Master thesis - Feedback presentation for physical activity coaching: by an embodied conversational agent or by a simple e-mail?

Adri Wiekens

UNIVERSITEIT TWENTE.

Contents

1 Introduction 9
- 1.1 Behavior modification . . . 9
- 1.2 Thesis outline . . . 9
- 1.3 Goals . . . 11
- 1.4 Hypothesis . . . 11
- 1.5 Measurements of success . . . 12

2 Background 13
- 2.1 SMARCOS ARTEMIS . . . 13
- 2.2 Goals of SMARCOS . . . 13
- 2.3 Philips DirectLife . . . 14
  - 2.3.1 Goals of DirectLife . . . 14
- 2.4 System details . . . 15
- 2.5 Integration with the projects . . . 15

3 Behavior change support system and embodied agents 17
- 3.1 Health assistance systems . . . 17
- 3.2 Use of mobile devices . . . 17
- 3.3 Emotional influencing . . . 17
- 3.4 Behavior adjustment psychology . . . 18
- 3.5 Transtheoretical Model . . . 19
  - 3.5.1 Pre-contemplation(not ready) . . . 19
  - 3.5.2 Contemplation(getting ready) . . . 19
  - 3.5.3 Preparation(ready) . . . 20
  - 3.5.4 Action . . . 20
  - 3.5.5 Maintenance . . . 20
  - 3.5.6 Stable behavior . . . 21
  - 3.5.7 Relapse . . . 21
  - 3.5.8 Stage transitions . . . 21
- 3.6 Transition . . . 21
  - 3.6.1 Transitional processes . . . 22
- 3.7 Weapons of persuasion . . . 24
  - 3.7.1 Reciprocation . . . 24
  - 3.7.2 Commitment and consistency . . . 24
  - 3.7.3 Social proof . . . 24
  - 3.7.4 Liking . . . 24
  - 3.7.5 Authority . . . 25
  - 3.7.6 Scarcity . . . 25
- 3.8 Virtual embodied agents in behavior change support systems . . . 25
- 3.9 Virtual embodied agents vs pure text systems . . . 25
- 3.10 Putting theory to practice . . . 26

4 System description 27
4.1 Comparing the consumer system with the experiment system 27
4.2 Key to success 29
4.3 BML 29
4.3.1 Text use model 29
4.3.2 Emotion use model 30
4.4 Elckerlyck 30
4.5 Feedback types 31
4.6 Components 31
4.7 Total experiment system setup 31
4.7.1 WebServer 32
4.7.2 User information extraction 32
4.7.3 Data transfer 33
4.7.4 E-mail server 33
4.8 input 33
4.8.1 Database link 34
4.9 Script generator further examined 34
4.9.1 output 35
4.9.2 Text model 38
4.9.3 Snippet fetching algorithm 38
4.9.4 BML results from scriptgenerator 39

5 Experiment 41
5.1 Procedure 41
5.1.1 Required system results 42
5.1.2 Statistical survey 42
5.2 Measures 43
5.2.1 Likeability 43
5.2.2 Perceived intelligence 43
5.2.3 Trust/fear 44
5.2.4 Effectiveness 44
5.2.5 Persuasiveness 44
5.2.6 Relapse sensitivity 44
5.2.7 Quality of coaching 45
5.2.8 Intention to use longer 45
5.3 Participants 45
5.4 Using the system as a client 47
5.5 E-mail response 49

6 Experiment results 51
6.1 Feedback sessions 51
6.1.1 Quality over quantity 51
6.2 Results 51
6.2.1 Likability 51

6.2.2 Perceived intelligence . . . 52
6.2.3 Fear . . . 53
6.2.4 Intention to use longer . . . 53
6.2.5 Quality of coaching . . . 53
6.2.6 General adherence and interest . . . 53
6.2.7 Performance of users . . . 54
6.3 User experiences . . . 54
6.4 Interpretation of results . . . 55

7 Discussion 57
7.1 Implication for hypothesis . . . 57
7.2 Recommendations . . . 57
7.2.1 Success of method . . . 57
7.3 Useful techniques that are unusable for practical reasons . . . 57
7.4 Techniques that are unusable for this concept . . . 58
7.4.1 What influences can invoke relapse? . . . 58
7.4.2 Commitment . . . 58
7.4.3 Recommitment . . . 59
7.5 Future work . . . 59
7.6 Opportunities . . . 60

8 Conclusion 61
8.1 Final thoughts . . . 61

9 Acknowledgments 62
9.1 Special thanks . . . 62

References 63

10 Appendix 66
10.1 Experiment figures . . . 86
10.1.1 Half way point . . . 86
10.1.2 Sixth week survey . . . 87
10.1.3 Performance of E-mail conversation . . . 87
10.1.4 Performance of E-mail conversation . . . 88
10.1.5 Performance of The embodied agent . . . 88

Abstract

"In this study we compare two ways to present feedback by a digital coaching system that helps clients to improve or sustain their physical activity level in daily life. One way is by means of a simple e-mail sent to the user; the other way is by means of an embodied conversational character that reads the feedback message aloud. The comparison is conducted by means of an experiment with 20 test subjects who used the coaching system, monitoring their physical activity by means of an accelerometer, for a period of 6 weeks. 10 of them received feedback by an embodied conversational agent, from here on refered to as ECA, and 10 by means of e-mail.

We measured effectiveness as well as perception of the feedback presentation by means of validated questionnaires. The overal conclusion was that there was a significant difference in measured quality of coaching in favor of the coaching system that made use of ECA coaching. The results also included that clients found the ECA more likable than the e-mail feedback. The perceived intelligence of the coach is significantly higher when feedback is presented by an ECA than when presented by e-mail. It seems that feedback presentation by an ECA has a positive effect on adherence. The ECA makes the client feel personally addressed by the coach. This was also demonstrated with a negative response from participants when the ECA mispronounced the participants name, showing that it is important that the ECA pronounces the name of the client in a correct way.

1 Introduction

There are many ways in which a person can change his or her behavior to a behavior that is closer to what he or she desires, and there are quite a lot of people who wish to change their behavior to increase their own health. In my thesis I will discuss the proceedings and results of an experiment that was conducted in order to produce and analyse a system that is aimed at behavior modifcation, which should improve the health of the people that use the system involved in the experiment. We wish to improve their health by sufficient physical activity which in turn has been shown to improve the general well being of a person[31].

This experiment was set up in cooperation with an existing system, namely the system used in Philips DirectLife. With DirectLife the participants make use of a behavior change support system, as seen before in other experiments[23]. And in this particular version they are introduced to an agent for health behavior change, similar systems have been demonstrated before by Timothy Bickmore[8]. This specific experiment is not just focussed on creating a behavior change support system, but to improve the quality of presentation method that is used in this behavior change support system. Can we improve the quality of the DirectLife program by giving a human like face to the system that is guiding the people following this program?

1.1 Behavior modification

The improvement of the quality of life is achieved in the system by changing the behavior of people, and in order to achieve the desired changes, we need to figure out what the methods we can use and what tactics to use to change the behavioral patterns. The basic concepts for these tactics and weapons are predominantly taken from the field of psychology. We want to influence peoples way of thinking by making use of their current state of mind, if a person wants to change we can support him or her to achieve this or if a person has not yet made up his or her mind then we can give information to help the person make a more informed decision. Changing the participants self image and making use of the expectancies of others is key to change the behavior of a person willing to change[25]. But at the moment the method that is mainly used, DirectLife, to achieve this goal is informing the users of the system by pure text, within this experiment we are going to look whether switching from pure text by e-mail to a form of communication that people can relate to better will result in a better behavior change support system. So we perform an experiment in which we are going to have the same system communicate to its users by two different means, one being the e-mail communication platform, while the other is an embodied conversational agent.

1.2 Thesis outline

In this section I will give a short introduction and overview for each of the following sections of the complete thesis.

2. **Background**
   Firstly I want to give a short outline of all the background information that wwere relevant for this research project. This will introduce the origin of this research, the practical application of it as well as the related projects of this experiment.

3. **Behavior change support system and embodied agents**
   This specific research combines several other fields of study, in this third chapter I will discuss which other papers and studies are used for assumptions and starting points in order to construct this specific study. It treats behavior change support systems and previous experiments that have been done in this field of development such as the many experiments performed by Timothy Bickmore. To better give an idea on how these systems work on human psychology I will also discuss several papers that have served as a base for the DirectLife system and the current experiment version, such as the works of Prochaska[15]. One of the last central themes in this research is the virtual embodied agent, in order to see what other behavior change support systems have done with embodied agents to achieve their goal.

4. **System description**
   This chapter deals with the technical details and describes the entire system set up. It first starts by introducting the system as it is currently running in a consumer environment. Then it gives a description of how the system can be adjusted and which components need to be added and/or modified in order to perform a comparing experiment between the two distinct forms of user interaction i.e. e-mail communication and embodied agent communication. After giving a full layout of the entire system I will further examine the more crucial components which are unique for this experiment in full detail, which should give enough information to understand how the system is able to perform the task at hand in the experiment.

5. **Experiment**
   This chapter will deal with all the elements of the experiment. I will deal with the procedures of the experiment setup, in which I describe how the experiment is setup, the measures I use to quantify my results, the participants of the study.

   In the following paragraphs I will then continue to explain how the data from the users is handled and processed into the proper form of input that can result into the needed human machine interaction.

6. **Results**
   To draw any conclusions from the experiment the last set of paragraphs in this chapter describes all the figures that resulted from the experiment and the forms that were provided to the participants. At first I will present the results in just numbers, ordered into two seperate groups for the e-mail engine and the embodied agent users. The following paragraphs will then discuss firstly the numbers should statiscially be processed, and then how the numbers and results can be interpreted and lastly what consequences these results have for the hypothesis that was put forward.

7. **Discussion**
   This chapter is used to discuss the outcome of the experiment, and put forwards some points of discussion on future work. Aside from that I will defend the elements that were purposely left out for practical reasons or that did not fit the profile of this experiment.

8. **Experiment conclusion**
   The overal conclusion of the research, discussing the found results.

9. **Acknowledgments**
   A graduation thesis cannot be completed with some assistance, here I will thank everybody who assisted me in completing it.

1.3 Goals

The aim of this research is to compare the two methods within the behavior change support system which have the same factual content but a different form of presentation in performing the function. On one hand we have the system where it makes use of pure text feedback in the form of e-mail interaction, while on the other hand the system is expanded with an additional embodied agent who will bring the messages to the user of the system. This will lead us to the following research question: is a virtual embodied agent better(as defined in the next subsection) than e-mail communication when trying to alter human behavior, within the setting of DirectLife? An experiment involving a research version of the current direct life system will be used to find out what the answer to this question is.

1.4 Hypothesis

The hypothesis that I wish to put forward and put to the test is that a system with an embodied conversational agent will perform better in coaching than a system that responds with coaching e-mail. In order to see what is "better" in this context we need to define several criteria on which we compare both systems and what units of measurement we then are going to use to see which of the two systems scores higher on each of the criteria. For better in this context I will use the following criteria:

1. Effectiveness - How effective are both systems, this is measured by which method has the higher increase in physical activity level are participants that recieve agent coaching more active than participants who recieve coaching by e-mail.

2. Persuasiveness - How well can the system persuade the user to follow the advice that it is giving to the user. This can be measured by how close are participants to reaching their set weekly objectives.

3. Relapse sensitivity - This can be measured by the percentage of people who do not finish the program or simply fail the set objectives of the system alround before quitting.

4. Attractiveness - How pleasant is the system to use, this is measured by ratings from the users by inquiry.

5. Perceived intelligence - Is the agent that the people are communicating with convincing as an intelligent person to talk to.

6. Trust/fear - Is there any bond forming between system and user, if there is a stronger forming of coach and coachee the system could possibly work more effectively this could be demonstrated by how much the person trusts the agent, this is measured by questionaire.

7. Quality of coaching - How do users perceive the quality of the coaching they have recieved. This will be tested by questionaire.

8. Intention to use longer - Are the participants in the experiment planning to use the system after the experiment has ended.

1.5 Measurements of success

The method of falsification will be an experiment with a small sample size (20 people); followed by an evaluation with the help of the coach quality check feedback forms, as well as personal interviews at the end of the project. The interviews will use a standardized set of questions to determine how well each feedback method scores on the earlier mentioned, as well as statistics that are generated by use of the system, such as how often people connect to the system and how often people view their received feedback.

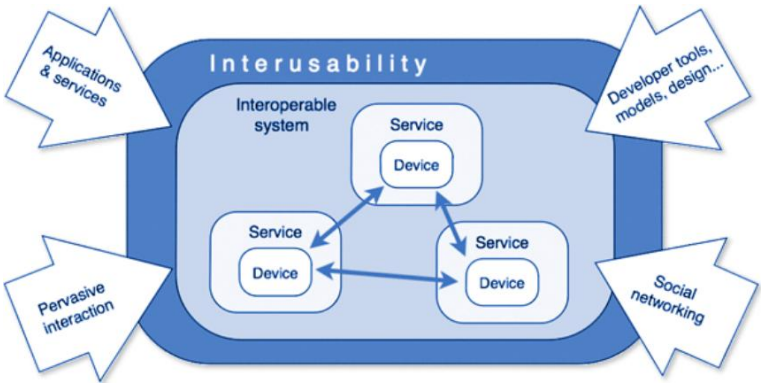


Figure 1: Smarcos overview

2 Background

This experiment is not an individual new piece of research, it is part of a collection of projects, researches and experiments that all involve interconnected embedded systems and immersive environments. The idea of all these systems is to make the gap between humans smaller and make the technology they use as smart as possible and to make technology adaptable to unexpected events. The name of the encompassing project is SMARCOS, and Royal Philips Electronics Incorporated is one of the companies participating in this project along with the University of Twente.

2.1 SMARCOS ARTEMIS

According to mission statement of SMARCOS (short for smart composites) the aim of the project is to help users of interconnected embedded systems by ensuring their interusability. The concept of this mission is derived from that fact that nowadays, many products connect with web services. SMARCOS allows devices and services to communicate in UI level terms and symbols, exchange context information, user actions, and semantic data. It allows applications to follow the user's actions, predict needs and react appropriately to unexpected actions. The use cases would be constructed around three complementary domains: attentive personal systems, interusable devices and complex systems control, for this research itself however we are particularly focussing on attentive personal systems.

2.2 Goals of SMARCOS

With the diversity of subjects tackled in the SMARCOS project setting, it is little suprising that their goals are also multilateral. Their ultimate goals were stated as:

1. Cost reduction of system design.
2. Reduction of development cycles.

3. Reduction of revalidation and recertification effort and time.

4. Improvement of reusability of concept accross multiple sectors e.g. interop devices between domains.

All this should result then in happier users, less redesign, more reusable models and novel services. [1]

2.3 Philips DirectLife

Philips DirectLife is the behaviour altering system that serves as a central theme for the experiment. It is a commercial product in which an triaxial accelerometer is used to keep track of the activity level and activity energy expenditure (AEE) of a system user. The information gathered by this accelerometer, which has been named activity monitor, is processed in combination with several personal profile attributes in order to calculate the activity status of the system user. Based on this profile and activity the system can provide information to a coach, which is at time of writing a real human being, who will then perform his task as coach by sending a coach message in the form of an e-mail to the system user.

The system is set up as an aide to people who wish to improve their own pattern of behaviour and activity, and in its current form is mostly aimed to benefit those people who are involved in work that requires little physical activity, or have otherwise a job that requires them to sit for extensive periods of time.

2.3.1 Goals of DirectLife

The goals of DirectLife vary widely from demonstrating that the company is interested in improving the wellbeing of its customers which can improve on the general attraction of the company brand name. Another point of interest is that it demonstrate that the company is acting within a modern timeframe, being green and health aware is a current issue, so demonstrating interest here and producing a product that can help people toward this more healthy way of Life is also benifitial toward the brand name. Another way that this last element works is used by the possible means of distribution, expecially in the United States.

However, the more interesting goal for this specific experiment is the goal Direct life users have with the system. The product is setup to stimulate its users to embed sufficient physical activity into their daily lifes, addressing the many well-known health benefits of an active lifestyle[31]. Next to that there is also the important factor that health is an important factor in wellness, and people that feel well usually perform wel, it has in fact been demonstrated that several health risks that can be reduced by a regular activity routine are a cause for absenteeism which is very damaging for work performance[27].

2.4 System details

The current system works based on personal feedback from human coaches. These coaches are informed by the system in the form of a profile what kind of participant they are dealing with at that moment, the profile will state which of the previously mentioned category the person is in. At the same time the participant carries with him the so called Activity Monitor, which is a battery charged triaxial accelerometer, with a magnetic connector which allows it to connect to a USB device that functions as a docking station. The dock can be connected to the computer via a USB port and as soon as the Activity Monitor is connected to the dock, it will synchronise its content with the server where it is processed on the base of the profile information, the processed data will then be provided to the human coach. This will enable the coach to provide accurate feedback to the participant, stimulating him or her to further improve the pattern of activity.

2.5 Integration with the projects

So the overal aim is to develop a new form of intelligent technology, more specific a method of presenting the information given by an existing system, that interacts directly with people, and makes use of the DirectLife system. The next chapter will deal with related work and will outline the theories behind the behaviour change support system involved in this new technology.

3 Behavior change support system and embodied agents

The system that we are examining is a behavior change support system which involves the DirectLife system as developed by Philips, which is used as a tool to test a method of conveying a message to a user. One of the two comparing methods of conveying a message involves communication via an embodied agent. Embodied agents have been used in the past as well for behavior change support systems[8][26], and in those previous researches it was demonstrated as an effective method of instruction. This chapter will discuss and elaborate further on similar experiments which have dealt with behavior change support systems, through embodied agents as well as other methods; and it will also deal with the psychological theories on which these systems are built.

3.1 Health assistance systems

Using embodied agents in behaviour alterating systems and health assistence systems has been tested before succesfully on several occasions. An important factor in developing such a system is that users of the system need to find the character more or less relatable in order for the persuasive tactics to work for example it is essential that the embodied conversational agent(ECA) has the ability to generate verbal and nonverbal output [11].

3.2 Use of mobile devices

The use of mobile devices to give personal feedback from an embodied agent on a mobile device is not a novel one. Back in 2002 [9] several researchers already developed a health advisor system that was developed for use on a PDA system. However, the difference with the system in that research was that the agent used was interactive and rendered real time, while the design in this current research is such that the feedback from the agent is rendered in advance, thus resembling the e-mail response much more accurate than it would be if the agent was real time interactive. Another important difference between previous research and current research is the strict seperation between textual and spoken information in the ECA environment. Raw data about performance that involve numbers and statistics in the current system are still displayed as text, while the real coaching aspect in the new system is uttered in the form of spoken (Dutch) text, without any form of subtitles or text balloons. This in order to give some extra feeling of authenticity to the agent.The current system in this experiment is designed as such that it can be adjusted to view on smartphones, and in its current form is already functional for some smart phones, making it comparable to the existing Bickmore[9] experiment and we hope to have results that are in line with the expectations that were created by the Bickmore experiment.

3.3 Emotional influencing

An important factor within behavior change support system is to deal with the emotions of the person that the system is trying to support while that person is trying to change his

or her personal behavior while using the system. There are a lot of ways in which systems try to influence other people by playing on their emotions. What these systems share is that they try to take a persons current state and from that state they try to convince the person to perform a specific action. The current state is not necessery an emotional one, emotions can also be used as a tool to convince the person to take the action you wish them to take. It has been demonstrated for example that just by playing on someones feelings of guilt can convince people to do actions they would normally not give in to[25], just as their happiness or grief[25]. In the current system that we are looking for we are making use of a simpler principle positive reinforcement, where positive behavior i.e. behavior that is closer to the desired changed behavior gets rewarded, which has been demonstrated effective in health counseling and disease prevention[14], making them associate the change in their behavior with a positive emotion.

In order to achieve our goal, i.e. the adjustment of the behaviour of the user, we are to employ several psychological tools. First of all, we need to classify in which category our user falls.

3.4 Behavior adjustment psychology

Since this is a comparative study about two destinct presentation methods of feedback within the same behaviour altering system, it is important to take a look at previous studies in which someone tried to alter or adjust behaviour of people with the same or similar systems. This current system makes use of behavior adjustment psychology the concept on how to achieve change through behavior adjustment psychology is illustrated in figure 2 and 3. We try to make users of a behavior change support system go through several stages of change which as a final results should have a permanent change of behavior to a behavioral pattern that is more beneficial to that person. This system that we are using here is also one that is developed for the internet, where it is not the first one. One of the most prominent areas for future healthcare improvement is the role of the web in fostering improved health and healthier lifestyles[24]. There is a wide range of diseases, health risks, addictions and psychological problems for which researches have reported positive results in managing them via online systems[28].

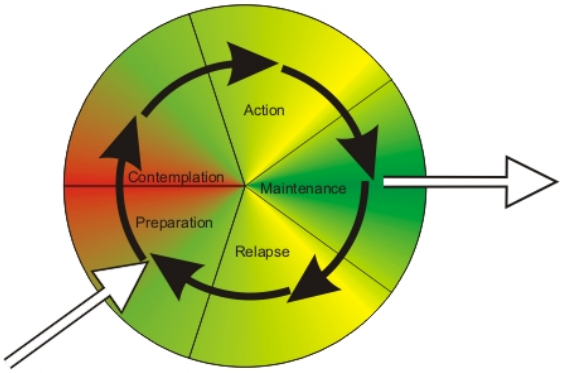


Figure 2: Stages of change

In a paper done by Harri Oinas-Kukkonen[23] they make a specific division between three different types of behavior change support systems. C-Change, B-Change and A-change where A is defined as the most difficult type of change and C is the least difficult. The C-change is simply the user complying to the requests of the system, in our case

the basic request of the system is that the user takes his activity monitor and docks it regularly. The user will be reminded of this each time there is a lack of information from the system. The B-change system is to make users comply with the actual goal of the system, this is a bit more difficult but is the actual aim, namely to make the participants more active. The A-change is to not just adjust the behavior of the person temporarely, but to actually change the mentality of the person using the system to make it more effective on the long run. The experiment system we are using is focussing mostly on the C- and B-Change, due to the lack of time of the experiment and given how difficult it is to establish A-change in such a short time.

3.5 Transtheoretical Model

The basic model that is used to see in what current state a participant of the experiment is and what methods can be used to get the person to act as is desired by the system is the Transtheoretical Mode as introduced by Prochaskal[15]. The stages in which the people involved in the experiments are devided are pre-contemplation, contemplation, preparation, action maintanence, stable behavior and relapse. By adjusting the pros and cons within each decision that a participant experiences before a change and influencing the self efficacy of each participant we can make a participant change his or her current state to another one, and if the system works properly then the new stage should be close to stable behavior.

3.5.1 Pre-contemplation(not ready)

Participants at this stage do not intend to start the healthy behavior in the near future (within 6 months), and may be unaware of the need to change. Here they still learn more about healthy behavior: they are encouraged to think about the advantages of changing their behavior and to feel emotions about the effects of their negative behavior on others. Precontemplators typically underestimate the advantages of changing, overestimate the disadvantages, and often are not aware of making such mistakes. These individuals are encouraged to become more mindful of their decision making and more conscious of the multiple benefits of changing an unhealthy behavior.

3.5.2 Contemplation(getting ready)

At this stage, participants are intending to start the healthy behavior within the next 6 months. While they are usually now more aware of the positive sides of changing their behavioral patterns, the negative sides of actually having to go through that change are still in balance with the positive elements. This ambivalence about changing can cause them to keep putting off taking action.

Participants of a DirectLife program that find themselves in this state learn about the kind of person they could be if they changed their behavior and learned more from people who behave in healthy ways. So to get a person to step up to the next stage, we need to demonstrate the behavior of other people and show to them how their health

is benefitial to them. They are encouraged to work at reducing the influences of the negative elements of changing their behavior. Important here is not to excessively point out that they need to change, since this highlights the downsides of changing much more than the positive influences that the change will have on their lives[15].

3.5.3 Preparation(ready)

Participants at this stage are ready to start taking action within the next 30 days. They take small steps that they believe can help them make the healthy behavior a part of their lives. For example, they tell their friends and family that they want to change their behavior, resulting in an increase of commitment to the change they are preparing for.

During this stage, participants are encouraged to seek support from friends they trust, tell people about their plan to change the way they act, and think about how they would feel if they behaved in a healthier way. Their number one concern is—when they act, will they fail? They learn that the better prepared they are the more likely they are to keep progressing. So for our system it is important to keep focus on social re-evaluation and generate a relationship of trust between the participant and the coach.

3.5.4 Action

Participants at this stage have changed their behavior within the last 6 months, and need to work hard to keep moving ahead. These participants need to learn how to strengthen their commitments to change and to fight urges to slip back. This is the stage where it is important to keep the decisional balance the way it is, namely in favor of proceeding with the change that has already been put into motion.

Strategies taught here include substituting activities related to the unhealthy behavior with positive ones, rewarding themselves for taking steps toward changing, and avoiding people and situations that tempt them to behave in unhealthy ways.

3.5.5 Maintenance

Participants at this stage changed their behavior more than 6 months ago. It is important for people in this stage to be aware of situations that may tempt them to slip back into doing the unhealthy behavior—particularly stressful situations.

Participants here learn to seek support from and talk with people they trust, spend time with people who behave in healthy ways, and remember to engage in alternative activities to cope with stress instead of relying on unhealthy behavior. Participants that are in this stage of change are basically undergoing a type A-change[23], which hopefully will result in a new stable pattern of behavior. In our experiment we use specific hints as weekly advice that demonstrate that falling back in unhealthy behavior results in long term health risks hopefully resulting in a process of consciousness-raising, and we also point out that they are not alone in the systems, that there are others there who they can compare themselves with.

3.5.6 Stable behavior

The ultimate objective with a behvior change system is to achieve new stable behavior in which the desired original change has become the new daily routine for the user. The change to the new behavioral pattern has been more than 6 months ago and at this point there is no longer a direct danger of relapse present. If the participant has achieved this form of change we can speak of an A-change.

3.5.7 Relapse

If during the stage of maintenence the new stable behavior is not achieved, it will result into a relapse of old habits. This will mean that the participant will have to re-evaluate the pros and cons of his or her current behavior in order to once again go through the entire process of changing.

3.5.8 Stage transitions

It has sometimes been argued that the transitions and categories within the state transitional diagram as used by Procheska are somewhat arbitrary[32]. To demonstrate that this model is still valid for use in this experiment we can show each of the steps involved during a transition occurs from one stage to another. One of the fundamental elements of achieving each stage change is the element of the participants believe in their own competence or in short their self-efficacy, making use of the predictive qualities resulting from studies on self-efficacy[29] we can use the knowledge about the indivuduals' self efficacy to stimulate change.

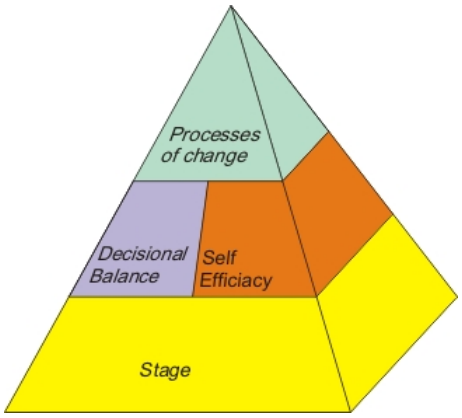


Figure 3: Graphic representation of behaviour adjustment psychology

3.6 Transition

When trying to persuade a person to do something, in this case change willingly on a personal level, there are several psychological methods that we can use that adhere to the principles involved the transtheoretical model.

3.6.1 Transitional processes

There are several processes that can start a change of state within the participants behavior. The ability to invoke on each process depends on several factors, namely what part of the psychology within the stage of change we wish to adjust, (decisional balance or the self efficacy), the current stage the person is in and the tools, both mechanical and psychological that we have to our disposal. We have the following processes:

1. **Helping Relationships**— This is the central process of the experiment involved in this research. It is about the participant finding people who are supportive of their change, by making use of a personal coach. We offer a coach to the participants of the experiment, one coach with one basic set of personal information about the participant and we give this coach the ability to provide the service of a helping relationship. Persuasive technologies developed to date have, for the most part, implemented persuasive techniques inspired by human-human interaction[13]. So the coach we offer here is completely based on the coaches of the current commercial DirectLife system, as well as the knowledge of a real life coach. It is very important that both systems try to achieve a close relationship with the participant since a message is more persuasive if it comes from a source with whom the listener perceives a greater affiliation[12].

2. **Consciousness-raising** — Increasing awareness via information, education, and personal feedback about the healthy behavior. By giving information in the proper way we wish to push forwards the advantages of changing while at the same time making aware that the disadvantages are less, and the difficulties during change can be overcome, thus shifting the decisional balance in the favor of change.

3. **Dramatic relief**—Feeling fear, anxiety, or worry because of the unhealthy behavior, or feeling inspiration and hope when they hear about how people are able to change to healthy behaviors. By making use of feelings about health problems and stating health risks in a weekly advice, we can use this process to convince a participant to take action to improve his or her daily pattern. This in order to change the daily pattern with a more healthy one. An example of this is that the system makes use of some simple medical references such as pointing out that a lack of steady movement can result in heart disease.

4. **Self-re-evaluation**—Realizing that the healthy behavior is an important part of who they are and want to be, this element can be played upon by using the right words when giving the weekly coaching. It is important here that we do not actively instruct the participant of the experiment to do something. Instead of giving instructions, the feedback will be more in order to give general health advice and how this can affect them personally.This should encourage the person to re-evaluate his or her own behaviour.

5. **Environmental reevaluation**— This is the effect of people realizing how their unhealthy behavior affects others and how they could have more positive effects

by changing their daily routines and activities to more healthy ones. We can make use of this mechanism in two ways. The first one is to point out the benefits to their environment by health tips, this must be done in such a way that it cannot be conceived as an instruction or insult in any way, since these would work counterproductive. The second way is already in use in the web-environment part of the current system, where the participants can sign up in friendship. This will show their environment that they are actively trying to increase their health.

6. **Social Liberation**—This is about the participant realizing that society is more supportive of the healthy behavior, this is a point that again can be easily pointed out as one of the advice of the week elements of the coaching. However since there were no direct articles that the coach could refer to, this element was not applied.

7. **Self-Liberation**— The method of self liberation has to do with believing in one's ability to change and making commitments and re-commitments to act on that belief, it is closely related to the psychological principle of consistency and commitment, it has been demonstrated by several researches that once an individual publically sets a goal and attempts to it, he or she will keep going to work toward this goal even if its against his or her own intuition later on[25]. The system makes use of these goals by giving the participants of the system the ability to broadcast their achievements or by making them public automatically, and at the same time letting the system set goals for them to adhere to.

8. **Reinforcement Managemen**t—This process deals with the increasing of the rewards that come from positive behavior and reducing those that come from negative behavior. During the weekly response coaching in this experiment we make use of basic forms of positive reinforcement to encourage the participant to move more. At the same time the feedback from other people that are involved in the same community of the experiment will also possitively encourage extra activity. At the same time not wearing the activity monitor or not moving enough will put the participant further away from the goals, and this is measured by the system and compared to the goals as they are registered in the system. Reading back on such negative results will generally be experienced as discouraging.

9. **Stimulus Control**—This is the avoidance or countering stimuli that elicit problem behaviors. We can invoke this process by using reminders and cues that encourage healthy behavior as substitutes for those that encourage the unhealthy behavior. The most basic ones in this system are the fact that particpants actually bring the activity monitor and docking it with their system so they can actively keep an eye on their behavior, and in that way convince themselves to adjust whenever they see that there is a unhealthy pattern. And from that participants will realize that they are doing something wrong, without the system having to explicitly point it out for them.

10. **Counter-Conditioning**—This is the process of substituting healthy ways of acting and thinking for unhealthy ways. This process is not further used in this

research experiment.

3.7 Weapons of persuasion

If we wish to evoke these transitional processes via another person (or in our case virtual human being) we need to examine what psychological mechanisms exist to do so. In "Weapons of persuasion" several strategies to do just that are pointed out[25].

3.7.1 Reciprocation

The concept of reciprocation is that when a person is providing you a favor, no matter how small or how big, the person automatically gets the feeling that he or she needs to repay that favor in some form. This feeling is not based on rational balance so you can encourage a person to provide a big favor as a response to only a small courtesy [25]. This concept can be used by a coach by attempting to give the participant the feeling that the service the coach is providing is sincere, and the best way to reciprocate is to follow the coach's instructions.

3.7.2 Commitment and consistency

After a statement is made a person will almost always try to stick to this statement, even if that person has been proven wrong or he has the "gut feeling" that he has been duped[25]. The system makes use of this system by setting a goal in advance that is optimized for this person, and then every week presenting this goal to the participant in order to try to keep the person consistent.

3.7.3 Social proof

If you see someone that is of the same "clique" as you, you are more likely to immitate[25]. The concept of social proof is very important in coaching or instruction environments, since it allows us to present to the coachee an coach that can serve this exact purpose. Other studies have shown that this is effective but only if the coach is realistic enough in behaviour and also uses varying methods to bring the information[22]. However, when trying to model a coach as realistic as possible there are some areas of realism known as the uncanny valley[2] that should be avoided, since on one side exaggerated emotion display proves effective, overly realistic display of an artificial human representation turns to invoke more negative feelings.

3.7.4 Liking

If you see someone that is "likable" you are more likely to follow his or her instructions[12]. This fact is closely related to the concept of social proof, and it is also demonstrated to work with embodied agent environments[26]. An effective coach will attempt to form a strong working relationship with the coachee, and this counselor-client relationship is

predictive of outcomes. When measuring likability for systems we can resort to a standard set of survey questions that have been demonstrated to be effective in analysing how well a person likes the agent that person is interacting with[5].

3.7.5 Authority

If the person has symbols or titels of authority, you are very likely to comply with instructions automatically, even if it goes against almost the very core of your being[25], something that was clearly demonstrated in the Milgram experiments in 1974[21]. Simple tricks can be used to give the actor in the system extra authority, giving the person to communicate with a labcoat has been proven to be effective as well as placing the actor in a proper environment such as an office or similar official surroundings.

3.7.6 Scarcity

If there is few of an item we want it more. If the item seems to be unobtainable soon, we want it the most. It has been demonstrated in several occassions that aquiring can be far more imporant than the actual possession[25]. Since this is a training program with an instructor, there is not a specific object that people can obtain in a specific form, so the only limiting factor in how we can make people desire the product more is by making them aware that they can only use it for a limited amount of time.

3.8 Virtual embodied agents in behavior change support systems

In the past, ECAs have been used in a wide variety of environments to support behavior change. There have been experiments in which they have been used to deal with depression where it turned out that 35 percent prefered an ECA over even a real doctor[16]. Next to that there have also been experiments in dealing with psychological illnesses such as schizophrenia where embodied agents were used to help patients[10].

Next to this there have also been experiments done that are closer to people managing their own personal health where they even managed to determine that the right kind of politeness were key to getting people to keep adhering to their new more healthy behavior and that it is crucial to avoid annoying interruptions [7].

3.9 Virtual embodied agents vs pure text systems

Other studies indicated that system that make use of embodied conversational agents are more effective at putting a message across that is remembered.[16] Making sure that the coaching message is remembered should serve as an indication that the embodied agent environment should be more effective at relaying the coaching message. Other studies have shown that in instruction environments, the addition of extra gestures such as gaze, blinking, nodding are vital for people to accept these systems[17]. Next to this there is a need for exaggerated emotional behavior, so when adding a face to an existing text environment, not only does the text have to be converted to a speech which is significantly different from an e-mail also there is a need to emphasize specific words of

importance, and we need to take control of the rate of speech of the system.
At the same time there are also other studies that have shown that when comparing a system involving an embodied conversational agent with the same system without the embodied agent, for learning it has neither a beneficial nor a detrimental effect[20]. So if we take these findings in account we would expect to have some positive influence from the embodied agent when it comes to coaching, if all the important elements of the systems are correctly setup. For example in an e-mail system the participant can read at his or her own speed, the embodied agent will perform the speech at a pre-defined speed which needs to be calibrated carefully. As mentioned before the same goes for nodding, blinking and gestures as well as gazing. All these should be properly setup to make the agent acceptable for the participant.

3.10 Putting theory to practice

Given this theoretical information we can continue to design a system that makes use of all the principles, to find out if the results are in line with the previous performed experiments.

4 System description

The current system exists out of several parts, the innerworkings of which we cannot describe in full detail due to the confidential nature of the internal working, instead we will discuss them on a level of functionality. The system relies on several subsystems, and a personal feedback coach. In figure 4 we can see a simple schematic outlining all functional parts of the system. It contains a way to present information to the user, by means of a website and by sending e-mails. It can receive information from the activity monitor through the internet, and by analysis through some undisclosed statistics and heuristics the system is able to catagorize system users. The system will provide a coach with a basic text and profile information, enabling the coach to send the user a personalized coaching e-mail assisting him or her in fullfilling the program.

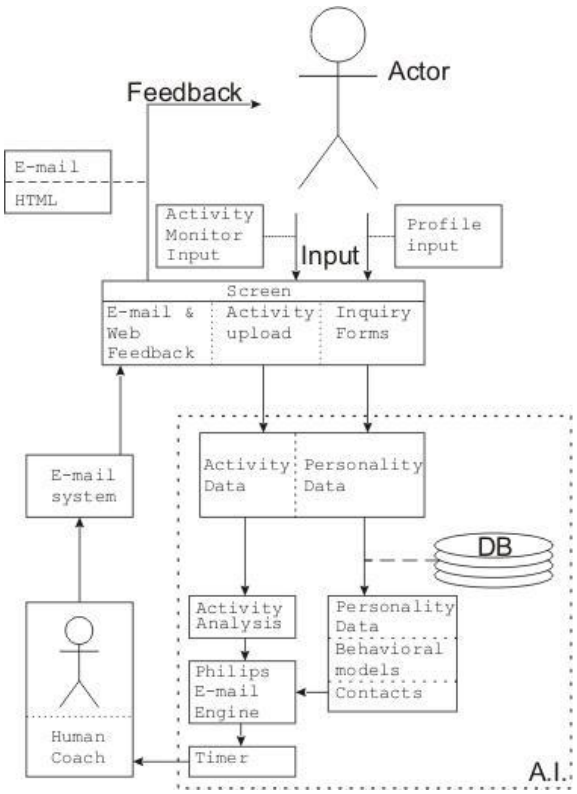


Figure 4: Consumer system

4.1 Comparing the consumer system with the experiment system

When comparing the two systems, as seen in Figure 4 and Figure 5, we notice several important differences. First of all the part of processing that is handled in Eindhoven is considered a black box from which we assume that correct information is obtained from the user input. Secondly we have the feedback method, where the consumer system

basically provides information to a human coach who then e-mails it for the user to read, the new system performs its role differently.

First of all, the Philips system here has a predefined XML output, which is placed on a data dump server, as seen in figure 7. This server will log all its input and informs the system operator, a human component within the system and in this experiment we will take on this role, that data is available to be processed. In case the data is meant for e-mail presentation the operator will check the e-mail on validity of the content, this is similar to the live system except that the e-mail is generated in full text and in most cases no human intervention is needed.

In case the data is meant for ECA presentation there will be a generated BML script created by the system from the XML, which will be checked in the same way by the same person who checked the e-mail presentation data. The script shall be performed by ECA and screen captured by the operator, and the result shall be put on the webserver which will inform the participant that new data is available.

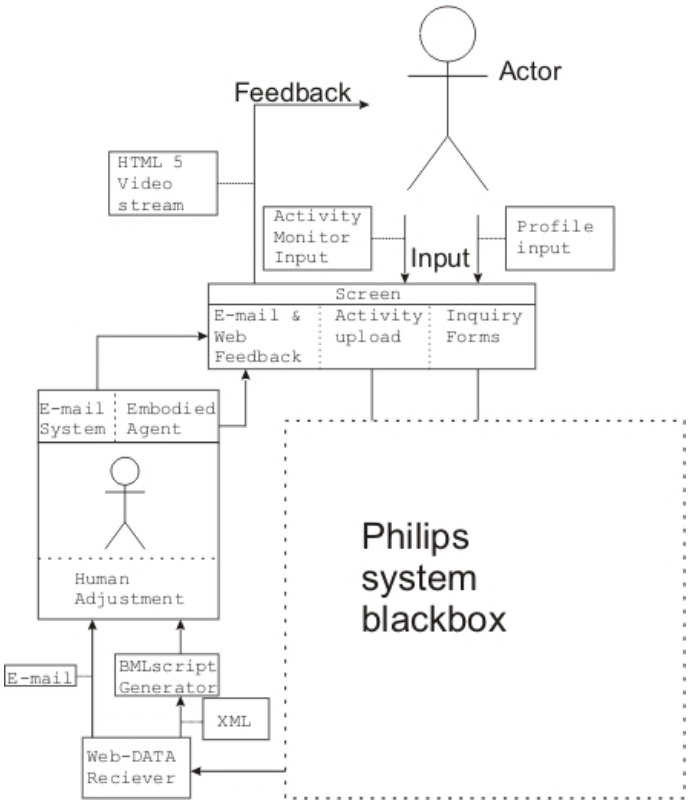


Figure 5: Experiment system

4.2 Key to success

When comparing this system to other previously developed behavior change systems they have for the most part been implemented with persuasive techniques inspired by human-human interaction. So when we try to create new persuasive technologies in order to most likely be succesful we need to create a system that resembles human-human interaction the most[13]. And at the same time we wish to create an intimite bond between the agent doing the communication and the participant within the experiment so we can get the desired affiliation[12]. So the key to success in building a new form of persuasive technology is to create interaction that A) simulations human - human behavior, and B) the technology, in this case an embodied agent, needs to be perceived with a greater affiliation.

4.3 BML

At some point in the history of development of XML, people tried to use this standard to create a formal construct to be able to describe behaviour[18]. This formal construct will then allow developers to write input for embodied agents that should result in output that represents comparable behavior and expressions. Seeing as how this should uniformly behave for all kinds of ECA, we can use this as output of our script generation process and as input for our embodied conversational agent, resulting in a script that should also be useful for other agent systems.

4.3.1 Text use model

The system that is currently being developed by Philips to automate their e-mail response to DirectLife users is called the "e-mail engine". During the analysis of the existing e-mail engine we saw that in order to create a proper text, the most effective way is to make use of a combination of snippets. These snippets are basically text fragments, that together can form a paragraph of a text, and a combination of paragraphs together can form a script for the embodied agent.

So in order to get a full functioning script we need to build this bit by bit. First we determine what kind of script it needs to be (i.e. collect a specific set of paragraphs of which the content suits

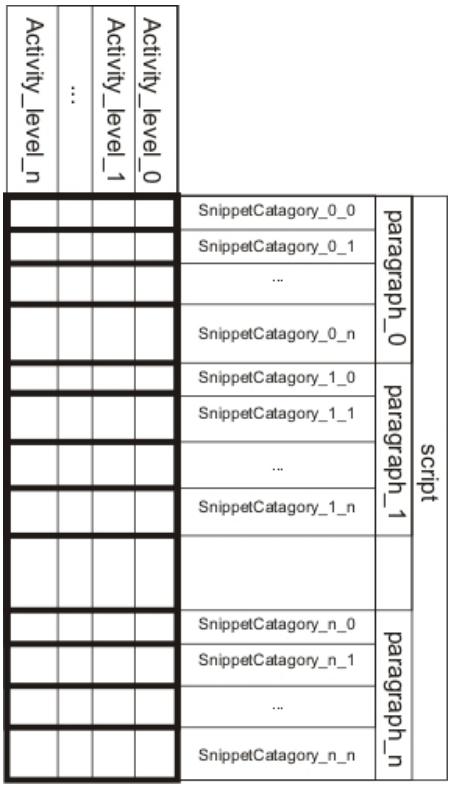


Figure 6: text use model

the person and the total function of the script). To be able to find the proper paragraph construction, we need two pieces of information, the first is the total function of the text, as in are we introducing a user to the system or are we giving a standard message or are we trying to do something else. The second is that we need to know who we are addressing. With the help of a personality analysis we are able to determine what kind of person this is and what kind of script build up we should use to achieve maximum effect.

After determining what the general construction of the script should be like, the next step is to determine the function of each of the fragments of the paragraphs. This is a simple linked list, since the function of the paragraph determines series of sentence categories to use. However when we look at the function of the text as a whole the tone of the message should be depending on the activity profile of the user as well, so out of a set of possible fragments we pick the one exact fragment type that is appropriate for the activity profile of this user.

Since this is an automated system, we risk that the user will hear the same message over and over again, in case he or she is performing in a constant pattern (as in the profile does not change, the function of the script category does not change, and the user performance does not change). To prevent repetative scripts, the last part to actually fetch the snippet is modified by a randomizer, by picking randomly from a set of three possible snippets that all carry the same message in the same tone, yet with the word choice slightly altered and the order of words slightly altered.

A schematic oversight of this text model can be found in figure 6, and an example of the output is demonstrated in section 4.8.1.

4.3.2 Emotion use model

The original coaching system makes use of a mild form of positive reinforcement conditioning. By always making the coaching message positive it encourages the person to continue going on, but this does endanger the effectiveness in giving correct feedback, since there is little use to compliment someone on doing absolutely nothing toward their set goal. So we present the most positive feedback when closest to a specific goal range under the right conditions.

4.4 Elckerlyck

An imporant part of the system is the part that actually realizes the BML script into an actual movie. The ECA that performs the role of coach in this behavior modification support system is the Elckerlyck agent [30] as developed by the university of Twente. This embodied agent is able to combine facial gestures with several different forms of text to speech engines, in this specific system we will be making use of Loquendo text to speech engine for Dutch language. It is able to express emotions, speak at various rates of pronunciation and is capable of showing several gestures based on BML input. By providing a proper script to the embodied agent we can create a video performance that suits the needs of the experiment. It has been demonstrated to be capable of real time

realisation of the scripts and it is also capable of being expanded with extra emotions, gestures and expressions if required.

4.5 Feedback types

When interacting with a behavior change system like this there are two distinct forms of feedback, one of these is the coach instruction feedback which is the main focus of this research while the other is the response to user questions. This second type is also something which perhaps can be automated but this is outside the field of this research.

4.6 Components

For this research we needed a suitable test environment. There were several restrictions on this test environment, first of all the most constricting factor: time. The total timeframe for complete system development of this experiment is less than 4 months. This means that making use of components that require alot of implementation from the developer side are not an option, all implementations need to be directly usable with a minimal amount of programming. There are several tools that lend themselves perfectly for this experiment, due to the amount of experience we have with them and the relative well fit they are in their most basic form for the task at hand. These tools are a set of programming languages and frameworks and open source components, as well as components developed by other people at the university. Out of all the useful components and tools that are available we have chosen the following:

1. **Programming languages**
   Java, as basic language for the script generation process and external communication with the database.
   Ruby, in combination with the rails framework, in order to provide access to the video content and manage safe log in and user accounts. All the Ruby on rails code will run on a Mongrel server which has demonstrated to be stable and reliable.

2. **Database**
   MySQL, since this is the database the developer has most experience working with, and is also suited to serve as the database background for the rails framework.

3. **Embodied agent**
   Elckerlyck, The ECA for realizing the BML script.
   Loquendo text to speech voice Saskia, since this is the only available text to speech that is female, dutch and can also be integrated with the BML realiser platform of Elckerlyck.

4.7 Total experiment system setup

Each of the components have been chosen for sake of practicality and because they would, in their current form, be easily integrated in the already existing consumer version of

the DirectLife system. The data-transfer layer is of little interest since it is of temporary use, so it was built up with robust but simple to built server. The videoserver in its testform was built up in such a way that it needed a secure login, such as would be used in the DirectLife current web system, and it can provide a basic structure that can easily be built in in each type of webserver, by using the existing frameworks for it that are well tested and completely open source it demonstrates itself to be quickly implemented, stable and secure. The scriptgenerator has been implemented in Java, this was done because of the necessery external software was written in java as well, which makes everything easier to integrate into one whole system, as well as for the sake of programmer experience, since it is language the developer has most hours of programming experience in. The choice for the Elckerlyck agent is a more complicated one as we have seen in the previous paragraph. The streaming video concept is one that can be used due to the choice for a-synchronous communication.

4.7.1 WebServer

This is a http server which makes use of session locks to communicate with participants of the subject. It is a Mongrel server programmed in Ruby and running on the Rails framework. Based on session locks the system can only send movies of the current user to him, attempts to view other movies will generally result in failure unless the locking system is somehow hacked. In order to further secure privacy, passwords of any user of the system will be encrypted and there is no viewable URL information that can compromise security. Once a participant has made an account on this server he needs to confirm this with e-mail feedback. After all this is done the viewer has access to lists of movie entries that are linked to his or her account, the content of the movies and the movie entries need to be added manually by the system administrator. In a full system this of course will be automated, but that would go beyond the range of this experiment. Once the participant has access to this movie list all he or she needs to do is click the desired coaching video and it will appear. The view of each movie is setup in HTML5 with a fallback to flash in case HTML5 functionality is missing on the client side. The choice for this medium is based on compatibility, HTML5 and flash together are supported by 95 percent of the browsers used and its also supported on mobile devices. In case even this support fails, there are links to three seperate formated downloadable versions to accommodate this last group of users. Any failure to view the movies aside from these are considered beyond the range of this experiment(participants in the experiment will be helped in order to make sure they can view the movies).

4.7.2 User information extraction

For the system to work, we require input extracted from the user profile and the activity data as provided by the experiment participant. During a standard interval the length of which is defined by the participant, the Philips system will generate XML data and an e-mail message. For the experiment this part of the system will be viewed to all purposes as a black box that gets input and produces the proper output as is required

by the rest of the system. The exact innerworkings of the system is confidential and therefore cannot and will not be further examined in this research.

4.7.3 Data transfer

After data-generation, this data will be sent via an http medium to the mongrel web-server at the university. Its function is relatively simple: it receives the information, makes an entry in a database that the information is received, then it generates two files. The two files are one e-mail file, that contains the e-mail data as produced by the Philips Direct Life e-mail engine and a second file that contains the XML data as produced by the data analysis system within the Philips system. By storing it in both a database and putting it into files we have a logging and back up system, and a convenient way to loading files into the script generator.

4.7.4 E-mail server

The system will make use of an e-mail server, which is an integrated part of the entire movie server that informs a user as soon as a new instructional coaching video is placed on their special video channel. This component makes use of the built in system of post-mail that exists within the Rails framework. It can function properly since the only requirement for correctly delivering mail via this system is to implement the tags that surround the content of the e-mail. For the users of the traditional system a similar system is used only here the e-mail will be send with full content of information regarding the coach instructions, as generated by the already implemented "Philips e-mail engine". Each time one of these e-mails is delivered an entry will be made into a database, allowing this experiment to monitor all send out data closely and keeping track of everything with the help of timestamps of database entries.

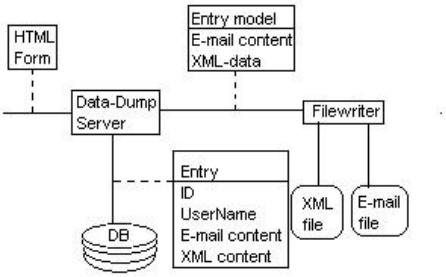


Figure 7: XML input handling

4.8 input

In order to obtain a script through this system, it requires a form of input that it can handle. This java program requires an XML schema as input, containing the following pieces of information:

1. Username

2. Date to put online
3. Calories
4. Percentage of goal reached
5. Stage of change
6. Activity type

One of the components of the experiment environment was the data receiving server, which by HTML forms received the data from Philips, and then took all the relevant content and created an XML file from this, the content is stored in a database, and stored in file format (both input and resulting xml output), and in order to always be informed the information is also sent by e-mail to an account that was specially created for this particular experiment. The special formatted XML can then be fed to the script generator, which in its turn will parse it to the proper parameters to use on the script generating algorithm.

4.8.1 Database link

This class is the mediator between the java program and the database which contains the actual text fragments and the hierarchy of the content of the database. The discription of the content of the database can be found below in the tabels. The database entries are divided over four different tabels, the top level tabel will pick the paragraphs of the script, while the second level tabel will determine which catagory of sentences the script will have, based on this activity information the it will get an index of the lowest level of tabel, which contains three entries for each key. To provide with extra variety within the system a random one will be picked from the three possiblities, the three possiblities will contain the same content, tone and method of addressing, yet the word order is slightly adjusted and with some words replaced with synonyms. Though the scripts are actually created on base of weeknumber, making them unique per week, so it is impossible for one person to recieve the same message twice, it is also rather unlikely (though no t impossible with the exception of the name of the individual) for two people to recieve the same movie, even if all the parameters are the same. The only exception in this is the received message upon a lack of data, since there are only a very small number of possible replies here.

4.9 Script generator further examined

The scriptgenerator program itself is a java application, that is built on top of an SQL database. The database provides a linked structure where all the script fragments can be found, while the application uses its internal structure to form a complete script from the information obtained from the database. To make sure this project is also suited for possible follow up experiments, the graphical user interface is designed to be

comprehensible and simple to use, and there is also a possibility to generate the scripts by commandline with the proper attributes.

Within the grafical user interface there are several settings that are built in for experiment sake, that are not supposed to be used in practice, but can be useful to optimize the system for actual use, these are the options to adjust user profile input values before they are parsed to the script generating algorithm.

4.9.1 output

Given the general design of the text built up model the eventual content can be described in several layers. The top layer is a story, a story consists out of paragraphs, paragraphs consist out of snippets which are sentences combined with BML fragments. The whole story needs to conform to the BML specifications. By making use of the parameters that are available within the system we can built up a text according to the following schedule:

Table 1: Story lineout paragraph level

week	**Paragraph ID**
0	0,1,2,3,4
1	0,5,11
2	0,6,11
3	0,7,12
4	0,8,13
5	0,9,13
6	0,10,14
6.1	0,15,16

This is the highest level of abstraction, it deals with what type of script it should return and this is based on the week of the program. For example if it is the beginning of the programming, week 0, it will then continue to the next table where it will attempt to fetch in order paragraph 0, 1,2,3 and 4.

Table 2: Content of paragraph.

Paragraph ID	**Info**	**Category ID**
0	BML + salutation	0
1	Intro movie paragraph1	1
2	Intro movie paragraph2	2
3	Intro movie paragraph3	3
Continued on next page		

Table 2 – continued from previous page

Paragraph ID	Info)	Category IDs
4	Intro movie paragraph4	4,19
5	Main first feedback	5,6
6	Main second feedback	5,7
7	main third feedback	5,8
8	Main fourth feedback	5,9
9	Main fifth feedback	5,10
10	Main sixth feedback	5,11
11	Closing advice week 1+2	12,16,19
12	Closing advice week 3	13,16,19
13	Closing advice week 4+5	14,16,19
14	Closing advice week 6	15,16,19
15	closing program Info	5,17
16	End of program Farewell	18,19

This is the second level of abstraction, it deals with the content of the paragraphs of the scripts, if we continue our previous example we see that paragraph 0 will return snippet category 0, paragraph 1, will return snippet category 1, paragraph 2 will return 2, paragraph 3 will return category 3, and paragraph 4 will return snippet catagory 4 followed by catagory 19. All of these will be fetched from the next level table.

Table 3: Snippet categories.

Category ID	Info	Snippet ID	stage of change
0	BML and Salutations	0	
1	Intro movie paragraph1	1	
2	Intro movie paragraph2	2	
3	Intro movie paragraph3	3	
4	Intro movie paragraph4	4	
5	Reference to text info	5	
6	Explaning data and giving	6-15	not ready
	short feedback	16-25	ready/stable
		26-35	stable/relapse
7	Explaning data and	36-45	not ready
	short feedback	46-55	ready/stable
		56-65	stable/relapse
8	Explaning data and	66-75	not ready
	short feedback	76-85	ready/stable
		86-95	stable/relapse

Continued on next page

Table 3 – continued from previous page

Category ID	Info	Snippet ID	stage of change
9	Explaning data and..	96-105	not ready
		106-115	ready/stable
		116-125	stable/relapse
10	Explaning data and..	126-155	not ready
		136-145	ready/stable
		146-155	stable/relapse
11	Explaning data and giving	156-185	not ready
	a short feedback	166-175	ready/stable
		176-185	stable/relapse
12	friendly remark	186-195	not ready
	Active advice	196-205	ready/stable
	Advice to upkeep pattern	206-215	stable/relapse
13	Explaning data and..	216-225	not ready
		226-235	ready/stable
		236-245	stable/relapse
14	Explaning data and..	246-255	not ready
		256-265	ready/stable
		266-275	stable/relapse
15	Explaning data and..	276-285	not ready
		286-295	ready/stable
		296-305	stable/relapse
16	see you next week	306	
17	Explaning data and..	307	not ready
		308	ready/stable
		309	stable/relapse
18	Explaning data and..	310	not ready
		311	ready/stable
		312	stable/relapse
19	Closing BML tags	313	

Here we have the content of each category, the actual snippets, these are divided over the total input range of personal parameters, which are the state of change of the person combined with the percentage of goal. Seeing how the introduction only has activity level our example will yield the snippet ID values 0,1,2,3,4,5. These will be taken to the next tabel of the database, where each key will return one of three random options. These options are equal in meaning and function, but have slight variations to make each script more unique.

Table 4: Snippet range.

Level	Info	Activity type
0	no significant activity	
1	Under performance	active
2	Adequate performance	active
3	good performance	active
4	Overperformance	
5	serious under performance	inactive
6	mild underperformance	inactive
7	Adequate performance	inactive
8	Good performance	inactive
9	Overperformance	

4.9.2 Text model

This is the central model class in the program. To create a text model, we need to provide the model with the proper parameters which are the weeknumber, the percentage of goal in activity, the stage of change, and the activity type. Based on this information it will contact the database link class, which will provide the proper information from the parameters. The text model consists out of a series of paragraphs, each paragraph by itself consists out of a set of snippetcategories, which all have a set function depending on what paragraph they are in.Then based on the original parameters (i.e. activity type, percentage of goal and stage of change) the right snippet is picked from the possible snippets within each category. If we take the total text and put each snippet in order as set by the paragraphs, we will end up with a full BML script that is able to serve as input to a BML realiser to result into a performance by an embodied agent.

4.9.3 Snippet fetching algorithm

The method by which a snippet is fixed is by determining a score which will be mapped on the total possiblities within the snippet catagories. Firstly we take a look at the percentage score, the total range is divided into five segments, starting from zero percent to 60 percent as insufficient activity to give an accurate response, with which the user will be advised to wear the monitor more and connect to the system more frequently. The next level is from 60 to 90 percent, where the activity is not enough to meet the goals as set by the program, but it is enough to give a form of advice. The third level is the around 91 to 110 percent score, where the participant is told that he or she made the mark, and then there are the 110 to 130 percent scores where there is a solid above goal performance that is complimented. and the last category is the higher than 130 percent score, which gets a similar reply as the normal overperformance with a suited warning as not to work too hard. On top of this basic range there we divide the people into

active an non active, where the activity is based on the initial pattern, giving us a range from zero upto nine. The last factor then is the stage of change, here we have three categories, before ready, ready/stable and relapse. For each stage of change category we set the same range of ten possibilities, resulting in a total of thirty possible snippets (maximum) per snippet category. The algorithm is adjusted in such a way that if there are fewer than thirty possibilities present it will scale the score of 0 to 29 on the total range of possibilities that is present at that time, for instance if we just have 1 possible category because of lack of information it will always result index zero.

4.9.4 BML results from scriptgenerator

The output of the script generator is a BML form that can be parsed by a BML realiser such as the one used within the Elckerlyck platform. The scripts that are generated in the experiment can be divided into four basic categories, first of all the introduction of the coach, which briefly explaines how to use the system and what to expect from it, this script has only one snippet set per snippet category in each paragraph, since activity information is still virtually non existant at that time, these follow from the week 0 and 1 scripts. The second category is the standard progress message, this consists out of an introduction, a reference to some numric information, some elaboration on the numbers, a possible reference to advice (not in all cases present, because someone who is not ready for this extra tip might not want it pointed out) and following advice, and then closing up with a standard see you next week remark, the scripts from week 2,4 and 5 are such scripts. The third type is the half way message, addressing the status of the participant half way through the program, and giving advice that is based on the knowledge about the person built up so far, these are build up from the week 3 scripts. The last category is the closing message, which depends on the end state of the person and not so much on the individual percentage score since it will no longer be possible to set percentage scores after that, giving one extra goal would not really serve a purpose since it is impossible to see whether the goal will be reached or not.

5 Experiment

The method to validate the difference in performance between the two feedback methods is the experiment. This chapter will deal with which parameters will be measured, how they are measured and how the resulting data will be processed.

5.1 Procedure

The first step was to get twenty participants to perform the experiment, which would be performed as a single blind study, where the participants would not be aware what forms of feedback exist within the system and only are informed about the form of feedback they themself can expect to personally receive from the system. At the end they are asked to evaluate the performance of the feedback given by the system. After all information(see section 5.3) is gathered and processed, each participant is given an activity monitor and a short manual on how to proceed with the experiment. This manual can be found in the appendix as figure 12 on page 59, it deals with several points amongst which the communication, where they were instructed to ignore all e-mail that are automatically generated by the administration system, rather than the coach of the experiment. Both parties within the experiment are being informed of a weekly update via their e-mail, in one group the mail will contain all the information of the last week, while the other group just receives a message that there is a new video waiting for them. Here the coach will give them the information of the previous week. The information that is in the video is exactly mapped on the information that can be found in the e-mail, this in order to avoid that one group is better informed than the other. All participants were introduced to the coach who was named Brenda which is an acronym for BML Realizing natural Direct-life Assistant, while one group only saw the name under each e-mail as sender, the other group would have a short introduction in which the embodied agent introduces herself and tells the participants that she will guide them further.

week 1

The next step is the first week of the actual experiment. In the first week, all participants will have to complete an assessment program. This is a training program that gives no daily feedback on the website, and at the end of the week the system will provide them with feedback about their activity pattern, this will be the first unique feedback message to all participants, it will deal with the results of the assesment and asks for all participants to start their usual programs then and there if there is enough information for the program to run on. If there is a lack of information participants will be asked to select a standard pre-defined program.

week 2 and 3

The two weeks after that are normal weeks in which the user can receive daily feedback from the system about their performance and at the end of the week they will get feedback from the coach informing them of their overal weekly performance as well as giving them an extra piece of advice that is selected for them based on their activity level and stage of change, as well as their direct life profile. For example people who are generally inactive and in danger of relapsing yet suddenly have a week with abnormally

high activity who run a risk of "crashing" will be adviced to stick to a pattern that they feel comfortable with and not to overdo it. Another example would be people who are still in the contemplation stage and are not doing too bad but are not reaching their goals yet either, these persons would be informed of health risks of general inactivity and the benefits of long distance walking.

week 4

After this week the half way point of the experiment is reached, and they will be given a feedback session that is reflective of their entire path so far. This is also the moment for the first feedback session for the experiment so all participants are at that point asked to fill in the second survey.

week 5 and 6

Once passed that week only two more feedback moments from the coach to the participant remain. For these last two weeks similar feedback will be given as for the second and third week. After these last weeks the final survey will be taken as a personal interview giving us the total overview of the systems performance with the two seperate coach feedback systems.

5.1.1 Required system results

In order to compare the new feedback presentation method we need clear numbers that can quantify the performance of both feedback presentation methods, there are several aspects of the system that need to be checked. We need to be able to compare the performance in sense of persuasiveness, relapse sensitivity, attractiveness and relatability. For relapse sensitivity we focus on the pure numbers of the test to see which group performs closer to their set goals on average, while persuasion is going to be measured by a combination of survey and numbers such as how close to their goal are the participants during each interval and their physical activity level. The aspects of attractiveness and relatability are examined purely by survey.

5.1.2 Statistical survey

To be able to form a final judgement on the success of the experiment and of each of the defined parameters (effectiveness, persuasiveness, relapse sensitivity, attractiveness, perceived intelligence, fear, quality of coaching and intention to use longer)we need to make use of a statistical survey with each of the participants. By making use of validated seven and five points scales we can check how attractive and persuasive the system is with a by making use of questions with a seven point scale that are designed to test these specific elements of an embodied agent system [4]. For both feedback methods the same surveys are used to obtain the scores of each feedback method, with the addition of several specific questions about the behaviour of the agent to see whether it was the behaviour of the agent that caused an anomaly, or that it is just the method of relaying the message that causes a difference in opinion about the system. The surveys are done by personal interview. This could cause a social biased result, but since neither group is informed of the existence of another group, and both groups are interviewed using the

same method, this influence should be present as strongly in both groups. The exact survey forms can be found in the appendix, page 64 to 77.

5.2 Measures

In this research the main focus was on quality of information by gathering as much data per person rather than going for large quantities of participants with relatively shallow information. For each participant we would measure along all the dimensions.

5.2.1 Likeability

This dimension is going to be measured with the aide of a validated set of survey questions to measure performance on attractiveness from an agent system. We will ask each participant to rate the system on a range of 1 to 7 on five Likert-scales that all deal with how attractive they find the agent with which they are communicating. Each of the questions should be answered similarly for the survey to be a valid representation of the attractiveness of the system. The average of each of the question will then serve as an indication how attractive the system is. So in this case the desireable outcome for the feedback method is to have a high score on attractiveness on the survey. The questions that they would be asked to rank on a one to seven likert scale would be, Dislike - like, unfriendly - friendly, unkind - kind, unpleasant - pleasant and awful - nice. Questionnaires that use these Likert scales (though they used five point scales) in other experiments they have been demonstrated to have a reliability Cronbach's alpha value of well over 0.7 [2][3].
As an addition to the attractiveness of the agent, the attractiveness to the system as a whole will be measured by a set of questions relating to how interested the participants are in actually buying the system. The more interested participants are to buy the system, the more attractive it is rated.
In addition to this we can also measure for the video group how often and how quickly they watch their video feedback after it has been announced on their e-mail that they have a coaching message waiting for them in their inbox. Since we do not have a method of measuring the responsiveness towards the e-mail group we will have to accept this last set of measurements as a interesting sidenote, but it has no measured element from the e-mail counterpart that it can be compared to.

5.2.2 Perceived intelligence

This will be measured in exactly the same way as attractiveness. The questions are again a validated set of questions relating to percieved intelligence. The idea behind this is that a system that is considered more intelligent and capable of giving correct coaching messages will be accepted more readily by participants as a suitable coach. So the desireable outcome for a system is a high amount of percieved intelligence. We used the following questions on a seven points likert scale to assess the perceived intelligence:

Incompetent - competent, ignorant - knowledgeable, irresponsible - responsible, unintelligent - intelligent and foolish - sensible. These Likert scales have been demonstrated to be reliable in other experiments with a Cronbach's alpha value of 0.75 [3] and 0.769[6].

5.2.3 Trust/fear

This is again measured by using a validated set of questions relating to the amount of fear that the participants experience while interacting with the agent. If the participants feel more anxious or nervous while recieving feedback of the information it can indicate a certain amount of fear or a lack of trust from the participant towards the coach. This could have a negative impact on the performance of the coach. So the desirable outcome for a good system is a low amount of fear. The questions we used for these are those of perceived safety, which are anxious - relaxed, agitated - calm and quiescent - surprised. These scales have been demonstrated to be reliable with a Cronbach's alpha value of 0.91 with a five points Likert scale[19].

5.2.4 Effectiveness

The effectiveness of the system is determined by the physical activity level of all the participants. This is an absolute number that represents how active each participant is. If we have a continuous set of activity levels for each participant, we can then see how well they have improved over the duration of the program, and we can compare whether there is any significant difference between both feedback methods in sense of effectiveness. So the best outcome for a system is a constant increase of physical activity level for each of the participants within that specific group.

5.2.5 Persuasiveness

Each user will receive a goal to reach during each day of the program. This goal is then set as 100 percent, and the idea is to measure how capable both feedback methods are at keeping the participant performing steadily above 100 percent. The best outcome for each feedback method would be that the participants of the group have a score that is constantly above 100 percent and at a steady high level.

5.2.6 Relapse sensitivity

This we can measure in the same fashion as the effectiveness. If we can accurately measure the physical activity level of each individual participant, we can also see which participants will fall back onto their original level of activity, which should indicate a relapse into old habits moment. Another measurement is to see how many participants quit before the end of the experiment. The desireable outcome of this measurement is to have a low percentage of participants who exhibit signs of relapse.

5.2.7 Quality of coaching

This is measured by a standard survey as set up by Philips, which they have used in the past to measure the quality of coaching of their DirectLife system. We will give both groups of users the same survey as used before by Philips and compare the two with each other. The questions on quality of coaching can be found in appendix page 70,71 and 72.

5.2.8 Intention to use longer

At the end of the experiment all users will be asked in their post experiment survey on whether they could continue the current experiment if needed, whether they would repeat the current experiment and whether they approved of the duration of the experiment. This will be rated on a 7 point likert scale, and compared between both the ECA group and the e-mail group.

5.3 Participants

The people approached to participate in it are from the original target audience i.e. people with occupations that require them to sit for extensive periods of time, or are in any other ways attempting to get a more active balanced life pattern. This resulted in two groups, all over twenty years of age, half of them would be asked to use the already existing e-mail system, while the other half is going to use the video feedback system. Both groups were randomly chosen and were before the end of the experiment unaware of the existence of the other feedback presentation method. During the experiment technical difficulties resulted in the early withdrawel of five participants reducing the ammount of effective participants from 20 to 15. When we look at how the group is divided by gender and age more exactly we get the following result:

Table 5: Statistic distribution of participants

Age	Males ECA	Females ECA	Males E-mail	Females E-mail	total
20-35	3	2	5	1	11
35-50	0	1	1	0	2
50+	1	1	0	0	2
Sum:	4	4	6	1	15

Next to the basic information about age and gender they were also required to give information about their height and weight in order for the system to be able to calculate the amount of calories burned during the training program.

Another important factor to be able to more accurately evaluate the results and opinions of the participants is the effect of previous experiences of participants with

Weight	F	M	total
45-60	2	0	2
61-75	3	2	5
76-85	0	4	4
85+	0	4	4

(a) Height

Height	F	M	total
150-165	2	0	2
166-180	3	1	4
181-190	0	6	6
191+	0	3	3

(b) Weight

BMI	F	M	total
17-19	1	0	1
19-21	1	1	2
21-23	3	2	5
24+	0	7	7

(c) BMI

Table 6: Callory burning information

similar systems, and with the most important part of the system, namely the embodied conversational agent. It is interesting to see whether people with alot of experience generally judge it differently (i.e. more positively or more negative) than their counterparts with little to no experience.

Experience	participants ECA	E-mail
none	5	0
little	0	0
moderate	1	4
Well informed	1	3

(a) Embodied agents

Experience	participants ECA	E-mail
none	5	0
little	1	3
moderate	1	2
Well informed	1	2

(b) Behavior modifaction systems

Table 7: Experience tabels

The last part of information that the participants in the experiments were asked to provide is an estimate of their current state of fitness as well as the motivation for them to use the system. In this question there were multiple(up to three) possible selectable answers, but one answer that stood out as chosen most by far was the answer "just for fun". Which after further asking most participants when asked for their interest at the start of the experiment indicated that they were only interested in assistence in completing the experiment, rather than actually really using the system for its actual intentions. Nonetheless the second most popular answer here was "to increase general health" so the secondary objective of the users of the system was what the system is designed for.

When we look at how fit the users of the system estimate themselves, measured by the three questions about their activity in the intake, then we can see the following:

Fitness	participants ECA	E-mail
no activity	0	0
little	2	2
moderate	2	1
very active	4	4

So in total both groups have similar knowledge of the system in use, and similar fitness. Due to the drop outs within both groups the average age in the ECA group is quite high compared to e-mail group and the male to female ratio in both groups is quite different as well seeing as how 2 out of three of the females dropped out in group 2 while all the drop outs in group one were men.

5.4 Using the system as a client

When a participant of the system is making use of the video feature, the participant will be informed by e-mail whenever he or she has a new message available. This e-mail will then tell the user to go to the website to log in.

Figure 8: log in screen

This screen is a copy of the already existing DirectLife screen with some minor adjustments to show that this was a demonstration server and not the same. After loging in the participant would then get an index of all the received messages, with a color difference between viewed and unviewed messages.

In this index the participant can select a specific video that he or she wants to view at that time, it is also possible to remove the older videos if the participant wishes to no longer have access to it. The system in fact will only remove the possiblity to link to the video, the video itself will not be removed until explicitly to do so by the system administrator. After selecting a video the user will be taken to the video view screen.

During some initial experimentation with early video versions it was discovered that pronouncing the basic statistics namely percentage of goal reached, amount of calories

Figure 9: video index

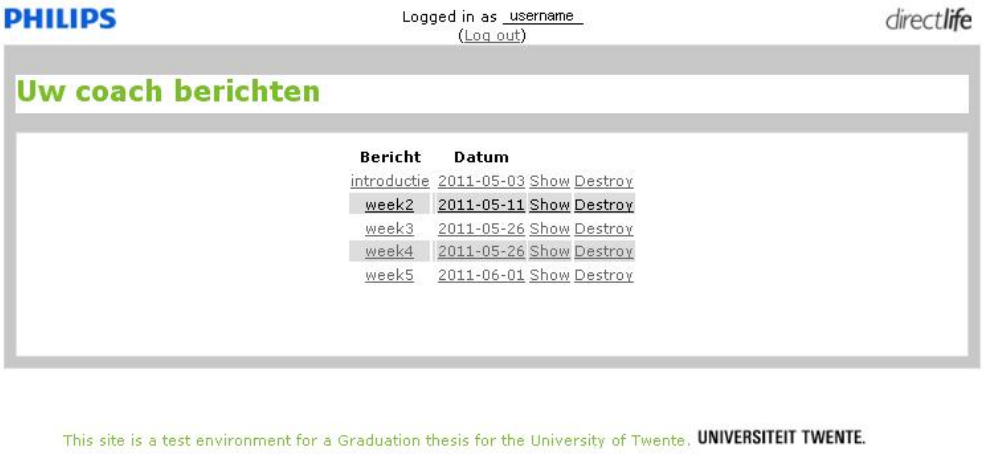


Figure 10: e-mail example

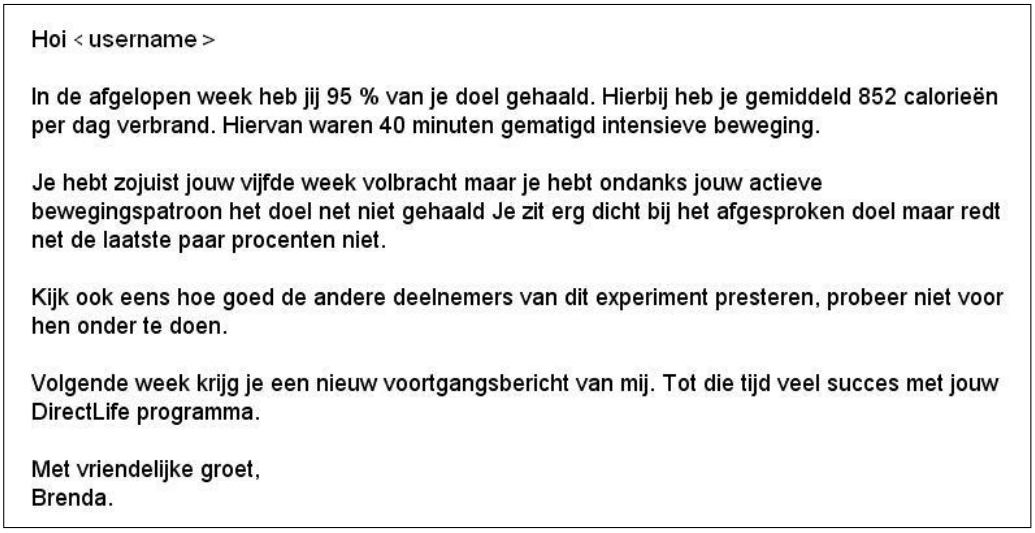
burned, high and low intensity minutes, was by all initial tests considered boring, and making the message unnecessery longwinded. Since the ultimate aim of the current embodied agent implementation was to integrate the agent with the original site, the experiment system has a similar sheet of weekly results with a performance bar that

changes color. The embodied agent will refer to the information during the video.

5.5 E-mail response

The e-mail response is typically built up in a similar fashion as the script, so basically we have four distinct types of e-mail, firstly the introduction, which is a long text introducing the coach, be it purely as the name Brenda, and information on how to use the system, the second type is the feedback on the assesment week, the third type is the standard progress mail, giving the calories, activity duration and percentage of goal, followed by a short explanation, and the last paragraph will contain a basic health advice that is selected specifically for that persons profile and performance. In the following image we can see an example of such an e-mail as sent by the system.

Figure 11: E-mail coaching


Hoi < username >

In de afgelopen week heb jij 95 % van je doel gehaald. Hierbij heb je gemiddeld 852 calorieën per dag verbrand. Hiervan waren 40 minuten gematigd intensieve beweging.

Je hebt zojuist jouw vijfde week volbracht maar je hebt ondanks jouw actieve bewegingspatroon het doel net niet gehaald Je zit erg dicht bij het afgesproken doel maar redt net de laatste paar procenten niet.

Kijk ook eens hoe goed de andere deelnemers van dit experiment presteren, probeer niet voor hen onder te doen.

Volgende week krijg je een nieuw voortgangsbericht van mij. Tot die tijd veel succes met jouw DirectLife programma.

Met vriendelijke groet,
Brenda.

6 Experiment results

The original intent was to make use of two sets of connected experiment results. On one side we have the data that is produced by the system, while the rest of the information is obtained through the inquiries. This was the original intent, however through failure of the activity monitor hardware (only 2 of the original 24 testmodels went through the full experiment duration without the battery failing on the participant, with a large quantity of the activity monitors having failing batteries at some point of the experiment the output of the system fell back to the baseline zero activity output. So this removes the physical activity level and the percentage of goal level of the output elements of our experiments.

6.1 Feedback sessions

There were two feedback sessions during the entire experiment, one of the two was done via e-mail in which the participants were asked to fill in a pdf form and reply to give an indication of their opinion of the system so far. The second feedback session was a personal interview with questionnaire where both the participant of the experiment and the person overseeing the experiment were present. This choice for this was done by attempting to combine maximum results with practical feasibility.

6.1.1 Quality over quantity

The design of the experiment was such that in order to keep it manageablee the amount of participants within the experiment group could not be too high, but to still get results that were statistically sound, we need to make use of validated questionnaires and measure all activity that was possible to measure and significant for the outcome of the experiment.This means that for each feedback session all questions needed to be verifiable, and important, and all the activity of the participant needs to be registered as well. Aside from this, each piece of information that is sent to the participants of the experiment is archived in a database and timestamped, and the e-mail address that is used for the participant-system communication is also used for question feedback. Each of the responses to this e-mail address is forwarded and archived giving us a complete overview of all available information. The full details of all results can be found in the tables at the end of the appendix page 78.

6.2 Results

For the following dimensions of the experiment we have had enough results to draw a conclusion.

6.2.1 Likability

If we compare both methods of presentation for likability then we found that during the half way period that the performance of the ECAis evaluated with a 5.275 mean

value with a variance of 1.165 on a seven points Likert scale while the e-mail system receives a 4.143 with a 0.61 variance. At the post experiment survey when asked the same set of questions the average likability of both systems increased, in case of the ECA group to 5.625 with a variance of 0.78 and the e-mail group to 4.62 with a variance of 1.07. During the last survey the ECA group was on all elements of the survey on likability more positive and more consistent than during the half way survey. When we asked the ECA group how then would estimate for themselves the change of opinion on the experiment, half of them recognized that they enjoyed it more in the end. When we compare the total sets of numbers by T-test we can see that there is a significant difference in likability between the e-mail group and the ECA group with a p value of 0,02 at the half way point and 0.035 at the post experiment point.

When we look at the individual results of both presentation methods we also see that the opinion about the presentation methods became more positive during the experiment, though there was only a significant difference on the Likert scale on friendly- unfriendly with a p value of 0.038 in a paired T-test for the ECA group, and for the e-mail group the only significant difference was on the Like-dislike scale with a p value of 0.008. Participants judged the behavior of the ECA as more natural and realistic by the end of the experiment than during the half way point with an increase of an average 0.5 point on a seven point Likert scale, though there was no change to the system or production method of the presentation. The same was for the e-mail group, by the end of the experiment they were more positive about the language use and general feedback than during the half way point. This change in opinion though consistent is small and can be caused by external causes or random influence.

When checking for internal consistency and reliability of the figures we get a Cronbach's Alpha value of 0.896 at the half way point for likability and 0.905 for the post experiment test, indicating a high reliability.

6.2.2 Perceived intelligence

When we compare both presentation methods for perceived intelligence we find that during the half way survey that the mean value of the response is 5.00 for the ECA with a variance of 1.01 and for the e-mail group 4.17 with a variance of 0.83 and a Cronbach's alpha value of 0.776. While the perceived intelligence increased and became more consistent in the responses from the ECA group it remained fairly similar for the e-mail group and became more varying by the end. The mean value of the perceived intelligence of the ECA presentation method was 5.13 with a variance of 0.65 while the mean value of the response of the e-mail group was 4.14 with a variance of 1.49. When we compare these numbers by T test we can see that the final result is a significant difference with a p value of 0.05 with a Cronbach's alpha value of 0.876. During the half way point the difference is not large enough to be conclusive.

6.2.3 Fear

On the subject of how fearfull the users are toward the agent when they received the coaching the data is a bit less consistent then with the likability and perceived intelligence. The mean value of the response from the ECA group were 5.6 on a 7 points likert scale with a vartiation of 0.95, where a score of 1 is the most fearful, while the e-mail group had a response with a mean value of 4.31 with a variance of 0.73 with a Cronbach's alpha value of 0.835. By the end of the experiment however this completely changed where the mean value of the ECA dropped very slightly to 5.5 with a variance of 0.57 while the e-mail group increased to 5.23 with a variance of 0.77, with a reliability Cronbach's alpha value of 0.859. So while at the half way point it seems that users are more fearful of the e-mail messages, by the end of the experiment there is no significant measurable difference between both groups. When we compare each scale individually by a paired T-test between the half way survey and the end survey then we can see only a significant drop in comfortable-uncomfortable scale with a p value of 0.026 for the ECA group, while at the e-mail group there is a significant increase on the scale comfortable-uncomfortable with a p value of 0.049.

6.2.4 Intention to use longer

When we look at the intention to use longer or repeat the experiment and compare the two on how they answered survey questions on this, the response from the ECA group had a mean value of 5.06 on a scale of 1 to 7 with a variation of 2.86 while the mean value of the e-mail group was 4.06 with a variation of 3.62. These results exclude the three members that indicated that they had not done all six weeks of the course. Here we see that the mean value of the ECA group is higher, but the variation in both groups is so high that it is quite possible caused by random influences other than the presentation method.

6.2.5 Quality of coaching

The quality of coaching was measured after completion of the entire experiment. This is the part where we find the largest difference between the ECA method of presentation and the e-mail method of presentation. The mean value on 21 questions on a scale of 1 to 7 for the ECA group was 4.24 with a variation of 1.63 while the e-mail group response had a mean value of 2.88 with a mean value of 0.81. Here we see a significant difference with a p value of 0.016. In tabel 9 we can see the full results per question on how each question was answered compared between both the ECA group and the E-mail group.

6.2.6 General adherence and interest

The overall adherence to the system was heavily influenced by the way the experiment was set up. One of the methods used to measure the adherence amongst the video users was to check the latency between the information given that a video was available and the actual watching of the video by that user. It turned out that since some users used

an e-mail adress that was related to their profession that they did not watch the movies at a regular base, since they were informed that they had a movie at work but since they were working it was not possible to watch the feedback video at that time.

Next to this, there was the factor of uptime, during the third week of the experiment, the ability to contact the DirectLife server was interrupted for several days to nearly a week for most users which resulted in a large decline in interest and even resulted in people not watching the videos at all.The e-mails in this case were still read but with less interest, users actually sent e-mails at this time that the lack of server functionality made them stop checking the content of their messages. From this we can conclude that the overal system uptime and the address of the participant were of a far larger influence to the adherence and interest of the participant than the method of communication.

Also when we look at the results from the measured interest we see that there was an increase in likability towards the system, as more users rated it as more likable,yet at the same time when asked whether they would actually buy it, non of the participants trusted the system enough to be worth the money, though this was often due to technical failures during the experiment, and had little to do with the coach performance. When we look at the groups individually than the interest of the system changes differently, with the e-mail users one person changed his opinion to a more negative one and the majority of the group felt consistently the same about the system, and one person consistently liked it more, while one single person was disappointed by the systems both in the third as in the sixth week, exact numbers of the response can be found in the appendix in the last page.

6.2.7 Performance of users

When we look at the performance of all the users, it is very difficult to get a clear image of how they worked towards their goals. The amount of data that was corrupted by malfunctioning batteries was simply too great and caused too many weeks to yield 0 percent of goal reached and a physical activity level of zero or near zero. Seeing as how incomplete this data is we will not make any further attempt on drawing any conclusions from it.

6.3 User experiences

There have also been some interesting replies on the open questions, which also explain some of the deviations in user experience, for example one of the users said that the amount of trust towards the system has very little to do with the communication of the system, but rather that he "does not trust any commercial cooperation with more information than is necessary for the system to function". Another often heard complain was that the webserver was "Down too often" and "processing data takes too long". These last two have little to do with the user experience as an effect of the communication, but do have a negative impact on the overall user experience. A typically Dutch response was also given when asked about the interest to buy the commercial variant, five out of 15 interviewed persons in the end survey responded with "Not at that price". When

asked about the performance of the text to speech client three responses were "She mispronounced my name, that annoyed me." indicating that they do feel like it is aimed toward them and it is disturbing when pronounciation goes wrong.

6.4 Interpretation of results

For the most part, the sample sizes of seven and eight active participants per group were insufficient to show conclusive significant differences between both groups. However we can see that for both groups over time their opinion changed positively. Most participants appreciated the system more (either e-mail group or ECA-group) at the end of the experiment than at the beginning. At the same time the quality of coaching, attractiveness, percieved intelligence and fear have all been evaluated in favor of the ECA. Next to that, at the end of the experiment we can see that the variation on most points drops making the opinions of the participants more likeminded, this could be an indication that a longer experiment with more participants would have given an even more accurate result with less variation, however this is pure speculative.

7 Discussion

There are some points within this experiment that still need to have an exact conclusion and require further testing, in this chapter I would like to put forward how to further handle the hypothesis, to get a more complete image of it all.

7.1 Implication for hypothesis

We previously defined system of feedback A is better than system of feedback B if it performs better on the elements of effectiveness, persuasiveness, relapse sensitivity, attractiveness, percieved intelligence and trust versus fear. When we look at the results from the experiment that is is not clear which of the two feedback methods is more effective or persuasive, there is not enough data to support or reject the idea whether there is a difference and in favor of which feedback system that difference would be. However on there were some notable differences on the other elements. The outcome of the experiment strongly indicates that the feedback system with an ECA is percieved as more pleasant in use but both systems would improve significantly if they functioned as properly as the consumer version.

7.2 Recommendations

This results from this experiment are restricted because of the proof of concept setup of the whole setup. If we were to adjust the system and perform a larger experiment, with more participants and a continuous guarantee on functionality for all the hardware such as in a professional environment we could get more accurate results, next to that there are a few other points that needs to be addressed.

7.2.1 Success of method

A question that is more important that still needs to be answered is the effectiveness of the coaching methods. Do participants of DirectLife who use the virtual embodied agent coach actually perform better than a group of people that are coached by means of e-mail? To get a more definitive answer on this question the experiment needs to be performed again, for a more extensive period and with more participants for a better control. Along with this there is another element that needs to be verified and that is the frequency with which the feedback is given. During the experiment the weekly feedback from the coach is far more intense than a standard training program, however when people were asked specifically about the frequency of the feedback all were very happy with this and would not have prefered more or less feedback, though the timing of the feedback is something most would have liked more control over.

7.3 Useful techniques that are unusable for practical reasons

Some of the previous named weapons and processes might not be usable due to the limitations as provided by the test setup. For example if we wanted to make the virtual

coach more similar to the user as a form of social proof use, then we cannot do this since we are at the time limited to the Elckerlyck embodied agent which cannot be adjusted to all possible user profiles (yet). Some other techniques were used but could have been used more effectively if there was more time to develop these, for example, with the embodied agent there is a need to automatically generate speech and for the experiment we were restricted to one specific text to speech generator, namely the Saskia loquendo voice. If we wanted to make more use of authority it would have been desirable to get a speech synthesizer that is capable of giving a speech in in this more suiting way, or the existing one needed to be "tweaked" extensively to more closely give the performance needed.

7.4 Techniques that are unusable for this concept

Some of the techniques that exist are practically unusable in the context of this subject, their basic principles simply are not part of this experiment environment. For example the scarcity tactic. There are no resources available that we can show as being in short supply in order to convince the user to perform better in his behaviour, except perhaps with a time indicator to demonstrate that the user has only a small portion of time left to achieve his goal. However since time cannot be gained back it might not give the proper result and could very likely just stress the user out.

Next to that it has shown in previous direct life experiments that actively instructing the people on an activity while pointing out a specific day to do it, is counter productive.

7.4.1 What influences can invoke relapse?

In a group where succes is not 100 percent, there is bound to be failure. In this case failure would be indicated by relapse of the system user into old activity patterns. The duration of this experiment however is a bit on the short side to be able to see effects of relapse, however technical problems can cause a significant loss in interest in the use of the system causing relapse. Basically whenever the decissional balance of the participant in shifted in such a way that it no long benefits change after the user has already gone through the whole process of change we will notice relapse. Under normal circumstances where the system is working properly and the participants have also invested their money in the use of the system it is far more likely that this kind of relapse will occur after the first six weeks than within.

7.4.2 Commitment

Because it is known to the participants that they have entered an experimental version of the DirectLife system and because of the social convinience nature of the participants group, the statistics of commitment are likely to have been distorted when compared to normal. Aside from that basic fact, there is the fact of hardware malfuction that played a large influence. All except for one of the drop outs within the experiment have commented that they did not feel like participating any further due to either the

failure to connect the activity monitor to the server, or the lack of results from using the monitor(due to the failing battery). The fact that four people still finished their program though the last two weeks their device was not actually functioning anymore is a demonstration of one of the weapons of influence, namely commitment to the experiment. To see whether the virtual embodied agent differs significantly from an e-mail system we would have to compare a more realistic setting where the participants commitment is not toward an experiment, but completely to their own improvement.

7.4.3 Recommitment

A somewhat surprising outcome of the experiment is the existence of recommitment in such a short timespan. While it is true that in the current live system there is a tendency for users to take a break from the system after a period of 5 to 6 weeks. Due to the problems that existed with the batteries of the activity monitor, and the limited availability of the webserver in Eindhoven there were four individuals that stopped for one week, but who turned out much more motivated after their faulty activity monitor had been replaced with a correctly functioning one, all four of whom turned out to be active in the embodied agent portion of the experiment. Seeing as how the group of four on a testgroup of 15 is too small to draw any serious conclusions, it seems possible that the embodied agent is more effective at creating a feeling for recommitment.

7.5 Future work

With the completion of this specific experiment it is also time to take a look at what it can bring in the future. First of all due to the technical difficulties experienced in this experiment it is clear that in order to get more definitive results the experiment needs to be performed again, over a longer period, with a larger group of participants and with better functioning hardware. That being said there are some point aside from the technical and statistical aspects that were pointed out by the participants and were found during the experiment.

1. **3D graphics** - the interface in its current webform has been used since it is the quickest multiplatform method of working. It is pre-rendered and therefor a 2D movie application. By making use of a program that only needs BML as input the platform can become quicker to load and more interactive, since it would be directly 3D instead of 2D.

2. **Text variation** - Though this was largely influenced by the lack of function of the activity meter, participants in both groups complained about the repetition of sentence fragments. This was generally experienced as annoying, though none of the same sentences were actually used twice at the same person. Because the general idea of the sentence varied little they still experienced it as repetition. So increasing the variety of responses or reducing the amount of times that the system provides responses could fix this. An important point is here that the actual live

system only provides feedback with one third of the frequency of the experiment system, so in reality this problem might not even really exist.

3. **mobile support** - Though the experiment was initially set up to include the support of mobile phone and similar devices, when it proved difficult to include text information as overlay in such a device this element was no longer paid any attention to. However with the development of a separate mobile device application as for android, IOS, and possibly symbian, combined with push messages full mobile phone support is quite possible to create.

4. **Interactivity** - This was a point that users of the embodied agent group pointed out on several levels, the most basic level was that they were not allowed to choose their coach, while if there was the time to develop several agents and text to speech variations, this should be easily realized. Secondly there was the basic complaint about the inability of the embodied agent to pronounce names, especially if the origin was not basic Dutch, the name "Bozhi" which has a chinese origin was impossible for the agent to realise, as well as the more standard name "Auke" and "Elly". Some interactive way to correct the pronounciation of the persons name could be quite useful here such as allowing the participant to self phonetically write their name for the embodied agent. This last method would perhaps not result into a 100 percent correct match, but it would generate a pronounciation that the participant self would find acceptable, (such as when you go to a foreign country and people find it hard to pronounce your name you can suggest a shorter more easy version to them that is also acceptable to you). The last level and highest level of interactivity that was asked about was to talk to the agent itself. If the coach feedback was not a prerendered movie but a real time application with some basic communicative capabilities it would be possible and quite desireable to for example insert the frequently asked question system in the embodied agent.

7.6 Opportunities

There are alot of opportunities in the use of embodied agents as a coach, when implemented correctly they can be successful, more successful than pure text agents. And in some cases due to their nature of being a computer program they could even surpass a human coach since they can be available 24/7.

8 Conclusion

We wish to draw a clear conclusion to whether or not the feedback by an embodied agent is better than feedback by e-mail, where the e-mail is generated automatically as well. To see what constitutes as "better" this has been done on several elements.

The experiment so far gives an indication that it generally is percieved as a more attractive with a P value of 0.018 at the half way point 0.035 at the end , and is percieved to be more intelligent with a P value of 0.05 at the end, when compared to a system that provides the same information by e-mail.

When asked for an opinion of the quality of coaching participants strongly favored the ECA as well with a P value of 0.016. Since both groups were provided the same information just by a different method, this should be an indication that users prefer to get a coaching message spoken by somebody that looks at them rather than by a message that they need to read.

When we look at the results for how fearful the participants are towards the messages there is a less clear difference between both feedback methods, the difference that exists at the end of the experiment is still in favor of the ECA but it is smaller than that of attractiveness and percieved intelligence though at the half way point there seemed to be a significant difference in favor of the ECA, this might have been caused by other random influences.

On the point of relapse sensitivity there were somewhat more remarkable results. Due to the nature of the change that the experiment deals with we expected little to no effects on relapse and recommitment. However the influence of hardware during the experiment did lead to people quitting before the end of the experiment and before the experiment was over recommitting to it. The recommitters were all part of the ECA group. Though this is hardly significant and could very easily be contributed to random chance it did stand out.

For the intent of future use, the results of the surveys vary far to strong to form a conclusion, the small difference in favor of the ECA can easily have been caused by random c hance.

On the points of effectitveness and persuasiveness there was unfortunately not a complete enough dataset to judge wether one system was superior to the other or whether they differed at all.

8.1 Final thoughts

So we cannot clearly say from this experiment that the ECA coaching method is better than the e-mail coaching method. However it does indicate that it is more attractive, perceived as more intelligent and the general quality of coaching is evaluated as higher.

9 Acknowledgments

First of all I wish to thank my good friend Rolf Oostebring for sacrificing some of his free time to assist me with the creation of coaching feedback text, his experience as a profesional coach and instructor was invaluable.

Next to that I wish to express my gratitude towards all participants of the experiment, who took the effort to completely voluntarily participate in my experiment for six weeks.From their response I have seen that it was not always as easy to keep up all the way to the end of the experiment, and I am glad that enough stayed as long as they did for me to get proper results.

I would also like to thank Aart van Halteren who provided guidence and support from Philips, and who provided all the needed software and hardware support to keep the experiment going.

I would also like to thank Betsy van Dijk, Randy Klaassen en Rieks op den Akker for their guidence throughout this research project.

I would also like to mention Hendri Hondrop whose tireless efforts in obtaining loquendo licenses helped me keep the experiment going for as long as it took.

9.1 Special thanks

Special thanks goes out to my parents and family who supported me in many ways throughout the duration of my studies and who have always stood behind me. If it was not for their support I would not have been able to finish my studies.

在此，我想由衷的感谢游博之的父母，游小铁先生和李艳霞女士。感谢他们无私奉献和对我们的全力支持。没有他们支持，他们的女儿就无法陪我在荷兰就学，在精神上和经济上支持我。我深知任何形式都无法回馈他们对我们的贡献，都无法表达我对他们感谢。

And last but definitly not least special thanks go out to 博之 who has shown more patience with me and who has put up with more from me than anybody,

感谢我的未婚妻，感谢她无限的耐心以及给予我的精神支持

.

References

[1] Smarcos artemis homepage. Online webpage. http://www.smarcos-project.eu/ Last viewed 16 juni 2011.

[2] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita. Is the Uncanny Valley an Uncanny Cliff? In *Proceedings of the 16 th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2007*, pages 368–373. IEEE, 2007.

[3] Christoph Bartneck, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. My robotic doppelganger - a critical look at the Uncanny Valley. In *18th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN2009*, pages 269–276. IEEE, November 2009.

[4] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, 1(1):71–81, January 2009.

[5] Christoph Bartneck, Dana Kulic, Elizabeth Croft, Susana Zoghbi, C. Bartneck, D. Kulic, E. Croft, S. Zoghbi, E. Croft, and S. Zoghbi. Original paper measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots, 2008.

[6] Christoph Bartneck, Marcel Verbunt, Omar Mubin, and Abdullah A. Mahmud. To kill a mockingbird robot. In *2nd ACM/IEEE International Conference on Human-Robot Interaction*, pages 81–87, Washington DC, 2007. ACM Press.

[7] Timothy Bickmore, Daniel Mauer, Francisco Crespo, and Thomas Brown. Persuasion, task interruption and health regimen adherence. In *Proceedings of the 2nd international conference on Persuasive technology*, PERSUASIVE'07, pages 1–11, Berlin, Heidelberg, 2007. Springer-Verlag.

[8] Timothy Bickmore, Daniel Mauer, Francisco Crespo, and Thomas Brown. Negotiating task interruptions with virtual agents for health behavior change. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 3*, AAMAS '08, pages 1241–1244, Richland, SC, 2008. International Foundation for Autonomous Agents and Multiagent Systems.

[9] Timothy W. Bickmore and Rosalind W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.*, 12:293–327, June 2005.

[10] Timothy W. Bickmore, Kathryn Puskar, Elizabeth A. Schlenk, Laura M. Pfeifer, and Susan M. Sereika. Maintaining reality: Relational agents for antipsychotic medication adherence. *Interacting with Computers*, 22:276–288, 2010.

[11] Justine Cassell, Tim Bickmore, Lee Campbell, Hannes Vilhjalmsson, and Hao Yan. Conversation as a system framework: Designing embodied conversational agents, 2000.

[12] Robert B. Cialdini and Noah J. Goldstein. Social influence: Compliance and conformity. *Annual Review of Psychology*, 55:591–621, 2004.

[13] B. J. Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002, December 2002.

[14] Rachel M Henke, Ginger S Carls, Meghan E Short, Xiaofei Pei, Shaohung Wang, Susan Moley, Mark Sullivan, and Ron Z Goetzel. The relationship between health risks and health and productivity costs among employees at pepsi bottling group. *Journal of occupational and environmental medicine American College of Occupational and Environmental Medicine*, 52(5):519–527, 2010.

[15] Colleen A. Redding Verna Burden Nancy Perrin Michael Leo Marna Flaherty-Robb Janice M. Prochaska James O. Prochaska, Susan Butterworth. Initial efficacy of mi, ttm tailoring and hris with multiple behaviors for employee health promotion. volume 46, pages 226–231. Elsevier Inc, 2007.

[16] Robbert jan Beun, Eveliene De Vos, and Cilia Witteman. Embodied conversational agents: Effects on memory performance and anthropomorphisation. In *Intelligent Virtual Agents*, pages 315–319, 2003.

[17] Kristinn R. Thorisson Justine Cassell. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13:519 – 538, 1999.

[18] Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn Thórisson, and Hannes Vilhjálmsson. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. *Intelligent Virtual Agents*, pages 205–217, 2006.

[19] Dana Kulić and Elizabeth Croft. Pre-collision safety strategies for human-robot interaction. *Auton. Robots*, 22:149–164, February 2007.

[20] Jan Miksatko, Kerstin H. Kipp, and Michael Kipp. The persona zero-effect: evaluating virtual character benefits on a learning task with repeated interactions. In *Proceedings of the 10th international conference on Intelligent virtual agents*, IVA'10, pages 475–481, Berlin, Heidelberg, 2010. Springer-Verlag.

[21] Stanley. Milgram. *Obedience to authority : an experimental view / Stanley Milgram.* Harper & Row, New York :, 1974.

[22] Yohei Murakami, Yuki Sugimoto, and Toru Ishida. Modeling human behavior for virtual training systems. In *In AAAI*, pages 127–132, 2005.

[23] Harri Oinas-Kukkonen. Behavior change support systems: A research model and agenda. In Harri Oinas-Kukkonen Thomas Ploug, Per Hasle, editor, *Persuasive Technologies*, pages 4–14. Springer-Verlag Berlin Heidelberg, June 2010.

[24] Elin Olsen and Pål Kraft. epsychology: a pilot study on how to enhance social support and adherence in digital interventions by characteristics from social networking sites. In *Proceedings of the 4th International Conference on Persuasive Technology*, Persuasive '09, pages 32:1–32:6, New York, NY, USA, 2009. ACM.

[25] Robert B. Cialdini PH.D. *Influence - The psychology of persuasion*, chapter 2,3,4,5,6, pages 23–52,53–86,97–135,136–166,168–212. HarperCollins e-books, March 2009.

[26] Daniel Schulman and Timothy Bickmore. Persuading users through counseling dialogue with a conversational agent. In *Proceedings of the 4th International Conference on Persuasive Technology*, Persuasive '09, pages 25:1–25:8, New York, NY, USA, 2009. ACM.

[27] Daniel G. Spencer and Richard M. Steers. The influence of personal factors and perceived work experiences on employee turnover and absenteeism. *Academy of Management Journal*, 23(3):567–572, September 1980.

[28] V. Strecher and S. Nolen-Hoeksema. Internet methods for delivering behavioral and health-related interventions (eHealth). *Annu. Rev. Clin. Psychol.*, 3:53–76, 2007.

[29] Victor J. Strecher, Brenda McEvoy DeVellis, Marshall H. Becker, and Irwin M. Rosenstock. The Role of Self-Efficacy in Achieving Health Behavior Change. *Health Education & Behavior*, 13(1):73–92, March 1986.

[30] H. van Welbergen, D. Reidsma, Z.M. Ruttkay, and J. Zwiers. Elckerlyc - a bml realizer for continuous, multimodal interaction with a virtual human. *Journal on Multimodal User Interfaces*, 3(4):271–284, 2010. ISSN=1783-7677.

[31] Darren E.R. Warburton and Shannon S.D. Bredin. Health benefits of physical activity. *Canadian Medical Association Journal*, 175:777, 2006.

[32] R. West. Time for a change: putting the Transtheoretical (Stages of Change) Model to rest. *Addiction*, 100(8):1036–1039, 2005.

10 Appendix

Enviromnent details followed by all used inquiries.

script generator environment:
Java SE Runtime Environment 1.6.0_22-b04
java version: 1.6.022
Web environment:
Ruby version : 1.8.7
Gems version: 1.3.7
Rails version: 2.3.5
Webserver:
Mongrel
BML Realizer platform
Elckerlyk: Amandia BML realizer demo environment
Text to speech:
Loquendo version 7.3. stem Saskia (Dutch)
Loquendo version 7.2 stem Saskia(Dutch) after week 4 of the experiment due to licensing.

Figure 12: Users manual

Gebruikshandleiding

U staat op het punt om te beginnen met uw persoonlijke Direct Life programma. Omdat dit niet een programma is dat onder het standaard Direct life systeem van Philips valt, zijn er een aantal punten waarop u moet letten gedurende dit programma.

- Feedback: Alle feedback binnen het systeem dient afkomstig te zijn van Direct.life.coach@gmail.com. Het is mogelijk dat er vanuit de testomgeving enkele andere e-mails gegenereerd worden. U wordt vriendelijk verzocht deze te negeren aangezien ze geen nuttige informatie voor u bevatten.

- Website gebruik 1: Tijdens uw programma wordt u ingelogd op de systeem website van Direct Life zodra u gebruik maakt van het USB-dock. Deze site is NIET de standaard commerciële site, maar een experimentele test omgeving. Probeer daarom niet contact te zoeken met de mensen van de commerciële omgeving, dit kan voor problemen zorgen omdat zij niet bekend zijn met u als gebruiker.

- Website gebruik 2. Tijdens uw programma heeft u ook de mogelijkheid om gebruik te maken van het tabblad met de titel “mijn coach”. Dat tabblad geeft informatie bevat die bruikbaar is voor de gebruikers van het standaard systeem. U maakt geen gebruik van het standaard systeem, en daarom zal dit onderdeel van de website voor u ook geen nuttige informatie bevatten. U wordt daarom ook vriendelijk verzocht geen gebruik te maken van deze pagina.

- Website gebruik 3: De tweede website waarvan u gebruik kunt maken is de video website, iedere keer zodra u een video boodschap heeft, zult u per e-mail hierover op de hoogte gebracht worden. U kunt vervolgens inloggen met de gegevens die bij deze handleiding zijn meegegeven.

Dit zijn de punten waar u rekening mee moet houden. Om nu met uw programma te kunnen beginnen, kunt u uw Activity monitor met uw computer docken, en vervolgens inloggen met uw gegevens. Doorloop alle stappen en klik vervolgens zodra u de keuze heeft op “Start assesment now”.

Figure 13: Consent form dutch edition

Toestemmingsverklaring

U wordt hierbij gevraagd deel te nemen aan een experiment dat deel uitmaakt van mijn (Adri Wiekens) afstudeerprogramma. Dit onderzoek is in samenwerking met Philips, en betreft het zogenaamde Direct Life systeem, een systeem dat erop gericht is om mensen te helpen met het ontwikkelen van een gezonder bewegingspatroon. Om deel te kunnen nemen aan dit onderzoek dient u in te stemmen met de volgende voorwaarden, indien u hiermee instemt verzoek ik u het akkoord aan te vinken:

Enquête gegevens

Tijdens dit experiment zult u op drie momenten worden gevraagd deel te nemen aan een enquête. De informatie die u invult op deze documenten zal strikt vertrouwelijk behandeld worden en niet gedeeld worden met derden. Tijdens het invullen zijn er geen foute antwoorden, De interesses tijdens dit onderzoek zijn enkel in uw persoonlijke meningen en inzichten. De gecombineerde resultaten van het onderzoek zullen verwerkt worden in een verslag, welke publiekelijk toegankelijk is. Geen enkel element van de persoonlijke informatie zal tot u te herleiden zijn.

Akkoord O

Activiteitsdata

Tijdens het verloop van dit experiment wordt van u verwacht dat u de activiteitsmeter zoveel mogelijk bij u hebt. De gemeten persoonlijke informatie zal verder nooit gepubliceerd worden en enkel dienen om de functionaliteit van het systeem aan te sturen en om enkele statistische inzage te creëren.

Akkoord O

Vrijwillig

Deelname aan dit experiment is volledig op vrijwillige basis, u kunt zich gedurende het experiment op ieder moment terugtrekken, zonder opgegeven reden.

Akkoord O

Ondergetekende heeft alle voorwaarden gelezen en gaat akkoord met alle opgestelde voorwaarden en wenst deel te nemen aan dit onderzoek.

Akkoord O

Naam deelnemer : _____ Naam : Adri Wiekens

Handtekening : _____ Handtekening : _____

Datum : ___/___/___ Datum : ___/___/___

Indien u aan het einde van het onderzoek een korte samenvatting van de onderzoeksresultaten wenst te ontvangen vul dan hier uw E-mail adres in.

E-mail adres: _____

Figure 14: Consent form english edition

Informed consent form

With this form you are officially invited to participate in an experiment as part of my (Adri Wiekens) graduation thesis. Within this research there is a cooperation with Philips with as main subject their so called "Direct Life" system, a system which aim is to assist in the adjusting the behavior of participants into a more healthy pattern of activity. In order to take part in this research you are required to consent with the following conditions. Please tick the consent circles if you agree with the posed conditions.

Inquiry data

In the course of this experiment you will be asked to answer a total of three questionnaires forms, one at the beginning of the experiment, one during the half way point, and one at the end of this experiment. The information that you share on these inquiries will be treated with strict confidentiality and will not be shared with third parties. During the filling in of the answers of each questionnaire, it is important to realize that there are no "wrong answers." The main focus of this research is on the personal opinions and insights of the participants. The combined total results of the inquiry will be processed into the final thesis report, which will be available publically. The given personal information will not be traceable to the participant of this experiment.

Agree O

Activity data

During the experiment you are asked to wear your activity monitor as much as possible. The measured personal data by the activity monitor will never be published. The personal information will never be published and only serves as data feed to the system and to give some extra statistic insight.

Agree O

Voluntary

Participation in this experiment is completely voluntary and the participant is free to withdraw at any time, without giving any reason.

Agree O

Undersigned participant has read all and understood all elements of this consent form and agrees to participate in this study.

Agree O

Participant's name	: _____	Researcher: Adri Wiekens	
Signature	: _____	Signature	: _____
Date	: ___/___/___	Date	: ___/___/___

In case you would like to receive a summary of the research results at the end of this experiment you can leave your E-mail address here:

E-mail adres: _____

Figure 15: pre-experiment inquiry form - page 1

Voornaam:	……	Straatnaam:	……
Achternaam	……		
Aanspreek naam:	……	Huisnummer	……
E-mail adres:	……	Postcode	……
wachtwoord[1]:	……	Woonplaats	……
AM nummer		Gewicht	… kg
Geboorte datum	…./…./19…	Lengte	…, …. m

Wat is voor u een reden om uw bewegingspatroon aan te passen? (meerdere antwoorden zijn mogelijk).	o Stressreductie o Energie verhogen o Voor de lol o Een uitdaging zoeken o Bewonderd worden. o Sociale bezigheid o Winkansen verhogen voor sporten o Op dokters advies o Ziekte preventie o Algemene gezondheid verbeteren o Gewichtscontrole o Er beter uitzien o Fitheid verhogen o Leniger/flexibeler worden
Wenst u dat uw Direct Life profiel zichtbaar of onzichtbaar is?	o Zichtbaar o Onzichtbaar
Bent u gemiddeld genomen redelijk actief voor meer dan vijf keer per week 30 minuten per dag? :	o Ja o Nee
Bent u de laatste zes maanden gemiddeld genomen actief geweest?	o Ja o Nee
Wilt u actiever worden?	o Ja o Nee
Waar bent u van plan de Activity Monitor[2] te gaan dragen?	o Broekzak o Ketting o Borst o Riem

[1] Dit is het wachtwoord van uw experimentele Direct Life, deze data zal gecensureerd worden opgenomen in het verslag, echter het is nog steeds niet verstandig om hier een privacy gevoelig wachtwoord in te vullen aangezien het via papier verwerkt wordt vul dus NIET hier uw eigen e-mail wachtwoord of iets soortgelijks in..

[2] Dit is het apparaat dat uw activiteit meet en dat u op uw computer aan kunt sluiten om uw bewegingsdata door te geven aan het systeem.

Figure 16: pre-experiment inquiry form - page 2

Het al bestaande Philips Direct life system kost € 99 denkt u na afloop van dit onderzoek geïnteresseerd te zijn in dit systeem? Kies een mogelijkheid.	o Ja, want ik wil mijn bewegingspatroon aanpassen. o Ja, want het lijkt me wel leuk om te doen. o Ja, maar om een andere reden, namelijk ……. o Nee, want ik ben tevreden over mijn bewegingspatroon. o Nee, voor een experiment is leuk, maar meer hoeft voor mij niet. o Nee, want het lijkt mij niet echt leuk. o Nee, voor een andere reden namelijk ……. o Misschien, hangt af van hoe het experiment verloopt.
Kunt u een schatting geven over uw ervaring en kennis van Embodied agent[3] systemen? Doe dit op een schaal van 0 tot 10 waar 0 totaal geen ervaring is en 10 zeer veel ervaring en kennis.	0 1 2 3 4 5 6 7 8 9 10 O O O O O O O O O O O
Hoeveel kennis/ervaring heeft u met systemen voor gedragsverandering[4]?	0 1 2 3 4 5 6 7 8 9 10 O O O O O O O O O O O
Welke besturingssystemen gebruikt u het meest tijdens dit experiment? (meerdere antwoorden mogelijk)	o Windows XP of ouder o Windows Vista o Windows 7 o OSX o Linux o Android o IOS (Iphone / Ipad) o Symbian o Anders namelijk: …….. :
Welke Webbrowser bent u van plan te gebruiken voor dit systeem?	o Internet Explorer o Firefox o Chrome o Safari o Anders namelijk:

[3] Een embodied agent is een lichamelijke representatie van een systeem waarmee u communiceert, voorbeelden hiervan zijn interactieve robots of avatars in computer omgevingen. Indien u vrijwel geen van dit soort systemen kent kies dan 0.

[4] Het Direct Life systeem is een gedragsaanpassing systeem. Met systemen als deze worden mensen geholpen die wensen hun standaard gedrag aan te passen.

Figure 17: Half way survey - page 1

Voornaam: …… Achternaam …… E-mail adres: ……	
Gewicht	… kg

Je bent nu halverwege het Direct Life experiment programma, deze enquête is bedoeld om een verder inzicht te ontwikkelen in jouw mening en gebruik van het systeem.

Wat voor indruk heb je van jouw coach(Brenda) na haar feedback?

Algemene indruk Brenda

Ik vind haar:

Leuk	1 2 3 4 5 6 7 O O O O O O O	Verschrikkelijk
Incompetent	1 2 3 4 5 6 7 O O O O O O O	Competent
Goed geinformeerd	1 2 3 4 5 6 7 O O O O O O O	Onwetend

Terwijl zij mij feedback gaf voelde ik mij:

Comfortabel	1 2 3 4 5 6 7 O O O O O O O	Onaangenaam
Gespannen	1 2 3 4 5 6 7 O O O O O O O	Ontspannen

Figure 18: Half way survey = page 2

Ik vind haar:

Onverantwoordelijk	1 2 3 4 5 6 7 O O O O O O O	Verantwoordelijk
Onvriendelijk	1 2 3 4 5 6 7 O O O O O O O	Vriendelijk
Onintelligent	1 2 3 4 5 6 7 O O O O O O O	Intelligent
Aardig	1 2 3 4 5 6 7 O O O O O O O	Onaardig

De indruk die ik van mijn coach heb is:

Ik mag haar	1 2 3 4 5 6 7 O O O O O O O	Ik mag haar totaal niet
Ze is aangenaam	1 2 3 4 5 6 7 O O O O O O O	Ze is Onaangenaam
Gunstig voor mij	1 2 3 4 5 6 7 O O O O O O O	Ongunstig voor mij
Ze is onbenaderbaar	1 2 3 4 5 6 7 O O O O O O O	Ze is toegankelijk

Figure 19: Half way survey page 3

Terwijl zij mij feedback geeft voel ik mij::

Nerveus	1 2 3 4 5 6 7 O O O O O O O	Zeker
Rusteloos	1 2 3 4 5 6 7 O O O O O O O	Kalm

De feedback speeches die de coach aan mij gegeven heeft ervaar ik als:

Persoonlijk	1 2 3 4 5 6 7 O O O O O O O	Generiek / Algemeen

Figure 20: Half way survey page 4

Kan je hier aangeven hoe je de coach beoordeelt in algemeen gedrag?

Kijk gedrag									
Ik vond het kijkgedrag van de embodied agent:	Nep	1	2	3	4	5	6	7	Realistisch
		O	O	O	O	O	O	O	
Als ik de feedback van de embodied agent bekijk dan vind ik het kijkgedrag:	Machineachtig	1	2	3	4	5	6	7	Mensachtig
		O	O	O	O	O	O	O	
De manier waarop de embodied agent mij aankeek vond ik:	Aangenaam	1	2	3	4	5	6	7	Onaangenaam
		O	O	O	O	O	O	O	
Kan je een algemene beschrijving geven over waar je het kijkgedrag van de coach het beste mee te vergelijken vindt? (bijvoorbeeld een nieuwslezer of een echte coach of een ander embodied agent presentatie) ….. …….. …….. …….. …….. …….. …….. …….. …….. …….. …….. …. …. …. …. ….. …….. …….. …….. …….. …….. …….. …….. …….. …….. …….. …….. ……..

Taalgebruik en stemgedrag									
De manier waarop de coach tegen mij sprak ervoer ik als:	Machineachtig	1	2	3	4	5	6	7	Mensachtig
		O	O	O	O	O	O	O	
Het taalgebruik van de embodied agent ervoer ik als:	Aangenaam	1	2	3	4	5	6	7	Onaangenaam
		O	O	O	O	O	O	O	
Waar ik mij het meest aan stoorde:….. ….. …. ….. …….
Wat ik het meeste kon waarderen::….. ….. …. ….. …….

Figure 21: Half way survey page 5

Systeem feedback:

Geef een cijfer aan de algehele kwaliteit van de feedback, geef ook aan waarom.	1 2 3 4 5 6 7 O O O O O O O Waarom? ……
Geef een cijfer voor de hoeveelheid informatie die gegeven wordt tijdens de feedback, 1 is veel te weinig, 4 precies goed, 7 veel te veel	1 2 3 4 5 6 7 O O O O O O O Waarom? ……

Opinie veranderingen: Tijdens uw eerste inschrijving heeft u aan moeten geven hoe uw interesse is ten op zichte van dit systeem. Is dit veranderd?	O Ja, ik vind het nu leuker/beter O Ja, het valt me tegen O Nee, mijn mening is hetzelfde
Geef een cijfer over of u het systeem zou aanraden voor andere mensen om te gaan gebruiken, 1 is ik zou het zeker niet doen, 7 ik zou het zeker doen.	1 2 3 4 5 6 7 8 9 10 O O O O O O O O O O Waarom? ……
Verdere algemene opmerkingen: ……..	

Figure 22: Post experiment survey - page 1 - E-mail version

Voornaam:	……
Achternaam	……

Selfefficacy

Geef alsjeblieft aan hoeveel vertrouwen je erin hebt dat je actief kunt zijn in de volgende situaties

Helemaal geen vertrouwen	1
Een beetje vertrouwen	2
Redelijk vertrouwen	3
Wel vertrouwen	4
Erg veel vertrouwen	5

	1	2	3	4	5
Wanneer ik moe ben.	O	O	O	O	O
Wanneer ik een slecht humeur heb.	O	O	O	O	O
Wanneer ik het gevoel heb dat ik geen tijd heb.	O	O	O	O	O
Wanneer ik op vakantie ben.	O	O	O	O	O
Wanneer het regent of sneeuwt.	O	O	O	O	O

Figure 23: Post experiment survey - page 2 - E-mail version

Quality of coaching

Op dit formulier wordt je gevraagd over de kwaliteit van de coaching die je hebt ontvangen, geef bij deze vragen aan hoe vaak de uitspraak bij jou van toepassing is over een bereik van 1 tot en met 7, waar 1 nooit betekend en 7 altijd.

Nooit	1
Zelden	2
Soms	3
Best vaak	4
Vaak	5
Heel vaak	6
Altijd	7

	1	2	3	4	5	6	7
Mijn coach geeft advies over hoe ik onder druk of in moeilijke situaties kan presteren.	O	O	O	O	O	O	O
Mijn coach geeft advies over hoe ik kan doorzetten.	O	O	O	O	O	O	O
Mijn coach geeft advies over hoe ik vertrouwen kan blijven houden in mijn vaardigheden.:	O	O	O	O	O	O	O
Mijn coach geeft advies over hoe ik positief over mezelf kan blijven.	O	O	O	O	O	O	O
Mijn coach geeft advies over hoe ik gericht kan blijven (op mijn activiteitsdoelen).	O	O	O	O	O	O	O
Mijn coach reikt me strategieën aan om mijn doelen te halen.	O	O	O	O	O	O	O
Mijn coach houdt mijn voortgang richting mijn doel in de gaten.	O	O	O	O	O	O	O

Figure 24: Post experiment survey - page 3 - E-mail version

	1	2	3	4	5	6	7
Mijn coach Helpt me om korte-termijn-doelen te stellen en actieplannen te maken.	O	O	O	O	O	O	O
Mijn coach helpt me om richtdatums te kiezen om (tussen)doelen te behalen	O	O	O	O	O	O	O
Mijn coach geeft me ondersteuning om mijn doelen te behalen.	O	O	O	O	O	O	O
Mijn coach laat zien dat hij/zij me begrijpt.	O	O	O	O	O	O	O
Mijn coach kan goed luisteren.	O	O	O	O	O	O	O
Mijn coach is gemakkelijk benaderbaar over eventuele persoonlijke	O	O	O	O	O	O	O
Mijn coach laat zien dat hij/zij om me geeft (dus ook andere aspecten van mijn leven en niet alleen lichamelijke activiteit)	O	O	O	O	O	O	O
Mijn coach kan ik mijn persoonlijke problemen toevertrouwen.	O	O	O	O	O	O	O
Mijn coach gaat vertrouwelijk om met mijn persoonlijke problemen.	O	O	O	O	O	O	O
Mijn coach helpt me om mijn activiteitsprestaties te herkennen en vieren.	O	O	O	O	O	O	O

Figure 25: Post experiment survey - page 4 - E-mail version

Mijn coach helpt me om motivatie en inspiratie uit anderen te halen	1 2 3 4 5 6 7 O O O O O O O
Mijn coach helpt me te ontdekken welke dingen het best voor mij werken om actiever te worden.	1 2 3 4 5 6 7 O O O O O O O
Mijn coach heeft de juiste kennis en vaardigheden om goede coaching te geven.	1 2 3 4 5 6 7 O O O O O O O
Mijn coach geeft coaching van goede kwaliteit.	1 2 3 4 5 6 7 O O O O O O O

Figure 26: Post experiment survey - page 5 - E-mail version

Algemene indruk Brenda

Ik vind haar:

Leuk	1 2 3 4 5 6 7 O O O O O O O	Verschrikkelijk
Incompetent	1 2 3 4 5 6 7 O O O O O O O	Competent
Goed geinformeerd	1 2 3 4 5 6 7 O O O O O O O	Onwetend

Terwijl zij mij feedback gaf voelde ik mij:

Comfortabel	1 2 3 4 5 6 7 O O O O O O O	Onaangenaam
Gespannen	1 2 3 4 5 6 7 O O O O O O O	Ontspannen

Figure 27: Post experiment survey - page 6 - E-mail version

Ik vind haar:

Onverantwoordelijk	1 2 3 4 5 6 7 O O O O O O O	Verantwoordelijk
Onvriendelijk	1 2 3 4 5 6 7 O O O O O O O	Vriendelijk
Onintelligent	1 2 3 4 5 6 7 O O O O O O O	Intelligent
Aardig	1 2 3 4 5 6 7 O O O O O O O	Onaardig

De indruk die ik van mijn coach kreeg was:

Ik mag haar	1 2 3 4 5 6 7 O O O O O O O	Ik mag haar totaal niet
Ze is aangenaam	1 2 3 4 5 6 7 O O O O O O O	Ze is Onaangenaam
Gunstig voor mij	1 2 3 4 5 6 7 O O O O O O O	Ongunstig voor mij
Onbenaderbaar	1 2 3 4 5 6 7 O O O O O O O	Toegankelijk

Figure 28: Post experiment survey - page 7 - E-mail version

Tijdens het ontvangen van de feedback voelde ik mij

Nerveus	1 2 3 4 5 6 7 O O O O O O O	Zeker
Rusteloos	1 2 3 4 5 6 7 O O O O O O O	Kalm

De feedback speeches die de coach aan mij gegeven heeft ervoer ik als:

Persoonlijk	1 2 3 4 5 6 7 O O O O O O O	Generiek / Algemeen

Hoe groot schat je de toegevoegde waarde van de e-mail communicatie ten opzichte van de web-interface?	Overbodig	1 2 3 4 5 6 7 O O O O O O O	waardevol

Experiment duur			
Het experiment geef aan wat je vond van de duur van het experiment.	Te lang	1 2 3 4 5 6 7 O O O O O O O	Te kort
Hoeveel moeite had je met het afronden van de volledige duur	Weinig	1 2 3 4 5 6 7 O O O O O O O	Veel
	O Ik heb heb de volledige 6 weken gedaan O Ik ben afgehaakt voordat het programma af was		
Ik zou graag het programma nog langer doorlopen.	Waar	1 2 3 4 5 6 7 O O O O O O O	Onwaar
Ik zou graag het programma nog langer doorlopen.	Waar	1 2 3 4 5 6 7 O O O O O O O	Onwaar
Verdere opmerkingen over de duur van mijn programma. ….. …….. …….. …….. …….. …….. …….. …….. …….. …….. …. …. …. …. ….. …….. …….. …….. …….. …….. …….. …….. …….. …….. …….. …….. …….. ……..			

Figure 29: Post experiment survey - page 8 - E-mail version

Systeem feedback	
Geef een cijfer aan de algehele kwaliteit van de feedback, geef ook aan waarom.:	1 O · 2 O · 3 O · 4 O · 5 O · 6 O · 7 O · 8 O · 9 O · 10 O Waarom? …………………………………………………
Geef een cijfer voor de hoeveelheid informatie die gegeven wordt tijdens de feedback, 1 is veel te weinig, 4 precies goed, 7 veel te veel	Te weinig · 1 O · 2 O · 3 O · 4 O · 5 O · 6 O · 7 O · Te veel

Opinie veranderingen: Tijdens uw tweede inschrijving heeft je aan moeten geven hoe je interesse is ten op zichte van dit systeem. Is dit veranderd?	O Ja, ik vind het nu leuker/beter O Ja, het valt me tegen O Nee, mijn mening is hetzelfde
Geef een cijfer over of u het systeem zou aanraden voor andere mensen om te gaan gebruiken, 1 is ik zou het zeker niet doen, 7 ik zou het zeker doen.	afraden · 1 O · 2 O · 3 O · 4 O · 5 O · 6 O · 7 O · Aanraden
Verdere algemene opmerkingen ……..	

Figure 30: Post experiment survey - page 8 - Video version

Kun je hier aangeven hoe je de coach beoordeeld in algemeen gedrag?

Kijk gedrag									
Ik vond het kijkgedrag van de embodied agent:	Nep	1	2	3	4	5	6	7	Natuurlijk
		O	O	O	O	O	O	O	
Als ik de feedback van de embodied agent bekijk dan vind ik het kijkgedrag:	Machineachtig	1	2	3	4	5	6	7	Mensachtig
		O	O	O	O	O	O	O	
De manier waarop de embodied agent mij aankeek vond ik:	Aangenaam	1	2	3	4	5	6	7	Onaangenaam
		O	O	O	O	O	O	O	

Taalgebruik en stemgedrag									
De manier waarop de coach tegen mij sprak ervoer ik als:	Machineachtig	1	2	3	4	5	6	7	Mensachtig
		O	O	O	O	O	O	O	
Het taalgebruik van de embodied agent ervoer ik als:	Aangenaam	1	2	3	4	5	6	7	Onaangenaam
		O	O	O	O	O	O	O	

10.1 Experiment figures

Table 8: Self-efficacy

Question	$\overline{X}$ **group 1**	σ **group 1**	$\overline{X}$ **group 2**	σ **group 2**	Δ
1	3.25	0.5154	1.8571	1.069	+1.393
2	2.875	0.4828	2.8571	0.800	+0.018
3	2.625	0.8377	3.1429	1.464	-0.518
4	3.625	1.2557	4.4286	0.787	-0.804
5	2.875	1.2516	3.5714	0.787	-0.696

Table 9: Quality of coaching

Question	$\overline{X}$ **group 1**	σ **group 1**	$\overline{X}$ **group 2**	σ **group 2**	Δ
1	3.25	0,898	1.714	0.829	+1.536
2	3.75	0.657	3.143	0.922	+0.607
3	3.625	1.297	3.429	0.778	+0.196
4	3.875	0.970	3.429	1.327	+0.446
5	4.625	0.838	4.000	1.732	+0.625
6	4.375	0.672	4.143	0.922	+0.232
7	4.375	0.672	4.714	1.039	-0.340
8	5.125	0.483	2.571	1.017	+2.554
9	4.375	0.800	2.429	1.443	+1.946
10	4.375	0.493	3.143	1.251	+1.232
11	4.75	1.280	2.143	1.274	+2.607
12	4.125	1.386	2.000	0.957	+2.125
13	4.375	1.405	2.286	0.508	+2.089
14	3.5	1.931	2.571	1.362	+0.929
15	3.5	1.931	1.571	0.606	+1.929
16	5.125	1.268	2.000	0.957	+3.125
17	4.375	0.493	3.000	1.414	+1.375
18	3.375	0.672	2.714	1.765	+0.661
19	4.125	0.544	3.000	0.957	+1.125
20	4.75	0.125	3.286	0.705	+1.464
21	5.25	0.718	3.286	1.353	+1.964

10.1.1 Half way point

Table 10: agent performance

Question	$\overline{X}$ **group 1**	σ **group 1**	$\overline{X}$ **group 2**	σ **group 2**	Δ
Likability	5.275	0.511	4.143	0.226	+1.321
Percieved intelligence	5.000	1.050	4.171	0.396	+0.829
fear	5.600	0.930	4.314	1.286	+1.286

Table 11: Opinion change

Group	**Appreciate more**	**Appreciate same**	**Appreciate less**
Group 1	2	4	2
Group 2	1	5	1

10.1.2 Sixth week survey

Table 12: Agent performance

Question	$\overline{X}$ **group 1**	σ **group 1**	$\overline{X}$ **group 2**	σ **group 2**	Δ
Likability	5.625	0.884	4.629	1.036	+0.996
Percieved intelligence	5.125	0.807	4.143	1.220	+0.982
fear	5.500	0.756	5.229	0.875	+0.271
added value	5.375	1.297	3.429	2.149	1.946

Table 13: Opinion change

Group	**Appreciate more**	**Appreciate same**	**Appreciate less**
Group 1	4	4	0
Group 2	1	4	2

10.1.3 Performance of E-mail conversation

Table 14: e-mail engine performance

Question	$\overline{X}$ **week 6**	σ **week 6**
human - machine	3.714	1.380
pleasant - unpleasant	4.429	1.134
realistic - fakse	4.286	1.254

10.1.4 Performance of E-mail conversation

Table 15: e-mail engine performance

Question	$\overline{X}$ **week 6**	σ **week 6**
human - machine	3.714	1.380
pleasant - unpleasant	4.429	1.134
realistic - fakse	4.286	1.254

10.1.5 Performance of The embodied agent

Table 16: agent performance

Question	$\overline{X}$**week 3**	σ **week 3**	$\overline{X}$ **week 6**	σ **week 6**	Δ
Realism gaze	4.25	1.832	4.75	1.488	+0.5
Humanlike gaze	4.125	1.959	4.125	1.126	0
Pleasantness gaze	4.125	1.458	4.875	1.126	+0.75
Humanlike language	4.25	1.753	4.75	0.886	+0.5
pleasantness language	4.375	1.188	5.125	0.835	+0.75