Reasons to continue or stop using a virtual coach for quitting smoking and increasing physical activity: A mixed-methods analysis.

Nadyne Aretz
Supervisors: Nele Albers, Willem-Paul Brinkman
Intelligent Systems department, Delft University of Technology, The Netherlands

A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

Abstract

The increase in health problems due to smoking or lack of physical activity has created the incentive to develop a virtual coach to help people with smoking cessation and increase physical activity. In this study, a mixed-methods analysis (thematic analysis, quantitative analysis, and a literature study) was carried out to find out what the reasons are for wanting to stop or continue using this virtual coach. By using triangulation, themes were found. The three themes found are the authenticity of the interaction, the caring character of the virtual coach, and the content of the interactions. Based on these themes, recommendations have been given on how to improve the virtual coach. The recommendations suggested are to improve the interactivity of the chat between the virtual coach and the participants, tailor the content of the interactions to the participant's personal needs, and make use of gamification approaches to increase motivation.

1 Introduction

In media, a reoccurring theme of experts pleading for people to start living a healthier lifestyle can be seen (e.g., [1], [2]). Not only is it important to make people more aware of healthy lifestyle choices, but it is also necessary. For example, smoking does not only influence the health of a smoker itself but also the people around them [3].

To help tackle this problem the Perfect Fit initiative was brought into life. "The initiative strives to design a method that is as well flexible as personalized to help people quit smoking and increase their physical activity" [4]. As part of this initiative, an experiment has been carried out to gather data for and test a reinforcement learning approach for motivating people to do preparatory activities for quitting smoking and increasing physical activity [5].

This experiment has been designed and implemented by members of the interactive intelligence group of the intelligent systems department of the TU Delft. A conversational agent Sam who acts as a virtual coach was implemented [6]. After a screening questionnaire and a pre-questionnaire, participants (recruited from the online recruiting platform Prolific [7]) had multiple sessions with the virtual coach Sam. After these sessions, a post-questionnaire was held. The goal of this questionnaire was to collect data about the acceptance of the agent Sam by the participants. The whole experiment structure can be found in Albers and Brinkman [5].

The challenge is now to understand and analyze the data gathered by this experiment to gain insights into what aspects influence the acceptance of the virtual coach by the participants and how this might correlate with certain user characteristics or user groups. This is done by carrying out a mixed-method analysis that tries to answer the research question: **"What are reasons for wanting to stop or continue using the virtual coach Sam?"**.

In similar research, it can be seen that the reasons for wanting to continue or stop with a virtual coach play an influence in the final acceptance of a virtual coach by people. For example, in a study where researchers looked at the development, usability, acceptability, and feasibility of a smartphone app promoting an active lifestyle in lower-educated working young adults [8]. The researchers found that people were more inclined to continue using the app because a tracker that the participants had to use in combination with the app was quite comfortable to wear. This was one of the reasons that a higher acceptance rate of the app was measured.

Research like mentioned above and others (e.g., [9]–[13]) were used to substantiate the findings of this study. The main results of the study were three themes, the authenticity of the interaction, the caring character of the virtual coach, and the content of the interactions. Based on these results recommendations were given on how the virtual coach Sam could still be improved. The recommendations suggested are to improve the interactivity of the chat between the virtual coach and the participants, tailor the content of the interactions to the participant's personal needs, and make use of gamification approaches to increase motivation.

In the remaining part of the paper the methodology used to find these results will be discussed (Section 2). Also, the findings will be explained in more detail (Section 3). In the last part of the paper, the ethics and improvements that could be made to this study will be discussed together with the conclusion (Sections 4, 5, and 6).

2 Methodology

An inductive, latent thematic analysis was carried out during this study [14]. In this study, the phases of the thematic analysis were divided into four steps: data preprocessing, coding of the data, finding and analyzing themes, and reporting the results in the form of this paper. Steps 1-3 together with how the data was collected are explained in the sections below.

2.1 Data collection

The data used for this study has been collected by Albers and Brinkman [15] through the online recruiting platform Prolific [7]. A pre-screening questionnaire was held to determine the eligibility of the participants for the study. If participants passed the pre-screening, they were invited for a pre-questionnaire. The pre-questionnaire contained questions to gather more information for example about the quitting self-identity, personality type, or physical activity identity. Parts of the data collected from the pre-screening and the pre-questionnaire were used during the quantitative data analysis. In Section 3, the exact data used is depicted in Table 3, 5, 7, and 8.

After the pre-questionnaire, participants were invited to have five sessions with the virtual coach Sam. According to the study design [5], during these sessions, participants had conversations with the virtual coach Sam where it would propose preparatory activities for smoking cessation and increasing physical activities. After that Sam would try to convince people to do the preparatory activity and ask in the next session about the experience of doing the activity.

At the end of the sessions with Sam, participants filled in a post-questionnaire. In the post-questionnaire, participants

had to answer how hard it was to do the activities and how motivated they were to do the activities. In the remaining part of the post-questionnaire, participants had to answer six questions about the acceptance of the virtual coach Sam. They had to give both free-text responses and a numerical satisfaction rating.

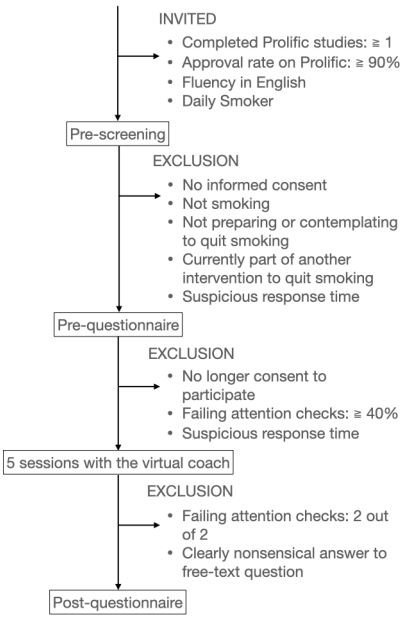

Figure 1: Basic layout of the experiment structure based on the figure "Experiment_structure.PNG" in Albers and Brinkman [5].

A basic layout of the experiment structure and reasons for exclusion of participants is illustrated in Figure 1 based on the figure "Experiment_structure.PNG" in Albers and Brinkman [5]. At the end of the experiment, after exclusion, 500 participants were left. All the data, from participants who were excluded, was removed from the pre-screening and the pre- and post-questionnaire. Lastly, the data was anonymised.

2.2 Data preprocessing

During this study, not all the data that was collected by Albers and Brinkman [5] was used. The research question of this study was based on one of the acceptance questions of the post-questionnaire and therefore, next to the data about how hard it was to do the activities and how motivated participants were, only the free-text responses and ratings concerning "Do you prefer continuing or stopping to work with the conversational agent Sam?" were used from the post-questionnaire.

To be able to start with the thematic analysis, the quantitative data had to be preprocessed. Some of the data had to be averaged and reversed. During the data collection, the participants had to fill in multiple psychometric tests. Psychometric tests are tests designed to show for example participants' personality types or cognitive abilities. The results of these tests had to be processed. To test the reliability of these psychometric tests, the Cronbach's alpha was calculated (Table 1). According to George and Mallery [16, p. 231], all the psychometric tests have a good to an excellent score.

Psychometric test	Score	Interpretation
Physical activity identity	0.89	Excellent
Quitting self-identity	0.76	Good

Table 1: Cronbach's alpha score for the psychometric tests.

2.3 Coding of the data

The first step of the thematic analysis was to get familiar with the 500 free-text responses and to write down some initial codes that could be used. The next step was to start making an initial coding scheme. In total four coding sessions were performed to code every one of the 500 free-text responses and make a semi-final coding scheme.

To introduce less bias a double coder, who also worked on a similar research question, was introduced to the coding scheme. First, the double coder was trained. For the first five free-text responses, a demonstration was given on how to code them. The next fifteen free-text responses were then coded by the double coder. After coding each response, the original coder would give feedback. After the training session, the double coder coded the remaining responses.

To measure the interrater reliability the Cohen's kappa was calculated. The value of the Cohen's kappa was 0.60 which signifies a moderate agreement between the coders [17]. To increase the agreement, the original coder and double coder discussed all the codes for the free-text responses that they disagreed on. Some codes were removed entirely from the coding scheme, while others were merged or the names were changed. This resulted in the final coding scheme which was used for the remainder of the thematic analysis and which can be found in [18].

Codes that were not used by the double coder and from which the meaning of the code was already represented by other codes were removed. For example, after discussion with the double coder, the code "empathetic" was removed because the code "understanding" already covered it.

Secondly, when the names of the codes did not correctly represent the meaning of the codes these names were changed. For example, after discussion, it became clear that the code "not authoritative enough as bot" was too vague. In the final coding scheme, it was changed to "cannot respect a bot".

Lastly, the merging of the codes happened when codes were very similar to each other and could be covered under one new code. This happened, for example, with the codes "non-judgemental" and "cannot get offended". These codes were merged under the new code "Unbiased".

2.4 Finding and analyzing themes

The next step of the thematic analysis is to find themes in the data and analyze them. In this study, the themes found were devised from the final coding scheme and the quantitative data. Finding themes and analyzing them is an iterative process. After each round of analysis, it is important to go back to the data to see if all the themes are still relevant [19].

To analyze the themes triangulation was used, more specifically method triangulation was used. Method triangulation is

using thematic analysis, quantitative data analysis, and a literature study together to find themes and explain them.

The goal of the literature study was to search for similar research to this study to use their findings for hypotheses and confirmation of the results of this study. For example papers about acceptance of virtual coaches and acceptance of technology. To search for the literature, pearl growing, snowballing and citation searching was used [20][21].

For the quantitative data analysis parts of the free-text responses, corresponding satisfaction ratings, and the data collected in the pre-screening and the pre- and parts of the post-questionnaire were used. The exact data used together with the results can be found in the 4TU research database [18]. Before the correlation analysis was started, based on the themes found, similar research, and general knowledge, hypotheses were made. Based on these hypotheses and with the help of correlation analysis, correlations were found between themes and the satisfaction rating and between themes and the other quantitative data.

To calculate the correlations, the Pearson correlation coefficient was used [22]. Before the correlation analysis could be done between the themes and the user characteristics, the data of the themes first needed to be translated to ordinal data. This was done by summing the number of codes (that were a part of this theme) present in a response. This data was then used to correlate against the user characteristic data.

In the correlation tables of Section 3, two values will be used. The R stands for the Pearson's correlation coefficient and the P stands for the probability of this R occurring when the null hypothesis is true. The R is deemed statistically significant when the $P < 0.05$ [23]. To interpret the correlation coefficient the Dancey and Reidy interpretation in Table 1 in Akoglu [22] was used.

3 Findings

In the next subsections, the findings will be explained for the themes "authenticity of the interaction", "caring character of the virtual coach", and "content of the interactions". Lastly, the findings between the numerical satisfaction rating given by the participants and the user characteristics will be explained in more detail.

3.1 Authenticity of the interaction

Name of code	Percentage
Cannot respect a bot	2%
Did not feel authentic	26%
Not interactive enough	6%
Not the right format	6%
Prefers using other services	16%
Indicates wanting to stop	44%

Table 2: Codes depicted with the percentage that the code was used in the responses covered by the theme "Authenticity of the interaction".

"I think that real professionals are way better than a scripted dialogue", (P1). One of the themes that emerged is the authenticity of the interactions with the virtual coach. Multiple participants (P1 - P13) indicated that it felt like talking to a non-human agent based on the interactions they had with the coach. Examples of the interaction not being authentic is that it felt like "a scripted dialogue" (P1) and another participant saying that it "Didn't feel I could identify with the response choices given. everything pointed to the fact it was a bot..." (P2). This theme covers around 27% of the total responses and in Table 2 are the codes depicted that this theme consists out of.

User characteristic	R	P
Satisfaction rating	-0.61	< 0.05
Extraversion	-0.02	0.68
Openness to experience	-0.08	0.06
Motivation towards doing the activities	-0.21	< 0.05
Ease of doing the activities	-0.16	< 0.05

Table 3: Pearson correlation coefficients (R) and the corresponding significance value (P) between the theme "authenticity of the interaction" and a subgroup of user characteristics.

Multiple studies (e.g., [24]–[27]) indicate that people interact with computers (or robots) in the same way as to how they would interact with other people. Thus when participants do not like the interaction with the virtual coach, one might expect that there would be a negative correlation with the satisfaction rating given by the participants. Indeed, a negative correlation of -0.61 ($P < 0.05$) was found between this theme and the numerical satisfaction rating given by the participants.

Based on the big five personality traits extraversion "indicates the extent that someone seeks interaction with their environment" and openness to experience indicates "one's willingness to try new things" [28]. The expectation would be that when participants scored higher on extraversion that there would be a negative correlation with the theme, based on that people who are extroverted value interaction more.

For openness to experience the expectation would be the other way around, namely a positive correlation between this characteristic and the theme. This was based on the assumption that when people are more willing to try something they base less value on the interaction. In the end, not a significant relationship could be found between these user characteristics and the theme, see Table 3.

As indicated in two studies where the relation between motivation, interaction, and learning was studied, interaction influences the motivation of participants and their performance [29][30]. Although these studies were mainly focused on researching learning and e-learning, their results still indicate the importance of interaction for participants. Based on these results a negative correlation would be expected between the theme and the motivation towards doing the activities and the ease of doing the activities. As shown in table 3, in both cases

a weak negative correlation was found.

Recommendations. Based on the results of the literature study and correlation analysis, it became clear that the authenticity of the interaction is an interesting factor to look at. One way to increase the interaction between the virtual coach and the participants is to focus more on the interactivity (or two-way interaction) between the virtual coach and the participants [25][31]. Another way could be to focus more on the social rules of human-human interaction to improve the interaction between the virtual coach and the participants [25].

3.2 Caring character of the virtual coach

Name of code	Percentage of theme
Comforting	9%
Engaging	7%
Friendly	13%
Motivating	47%
Patient	1%
Polite	2%
Supporting	9%
Unbiased	6%
Understanding	7%

Table 4: Codes depicted with the percentage that the code was used in the responses covered by the theme "Caring character of the virtual coach".

"...being in contact with Sam gave a feeling of support it helped a great deal towards the motivation to carry on", (P14). The second theme that emerged based on the responses and the final coding scheme is the positive influence the caring character of the virtual coach had on the participants. Multiple participants (P14 - P23) indicated that they felt understood and supported while working with Sam due to his caring characteristics. This theme covers around 20% of the total responses and in Table 4 are the codes depicted that this theme consists out of.

Quantitative data	R	P
Satisfaction rating	0.29	< 0.05
Participant's age	0.10	< 0.05
Motivation towards doing the activities	0.23	< 0.05

Table 5: Pearson correlation coefficients (R) and the corresponding significance value (P) between the theme "Caring character of the virtual coach" and a subgroup of user characteristics.

In several studies where the social connection between virtual agents and participants was studied, it became clear that a caring and social agent was more readily accepted and resulted in long-term usage by the participants [32]–[34]. This would suggest a positive correlation between the satisfaction rating and this theme. In table 5, it is shown that a positive correlation was found between the theme and the satisfaction rating given by the participants. This signifies a weak relationship.

In one study, the researchers looked specifically at the acceptance of a relational agent for older adults [32]. They indicated that the social nature of the agent helped with the acceptance. Therefore, it was interesting to look at if any relationship could be found between this theme and the participant's age. In the end, a weak relationship was found, see table 5.

In the same study, the researchers also indicated that an increase in motivation was seen. Therefore, the expectation for this study was that there would be a positive correlation between this theme and the motivation towards doing the activities. Indeed a weak relationship of 0.23 ($P < 0.05$) was found between the two.

Recommendations. To even further increase the caring character of the virtual coach an appropriate use of empathetic behavior could be implemented [25][35][36]. One thing to watch out for is the repetitiveness in the interaction between the virtual coach and the participants [32].

3.3 Content of the interactions

Name of code	Percentage of theme
Clear explanations	4%
Guidance	8%
Helpful	56%
Informative	8%
Interesting	12%
Proposes good ideas or activities	12%

Table 6: Codes depicted with the percentage that the code was used in the responses covered by the theme "Content of the interactions".

"It gave me helpful tips that I can use and will help motivate me to quit smoking", (P26). The third theme that emerged based on the responses and the final coding scheme is the content of the interactions with the virtual coach. Multiple participants (P24 - P47) indicated that the ideas and activities given to them during the interactions helped them and that they felt guided. This theme covers around 45% of the total responses and in Table 6 are the codes depicted that this theme consists out of.

Quantitative data	R	P
Satisfaction rating	0.37	< 0.05
Physical activity identity	0.07	0.12
Ease of doing the activities	0.11	< 0.05
Motivation towards doing the activities	0.24	< 0.05

Table 7: Pearson correlation coefficients (R) and the corresponding significance value (P) between the theme "Content of the interactions" and a subgroup of user characteristics.

In a study where a multi-stakeholder approach to eHealth development was researched, the researchers found that simplicity and guidance were valued by the participants [37]. Therefore, for this study, a positive correlation between the theme and the satisfaction rating given by the participants was suspected. After analysis, a weak, but significant relation was found between the two, see Table 7.

Based on the idea that part of the activities proposed by the virtual coach are physical activities, a positive correlation between the physical activity identity and the theme was suspected. As depicted in Table 7, not a significant relationship could be found.

Multiple responses of the participants indicated that the feeling of being guided motivated them. For example, the statement from P31 "Sam was a very helpful guide and motivated me and made me think of good habits to get away from the idea of smoking". Other participants liked that everything was explained clearly, like P48 "Sam explains clearly its minds, the activities and the final goals of each step".

Based on these responses a correlation was suspected between this theme and the user characteristics motivation towards doing the activities and ease of doing the activities. As shown in Table 7, a weak relationship was found between the theme and the ease of doing the activities and the theme, and a weak relationship was found between motivation towards doing the activities.

Recommendations. Based on the literature three recommendations can be given. Firstly, tailoring would be suggested [38]. This could be done by making the messages of the virtual coach more personal for the participants. For example, by calling the participants by their name at the beginning of a message or by content matching. With content matching, the content of the interactions is adapted based on user characteristics.

The second and third recommendations suggest making use of the mechanisms of recognition of ability and visibility of the efforts to quit smoking. These mechanisms could be part of a gamification strategy. In a similar study, participants responded that an award gifted to them via the platform that they were using gave them a confidence boost [39]. The same study reported also that participants were motivated by being able to see the progress of their efforts to quit smoking on a progress page.

3.4 Satisfaction rating

User characteristic	R	P
Participant's age	-0.03	0.56
Quitting self-identity	0.07	0.11
Motivation towards doing the activities	0.47	< 0.05
Ease of doing the activities	0.32	< 0.05

Table 8: Pearson correlation coefficients (R) and the corresponding significance value (P) between the satisfaction rating given by the participants and a subgroup of user characteristics.

Next to looking into correlations between themes and user characteristics, this study also looked at possible correlations between the numerical satisfaction rating that participants had to give and a subgroup of user characteristics. The results are depicted in Table 8.

Multiple studies (e.g., [40], [41]) researching similar eHealth systems, reported that no direct influence of age could be determined for the acceptance of an eHealth system. In this study, indeed a significant relationship between age and the satisfaction ratings could not be found, see Table 8.

In a study where the use of eHealth systems based on the motivation to quit smoking of participants was studied, it became apparent that motivated participants make more use of eHealth technology than unmotivated participants [42]. Based on this finding, a positive correlation between the satisfaction rating and the quitting self-identity would be expected, as the quitting self-identity is an indication of the intention to quit smoking [43]. However, in this study, a significant relationship could not be found between the satisfaction rating and the quitting self-identity, see Table 8.

Another expectation would be that participants that are motivated to quit would also be more motivated towards doing the activities. Therefore a positive correlation would be expected between the satisfaction rating and motivation towards doing the activities. A positive correlation of 0.47 ($P < 0.05$) was found, which signifies a moderate relationship.

In another study, it was found that the effort expectancy influences the behavior intention to use the eHealth system, which the researchers were studying, greatly [44]. The effort expectancy is "the degree of ease associated with the use of the system" [45]. Therefore, for this study, a closer look was taken at the ease of doing the activities. In the end, a weak correlation between the satisfaction rating and the ease of doing the activities was found.

Recommendations. Based on the findings, it would be interesting to look into why participants were motivated towards doing the activities and what influences exactly the ease of doing the activities. By answering these two questions, improvements can be found for the virtual coach.

4 Responsible research

During this study, multiple attempts have been made toward responsible research. The data used for the analysis has been collected by Albers and Brinkman. The process of how this data was gathered has been documented in detail [5] and the Human Research Ethics Committee of Delft University of Technology granted ethical approval for the research (Letter of Approval number: 1523) [15]. Also, during the experiment on multiple occasions participants have been asked for their informed consent. Participants could stop with the experiment at any moment.

To recruit a big group of participants, Prolific was used [7]. Prolific is an online recruiting platform. One of the advantages of using Prolific was that it helps with gathering a diverse participant pool. In this way, participants from all over the world and different cultures, age groups, and genders could participate in the experiment.

One of the drawbacks of using Prolific is that people were paid for taking part in the experiment. This could increase the motivation of the participant to do well in the experiment and therefore positively enhance their answers. Another side of this could be that people filled in their questions with minimum effort. Both of these behaviors could influence the trustworthiness of the answers. In an ideal situation, participants would take part in an experiment without a monetary benefit. However, in that case, it would be hard to find enough people willing to participate in an experiment.

Another side of responsible research is to be transparent about how the research was carried out. This has as goal to make it easier to reproduce the research. As part of this approach, in the methodology, all the steps taken during this study have been described. Also, all the intermediate and final results together with this paper have been made publicly accessible via the 4TU research database [18]. The 4TU research database ensures that the storage, sharing, and publishing of the research data conforms to the FAIR data principles [46].

To reduce personal bias, investigator triangulation, as well as method triangulation, has been performed. As part of the investigator triangulation, double coding has taken place as described in section 2.2. To reduce personal bias during the data analysis, triangulation of the thematic analysis, quantitative data analysis, and the literature study was used.

During the literature study, attention was paid to the sources of the literature. Next to that, an attempt was made to use multiple sources to increase the trustworthiness of the arguments used in the literature. It was also important to check if the literature used was still relevant, by checking if there were additional results found or if certain results were refuted.

To watch out for data dredging [47], multiple measures were taken during the correlation analysis and the reporting of the results. During the correlation analysis, this was done by first thinking of some possible hypotheses which could result in correlations. These hypotheses were based on similar research, general knowledge, and indications given in the free-text responses. This was to counter the possibility of data mining.

When reporting the results, not only results have been shown that are statistically significant, but also results are shown that are not statistically significant. If only the statistically significant results were shown, then this is a form of cherry-picking which is bad practice.

5 Discussion

As in any study, this study has limitations that could have influenced the reliability and quality of the results. As mentioned in Section 4, the use of an online recruiting platform like Prolific could have influenced the data due to people receiving a monetary benefit for partaking in the study. Another factor that could have influenced the data is that the participants were recruited online, so they had already some affinity with technology. This could have biased the participant's view towards a virtual coach. Both these factors could have influenced the reliability of the results.

During this study, the focus on reducing researcher bias was important. Firstly, a second coder was asked to code the free-text responses, as described in Section 2.3. Although this already helped reduce bias, ideally, if time had permitted, another round with a third coder could have even reduced the bias more. Secondly, triangulation was used for analysing the data. This was another measure to reduce bias by using the results of the coding of the data, the quantitative data (user characteristics), and literature. With the reduction of bias, the quality of this study was safeguarded.

Next to these limitations, time was certainly of influence during this study. If the study would have taken place over a longer time more themes could have been worked out in more detail. For example, a more detailed analysis could have been done, and as a next step regression analysis could have been performed. With as result that the quality of the results would have improved.

Lastly, the challenge during the coding process was to not get lost in details but to keep clear and concise. To improve this study, it would have been better to keep the codes attributed to the free-text responses more general and described more clearly. This would have helped the double coder to get a better grasp of the meaning of the codes. Next to that, it would also have helped with finding themes and therefore the quality of the results would have been improved.

6 Conclusion

To conclude, based on the free-text and numerical satisfaction rating responses from the participants, the quantitative data and the literature multiple themes were found during the mixed-methods analysis. These themes helped answer the question: "What are the reasons for wanting to stop or continue using the virtual coach Sam?". The method used for this study was triangulation where a thematic analysis, quantitative data analysis (correlation analysis), and a literature study were used.

In the end, three themes were found: authenticity of the interaction, caring character of the virtual coach, and content of the interactions. Based on the correlations found between these themes and the user characteristics and the correlations found between the satisfaction rating of the participants and user characteristics, multiple recommendations were given for improving the virtual coach.

The following recommendations were given:

- Improve the interactivity (two-way interaction) between the virtual coach and the participants.
- Focus more on the social rules of human-human interaction to improve the interaction between the virtual coach and the participants.
- Increase the caring character of the virtual coach by appropriate use of empathetic behavior.
- Watch out for repetitiveness in the conversation between the virtual coach and the participants.
- Pay more attention to tailoring the content of the interaction to the participant.
- Incorporate a mechanism to recognize the ability of the participant of completing certain tasks.

- Incorporate a mechanism to visualize the participants' efforts to quit smoking.

For future work it would be interesting to take a closer look at the reasons why participants were motivated towards doing the activities and what influenced the ease of doing the activities.

Acknowledgement

Special thanks to Nele Albers and Willem-Paul Brinkman for guiding the process of this study and giving feedback where necessary. Also thanks to Omar Sheasha and Jaap Dechering for preprocessing most of the quantitative data and for calculating the Cronbach's alpha scores for the user characteristic tests. For the double coding and giving multiple times feedback during this study, a thank you is given to Arsen Ekinci and to Mahira Ali.

References

[1] NOS, *Overgewicht, roken en drinken in drie jaar nauwelijks gedaald*, https://nos.nl/artikel/2420676-overgewicht-roken-en-drinken-in-drie-jaar-nauwelijks-gedaald, [Accessed: 2022-04-22], 2022.

[2] N. en Co, *Gezonde levensstijl is nu echt noodzakelijk*, https://www.nporadio1.nl/nieuws/binnenland/840845ad-996e-4cd5-82df-d2d7f877be1b/obesitas, [Accessed: 2022-04-22], 2020.

[3] S. Voormolen, *Vijftig rokers doden één niet-roker*, https://www.nrc.nl/nieuws/2020/03/17/vijftig-rokers-doden-een-niet-roker-a3994021, [Accessed: 2022-04-22], 2020.

[4] Commit2Data, *Perfect fit - targeting key risk factors for cardiovascular disease in at-risk individuals using a personalized and adaptive approach*, https://commit2data.nl/projecten/perfect-fit, [Accessed: 2022-04-22], 2020.

[5] N. Albers and W.-P. Brinkman, *Perfect fit - experiment to gather data for and test a reinforcement learning-approach for motivating people*, May 2021. DOI: 10.17605/OSF.IO/K2UAC. [Online]. Available: osf.io/k2uac.

[6] N. Albers, *Reinforcement learning-based persuasion for a conversational agent to support behavior change*, https://github.com/PerfectFit-project/virtual_coach_rl_persuasion_algorithm#reinforcement-learning-based-persuasion-for-a-conversational-agent-to-support-behavior-change, [Accessed: 2022-04-22], 2021.

[7] Prolific, *Quickly find research participants you can trust*, https://www.prolific.co.

[8] D. Simons, I. De Bourdeaudhuij, P. Clarys, K. De Cocker, C. Vandelanotte, and B. Deforche, "A smartphone app to promote an active lifestyle in lower-educated working young adults: Development, usability, acceptability, and feasibility study," *JMIR Mhealth Uhealth*, vol. 6, no. 2, e44, Feb. 2018, ISSN: 2291-5222. DOI: 10.2196/mhealth.8287. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/29463491.

[9] A. Tironi, R. Mainetti, M. Pezzera, and N. A. Borghese, "An empathic virtual caregiver for assistance in exer-game-based rehabilitation therapies," in *2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH)*, 2019, pp. 1–6. DOI: 10.1109/SeGAH.2019.8882477.

[10] F. Mostajeran, F. Steinicke, O. J. Ariza Nunez, D. Gatsios, and D. Fotiadis, "Augmented reality for older adults: Exploring acceptability of virtual coaches for home-based balance training in an aging population," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–12, ISBN: 9781450367080. [Online]. Available: https://doi-org.tudelft.idm.oclc.org/10.1145/3313831.3376565.

[11] H. Wang, D. Tao, N. Yu, and X. Qu, "Understanding consumer acceptance of healthcare wearable devices: An integrated model of utaut and ttf," *International Journal of Medical Informatics*, vol. 139, p. 104 156, 2020, ISSN: 1386-5056. DOI: https://doi.org/10.1016/j.ijmedinf.2020.104156. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1386505619311438.

[12] J. Zhang, Y. J. Oh, P. Lange, Z. Yu, and Y. Fukuoka, "Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: Viewpoint," *J Med Internet Res*, vol. 22, no. 9, e22845, Sep. 2020, ISSN: 1438-8871. DOI: 10.2196/22845. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/32996892.

[13] C. Free, R. Whittaker, R. Knight, T. Abramsky, A. Rodgers, and I. Roberts, "Txt2stop: A pilot randomised controlled trial of mobile phone-based smoking cessation support," *Tobacco control*, vol. 18, pp. 88–91, May 2009. DOI: 10.1136/tc.2008.026146.

[14] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, pp. 77–101, Jan. 2006. DOI: 10.1191/1478088706qp063oa.

[15] N. Albers, M. A. Neerincx, and W.-P. Brinkman, "Acceptance of a Virtual Coach for Quitting Smoking and Becoming Physically Active: Dataset," May 2022. DOI: 10.4121/19934783.v1. [Online]. Available: https://data.4tu.nl/articles/dataset/Acceptance_of_a_Virtual_Coach_for_Quitting_Smoking_and_Becoming_Physically_Active_Dataset/19934783.

[16] D. George and P. Mallery, "Spss for windows step by step: A simple guide and reference. 11.0 update (4th ed.)," in Boston: Allyn & Bacon., 2003.

[17] M. McHugh, "Interrater reliability: The kappa statistic," *Biochemia medica : časopis Hrvatskoga društva medicinskih biokemičara / HDMB*, vol. 22, pp. 276–82, Oct. 2012. DOI: 10.11613/BM.2012.031.

[18] N. Aretz, N. Albers, and W.-P. Brinkman, *Data and results for the bachelor thesis: Reasons to continue*

or stop using a virtual coach for quitting smoking and increasing physical activity: A mixed-methods analysis. Jun. 2022. DOI: 10.4121/19936487. [Online]. Available: https://data.4tu.nl/articles/dataset/Data_and_results_for_the_bachelor_thesis_Reasons_to_continue_or_stop_using_a_virtual_coach_for_quitting_smoking_and_increasing_physical_activity_A_mixed-methods_analysis_/19936487/0.

[19] J. Caulfield, *Een introductie tot thematische analyses met voorbeelden*, https://www.scribbr.nl/onderzoeksmethoden/thematische-analyse/, [Accessed: 2022-04-24], 2021.

[20] L. Guides, *How to search for information?* https://buas.libguides.com/c.php?g=670074&p=4776470, [Accessed: 2022-05-30].

[21] TUlib, *Making a search plan*, https://tulib.tudelft.nl/searching-resources/making-a-search-plan/, [Accessed: 2022-04-24].

[22] H. Akoglu, "User's guide to correlation coefficients," *Turkish Journal of Emergency Medicine*, vol. 18, no. 3, pp. 91–93, 2018, ISSN: 2452-2473. DOI: https://doi.org/10.1016/j.tjem.2018.08.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2452247318302164.

[23] MedCalc, *Correlation coefficient*, https://www.medcalc.org/manual/correlation.php, [Accessed: 2022-05-30].

[24] C. Nass and Y. Moon, "Machines and mindlessness: Social responses to computers," *Journal of Social Issues*, vol. 56, pp. 81–103, Mar. 2000. DOI: 10.1111/0022-4537.00153.

[25] M. de Graaf, S. Allouch, and J. A. Van Dijk, "What makes robots social?: A user's perspective on characteristics for social human-robot interaction," Oct. 2015. DOI: 10.1007/978-3-319-25554-5_19.

[26] K. Lee, N. Park, and H. Song, "Can a robot be perceived as a developing creature?" *Human Communication Research*, vol. 31, pp. 538–563, Jan. 2006. DOI: 10.1111/j.1468-2958.2005.tb00882.x.

[27] s. Kwak, Y. Kim, E. Kim, C. Shin, and K. Cho, "What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot," Aug. 2013, pp. 180–185. DOI: 10.1109/ROMAN.2013.6628441.

[28] A. Lim, *The big five personality traits. simply psychology.* www.simplypsychology.org/big-five-personality.html, [Accessed: 2022-06-05].

[29] R. Liu and Y.-L. Chiang, "Who is more motivated to learn? the roles of family background and teacher-student interaction in motivating student learning," *The Journal of Chinese Sociology*, vol. 6, Dec. 2019. DOI: 10.1186/s40711-019-0095-z.

[30] S. Essam and J. Alammary, "The impact of motivation and social interaction on the e-learning at arab open university, kingdom of bahrain," *Creative Education*, vol. 4, pp. 21–28, Nov. 2013. DOI: 10.4236/ce.2013.410A004.

[31] G. Kreps and L. Neuhauser, "New directions in ehealth communication: Opportunities and challenges. patient education and counseling, 78(3), 329-336," *Patient education and counseling*, vol. 78, pp. 329–36, Mar. 2010. DOI: 10.1016/j.pec.2010.01.013.

[32] T. Bickmore, L. Caruso, K. Clough-Gorr, and T. Heeren, ""it"s just like you talk to a friend' relational agents for older adults," *Interacting with Computers*, vol. 17, pp. 711–735, Dec. 2005. DOI: 10.1016/j.intcom.2005.09.002.

[33] T. Bickmore and R. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Trans. Comput.-Hum. Interact.*, vol. 12, pp. 293–327, Jun. 2005. DOI: 10.1145/1067860.1067867.

[34] T. Bickmore, "Relational agents: Effecting change through human-computer relationships," Mar. 2003.

[35] J. Klein, Y. Moon, and R. Picard, "This computer responds to user frustration:: Theory, design, and results," *Interacting with Computers*, vol. 14, pp. 119–140, Nov. 2001. DOI: 10.1016/S0953-5438(01)00053-4.

[36] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Empathic robots for long-term interaction," *International Journal of Social Robotics*, vol. 6, pp. 329–341, Aug. 2014. DOI: 10.1007/s12369-014-0227-1.

[37] L. Breeman, M. Keesman, D. Atsma, *et al.*, "A multi-stakeholder approach to ehealth development: Promoting sustained healthy living among cardiovascular patients," *International Journal of Medical Informatics*, vol. 147, p. 104 364, Dec. 2020. DOI: 10.1016/j.ijmedinf.2020.104364.

[38] A. Dijkstra, "Personalization/computer-tailoring in persuasive technology: Tailoring ingredients target psychological processes," p. 7,

[39] L. Struik, J. Bottorff, N. Baskerville, and J. Oliffe, "The crush the crave quit smoking app and young adult smokers: Qualitative case study of affordances," *JMIR mHealth and uHealth*, vol. 6, e134, Jun. 2018. DOI: 10.2196/mhealth.9489.

[40] S. Hennemann, M. Beutel, and R. Zwerenz, "Ready for ehealth? health professionals' acceptance and adoption of ehealth interventions in inpatient routine care," *Journal of Health Communication*, vol. 22, pp. 274–284, Feb. 2017. DOI: 10.1080/10810730.2017.1284286.

[41] L. Liu, A. Miguel Cruz, A. Rios Rincon, V. Buttar, Q. Ranson, and D. Goertzen, "What factors determine therapists' acceptance of new technologies for rehabilitation – a study using the unified theory of acceptance and use of technology (utaut)," *Disability and rehabilitation*, vol. 37, no. 5, pp. 447–455, 2015, ISSN: 0963-8288. DOI: 10.3109/09638288.2014.923529.

[42] B. Borrelli, Y. K. Bartlett, E. Tooley, C. J. Armitage, and A. Wearden, "Prevalence and frequency of mhealth and ehealth use among us and uk smokers and differences by motivation to quit," *J Med Internet Res*, vol. 17, no. 7, e164, Jul. 2015, ISSN: 1438-8871.

DOI: 10.2196/jmir.4420. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/26149323.

[43] B. Van Den Putte, M. Yzer, M. Willemsen, and G.-J. de Bruijn, "The effects of smoking self-identity and quitting self-identity on attempts to quit smoking," *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, vol. 28, pp. 535–44, Sep. 2009. DOI: 10.1037/a0015199.

[44] K. Ghorai and P. Ray, "Factors affecting the acceptance of mobile based multi-feature service for smoking cessation using utaut," Apr. 2019.

[45] "User acceptance of information technology: Toward a unified view," *MIS Quarterly*, vol. 27, no. 3, pp. 425–478, 2003, ISSN: 02767783. [Online]. Available: http://www.jstor.org/stable/30036540 (visited on 06/04/2022).

[46] Force11, *The fair data principles*, https://force11.org/info/the-fair-data-principles/, [Accessed: 2022-06-08.

[47] R. Awati, *Data dredging (data fishing)*, https://www.techtarget.com/searchdatamanagement/definition/data-dredging, [Accessed: 2022-06-01.