Designing a Virtual Lifestyle Coach for Elderly People: A User Study on the Effect of Visual Realism

Geanne Woertink

July 2020

University of Twente

BSc Creative Technology

Supervisor: Job Zwiers

Critical Observer / Client: Silke ter Stal

Roessingh Research and Development

Abstract

Embodied Conversational Agents (ECAs) can be used to motivate elderly to adopt a healthier lifestyle, it is relevant to optimise the ECAs for elderly by adapting their preferences. The aim of this research is to get insight in the effect of an ECA's visual realism on elderly's perception of the personality of the ECA as a lifestyle coach. Short interactions with three static agent images, differing in the amount of visual realism, were created and iterated during one-to-one co-design sessions with five elderly. Respondents (n = 64) who interacted with the resulting prototypes could rate the ECA's characteristics – friendliness, reliability, empathy, expertise, communicative, seriousness – and select their preferred ECA. A part of the respondents (n = 10) were interviewed for detailed understandings supporting these impressions. Our results suggest that (1) people have a clear preference for a photorealistic agent image, (2) the agent's visual realism affects the perception of the characteristics of the agent image and that (3) people do not seem to feel that they have talked to a person while interacting with the agent. In addition, the interviews showed that audio and animation, preferably as realistic as possible, could improve the experience of talking to a real person. Future work could do a similar larger-scale research, and research the effect of the realism of the audio and animation.

Acknowledgements

First of all, I would like to thank my supervisor Job Zwiers and my critical observer and client Silke ter Stal, from Roessingh Research and Development, for their supervision; they have helped me think critically and explorative and supported me throughout my research. Secondly, I would like to thank Alieke van Dijk for sharing her expertise on psychology and coaching to help me improve my questionnaire and research.

I also want to thank the people from the co-design session group, Aukelien, Cobi, Henk, Jan, and Jelte, for joining the three co-design sessions and helping me create and iterate the final mock-up. Finally, I want to thank all participants who joined my research, especially the ten people who were willing to elaborate on their answers in an interview; Baukje, Eric, Gepke, Henriëtte, Janny, Jeanet, Klaas, another Klaas, Mattie, and Roelof.

Table of Contents

1. Introduction

Lifestyle coaches help people to make changes in their lives to enhance their wellbeing. This type of coaching has become a popular means of helping non-clinical populations to set and reach goals with enhancing their wellbeing as goal [1]. Lifestyle coaching would benefit elderly people to help them meet their recommendations for physical activity established by the World Health Organisation (WHO). One of these recommendations is that they should do physical activities at least 150 minutes per week [2]. Care for elderly and their health is increasing as the population is aging [3]–[5]. To keep up with this development, technological possibilities that can replace or add to this demanding care are being investigated.

Additions to the Health Care Professionals (HCPs) may be Embodied Conversational Agents (ECAs): virtual characters that simulate face-to-face conversations with users [6]. ECAs can be used to persuade and motivate users to adopt a healthier lifestyle. They are proved to be helpful for improving health behaviours of people [7]. The use of ECAs in the context of eHealth has been researched a lot lately, also in the context of lifestyle coaching although this is a novel field [7]–[10]. These studies investigate the user's perception of the personality of such an agent, which influences the willingness of people to take its advice. However, the target group of elderly people is quite unexplored although they would benefit from a coach. It is relevant to adapt the ECA in an accessible way for elderly people. It is therefore relevant to research how ECAs work best for them. In this study, the focus is on the aspect of visual realism. This aspect is already researched several times [11]–[14], but not yet for the goal of a lifestyle coach for elderly people, while this target group might have different ways to perceive such a virtual coach.

Therefore, the main goal of this research was to get insight in the effect of visual realism on the perception of the personality of the ECA in the form of a lifestyle coach for elderly people. A human-centred investigation was done with two-dimensional ECAs in a mobile app interface where the features of the agents differ in the amount of realism.

Our results suggest that (1) people have a clear preference for a photorealistic agent image, (2) the agent's visual realism affects the perception of the characteristics of the agent image and that (3) people do not seem to feel that they have talked to a person while interacting with the agent. In addition, the interviews showed that audio and animation, preferably as realistic as possible, could improve the experience of talking to a real person.

2. State of the Art

To do a relevant research on the effect of realism, findings of existing literature and researches are considered. This chapter describes the findings of how ECAs can be used as lifestyle coaches, what design guidelines are researched for their visual appearance, and what is already known about the effect of visual realism on ECAs.

2.1 ECAs as lifestyle coaches

An example of an ECA can be found in Figure 1. This is an app of Sensely, that uses a virtual agent to have a "conversation" with patients to help them manage their health as they deal with a chronic congestive heart failure [15].

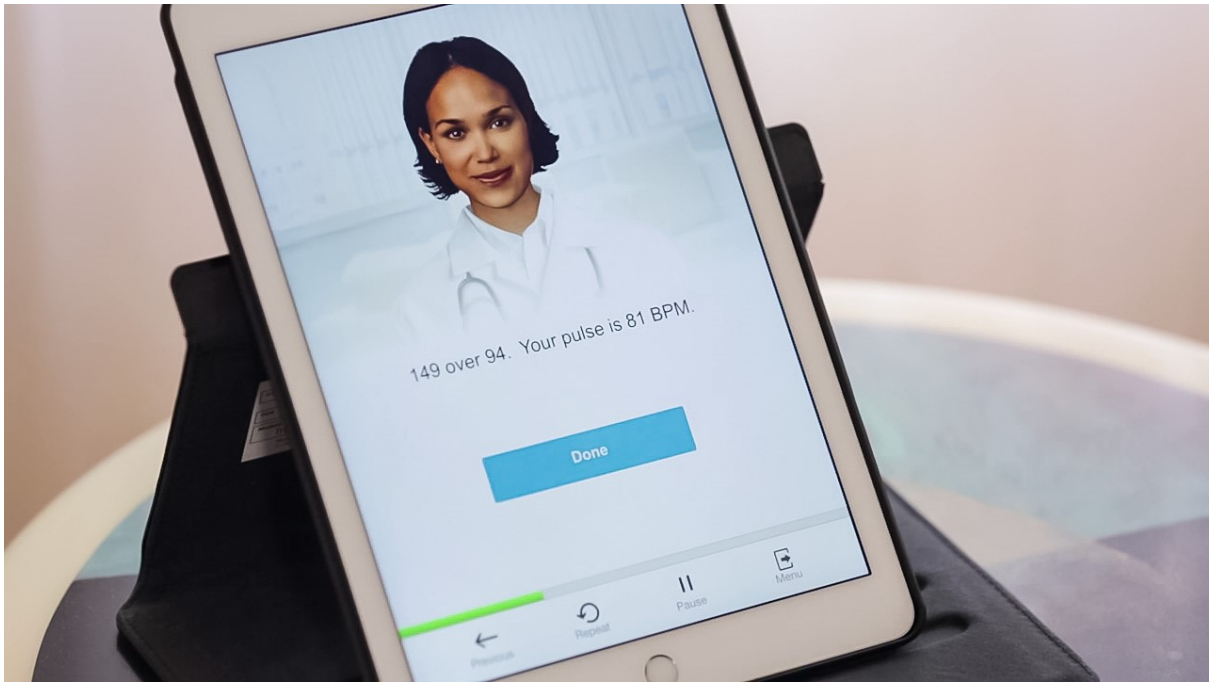


Figure 1 Example of an Embodied Conversational Agent in the context of health

To design and develop an ECA as a lifestyle coach, it is important to know that an ECA is a virtual Health Care Professional (HCP). Firstly, the competences of human HCPs to achieve behaviour change are discussed to consider how ECAs can also cover these or even overcome them. Brinkman [5] divides support for behaviour change in four human competences: monitoring, cognition, affect, and behaviour. When HCPs support these, they satisfy humans in their needs for awareness, understanding, motivation and the execution of their desired behaviour. Brandt et al. [9] analyse how HCPs perceive eHealth coaching, and find that the HCPs wanted to get to know the patient first and prefer to provide both relational communication and goal-oriented coaching. The study concludes that successful eHealth coaching requires establishing and maintaining an empathic relationship with the patient. This ties in with Moyers et al. [16] who state it is important for a HCP to be empathic and reliable.

However, in the long-term, an empathic relationship might not be necessary to maintain the gained results. Green et al. [1] conclude that gains attained as a result of a lifestyle coaching intervention can be maintained up to 30 weeks later in the absence of an ongoing weekly coaching intervention. Their research is about a cognitive-behavioural, solution-focused lifestyle coaching intervention. The techniques taught in the lifestyle coaching programme were aimed at the participant being able to continue to self-coach or peer-coach, for example, by helping the participants to determine possible routes to their goal and thereby increase pathways thinking. Green et al. therefore suggest that self-coaching techniques may increase self-regulation skills with minimal contact to regulate the process. This implies that empathy and contact is not necessary in the long-term because the gains can be maintained in the absence of this weekly coaching intervention.

To achieve behavioural change, several features of an ECA are found important. Features of an ECA specifically (not an HCP, but a virtual coach) that are considered important are the ability to maintain an empathic relationship [17], [18], their visual appearance [8], [19], [20], and the variation in how the ECA looks [8]. However, ter Stal et al. [18] conclude in their literature review that little research was performed on the agent's looks, and a consensus on design features of ECAs in eHealth is far from established.

Other relevant variables to take into account when developing an ECA as lifestyle coach, regarding to op den Akker [8], are their ease-of-use, visual appeal, clear representation and a rewarding user experience, motivation and effective coaching and tailored feedback. Added values of an ECA with respect to an HCP are the 24/7 availability [5], and that they are an effective way of human-computer interface for elderly people [17]. Furthermore, it is important to add that the overall attitude of elderly towards ECAs is positive. Van Wissen et al. [13] conclude that elderly people were positive about their interaction with the ECA and have a positive attitude towards the development of an ECA to support health-related tasks.

Concluding, when designing and developing an ECA as lifestyle coach for elderly people, more research is needed. However, literature shows that the capability of an empathic relationship and the visual appearance of an ECA has influence on behavioural change. In addition, their 24/7 availability could also be beneficial. Since little research has been done with the target group of elderly people, no valid conclusion can be drawn for them specifically, although the overall attitude of this group towards healthcare ECAs is positive.

2.2 Effect of visual appearance

Since it turns out that the visual appearance of ECAs has an important influence on behavioural change, this is further discussed in this section. In the recent literature review of ter Stal et al. [18], they conclude that little research was performed on the agent's looks. These articles are

mainly focused on realism, agent's clothing and body shape, demographics and similarity and task dependence. Section 2.3 elaborates on the effect of realism since that is the main topic of this research. The other subcategories are shortly discussed with respect to elderly people, although little research is done for this target group.

2.2.1 Clothing and body shape

First of all, Parmar et al. [21] conclude that a professionally dressed agent, dressed in a white coat and wearing a stethoscope, is preferred above a casually dressed agent when the goal is to motivate users in making healthcare decisions and feel more confident about it. It is positively rated on its characteristics (e.g. credibility, trustworthiness, reassurance, caring and friendliness), relation with the user and intention to use. Since the age of the participants was not mentioned in the article, this preference might differ for other demographic groups. Concerning the agent's body shape, literature shows mixed results. Some show a preference for attractive agents above unattractive agents (regarding the subjective interpretation of participants on the feature "attractive", often the ideal picture of their own body) [22]–[24], while other research shows a preference for non-ideal, fatter agents above ideal, slim agents [25]. However, these researches were held with participants around the age of 25, so this might not be applicable for older adults.

2.2.2 Demographics

Literature does not show a clear consensus when it comes to preference for a particular gender or age [18]. The results of Wissen et al. [13] indicate that elderly people have slight preference for a younger appearance over an older one. Next to this, Zhou et al. [26] argue that an agent having the same cultural background as the user is more positively rated on its relationship with the user compared to an agent with a different cultural background. The ages of their participants are 18-90, with a mean of 49. However, this research was about user attitudes towards a hospital virtual nurse agent, so not necessarily about behavioural change. Robertson et al. [27] suggest that, when designing an health decision aid, there is no optimal set of appearance parameters that will be appealing and acceptable to every user. Rather, they think that the ECA's demographic appearance, such as gender, age and race, should align with the target user population's demographics. Forlizzi et al. [28] conclude as a design consideration that gender is a primary design feature which should be a critical consideration in the design of ECAs. They claim that the role that gender plays in a human-computer interaction cannot be underestimated.

In the research of ter Stal et al. [19], the gender, age and role is researched in the context of eHealth, focused on the first impression of the agent image. They conclude that both a general

and elderly population seem to prefer images of young, female agents over old, male agents. Furthermore, the agent's design features affect people's first impressions of the following characteristics: friendliness, expertise, reliability, involvement and authority of the agent image and the likeliness to follow the agent's advice. These results imply for both the general and elderly population, although some differences in perception exist. Therefore, it might be beneficial to adapt the design to the user. Besides this, it is interesting to mention that their focus groups showed that elderly people believe that dynamic design features (e.g. friendliness and reliability) are more important than the agent's static design features (e.g. gender, role and age). However, it appeared that the static design features do affect the perception of the agent, since the participants preferred young females over old males. This was at first glance and shows that elderly people also have preferences for static design features.

2.2.3 Similarity and task dependence

Several studies show that the agent's demographics should be similar to the looks of the user, and that the perception of their personalities depends on the task of the agent. Zhou et al. [26] show that similarity seems to influence the perception of the characteristics of and preferences for particular agents. Next to this, studies show that preference for particular agents and perception of their personalities depend on the task of the agent [29], [30]. However, this was not yet researched with elderly people.

2.3 Effect of realism

The effect of realism of ECAs has been researched multiple times for different goals and target groups. To research this effect again in the context of a lifestyle coach for elderly people, it is useful to get insight in these researches and take their methods and results into account.

2.3.1 Visual realism

When researching the effect of realism, it is important to first find out which factors influence the perception of how realistic an image is. Fan et al. [31], [32] did two studies (with different researchers) on visual realism, where visual realism of an image is how real it appears. In their first study, in 2012 [31], they compared real photos with computer-generated (CG) images. One key finding is that their results suggest that shading may be more important than colour for judgments about visual realism, because participants discriminate between photos and CG images best for original images, less well for grayscale images and worst for reflectance images. In this study, reflectance accounts for the colour of the surface when 3D geometry effects are removed. Fan et al. also researched the differences between experts and laypersons in perceiving these images, and it turned out that details on skin texture (such as wrinkles or

freckles) may be more effective in stimulating realism perception among audiences of laypersons. Experts seem to focus more on skin glossiness.

In the research of Fan et al. in 2014 [32], they further investigated these findings. In this study, they once more find that shading is more important than colour in provoking realism perception. They also investigate the relative importance of face parts to visual realism. It turns out that firstly the eyes are the most important, then the mouth, and finally the nose. So, if wanting a more realistically perceived image, these face parts should get extra attention for their realism.

2.3.2 Effects of render style

McDonnell et al. [14] investigate the effects of render style on the perception of virtual humans, to discover whether there are differences in how truthful we perceive real and virtual humans. Does rendering style influence trust? Their starting point is the movie and game industry, although they suggest that their findings could also be beneficial for other applications, like advertising and virtual training applications. Ten render styles, ranging from abstract to realistic, were evaluated on social aspects, such as friendliness, trustworthiness, and appeal of the character. The characters can be found in Figure 2.

The results show that toon shaded (more abstract cartoon versions), and highly realistic models were best received across the different comparisons. Cartoon characters were considered highly appealing and were rated as more friendly than realistic styles. McDonnell et al. suggest these cartoon characters might therefore be more appropriate for certain virtual interactions like in rehabilitation simulations. Therefore, for the current research in lifestyle coaches, cartoon characters might work. Although not all cartoon characters might work, since one of the more abstract versions (Toon Bare) was considered quite unappealing and evoked negative reactions across most of the tested scales. Besides this, since the participants were so focused on the task, the appearance of the character did not sway them. Therefore, the study also concludes that the audio and animation contributed to the interpretation of the characters' intention, this would have even more influence than the render style.

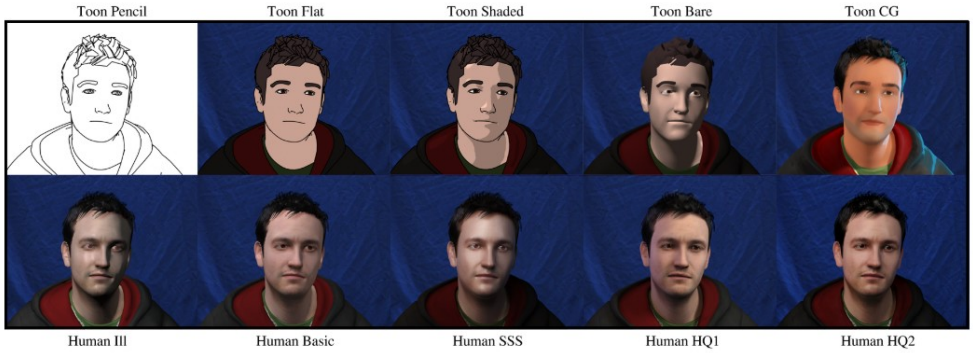


Figure 2 The ten render styles used in the experiments of McDonnell et al., ranging from abstract to realistic

2.3.3 Effects of a realistic appearance

Ring et al. [30] explored the effect of an agent's appearance and application domain on user perceptions of the agent. The used agents can be found in Figure 3. They assessed the user attitudes towards the agent on the following items: realism, appeal, familiarity, eeriness, friendliness, trustworthiness, easiness to interact with, desire to continue working with, likability, caring, appropriateness, and the quality of motion. Their results show that changes in an agent's appearance effects how users rate its friendliness, likability, caring, and appeal depending on the content of its dialogue (social or medical). The findings from this study suggest designing an agent may not be as simple as make the most realistic or cartoony agent possible. Their results suggest that in a purely medical system a highly realistic agent may be a better design, whereas for a social system, a cartoon-like agent might work better.

Figure 3 Realistic (left) and Toon Shaded (right) agent from the study of Ring et al.

In the study of Robertson et al. [27], the effect of the realism of an ECA in the context of a health decision aid is researched. They find that the most appropriate rendering style of an ECA is challenging to determine in this context. Presenting a cartoon or stylized character to appear friendly, empathetic, trustworthy, etc. may backfire due to user's sense of the seriousness of the subject. Their context was advising newly diagnosed cancer patients on treatment options and associated risks. They suggest that a stylized, cartoon ECA would be better accepted in a less

serious health context. It should be mentioned that this research did not include their designing methods and that the characters look less alike and fitting in the sequence of realistic to abstract than the characters of other researches. This might be caused by the difference in (skin) colour, expressed emotion and proportions (see Figure 4).

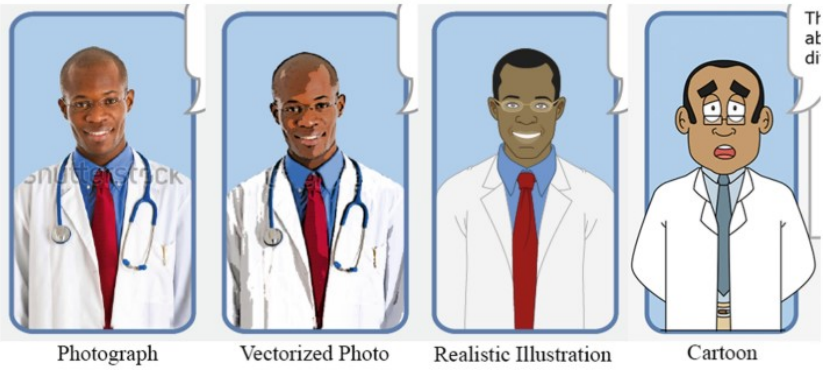


Figure 4 Rendering styles from the research of Robertson et al., ranging from photorealistic to abstract

Van Wissen et al. [13] researched the effect of similarity, familiarity and realism. Their findings show that participants preferred the realistic-looking ECA over the more stylized one. They conclude that when designing ECAs as health coaches, their appearance should be realistic in order to mimic human coaching relationships. However, it should be mentioned that the more stylized agent might not have been representable since they turned out to look (maybe unintentionally) unrealistically due to the style and colours (see Figure 5).

Figure 5 Appearances of the ECAs of the study of van Wissen et al.

2.3.4 Effects of character proportions

The proportions of characters have been explored in ways to understand and work around the uncanny valley [33]. This is the theory that when characters approach realistic similarity to humans, they stop being likable and instead become eerie, frightening, repulsive – "uncanny".

Changing the proportion of a cartoon character in respect to a realistic character can be a way to work around this effect [34]. The character's proportions should then be moved and structured outside the range of human. The character will not be perceived as human and will not judge them by the same rule as if they were. However, the uncanny valley shows to have more effect on moving characters rather than on still characters. This is partly because still characters are less familiar than moving characters, which causes a less deep uncanny-valley – a smaller gap in familiarity as near-human likenesses approach realism (see Figure 6). Therefore, this effect might not be relevant in the current study but could be considered.

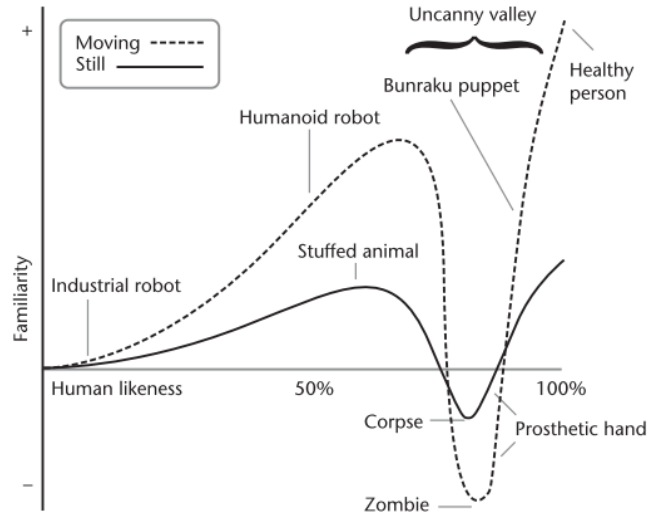


Figure 6 Masahiro Mori's graphical representation of the uncanny valley. The graph demonstrates a gap in familiarity as near-human likenesses approach realism and a significant difference between still and moving characters.

Ring et al. [30] compared a toon shaded agent with a disproportioned cartoon character (see Figure 7). They found that cartoon proportioned characters were rated as being more friendly regardless of task domain, but that more realistic characters were rated as more appropriate for medical tasks. However, this study was conducted with samples recruited from a crowdsourcing marketplace that may not generalize to any particular user demographic. Therefore, this does not have to apply for elderly people.

Figure 7 Human (left) and Cartoon (right) proportioned agents from the study of Ring et al.

2.4 Conclusion

Concluding the results of this literature research showed what to consider in the current study. Firstly, for designing an ECA as lifestyle coach for elderly people, it is important that they are (1) capable to establish and maintain an empathic relationship, that (2) there is attention for their visual appeal and that (3) they are easy to use.

Subsequently, the effect of their appearance was researched. The results show that (1) a professionally dressed agent might be preferred over a casually dressed agent, (2) demographics such as gender, age and race are important, but there is no clear consensus on this, (3) similar demographics as the user might be preferred, and (4) the perception of the personality of the agent is dependent on the task.

When wanting to achieve visual realism, studies [31], [32] show that shading is more important than colour, details on skin texture are important and the eyes play relatively the most important role of face realism of the face parts. After the eyes, the mouth and then the nose are the most important. The studies about the effect of realism showed that (1) toon shaded and highly realistic characters are received best on social aspects, (2) in a purely medical system a highly realistic agent may be a better design, whereas for a social system, a cartoon-like agent might work better. Furthermore, (3) stylized, cartoon ECA could backfire in a serious context, but might be better accepted in a less serious health context. And lastly the studies show that (4) cartoon proportioned characters were rated as being more friendly regardless of task domain, but that more realistic characters were rated as more appropriate for medical tasks.

Overall, the literature research shows that there is no clear consensus about the appearance, only that it is an important aspect in designing ECAs. The literature does mention some aspects to consider when designing an ECA, but not specifically for the target group of elderly people and the context of lifestyle coaching. However, it does show that it is a challenge to only research visual realism without the influence of other variables.

3. Research method

This chapter gives an overview on how the study was done and what to expect in this paper. In order to research the effect of realism of a lifestyle ECA on elderly people, iteration by co-design was used. This method was chosen to have a more elaborate prior research to make sure the only variable that is tested, is the visual realism. A scheme of this overall method can be found in Figure 8. Firstly, a literature research was done (see 2. State of the Art) on the possibilities for ECAs in the lifestyle coaching of elderly people and on the results of similar studies on ECAs and the effect of their realism. Subsequently, via co-design sessions an app interface with different agents was created. The first co-design resulted in the choice of the base agent which was used to make the three agents. During the other two co-design sessions, these agents and the app interface with the dialog were evaluated. These outcomes (three short interactions with the app with the three different agents) were connected to a questionnaire which was filled in by 64 elderly persons, this is the research phase. Lastly, these results were evaluated with respect to the perception of the personality of the agents. Findings from the literature, the co-design sessions and the research were combined into a discussion with recommendations for designing and developing a virtual lifestyle coach for elderly people, mainly with respect to visual realism.

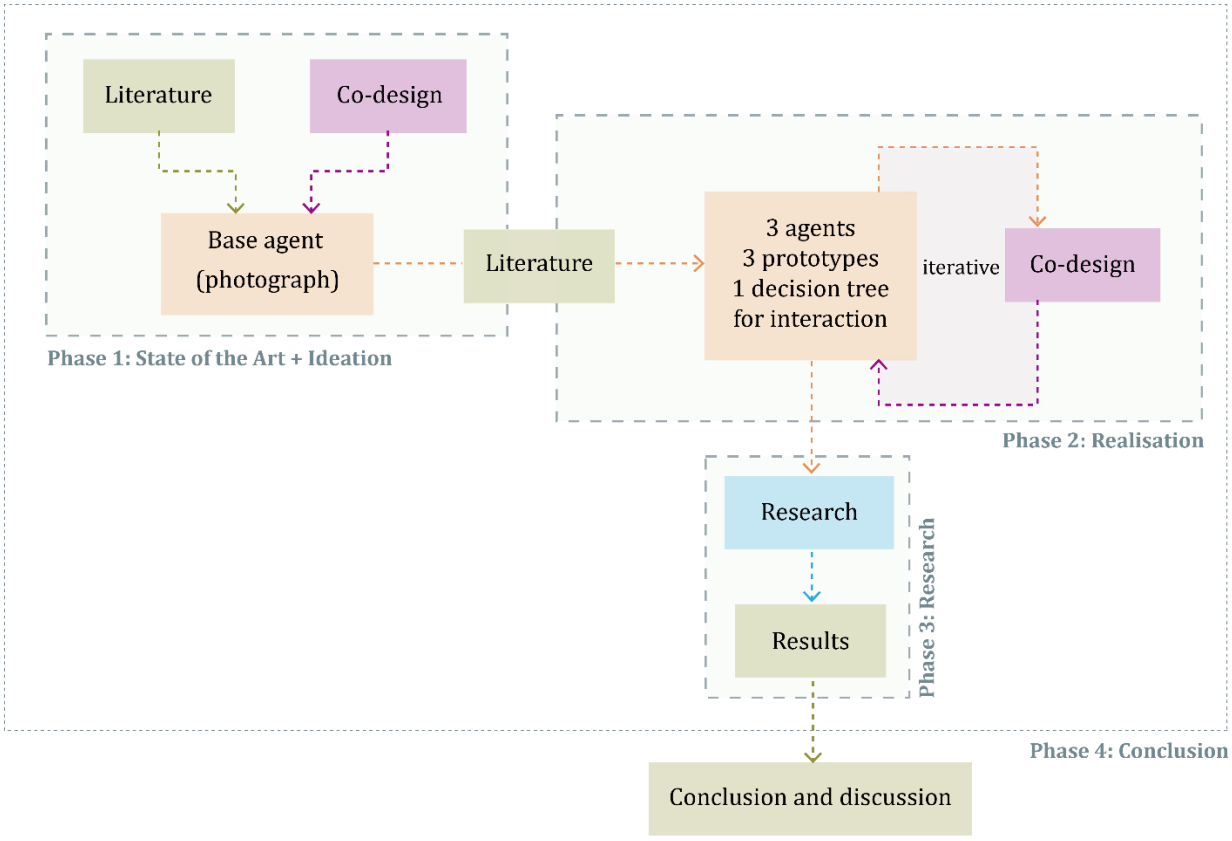


Figure 8 Overall method used in this research

3.1 Co-design

One-to-one co-design sessions were used to be sure the three tested agents were already partly approved and there are lesser biases that play a role. The co-design was in collaboration with five elderly persons. A base agent that was the most suitable considering the first co-design session and literature was chosen and the cartoon characters that did not go down well with elderly people were eliminated or adapted before the larger-scaled research. In this way, these limitations were minimized. These sessions resulted in a user interface of a mock-up phone application with a small interaction with a dialog. This interface was made in three forms, all with their own variant of an agent, differing in realism. The consent form that the participants signed can be found in Appendix A. By signing this form, they gave permission to record their audio and use their results and data for this study.

3.2 Research on the effect of visual realism

The user interfaces that resulted from the co-design sessions were added to a prototype and connected to a questionnaire. To research the effect of visual realism, quantitative and qualitative data was collected. A link that randomly leaded to one of the three prototypes was spread online and filled in by elderly people (n=64). Ten elderly participants agreed to an interview to continue to discuss their answers on the questionnaire. The results of these questionnaires gave insights in the effect of the visual realism of the ECA. The consent form that the ten participants signed can be found in Appendix A. By signing this form, they gave permission to record their audio and use their results and data for this study.

4. Ideation and requirements capture

In this chapter, preliminary research is described, largely based on findings from the literature study, in order to determine the more precise requirements for the tested ECAs.

4.1 Co-design session 1: choosing a photograph / base agent

4.1.1 Goal

Since the literature has no consensus about age, gender, race or role, the goal of the first co-design session was to select a photo that is suitable as a base agent for the remainder of this study. This base agent must fit the target group and topic. The target group needs a basis in the perception of helpful characteristics, such as trust or expertise [14], [19], [30]. Thus, the goal of this session was to get insight in the characteristics the participants find important and which person they prefer as a lifestyle coach.

4.1.2 Method

4.1.2.1 Participants

Nielsen et al. [35] show that for a valuable usability test, no more than five participants are needed, therefore, the sample consisted of 5 elderly people (3 males, 2 females; M_{age} = 66.8, SD = 11,53, ranging from 58 to 74 years), all the same race (Caucasian) and still living independently in their own homes in the Netherlands.

4.1.2.2 Materials

The photos of the ECAs which the participants could choose from, were selected from a website with free stock photos (freepik.com) and placed in Miro, an online software for UI design. The images' backgrounds were removed manually. The pictures were a mix of age (young/middle-aged/old), gender (male/female) and role (casual/doctor). The race was kept the same, and since the participants also have this same race, this should not influence the result [26], [27]. In Figure 9, the photos can be seen. For the session, they were mixed randomly. The setting for the session can be found in Figure 10. The photos were cropped just below the shoulders, since this composes a pleasing photograph [36]. Besides this, it avoids the problem that body poses need to be taken into account. To distinguish the photos, they are coded with 3 letters: M/F, Y/M/O, C/D. For example: the photo in the left top corner is Male, Young and Casual, so this photo is coded as MYC.

4.1.2.3 Procedure

After some context was offered in the form of an explanation of a possible interaction and the type of agent, the participants had to put the photos in order from most to least preferred.

Subsequently, a structured interview was held, to get to know their reasoning and why they chose this specific order. The structure of the interview can be found in Appendix B. The interview itself was via a phone call, which was recorded. The participants could place the photos in order with Miro, this process was screen recorded.

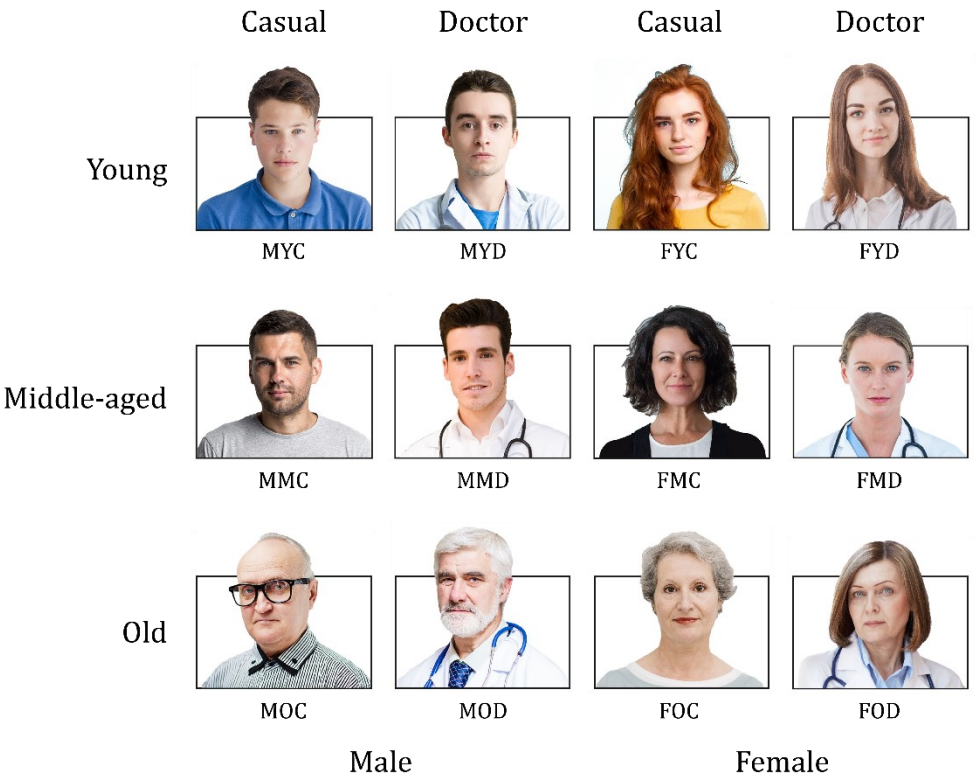


Figure 9 Photos that were selected to choose the base agent with the corresponding code

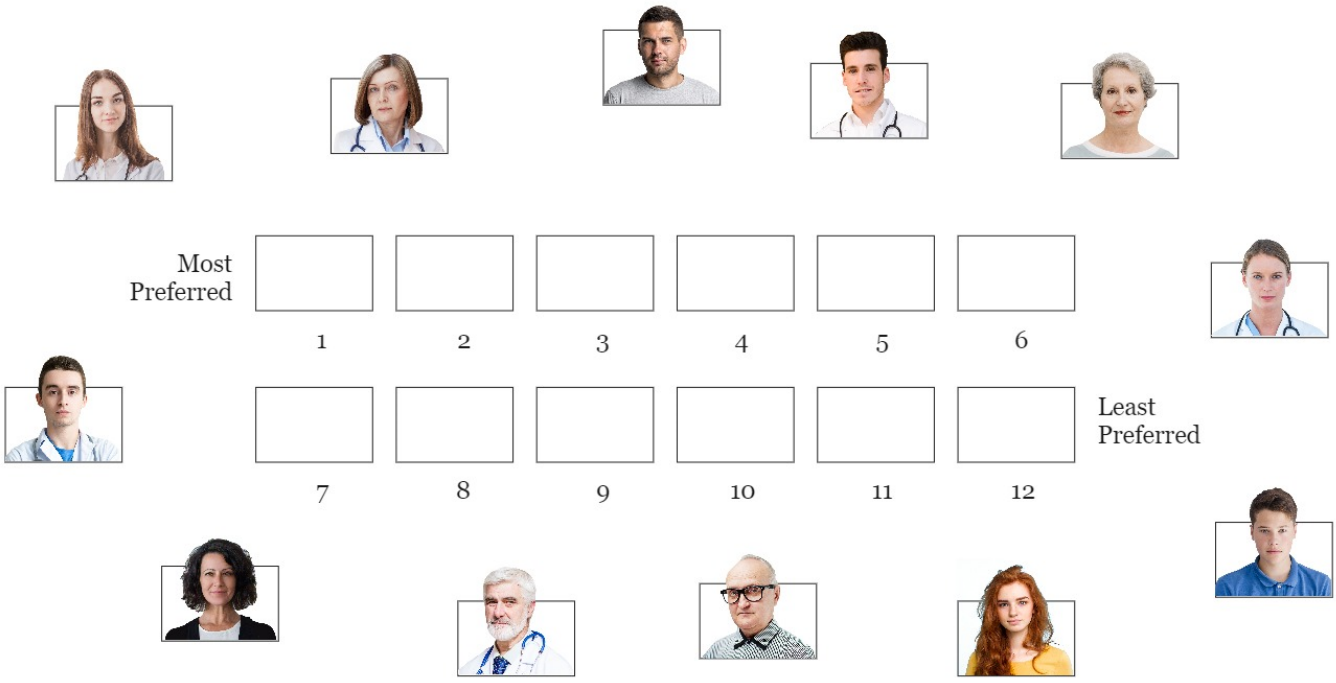


Figure 10 Setting in which the participants can make an order from most preferred to least preferred

4.1.3 Results co-design session 1

The goal of this session was to get insight in the characteristics the participants find important and which person they prefer as a lifestyle coach, and then choosing a photo to use as the base agent for the further research. The order of the photos that the participants ended up with can be found in Appendix C.

First, it is important to enlist the different characteristics participants came up with when they were asked to describe (in their own opinion) good characteristics of a lifestyle coach. The characteristics that came up most are showed in Table 1. Most of the participants said that a good lifestyle coach should be reliable, knowledgeable, communicative, serious, friendly and emphatic. E.g. participant B said: "He has to have knowledge; he has to be able to guide you in what you need to do". Furthermore, participants thought a good lifestyle coach should be balanced, patient, have authority, should look healthy him/herself. He/she should be sporty, honest, clear, competent, compassionate, open, calm and listening.

Table 1 Results characteristics a good lifestyle coach should have, as named by the participants

Characteristic / Participant	*A*	*B*	*C*	*D*	*E*	*Total*
Reliable / Trustworthy	X		X	X	X	*4*
Knowledgeable (expert)	X	X	X	X		*4*
Communicative	X		X		X	*3*
Friendly	X	X	X			*3*
Empathetic/Sympathetic		X	X	X		*3*
Serious	X			X		*2*

With these characteristics in mind, the orders in which they placed the photos are discussed below. What stands out is that participant E (female, 74 years old) did not fill in the order. She did not have a preference. "This doesn't mean anything to me, these are faces. I would not know why I would prefer someone; I think they all look friendly and reliable on first sight. I really wouldn't know a preference." She thought first impressions matter, but not on these pictures, but more in real life. In the results, her ranking was left out since she had equal preference for all photos.

The other participants had preferences, although they differed from each other. In Table 2, the division of ranking numbers per participant is laid out. The mean and standard deviations were calculated to get insight in the preferred photo and how these ranking numbers are spread

out from this average. The Male, Middle-aged Doctor scored the best together with the Male, Middle-aged Casual. The Male, Old Casual is in last place. Participant A mentioned that his top 4 was interchangeable, which implies that the other 3 following his number 1 (MMC, FMC and MMD) could also be on number 1.

Table 2 Ranking numbers per photo per participant (1 is the highest ranking, 12 the lowest), sorted on mean and standard deviation

Photo / Participant	*A*	*B*	*C*	*D*	*Mean*	*SD*
 **MMD**	**4**	**3**	**4**	**4**	**3,75**	**0,43**
 MMC	2	1	9	3	3,75	3,11
 FMD	1	5	8	2	4	2,74
 FYD	8	6	3	6	5,75	1,79
 FOD	5	10	1	7	5,75	3,27
 FMC	3	2	7	11	5,75	3,56
 FOC	6	7	2	9	6	2,55
 MOD	7	11	6	1	6,25	3,56
 MYD	10	8	5	5	7	2,12
 FYC	9	4	12	12	9,25	3,27
 MYC	11	9	11	10	10,25	0,83
 MOC	12	12	10	8	10,5	1,66

The first thing that stands out is that all casual versions in the table are lower than the ones with the same age group and gender but are dressed in a white coat and wear a stethoscope. This was confirmed by some of the participants. Participant C quoted: “Yeah, it does matter, that doctor's coat. It gives more confidence; he studied for it. I realise it is kind of old-fashioned, because I know it's does not necessarily mean that he has, but that's the way it works for me.”

The young, female versions were ranked higher than the old, male versions of the same role. But the middle-aged all score higher than the old or young versions. The results also show that in general, the middle-aged people are preferred above young or old, with a slight preference for old above young. Although they dislike an agent who is too young; participant A mentioned that “age does exude some experience”, which was also mentioned and thus confirmed by other participants.

Gender does not seem to have a big general effect, although it seems that the photos of women have a slight preference, since ranking 3 up until 7 are all female although the two photos at the top are men. Two participants (C (female) and D (male)) mentioned they preferred someone their own gender. They also preferred someone similar to themselves in age, they both put the old, doctor version of their own gender on the first place. The other two participants (both male) claimed not to prefer a specific gender.

Lastly, when looking at the characteristics that participants found important for a good lifestyle coach, reliability and expertise were mentioned the most. Followed up by good communication skills, friendliness, and empathy/sympathy.

4.1.4 Discussion

The goal of this session was to get insight in the characteristics the participants find important and which characters they prefer as a lifestyle coach, and then choosing a photo to use as the base agent for the further research. The preferences of the different demographics are shortly laid out and compared with literature from Chapter 2.

The results suggest that doctors are preferred over casual versions / laypersons. This is in line with Parmar et al. [21] and was confirmed by some of the participants. The white-coat effect is also a psychological principle, where the white coat indicates authority, friendliness and trust [37]. Besides this, the young, female versions were ranked higher than the old, male versions of the same role, which confirms the conclusion of ter Stal et al. [19] that both a general and elderly population seem to prefer images of young, female agents over old, male agents. However, it is neither convincing nor decisive, since the middle-aged all score higher than the old or young versions. The results show that in general, the middle-aged people are preferred above young or old, with a slight preference for old above young. This outcome is in line with Wissen et al. [13], who suggest that elderly people have slight preference for a younger

appearance over an older one. However, participants thought that age exudes experience, which made some agents too young.

Gender does not seem to have a big general effect, although it seems that the photos of women have a slight preference. Some participants mentioned to prefer similarity to themselves. This is line with Robertson et al. [27], who find that the ECA's demographic appearance, such as gender, age and race, should align with the target user population's demographics. However, the other two participants claimed not to prefer a specific gender.

Lastly, when looking at the characteristics that participants found important for a good lifestyle coach, reliability and expertise were mentioned the most. That these characteristics are preferred is in line with many of the aforementioned researches (see Chapter 2).

4.1.5 Conclusion

Concluding this first co-design session, it turns out that doctors are preferred over casual people / laypersons, middle-aged people are preferred over old or young, and gender does not seem to make a big difference. It is important for the lifestyle coach to be reliable, have expertise and good communication skills, be friendly and have empathy and sympathy. When considering the results and the discussion, we conclude that MMD (Male, Middle-aged, Doctor) overall scored best. This photo scored the highest average together with MMC, although when looking at the standard deviation, it turns out that with MMD, the preferences are less spread out. This implies that participants agreed that this man had a good base level of characteristics, he did not stand out for anyone, but was in the top 4 of all participants. Participant A even said that his top 4 was interchangeable, which implies MMD could also be his number 1. Therefore, the photo of MMD, shown in Figure 11, was taken as a starting point for the rest of this research because it overall scored best.

Figure 11 Male, Middle-aged, Doctor; overall best-scored photo

4.2 Design mock-ups

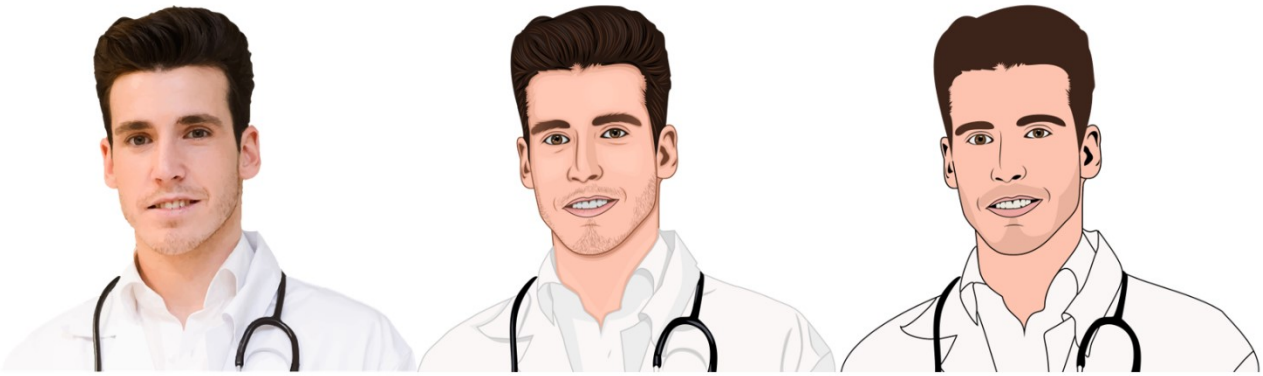
Based upon the previously selected base picture, three variations were designed, each with a different level of visual realism. We describe the process of creating these variations, and the rationale for the various design choices. Next, we discuss the design of the mock-up User Interface (UI).

4.2.1 Method agents

All the agents were made with Adobe Illustrator. The photorealistic one was made with vectorizing it automatically, and the other two were designed manually. When illustrating, the visual realism techniques of Fan et al. [31], [32] were considered. To achieve the perception of realism the shading seems to be important. Since the research is with laypersons, their study, which makes a difference between laypersons and experts, tells us face details such as wrinkles are also important. The eyes, mouth and nose need extra attention since they play a big role in visual realism. To make the cartoon, you do not want to stimulate too much visual realism, so the shading and details like the wrinkles were removed.

4.2.2 Results three versions of agents

The results of the designing of the agents can be found in Figure 12. These were used for the remainder of this study.



Photorealistic *Realistic illustration* *Cartoon*

Figure 12 First designs of agents

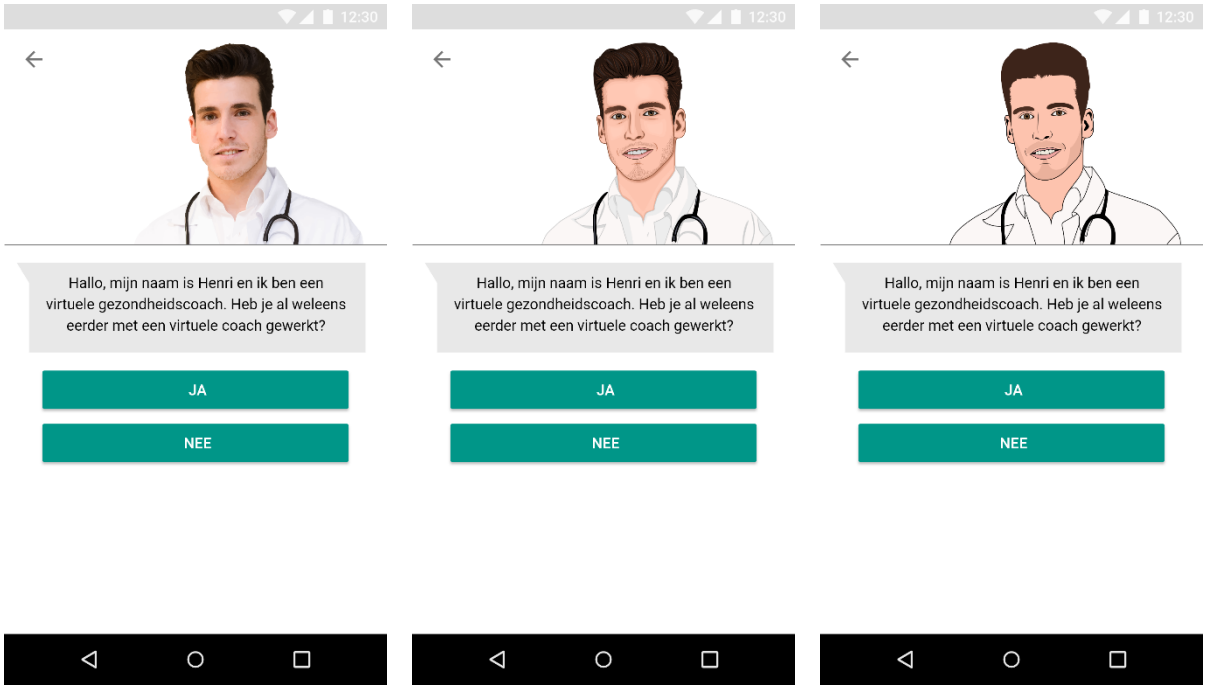
4.2.3 The UI design

To design a context in which users can understand the use of the agent, a User Interface (UI) was made. This design was made with Adobe XD and based on example scenarios received from the Roessingh Research & Development. The interface consists of a couple of screens of a mobile app through which users can navigate. The feeling of them using a real app can be simulated by these prototypes. One prototype was for the user evaluation, used with varying agent selections,

but for the rest the same interaction options, to keep most variables the same. The scenario that was used for the prototype can be found in Appendix D and was based on a neutral textual expression.

4.2.4 The resulting UI

The results of the designing of the prototypes can be found in Figure 13. These designs were improved, mainly in the sense of the appearance and usability, during the second and third co-design sessions with the feedback of the participants. All the screens used in this session can be found in Appendix E.



Photorealistic *Realistic illustration* *Cartoon*

Figure 13 First screens of first designs of the three prototypes

5. Realisation of the experimental set-up

In this chapter we describe the process from the design mock-ups from the last section until the final design. This final design consists of the three variants of the mock-up phone application, all with their own variant of the agent, which differs in visual realism.

5.1 Co-design session 2: feedback round agents and UI

5.1.1 Goal

Since a few studies about the effect of realism, like in the study of Robertson [27], discuss that the designs of their agents could have caused limitations, such as that it was not reliable due to a lack of similarity of the agents. We wanted to prevent similar problems with our own experimental set-up. The guiding principle was that we wanted to minimize the number of variables that potentially affect the judgement of ECAs by test subjects. What we really cared for was the level of realism as the variable of interest, to the exclusion of all other variables. Therefore, the goal of this second co-design session was to get feedback on the design mock-ups from the last section. Thereby, it was useful to already get insight in their preferences and how the agents look to them, so besides discussing the mock-ups, these questions were in the structured interview.

5.1.2 Method

5.1.2.1 Participants

The group of participants was the same as in Chapter 4.1.2, the group of five elderly people who joined the first co-design session. The participants are coded the same as in co-session 1. Thus, participant A in session 1 is the same person as participant A in session 2.

5.1.2.2 Materials

The materials needed for this session were the design mock-ups, these can be found in Figure 12 and 13. These are the first designs of the agents and prototype.

5.1.2.3 Procedure

A structured interview was held with the individual participants. The outline for this interview can be found in Appendix F and consisted of topics such as what they notice about the images, how the pictures looked to them, what qualities or characteristics they thought they had. They were also asked whether there were things that bothered them and which picture they preferred. The participants were called via Zoom, only the audio was recorded. During the call, the different images were shown, first individually (in the order: cartoon – photorealistic - realistic illustration) and then the three variations were shown next to each other. They were

asked questions about how the agent looks to them and who they prefer. A link to the UI was sent to them via mail, during the interview. They could open this on their phone and try it out. Then a short discussion followed where they could give feedback on it and on the characters. Since we are developing a mobile app, it is important that they can see and understand the images and app clearly on their mobile phone.

5.1.3 Results co-design session 2

The goal of this session was to get insight in the preferences of the co-design group and discussing the mock-ups with them. The results are discussed per subject, first the agents and then the user interface.

The cartoon was the first agent that was shown. The participants were asked what characteristics were fitting for this person, and they answered with similar characteristics as in the first co-design session. They found the agent trustworthy (4x), friendly (3x), young (3x) (participant D mentioned that the agent was too young in his opinion), open (2x), sympathetic (2x), knowledgeable, communicative, listening, and authoritative. Participant D also mentioned that the eyebrows were "heavy", although this was not negative, since it expressed authority. Furthermore, participants did not have comments on what stood out or looked disturbing.

After this, the photorealistic agent was shown. Although we did not intend this, three participants mentioned that the person on the photo looked a bit younger than the cartoon. Furthermore, the two female participants did not like the beard, participant C said: "I do not like such a beard, it looks unkempt.", she found that it was not fitting for a doctor. However, participant D, male, also mentioned the beard, but said it added something casual and that it suited him. Other comments about this photorealistic version were that he has friendly eyes, that the eyebrows looked different (not necessarily positive or negative) and that he looked more like a real human, less rigid. Participant B added that he looked more insecure than the cartoon, which made him somewhat less trustworthy. He thought this could have to do with the look of his mouth. Furthermore, the participants agreed that he looked like the same person and had mostly the same characteristics.

The last agent they got to see was the realistic illustration (the middle one of Figure 12). Participant A thought he had a similar appearance, participant B thought he was a mix of the photorealistic agent and the cartoon. He still found that this agent looked somewhat insecure, less than the photorealistic agent, but more than the cartoon. Participant C found that he has the same friendly eyes and smile, but that the beard looked a bit artificial. Participant D agreed to this finding. He added that he preferred this one the least of the three, although it did not make that much difference to him. Participant E mentioned that she found the eyes of the realistic illustration clearer, less dark.

When the participants saw the three agents next to each other, differences stood out more and the preferences of the participants differ. This can be seen in Table 3. It is important to look at some of their reasons for their preferences. Participant A and D preferred the photorealistic agent. Participant A said that "he looks best, the drawings are still drawings". Participant D mentioned that the other ones looked more like cartoon characters. And although he likes cartoons, he does not take them seriously: "They have to be humoristic, not trustworthy. I would not assume that they would tell me the truth". However, two other participants did prefer the cartoon agent. Participant B preferred him because it was sharper and clearer to him. He mentioned that the cartoon agent seems more secure and therefore more trustworthy. Participant C mainly preferred the cartoon over the others because of his unkempt beard and she liked the jacket of the cartoon the best. Participant E preferred the realistic illustration, she thought this one looked the most realistic/human-like, and that the eyes were clearer and less deep into the face. All participants mentioned that the differences were not very big, and that most of the characteristics counted for all three agents.

Table 3 Preferred agents per participant and total amount per agent

Preferred agent / Participant	*A*	*B*	*C*	*D*	*E*	*Total*
Photorealistic	X			X		*2*
Realistic Illustration					X	*1*
Cartoon		X	X			*2*

After presenting the agents, the User Interface itself was discussed. The main findings from this co-design to improve the user interface and experience were that (1) the language should be simple and clear, (2) the "OKÉ!" should change to "OK" to make it more readable and be placed higher so that scrolling would not be necessary, (3) multiple choice is good, but it would be nice to have the option to type for specific questions, and (4) it should be speech language, not written language. Participant E did not see the differences between the agents in the app, she is 74 years old, so the images might have been too small for her to distinguish. Furthermore, the buttons were clear, and it was clear that the text of the agent was in the speech bubble. Participant B mentioned that he liked the pertinent appearance of the app, just simple and clear.

5.1.4 Conclusion and discussion

The goal of this session was to get insight in the preferences of the co-design group and discussing the mock-ups with them. The preferences differed between the participants.

However, the sequence should be considered. Since the same sequence of cartoon-photorealistic-realistic illustration was chosen, this could have influenced or biased the results. Even as the fact that two weeks before, during the previous session, the participants all saw the photo. Overall, the participants had no outstanding disturbing or outstanding things about the agents, except for the beard, although opinions about the beard were also divided. Since several participants mentioned the language of the app, this is a point of attention to make a simple and clear dialog.

5.2 Co-design session 3: feedback on dialog

5.2.1 Goal

During co-design session 2 we discovered that a simple and clear dialog has an important influence on the experience of users with the coach, and therefore it might also have an important influence on their perception of the agents. Therefore, the goal of this third co-design session was to conclude with a simple and clear dialog within an interactive prototype to use during the user evaluation.

5.2.2 Method

5.2.2.1 Participants

The group of participants was the same as in Chapter 4.1.2 and 5.1.2, the group of five elderly people who joined the previous co-design sessions. The participants are coded the same as in co-session 1 and 2. Thus, participant A in session 1 and 2 is the same person as participant A in session 3.

5.2.2.2 Materials

The materials used for this session was the prototype. The screens were this time designed and prototyped in Figma. This online UI tool was chosen over Adobe XD because in Figma, you can add an external link, which is needed to forward to a questionnaire. The dialog used in the previous co-design session was simplified and extended. Since the participants agreed that multiple choice questions were practical, a decision tree was created with the help of a previous workshop of two lifestyle coaches [38]. This decision tree, a flowchart-style diagram, can be found in Appendix G and was used to make the screens of the prototypes. App.diagrams.net, an online diagram software, was used to make this flowchart. Subsequently, 15 screens were created for this session, the photograph was chosen as the agent. Since this session is about the interaction and dialogue, this agent was chosen randomly.

5.2.2.3 Procedure

Another structured interview was held with the participants. The outline for this interview can be found in Appendix H. The participants were called via Zoom, only the audio was recorded. The link to the prototype was sent to them and they were asked to click through. Then they could give their feedback on the formulation, text, and understandability of the interface. The outline of the structured interview was used as a guideline, but the goal was to get feedback. After every participant, the prototype was iterated. The feedback of the participant was processed directly so it could be iterated multiple times and the feedback was not all about the same prototype. In the results, these iterations are described.

5.2.3 Results co-design session 3

The exact changes made according to the feedback of the participants can be found in Appendix I. The first participant, participant A, mentioned that the sentences with an enumeration were not clear. It would be better to make a real enumeration by adding a bulleted list. This was implemented in the dialog about general tips. Besides this, he recommended to review the tips, since some of them were not clear or helpful. Furthermore, he added some textual changes and advised to remove the screen where you could get a tip for choosing between nutrition and physical activity. After this feedback, the text on this screen was merged with the previous screen.

Participant B added some textual changes and mentioned that he needed his reading glasses to read the text and that he would prefer to have the text one size larger. Participant C and E agreed that the prototype was clear and that the short sentences made it easy to read. This is in line with participant D, only he added some extra textual changes to make sure all sentences were easy to read. Furthermore, participant E mentioned that she recognized these types of questions and that this way of asking questions works well for her.

There was a short discussion about using “u” or “je”. This is about addressing someone formal (“u”) or informal (“je”) in Dutch. Participant C thought that “u” might have the preference for some elderly people. However, she and three other participants preferred “je” since this is the more informal way and felt less distanced. Participant D said: “When using “je”, I feel that I am more personally addressed.”

5.2.4 Conclusion and discussion

Overall, the participants were satisfied with the prototype and agreed that the dialog is clear and understandable. Participant B mentioned the font size was too small and Sakdulyatham et al. [39] conclude that the most appropriate font for elderly people (their participants were 60-69 years old) has the size of 16 pt. Therefore, the font size is changed from 14 pt. to 16 pt. The

textual changes of all participants were considered and changed in the prototype. The way how the user is addressed is still “je”, since most participants agreed that this way of addressing has their preference.

The goal of this third co-design session was to conclude with a simple and clear dialog within an interactive prototype to use during the user evaluation. With the help of the participants, the prototype was iterated and used for user testing.

5.3 Final mock-up

Concluding the iterative co-design method, a final mock-up in the shape of an interactive prototype was realised. A few screens can be found in Figures 14-17, the final decision tree can be found in Appendix J. The prototype consists of fifteen screens, three versions of this prototype were made, all with another version of the agent.

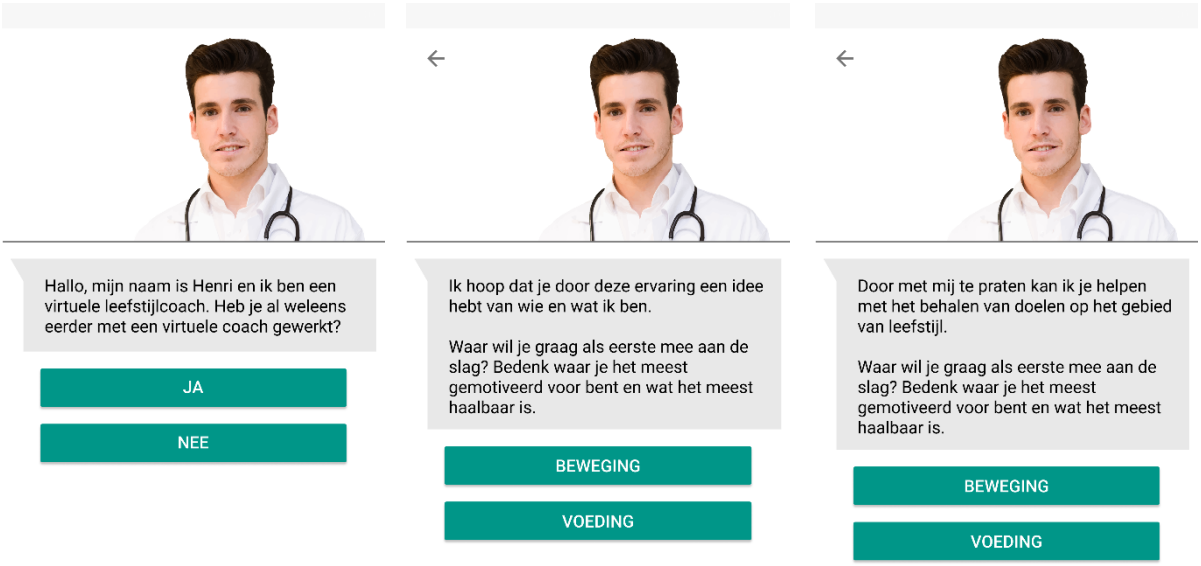


Figure 14 Screens version 1 (photorealistic version)

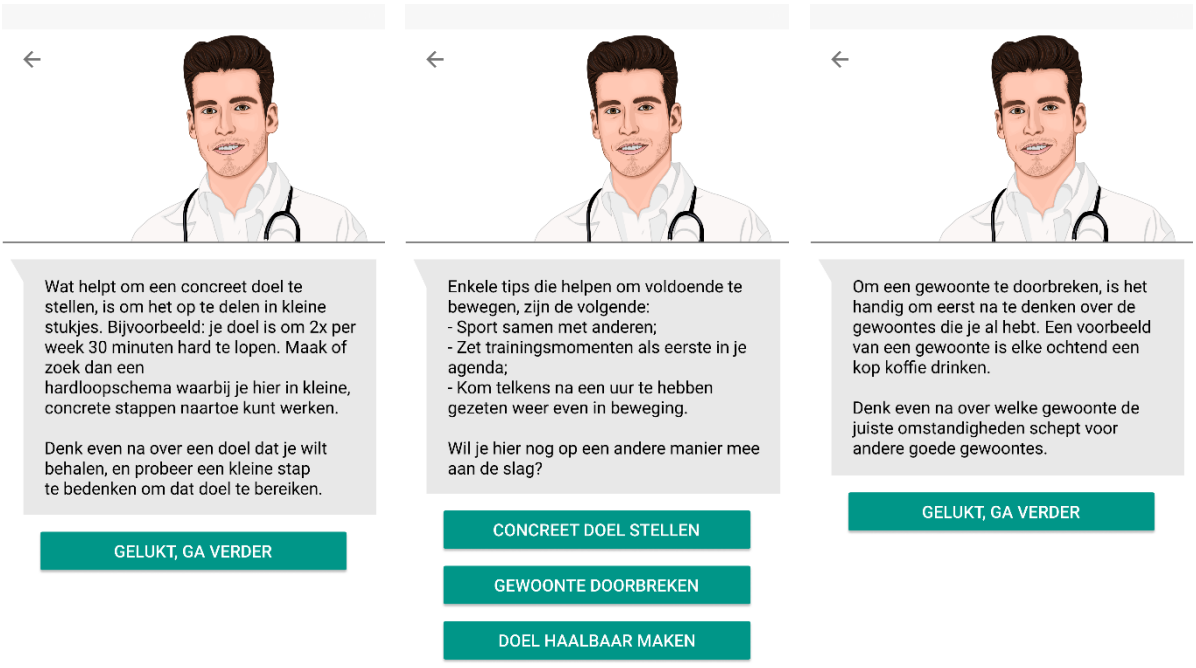


Figure 15 Screens version 2 (realistic illustration)

Figure 16 Screens version 3 (Cartoon)

Figure 17 End screen for all prototypes, user gets forwarded to questionnaire

6. Research on the effect of visual realism

6.1 Goal

The goal of this study is to get insight in the effect of visual realism on the perception of the personality of an ECA in the form of a lifestyle coach for elderly people. Therefore, the final mock-up was tested on users, this process and the results are described in this chapter.

6.2 Method main research

6.2.1 Scheme

Figure 18 shows the scheme of the method of the user testing and evaluation. This is the logistics of the upcoming research.

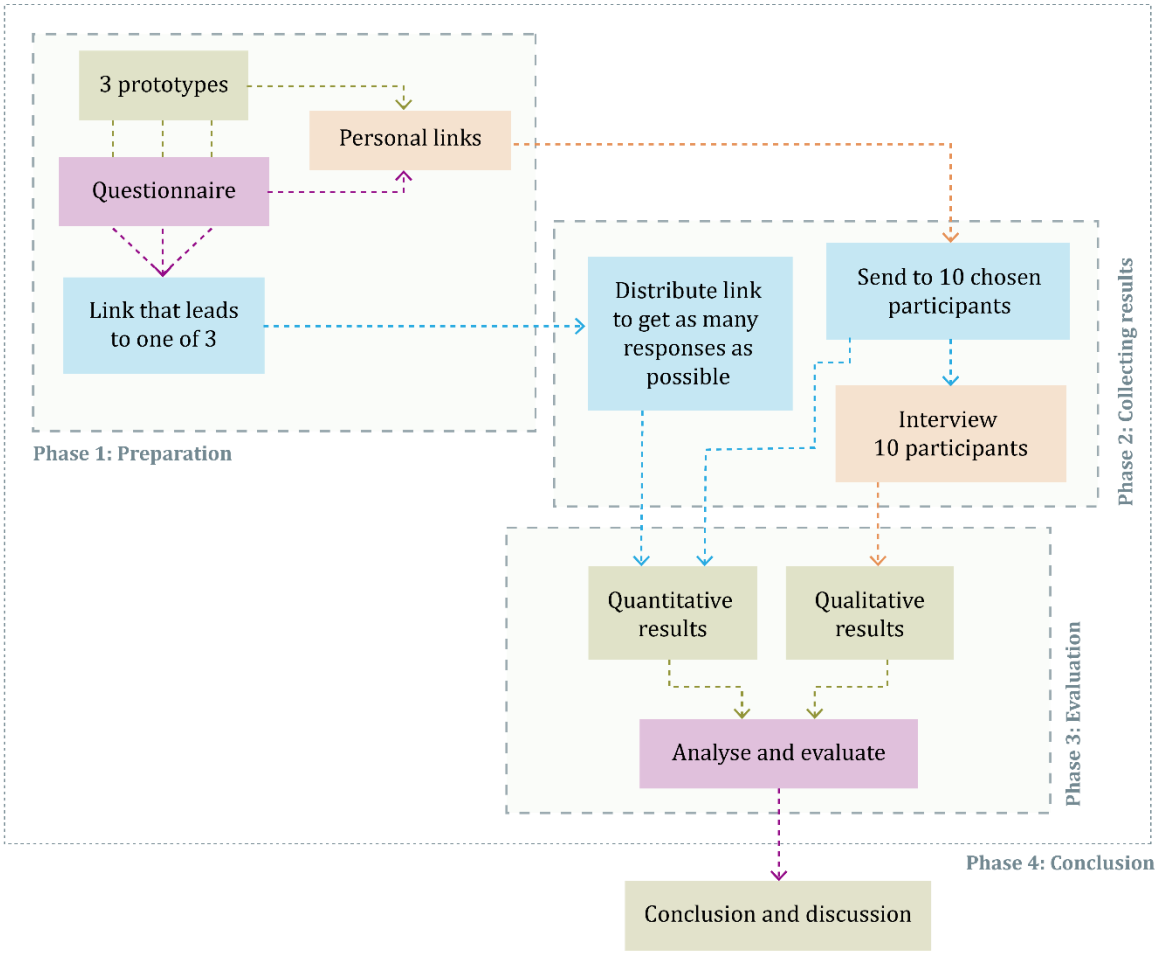


Figure 18 Scheme of method of research on the effect of visual realism

6.2.2 Participants

The sample consisted of 64 elderly people (27 males, 37 females; M_{age} = 63, SD = 6.0, ranging from 55 to 75 years). 10 participants of this group were invited in advance, with attention to a distribution of their gender and age, to further discuss their answers in an interview. This

sample consisted of 10 elderly people (4 males, 6 females; M_{age} = 62.3, SD = 5.2, ranging from 55 to 70 years). The details of the demographics per version and in total can be found in Table 4.

Table 4 Demographic information participants

				n	Mean	Min	Max	SD
Version	Photo	Gender	Female	13				
			Male	9				
			Total	22				
		Age			*62*	*55*	*73*	*6*
	Real. Ill.	Gender	Female	13				
			Male	9				
			Total	22				
		Age			*64*	*55*	*75*	*6*
	Cartoon	Gender	Female	11				
			Male	9				
			Total	20				
		Age			*63*	*55*	*74*	*6*
	Total	Gender	Female	37				
			Male	27				
			Total	64				
		Age			*63*	*55*	*75*	*6*

6.2.3 Materials

For this study, participants could interact with the prototype (final mock-up) by having a short dialog where they could answer themselves. After this, they could fill in a questionnaire. The questionnaire consisted of questions such as how the coach came across. The characteristics that were tested, were the characteristics that were mentioned two or more times during the first co-design sessions in section 4.1. The participants named characteristics they thought that fit a good lifestyle coach. The results can be found in Table 1 on page 21. Most questions of the questionnaire have a 5-point Likert scale, which is a type of psychometric response scale in which the respondents can specify their level of agreement in five points [40]. The participants could also choose between the three versions and explain why. Furthermore, they could leave any other comments or improvements for the app. The questionnaire was made with Qualtrics and its questions can be found in Appendix K.

The final mock-ups were made in Figma, an interface design tool, and for the ten participants that were interview afterwards. In the last screen of the mock-up, the participants got redirected to the questionnaire. To make sure participants did not know there were more versions of the prototype, a link was created that redirects randomly to one of the three

prototypes. To distinguish the answers on each prototype, they redirected to three different questionnaires with the same questions.

6.2.4 Procedure

The link that leads randomly, double-blind with a short segment of programming code, to one of the three prototypes was distributed via social media to get as many responses as possible. Personal links of the questionnaires were generated for the ten participants, to distinguish their responses from the other participants, since the others are anonymous. After all participants filled in the questionnaires, ten structured interviews were held. In these interviews, the participants elaborated their answers on some of the questions. The goal of these interview was to get more insight and qualitative data on why people thought the coach came across in a certain way, why certain characteristics fit him. Besides this, it was meant to get insight in whether people prefer another version when they know this is a possibility and hear them explain why.

The questionnaire data was exported from Qualtrics to Excel. The Excel file was imported into SPSS 26 statistics program to perform statistical analysis. All tests were performed using a 95% Confidence Interval.

6.3 Results

The main goal of this research is to get insight in the effect of visual realism on the perception of the personality of the ECA in the form of a lifestyle coach for elderly people. The data tables with the answers of the participants on the multiple-choice questions can be found in Appendix L.

6.3.1 Characteristics

To compare the versions on the perceived characteristics, a few tests were executed. First, the Shapiro-Wilk test was executed to see if the data was normally distributed. The data was not normally distributed (the significance is less than 0.05 in all cases, see Appendix M). Therefore, the independent-samples Kruskal-Wallis test was executed. For each agent characteristic (friendliness, expertise, empathy, reliability, communicative, seriousness) the mean and standard deviation together with the p-value for significance were calculated. As can be seen in Table 5, the differences between the versions were not statistically significant. A significance level of 0.05 was chosen, and all p-values were larger than this value. This indicates that the differences between the medians were not statistically significant. This means that the different ECAs were not perceived different from each other on their characteristics. However, that no statistical significance and normal distribution can be proven, might have to do with the small dataset of 20-22 participants per group.

Table 5 Mean and Standard Deviation together with the significance from the Independent-Samples Kruskal-Wallis test

	Version			
Characteristic	*Photo, M(SD)*	*Real. Ill., M(SD)*	*Cartoon, M(SD)*	*p*
Friendliness	3.86(0.56)	3.82(1.18)	3.95(0.95)	0.664
Reliability	3.68(0.78)	3.45(1.10)	3.65(0.75)	0.889
Empathy	3.14(0.71)	3.23(1.19)	3.50(0.61)	0.217
Expertise	3.41(0.73)	3.45(1.14)	3.35(0.88)	0.717
Communicative	2.95(0.65)	3.32(1.17)	3.20(0.77)	0.281
Seriousness	3.59(0.80)	3.68(1.09)	3.80(0.62)	0.613

Despite the small sample size, it can be helpful to get insight in the data. A diverging bar chart was made (Figure 19), with the neutral answers set apart. The division was made percentwise because the sample size of the cartoon was 2 participants smaller than the other two versions. The numbers inside the bars are the amount of responses, and the numbers in the circles are the means (where 1=Totally Disagree and 5=Totally Agree).

Figure 19 shows that most participants (50, 78,1%) agreed that their coach came across as if he was friendly. Percentwise the cartoon was rated the friendliest (16, 80%), followed by the photorealistic version and the realistic illustration (both 17, 77,3%). Participant F added in the interview that the coach (photorealistic version) came across friendly since he wanted to give advice. He also thought the coach came across reliable as he does not say weird things. Reliability also scored a high level of agreement overall (42, 65,6%). Again, the cartoon version scores the highest (15, 75%), with only one person totally disagreeing. Then comes the realistic illustration (14, 63%) and after that the photorealistic version (13, 59,1%). Participant D (photorealistic version) thought the white coat made the coach look reliable. However, participant C (realistic illustration) thought the coach came across as if he had sympathy, but his doctor-look made her feel like she was sick.

A characteristic that scored lower was empathy (27, 42%). Of the photorealistic version, most people voted neutral (14, 63%). Only 5 people (22,7%) agreed with that this version came across as if he were empathic. The realistic illustration (11, 50%) and the cartoon (11, 55%) scored a higher level of agreement. Participant F thought that to call the coach empathic went too far since it was still a static image who does not know you.

Another characteristic that scored low was communicative (21, 32,8%). Again, the photorealistic version scored the lowest amount of people who agreed to this statement (4,

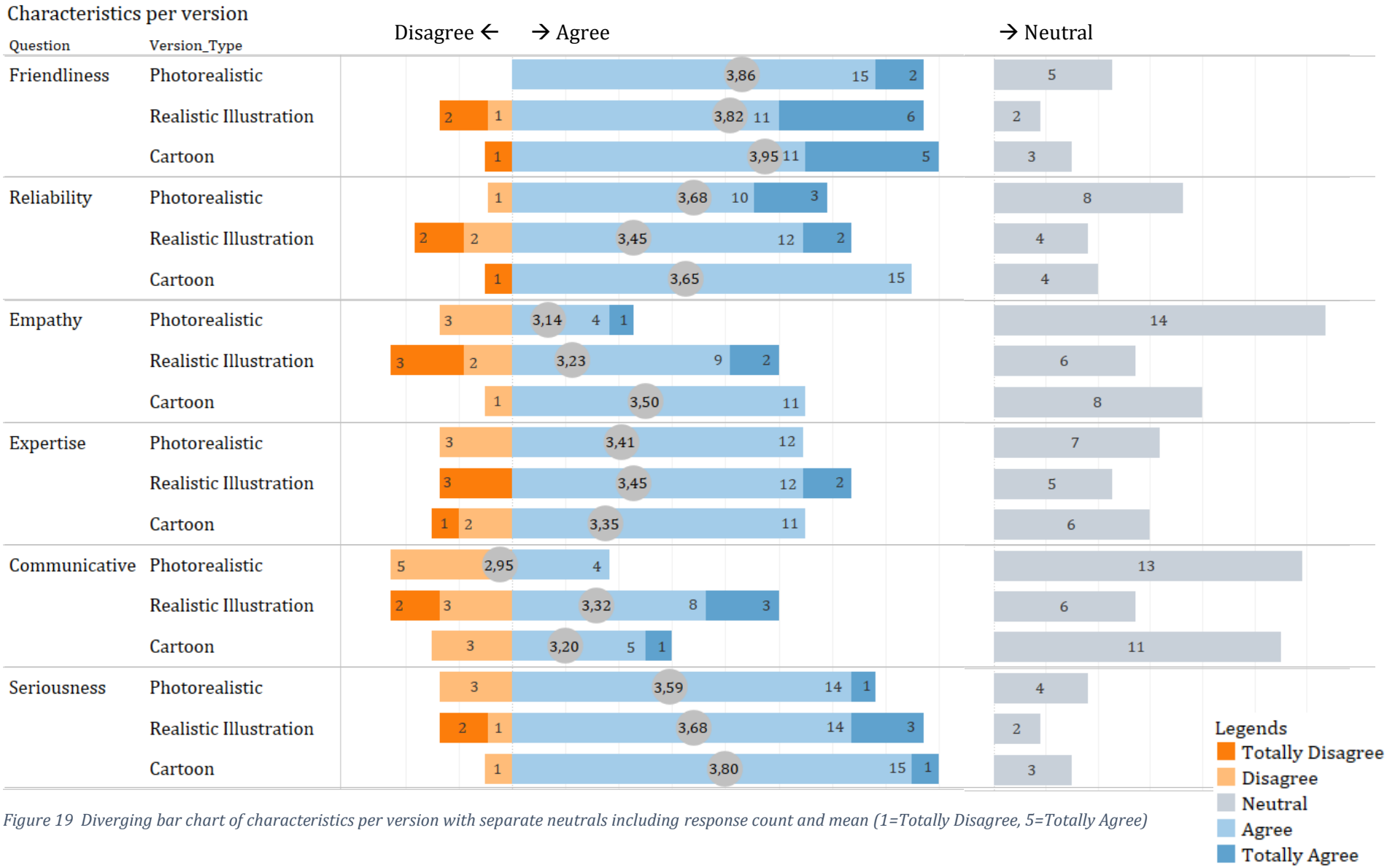


Figure 19 Diverging bar chart of characteristics per version with separate neutrals including response count and mean (1=Totally Disagree, 5=Totally Agree)

18%). It also has a mean of 2.95, which is the lowest mean of all combinations. The most agrees of the coach being communicative came from the realistic illustration (11, 50%) and then the cartoon (6, 30%). Participant A (photorealistic version) said: "Communicatively strong I did not agree, because I had no communication. I was expecting a question, and it was all in writing. I expected something verbally." Participant G (cartoon) did not find the coach communicative nor reliable or an expert since he directly commented on a "problem" while he did not know her situation.

Expertise scored somewhat above the half the sample size when counting the people who agreed (37, 58%). Per version this division was, first the realistic illustration (14, 63%), then the cartoon (11, 55%) and then the photorealistic version (12, 54.5%).

Lastly, most participants agreed that their coach came across serious (48, 75%). Percentwise the cartoon was rated the most serious (16, 80%), then the realistic illustration (17, 77.2%) and then the photorealistic version (15, 68.2%). Participant B (realistic illustration) thought the coach was looking serious because he looked neat, like he was serious about his job. She added that he had no earring or visible tattoos.

6.3.2 Preferences

Besides comparing the three different versions on the characteristics, it is also important to consider the preference of the participants themselves. In Table 6 the frequency table of these results can be found.

Table 6 Frequency table preference per version

Preferred version	**Version interacted with**							
	Photo		*Real. Ill.*		*Cartoon*		*Total*	
	Count	*%*	*Count*	*%*	*Count*	*%*	*Count*	*%*
Photo	16	72.7	17	77.3	14	70.0	47	73.4
Real. Ill.	3	13.6	1	4.5	4	20.0	8	12.5
Cartoon	3	13.6	4	18.2	2	10.0	9	14.1

Figure 20 shows a percentwise bar chart, which shows that most participants (47, 73.4%) preferred the photorealistic version. Figure 19 shows a histogram which visualizes Table 6. For the people who filled in the photorealistic version, the same amount of people preferred the realistic illustration and the cartoon. For the participants who filled in these versions, was a slight difference with a preference for the version they did not fill in themselves. Overall, the photorealistic version was preferred, participants B, C, G and J (all female) argued that they preferred this version because they had the feeling that there is a real person behind the app.

Participant A had her attention on the eyes and preferred the photorealistic version because his eyes were the most lifelike. Participants H and I could not really say what their argument was, but they thought the photorealistic version was the most appealing. Participant E preferred the photorealistic version because the eyebrows of the other two versions were too thick. He argued that the thicker the eyebrows, the more criminal the coach came across. Participant D said he was not so drawn to animations or drawings, which made the photorealistic version more his style. However, participant F preferred the cartoon because that one was more abstract, which made it clearer that it is a system instead of a person.

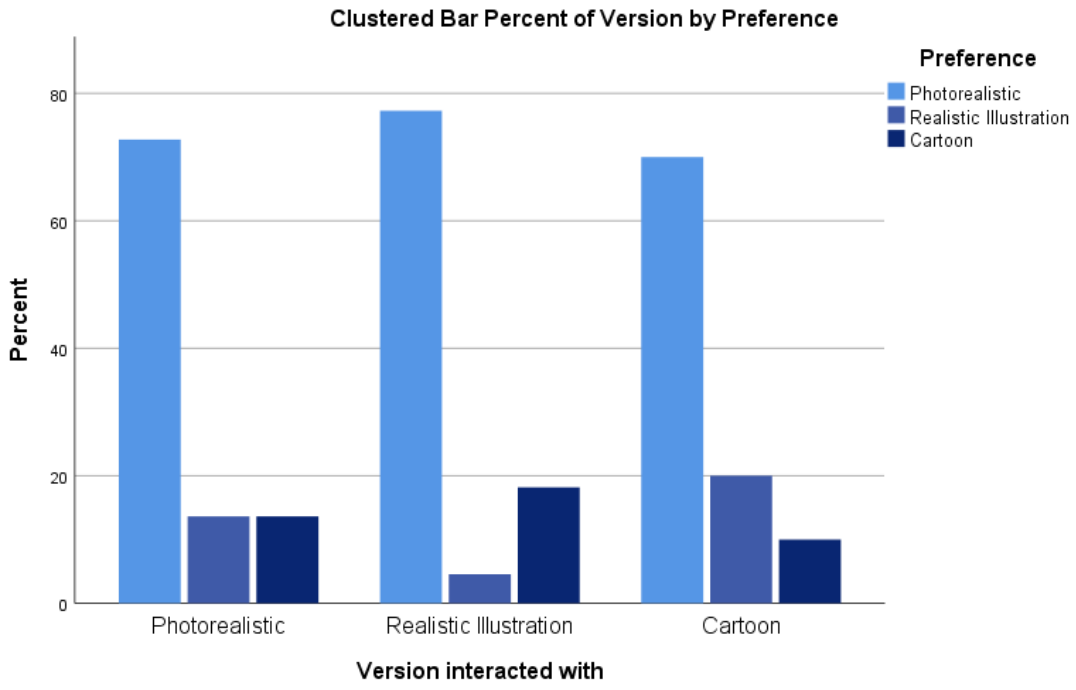


Figure 20 Bar Graph of Preference by Version

6.3.3 Feeling of talking to a person

Since it is also important that the participants have the idea that they talk to a person, which also gives them a reason to assign characteristics to this coach, this also needs to be considered. In Table 7, the results of this question can be found in a frequency table.

Table 7 Frequency table

Feeling of talking to coach	Version							
	Photo		*Real. Ill.*		*Cartoon*		*Total*	
	Count	*%*	*Count*	*%*	*Count*	*%*	*Count*	*%*
Totally disagree	4	18.2	4	18.2	4	20.0	12	18.8
Disagree	8	36.4	7	31.8	9	45.0	24	37.5
Neutral	6	27.3	7	31.8	2	10.0	15	23.4
Agree	3	13.6	3	13.6	4	20.0	10	15.6
Totally agree	1	4.5	1	4.5	1	5.0	3	4.7

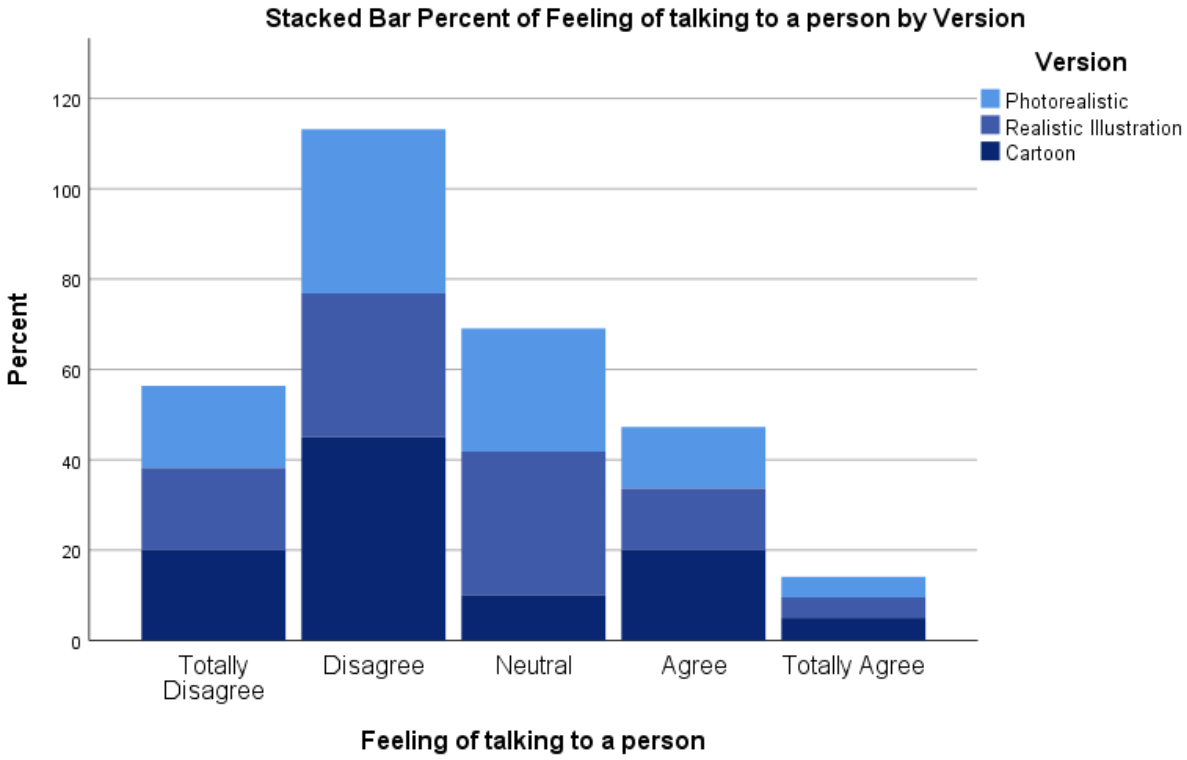


Figure 21 Stacked Bar Chart of Feeling of talking to a person by Version

Figure 21 shows a percentwise stacked bar chart which makes clear that most people disagrees with this feeling of that they are talking to a real person. Most statements have an equal division between the versions, only less people who filled in the cartoon version voted "Neutral". During the interviews, most participants mentioned that they did not have the feeling they really had a conversation with the coach. Participant C did not notice that she had to act as if he was her coach. She said "At first, I didn't know what the idea was. Because you do have that Henri in the picture as a drawing, after a few questions I realized that I had to pretend that man was my coach. Maybe because he was in the picture. Of course, I have often chatted with a company and then you do not have anyone in the picture, but in such a case I had the feeling even more that I

was talking to someone than I had now. With chatting someone reacts immediately." Participant D mentioned that with something online, he quickly has the feeling that the answers are not real-life, but statistically correct. Participant F added that for him it felt more like an intelligent webpage. "For me it's a kind of intelligent webpage in which I can read a piece in a handy way. If I would not be able to read so well or if I do not feel like it, it would be efficient. But with more personal questions you will of course get some better advice."

6.4 Conclusion

We found that, for the elderly population of 55-75 years old, (1) people seem to have a clear preference of the photorealistic version, (2) the agent's visual realism affects the perception of the characteristics of the agent image, although this is not statistically significant and that (3) people do not seem to have the feeling that they talked to a person when interacting with the coach. Furthermore, it could be that audio and animation would add to the experience and people would feel more like they are talking to a person which would affect their perception of the personality of the agent.

7. Final iteration by adding sound and animation

7.1 Goal

The goal of this final iteration was to test whether sound and/or animations makes people feel more like they are talking to a person and would influence their perception of the personality of the agent.

7.2 Method

7.2.1 Participants

The 10 participants who were interviewed during the user testing, also reacted to this final iteration. This sample consisted of 10 elderly people (4 males, 6 females; M_{age} = 62,3, SD = 5,23, ranging from 55 to 70 years). The participants were coded the same as in the user evaluation. Thus, participant A the user evaluation is the same person as participant A in this final iteration.

7.2.2 Materials

In Adobe After Effects, a short animation was created where the photorealistic version of the coach makes idle movements (such as slightly moving his head and eyebrows and blinking his eyes) and has a simple mouth animation. His mouth movements (open and closed) are synchronized with his speech. The audio was created with an online text-to-speech converter [41]. Figure 22 shows three stills of this final iteration. The animation and audio was made for the first two screens and exported to an mp4.

7.2.3 Procedure

The video of the animation with sound was shown to the participants at the end of the interview in which they elaborated on their survey answers. A short semi-structured interview was conducted in which the participants were asked whether these movements and sound added something comparing to a static image. They were also asked more specifically what adds the most, the sound or the movements.

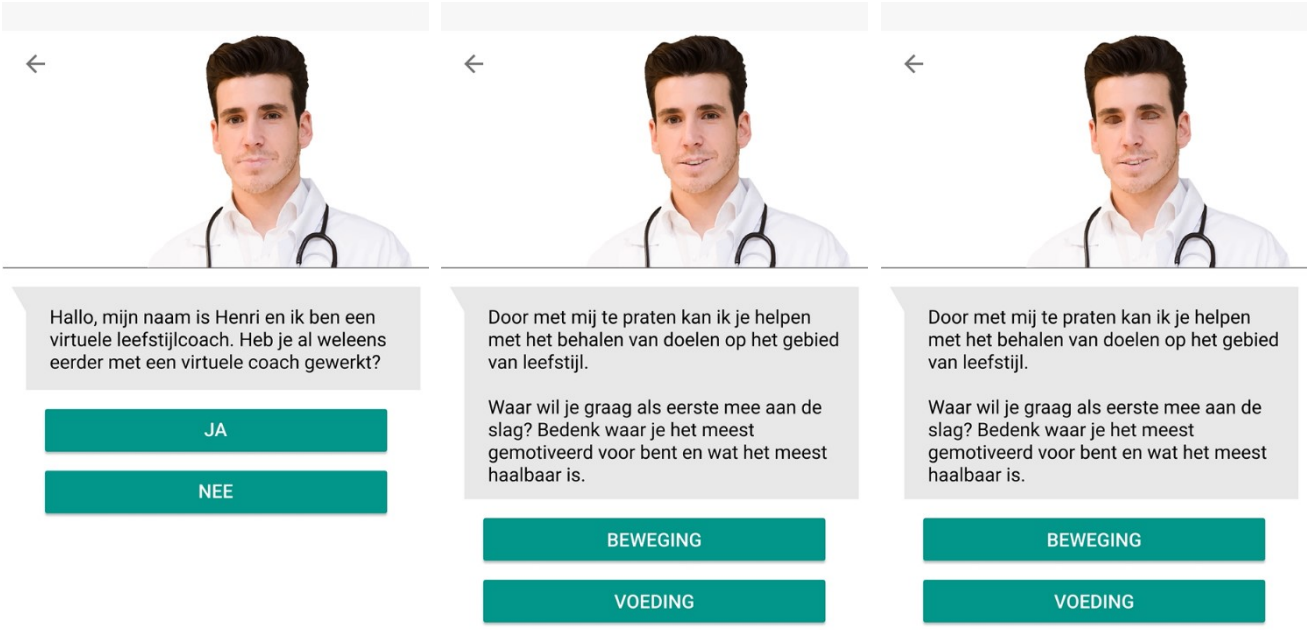


Figure 22 Stills of the animation

7.3 Results final iteration

At first, of the 10 participants, the division of their answers on whether they had the feeling that they were talking to a person is as followed: 5 participants (A, F, G, H, and I) voted that they did not have that feeling, while 4 participants (B, C, E and J) did have this feeling. One participant (D) answered neutral on this question. Table 8 shows the results per part when the animation and sound were added. As can be seen, half of the participants (B, D, F, G and J) thought the whole animation including sound and mouth and idle movements had added value in the sense of that people feel more like they are talking to a person. Participant B added that because of the whole image with all its movements came across as if the coach was more reliable.

Table 8 Added value per part per participant

Part / Participant	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*Total*
General	X	X	X	X	X	X	X	X		X	*9*
Sound	X	X	X	X	X	X	X			X	*8*
Idle movements	X	X		X	X	X	X			X	*7*
Mouth movements		X		X		X	X			X	*5*

The other 5 participants had one or more elements they did not find to have added value. Most participant (8, 80%) found that the sound had added value, and after that come the idle movements such as the eye blinking with 7 participants (70%). Only half of the participants thought that the mouth movements had added value.

Participant I thought that the iteration in general did not have added value, she would even prefer the static image over this animation and sound. She said “I think it is a bit made up. He does not say what is written, he just moves his mouth back and forth. It is not really someone talking to me. I think it's more likely to take off because it's so unreal.” Participant H agreed that it looked and sounded a bit made-up, only he thought that in general it would add value with reference to a static image. Participant C thought that only the sound added something. According to her, the movements make it somewhat more real although the sound really has added value.

Half of the participants did not think the movements of the mouth had added value. Participant A mentioned that the mouth was disturbing her, since the agent did not make the shape of the words with his mouth. She said “It is like being with one of those hand puppets. Then you are doing a little bit like this with your hand, it is chatting. But then you do not have the facial and motor skills.” Participant E added that he thought the agent did not open his mouth much and therefore it had not much added value. He thought that if the coach opened his mouth further, he would talk more clearly. He added that if your eyes are a little worse, this would matter. In addition, participant G suggested that an emoticon could be added to set up a mood. She argued that emotion would work well since empathy always works well.

7.4 Conclusion

The aim of this final iteration was to test whether sound and/or animations makes people feel more like they are talking to a person and would influence their perception of the personality of the agent. The results suggest that in general, the final iteration of adding animation and sound adds value to the coach in the sense of that people feel more like they are talking to a person.

8. Final conclusion and discussion

8.1 Conclusion

The main goal of this research was to get insight in the effect of visual realism on the perception of the personality of an ECA in the form of a lifestyle coach for elderly people. A human-centred investigation was done with two-dimensional ECAs in a mobile app interface where the looks of the agents differ in the amount of realism.

Our results suggest that (1) people have a clear preference for a photorealistic agent image, (2) the agent's visual realism affects the perception of the characteristics of the agent image and that (3) people do not seem to feel that they have talked to a person while interacting with the agent. In addition, the interviews showed that audio and animation, preferably as realistic as possible, could improve the experience of talking to a real person.

8.2 Discussion

8.2.1 Preference for the photorealistic coach

Our results indicate that respondents preferred the photorealistic coach when it comes to them choosing their preference themselves. The preference for the photorealistic version could be the result of the health coaching task being associated with a conversation with a real person. Several participants mentioned that the more lifelike the coach looked, the more they had the idea that they were talking to a real person. This is in line with the research of van Wissen [13]. They conclude that when designing ECAs as health coaches, their appearance should be realistic in order to mimic human coaching relationships. It is also in line with the results of Ring et al. [30] suggest that in a purely medical system a highly realistic agent may be a better design. However, they suggest whereas for a social system, a cartoon-like agent might work better. One can discuss whether this lifestyle coach context is medical or social. Our results are partly in line with the results of McDonnell et al. [14] whose study shows that toon shaded (more abstract cartoon versions), and highly realistic models were best received across the different comparisons. Besides this, participant E said that he could not see much differences between the versions as face recognition was not his strong side. It could be discussed whether the versions were too similar.

8.2.2 Characteristics

The results of our research on the perception on the characteristics of the agents was not proven statistically significant. Since the sample size was relatively small, it could become statistically significant if it were tested with a larger population. The results of McDonnell et al. [14] show that cartoon characters were considered highly appealing and were rated as more

friendly than realistic styles. In our results, the cartoon also scored the highest percentwise although the photorealistic version did not get any "disagrees" on this characteristic. It is also remarkable that the photorealistic version does not get rated higher on the characteristics while people give it their clear preference when they can choose between the versions. Even more so, the photorealistic version scored percentwise the lowest on every characteristic. This could be the result of that people had higher expectations of someone looking as realistic but is still a static image. This effect could cause the higher ratings of the cartoon since the participants would have lower expectations of this version in general. A less realistic character would be rated more empathic or friendly-looking, because you expect him to be less empathic or friendly-looking. But for a photorealistic agent, you know that this is not as empathic as he can be, since he looks like a real person, only static. Participant A expected sound and face and talking movements. This is related to the feeling of talking to a person, which is elaborated in the next section.

Besides these characteristics, context might also play a role. Participant A mentioned that she expected someone dressed in sporting clothes when she chose the option of physical activity. She said the doctor fitted the nutrition track but did not fit the other track. Therefore, the appearance of the coach might depend on the context, which is also in line with studies that show that preference for particular agents and perception of their personalities depend on the task of the agent [29], [30]. Thereby, appearance of the coach as a doctor was also not always appreciated. Participant G thought that the coach looked like a stereotype with his white coat and stethoscope. For her, this backfired since she knew that this was a technique to come across as more reliable. Participants C and J said that they did not find the stethoscope necessary, since it made the coach look like a doctor while he was a lifestyle coach and not a doctor.

Participant I did not find any characteristic fitting, she said that when there was a possibility, she would have answered "I don't know" for every characteristic. For her, this was caused by the brevity of the interaction. Therefore, the fact that the interaction was short could have influenced the ratings of the characteristics for more participants. Participant F added that he might have found the coach more reliable and empathic when the interaction would have been longer and more personal.

8.2.3 Feeling of talking to a person

Clearly the majority of participants did not feel like they were talking to a real person. This could have been influenced by the coach being a static image. As said, a participant expected movements and sound. Moreover, participant G thought that the coach showing emotion would also add something. This was confirmed by participant D, who thought the interaction was quite cold, the coach does not show or read emotions.

An explanation for the fact that few people had the feeling they were talking to a person could be found in the study of McDonnell et al. [14]. Their results show that participants were so focused on the task, that the appearance of the character did not sway them. Participant H said that he did not feel like it was a real conversation, more like a textual conversation. It could be discussed that the participants did not really give attention to the coach since they were so focused on reading the text. McDonnell et al. found that the audio and animation contributed to the interpretation of the characters' intention rather than the render style.

8.2.4 Adding sound and animation

The results of the final iteration suggest that in general, the final iteration of adding animation and sound adds value to the coach in the sense of that people feel more like they are talking to a person. However, the details of the animation, such as more realistic mouth movements and a more realistic sound could be improved. It could be discussed that the more realistic the character looks; the more realistic people expect the character to talk and move. This is discussed in the paper of McDonnell et al. [14]. They state that motion anomalies are considered more unpleasant on human than on cartoon render styles. They believe this was due to the fact that humans are inherently conditioned to analyse human faces and are therefore less forgiving of irregularities when a human photograph is applied to the model. Furthermore, emotion or emoticons could be added to set a mood. This could cause that the coach comes across more empathic.

8.3 Strengths and limitations research

Our research was user-centred and therefore, the elderly was taken into the process of designing and adapting the virtual coach. This was a strength, together with the elaborate prior research which minimized the influence of other variables when research only the visual realism. Despite these strengths, there were also some limitations of the research.

Our results show that the perceptions of characteristics of an agent relate to the visual realism of this agent. However, to value differences in perception of specific characteristics among the versions of these agents, it would be relevant to know how these characteristics are valued in the first place. By researching the "absolute importance" of the agent characteristics, we could estimate the relative importance of the different characteristics to each other and better value the differences. This might also have caused the differences between the results of the characteristics and the results when the participants could choose their preferred agent.

In addition, since the sample size was relatively small, the results were not statistically significant. Thereby, the sample size of the cartoon was somewhat smaller than the other two groups, which might have caused differences in results between the groups. Thereby, since a

participant mentioned that he could not see much differences between the versions, the three versions might have looked too similar to measure real differences.

Another factor that might have influenced the results of our research is that the agent designs indirectly differed on multiple design features. Although we tried to overcome this by an elaborate prior research with co-design sessions, some design features like the thickness of the eyebrows or the smoothness of the beard, were not constant over the designs. Therefore, this might have affected the ratings of the agent's characteristics and the preferences.

Moreover, the scenario that was used for the testing only included multiple choice questions and was quite short. This might have influenced the perception of the characteristics of the agents. Several participants mentioned the brevity of the interaction and told them the agent was not personally involved.

Thereby, the animations of the final iteration were not so advanced, influencing the realism of the movements and sound. This might have influenced the results on the added value of the sound and animations. This last research was with a small sample size, which influenced the significance of the results.

Although this study only focused on three different static agents designs differing on one dimension and designed for a lifestyle coaching context and for elderly people, our results may be generalized toward other contexts, such as the more serious health context or rehabilitation context..

8.4 Future work

Our research focused on 2D static images, and a smaller scaled research on some animation and sound. Future work could research the difference between 2D and 3D images as well as large-scale research on animation and sound. Since our results suggest that the realism of the sound and animations have influence on how people perceive the agent and the conversation, this could be further researched.

Future work could also research with a longer or even long-term interaction. Then, participants could get to know the coach better and this might influence their perception of its characteristics. Thereby, future work could develop a coach who is more personally involved, who also makes a profile of the user or has insight in his situation.

In addition, future work could research the effect of emotions or emoticons on the perception of the characteristics of the coach. Even as more realistic sound and animations, this might have an important influence. This could also be done with other elements like similarity, or a personalized coach.

References

[1] L. S. Green, L. G. Oades, and & A. M. Grant, "Cognitive-behavioral, solution-focused life coaching: Enhancing goal striving, well-being, and hope," *J. Posit. Psychol.*, vol. 1, no. 3, pp. 142–149, 2006, doi: 10.1080/17439760600619849.

[2] World Health Organization, "Physical Activity and Older Adults: Recommended levels of physical activity for adults aged 65 and above.," 2015. [Online]. Available: https://www.who.int/dietphysicalactivity/factsheet_olderadults/en/. [Accessed: 16-Mar-2020].

[3] CBS, "Forecast: 19 million inhabitants in 2039," 2019. [Online]. Available: https://www.cbs.nl/en-gb/news/2019/51/forecast-19-million-inhabitants-in-2039. [Accessed: 16-Mar-2020].

[4] U. E. Reinhardt, "Does the aging of the population really drive the demand for health care?," *Health Affairs*, vol. 22, no. 6. pp. 27–39, 2003, doi: 10.1377/hlthaff.22.6.27.

[5] W.-P. Brinkman, "Virtual Health Agents for Behavior Change: Research Perspectives and Directions," 2016.

[6] J. Cassell, J. Sullivan, E. Churchill, and S. Prevost, *Embodied Conversational Agents*. MIT Press, 2000.

[7] R. Klaassen, D. Heylen, and G. Huizing, "Designing and Developing Lifelike, Engaging Lifestyle Coaching Agents and Scenarios for Multiparty Coaching Interaction," in *CEUR Workshop Proceedings*, 2018, pp. 25–29.

[8] R. Op Den Akker, R. Klaassen, J. E. W. C. Van Gemert-Pijnen, O. Kulyk, and L. Van Gemert-Pijnen, "Personalized Virtual Coaching for Lifestyle Support: Principles for Design and Evaluation," 2014.

[9] C. J. Brandt, G. I. Søgaard, J. Clemensen, J. Søndergaard, and J. B. Nielsen, "Determinants of Successful eHealth Coaching for Consumer Lifestyle Changes: Qualitative Interview Study Among Health Care Professionals.," *J. Med. Internet Res.*, vol. 20, no. 7, p. e237, Jul. 2018, doi: 10.2196/jmir.9791.

[10] L. L. Kramer, S. Ter Stal, B. C. Mulder, E. de Vet, and L. van Velsen, "Developing Embodied Conversational Agents for Coaching People in a Healthy Lifestyle: Scoping Review," *J. Med. Internet Res.*, vol. 22, no. 2, p. e14058, Feb. 2020, doi: 10.2196/14058.

[11] S. Schindler, E. Zell, M. Botsch, and J. Kissler, "Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory," *Sci. Rep.*, vol. 7, Mar. 2017, doi: 10.1038/srep45003.

[12] A. L. Baylor and Y. Kim, "Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role," 2004, pp. 592–603.

[13] A. van Wissen, C. Vinkers, and A. van Halteren, “Developing a virtual coach for chronic patients: A user study on the impact of similarity, familiarity and realism,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9638, pp. 263–275, doi: 10.1007/978-3-319-31510-2_23.

[14] R. McDonnell, M. Breidty, and H. H. Bülthoff, “Render me real? Investigating the effect of render style on the perception of animated virtual humans,” *ACM Trans. Graph.*, vol. 31, no. 4, 2012, doi: 10.1145/2185520.2185587.

[15] “Design Thinking in Health Care: How a Virtual Nurse Helps Keep People Out of the Hospital - School of Design and Creative Technologies - The University of Texas at Austin.” [Online]. Available: https://designcreativetech.utexas.edu/design-thinking-health-care-how-virtual-nurse-helps-keep-people-out-hospital. [Accessed: 15-Jun-2020].

[16] T. B. Moyers, J. Houck, S. L. Rice, R. Longabaugh, and W. R. Miller, “Therapist empathy, combined behavioral intervention, and alcohol outcomes in the COMBINE research Project,” *J. Consult. Clin. Psychol.*, vol. 84, no. 3, pp. 221–229, 2016, doi: 10.1037/ccp0000074.

[17] T. W. Bickmore, L. Caruso, K. Clough-Gorr, and T. Heeren, “‘It’s just like you talk to a friend’ relational agents for older adults,” *Interact. Comput.*, vol. 17, no. 6, pp. 711–735, Dec. 2005, doi: 10.1016/j.intcom.2005.09.002.

[18] S. ter Stal, L. L. Kramer, M. Tabak, H. op den Akker, and dr. ir. H. Hermens, “Design Features of Embodied Conversational Agents in eHealth: a Literature Review,” *Int. J. Hum. Comput. Stud.*, Feb. 2020, doi: 10.1016/j.ijhcs.2020.102409.

[19] S. ter Stal, M. Tabak, H. op den Akker, T. 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,” *Int. J. Hum. Comput. Interact.*, 2019, doi: 10.1080/10447318.2019.1699744.

[20] A. L. Baylor, “The design of motivational agents and avatars,” *Educ. Technol. Res. Dev.*, vol. 59, no. 2, pp. 291–300, Apr. 2011, doi: 10.1007/s11423-011-9196-3.

[21] 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, IVA 2018*, 2018, pp. 301–306, doi: 10.1145/3267851.3267915.

[22] H. Nguyen and J. Masthoff, “Is it me or is it what I say? Source image and persuasion,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2007, vol. 4744 LNCS, pp. 231–242, doi: 10.1007/978-3-540-77006-0_29.

[23] A. Schmeil and S. Suggs, “How am i doing?- Personifying health through animated characters,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8519 LNCS, no. PART 3, pp. 91–102, doi: 10.1007/978-3-319-07635-5_10.

[24] P. Skalski and R. Tamborini, “The role of social presence in interactive agent-based persuasion,” *Media Psychol.*, vol. 10, no. 3, pp. 385–413, 2007, doi: 10.1080/15213260701533102.

[25] H. C. Van Vugt, E. A. Konijn, J. F. Hoorn, and J. Veldhuis, “Why fat interface characters are better e-health advisors,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006, vol. 4133 LNAI, pp. 1–13, doi: 10.1007/11821830_1.

[26] S. Zhou, T. Bickmore, M. Paasche-Orlow, and B. Jack, “Agent-user concordance and satisfaction with a virtual hospital discharge nurse,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8637 LNAI, pp. 528–541, doi: 10.1007/978-3-319-09767-1_63.

[27] S. Robertson *et al.*, *The visual design and implementation of an embodied conversational agent in a shared decision-making context (eCoach)*, vol. 9192. 2015.

[28] J. Forlizzi, J. Zimmerman, V. Mancuso, and S. Kwak, “How interface agents affect interaction between humans and computers,” in *Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces, DPPI’07*, 2007, pp. 209–221, doi: 10.1145/1314161.1314180.

[29] H. Nguyen and J. Masthoff, “Designing empathic computers: The effect of multimodal empathic feedback using animated agent,” in *ACM International Conference Proceeding Series*, 2009, vol. 350, p. 1, doi: 10.1145/1541948.1541958.

[30] L. Ring, D. Utami, and T. Bickmore, “The Right Agent for the Job? The effects of agent visual appearance on task domain,” 2014.

[31] S. Fan, T. T. Ng, J. S. Herberg, B. L. Koenig, and S. Xin, “Real or fake? Human judgments about photographs and computer-generated images of faces,” *SIGGRAPH Asia 2012 Tech. Briefs, SA 2012*, vol. 1, no. 212, pp. 3–6, 2012, doi: 10.1145/2407746.2407763.

[32] S. Fan, R. Wang, T. T. Ng, C. Y. C. Tan, J. S. Herberg, and B. L. Koenig, “Human perception of visual realism for photo and computer-generated face images,” *ACM Trans. Appl. Percept.*, vol. 11, no. 2, pp. 1–21, 2014, doi: 10.1145/2620030.

[33] M. Mori, K. F. MacDorman, and N. Kageki, “The uncanny valley,” *IEEE Robot. Autom. Mag.*, vol. 19, no. 2, pp. 98–100, 2012, doi: 10.1109/MRA.2012.2192811.

[34] T. Geller, “Overcoming the Uncanny Valley,” 2008.

[35] J. Nielsen and T. K. Landauer, “INR!RCHI A Mathematical Model of the Finding of Usability Problems.”

[36] M. Agrawala, W. Li, and F. Berthouzoz, “Design principles for visual communication,” *Commun. ACM*, vol. 54, no. 4, pp. 60–69, Apr. 2011, doi: 10.1145/1924421.1924439.

[37] G. L. Brase and J. Richmond, “The white-coat effect: Physician attire and perceived authority, friendliness, and attractiveness,” *J. Appl. Soc. Psychol.*, vol. 34, no. 12, pp. 2469–2481, 2004, doi: 10.1111/j.1559-1816.2004.tb01987.x.

[38] J. R. Jongeneel and A. van Duinen, “Doorbreek je gewoontes en kom in actie,” 2019.

[39] R. Sakdulyatham, S. Preeyanont, R. Lipikorn, and R. Watakakosol, “User Interface on Smartphone for Elderly Users,” *Int. J. Autom. Smart Technol.*, vol. 7, no. 4, pp. 147–155, 2017, doi: 10.5875/ausmt.v7i4.1339.

[40] V. R. Preedy and R. R. Watson, “5-Point Likert Scale,” in *Handbook of Disease Burdens and Quality of Life Measures*, Springer New York, 2010, pp. 4288–4288.

[41] “Free Text-To-Speech for Dutch language and MP3 Download | ttsMP3.com.” [Online]. Available: https://ttsmp3.com/text-to-speech/Dutch/. [Accessed: 23-Jun-2020].

9. Appendices

9.1 Appendix A: Information Brochures and Consent Forms (original in Dutch)

9.1.1 Co-design group

Dear reader,

Thank you for joining my research. As agreed, we will hold three sessions, which will take place online. My graduation project is about the design of a virtual lifestyle coach (think of a puppet on a screen with text next to it telling someone what to do, so more sports, healthy eating, etc.) for the elderly. I will especially ask for your opinion and interpretation about the online lifestyle coach in question. The sessions are always filled out differently, but in general it is an interview / conversation form in which you can give feedback and express your thoughts on something I will show. These can be portrait photos, characters that I have designed, or an app that you can test. The sessions will all take a maximum of one hour. At the end, you can request the results and the research and then you will get to know them. I would like to ask you to read and sign this consent form.

As a participant of these co-design sessions:
- You will be asked to call
- You will be asked to use a web app on your laptop
- You will be asked to use an app on your phone
- You will participate 3 times in a structured interview

Participation in these sessions is entirely voluntary. The audio of the interviews will be recorded and transcribed. These recordings will be stored in a safe place; only the researcher has access. When all usable data has been transcribed, the audio recordings will be removed correctly and only the transcription will be used. Anonymous quotes or fragments can be taken out to add to the thesis report. You can withdraw your consent at any time and participation will be terminated immediately. You are free to stop your participation in the research at any time for any reason. If you have any questions, you may contact me, or for any further questions with my supervisor or the ethics committee (contact details can be found at the bottom of this letter).

I have read the information on this form and have no further questions.

_____ _____

Signature participant Date

Kind regards,

Geanne Woertink
g.j.woertink@student.utwente.nl

Supervisor: Job Zwiers, j.zwiers@utwente.nl
Ethics committee: ethics-comm-ewi@utwente.nl

9.1.2 Research group

Dear reader,

Thank you for joining my research. As agreed, we will conduct an interview, which will take place over the phone. My graduation project is about the design of a virtual lifestyle coach (think of a puppet on a screen with text next to it telling someone what to do, so more sports, healthy eating, etc.) for the elderly. You will first interact with this coach via an app. After this you will fill in a questionnaire about this and then we will talk about this during the interview. I will mainly ask for your explanation and interpretation about the online lifestyle coach in question and your answers to the questionnaire. The interview will take a maximum of one hour. At the end you may request the results and the research and then you will get to know them. I would like to ask you to read and sign this consent form.

As a participant of this research:
- You will be asked to use an app on your phone
- You will be asked to complete a questionnaire
- You will be asked to call
- You will participate in a structured interview

Participation in this study is entirely voluntary. The audio of the interviews will be recorded and transcribed. These recordings will be stored in a safe place; only the researcher has access. Once all usable data has been transcribed, the audio recordings will be removed correctly and only the transcription will be used. Anonymous quotes or fragments can be taken out to add to the thesis report. You can withdraw your consent at any time and participation will be terminated immediately. You are free to stop your participation in the research at any time for any reason.

If you have any questions, please contact me, or my supervisor (Job Zwiers), the contact details can be found at the bottom of this letter.

Would you like independent advice about participating in this survey, or would you like to submit a complaint? If so, you can contact Petri de Willigen, secretary of the Ethics Committee (tel. 053-489 2085, ethics-comm-ewi@utwente.nl). This committee consists of independent experts from the university and is available for questions and complaints about the investigation.

I have read the information on this form and have no further questions.

_____ _____

Signature participant Date

Kind regards,
Geanne Woertink
g.j.woertink@student.utwente.nl

Supervisor: Job Zwiers, j.zwiers@utwente.nl

9.2 Appendix B: Outline co-design session 1 (original in Dutch)

Context

Imagine that you want to work on your health, but don't know where to start or how to do it. A virtual health coach can help you get started. A virtual coach is a digital person who, through conversations with the user, can provide support in achieving goals. Think about your nutrition, sleep or exercise.

First questions

- How old are you?
- Do you understand what a health coach means?
- What would you use a health coach for? Would you be interested in it yourself? (measuring whether someone has a base of interest)

Assignment

It is up to you to put these 12 pictures in order from the one you would prefer as a health coach to the one you would least prefer as a coach.

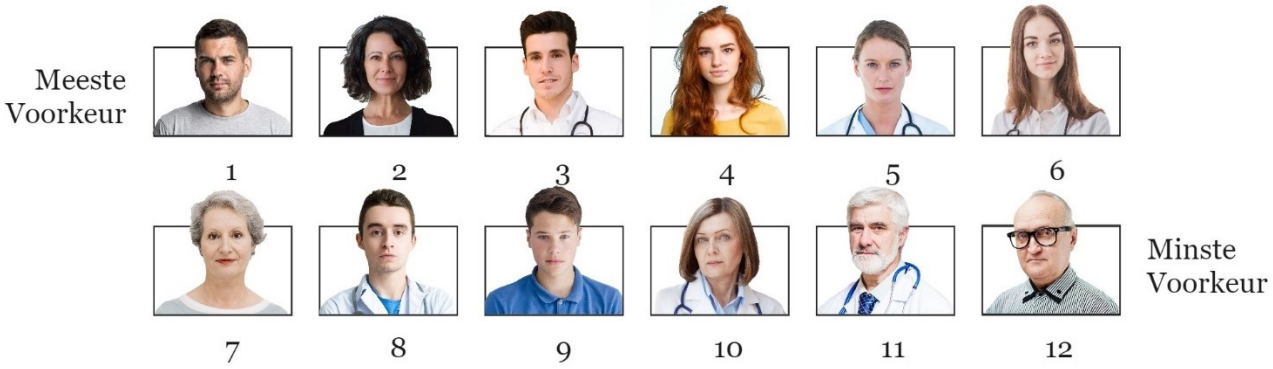
Structured interview

- Why did you choose this order?
- Was there someone who stood out for you and why?
    - So, what characteristics do you think this person has? (For example, can he/she be trusted, does he/she come across as if he/she has expertise, etc.)
- What characteristics do you think a good (according to you) health coach should have?
- To what extent did gender, age and role determine your choice?

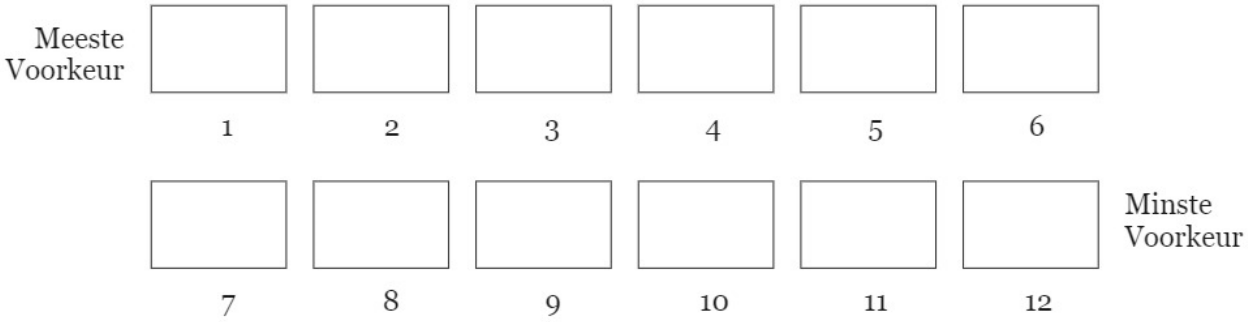
9.3 Appendix C: Results co-design session 1

Participant A (Male, 66):

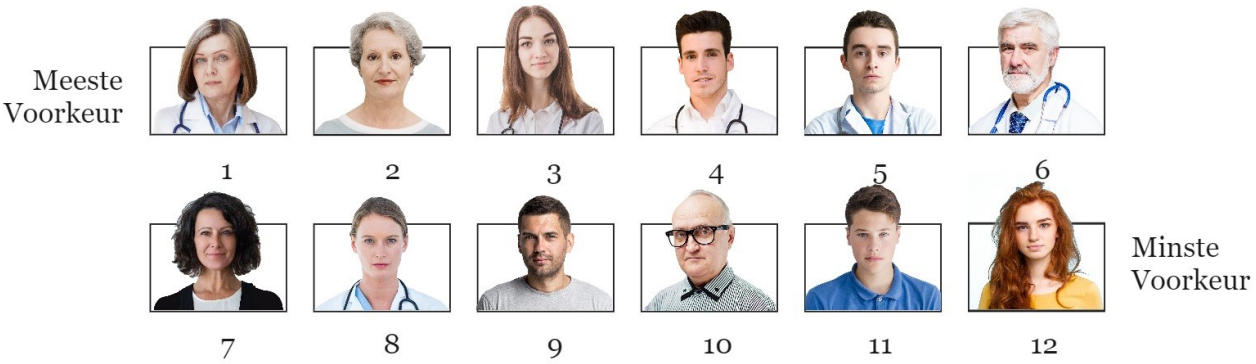
Participant B (Male, 58):



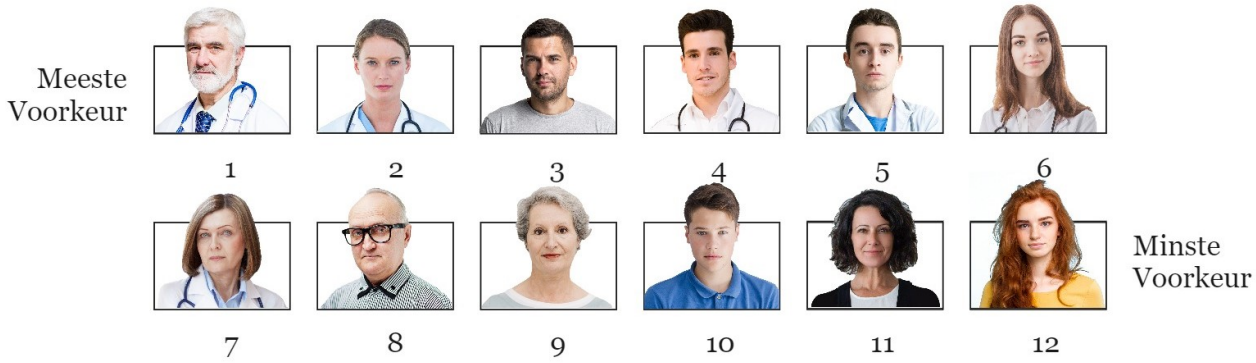
Participant C (Female, 74):



Participant D (Female, 68):



Participant E (Male, 68):



9.4 Appendix D: Scenario of user interface co-design session 2 (original in Dutch)

Virtual Agent (VA):

1. Hello, my name is Henri and I am a virtual health coach. Have you worked with a virtual coach before? (>2a&2b)

You (U):

2a. Yes. (>3a)

2b. No. (>3b)

VA:

3a. I hope you have an idea of what and who I am because of this previous experience.

U:

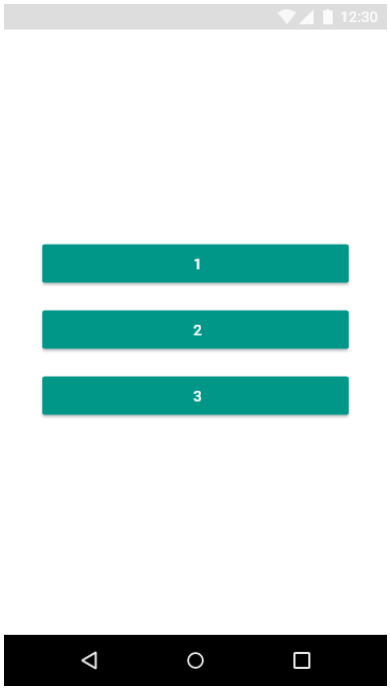
3aa. Yes (>startscreen)

3ab. No (>3b)

VA:

3b. A virtual coach is a digital person that can offer help with reaching goals through conversations with the user. (>startscreen)

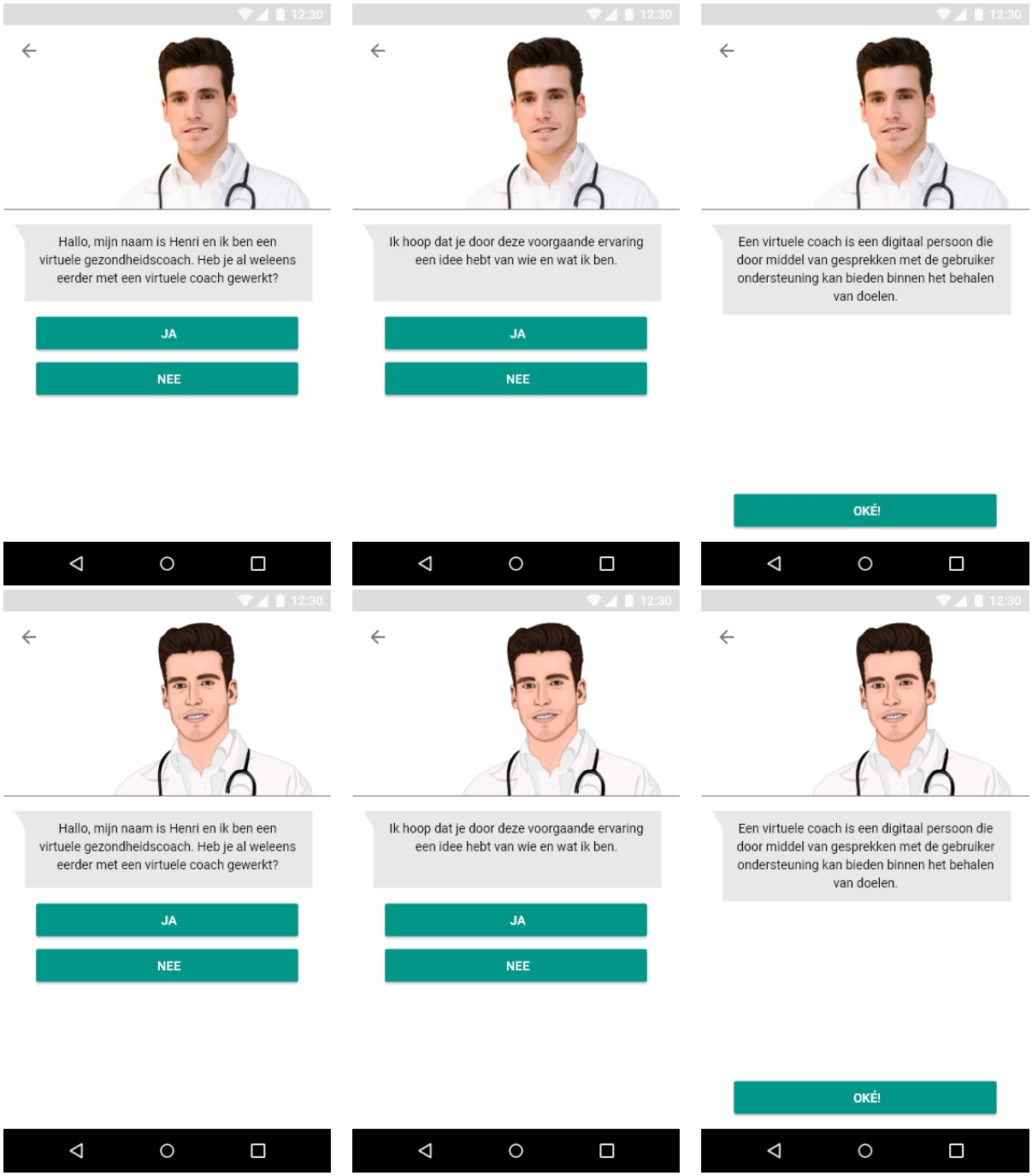
9.5 Appendix E: Screens user interface used in co-design session 2

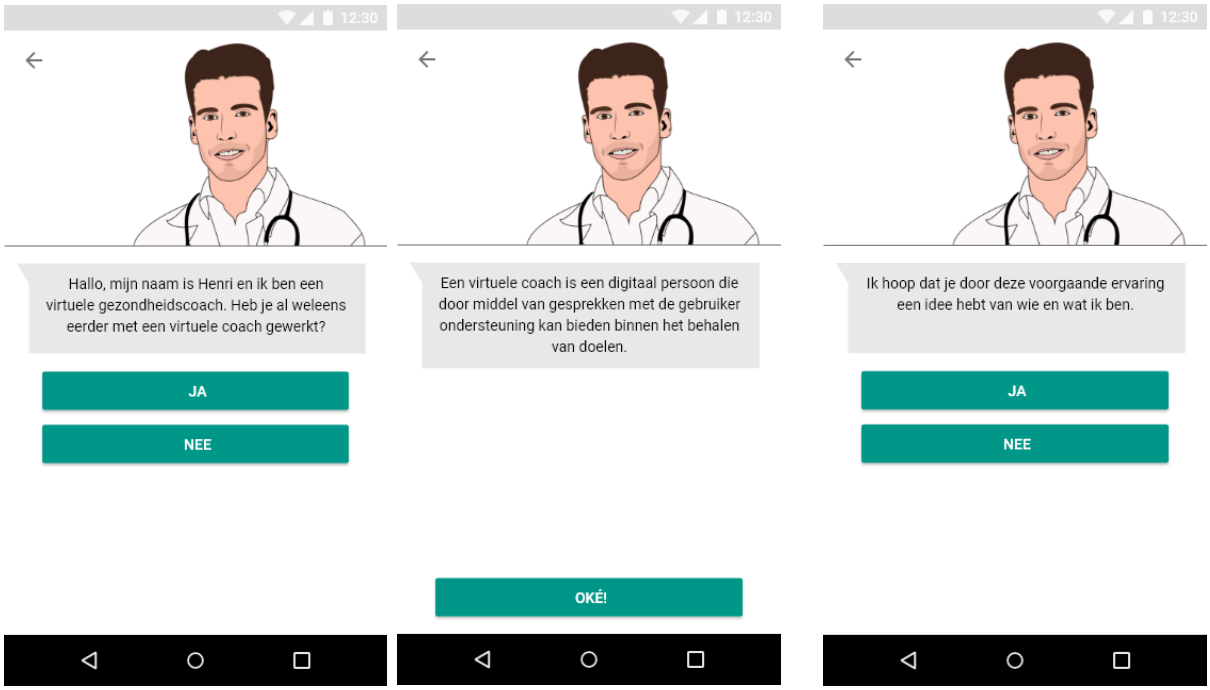


1 → Photorealistic version

2 → Realistic illustration version

3 → Cartoon version


12:30
Hallo, mijn naam is Henri en ik ben een virtuele gezondheidscoach. Heb je al weleens eerder met een virtuele coach gewerkt?
JA
NEE
12:30
Ik hoop dat je door deze voorgaande ervaring een idee hebt van wie en wat ik ben.
JA
NEE
12:30
Een virtuele coach is een digitaal persoon die door middel van gesprekken met de gebruiker ondersteuning kan bieden binnen het behalen van doelen.
OKÉ!
12:30
Hallo, mijn naam is Henri en ik ben een virtuele gezondheidscoach. Heb je al weleens eerder met een virtuele coach gewerkt?
JA
NEE
12:30
Ik hoop dat je door deze voorgaande ervaring een idee hebt van wie en wat ik ben.
JA
NEE
12:30
Een virtuele coach is een digitaal persoon die door middel van gesprekken met de gebruiker ondersteuning kan bieden binnen het behalen van doelen.
OKÉ!



9.6 Appendix F: Outline co-design session 2 (original in Dutch)

Structured interview

[sharing screen with the first picture: cartoon]

- What do you notice about this image?
- What do you think of this person?
- How does this person look to you? What qualities / characteristics does he have?

[show second picture: photorealistic]

- What do you notice about this image?
- What do you think of this person?
- How does this person look to you? What qualities / characteristics does he have?

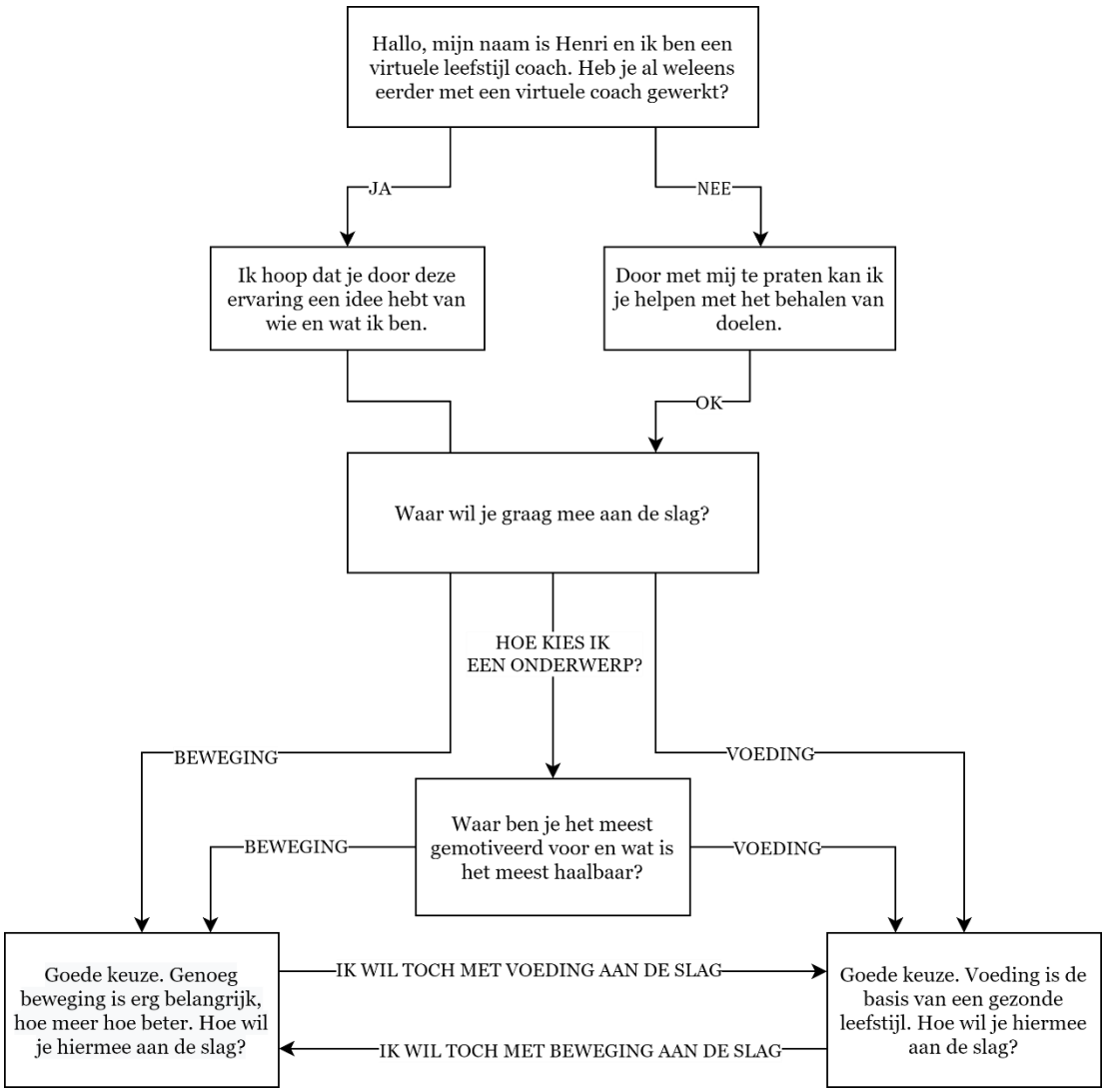
[show third picture: realistic illustration]

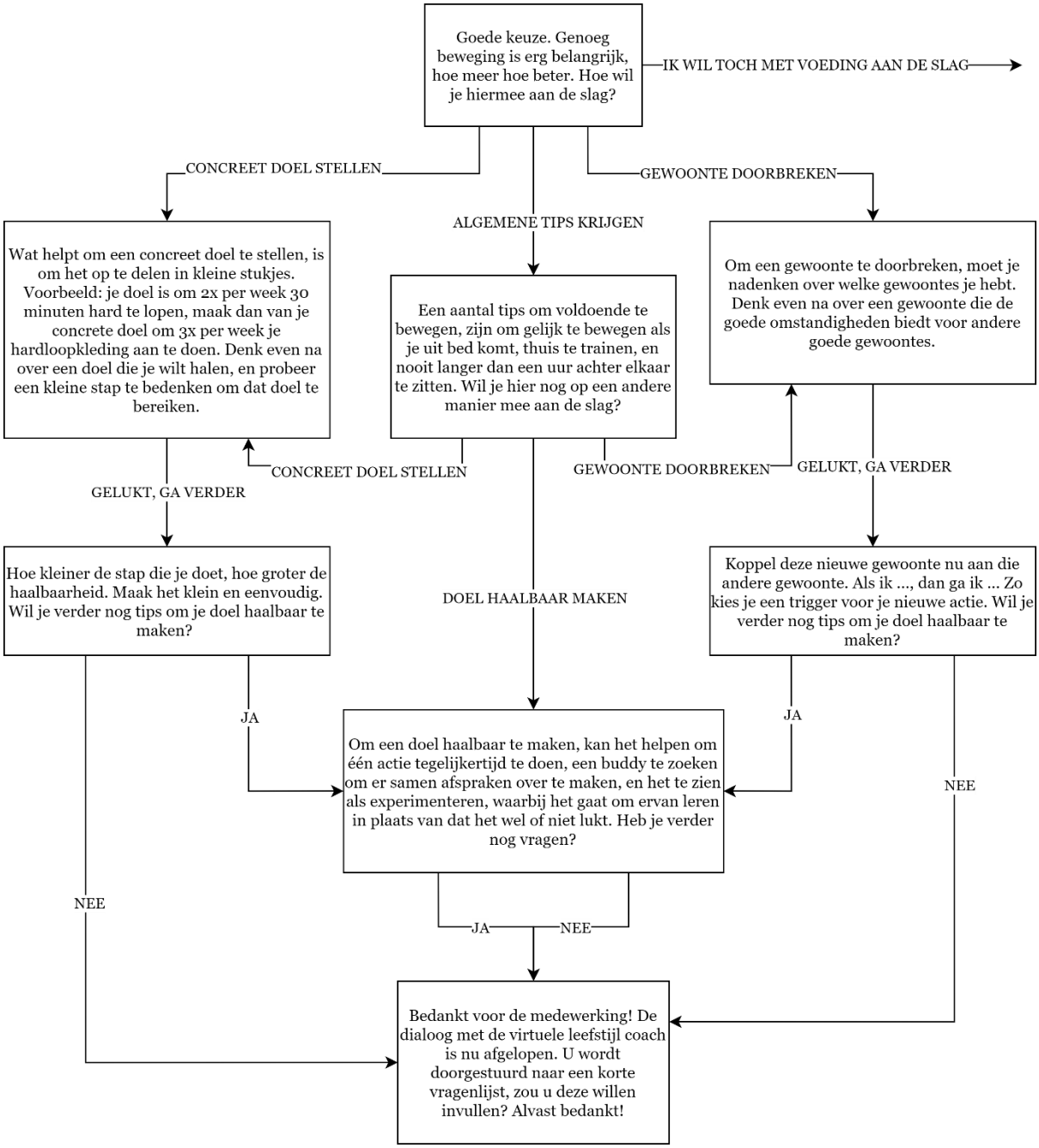
- What do you notice about this image?
- What do you think of this person?
- How does this person look to you? What qualities / characteristics does he have?

[showing all three images at once]

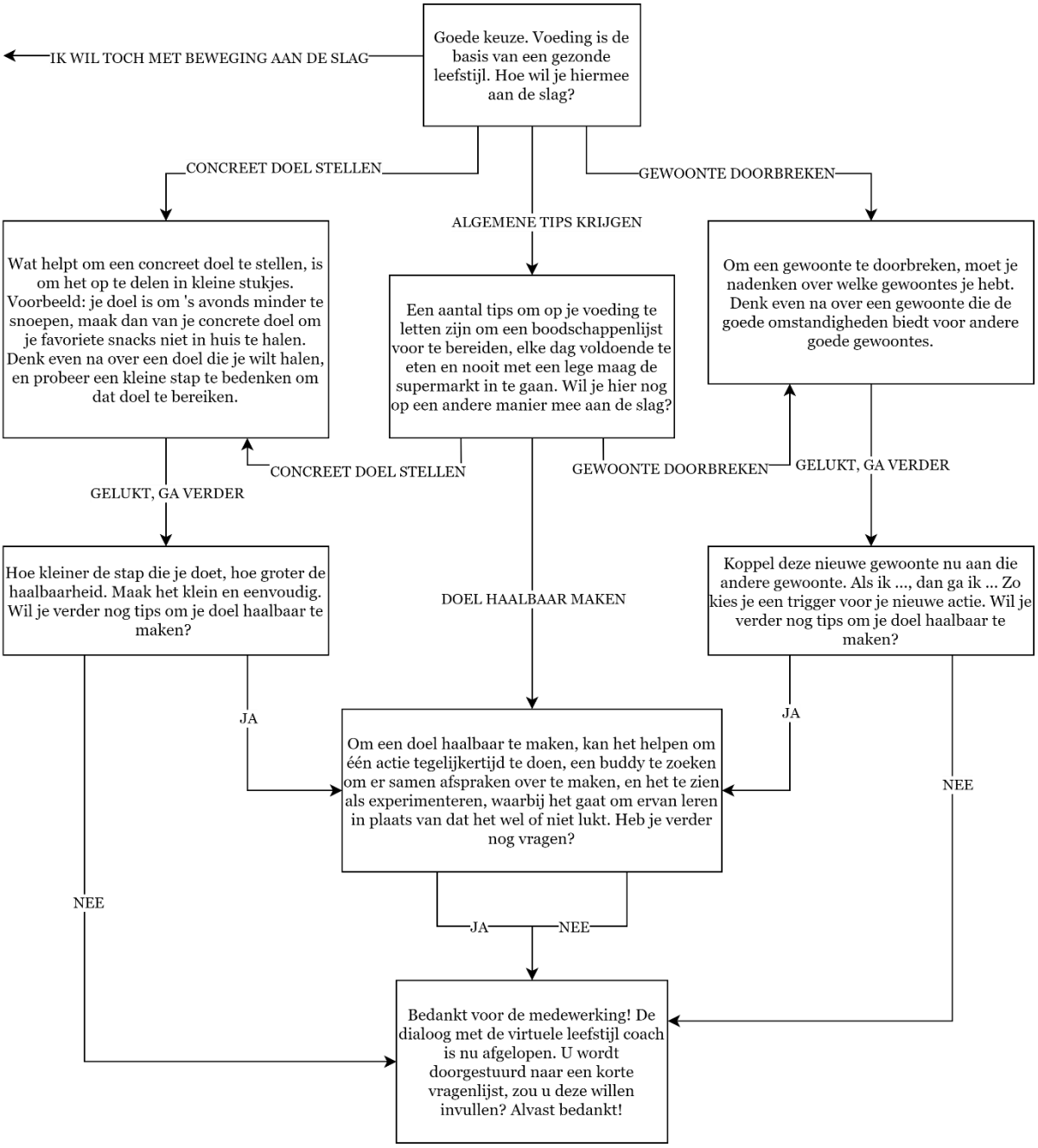
- What do these people differ in?
- What things do you notice about these people, are there things that bother you?
- How do they strike you now? The same way or is there a difference between them?
- Who do you prefer?

9.7 Appendix G: decision tree / dialog before co-design session 3

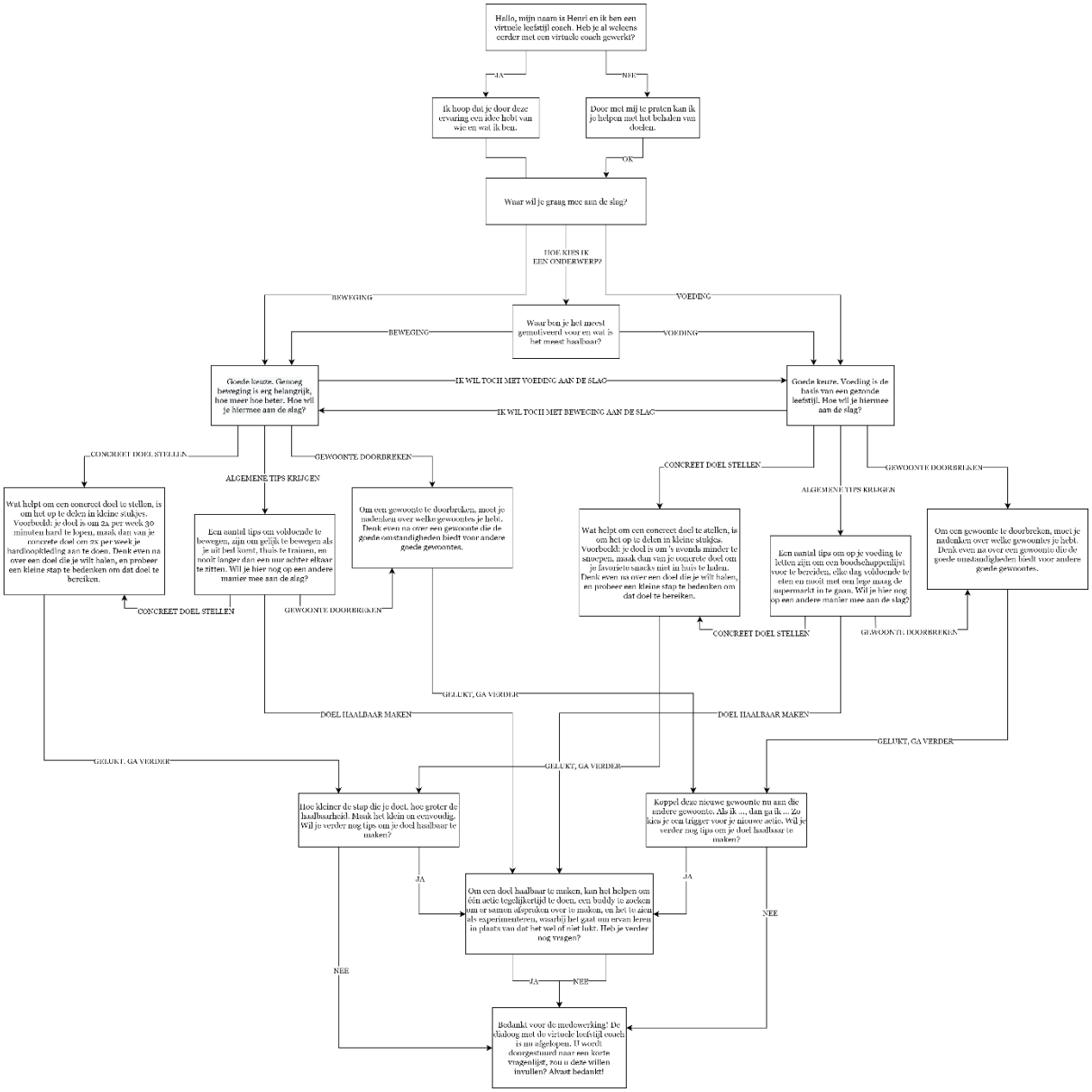




Part about physical activity



Part about nutrition



Overview of total interactions and needed screens

9.8 Appendix H: Outline co-design session 3 (original in Dutch)

Structured interview

- What do you think of the interface? Is it easy to use?
- What do you think of the information it contains and how do you find it? Is that clear?
- In general, what is your experience with the app?
- What would you like to change?

9.9 Appendix I: Iterations dialog – changelog

9.9.1 Changelog participant A

INTRODUCTIE

Eerder met coach gewerkt:

Ik hoop dat je door deze ervaring een idee hebt van wie en wat ik ben. Waar wil je graag als eerste mee aan de slag? Bedenk hiervoor waar je het meest gemotiveerd voor bent en wat het meest haalbaar is.

Niet eerder met coach gewerkt:

Door met mij te praten kan ik je helpen met het behalen van doelen. Waar wil je graag als eerste mee aan de slag? Bedenk hiervoor waar je het meest gemotiveerd voor bent en wat het meest haalbaar is.

DELETION of option “Hoe kan ik kiezen?”

DELETION of options “Ik wil toch met voeding aan de slag” and “Ik wil toch met beweging aan de slag”

CONCREET DOEL STELLEN

Beweging:

Wat helpt om een concreet doel te stellen, is om het op te delen in kleine stukjes. Voorbeeld: je doel is om 2x per week 30 minuten hard te lopen, maak ~~dan van je concrete doel om 2x per week je hardloopkleding aan te doen~~of zoek dan een hardloopschema waarbij je hier in kleine, concrete stappen naartoe kunt werken. Denk even na over een doel die je wilt halen, en probeer een kleine stap te bedenken om dat doel te bereiken.

GEWOONTE DOORBREKEN

Om een gewoonte te doorbreken, ~~moet je~~is het handig om eerst na te denken over ~~welke~~ de gewoontes die je al hebt. Een voorbeeld van een gewoonte is elke ochtend een kop koffie drinken. Denk even na over ~~een~~ welke gewoonte ~~die~~ de ~~goede~~ juiste omstandigheden ~~biedt~~ schept voor andere goede gewoontes.

Koppel dan een ~~deze nieuwe~~ gewoonte die je wilt aanleren aan ~~nu aan~~ die ~~andere~~ bestaande gewoonte. Bijvoorbeeld, ~~Als~~nadat ik een kop koffie drink, ~~…, dan~~ ga ik ~~…~~ buiten een rondje wandelen. Zo kies je een trigger voor je nieuwe actie. Wil je verder nog tips om je doel haalbaar te maken?

ALGEMENE TIPS KRIJGEN

Voeding:

~~Een aantal~~Enkele tips om op je voeding te letten zijn de volgende:

- Maak ~~om~~ een boodschappenlijst ~~voor te bereiden, elke dag voldoende te eten en~~
- Zorg dat gezonde dingen in het zicht liggen
- Ga nooit met een lege maag de supermarkt in ~~te gaan.~~

Wil je hier nog op een andere manier mee aan de slag?

Beweging:

~~Een aantal~~Enkele tips om voldoende te bewegen, zijn de volgende:

- ~~om gelijk te bewegen als je uit bed komt,~~Sport samen met anderen
- ~~thuis te trainen,~~Zet trainingsmomenten als eerst in je agenda
- ~~en nooit langer dan~~ Kom telkens na een uur achter elkaar te ~~zitten~~hebben gezeten weer even in beweging~~.~~

Wil je hier nog op een andere manier mee aan de slag?

DOEL HAALBAAR MAKEN

Om een doel haalbaar te maken, kan het helpen om met één ~~actie~~ doel tegelijkertijd bezig te ~~doen~~gaan ~~,~~en een buddy te zoeken om er samen afspraken over te maken~~.,~~ Verder kan het helpen om vooraf te bedenken wat belemmeringen zijn om zo alvast te plannen hoe je die voorkomt of hoe je erop wil reageren. ~~en het te zien als experimenteren, waarbij het gaat om ervan leren in plaats van dat het wel of niet luk~~t. Heb je verder nog vragen?

9.9.2 Changelog participant B

INTRODUCTIE

Hallo, mijn naam is Henri en ik ben een virtuele leefstijl~~-~~coach. Heb je al weleens eerder met een virtuele coach gewerkt?

Niet eerder met coach gewerkt:

Door met mij te praten kan ik je helpen met het behalen van doelen op het gebied van leefstijl. Waar wil je graag als eerste mee aan de slag? Bedenk hiervoor waar je het meest gemotiveerd voor bent en wat het meest haalbaar is.

CONCREET DOEL STELLEN

Voeding:

Wat helpt om een concreet doel te stellen, is om het op te delen in kleine stukjes. Voorbeeld: je doel is om 's avonds minder te snoepen, maak dan ~~van~~ je ~~concrete~~ doel concreet door ~~om~~ je favoriete snacks niet in huis te halen. Denk even na over een doel die je wilt halen, en probeer een kleine stap te bedenken om dat doel te bereiken.

ALGEMENE TIPS KRIJGEN

Enkele tips die helpen om voldoende te bewegen, zijn de volgende:

- Sport samen met anderen
- Zet trainingsmomenten als eerste in je agenda
- ~~-~~Kom telkens na een uur achter elkaar te hebben gezeten weer even in beweging

Wil je hier nog op een andere manier mee aan de slag?

AFSLUITING

Bedankt voor ~~j~~de medewerking! De dialoog met de virtuele leefstijl coach is nu afgelopen. Je~~U~~ wordt doorgestuurd naar een korte vragenlijst, zou je~~u~~ deze willen invullen? Alvast bedankt!

9.9.3 Changelog participant D

INTRODUCTIE

Waar wil je graag als eerste mee aan de slag? Bedenk ~~hiervoor~~ waar je het meest gemotiveerd voor bent en wat het meest haalbaar is.

CONCREET DOEL STELLEN

Bij beweging:

Wat helpt om een concreet doel te stellen, is om het op te delen in kleine stukjes. ~~V~~Bijvoorbeeld: je doel is om 2x per week 30 minuten hard te lopen~~,~~ M~~m~~aak of zoek dan een hardloopschema waarbij je hier in kleine, concrete stappen naartoe kunt werken. Denk even na over een doel dat~~ie~~ je wilt behalen, en probeer een kleine stap te bedenken om dat doel te bereiken.

Bij voeding:

Wat helpt om een concreet doel te stellen, is om het op te delen in kleine stukjes. ~~V~~Bijvoorbeeld: je doel is om 's avonds minder te snoepen, maak dan je doel concreet door je favoriete snacks niet in huis te halen. Denk even na over een doel dat~~ie~~ je wilt behalen, en probeer een kleine stap te bedenken om dat doel te bereiken.

ALGEMENE TIPS KRIJGEN

Bij beweging:

Enkele tips die helpen om voldoende te bewegen, zijn de volgende:

- Sport samen met anderen;
- Zet trainingsmomenten als eerste in je agenda;
- Kom telkens na een uur ~~achter elkaar~~ te hebben gezeten weer even in beweging.

Wil je hier nog op een andere manier mee aan de slag?

Bij voeding:

Enkele tips om op je voeding te letten zijn de volgende:

- Zorg dat gezonde dingen in het zicht liggen;
- Maak een boodschappenlijst;
~~- Zorg dat gezonde dingen in het zicht liggen~~
- Ga nooit met een lege maag de supermarkt in.

Wil je hier nog op een andere manier mee aan de slag?

GEWOONTE DOORBREKEN

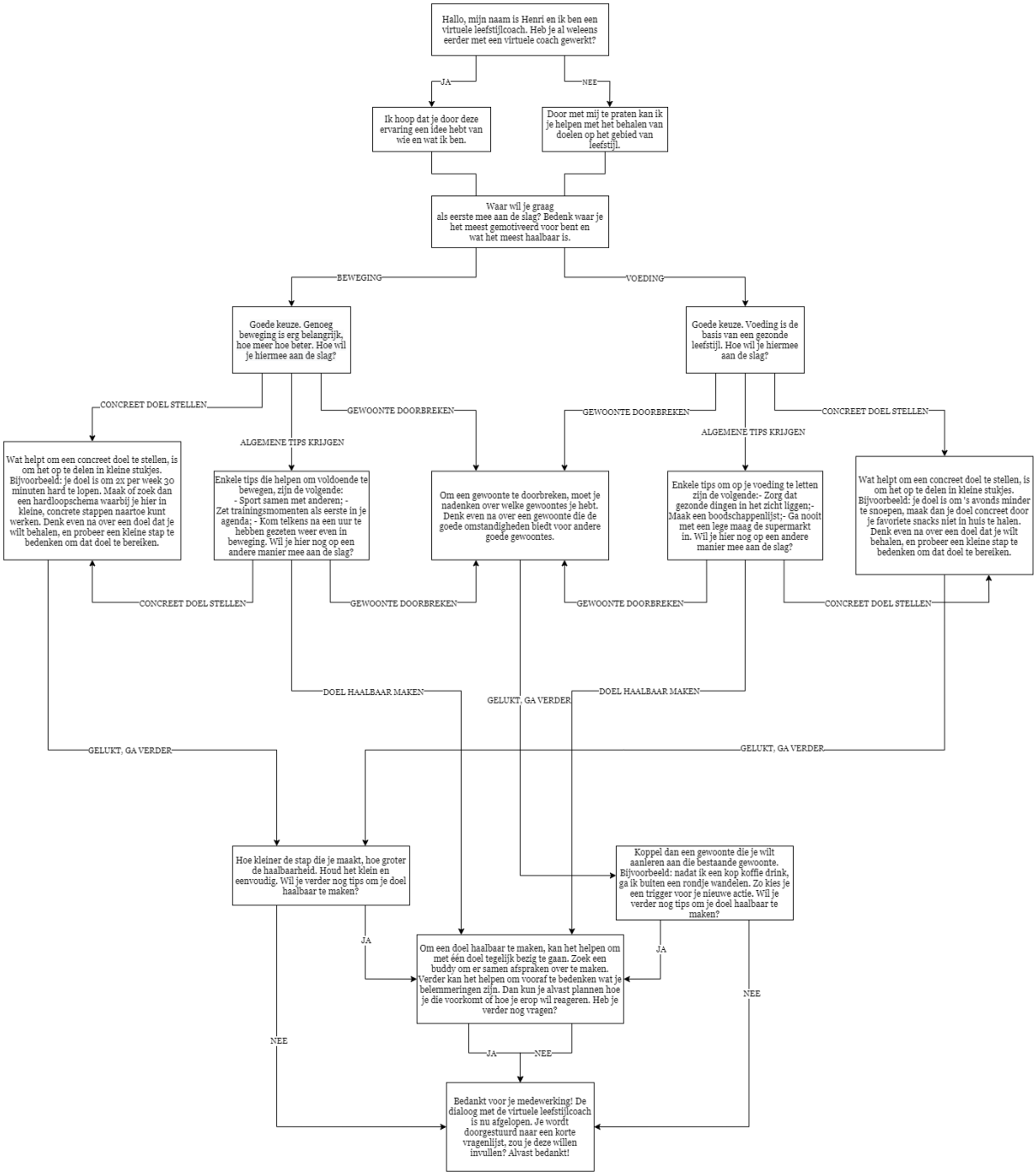
Koppel dan een gewoonte die je wilt aanleren aan die bestaande gewoonte. Bijvoorbeeld:~~,~~ nadat ik een kop koffie drink, ga ik buiten een rondje wandelen. Zo kies je een trigger voor je nieuwe actie. Wil je verder nog tips om je doel haalbaar te maken?

Hoe kleiner de stap die je ~~doet~~maakt, hoe groter de haalbaarheid. ~~Maak~~ Houd het klein en eenvoudig. Wil je verder nog tips om je doel haalbaar te maken?

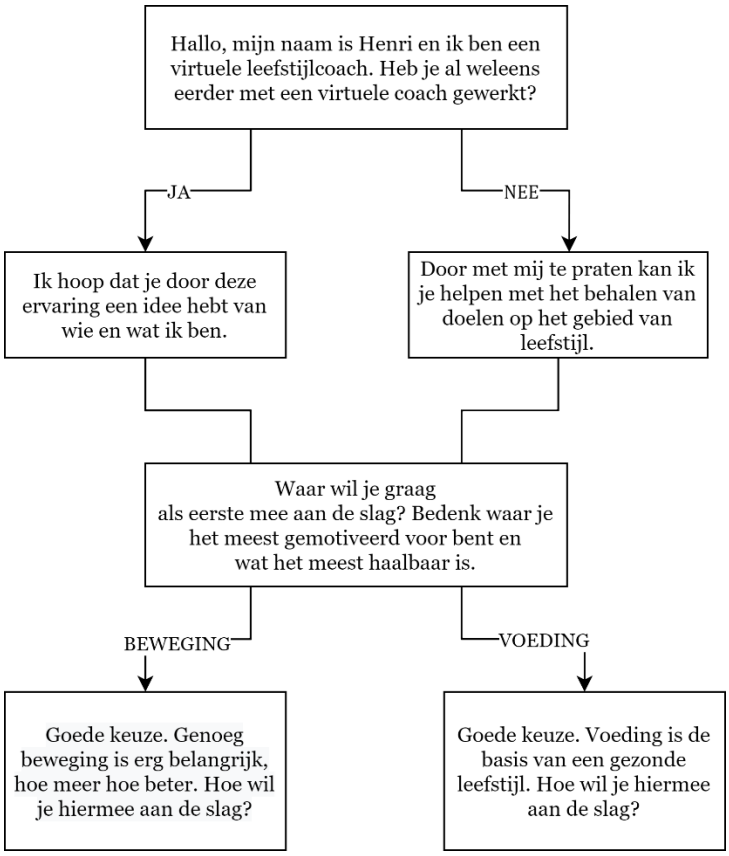
DOEL HAALBAAR MAKEN

Om een doel haalbaar te maken, kan het helpen om met één doel tegelijk~~ertijd~~ bezig te gaan. ~~en~~ Zoek een buddy ~~te zoeken~~ om er samen afspraken over te maken. Verder kan het helpen om vooraf te bedenken wat je belemmeringen zijn. Dan kun je ~~om zo~~ alvast ~~te~~ plannen hoe je die voorkomt of hoe je erop wilt reageren. Heb je verder nog vragen?

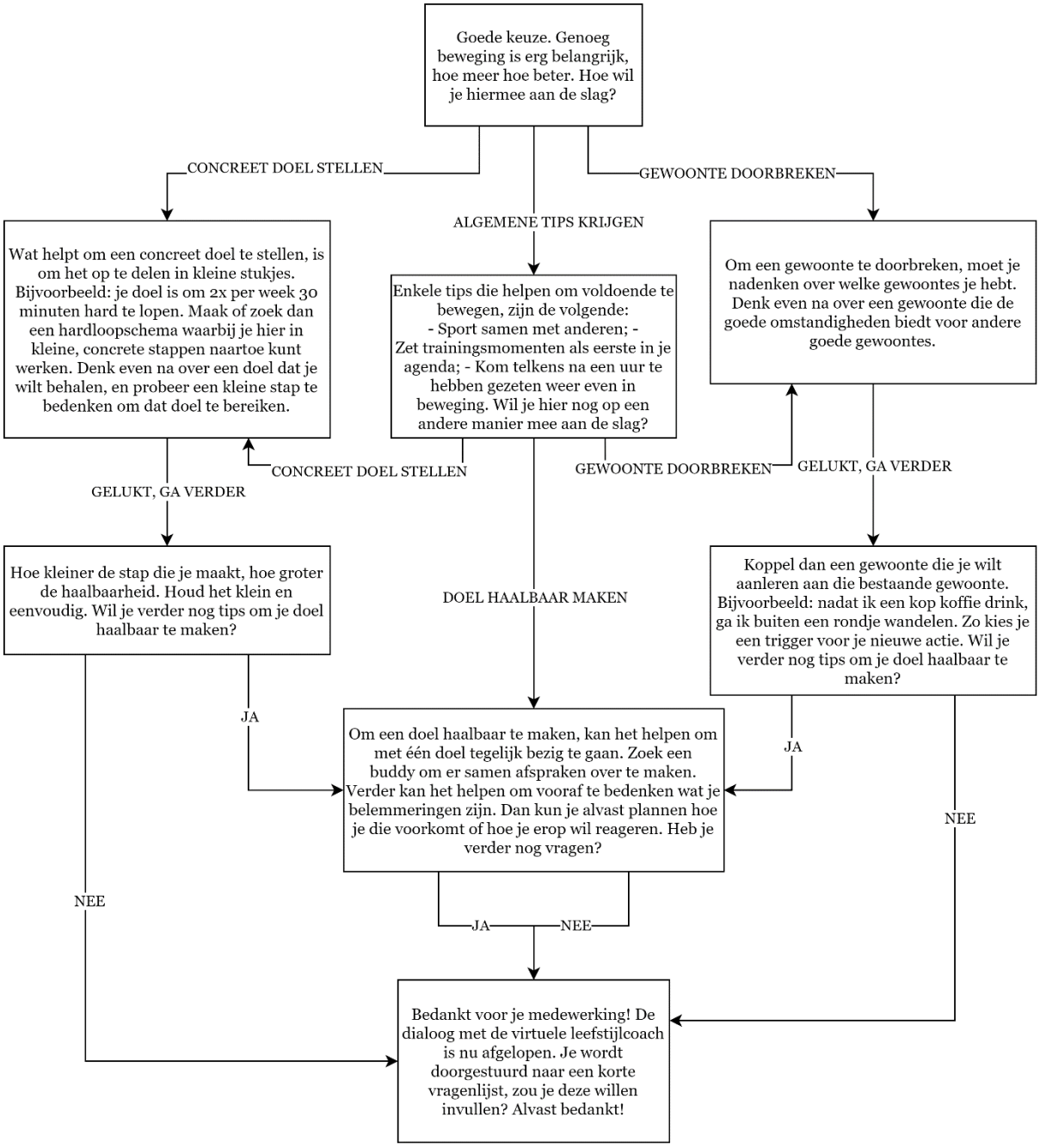
9.10 Appendix J: Final dialog



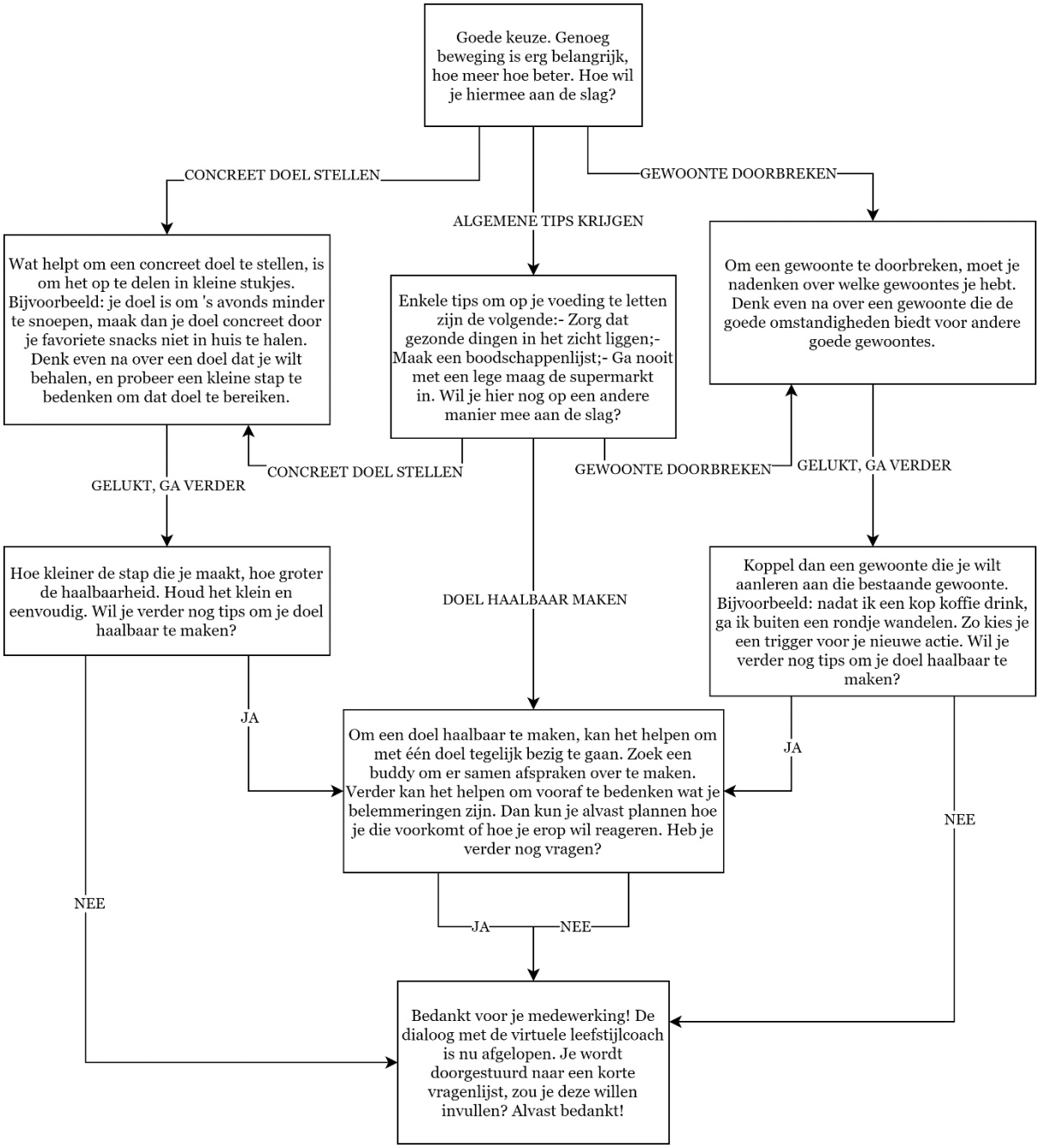
Overview decision tree



Introduction part



Part about physical activity



Part about nutrition

9.11 Appendix K: Questionnaire user evaluation (original in Dutch)

1. How often do you use mobile applications?

Very often - Often - Frequently - Occasionally – Never

2. How did you feel about communicating with the lifestyle coach?

Very difficult - Difficult - Not difficult and not easy - Easy - Very easy

3. The coach comes across as if he is... (this is a matrix question with a 5-point Likert scale question: Totally disagree - Disagree – Neither agree nor disagree - Agree - Totally agree)
    - Friendly
    - Reliable
    - Empathic
    - Knowledgeable
    - Communicatively strong
    - Serious
4. To what extent did you have the feeling you were talking to Henri as a person?

Not at all – Not – Neutral – Much – Very much

5. Which of these three coaches do you prefer? (multiple choice with the images of the three agents)
    a. Why?
6. What improvements do you have for the app or the coach himself?
7. What is your age?
8. What is your gender?

9.12 Appendix L: Data Tables and descriptive statistics

ResponseID	Version	Age	Gender	Frequency	Easy communicating	Friendly	Reliable
1	1	70	Female	4	4	4	4
2	1	73	Male	4	3	4	3
3	1	59	Male	2	5	4	4
4	1	71	Female	3	3	4	5
5	1	71	Female	3	4	3	3
6	1	57	Female	5	3	4	4
7	1	72	Female	4	3	5	5
8	1	60	Female	3	5	5	5
9	1	62	Male	3	4	4	4
10	1	60	Female	1	3	3	3
11	1	60	Female	3	3	3	3
12	1	61	Female	4	4	4	2
13	1	63	Male	2	5	4	3
14	1	57	Male	4	1	3	4
15	1	66	Female	2	4	4	4
16	1	62	Female	5	5	3	3
17	1	63	Female	4	3	4	4
18	1	59	Male	1	5	4	3
19	1	57	Female	3	4	4	4
20	1	55	Male	2	4	4	4
21	1	58	Male	3	4	4	4
22	1	57	Male	4	4	4	3

23	2	58	Female	4	5	5	4
24	2	73	Female	3	3	5	5
25	2	71	Female	1	5	1	1
26	2	67	Male	4	4	4	3
27	2	61	Female	3	3	4	4
28	2	67	Male	3	3	4	4
29	2	67	Female	4	4	4	4
30	2	66	Female	1	4	5	5
31	2	59	Male	3	5	3	2
32	2	56	Female	4	4	5	4
33	2	56	Female	3	3	4	3
34	2	58	Male	3	4	4	4
35	2	61	Female	1	4	4	4
36	2	65	Female	2	5	4	4
37	2	58	Male	1	5	4	4
38	2	66	Female	2	3	5	4
39	2	66	Male	3	3	3	3
40	2	68	Male	4	4	4	4
41	2	75	Female	4	3	1	1
42	2	68	Male	1	4	2	2
43	2	55	Male	1	5	5	4
44	2	57	Female	2	3	4	3
45	3	58	Female	2	5	4	4
46	3	73	Male	1	4	4	4
47	3	65	Male	4	3	3	3
48	3	66	Female	3	5	5	4
49	3	55	Female	3	3	5	4
50	3	61	Female	1	4	1	4
51	3	61	Male	3	4	5	4
52	3	58	Male	3	5	5	4
53	3	66	Male	2	4	4	4
54	3	64	Female	2	4	4	4
55	3	58	Female	3	4	3	3
56	3	70	Male	4	2	4	4
57	3	66	Male	1	4	5	4
58	3	70	Male	2	4	4	3
59	3	66	Female	1	3	4	3
60	3	56	Female	5	5	4	4
61	3	55	Male	2	4	4	4
62	3	60	Female	4	4	3	1
63	3	74	Female	3	3	4	4
64	3	65	Female	3	4	4	4

ResponseID	Empathic	Expert	Communicative	Serious	Feeling of talking to a person	Preference
1	3	4	3	3	2	1
2	3	4	2	4	2	2
3	2	4	3	4	2	1
4	4	3	4	5	3	1
5	3	2	2	3	1	2
6	4	4	4	4	4	1
7	3	4	3	4	3	1
8	5	4	4	4	5	3
9	3	4	3	4	2	1
10	3	3	3	4	3	1
11	3	3	3	3	1	2
12	2	2	3	2	2	1
13	4	3	3	4	2	1
14	3	4	3	4	3	1
15	3	3	2	4	2	1
16	3	3	3	2	1	1
17	3	4	3	4	4	1
18	3	4	3	3	4	1
19	3	3	3	4	3	1
20	2	2	2	2	1	3
21	4	4	4	4	2	3
22	3	4	2	4	3	1
23	4	4	4	5	5	1
24	5	4	5	4	1	1
25	1	1	1	1	1	1
26	3	4	4	4	3	1
27	4	4	4	4	4	3
28	4	4	2	4	2	1
29	4	4	4	4	3	1
30	5	5	5	5	3	1
31	3	1	2	3	1	1
32	4	4	3	4	4	2
33	3	3	3	4	3	1
34	2	4	3	4	3	1
35	4	4	4	4	3	1
36	3	4	4	4	2	3
37	1	4	2	4	2	3
38	4	5	5	4	4	1
39	3	3	3	3	2	1
40	4	4	4	4	2	3
41	1	1	1	1	2	1
42	2	3	3	2	2	1
43	4	3	4	5	1	1

44	3	3	3	4	3	1
45	4	4	4	4	4	1
46	3	4	3	4	2	1
47	3	3	3	3	3	1
48	4	4	4	4	5	1
49	4	4	3	4	4	2
50	4	4	4	4	2	1
51	4	4	5	5	4	1
52	4	3	3	4	2	1
53	3	3	3	4	2	1
54	3	4	3	4	2	1
55	2	2	2	2	1	1
56	4	4	3	4	2	1
57	4	3	4	4	4	1
58	3	3	3	3	2	2
59	3	2	2	4	1	2
60	4	4	4	4	2	3
61	4	4	3	4	1	2
62	3	1	2	3	1	1
63	3	3	3	4	2	3
64	4	4	3	4	3	1

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Version	64	1	3	1,97	,816	,058	,299	-1,494	,590
Age	64	55	75	63,09	5,678	,366	,299	-,997	,590
Frequency	64	1	5	2,78	1,161	-,121	,299	-,907	,590
Easy communicating	64	1	5	3,86	,852	-,517	,299	,647	,590
Friendly	64	1	5	3,88	,917	-1,401	,299	2,871	,590
Reliable	64	1	5	3,59	,886	-1,213	,299	1,890	,590
Empathic	64	1	5	3,28	,881	-,591	,299	,618	,590
Expert	64	1	5	3,41	,921	-1,162	,299	1,011	,590
Communicative	64	1	5	3,16	,895	-,043	,299	,059	,590
Serious	64	1	5	3,69	,852	-1,407	,299	2,220	,590
Feeling of talking to a person	64	1	5	2,50	1,113	,464	,299	-,517	,590
Preference	64	1	3	1,41	,729	1,481	,299	,592	,590
Valid N (listwise)	64								

9.13 Appendix M: Normality tests

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Version	Statistic	df	Sig.	Statistic	df	Sig.
Friendly	1	,369	22	,000	,733	22	,000
	2	,334	22	,000	,785	22	,000
	3	,321	20	,000	,779	20	,000
Reliable	1	,249	22	,001	,863	22	,006
	2	,326	22	,000	,819	22	,001
	3	,431	20	,000	,534	20	,000
Empathic	1	,349	22	,000	,801	22	,001
	2	,242	22	,002	,873	22	,009
	3	,345	20	,000	,723	20	,000
Expert	1	,335	22	,000	,742	22	,000
	2	,320	22	,000	,774	22	,000
	3	,321	20	,000	,749	20	,000
Communicative	1	,300	22	,000	,793	22	,000
	2	,220	22	,007	,907	22	,040
	3	,303	20	,000	,850	20	,005
Serious	1	,378	22	,000	,755	22	,000
	2	,388	22	,000	,734	22	,000
	3	,427	20	,000	,676	20	,000

a. Lilliefors Significance Correction