Embodied Conversational Agents' Appearance for Health Assessment of Older Adults: An Explorative Study

Silke ter Stal, Marijke Broekhuis, Lex van Velsen, Hermie Hermens, Monique Tabak

Submitted to: JMIR Human Factors
on: May 12, 2020

Table of Contents

Embodied Conversational Agents' Appearance for Health Assessment of Older Adults: An Explorative Study

Silke ter StalMSc, ; Marijke BroekhuisMSc, ; Lex van VelsenPhD, ; Hermie HermensProf Dr, ; Monique TabakPhD,

Corresponding Author:
Silke ter StalMSc,
Phone: +31088 0875 777
Email: s.terstal@rrd.nl

Abstract

Background: Embodied conversational agents (ECAs) have great potential for health applications, but are rarely investigated as part of such applications. To promote the uptake of health applications we need to understand how the design of the ECAs can influence the preferences, motivation and behavior of its users.

Objective: This is a first study that investigates how the appearance of an ECA implemented within a health application affects the users' likeliness of following agent advice, their perception of the agent characteristics, and feeling of rapport. In addition, we assessed usability and intention to use.

Methods: The ECA was implemented within a frailty assessment application in which three health questionnaires were translated into agent dialogues. In a within-subject experiment, questionnaire dialogues were randomly offered by a young female agent or an old male agent. Participants were asked to think aloud during interaction. Afterwards, they rated the likeliness of following the agent's advice, agent characteristics, rapport, usability and intention to use, and participated in a semi-structured interview.

Results: Twenty older adults (72.2±3.5 years) participated. The old male agent was perceived as more authoritative than the young female agent (P=0.03), while no other differences were found. The application scored high on usability (median 6.1) and intention to use (median 6.0). The participants indicated they did not see an added value of the agent to the health application.

Conclusions: In conclusion, agent age and gender little influence users' impressions after short interaction, however, remain important at first glance to lower the threshold to interact with the agent. Thus, it is important to take the design of ECAs into account when implementing them into health applications.

(JMIR Preprints 12/05/2020:19987)
DOI: https://doi.org/10.2196/preprints.19987

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Embodied Conversational Agents' Appearance for Health Assessment of Older Adults: An Explorative Study

Abstract

Background: Embodied conversational agents (ECAs) have great potential for health applications, but are rarely investigated as part of such applications. To promote the uptake of health applications we need to understand how the design of the ECAs can influence the preferences, motivation and behavior of its users.
Objective: This is a first study that investigates how the appearance of an ECA implemented within a health application affects the users' likeliness of following agent advice, their perception of the agent characteristics, and feeling of rapport. In addition, we assessed usability and intention to use.
Methods: The ECA was implemented within a frailty assessment application in which three health questionnaires were translated into agent dialogues. In a within-subject experiment, questionnaire dialogues were randomly offered by a young female agent or an older male agent. Participants were asked to think aloud during interaction. Afterwards, they rated the likeliness of following the agent's advice, agent characteristics, rapport, usability and intention to use, and participated in a semi-structured interview.
Results: Twenty older adults (72.2±3.5 years) participated. The older male agent was perceived as more authoritative than the young female agent (P=0.03), while no other differences were found. The application scored high on usability (median 6.1) and intention to use (median 6.0). The participants indicated they did not see an added value of the agent to the health application.
Conclusions: In conclusion, agent age and gender little influence users' impressions after short interaction, however, remain important at first glance to lower the threshold to interact with the agent. Thus, it is important to take the design of ECAs into account when implementing them into health applications.

Keywords: embodied conversational agent; appearance design; health status assessment; older adults; eHealth

Introduction

When people get older, people might suffer from frailty: a decline in functional and cognitive functions, such as a decrease in walking speed, balance control and working memory [1, 2]. Through eHealth, frailty can be assessed using digital questionnaires. By doing so, a large population can be targeted, including those who are less mobile and face difficulties in seeing a caregiver to perform frailty assessment. In addition, when performing frailty assessment digitally, the assessment can be performed on a regular basis, can be dynamically adapted based on information provided by the user, and can immediately provide the user with the results of the assessment. Subsequently, an eHealth application can coach the user in a personalized way towards a healthy lifestyle based on the outcomes of the frailty assessment. Research shows that collecting health data using a digital survey does not affect test reliability with respect to a paper version [2, 3, 4] and several studies showed similar results for a population of older adults [5, 6] In addition, Fanning & McAuley showed that older adults may accept a tablet for health surveys and Van Velsen et al. showed that older adults preferred a tablet survey to a paper survey.

Research shows that the older and more frail adults get, the more they become non-respondents to questionnaires [7, 8], whereas refusal of face-to-face interviewing is less present in this population

[7]. To overcome the problem of lack of face-to-face interaction in a digital frailty assessment, an embodied conversational agent (ECA) could provide an alternative. Embodied conversational agents are more or less autonomous and intelligent software entities with an embodiment used to communicate with the user [9]. By interacting with the user face-to-face, embodied conversational agents can build trust and rapport – a close and harmonious relationship –, leading to companionship and long-term continual use [10].

To establish trust and rapport with the agent, it is important that users have a positive impression of the agent. These impressions can be shaped by both static characteristics [11] and dynamic characteristics [11, 12]. Static characteristics mostly relate to an agent's visual appearance – often tested using the so-called *zero acquaintance approach*: a person observing the agent without interacting with the agent . Dynamic characteristics include an agent's verbal and non-verbal behaviors – often tested using the *thin slicing approach*: a person drawing inferences about an agent's personality based on short excerpts of social behavior [13].

Although embodied conversational agents have the potential to be used as eHealth applications, such as digital frailty assessment, little is known about how these agents should be designed and how the design affects our impressions of the agents; no design guidelines exist [14]. In a previous study, ter Stal, Tabak, et al. identified people's first impressions of various agents, differing in age, gender and role, using a so-called *zero acquaintance approach*: there was no interaction involved; participants rated static agent images at first glance [15]. The study shows that characteristics of older and male agents were perceived differently from characteristics of young and female agents respectively. In addition, older adults seem to prefer a young female over an older male agent. Also, other research focused on users' perceptions of static agent images at first glance [16, 17, 18] showing that the agent's gender and role affect the user's perception of the agent. However, little research focuses on what people's impressions are after short interaction with the agents, and how the design of the agents affects these impressions. Therefore, research is needed to investigate how the design of an agent affects the users' impressions of the agent during and after actual interaction (using a *thin slicing approach*).

The aim of this study is *to assess how an agent's appearance, in particular age and gender, affects the users' likeliness of following agent advice and the users' perception of the agent's characteristics and feeling of rapport after short interaction with the agent*. This study builds upon previous work [15] by studying users' impressions of agents at first glance (using a zero acquaintance approach) and after a short interaction with the agents (using a thin slicing approach) instead of studying their impressions only at first glance. As a secondary aim, we investigate the potential of a frailty assessment application with an agent by *evaluating its usability and intention to use*.

Methods

The Frailty Assessment Application

The ECA under study was embedded within a frailty assessment application. This is a web application, developed as part of a larger platform designed to counter frailty by offering older adults training modules in the domains of healthy nutrition, and physical and cognitive training to maintain a healthy lifestyle [19]. Use and continued use of the platform is stimulated by integrating gamification elements. In this study, we focused on the stand-alone frailty assessment application.

The frailty assessment application consisted of an index page (Figure 1) and a dialogue page (Figure 2). On the index page, an agent was displayed next to a blackboard. The blackboard provided a list of

available dialogues: an introductory small talk dialogue, a dialogue incorporating a questionnaire to assess an aspect of the older adult's health, or a small talk dialogue explaining the results of the questionnaires. When finishing a dialogue, the user returns to the index page. Before performing the questionnaire dialogues, only the introductory small talk was available on the blackboard. In this dialogue, the users were introduced to the agent and the goal of the frailty assessment. Afterwards, the questionnaire dialogues were unlocked and shown on the blackboard. For the questionnaire dialogues, three validated questionnaires were implemented to assess the older adult's frailty status covering multiple health domains:

- 36-Item Short-Form Health Survey (SF-36)
- Alzheimer Disease Detection (AD8)
- Mini-Nutritional Assessment (MNA-SF)

The SF-36 [20] contains 36 multiple-choice questions related to different topics (e.g., physical functioning, social functioning). The AD8 [21] tests for functional decline in memory using eight yes or no items. The MNA-SF [22] tests for malnutrition by six multiple-choice questions related to nutrition and weight. We translated the three frailty assessment questionnaires into dialogues between the agent and older adults. After performing the questionnaires, the result dialogue was unlocked on the blackboard. In this dialogue, the users received the outcomes of the assessment.

Only one dialogue was available at a time. When clicking on the start button of a dialogue, the dialogue page opened (Figure 2) A dialogue consisted of multiple dialogue steps. Each dialogue step consisted of a statement by the agent and one or more reply options that could be selected by the user. The statement by the agent was shown in the white box with the orange border and the reply options for the user were listed in the black box. After finishing a dialogue with the agent, the user returned to the index page and the available dialogues list on the blackboard was updated.

Figure 1. The starting page of the frailty assessment (on the left: Sylvia, on the right: Egbert).

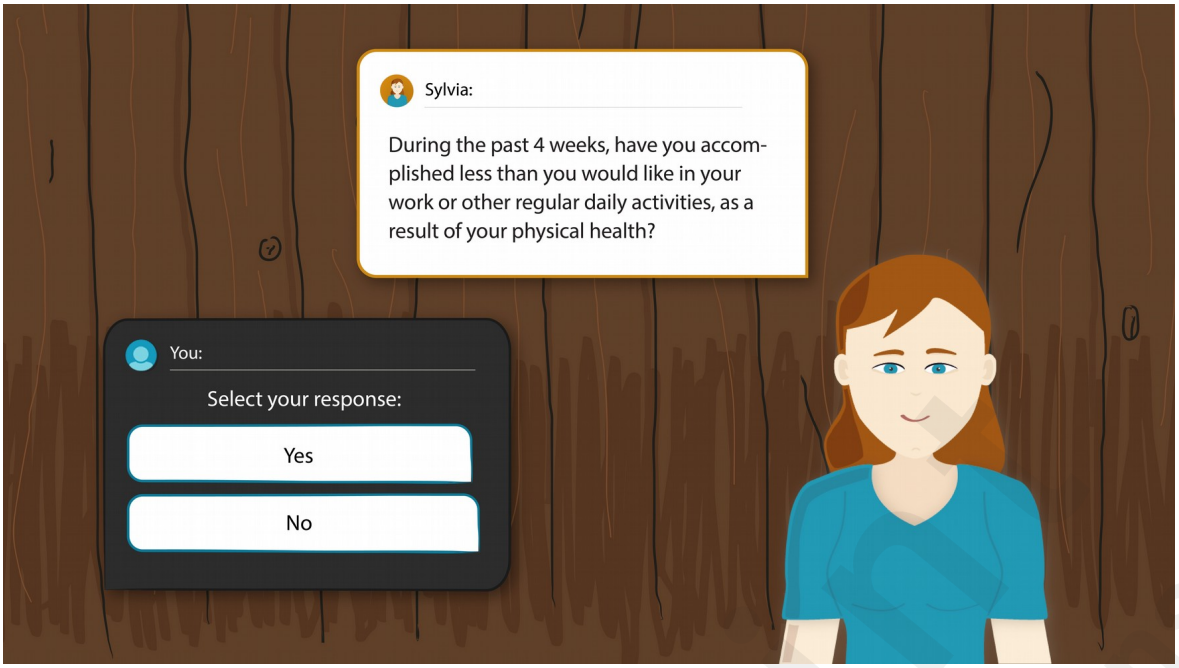


Figure 2. The dialogue page with the agent Sylvia.

The agents used in the frailty assessment application are displayed in Figure 3. The agent on the left is Sylvia, the young, female peer agent. The agent on the right is Egbert, the older, male peer agent. By a peer agent, we mean an agent who is not a medical expert. The agent designs were selected based on findings from a previous study [15]. In this previous study, static images of eight agents were evaluated. The agent images differed on three features: age (young or old), gender (male or female) and role (expert – having a high level of health expertise – or peer – having a low level of health expertise). In an online questionnaire, images of the agents were shown to the participant at once. Then, the participant selected the agent that he or she preferred most (to be his or her health coach) at first glance. Afterwards, the participant rated a set of characteristics for every agent. Results showed that a young female agent was preferred most, whereas an older male agent was preferred least, in both a general and elderly population (i.e. these designs were extremes in terms of user preference). This study builds upon the previous study by evaluating users' impressions of these two agents, both, at first glance and after short interaction with the agents, instead of solely at first glance. For both agents, a blinking eyes animation was implemented. In addition, when the agent "spoke" (i.e. when a new dialogue step was loaded), a mouth animation of a fixed duration was played.

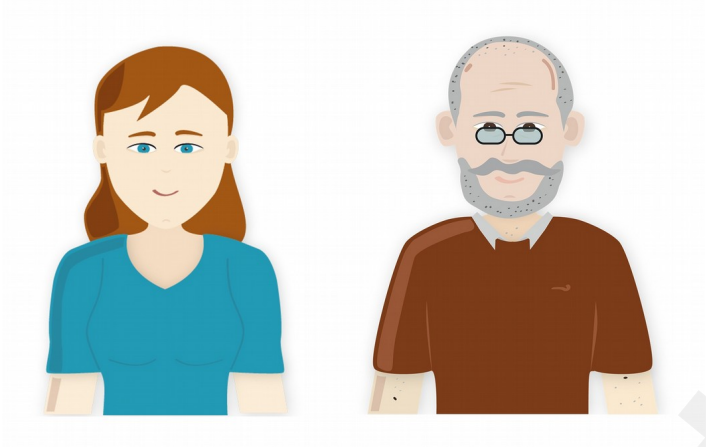


Figure 3. The two different agents used during the experiment.

Study Design

We applied a within-subject design in which we counterbalanced the order in which the agents were presented to the participants. Half of the participants started the frailty assessment with the young, female peer agent (A) and finished with the older, male peer agent (B) (as seen in Figure 4 - I). The other half of the participants were first presented with the older male peer agent (B), followed by the young, female peer agent (A) (as seen in Figure 4 - II).. The study was performed in a lab-setting, taking place either at a research institute, or at a local physiotherapy practice. The nature of this general study with healthy volunteers from the general population does not require formal medical ethical approval, according to Dutch law. All participants provided their informed consent.

Figure 4. Study design including randomisation process.

Participants

Participants should be aged 65 years or above and fluent in the Dutch language in order to be included. In addition, they should be cognitively able to work with an ECA, as assessed via the Mini-Mental State Examination (MMSE), scoring at least 23 out of 30 points [23]. We recruited the respondents via a Dutch panel of adults that indicated they were interested in participating in research on eHealth. In addition, participants were recruited via a local physiotherapy practice.

Measurements

Questionnaires

Before interacting with the frailty assessment application, the participant filled out the pre-interaction questionnaire gathering the participant's gender, date of birth, education, housing status, technology literacy, health literacy and state of change for nutrition and physical activity [24].

After interacting with each agent (Figure 4), the participant filled out the post-interaction questionnaire. To investigate the *effect of the agent's appearance* we assessed:

- Likeliness of following the agent's advice (on a 7-point Likert scale)
- Ratings agent characteristics (all on 7-point Likert scales): friendliness, authority, involvement, reliability, intelligence
- Rapport scale agent (all on 7-point Likert scales) by Acosta & Ward [25]: emotional rapport, cognitive rapport, helpfulness, trustworthiness, likeability, naturalness, enjoyableness, humanlikeness, persuasiveness, recommend-ability

Secondarily, we investigated the *usability* of the frailty assessment application and the *intention to use* the frailty assessment application both on a single 7-point Likert scale.

Thinking Aloud

In order to triangulate the quantitative data, participants were asked to think aloud while interacting with the frailty assessment application. Audio was recorded and screen captures were taken. The researcher did not help or support the participant, but only reminded the participant to think out loud when he or she became forgot to.

Interviews

At the end of the session, the participant was interviewed. The interview was semi-structured and guided by asking the user's opinion regarding positive and negative aspects, around *the effect of the agent's appearance, the usability of the frailty assessment application and intention to use the frailty assessment application*.

Data Analyses

SPSS 25 was used to perform statistical analyses. Since the underlying data was non-parametric, for all relations testing differences between the two agents, a Wilcoxon Signed Ranks test was conducted. All tests used a 95% Confidence Interval. All variables were tested for statistically significant differences between the two agents by means of a model consisting of Wilcoxon Signed Ranks tests for cross-over designs. Effect size was calculated by $r = Z/\sqrt{N}$, using 0.1, 0.3 and 0.5 as cut-off values for a small, medium and large effect respectively.

The audio recordings of the thinking aloud sessions and interviews were transcribed and then inductively thematically analyzed. In addition, screen captures of the interaction with the frailty assessment application were aligned with the audio recordings. This way, the screen captures were used to verify the thoughts of the participants on the audio recordings. All themes were coded using ATLAS.ti 8, based on an empirical method, proposed by [26]. One researcher (StS) created a first coding scheme, based on the data, and then labelled the transcripts. A second researcher (MB) used the coding scheme to code a subset of the data, so that a discussion could be held between the first and second coder for improving the coding scheme. The procedure of creating a first coding scheme,

labelling the data by two researchers and discussing the coding scheme, was repeated a second time, leading to a final coding scheme. The final coding scheme was used by the first coder to code all data for final analyses. The final coding scheme contained the following codes: agent characteristics, appearance agents, interaction with agents, preference agent, content questionnaires, language usage in dialogues, presentation information, interaction with application, design, navigation, general computer interaction and intention to use.

Results

Participants

Twenty-one participants participated (Table 1). One participant was not able to complete the protocol, due to a lack of computer experience, and was excluded. The average age of participants was 72.2 years (SD = 3.5 years). Thirteen males and seven females participated. Ten participants started with the young, female agent and ten participants started with the older, male agent.

Table 1. Participant demographics.

Demographic		n
Education	Elementary school	1
	High school	1
	Vocational education	8
	College	6
	University	4
Living situation	Living alone	1
	Living with a partner	19
Stage-of-change nutrition	Maintenance	18
	Precontemplation	2
Stage-of-change physical activity	Maintenance	13
	Action	3
	Contemplation	1
	Precontemplation	2
	Unknown	1
Technology literacy	Moderate or high technology literacy level	20
Health literacy	Moderate or high health literacy level	19
	Low health literacy level	1
Physical limitations	No risk at facing physical limitations	9
	Risk at facing physical limitations	10
	Already faced physical limitations	1
Cognitive limitations (MMSE)	No risk at facing cognitive limitations (score ≥ 23)	19
	Risk at facing cognitive limitations (score < 23)	1

Agent Appearance

Table 2 shows the questionnaire results regarding a) the likeliness of following the agent's advice, b) the users' perceptions of the agent characteristics (e.g. friendliness, expertise) and c) the users' feeling of rapport (e.g. emotional rapport, helpfulness) for both agents. Corresponding box plots can be seen in Figure 5 and Figure 6. For the ratings of the likeliness of following the agent's the advice, no significant difference between Egbert and Sylvia was found. However, Egbert was rated significantly more authoritative than Sylvia (P = 0.03), resulting in a medium effect size (r = 0.344). No significant differences were found between the agents for all other agent characteristics and the rapport scale items.

Table 2. Results of the Wilcoxon Signed Ranks tests (N = 19 or 20) comparing the mean ranks of the ratings of likeliness of following the agent's advice, agent characteristics and rapport scale items.

	Median Egbert (Q1 – Q3)	Median Sylvia (Q1 – Q3)	Z	*P*
Likeliness of following advice	5.0 (3.3 – 6.0)	6.0 (4.0 – 6.0)	-1.613	0.11
Agent characteristics				
Friendliness	6.0 (5.0 – 6.0)	6.0 (5.0 – 6.0)	-0.264	0.79
Expertise	5.0 (4.0 – 6.0)	5.0 (4.0 – 6.0)	-0.966	0.33
Reliability	5.0 (4.0 – 6.0)	5.0 (4.0 – 6.0)	-0.276	0.78
Authority	2.0 (2.0 – 4.0)	2.0 (1.0 – 4.0)	-2.121	*0.03
Involvement	4.5 (4.0 – 6.0)	5.0 (4.0 – 6.0)	-0.158	0.88
Rapport scale				
Emotional rapport	4.0 (2.0 – 5.0)	4.0 (3.0 - 5.0)	-1.310	0.19
Cognitive rapport	4.0 (4.0 – 5.0)	5.0 (3.3 – 5.8)	-0.829	0.41
Helpfulness	5.0 (4.0 – 6.0)	5.0 (4.0 – 6.0)	-0.877	0.38
Trustworthiness	5.0 (4.0 – 6.0)	5.0 (4.0 – 6.0)	0.000	1.00
Likeability	6.0 (4.0 – 6.0)	6.0 (4.3 – 6.0)	-0.604	0.55
Naturalness	5.0 (4.0 – 6.0)	5.0 (4.0 – 6.0)	-0.491	0.62
Enjoy-ability	5.0 (3.0 – 6.0)	4.0 (4.0 – 6.0)	-0.182	0.86
Human-likeness	4.0 (3.3 – 6.0)	4.5 (3.3 – 5.0)	-0.486	0.63
Persuasiveness	5.0 (4.0 – 6.0)	5.0 (4.0 – 6.0)	-0.942	0.35
Recommendability	5.0 (4.0 – 6.0)	5.0 (4.0 – 6.0)	-0.368	0.71

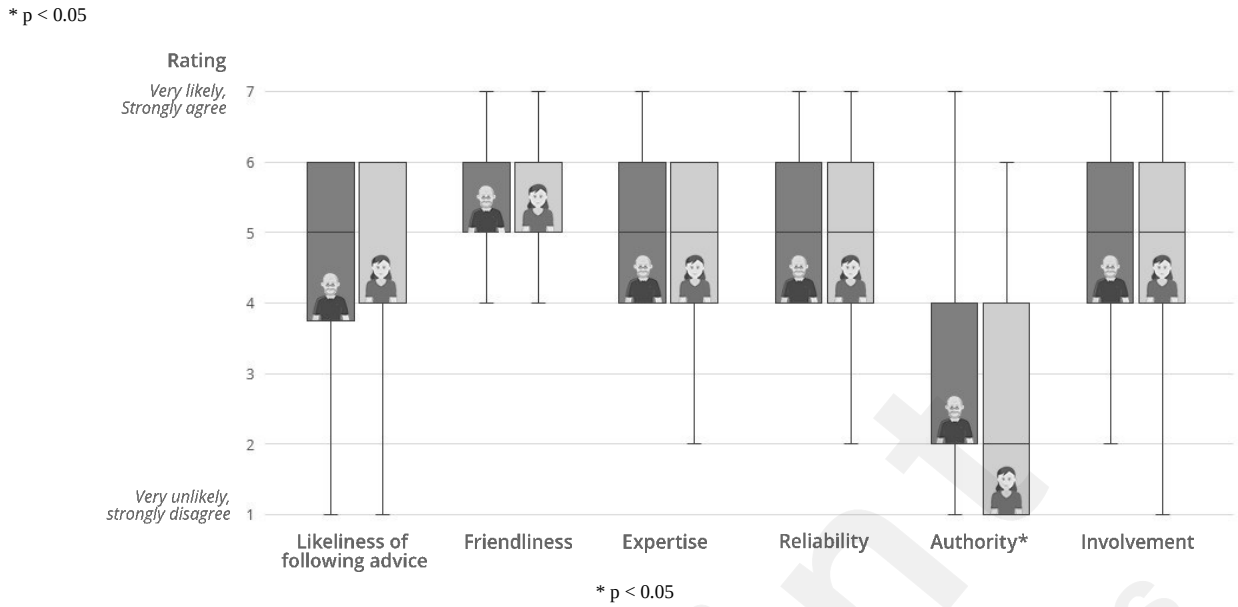


Figure 5. Box plots showing the ratings of the likeliness of following advice and characteristics of the two agents.

Figure 6. Box plots showing the ratings of the rapport scale items of the two agents.

The analysis of the thinking aloud sessions and interviews resulted in the following themes with respect to the effect of the agent appearance: *agent characteristics*, *appearance agents*, *interaction with the agents and preference agent*. The rest of the section describes the results per theme.

Users' Perception of the Agent's Characteristics

A few participants indicated that they had trouble getting an impression of the agents' personalities or found it difficult to connect personality to embodied conversational agents in general. A few others perceived the agents as natural and not artificial. On the other hand, the majority did not perceive the agents as human: they perceived the agents as cartoons, static dolls, computers or machines. As one of these participants indicated:

> "It is a computer, it is still interaction from a distance, it does not become personal, it does not have any personality, I do not feel a connection." – Male, 68 years

Another participant said:

> "The agents remain computers, you cannot call them friendly or unfriendly, they are computers and I do not connect any human characteristics to them." – Male, 78 years

In the interviews, some participants indicated they did not perceive the agents differently with respect to their personality. A few participants explained that both agents used friendly language, whereas others argued the agents were friendly, since they responded in a way that fit the situation and provided compliments. In addition, a participant explained that both agents were not too young, nor too old and seemed to be modern people due to responses, such as "gosh, how nice". Also, this participant said he liked that the agents were not too young, since a young agent would not have much experience. One participant particularly indicated that the female agent was friendlier than the male agent was, whereas another participant believed that the male agent was higher educated and more intellectual than the female agent.

Users' Perception of the Agent's Appearance

A participant indicated that the agents looked like cartoons or drawings, whereas she preferred the agents to look like real humans. This participant also indicated that the blinking eyes and mouth animation were distracting.

The rest of the comments related to the appearance of either one of the agents. One person particularly mentioned the female agent having a friendly face, whereas all other comments related to the male agent. The appearance of the older male agent evoked several associations, such as the agent looking old, and, therefore, unhealthy. Others associated the older male agent with a scientific staff member, a nerd or a male of the type of "wearing sandals with socks", because of his glasses and popular beard. Participants preferred an energetic, spontaneous person and one that is more neutral and clean-shaven. One participant did not like the male agent, because he associated the agent with his or her uncle, having a similar name: a spoiled man with whom you would not be able to connect. Another participant found the male agent more distracting than the female agent, because of its glasses.

Users' Perception of the Interaction with the Agent

Several participants explicitly indicated that they expected or would like the agent to speak. One participant expected the agent to speak due to its mouth animation, whereas another had this expectation, since humans interact via speech in real conversations. Another participant pointed out that, due to the absence of agent speech, the user has to multi-task: the user simultaneously has to read and answer the questions and pay attention to the agent. Therefore, she would like the agent to speak, as illustrated by the thought:

> "Well, I have to read what you say to me, but instead open your mouth yourself!" – Female, 73 years

Other opinions on the interaction with the agent focused on the naturalness of the interaction with the agent, as illustrated by a participant saying:

> "It felt as if there was a real human in front of me." – Female, 71 years

Another participant described the interaction as actually talking to someone and yet another participant described the interaction as having a phone call, in which someone is checking how you are doing. Some participants were less positive. A few participants especially said that the interaction with the agents was impersonal. As one participant described:

> "Actually, I do not have the feeling I am really communication with someone." – Female, 65 years

Another participant said that she did not take part in a conversation, but was simply reading and answering questions. This participant did not establish a connection with the agents, as explained by:

> "I barely know her" – Female, 65 years and "Understanding each other? Then one would expect interaction." – Female, 65 years

Last, some comments related to the implemented small talk. On the one hand, some participants seemed to like the small talk, reflected by them laughing. On the other hand, a participant was irritated by the implemented small talk, she felt being treated like a child.

Agent Preference

The majority of the participants indicated they did not prefer one of the two agents. Most of them indicated they did not have a preference, since they perceived the agents similar. Some even did not remember that they interacted with two different agents. However, some participants did show a preference. Some participants preferred the female agent, either because they believed the agent was friendlier or discussed a more interesting topic. Only one participant preferred the male agent, but could not argue why.

Usability and Intention to Use Frailty Assessment Application

The questionnaire results show that *the usability of and intention to use the frailty assessment application were high*: the 20 usability ratings displayed a median of 6.1 (IQR = 6.1 – 7.0) and the 20 intention to use ratings displayed a median of 6.0 (IQR = 4.0 – 6.0) on a 7-point Likert scale.

During the thinking aloud session and interviews participants commented pointed out usability issues of the frailty assessment application or provided suggestions for improvements of the application. The following themes were identified: content questionnaires, language usage in dialogues, presentation information, interaction with application, design, navigation and general computer interaction. Figure 7 shows the frequencies of the topics on which the participants commented.

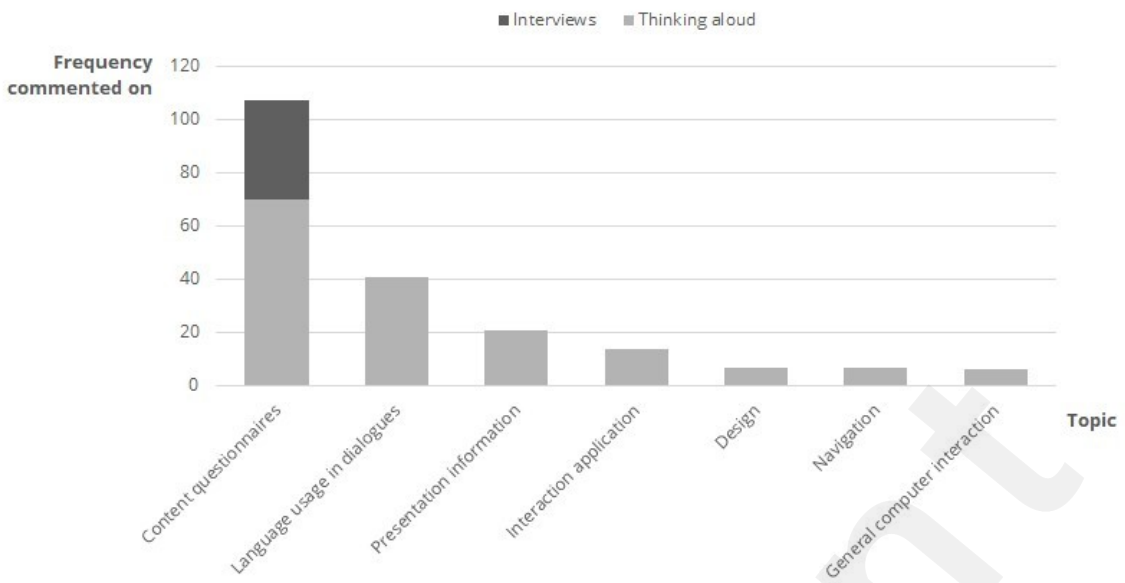


Figure 7. Frequencies comments on the frailty assessment application during both the interviews and thinking aloud method.

The majority of the comments or suggested improvements related to the content of the questionnaires and to the language in the application. The majority of the participants found that the questionnaires did not fit their personal situation, contained a lot of repetition or ambiguity. Participants suggested adapting the questionnaires according to previous answers given. In addition, participants commented on the language used: words being ambiguous, too popular or too old fashioned, unnecessary, patronizing or not being known by people with a lower education or older adults. Furthermore, participants commented on the length and structure of the sentences and pointed out spelling mistakes. A participant suggested to adapt the language used in the application to the education of the user. Considerably fewer comments related to the presentation of information, interaction with the application, the design or navigation of the application or general computer interaction. As an example, with respect to navigation, some participants indicated they would like to go back to a previous dialogue step.

With respect to the intention to use, the thinking aloud sessions and interviews showed that, on the one hand, the minority of the participants would like to use the application. A participant indicated he would not use the application him or herself, but would recommend the application to others who might benefit from it. On the other hand, some participants clearly indicated they would not use the application. The majority of the participants indicated that the agents did not have any benefit to the application. The argument supporting this opinion was that the application was not personal, since the answers that could be given were limited, and the possibility to explain them was missing. In order for the application to be beneficial, a participant believed that the application should also provide advice on what actions the user should perform to become more healthy. Another participant explicitly stated that he would use the application when the text was replaced by speech.

Discussion

Our results show that *the appearance of an agent, in particular age and gender, affects users'*

perceptions of the agent's authority, but does not affect users' perceptions of other agent characteristics, users' feelings of rapport and users' likeliness of following agent advice. Compared to a young female agent, an older male agent is only seen as more authoritative. These results are not in line with our expectation that agents are perceived differently after short interaction with a user. To the best of our knowledge, there is no existing research comparing users' impressions of agents at first glance with those after short interaction. But, research shows that in human-human interaction, first impressions, formed within milliseconds [27], are difficult to loose. Therefore, we assumed that the differences in perceptions of characteristics of a static image of a young female agent and an older male agent as found in a previous study [15], would still be present after short interaction with these agents. An explanation for this inconsistency could be that impressions in human-agent interaction differ from impressions in human-human interaction. Users' judgements of agents may modify with ongoing interaction, as research shows that agents do have a second chance to make a first impression [12, 28]. Therefore, differences in perceptions of both agents may have been present at first glance, but disappeared after interaction. Further research is needed to confirm this finding. Future research could study users' perceptions of an agent's characteristics with a larger study population. Eventually, agents will be used in a long-term setting, therefore, it is interesting to research not only users' perceptions at first glance and after short-term interaction, but also after long-term interaction.

How do we explain the difference in perception of the agents' authority after short interaction? Although research on short-term interaction with an agent indicates that an agent's appearance, such as its clothing [17], racial concordance with the user [29, 30] and similarity with the user [29, 31], could affect users' perception of the agent, to the best of our knowledge, there is no research on an agent's authority after short interaction in particular. From a previous study [15], we see that at first glance, indeed, static images of male and older agents are seen as more authoritative than female and young agents respectively. In addition, the study shows that the differences found in authority are often higher compared to differences found for other characteristics tested, which could explain why the difference in authority level is still present after short interaction. However, since we did not control the age and gender of the agents in this study independently, it is difficult to say whether the difference in perception of the agents' authority is caused by the agents' age or gender in particular, or solely by the combination. Future research could study which factor(s) actually control the difference, researching the users' perceptions of agents' authority by independently controlling the age and gender of the agents. In addition, future research could study how an agent's authority is perceived after long-term interaction.

We expect that the effect of the first impression established by an agents' age and gender on the impression after short interaction is small compared to the effect of other design features, such as the content and language of the messages, (absence of) agent speech and the amount of embodiment. Our study shows that the majority of the participants did not perceive the agents as humans, but as machines or cartoons, and found the interaction with the agents impersonal or artificial. They did not have the feeling of being in a conversation. These perceptions may point out that the users had a negative adaptation gap [28]): a user overestimating the competency of an agent, creating a negative gap between the expected and actual competency of the agent, resulting in the user being disappointed. This negative adaptation gap may have been caused by the content and language of the agents' messages, the agent lacking speech or the agent having little embodiment, as supported by the remarks made by the participants during the thinking aloud sessions and interviews. Therefore, we believe it is important to manage the users' expectations of the agent characteristics and functionality upfront. Ensure that the users' expectations match the actual capabilities of the agent by explaining what the users can expect from the agent. Future research could study how an agent's content, language, speech and embodiment affect users' perception of the conversation with the agent

(e.g. how these factors could make the conversation with an agent more human-like).

Although our study shows that an agent's age and gender has little effect on the users' impression of this agent after short interaction, we believe that adapting these features to the user is important, since they affect the user's impression of the agent at first glance [16, 18, 32] and research shows that people with favorable impressions of someone tend to interact more with that person than others having unfavorable impressions [33]. Selecting an agent with the right age and gender could thus lower the threshold to interact with the agent and to use the application.

Second, our results show that overall, *the usability of the developed frailty assessment application is judged positively*. Yet, the content and language of the questionnaires could be improved, since the majority of usability issues relate to either the content or language of the questionnaires. We suggest tailoring the content and language towards the personal characteristics of the user, as confirmed by existing research [34], and adapt the content to previous answers given by the user.

Third, *not all participants show an intention to use the application*. Research indicates that older adults put effort into learning new digital technologies as long as they believe they are worthy of time and dedication: when the technology can be used to keep in touch with others to foster relationships [35]. Similarly, research shows that elderly value applications that address a social problem [36]. The application used in our study did not address a social problem, which could have resulted in some participants not seeing the added value of the application and not showing an intention to use the application. In addition, elderly's intention to use digital technologies is, next to the quality of the technology itself, affected by their personal context (e.g. their ability to concentrate) and social context (e.g. whether family is around to provide technical support) [36]. Both factors might have affected participants' intention to use the frailty assessment application in our study.

More specifically, the majority of the participants does not believe the agent to be of added value to the frailty assessment application. Therefore, we suggest to update the design of the agent. We believe that the agent should convey additional information to its message in text via its embodiment. Existing research provides evidence for implementation of animations of the agent's embodiment, showing that animations positively affect the users' impression of the agent [37, 38, 39] and the interaction time [12, 38]. In addition, the use of speech is recommended, since it could increase the sense of personality of the agent [40] and could be used to describe feelings [41]. Especially low-literate users could benefit from multiple output modalities [42]. Furthermore, participants indicated they would like the application to provide advice on what actions they should perform in order to life healthy. We see an opportunity for using the agent to provide this advice. As an example, the agent could show videos of exercises to improve physical strength.

Strengths and Limitations

This is the first study that specifically evaluates effects of the agent's appearance after short interaction with the agent. In addition, this study uses actual health content, which is scarce in research on agent design.

Our study also has some limitations. First, the negative adaptation gap between the users' expectations of the agent's capabilities and the agent's actual capabilities, might indicate that the application used might not have been mature enough. The agent conveyed the majority of the information via text. The participants might have been focused on reading the text, and, therefore,

might have paid little attention to the agent, resulting in the participants having difficulties in creating impressions of the agent characteristics and establishing rapport. Second, the interaction time with the agents might have been too short to create impressions of the agent characteristics and to establish rapport. Third, although we found a difference in the users' perceptions of authority of the young female and the older male ECA, it is difficult to specify whether this difference was caused by either the ECA's gender or age, since these factors were not independently controlled in the study.

Based on these findings, we have created a set of recommendations for the design of digital health assessment with ECAs, as listed in Textbox 1. We believe these recommendations would increase the users' intention to use such an application.

Conclusions

Our study shows that an agent's appearance, in particular age and gender, only affects users' perceptions of the agent's authority after short-term interaction. We conclude that adapting the agent's age and gender to the users' preferences seems to be important to lower the threshold to interact, whereas the content and language of the agent's messages and agent speech and embodiment seem to be important factors for the users' impressions of the agent after short interaction.
We believe that ECAs have potential to be used in digital frailty assessment, but future research is needed. Future research could study users' perceptions of agents after long-term interaction, study whether users' perceptions of an agent's authority are related to an agent's age or gender in particular and study how an agent's content, language, speech and embodiment affect users' perceptions of the conversation with the agent.

Textbox 1. Recommendations for future research on digital frailty assessment with ECAs.

Towards Digital Frailty Assessment with ECAs: Recommendations for Future Research

Develop, implement and evaluate the following aspects for an ECA for digital frailty assessment:

Agent Design Implications

- *Create sufficient embodiment* – Convey empathy or emotion, using the agent's embodiment. This way, the agent design may positively affect the user's impression of the agent and interaction time.
- *Let the agent speak* – Reduce the user's cognitive load by providing the agent messages in speech. This way, the agent design may positively affect the user's impression of the agent.
- *Adapt the agent appearance to the user* – Select an agent appearance that fits the age and gender of the user. This way, the agent design may lower the threshold to start using the application.

Prerequisites Frailty Assessment

- *Adapt the content to the user's personal situation* – Take into account the user's personal situation, such as the user's disabilities and living situation. Adapt the questionnaire so that the user does not get confronted with questions that cannot be answered correctly, since it does not apply to the user's situation.
- *Adapt the questionnaire to previous answers* – Save the answers given by the user and adapt the questionnaire accordingly. This way, users do not have to answer questions that are not applicable to them.
- *Use language that fits the user's educational level* – Adapt the agent's language based on the educational level of the participant, so that the language is neither too simple nor complex.

Acknowledgements

This work was supported by FRAIL (Interventions on frailty and ageing risks for elderly people based on ICT tools), funded by the Eurostars-2 Programme (no.10824).

Conflicts of Interest

None declared.

Abbreviations

AD8: Alzheimer Disease Detection
ECA: embodied conversational agent
IQR: inter quartile range
MMSE: Mini-Mental State Examination
MNA: Mini-Nutritional Assessment
SF-36: 36-Item Short-Form Health Survey

References

[1] J. O. Malva and J. Bousquet, “Operational definition of active and healthy ageing: Roadmap from concept to change of management,” *Maturitas 84*, pp. 3-4, 2016.

[2] B. D. Bliven, S. E. Kaufman and J. A. Spertus, "Electronic collection of health-related quality of life data: Validity, time benefits, and patient preference," vol. 10, no. 1, p. 15–21, 2001.

[3] K. T, M. P, H. H, D. K, D. A. G, A. T and U. T, "Performance of health status measures with a pen based personal digital assistant," *Annals of the Rheumatic Diseases*, vol. 64, no. 10, p. 1480–1484, 2005.

[4] H. R, S. A, M. K, F. G and K. W, "Patient difficulty using tablet computers to screen in primary care," *Journal of General Internal Medicine*, vol. 23, no. 4, p. 476–480, 2008.

[5] L. Van Velsen, S. Frazer, A. N’Dja, N. Ammour, S. Del Signore, G. Zia and H. Hermens, "The reliability of using tablet technology for screening the health of older adults," *Studies in Health Technology and Informatics*, vol. 247, p. 651–655, 2018.

[6] J. Fanning and E. McAuley, "A comparison of tablet computer and paper-based questionnaire," *Journal of Medical Internet Research*, vol. 16, no. 7, p. 1–12, 2014.

[7] R. Hébert, G. Bravo, N. Korner-Bitensky and L. Voyer, "Refusal and information bias associated with postal questionnaires and face-to-face interviews in very elderly subjects," *Journal of Clinical Epidemiology*, vol. 49, no. 3, p. 373–381, 1996.

[8] J. A. Hardie, B. P. S. and O. Mørkve, "Non-response bias in a postal questionnaire," *Scandinavian Journal of Public Healt*, vol. 31, no. 6, p. 2003, 411–417.

[9] Z. Ruttkay, C. Dormann and H. Noot, "Embodied conversational agents on a common ground: A framework for design and evaluation," in *From Brows to Trust: Evaluating Embodied Conversational Agents*, 2004, pp. 27-66.

[10] L. P. Vardoulakis, L. Ring, B. Barry, C. L. Sidner and T. W. Bickmore, "Designing Relational Agents as Long Term Social Companions for Older Adults," in *International Conference on Intelligent Virtual Agents*, 2012.

[11] A. Cafaro, H. H. Vilhjálmsson and T. Bickmore, "First impressions in human-agent virtual encounters," *ACM Transactions on Computer-Human Interaction*, vol. 23, no. 4, pp. 1-10.

[12] K. Bergmann, F. Eyssel and S. Kopp, "A second chance to make a first impression? How," in *International conference on intelligent virtual agents*, Berlin, Heidelberg, 2012.

[13] O. Vartanian, K. Stewart, D. R. Mandel, N. Pavlovic, L. McLellan and P. J. Taylor, "Personality assessment and behavioral prediction at first impression," *Personality and Individual Differences*, vol. 52, no. 3, p. 250–254, 2012.

[14] S. ter Stal, L. L. Kramer, M. Tabak, H. op den Akker and H. Hermens, "Design Features of Embodied Conversational Agents in eHealth: a Literature Review," *International Journal of Human-Computer Studies*, 2020.

[15] S. ter Stal, M. Tabak, H. op den Akker, B. Beinema and H. Hermens, "Who do you prefer? The effect of age, gender and role on users’ first impressions of embodied conversational agents in eHealth," *International Journal of Human Computer Interaction*, 2019.

[16] J. Forlizzi, J. Zimmerman, V. Mancuso and S. Kwak, "How interface agents affect interaction between humans and computers," in *Conference on Designing Pleasurable Products and Interfaces*, 2007.

[17] D. Parmar, S. Olafsson, D. Utami and T. Bickmore, "Looking the part: the effect of attire and setting on perceptions of a virtual health counselor," in *Proceedings of the 18th international*

conference on intelligent virtual agents.

[18] J. Zimmerman, E. Ayoob, J. Forlizzi and M. McQuaid, "Putting a face on embodied interface agents," in *Designing pleasurable products and interfaces*, 2005.

[19] A. F. A. Noorman-de Vette, Designing Game-Based Ehealth Applications Strategies for Sustainable Engagement of Older Adults, 2019.

[20] K. van der Zee and R. Sanderman, "Het meten van de algemene gezondheidstoestand met de rand-36," 1993. [Online]. Available: https://www.umcg.nl/SiteCollectionDocuments/research.

[21] J. E. Galvin, C. M. Roe, M. A. Coats and J. C. Morris, "Patient's Rating of Cognitive Ability," *Archives of Neurology*, vol. 64, no. 5, p. 725, 2007.

[22] L. Z. Rubenstein, J. O. Harker, A. Salvà, Y. Guigoz and B. Vellas, "Screening for undernutrition in geriatric practice: Developing the Short-Form Mini-Nutritional Assessment (MNA-SF)," *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, vol. 56, no. 6, p. 366–372, 2001.

[23] R. Kok and F. Verhey, Gestandaardiseerde Mini-Mental State Examination, Janssen, 2002.

[24] J. O. Prochaska and W. F. Velicer, "The transtheoretical model of health behaviour change," *American Journal of Health Promotion*, vol. 12, no. 1, pp. 38-48, 1997.

[25] J. C. Acosta and N. G. Ward, "Achieving rapport with turn-by-turn, user-responsive," *Speech Communication*, vol. 53, no. 9-10, pp. 1137-1148, 2009.

[26] C. Pope and N. Mays, "Qualitative methods in health research," *Qualitative research in health care*, p. 1–11, 2006.

[27] M. Bar, M. Neta and H. Linz, "Very first impressions," *Emotion*, vol. 6, no. 2, p. 269–278, 2006.

[28] T. Komatsu, R. Kurosawa and S. Yamada, "How does the difference between users' expectations and perceptions about a robotic agent affect their behavior?: An adaptation gap concept for determining whether interactions between users and agents are going well or not," *International Journal of Social Robotics*, vol. 4, no. 2, 2012.

[29] S. Zhou, T. Bickmore, M. Paasche-Orlow and B. Jack, "Agent-user concordance and satisfaction with a virtual hospital discharge nurse," in *International conference on intelligent virtual agents*, 2014.

[30] S. Zhou, Z. Zhang and T. Bickmore, "Adapting a persuasive conversational agent for the Chinese culture," in *International conference on culture and computing (culture and computing)*, 2017.

[31] A. V. Wissen, C. Vinkers and A. V. Halteren, "Developing a virtual coach for chronic patients: a user study on the impact of similarity, familiarity and realism," in *International conference on persuasive technology*.

[32] H. Nguyen and J. Masthoff, "Is it me or is it what I say? Source image and persuasion," in *International conference on persuasive technology*, 2007.

[33] H. H. Kelley, "The warm-cold variable in first impressions of persons," *Journal of Personality*, pp. 431-439, 1950.

[34] S. Beukema, L. van Velsen, S. Jansen-Kosterink and J. Karreman, "“There is something we need to tell you . ”: Communicating health-screening results to older adults via the internet 1," *Telemedicine and e-Health*, vol. 23, no. 9, 741–746.

[35] S. Lindley, R. Harper and A. Sellen, "Desiring to be in Touch in a Changing Communications Landscape: Attitudes of Older Adults," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*, 2009.

[36] J. Waycott, F. Vetere, S. Pedell, A. Morgans, E. Ozanne and L. Kulik, "Not for me: Older adults choosing not to participate in a social isolation intervention," in *CHI Conference on Human Factors in Computing Systems*, 2016.

[37] A. L. Baylor and J. Ryu, "The effects of image and animation in enhancing pedagogical agent persona," *Journal of Educational Computing Research,* vol. 28, no. 4, p. 373–394, 2003.

[38] S. H. Kang, A. W. Feng, A. Leuski, D. Casas and A. Shapiro, "The Effect of An Animated Virtual Character on Mobile Chat Interactions," in *3rd International Conference on Human-Agent Interaction,* 2015.

[39] A. Cowell and K. Stanney, "Embodiment and interaction guidelines for designing credible trustworthy embodied conversational agents," in *International workshop on intelligent virtual agents*, 2003.

[40] C. Nass and K. M. Lee, "Does computer-generated speech manifest personality? An experimental test of similarity-attraction," in *Conference on Human Factors in Computing Systems*, 2000.

[41] G. Veletsianos, C. Miller and A. Doering, "Enali: a research and design framework for virtual characters and pedagogical agents," *Journal of Educational Computing Research,* vol. 41, no. 2, p. 171–194, 2009.

[42] I. M. Thies, "User interface design for low-literate and novice users: Past, present and future," *Foundations and Trends in Human-Computer Interaction,* vol. 8, no. 1, p. 1–72, 2014.

[43] L. P. Fried, C. M. Tangen, J. Walston, A. B. Newman, C. Hirsch, J. Gottdiener, T. Seeman, R. Tracy, W. J. Kop, G. Burke and M. A. McBurnie, “Frailty in older adults: evidence for a phenotype.,” *Journal of Gerontology: Medical Sciences,* no. 56A, pp. 146-156, 2001.

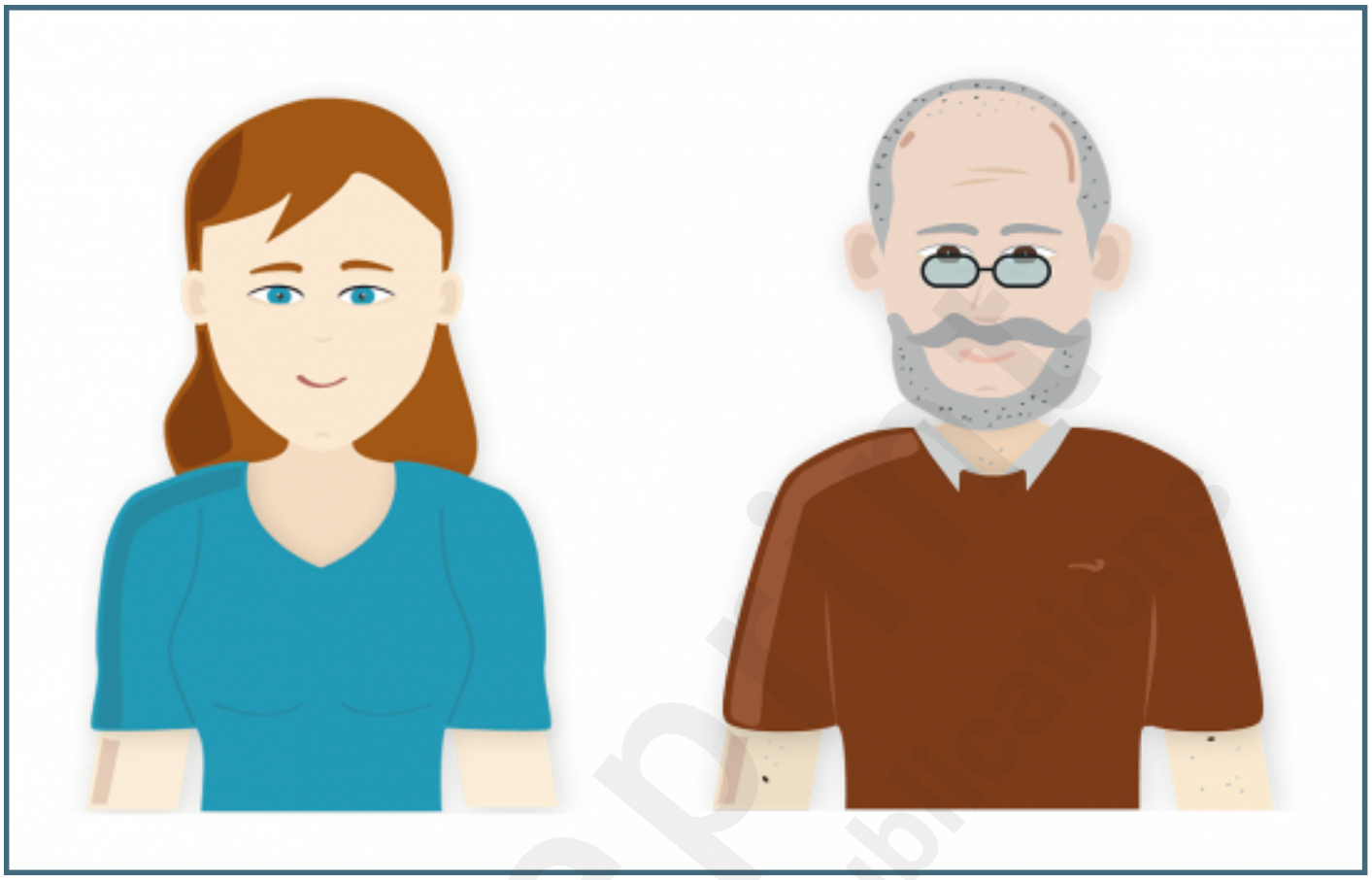
Supplementary Files

Figures

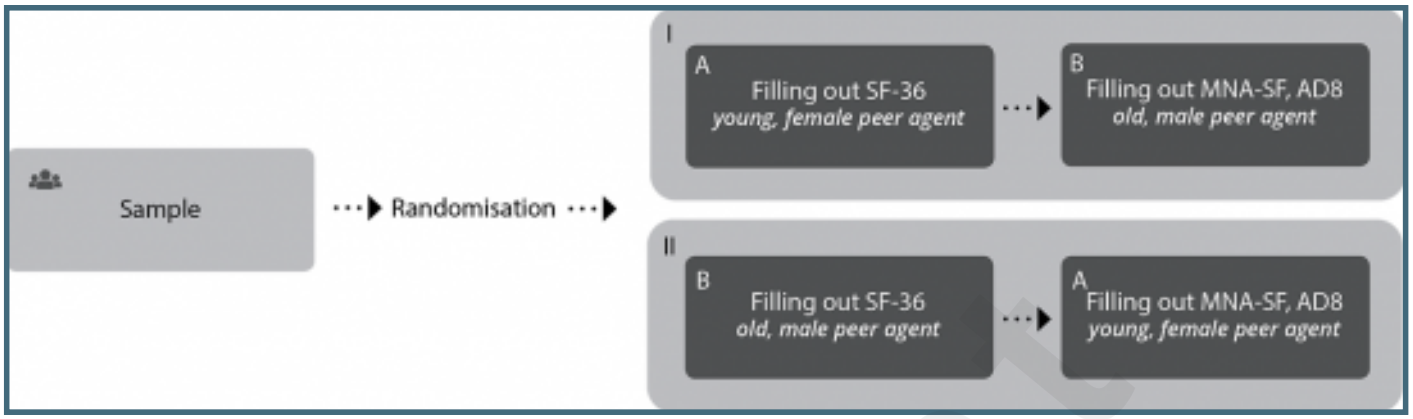
The starting page of the frailty assessment (on the left: Sylvia, on the right: Egbert).

The dialogue page with the agent Sylvia.

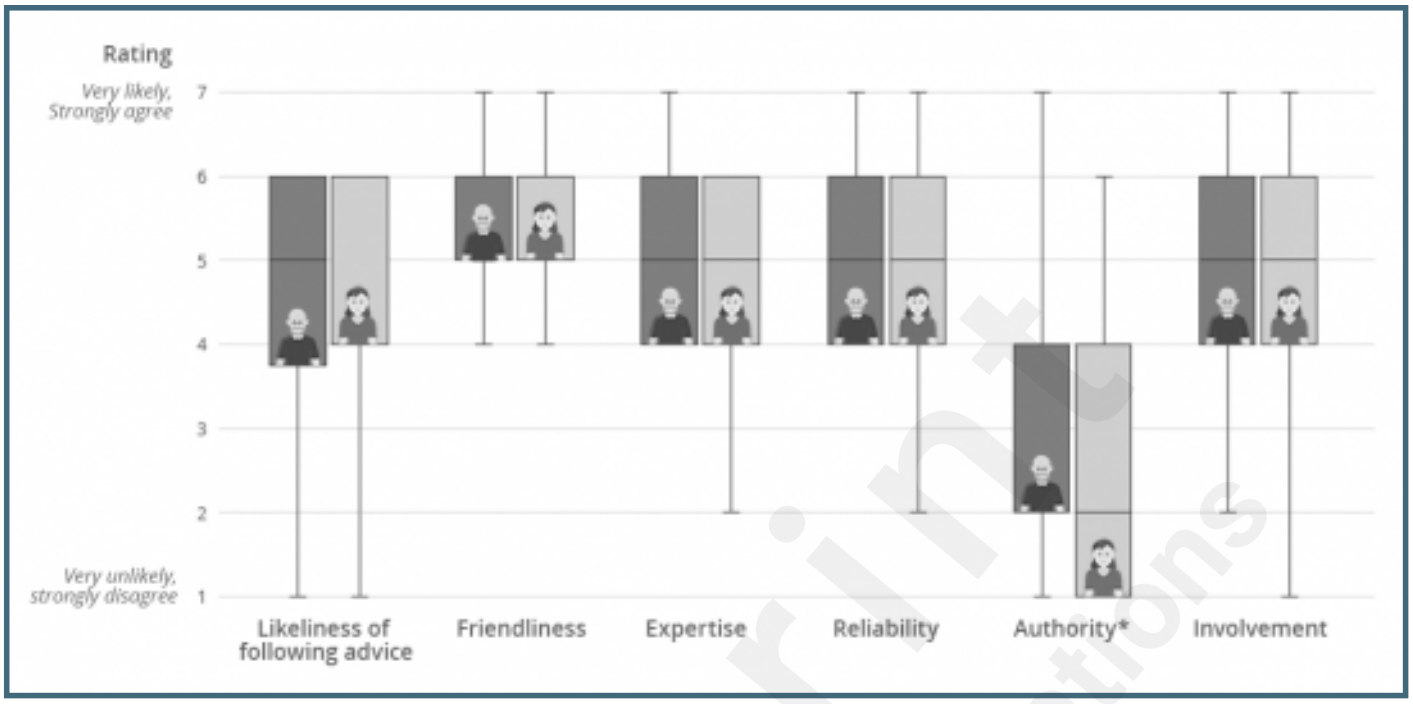
The two different agents used during the experiment.



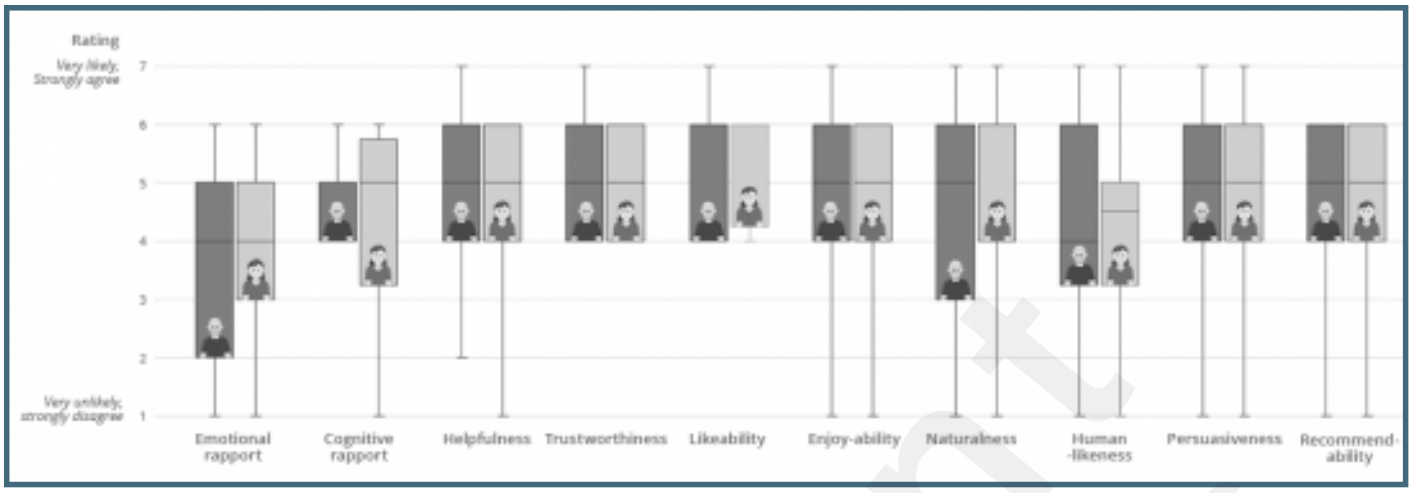
Study design including randomisation process.



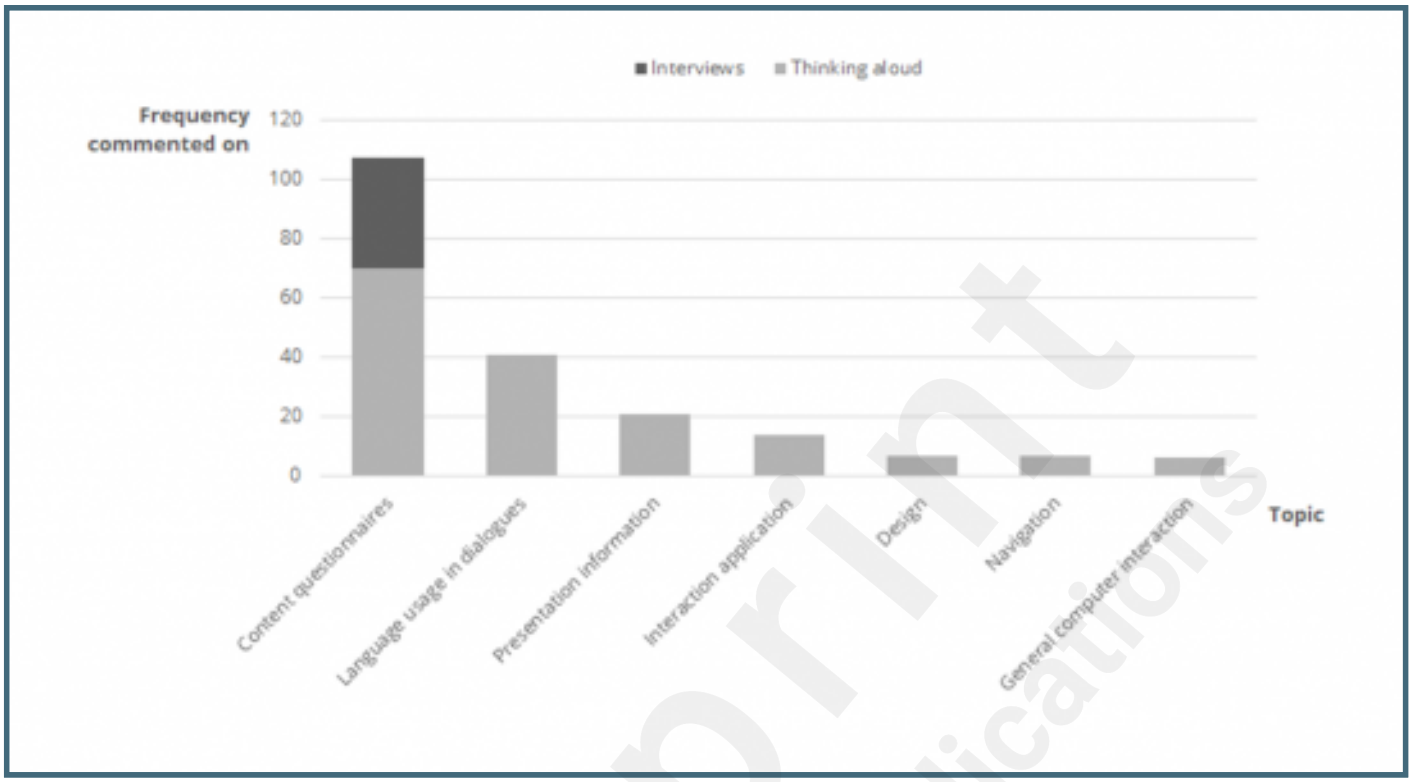
Box plots showing the ratings of the likeliness of following advice and characteristics of the two agents.



Box plots showing the ratings of the likeliness of following advice and characteristics of the two agents.



Frequencies comments on the frailty assessment application during both the interviews and thinking aloud method.



TOC/Feature image for homepages

TOC/Feature image.