Design guidelines and user scenarios for an AI-based emotion detection and behaviour change support system for social robots at home

Deliverable 3.1, V0.2

Project title:	AI-based Emotion Detection and Behaviour Change Robot Support System to Benefit Aging Well
Project acronym:	AID2BeWell
Duration:	02/2022 - 10/2022
Project number:	AAL-2021-8-127-SCP
Dissemination level:	Public

Authors: Oliver Jung, Salzburg Research Forschungsgesellschaft mbH, AT

Contributors: -

Versions

Version	Date	Changes	Author(s)
V0.1	21.12.2022	Outline and structure of report, inclusion bullet points and unstructured content	Oliver Jung
V0.2	22.01.2023	Inclusion of spelled out content for user scenarios as well as general, software and hardware guidelines	Oliver Jung
V1.0	31.01.2023	Integration of feedback	Oliver Jung

Project summary

Active and healthy ageing includes promoting physical health for mobility and independence, but also mental health for well-being and high quality of life. In this aging society leveraging technology to support active and healthy ageing to reduce the risk of depression or health problems is challenging but promising. There is a wide range of lifestyle interventions that can promote healthy aging when done in the most effective way. A key technology that could help older adults in their home environment is a robot platform that delivers adaptive personal behaviour change suggestions promoting well-being. Recent robots for older users are usually simple systems primarily combating loneliness or specific tasks (taking medications) or complex systems aimed at larger health care organisations. With this proposal, we intend to create an AI-based emotion detection behaviour change personal robot support system (ROSS).

With a strong focus on end-user requirements, we will find the best way to develop ROSS to benefit the older user. Emotion recognition will go further than basic emotions by including complex emotions and longer mood states based on arousal and valence levels. This will ultimately allow for more empathic and beneficial human-robot interactions. The innovative nature of AID2BeWell lies in (1) the primary end-user focus and involvement, (2) beyond the state of the art AI emotion and mood recognition, and (3) the integration of Just-In-Time Adaptive Interventions (JITAIs). This project could be the key to bringing affordable social robots to the market to help individuals age well.

Contents

1. Introduction

This document describes the design guidelines for an AI-based emotion detection and behaviour change support system for social robots at home. It covers (1) the problems that can arise when developing such a social robot system, (2) how to approach these, and (3) what are the lessons learned. The findings stem from literature research, previous projects, the user requirements documented in D1.1, the development phase and the two evaluation phases.

2. User scenarios for a social robot

D1.1 already elaborated on (1) design and functionalities, (2) potential usage scenarios, and (3) practical handling as concluded from the focus group discussion findings. Consequently, this deliverable will focus on challenges and required changes that did arise due to technical limitations and in-field results. More information about the detailed approaches to follow in the software and hardware design guidelines.

<u>Design and functionalities</u>

- Individual naming of the robot
  Since virtual keyboard entry is not user-friendly on a 7 inch touchscreen and speech-to-text tends to be unreliable when it comes to names we decided to provide a pre-defined selection in order to name the robot. The names have been equally mixed between female, male and gender-neutral options.
- Better contrast / different colouring of LEDs for displaying emotions
  The LEDs of Q.bo only allow for the colours blue, green and red. We selected blue for the mouth LEDs since it provided the best visibility and is not associated to a positive or negative implication. Furthermore, we reduced the amount of displayed emotions to a well-known positive, neutral and negative smiley visualization.
- Different interaction and communication modes
  We provided the output via text as well as speech and the input via touch as well as speech (wherever possible). Text size and voice have been customizable.
- Emotion recognition and reactions
  The answers of the robot as well as the decision trees and interventions have been adapted to respect the emotions of the user, e.g. choosing motivating phrases when detecting negative emotions.
- Connectivity with other devices
  While the connectivity with other devices, such as health trackers or smart home appliances, might serve beneficial the integration and usage of such data was out of scope of the project due to its limited duration and the study design.
- Integration of external services
  While the integration of external services, such as a shared calendar or grocery list, would add to the usage scenarios this was also out of scope of the project due to requiring more elaborate authentication methods.

<u>Potential usage scenarios</u>

- The robot's role as a motivator
  This scenario has mainly been addressed by the usage of simple JITAIs and emotion recognition. The robot would (1) promote interaction by addressing the user at suitable points of time (whenever the users face is within view of the robot during the day and the last interaction took place some time ago), (2) provide motivational feedback based on the mood of the user and the state of the current exercise, and (3) present interventions tailored to the needs and previous interaction of the user with various alternatives.
- The robot's role as a communication partner
  This scenario has mainly been addressed by the usage of text-to-speech and speech-to-text services as well as emotion recognition. The speech services should allow for a natural interaction and dialogue flow. A dedicated decision tree branch has been designed for conversations with the user. The possible answers take emotional changes into account.

Practical handling

- Setup and support
  When thinking into the direction of commercialization a very straightforward setup as well as a 24/7 support hotline is required. For the prototype evaluation the soft- and hardware has been equipped with plenty of automatic (re)start features and fallback options. The setup at the study site or users home has been performed by the end-user organizations. A tutorial was presented upon starting the first interaction and printed out as a handbook.
- Power supply
  For long-term use or combination with a mobile platform a powerful rechargeable battery combined with a display of charging state and low power warnings or automatic charging station is advisable. For the static setup during the prototype evaluation the usage of a socket was sufficient.

3. General and content-related guidelines

Before going into detail regarding concrete design guidelines for the software and hardware we will point out a few general and content-related topics to take into account when designing a robot support system that did not yet arise in the user scenarios.

Elaborate decision trees & JITAIs

Each branch of the decision tree should offer a multitude of possible paths in order to promote frequent interaction as well as keeping the topic interesting in the long run. Furthermore, each leaf should offer a multitude of interventions so the robot can learn about user preferences and adapt the provided information. The longer the typical usage duration of a topic, the more variation has to be provided. At the same time the number of decision tree layers and answer options should be limited for a pleasant user experience.

Reliable speech recognition

Speech recognition is possibly the most important feature when it comes to interaction design with a social robot. The topic at hand is extremely complex. Usually there is a trade-off between reliability, quality, costs and data protection. First of all, the hardware should not hinder the achievable results. Furthermore, one has to decide between an online or offline solution. Also, a choice has to be made between ready-to-use solutions with pre-trained models and frameworks for using custom training data. There is no approach that fits every need but the market offers plenty of solutions.

Target groups

The complexity of the usage, the scope of the exercises and the wording of the whole application should be tailored to the target group. Ideally, the application offers various presets of different complexity and multiple wording options for each text node in order to allow for a great degree of customization and tailor to the needs of each individual user.

Gamification

Gamification means the integration of playful elements into a non-gaming context with the aim of changing the behaviour of users and increasing their motivation. Typical game elements include, for example, ranking lists with the scores achieved, awards for achieving certain goals, the progress display of tasks or the mutual evaluation of users. Gamification was proven to be a reliable tool for achieving long-term motivation for any user group.

Data protection

Data protection should be integrated throughout the whole application by design. Apart from complying with the GDPR it is essential to keep the user informed whenever their data is used for any purpose and be transparent about the need for data collection in order to raise acceptance. In general, data collection should be as minimal as possible and only cached locally and/or anonymized wherever possible.

Funding models

For everyday use of social robots, both the acquisition and maintenance costs must be within an affordable range for the intended target group. There are various funding and/or renting models conceivable. Those will be addressed in detail in D3.2 – Report on evaluation of business potential.

Use cases

There are numerous relevant use cases for a social robot support system that have been outlined in D1.1. Apart from the envisioned user scenarios and functionality such a system can have multiple fields of application, e.g. support of nursing staff, preventive measures in the health sector, support in health management for chronic diseases, profiling to support healthy lifestyles, and many more.

Gender-related recommendations

A number of studies confirm that - just as in human-to-human communication - communication between a human and a robot is influenced by gender stereotypes. Subtle gender cues, such as the robot's voice or hair length, influence the perception of the robot as female or male and the clichéd skills or tasks of the robot associated with these categorizations. Consequently, it is advisable to stick to a gender-neutral design for any exterior or immutable components, provide gender-neutral options and wordings within the application and let the user decide for their personal preferences.

4. Software design guidelines

Decision trees

As previously mentioned each branch of the decision tree should offer a multitude of possible paths and each leaf should offer a multitude of interventions so the robot can learn about user preferences and adapt the provided information. Due to the very limited evaluation period this only applied to our project partially. The decision trees have been split into four main areas with multiple exercises/topics for each area:

- Games
    - Memory: A typical game of memory on a 4x4 board
    - Reaction: A game where the user is supposed to click appearing disks as fast as possible
    - Agility: A “Simon says” game with colour combinations
- Wellbeing
    - Meditation: A guided meditation video
    - Positive emotions: Various information about positive emotions originating from the HAP model
    - Positive events: Various information about positive events originating from the HAP model
    - Gratitude: A short exercise for thinking about gratitude
    - Personal strengths: A short exercise for thinking about personal strengths
- Exercise (Body & soul)
    - Mindfulness: A guided video about mindfulness
    - Yoga: A guided seated yoga exercise
    - Agility: A short Japanese agility exercise that can be performed seated or standing (Radio taiso)
    - Negative thought: A short exercise for avoiding negative thought (5-4-3-2-1 method)
- Conversation
    - Leisure: A structured conversation with up to four layers regarding leisure activities
    - Mood: A structured conversation with up to four layers regarding today’s mood
    - Stress: A structured conversation with up to four layers regarding today’s stress level
    - Sleep: A structured conversation with up to four layers regarding today’s sleep (problems)
    - Loneliness: A structured conversation with up to four layers regarding feeling lonely

Due to the limited evaluation period there has been a maximum of three interventions for the same leaf. Nevertheless, emotion-based triggers have been implemented in order to offer more variety and serve as JITAIs. Those will be addressed in more detail in the paragraph about facial analysis.

While the users enjoyed most of the tasks and had plenty to try out they pointed out that more variation and proactive interaction would be required for long-term engagement.

Speech interaction

The importance and complexity of reliable speech recognition has been outlined in the previous chapter. For our project, we decided to use a ready-to-use solution with pre-trained models because of a very limited implementation period. It should be integratable via an API and offer support for both German and Dutch. At the same time, servers should be located in the EU and the provider should have a good standing and compliance with GDPR. Following these constraints, we decided to use the speech-to-text and text-to-speech services by IBM Watson.

In general, the results have been very satisfying in both languages. Unfortunately, the internet dependency and bad mobile data coverage during the second pilot phase caused delays and unpredictable behaviour for some users. Also, background noise was hard to filter with the used hardware. For an actual in-home setup a stable WiFi connection or an offline solution are crucial.

Facial analysis

The usage of facial analysis was crucial for the design of emotion-based triggers and JITAIs. A key element of the project was the development of a FaceReader SDK for Linux. Due to the limited processing power and training data available, the SDK has been targeted at detecting the presence of a face and the valence level over a given period of time, which worked reliably during testing the second evaluation phase.

Also, the algorithm had to be optimized for the camera model and the relatively short distance between the camera and the user.

Open source

Wherever possible open source software has been used during development and will be published as such as long as it does not interfere with any IP. Permissive licences are always preferred.

GDPR

The software and any other survey or research material has been designed in compliance with GDPR. A data management plan was developed and users were clarified about the data usage and their rights concerning information and deletion.

Furthermore, the hardware and software indicated whenever audio or video recordings were being made and discarded any cached data right after processing – apart from minimal anonymized log data required for further development.

5. Hardware design guidelines

Processing power

The main processing unit of the social robot Q.bo used in our project is a Raspberry Pi 3rd Generation. While this is sufficient for using the built-in hardware and running some basic scripts, any more complex calculations can cause significant overloads.

Since we required more computation power – especially for the facial analysis and additional hardware – we added an Intel NUC (in the cheapest variant with a Celeron CPU and 8GB of RAM) to the setup.

The communication between the NUC and Q.bo took place via a Websocket. While a serial connection via the LAN port would have been possible, the Websocket causes very little traffic since it only transfers basic commands between the units and the internet connectivity was already in place for the speech recognition software.

The additional hardware and cabling was hidden inside an unobtrusive box next to the robot during the second pilot phase.

Camera

The built-in camera(s) of the social robot Q.bo are relatively cheap – consequently, they are using a small sensor and got a limited field of view. While this is sufficient for basic tasks, it serves a huge problem for the facial analysis, which needs to use a high resolution image showing as much of a face as possible in order to generate meaningful results.

In order to achieve these requirements a high definition camera with a 2.9" IMX322 sensor and a distortion free 100-degree lens was added to the setup.

It was attached to the robot head using a self-designed 3D-printed mount. It was positioned right between the existing eyes for an optimal viewing height and angle and hidden underneath a headband during the second pilot phase.

Sound

The built-in microphone and speaker of the social robot Q.bo are relatively cheap resulting in a low recording quality and tinny sound output at higher volumes.

The quality of the speakers was improved by using an amplifier provided by the robot manufacturer and attaching an external speaker using the jack connection. In doing so, the volume could also be adjusted on the speakers and not only at the software level.

The recording quality is essential for the speech-to-text quality. In order to avoid yet another external piece of hardware, we chose to use the microphone integrated in the third, added camera. The recognition results improved drastically in doing so.

Open hardware

One of the main selling points of the social robot Q.bo is the usage of open hardware – namely Arduino and Raspberry. Unfortunately, the Arduino board is actually a custom-made variant running all the hardware connections to the head (servos, LEDs, etc.). At the same time the versions used for the Raspberry hardware and software are outdated but cannot be easily replaced due to the dependency on the connection to the head and the drivers used for the communication between both boards.

While we fully support the usage of open hardware, such design decisions eliminate their benefit. In order to actually make use of the setup for any current and/or more complex application one would have to replace most components apart from the casing and servos.

Mobile platform

While we developed a mobile platform for a previous project, it was not required for the use cases of our project and is not straightforward to replicate. In general, a mobile platform consists of quite a few hardware and software components, such as a drive, 3D cameras and/or lidars, a separate powerful processing unit (preferably GPU-based), a separate power supply, SLAM algorithms, etc.

A mobile platform should be part of a robot design from the beginning in order to achieve optimal results.

Power supply

For the (simulated) in-home setups of both pilot phases, power could easily be supplied using a socket. Unfortunately, this limits the use cases to stationary ones and makes moving between rooms or users way harder. For a mobile use case, a rechargeable battery would be required. It has to be relatively powerful in order to supply processing units and servos at the same time. In order to optimize the user experience low battery warnings and an automatic charging station are preferable.

6. Summary and conclusion

This report presents the key findings regarding design guidelines for an AI-based emotion detection and behaviour change support system for social robots at home in the project AID2BeWell.
While we were able to collect multifaceted requirements for a robot support system in D1.1 and tried to do them justice, we had to cut corners due to the limited implementation time and insufficient initial hardware setup.

The main lessons learned are:

- Take content design (decision trees and interventions) serious and offer plenty of variety
- Keep users engaged by proactive interaction and gamification
- Allow for a huge degree of customization
- Tests speech interaction under real conditions
- Get hardware that suits your needs right from the start

In general, it is advisable to design and build a robot that fits the needs of your use case(s) from scratch for more extensive projects or collaborate with companies building such systems.
We will try to apply our findings in a follow-up project and work towards commercialization.