, which was also evident in the course.

At first the participant was not sure what he would need a toothbrush for, then again uncertainly about the tap of the water. After unscrewing and cleaning the hose, everything went smoothly. When opening the water tap, the participant asked that it should be perpendicular to the wall or horizontal with the washing machine.

Emotions in S12.   From the beginning, expressions of fear and surprise alternated. It was clear from the participant that he was not entirely sure what he was doing. The turning point came after unscrewing the supply hose.

The assistant responded to several signs of uncertainty. For example, when preparing aids or choosing a hose.

The crucial reaction was to make sure that the repair had the hardest part (luck detected). This gave the user a feeling of relief and his next steps were confident and fast.

Number of unexpected misunderstanding speeches: 1

Scenario S13.   After asking for help, the participant gradually went through all 4 options. First, he checked the drain filter, the supply hose and the drum of the washing machine. Uncertainty was seen in each of these inspections.

When checking the drain hose, the participant decided that a small pull did not matter. Subsequent questions about certainty in their answer made the participant upset. He told the specialist that the supply hose had probably burst.

Emotions in S13.   Uncertainty in the first inspections led to preventive emotional talk, which warned that the drum must not be opened, the filter and the supply hose must not be unscrewed. The frustration and uncertainty

of checking the drain hose led to a detailed path that finally convinced the participant that there was something wrong with the drain hose.

The unexpected misunderstanding occurred only once, but the participant apparently overheard it and continued.

Number of unexpected misunderstanding speeches: 1

Reaction to unexpected misunderstanding.   The participant ignored the unexpected misunderstandings and considered them normal.

Discussion.   Although the participant seemed more frustrated as the scenarios progressed, his assessment in the final discussion was positive. Immediately after the last scenario, he felt good and satisfied. Communication with his assistant seemed very natural to him. He didn't notice the reaction to the current emotions, but if it should help with a greater personalising of the conversation, then he is definitely in favor.

He considers the occurrence of unexpected misunderstandings to be a normal matter. There is also noise in the conversation between people and sometimes you have to ask again. The participant would like to have a similar assistant at home. Best for other types of repairs as well - dishwasher, dryer or car.

■ Participant B4

Emotion	S11	S12	S13
Anger	3	0	1
Happiness	1	0	2
Sad	0	0	0
Fear	1	0	1
Surprise	1	0	1

Table 4.5: Emotions counters of participant B4

Scenario S11.   From the beginning of the test, the participant gave the impression that he wanted to get it over with. Checking the height of the drain hose, he reported that it would probably be too high and proceeded to the filter. He hesitated for a moment about where the drain filter cover was located. The water drain went without any problems, the participant used the method of substantiating the washing machine and draining the water at once.

A minor problem occurred only when the participant wanted to screw the filter back on. He did not ask for help or talk. After a while, the lock was brought back when the assistant asked if the participant needed help. The rest of the repair went without complications.

Emotions in S11. As already mentioned, the participant entered the test with the feeling that he didn't care, all to get it done as soon as possible. This also affected the detection of emotions. The application had trouble finding any at all.

The assistant only responded to the guidance where the drain filter is located, as the application detected fear and surprise when responding.

A total of 3 unexpected misunderstandings occurred during the scenario. One time the error occurred inappropriately and the participant did not know how to react to it. For the other two, he automatically repeated what he needed.

Number of unexpected misunderstanding speeches: 3

Scenario S12. In this case, it was visible that the participant had probably cleaned or replaced the supply hose at the washing machine before. He subsequently confirmed this in the discussion.

He proceeded exactly step by step and performed the individual tasks very quickly. The cleaning of the sieve was not very thorough and the participant immediately started screwing it back on. He then automatically opened the water. Everything went without complications.

Emotions in S12. Unfortunately, no emotions were detected in this scenario. Their levels were very low.

An unexpected misunderstanding appeared a total of twice, but right at the beginning of the same speech. The participant thought that the sentence needed to be formulated differently.

Number of unexpected misunderstanding speeches: 2

Scenario S13. Prior to this experiment, we assured the participant that this was the last scenario. This probably put the participant in a better mood. The participant handled all the checks very quickly and answered clearly. He did not open the drum or unscrew the supply hose and filter.

He did not reveal the defect at first. He succeeded on the second attempt. Then he informed the specialist on the phone that the drain hose was broken and did not add anything more.

Emotions in S13. The participant's behavior was influenced by the fact that he was looking forward to the end of the test. Therefore, for most of the test, he felt happy, to which there is no reaction in this case, as it is not expected in a flooded bathroom. Fear and surprise occurred during the inspection of the drain hose, and thanks to that, a uncertain path was later launched.

The unexpected misunderstanding occurred a total of twice and the participant took it automatically and repeated his words.

Number of unexpected misunderstanding speeches: 2

Reaction to unexpected misunderstanding. A record 5 errors occurred in this test as there was a higher level of unexpected misunderstanding. In most cases, the participant just repeated his words. One time the speech came inappropriately.

Discussion. After the overall test, the participant admitted that he did not feel very well because he did not like such tests. He immediately admitted that in the second scenario he knew exactly what to do, as he was exchanging the supply hose at the washing machine at home.

He considered communication to be more natural and compared it to Siri, which he thinks is more natural. He didn't notice the emotions reactions at all, and he doesn't think it might be the thing for conversation assistants.

He noticed unexpected misunderstanding, but said that it seemed very natural to him, because even in the real world, people often did not understand each other. He considers the usability of the assistant to be very useful and thinks that is the future.

"You just need to catch all the flies."

■ Participant B5

Emotion	S11	S12	S13
Anger	1	1	5
Happiness	2	3	0
Sad	3	0	1
Fear	6	2	3
Surprise	7	2	3

Table 4.6: Emotions counters of participant B5

Scenario S11. From the beginning of the conversation, the participant became uncertain about how to respond and was talking very slow. Right after the start, the participant asked if the error was about water drain system. The drain hose was inspected without complications.

At the opening of the filter, it was clear that the participant is careful and slows down at every step. He asked for a rag and a container. The filter was then tightened and water was poured into a larger container. The filter

was cleaned smoothly, but immediately after that there was a problem with screwing in the filter. The participant did not succeed and for a long time he tried to tighten the filter himself. Despite several additional questions and speeches from the assistant, the filter could not be screwed on. In the end, the participant failed to tighten the filter and needed the assistance of a moderator. Then the washing was finished just fine.

Emotions in S11.   This course of the script was full of emotion. The application was able to detect fear and surprise several times together. As a result, there have been several emotionally bound speeches, such as reassurance that the filter can be easily screwed back on quickly. In addition to the numerous reactions to the participant's uncertainty, there was also a reaction to happiness when the assistant praised the participant, that the hardest work is already done.

Number of unexpected misunderstanding speeches: 3

Scenario S12.   In this scenario, the participant asked if he should turn off the water system when he wanted to clean the filter. After the first speeches, he was already following the right procedure.

The participant found the hose without any problems and still demanded that the water should be closed. However, the assistant went gradually, but the participant did not mind. The supply hose was unscrewed and cleaned in quick succession. The cleaning was done very thoroughly. There was no problem screwing the hose back in, and the participant eventually just opened the water tap.

Emotions in S12.   In terms of emotions, it was a poorer scenario. Reactions to uncertainty have appeared. Luck was recognized the most this time. The assurance, that the most difficult part of the repair is over, gave the participant confidence and he was able to complete the scenario in quick succession.

Number of unexpected misunderstanding speeches: 1

Scenario S13.   The participant remembered one of the real situations and was probably upset. The participant did not open the washing machine door, did not unscrewed the drain filter or did not open the supply hose. Probably due to incomplete tightening of the supply hose in the previous scenario, a defect was also detected at this point. On the other hand, the defect in the drain hose was not detected for the first time, but the defect was detected immediately after reassurance. The technician was told that the drain hose was probably cracked or incorrectly fastened. The participant did not report a fault with the supply hose.

Emotions in S13.   At the beginning of the scenario, there was anger. As the participant stated in the discussion, he has already experienced such a situation and it probably upset him. In each of the individual inspections,

the participant was uncertain, so the reactions came in a uncertain path.

Number of unexpected misunderstanding speeches: 1

Reaction to unexpected misunderstanding. The reaction to the unexpected misunderstanding was natural, the participant had no problem with it. He was not surprised by the higher number of such errors.

Discussion. The participant feels rather satisfied, according to him the assistant had a pleasant reassuring voice. He considers it an advantage to be the absence of a physical technician who could look bad at certain tasks. This is how it can be done in the peace of home. Participant compared it to the manual - does not have to deal with the manual.

The participant did not notice the reaction to the emotions, and the important thing for him was that no one made a fool of him.

The participant considered random mistakes to be his mistakes because he was ahead. The overall usability is pleasant, no instructions, no reading, only one listening of manual. According to the participant, it is even more detailed than the manual.

■ Participant B6

Emotion	**S11**	**S12**	**S13**
Anger	0	0	1
Happiness	3	2	0
Sad	1	0	0
Fear	2	1	1
Surprise	1	3	3

Table 4.7: Emotions counters of participant B6

Scenario S11. The participant immediately thought that there could be a problem with the drain filter. However, his assistant also had him check the drain hose. During the course, the participant often went ahead and the water flowed out a little.

Further steps were also taken in advance. However, the filter failed to screw on the first try, so the participant asked for advice. After a while, the filter was finally tightened and then the drain program started.

Emotions in S11. It has been clear since the participant came to the room that he is a naturally happy person. It also affected the emotions. For the most part, luck prevailed.

The detailed path was not triggered, although uncertainty was detected.

The reason was detection at a point where no detailed path was needed.

Number of unexpected misunderstanding speeches: 1

Scenario S12.   In this scenario, the participant proceeded exactly as it should be. He once had a supplementary question on the supply hose tap location. He cleaned the sieve thoroughly, screwing the hose back on and opening the water went without any problems.

Emotions in S12.   As in the first scenario, mainly luck was detected. Fear once occurred when the user was not sure if he had the right water tap. The reaction to happiness was used once at the point where half of the work was done. The participant did not respond to this speech, he just smiled.

Number of unexpected misunderstanding speeches: 1

Scenario S13.   The participant checked the washing machine drum, which he did not open. This was followed by an inspection of the drain filter and the supply hose. As a result, a warning was applied to the supply hose because uncertainty was detected. The participant managed to find the defect immediately and described the course of the inspection on the phone.

Emotions in S13.   This time, the participant tried to make the situation better, and the application also recognized the anger at the beginning. Then the surprise was repeated, because the participant was not sure if he should close the water. Uncertain path was used.

Number of unexpected misunderstanding speeches: 2

Reaction to unexpected misunderstanding.   The participant always repeated what he wanted to say and tried to reformulate it a bit.

Discussion.   The participant felt fine, he was rather satisfied. The assistant was rather natural and mainly reacted just fine. The participant did not notice the reaction to the emotions, but he likes that the assistant stays calm and concentrates on the task. It is also positive that the assistant does not deepen bad emotions.

He didn't notice the unexpected misunderstanding at all. The overall usability is very beneficial.

■ **Participant B7**

Emotion	S11	S12	S13
Anger	0	0	1
Happiness	1	3	0
Sad	0	0	2
Fear	1	0	2
Surprise	2	0	3

Table 4.8: Emotions counters of participant B7

Scenario S11. The participant was not very communicative during the conversation. At first the participant looked for a drain hose, the assistant advised him. The participant did everything in advance and it was clear that he had cleaned the filter before. He also managed to unscrew the whole filter and the water leaked out. The return of the filter did not go without minor problems. However, it was not necessary to add a detailed path, as the participant managed to hit the groove correctly.

Emotions in S11. The participant was satisfied from the beginning and it felt like a game for him. Unfortunately, he was also not very communicative, so it was very difficult to record emotions. No emotional reactions were used.

Number of unexpected misunderstanding speeches: 1

Scenario S12. This scenario was one of the fastest of all. The participant responded minimally (used mostly "yes" or "no") and had all the steps done very quickly. He had no problem finding the right hose, it didn't take him long to close the water or unscrew the hose. The sieve cleaning could be done a little more thoroughly, but the important thing is that it was done. Screwing in the hose and turning on the water again was done in advance.

Emotions in S12. The application only detected happiness, as the participant was satisfied that he was doing great and quickly. No emotional reaction was used.

Number of unexpected misunderstanding speeches: 2

Scenario S13. There was uncertainty about the washing machine drum as to whether water was leaking. Unfortunately, the assistant did not manage to prevent the participant from opening the door. The filter was inspected without any problems and the subsequent inspection of the supply hose was carried out quickly and confidently. However, a check of the drain hose was performed in a similar way, when the first two attempts failed to detect an error. The third attempt was already successful and the participant successfully ended the inspection and a said all about inspection and possible damage to the phone.

Emotions in S13.   A speech was used on the washing machine drum, which responded to uncertainty (fear, surprise). A similar uncertain path was used for the supply hose. During the second inspection of the drain hose, fear and surprise were detected, so a uncertain path was started. As a result, the participant was able to complete the inspection.

Number of unexpected misunderstanding speeches: 1

Reaction to unexpected misunderstanding.   The participant always just repeated the previous sentence. He acted as if his assistant had just overheard him.

Discussion.   It was interesting at the beginning of the discussion that the participant stated:

"As economical as I am, I use the drain water to the flowers. But sometimes the drain hose pops out and I have a flooded bathroom".

The participant feels rather satisfied after the interview. He considers the assistant to be beneficial for people who have no experience with similar repairs.

He didn't notice the reaction to the emotions, but the cameras seem very invasive. He realized the unexpected misunderstandings, but did not respond to them.

■   **Participant B8**

Emotion	S11	S12	S13
Anger	4	1	4
Happiness	1	5	2
Sad	0	0	1
Fear	2	2	4
Surprise	2	1	4

Table 4.9: Emotions counters of participant B8

Scenario S11.   The participant followed exactly the instructions of the assistant. We could see that he didn't want to do anything alone.

At the drain filter, the participant began to respond uncertainly. He divided the sentences a lot and sometimes the assistant jumped in. The participant therefore asked the assistant to slow down the process.

The water was drained from the washing machine into a small container and poured into a large one. The filter was cleaned properly and returning

the filter was again with difficulty. The participant had a problem guiding the filter into the groove. In the end, everything worked out and the participant finished the laundry.

Emotions in S11.   This participant was very rich in emotions. Apart from his sadness, the application detected all other emotions. The biggest representation had anger, mainly because the assistant was too fast. As is customary, the uncertain path was used as a response to fear and surprise couple times. Also dangerous path was used as a response to anger, but did not have any effect.

Number of unexpected misunderstanding speeches: 2

Scenario S12.   The participant chose the path by moving the washing machine to get to the drain hose. Step by step, he went through unscrewing and cleaning the sieve to return the supply hose. He had no problem screwing in the hose and opening the water.

Emotions in S12.   The participant was satisfied and smiling from the beginning of the scenario. This was also confirmed to us by the application that detected happiness. Based on this emotion, a later part of the scenario triggered the speech that the repair was already halfway through. It just encouraged the participant to smile even more. Although other emotions were detected as well, there was no opportunity to respond to them.

Number of unexpected misunderstanding speeches: 1

Scenario S13.   The individual inspections were carried out slowly by the participant. When checking the drain hose, which was probably jammed (test error), the participant could not find out that there was a fault. In a subsequent phone call, he only described that his bathroom was flooded and that he did not know where the fault was.

Emotions in S13.   The assistant responded mainly to uncertainty (fear and surprise). He chose a uncertain path for each of the inspections. Unfortunately, even a uncertain path in the case of the drain hose was not enough and the defect was not found.

Anger was also detected during the conversation, which reappeared in response to the assistant's speed.

Number of unexpected misunderstanding speeches: 1

Reaction to unexpected misunderstanding.   The participant tried to reformulate his original statement a little differently so that the assistant would understand him better.

69

Discussion.   The participant was rather dissatisfied and stressed, as he does not like any repairs. However, he thinks the assistant can help in these cases.

The participant did not notice any reaction to his emotions, but the cameras around him stressed him. However, he himself admitted that he thought that sensing emotions could make a good contribution to the safety of such processes.

It seemed to the participant that the assistant was moving too fast. He noticed unexpected misunderstandings, but it seemed normal to him. According to him, it happens in the real world all the time.

▮ Participant B9

Emotion	S11	S12	S13
Anger	1	1	5
Happiness	3	4	0
Sad	1	0	2
Fear	0	1	1
Surprise	5	3	5

Table 4.10: Emotions counters of participant B9

Scenario S11.   The participant entered the test with great surprise. He admired the washing machine all the time and the fact that it is perfectly created. The scenario went through without much difficulty, he was afraid to take individual steps, because he was afraid to damage the washing machine.

He unscrewed the filter completely and failed to return it. That's why some water spilled on the rag. As the participant tried to bring the filter back in a hurry, he succeeded and after cleaning it was no longer difficult for him.

Emotions in S11.   Although the application detected many emotions to us, a minimum of them were related to the scenario. The surprise appeared alone or in combination with happiness when admiring the washing machine. Only one message in response to emotions was used in the scenario. When it was necessary to drain the water, the participant was offered to call for help (anger detected). It was rejected.

Number of unexpected misunderstanding speeches: 1

Scenario S12.   This scenario went without problems. The participant asked for help, prepared things for cleaning, turned off the water, and unscrewed the hose. The cleaning of the sieve was thorough and the subsequent screwing in was very straightforward. The script was done very quickly.

Emotions in S12.   The emotions that were detected in this scenario were still influenced by participant's enthusiasm for the washing machine model. The assistant only responded once when he told the participant that they were halfway there. The participant only replied that it was good.

Number of unexpected misunderstanding speeches: 0

Scenario S13.   The participant went through this scenario very confidently and quickly. It was clear that he put himself into the scenario perfectly and was in a hurry to find the error. He found the cause of the problem correctly and described the specialist detected fault on the phone.

Emotions in S13.   The application detected a lot of anger. As a result, the assistant was able to end the conversation quickly and call someone without much explanation. As a result, the user's frustration could not be deepen further.

Number of unexpected misunderstanding speeches: 0

Reaction to unexpected misunderstanding.   An unexpected misunderstanding occurred in only one case at the beginning of a conversation. The participant only repeated what he said previously.

Discussion.   The participant feels good and very satisfied. He was fascinated by the model of the washing machine and enjoyed the scenarios. He evaluates communication with the assistant as neutral, because he found it very talkative.

The participant stated that he thought the assistant responded to his emotions by giving a detailed path. Which, of course, did not happen once in this case. The participant also noticed an unexpected misunderstanding.

According to the participant, the overall usability is great. It would be even better if the assistant was able to call the service himself.

◼ Participant B10

Emotion	S11	S12	S13
Anger	5	2	4
Happiness	2	6	2
Sad	8	2	2
Fear	6	0	2
Surprise	2	0	3

Table 4.11: Emotions counters of participant B10

71

Scenario S11. This conversation was probably the most natural. The participant answered clearly and with the whole sentences. He was not afraid to ask inquisitive questions. He did not overtake the assistant and always waited for his reaction.

There were some problems during the scenario. The filter did not return to the washing machine and some water spilled on the prepared rag when the water was drained. All the dirt were removed and the filter was screwed on after a long time - troubles with groove.

Emotions in S11. The participant was afraid of what would happen. Therefore, the application detected fear mainly from the beginning of the scenario. Anger also began to appear due to the unsuccessful return of the filter to the groove.

The assistant reacted mainly to the uncertainty when opening the filter. Unfortunately, it did not prevent a small water leak.

Number of unexpected misunderstanding speeches: 2

Scenario S12. As usual, the script went smoothly. The participant followed exactly the assistant. The assistant did not understand the participant a few times, but participant always reformulated his sentences so that the assistant could understand. The assistant also advised the participant which hose to use.

Emotions in S12. The participant was mainly happy during the scenario. Fulfilling every next step motivated him more and more. Based on the detection of happiness, a speech was used, which announces that we are already behind the harder half of the work.

The application also detected anger. It always appeared as a reaction to unexpected misunderstanding. However, the level of anger was not high.

Number of unexpected misunderstanding speeches: 2

Scenario S13. During the beginning of the test, a comical situation arose in which the participant did not understand the assistant and the assistant did not understand the participant. The assistant finally repeated the last speech and the script continued.

The first three inspections (drum, filter and supply hose) went well, the problem occurred with the drain hose. The participant did not correctly detect the defect. He described on the phone that he had a flooded bathroom, but did not know the cause.

Emotions in S13. The comic situation slightly irritated the participant, so our application detected anger. Anger also arose when the drain hose

was inspected as the assistant insisted on re-checking. Furthermore, the participant also had uncertainty, so a uncertain path was used for the washing machine drum and the washing machine filter.

Number of unexpected misunderstanding speeches: 2

Reaction to unexpected misunderstanding. The participant always formulated his speech a little differently so that his assistant could understand him better. It made him a little angry.

Discussion. The participant feels fine and rather satisfied. He thinks that communication with an assistant cannot be completely natural.

The participant did not notice the reaction to the emotions, but he liked it when the assistant praised him. According to the participant, the assistant behaved correctly in case of misunderstanding. He expects the assistant not to know all the intents, so sometimes he doesn't understand.

The assistant is useful for a person who has never encountered repairs and for more experienced people it can be some guide.

■ Participant B11

Emotion	S11	S12	S13
Anger	5	3	3
Happiness	0	1	2
Sad	2	1	1
Fear	3	4	3
Surprise	0	3	1

Table 4.12: Emotions counters of participant B11

Scenario S11. The participant was not sure if he should talk to the assistant at the beginning. Then the participant asked if he should turn off the washing machine.

The drain hose was checked correctly and the participant prepared a smaller container. Then he did not manage to pour water into bigger container in time and some water spilled on the rag placed under the washing machine.

The return of the filter to the washing machine also failed because the participant had a problem with keeping the filter in the place. He managed to direct it there after several attempts.

Emotions in S11. The participant, although he was willing to help, wanted to get over it. The emotions of anger, sadness, and fear were most detected.

73

Fear was detected especially when the participant was guiding the filter into the groove. Based on these emotions, the participant was guided to the right filter door.

Number of unexpected misunderstanding speeches: 3

Scenario S12.   The participant asked for help with cleaning the filter. The assistant helped find the right hose and the participant then closed the water. The hose was unscrewed and cleaned. It then took the participant a moment to screw the hose back on, as he was a little weaker because of his age.

Emotions in S12.   The emotions from the previous scenario are similar, now we have happiness and surprises as well. This led to the launch of a uncertain path and also praise for the good work.

Number of unexpected misunderstanding speeches: 0

Scenario S13.   When checking the drum of the washing machine, it was necessary to remind participant that the he should not open the drum of the washing machine. Unfortunately, that happened anyway. Other checks have already passed very certainly and quickly. He discovered the defect at first try and when calling a specialist he described only the most probable cause.

Emotions in S13.   As already mentioned, a uncertain path has been launched on the basis of those emotions. In addition to fear, sadness and surprise, there was also joy and anger. The anger was caused by an unexpected misunderstanding.

Number of unexpected misunderstanding speeches: 1

Reaction to unexpected misunderstanding.   The participant always repeated his last speech. Unfortunately, this unexpected misunderstanding made the participant angry.

Discussion.   The participant felt good and satisfied after the test, but was nervous before the test. The communication seemed quite natural to him.

The participant did not notice the reaction to the emotions, but the cameras made him nervous. He wouldn't want anyone to watch him.

The participant did not like the unexpected misunderstanding very much, but he evaluates the overall usability of assistant very positively. He would welcome such a washing machine himself at home.

■ **Participant B12**

Emotion	S11	S12	S13
Anger	3	4	4
Happiness	3	4	1
Sad	1	1	3
Fear	1	2	1
Surprise	0	2	3

Table 4.13: Emotions counters of participant B12

Scenario S11.   The participant was very communicative and all the answers sounded very natural. Unfortunately, he did not wait for the assistant and unscrewed the filter straight away. The water leaked out but he did not mind it at all. When he tried to return the filter, he failed. The assistant advised him how to do it and then he succeeded.

Emotions in S11.   The participant stepped out of his comfort zone so he can participate. Because of this, anger appeared several times at the beginning of the test. In addition, the spilled water also encouraged it a bit. However, the found defect changed the participant's mood and our application began to detect happiness.

No emotional reactions were used.

Number of unexpected misunderstanding speeches: 0

Scenario S12.   At the beginning of the scenario, an unexpected misunderstanding was used twice in a row when preparing the props. Subsequent supplementary questions about the correctness of the hose and the closed tap did not improve the participant's mood much. The hose was unscrewed and cleaned quickly. The participant completed the scenario without difficulty.

Emotions in S12.   The participant came into the script in a good mood, so praise was used for wanting to clean the filter. Subsequently, however, an accidental mistake caused frustration in the participant. The joy of the unscrewed hose, which was tied to the speech that the hardest part was done, was overcome by this.

Number of unexpected misunderstanding speeches: 3

Scenario S13.   The participant checked all possibilities. At the washing machine drum, the participant was not entirely sure if it was not flowing. He didn't know how to find out. The participant found out the fault correctly and then quickly called the service. In the description to the specialist, participant only mentioned that the problem is somewhere in the drain hose.

Emotions in S13. Due to the participant's good understanding of the scenario, anger was detected.

Uncertainty was again detected during the drum inspection, so the speech that the drum must not be opened was used.

Number of unexpected misunderstanding speeches: 1

Reaction to unexpected misunderstanding. In the first scenario, an unexpected misunderstanding occurred immediately after other, so it upset the participant. Nevertheless, he always reformulated his requests.

Discussion. The participant was rather satisfied and said it was fun. Communication with the assistant was rather natural, but person would understand better.

He didn't notice the reaction to the emotions. He does not need an assistant with emotions, it is important for him that the assistant could help him.

Unexpected misunderstanding is completely natural and the voice of the user must be learned.

The usability is great and the participant assumes that it would be implemented in all devices.

■ 4.4.3 Post-test Evaluation

After the end of each test, we started a discussion with the participants, where we tried to get information about the assistant and participant's point of view. We asked about satisfaction with the assistant, how natural the communication was and the usefulness of the assistant. We also wanted to know if they noticed the reactions to their emotions and their view of emotion sensing.

The table 4.14 shows the answers to each question using numbers. Number one represents the best value, and the higher the number, the more negative the answer.

Participant	Satisfaction	Natural	Emotions	Usefulness
B1	2	2	No	1
B2	2	3	No	1
B3	2	1	No	1
B4	4	2	No	1
B5	2	1	No	1
B6	2	2	No	1
B7	2	2	No	2
B8	4	3	No	2
B9	1	3	Maybe	1
B10	2	2	Maybe	2
B11	2	2	No	1
B12	2	2	No	1

Table 4.14: Post-test evaluation

As can be seen, most participants rate the assistant positively. There is only a minimal number of more negative responses (participant B4 and B8) and these were mostly caused by the induced stress. Two participants (B9, B10) stated that they may have noticed a response to emotions, but only one (B9) noticed it correctly.

Opinions on sensing emotions varied greatly. In general, however, the cameras at the washing machine caused stress and discomfort to the participants. One of the participants stated that he does not need an assistant who monitors emotions, but an assistant who will help with the work (B12). However, one participant (B8) also said that capturing emotions can be beneficial and can prevent bad situations.

In the end, we were also interested in the opinion of an unexpected misunderstanding. Although the unexpected misunderstanding had a negative effect on some people (B8, B12), almost everyone (B3, B4, B5, B8, B10, B12) said it would come naturally to them. In the real world, we don't understand everything as well.

4.4.4 Comparison of the time aspect

In this section, we focus on the times of individual participants spent on different scenarios.

Participant	S11	S12	S13
B1	527	255	141
B2	494	159	285
B3	404	168	242
B4	582	368	155
B5	943	285	242
B6	382	163	178
B7	262	119	149
B8	349	139	213
B9	522	172	120
B10	487	253	161
B11	393	182	123
B12	211	209	159

Table 4.15: Times in seconds of participants in scenarios

In the table 4.15 we see the times of the individual participants. The first four participants represent the category of young people under 65 years old. We can compare the results with older people. The remaining lines represent the category of seniors over 65 years old. For a better idea, we projected the times into a graph 4.21.
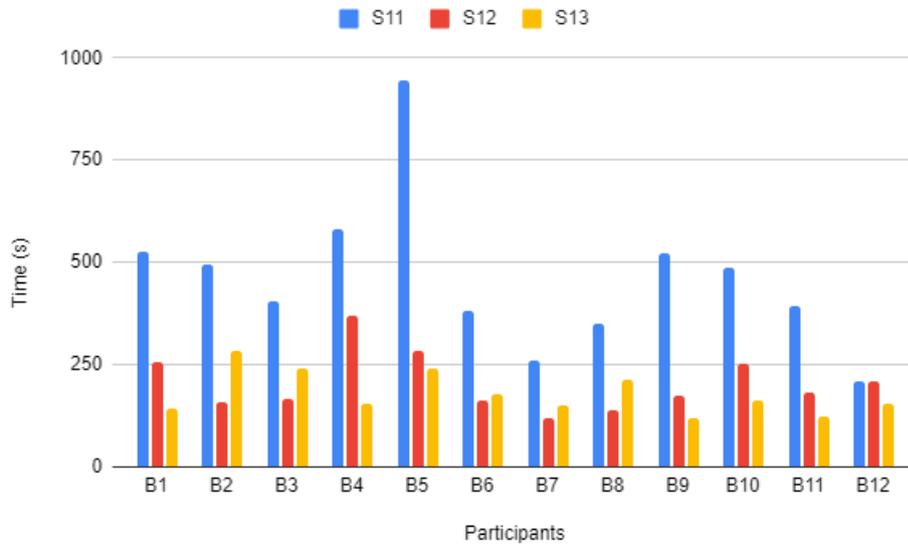
Figure 4.21: Times in seconds of participants in scenarios

From the graph 4.21 we can recognize at first glance one outlier for the S11 scenario. The B5 participant, who spoke very slowly, so the scenario is very long. On the contrary, the B12 participant went through the scenario very quickly. The difference between the fastest and slowest participant for the S11 scenario is 732 seconds.

We also compared the individual scenarios with the previous experiment
[30]. In this case, we compare only participants from the category of seniors.
We compare the mean ($\overline{x}$) and standard deviation (σ).

■ S01 vs. S11 - Error E20

In this section, we compare scenario S01, which was created in 2.4.1, and
scenario S11, which contains emotion paths.

$\overline{x}$ **S01**	$\overline{x}$ **S11**	σ **S01**	σ **S11**
479,3	443,6	66,0	226,7

Table 4.16: Error E20 time comparison

In the table 4.16 we see that the average time of scenario S11 is exactly
35,7s better than in scenario S01. We attribute this fact to the early detection
of uncertainty and the rapid delivery of a detailed path. The scenario is thus
accelerated at several points, which can be seen in the overall average.

If we take the standard deviation, it is significantly higher for the S11 scenario.
The reason is the presence of the already mentioned outlier.

■ S02 vs. S12 - Cleaning of the filter

In this section, we compare scenario S02, which was created in 2.4.1, and
scenario S12, which contains emotion paths.

$\overline{x}$ **S02**	$\overline{x}$ **S12**	σ **S02**	σ **S12**
204,7	190,3	24,4	56,2

Table 4.17: Cleaning the filter time comparison

This time, the average transit time of the scenario is only 14.4 seconds
better. There is the least reaction to emotions in the scenario. The biggest
benefit here was the compliment for unscrewing the hose. The participant
was probably more motivated to complete the task.

The standard deviation is not as drastic as in the previous scenario, but
is still lower for the emotionless scenario.

■ S03 vs. S13 - Flood

In this section, we compare scenario S03, which was created in 2.4.1, and
scenario S13, which contains emotion paths.

$\overline{x}$ **S03**	$\overline{x}$ **S13**	σ **S03**	σ **S13**
206,9	167,8	38,4	42,2

Table 4.18: Flood time comparison

There is quite a big difference in the scenario for a flooded bathroom. The average of scenario S13 is 39.1 seconds faster than the average of scenario S03. This is the highest difference achieved so far, which we attribute to the addition of detailed paths linked to emotions during individual inspections.

The standard deviation is very similar in this case.

■ Comparison of age groups

Group	$\overline{x}$ S11	$\overline{x}$ S12	$\overline{x}$ S13	σ S11	σ S12	σ S13
> 65 years	443,6	190,3	167,8	226,7	56,2	42,2
< 65 years	501,8	237,5	205,75	74,6	97,2	69,2

Table 4.19: Flood time comparison

As we can see, the younger participants progressed more slowly. This is mainly due to the fact that several seniors (such as B5 and B6) have sometimes done some scenarios in real life, while young people no longer do these activities with maintenance and repair of washing machines themselves and call technician for help.

■ 4.4.5 Findings

■ Uncertain path

This path has worked very well. Together, it has been used the most across scenarios. In the first scenario 6 times, in the second 3 times, and in the last 10 times.

If we combine the use of the path with the time aspect of the scenarios, then we see that it was in the first and third scenarios (where the uncertain path was used the most times) that our time improved. Although improving time is not absolutely essential

The main benefit is mainly attributed to the fact that the assistant, thanks to the uncertain path, was able to prevent poorly performed or not performed tasks at all, to which the participant would have to return anyways in a course of scenario.

■ Happy path

This path, although it had no effect on time, was able to lift people's spirits during the scenario and thus encourage further action. This path was used by the assistant primarily in the second scenario. She appeared in 10 out of 12 participants.

After each use of the happy path, the application began to detect more emotions of happiness. It was clear from the participants themselves that the speech made them happy.

◼ Dangerous path

Reaction to anger is hard thing to do, which is also confirmed by the low numbers of use of this path. In the first scenario, this path was used by the assistant 3 times, in the second and third only once.

The reaction was often inappropriate, so it was not used as often by the assistant. At the same time, each user has different reactions to the prepared speeches. Some may be upset by the speech, and others may be reassured. However, it cannot be taken globally for all participants.

◼ Other paths

Other paths and speeches, such as appeasement path, decision point, or global emotion speeches, have been used very little. For the appeasement path and decision point, this is due to the low detection of the necessary emotions at a given point in the scenario.

Global emotion speeches were used by only two participants, but no response came from them. This can be attributed to the fact that each scenario is different and it is difficult to respond in general to the different emotions in a given scenario. It would be always better to have a specific speech ready for each scenario.

◼ Emotion detection

Emotion detection worked great and confirmed that users experience different emotions in each scenario. In addition, each user responds differently at different points in the scenario. Therefore, it is important for the assistant to capture these emotions and be able to respond to them.

Emotion	S11	S12	S13	SUM
Anger	32	20	36	88
Happiness	23	34	12	69
Sad	17	5	16	38
Fear	28	17	23	68
Surprise	31	20	36	87

Table 4.20: Global emotion counters

We can see from the table 4.20 that all emotions were represented in all scenarios. In the first scenario, anger, surprise, and fear prevailed. In the second scenario, happiness was the most and sadness was minimal. In the

third scenario, anger, and surprise were combined the most.

We also see that the longer the scenario, the more emotions were detected (the emotions were detected each time when the participant answered). In total, across the scenarios, the application detected the most anger and surprise.

■ Cameras

Most participants said that the cameras are very invasive. The participants generally had a problem with the fact that we were still watching them and that made them very stressed.

We were already afraid of this fact before the experiment but we were willing to sacrifice it for the sake of the experiment.

■ Unexpected misunderstanding

With this approach, in the prototype, we try to get closer to fully automatic virtual assistants and also to real life, where we cannot always understand the other. This benefit will bring us a more natural form of experimentation.

In the experiment, the unexpected misunderstanding fit into the scenarios quite naturally. Although each participant reacted differently, most of them stated in discussion that this is a completely natural thing that also occurs in the real world. In the event of an unexpected misunderstanding, anger often appeared within the emotions. Although most people say it's natural for them, this anger is completely automatic. It naturally frustrates a person when they have to repeat something several times. In addition, they raise their voices, which is generally a sign of anger.

In the test, we used a maximum of 3 unexpected misunderstandings with a probability of 8%. These numbers seem to be ideal, as the random error usually spread to different places and it is not very often that two occur immediately after each other. When setting this level before the experiment itself, we found that if the probability is higher, the errors occur mainly at the beginning of the conversation and not at all after exceeding the maximum value.

■ 4.4.6 Answers to research questions

■ Q1 - What are the most common emotions in washing machine repair/maintenance scenarios?

As we can see from the table 4.20, the most common emotions are **anger** and **surprise**. Anger was mainly caused by the nature of the scenarios, as for example a flooded bathroom or a non-functioning washing machine are not pleasant.

Surprise appeared mainly because a person does new tasks (mostly), and thus the whole process is surprising for him. However, this emotion usually appeared in connection with fear or sadness. An uncertain path was linked to these combinations.

Q2 - What effect do emotions have on the course of the experiment?

The effect on the course of the scenario depends on what emotion it is.

Anger. It has a negative effect on the course of the scenario. We have seen that the participants who showed this emotion acted recklessly and too quickly. It is hard to respond to anger and the assistant had trouble working with this emotion properly.

Happiness. Happiness has a very positive effect on passing the scenario. Participants in whom luck was detected performed most operations without problems. Although it has no time effect, the assistant is more user-friendly when reacting to happiness.

Sadness, fear, and surprise. These emotions have a big influence on the course of the scenario. Together, this is uncertainty. In this case, participants are not sure that they have done something right or what to do. This can slow down the scenarios.

Q3 - What is the option of users on sensing emotions?

Participants didn't like being watched. In the final discussion, several participants complained about the presence of cameras. It is very stressful and invasive for them. Participants do not see much importance in sensing emotions. Detecting emotions from the user's voice would be less invasive.

4.4.7 Suggestions

Voice emotion recognition

Most participants said that the cameras were invasive, stressful, or uncomfortable. In the future, it would be good to do research where emotions are captured from the voice. Although it will be difficult to acquire emotions in real time.

High-quality sensing of emotions from the person can bring a breakthrough in voice assistants, as even in this experiment we found that emotions have a great influence on conversation. If there was an assistant who could adapt to them properly, then unpleasant situations could be avoided and, for example, scenarios can be carried out better and more accurately.

■ Virtual voice assistant

In the first part of this thesis, we tried to create a virtual voice assistant using available tools. Although it has failed, it remains one of the future goals of the future crisis-solving voice assistants.

A fully virtual form combined with a good detection of emotions and reactions to them can be a great benefit for ordinary users.

■ Use assistant in a different devices or equipment

Participants B7 and B12 stated that they assume that the assistant we are currently developing will not be available only for washing machines. They said they would like it in other facilities in the house. It could be a portable box that has this feature.

Another goal for future development may be to extend the scope to other devices and equipment. There is, for example, a dryer, a dishwasher, or even a car. Everyone must have experienced that some of these things were broken.

Chapter 5

Conclusion

This work was motivated by the possibility of using emotions as a tool for the voice assistant. The purpose of this tool was to help improve user-friendliness and bring the assistant closer to human behavior. In the research part of the work, we dealt with the possibilities of capturing emotions, existing voice assistants, and outlining the issue of emotions.

The first goal of this work was to create a virtual voice assistant. This part of the work dealt with the implementation of a virtual voice assistant and a subsequent experiment to test its functionality. We managed to implement such a virtual voice assistant that worked completely automatically. The pilot testing proved that the currently available technology was not well prepared for real-life usage and therefore user testing failed. Although we could see a potential in using a virtual assistant, its stability is very important. It was necessary to choose another way to test the emotions of the voice assistant.

The main goal of this work was to create an emotional module into an existing prototype. In this part, the module was designed and implemented together with an application for real-time emotion detection. From a subsequent experiment, we found that the uncertain path became a great benefit, thanks to which the participants were able to complete the tasks more safely. An interesting finding is also the use of happy path, which managed to tune the user to a better continuation of the scenario. Although the emotion tracking method seemed to be beneficial, the use of the camera method was not well received by the participants and may be replaced in the future by capturing emotions from the voice.

Bibliography

[1] *Affectiva SDK: A cross-platform realtime Multi-Face Expression Recognition Toolkit.* Mar. 2017. URL: https://www.affectiva.com/science-resource/affdex-sdk-a-cross-platform-realtime-multi-face-expression-recognition-toolkit.

[2] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network". In: *2017 international conference on engineering and technology (ICET)*. Ieee. 2017, pp. 1–6.

[3] Tawfiq Ammari et al. "Music, Search, and IoT: How People (Really) Use Voice Assistants." In: *ACM Trans. Comput. Hum. Interact.* 26.3 (2019), pp. 17–1.

[4] Jerome R Bellegarda. "Spoken language understanding for natural interaction: The siri experience". In: *Natural interaction with robots, knowbots and smartphones* (2014), pp. 3–14.

[5] Dario Bertero et al. "Real-time speech emotion and sentiment recognition for interactive dialogue systems". In: *Proceedings of the 2016 conference on empirical methods in natural language processing*. 2016, pp. 1042–1047.

[6] Chris M Bishop. "Neural networks and their applications". In: *Review of scientific instruments* 65.6 (1994), pp. 1803–1832.

[7] TechTarget Contributor. *What is google assistant? - definition from whatis.com.* Aug. 2017. URL: https://www.techtarget.com/whatis/definition/Google-Assistant.

[8] Cosanlab. *Affectiva Emotion detector.* Dec. 2017. URL: https://github.com/cosanlab/affectiva-api-app/blob/master/affectiva_emotion_detector.html.

[9] *Czech team of AI developers for Amazon Echo.* URL: http://alquistai.com/.

[10] Poorna Banerjee Dasgupta. "Detection and analysis of human emotions through voice and speech pattern processing". In: *arXiv preprint arXiv:1710.10198* (2017).

[11] *Emotion analysis.* URL: https://www.noldus.com/applications/emotion-analysis.

[12]   *Emotion definition and meaning: Collins english dictionary.* URL: https://www.collinsdictionary.com/dictionary/english/emotion.

[13]   Paul Green and Lisa Wei-Haas. "The rapid development of user interfaces: Experience with the Wizard of Oz method". In: *Proceedings of the Human Factors Society Annual Meeting.* Vol. 29. 5. SAGE Publications Sage CA: Los Angeles, CA. 1985, pp. 470–474.

[14]   *IRobot®: Robot Vacuum and Mop.* URL: https://www.irobot.com/.

[15]   Neha Jain et al. "Hybrid deep neural networks for face emotion recognition". In: *Pattern Recognition Letters* 115 (2018), pp. 101–106.

[16]   Soheil Kianzad et al. "Accountability-Aware Design of Voice User Interfaces for Home Appliances". In: *Graphics Interface 2021.* 2020.

[17]   Pavani Kollamudi et al. "Smart virtual assistant". In: *Materials Today: Proceedings* (2021).

[18]   Kiljae Lee, Kyung Young Lee, and Lorn Sheehan. "Hey Alexa! A magic spell of social glue?: Sharing a smart voice assistant speaker and its impact on users' perception of group harmony". In: *Information Systems Frontiers* 22.3 (2020), pp. 563–583.

[19]   Gustavo López, Luis Quesada, and Luis A Guerrero. "Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces". In: *International conference on applied human factors and ergonomics.* Springer. 2017, pp. 241–250.

[20]   Ingrid Lunden. *Amazon gets into voice recognition, buys Ivona software to compete against Apple's Siri.* Jan. 2013. URL: https://techcrunch.com/2013/01/24/amazon-gets-into-voice-recognition-buys-ivona-software-to-compete-against-apples-siri/.

[21]   Yong Ma, Heiko Drewes, and Andreas Butz. "How Should Voice Assistants Deal With Users' Emotions?" In: *arXiv preprint arXiv:2204.02212* (2022).

[22]   Martin Magdin and F Prikler. "Real time facial expression recognition using webcam and SDK affectiva". In: (2018).

[23]   Lisandra Maioli. *UX for conversion: Understand your user.* Mar. 2019. URL: https://blog.fluidui.com/ux-for-conversion-understand-your-user/.

[24]   J Medeiros. "Virtual Assistants Can Detect Your Bad Mood and Do Something About It". In: *Voice Summit Blog* 7 (2018).

[25]   *Microsoft.* URL: https://support.microsoft.com/en-us/topic/what-is-cortana-953e648d-5668-e017-1341-7f26f7d0f825.

[26]   Jeanne Nakamura and Mihaly Csikszentmihalyi. "The concept of flow". In: *Flow and the foundations of positive psychology.* Springer, 2014, pp. 239–263.

[27] Jakob Nielsen and Thomas K Landauer. "A mathematical model of the finding of usability problems". In: *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. 1993, pp. 206–213.

[28] FATHIMA ANJILA PK. "What is Artificial Intelligence?" In: *"Success is no accident. It is hard work, perseverance, learning, studying, sacrifice and most of all, love of what you are doing or learning to do".* (1984), p. 65.

[29] Promethist.ai. *Conversational AI platform.* URL: `https://flowstorm.ai/`.

[30] Talácko Rudolf. "Konverzační asistent pro chytrou domácnost". MA thesis. České vysoké učení technické v Praze. Vypočetní a informační centrum., 2020.

[31] Stephen Samuel and Stefan Bocutiu. *Programming kotlin.* Packt Publishing Ltd, 2017.

[32] *Sentiment.* URL: `https://dictionary.cambridge.org/dictionary/english/sentiment`.

[33] *Smart pet care.* Mar. 2022. URL: `https://www.teslasmart.com/en/products/pet-care/`.

[34] *What is a voice assistant?* URL: `https://www.atinternet.com/en/glossary/voice-assistant/`.

[35] Xianda Zhou and William Yang Wang. "Mojitalk: Generating emotional responses at scale". In: *arXiv preprint arXiv:1711.04090* (2017).

Appendix A

List of abbreviations

- XML - Extensible Markup Language

- CNN - Convolutional Neural Network

- NN - Neural networks

- AI - Artificial Intelligence

- HTML - Hypertext Markup Language

- CSS - Cascading Style Sheets

- SDK - Software development kit

- ASR - Automatic speech recognition

Appendix B

Virtual voice assistant - Pre-Test Form

Testing of an virtual voice asistant

Hi,

I would like to ask you to fill out this form in order to test the virtual voice assistant.
By filling out the form, you agree to the processing of your personal data for the purposes
of the diploma thesis. The personal data will not be published anywhere.

Thank you

Marek Klement

1. Your name

    _____

2. Level of english

    Označte jen jednu elipsu.

    ⬭ B1 or worse

    ⬭ B2

    ⬭ C1 or better

3. What is your gender?

    Označte jen jednu elipsu.

    ⬭ Male

    ⬭ Female

    ⬭ Other

4. What is your age?

    _____

5.   Would you say that you are a technical type?

Označte jen jednu elipsu.

( ) surely yes

( ) probably yes

( ) not sure

( ) probably not

( ) surely no

6.   How would you rate your manual skill level?

Označte jen jednu elipsu.

( ) I leave all repairs to someone more experienced.

( ) I manage minor repairs myself.

( ) I'm not afraid to do more demanding things.

( ) I'm a handyman.

7.   Do you have any experience with repairing the washing machine?

Označte jen jednu elipsu.

( ) No

( ) Yes

8.   In case of an answer yes on previous question, what experience do you have?

_____

_____

_____

_____

_____

9.   Do you think you are communicative?

   Označte jen jednu elipsu.

   ◯ I love to talk to other people and I am not afraid to ask for help.

   ◯ Mostly yes, but sometimes it can be hard.

   ◯ Mostly no, but I am able to if I need help.

   ◯ Not at all.

10.   Do you have any experience with virtual conversation assistants?

   Označte jen jednu elipsu.

   ◯ None.

   ◯ I tried but not using anymore.

   ◯ I am an occasional user.

   ◯ I am and active user.

11.   Are you interested in participating in a practical experiment (Karlovo náměstí -
        CTU - building E)?

   Označte jen jednu elipsu.

   ◯ Yes

   ◯ No

12.   Email

   _____

13.   Telephone number

   _____

Obsah není vytvořen ani schválen Googlem.

Google Formuláře

Appendix C

Virtual voice assistant - English session guide

Session guide

Introduction

Hello, let me welcome you to the testing of the virtual voice assistant usage. My name is Marek Klement and I am a student of master's degree in Human-computer interaction (HCI) at the Faculty of Electrical Engineering, Czech Technical University in Prague. As part of my diploma thesis, I am creating a virtual voice assistant, which we will be testing today.

Informed consent

Before we begin, I must ask you to fill [informed consent](#) to participate in this study.

Please read this document carefully here. If you do not understand something or you are curious about anything, please do not hesitate to ask. If you do not have any further questions, I will ask you to fill in your name, today's date and signature.

Experiment introduction

Now let me briefly present the experiment we will be dealing with today. We are currently developing a virtual voice assistant for the smart home, specializing in crisis management and troubleshooting. Virtual Voice Assistant is a software program that works on the automated principle of communication of a participant with a virtual assistant. Although the assistant is automatic, it is not possible to support the full range of possible answers. Because of this, you may occasionally come across phrases like "I don't know what that means?" or "Can you please repeat that?". Which does not mean that you have to limit yourself in your questions in any way. Those questions can help us improve the prototype and prepare it for more situations.
In total three test scenarios await you in the experiment itself. In our case, it will be a simulation and solution of various faults, which you will try to solve with the help of a virtual voice assistant.

In order for a smoother experiment, I have prepared a small wooden model of the washing machine for testing purposes, on which we will carry out repairs. You can now view or touch the washing machine, but please do not maintain anything yet.

I would also like to introduce you to small props that you can use during the experiment. They will be placed behind you in a prepared locker and you can pick them at any time. There are two containers, a rag, a toothbrush and foam cubes.

Finally, I would like to describe the form of communication with the assistant. You can think of communication as if you need advice on a workflow and ask someone for help over the phone. The virtual voice assistant does not see what you are doing, so for it to work properly, you should not be afraid to talk more and calmly think out loud or say what you are doing.

The goal of the experiment

The aim of the experiment is to improve the virtual voice assistant and compare the test results with a prototype of the non-automatic conversation assistant. If something goes wrong, then don't worry, it's not your fault, it's the fault of the assistant itself.

Equipment to use

- washing machine
- laptop with the assistent
- two containers
- foam cubes
- old toothbrush
- contact list
- manual for the washing machine

Scenarios

We currently have 4 scenarios for repairing the washing machine. Each scenario solves a slightly different problem, which sets when the washing machine breaks or when a maintenance action is required. The links lead to a site where the scenario can be tried.

- [S00](#) - Introduction
- [S01](#) - Error E20 and cleaning the filter
- [S02](#) - Cleaning the filter of natural hose
- [S03](#) - Too much water error

Introduction of scenarios

Introduction

The first of all scenarios is a short introduction. This is mainly a test of communication with the assistant, getting used to his voice and adjusting the required volume. In this scenario, you will also experience the basic elements of the conversation. Try everything you can. This scenario also contains following possibilities, so the participant can try them before harder scenarios:
- Stopping the conversation at any point

- Moving to another step
  - Going back in the conversation

There will be no output from this scenario.

Error E20

This scenario is the first non-simple. Imagine that it is Friday afternoon and you want to go on a family vacation on weekends. You decide that you want to start the washing machine with a delayed start before the vacation so the laundry is ready to be hung in the morning. But when you come to the washing machine first thing in the morning, the error E20 is visible on the display. You should figure out what needs to be done with the help of a conversation assistant.

Cleaning the supply hose filter

You are drawn into a situation where, after one of the washing machine failures, you read that these situations can be prevented by regular cleaning of the supply hose filter. In this situation,you decided to ask for help with cleaning the voice assistant. Together you go through the whole process.

Water error

In this case, you witnessed a very unpleasant situation when the bathroom was flooded. Your task, together with the voice assistant, is to find out the cause of the flooding of the bathroom.

Post-test interview

Finally, I would like to ask you a few questions that relate to the evaluation of individual aspects of the virtual voice assistant.

1. How do you feel in general?
   a. Very satisfied
   b. Satisfied
   c. Neutral
   d. Dissatisfied
   e. Very dissatisfied
2. How smooth do you think the conversation with the assistant was?
   a. Very smooth
   b. Smooth
   c. Neutral
   d. Not smooth enough
   e. Not smooth at all
3. How would you rate the comprehensibility of communication with the conversation assistant?
   a. Very comprehensible
   b. Comprehensible

  c. Neutral

  d. Incomprehensible

  e. Very incomprehensible

4. How would you rate the level of detail of the troubleshooting instructions given by the assistant?

  a. Very satisfied

  b. Satisfied

  c. Neutral

  d. Dissatisfied

  e. Very dissatisfied

5. If so, what level of detail would you prefer? Higher or lower?

6. How would you rate the overall usefulness of the conversation assistant in helping to deal with repair and other crisis situations?

  a. Very useful

  b. Mostly useful

  c. Neutral

  d. Mostly useless

  e. Very useless

Appendix D

Prototype voice assistant - Czech session guide

Průvodce testováním

Úvod

Dobrý den, rád bych Vás přivítal na testování použitelnosti prototypu hlasového asistenta se zaměřením na emoce subjektu. Jmenuji se Marek Klement a jsem studentem magisterského studia Interakce člověka s počítačem (HCI) na Fakultě elektrotechnické, ČVUT. Součástí mé diplomové práce je úprava prototypu hlasového asistenta, kterého budeme dnes testovat.

Informovaný souhlas

Než začneme, musím Vás požádat o vyplnění a podepsání informovaného souhlasu, aby jste se mohl studie zúčastnit.

Prosím, přečtěte si pečlivě všechny body. Pokud něčemu nebudete rozumět, nebo budete mít doplňující otázky, neváhejte se zeptat. Pokud nemáte žádné další otázky, poprosím Vás o Vaše jméno, dnešní datum a podpis.

Představení experimentu

Nyní mi dovolte vám stručně představit experiment, který tu dnes budeme dělat. V současné době vyvíjíme virtuálního hlasového asistenta pro chytré domácnosti se specializací na krizové jednání při údržbě a opravě spotřebičů. Virtuální hlasový asistent je software, který pracuje na principu komunikace člověka s virtuálním asistentem.

Za účelem vytvoření takového asistenta vzniká několik prototypů, díky kterým můžeme funkci asistenta simulovat a zjistit tak potřebné informace. Prototyp, který dnes budeme testovat se specializuje na emoční aspekt konverzace.

Ačkoliv se jedná pouze o prototyp, snažte se s asistentem komunikovat tak, jako by jste hovořili s plně automatizovaným virtuálním asistentem, jako je například Google asistent, Siri nebo Alexa. Nebojte se asistenta zeptat na to, co Vás zajímá. Tyto dotazy nám pomohou vylepšit prototyp i následného virtuálního asistenta tak, aby měl co největší rozsah.

Dohromady si dnes projdete v rámci experimentu tři scénáře. V tomto případě se bude jednat o simulaci opravy a údržby pračky. S touto činností Vám pomůže právě hlasový asistent.

Za účelem jednodušší představy opravy jsem zde připravil dřevěný model pračky, na kterém budete opravu simulovat. Nyní si můžete pračku prohlédnout a osahat. Zatím však prosím nic neopravujte a neprovádějte.

Rád bych Vám také rád představil několik dalších drobnějších rekvizit, které můžete využít v průběhu testu. Po celou dobu testování budou tyto rekvizity přítomny vedle pračky a Vy je můžete kdykoliv využít. Nachází se zde dvě nádoby, hadr, starý kartáček a několik pěnových kostek.

Nakonec bych Vám rád představil formu komunikace s asistentem. Komunikaci si lze představit jako rozhovor mezi Vámi a rádcem přes opravu praček. Asistent Vás provede krok za krokem opravou/údržbou a vy se ho můžete doptávat na detaily, nebo mu pouze odpovídat. Asistent však nevidí to, co vidíte vy a je třeba mu situaci popsat. Nebojte ptát na podrobnosti a přemýšlet nahlas. Velmi důležitým aspektem jsou také vaše emoce, proto se snažte vžít do situace co nejpřesněji a ve vašich promluvách se nebojte využívat tón a zabarvení hlasu podle dané emoce.

Cíle experimentu

Hlavním cílem experimentu je vylepšení prototypu hlasového asistenta a virtuálního hlasového asistenta. Rozhodně není cílem Vás vydeptat a naštvat. Pokud se něco nepovede, nikdy to není Vaše chyba, ale chyba prototypu.

Pomůcky k použití

- pračka
- dvě nádoby
- pěnové kostky
- starý kartáček
- obrázek s popisem pračky
- prototyp hlasového asistenta

Scénáře

Jak již bylo zmíněno, projdeme zde společně 3 scénáře. Než však začneme, bude dobré si konverzaci vyzkoušet na drobném úvodním scénáři. Každý scénář řeší trochu jiný problém s pračkou, ať už se jedná o její opravu nebo údržbu. Jednotlivé scénáře je možno si projít také v virtuálním emočním asistentovi. Následující odkazy vedou právě na jednotlivé scénáře.

Představení scénářů

Úvod

Nultým scénářem je krátký úvod. Účelem tohoto scénáře je představení formy komunikace s asistentem, zvyknutí si na jeho hlas a upravení hlasitosti zvuku.

Error E20

Tento scénář je prvním netriviálním scénářem. Představte si, že je pátek odpoledne a Vy chcete jet o víkendu na rodinnou dovolenou. Rozhodnete se, že dáte prát prádlo s odloženým startem, aby bylo možné ho pověsit ihned ráno před odjezdem. Když však přijdete k pračce, na obrazovce svítí nápis "Error E20". Vaším úkolem je zjistit, o jakou závadu se jedná a co je třeba udělat, aby byla odstraněna. K tomu Vám pomůže hlasový asistent.

Preventivní čištění filtru přívodní hadice

Představte si situaci, kdy po jednom ze selhání pračky si přečtete, že je dobré preventivně čistit filtr přívodní hadice. Proto se rozhodnete to udělat za pomoci hlasového asistenta. Společně procházíte procesem čištění filtru.

Vytopená koupelna

V tomto případě zažijete velmi nepříjemnou situaci, kdy přijdete do koupelny a ta je vytopená. Ihned zjistíte, že příčinou je právě pračka. Vaším úkolem je zjistit která závada na pračce zavinila vytopení koupelny a následně problém popsat technikovi do telefonu. S odhalením problému Vám pomůže hlasový asistent.

Rozhovor a diskuse po testu

Vše je hotovo a já bych se Vás rád nakonec zeptal na několik dotazů, týkajících se právě hlasového asistenta.

1. Jak se cítíte?
    a. Velmi spokojený
    b. Spíše spokojený
    c. Neutrálně
    d. Spíše nespokojený
    e. Velmi nespokojený
2. Jak přirozená podle Vás komunikace s asistentem byla?
    a. Velmi přirozená
    b. Spíše přirozená
    c. Neutrálně
    d. Nedostatečně přirozená
    e. Absolutně nepřirozená
3. Všiml/a jste si v nějakém scénáři reakce na vaše aktuální emoce?
    a. Ano
    b. Možná
    c. Ne
4. Jak byste ohodnotil/a emoční cítění asistenta?
    a. Velmi dobré
    b. Spíše dobré
    c. Neutrálně

d. Spíše špatné
   e. Velmi špatné
5. Kde přesně jste si toho všimnul/a a jak to na Vás působilo?
6. Jak na Vás působila chybovost asistenta?
7. Jak byste ohodnotil celkovou použitelnost využití?
   a. Velmi užitečné
   b. Spíše užitečné
   c. Neutrální
   d. Spíše nevyužitelné
   e. Zcela nevyužitelné