Chapter 8
Virtual Coaches for Healthy Lifestyle

H.J.A. op den Akker, R. Klaassen and A. Nijholt

Abstract Since the introduction of the idea of the software interface agent the question recurs whether these agents should be personified and graphically visualized in the interface. In this chapter we look at the use of virtual humans in the interface of healthy lifestyle coaching systems. Based on theory of persuasive communication we analyse the impact that the use of graphical interface agents may have on user experience and on the efficacy of this type of persuasive systems. We argue that research on the impact of a virtual human interface on the efficacy of these systems requires longitudinal field studies in addition to the controlled short-term user evaluations in the field of human computer interaction (HCI). We introduce Kristina, a mobile personal coaching system that monitors its user's physical activity and that presents feedback messages to the user. We present results of field trials ($N = 60$, 7 weeks) in which we compare two interface conditions on a smartphone. In one condition feedback messages are presented by a virtual animated human, in the other condition they are displayed on the screen in text. Results of the field trials show that user motivation, use context and the type of device on which the feedback message is received influence the perception of the presentation format of feedback messages and the effect on compliance to the coaching regime.

H.J.A. op den Akker · R. Klaassen · A. Nijholt (✉)
Human Media Interaction, University of Twente,
217, 7500 AE Enschede, The Netherlands
e-mail: a.nijholt@utwente.nl

H.J.A. op den Akker
email: h.j.a.opdenakker@utwente.nl

R. Klaassen
e-mail: r.klaassen@utwente.nl

A. Esposito and L.C. Jain (eds.), *Toward Robotic Socially Believable Behaving Systems - Volume II*, Intelligent Systems Reference Library 106,
DOI 10.1007/978-3-319-31053-4_8

8.1 Introduction

We are at the coffee corner of Jones' office. Jones has a tough day and he is in need now of another coffee. Facing the machine Jones routinely presses the GO button. The machine says "select your item" and Jones presses the coffee-with-sugar button. When the machine requests authorization he holds his card in front of the machine's card reader. Jones expects that the machine will now ask him to wait a while. Instead, it says: "Mr. Jones, may I remind you that this is your fourth coffee today. The weather is nice. A short walk will do you well. That is healthier than another coffee."

Clearly, Jones' coffee machine is not just a coffee machine. It is integrated into a pervasive system that also keeps record of his daily consumptions. It uses information about his physical activity, that it has read from his step counter, a built-in function of his smart watch. It knows about the local weather. The system uses all this information to carry out its task which is not just to serve coffee on demand but to care about mr. Jones' well being. The machine has become a representative agent of a system that tries to nudge its clients to live a more healthy life. This is an example of a *behavior change support system* or personal coaching system. It uses persuasive strategies and principles from captology, the study of persuasive technology, to bring about its intended effect on its users.

In their 1993 paper "Learning Interface Agents" Pattie Maes and Robyn Kozierok describe an *intelligent interface agent* that learns how to sort messages into the different mailboxes created by the user. The authors use the metaphor of the "personal assistant" that is collaborating with the user in the same work environment [41]. The interface agent learns by continuously "looking over the shoulder" when the user is performing actions (p. 461). It stores situations (information about incoming e-mail) with the actions performed by the user in these situations. When a new mail arrives the agent suggests an action based on the similarity between the new situation and previously seen cases. The agent also learns from the user who gives either explicit feedback, e.g., by correcting the agent's action, or implicit feedback, e.g., by ignoring a suggestion by the agent.

Since the seventies of the 20th century we witness several changes in the paradigm for human computer interaction and in the metaphors we use when we think and talk about interacting with computers. First, the computer changed from a computing device that has to be programmed by expert users and scientists through a command line interface to a communication medium that could also be used by nonexpert users. Alan Kay's Dynabook, designed in 1968, was intended for children! [35]. From the desktop metaphor according to which the user is "navigating the information space" by means of graphical items, icons representing administrative functions of the computer, the paradigm shifted to that of the conversation. Today it is quite common to think about the computer as a social agent. The personal computer has become a person itself.

The discussion between proponents of interface design methods within the intelligent agent paradigm, e.g., Pattie Maes, and opponents of this paradigm, e.g., Ben Schneiderman, advocates of the direct manipulation or desktop metaphor, boils down

to a common concern about who is in control of the system and who is responsible for what the system does [61]. Systems become more and more complex and proponents argue that interface agents that understand the user and act in a proactive way make systems accessible to nonexpert users. The interface should provide intuitive means of natural interaction and software agents should act on behalf of their users. Opponents see the danger of delegating tasks to proactive autonomous software agents. The system may run out of control of the user. The problem is who is responsible then.[1]

A difference between the personal assistant of Kozierok and Maes [41] and the social caring agent behind Jones' coffee machine is in the types of work spaces in which the collaboration between system and user takes place. Maes' personal assistant is concerned with a very specific information processing task. The system behind Jones' coffee machine tries to support Jones to change his sedentary life style -typical for most office workers that look at computer screens for many hours a day- into a more healthy one. Daily personal life has become a design space. Machine learning techniques such as the one Maes and Kozierok implemented in their personal assistant are also used in Jones' lifestyle change support system. After some time it will learn if Jones likes to be mothered by the system. The acceptance of this type of personal persuasive systems is a real issue. The feeling of being in control appears to be an important factor.

Using interface agents does not necessarily mean that they are anthropomorphized and that they should be graphically presented as a human in the interface. It is, however, interesting to see that the graphical interface of the interactive learning assistant contains a caricature that conveys the functional and cognitive state of the agent to the user. Where computers make computations of which only the outcome is interesting for the user, the personal assistant approaches the user in a social way: it asks questions, suggests solutions, it even tries to persuade the user to take a break instead of having another coffee. As soon as the technology is there to graphically represent the agent on the interface it seems hard to resist to do so.

[1] From the very beginning of the rise of captology ethical implications of persuasive technology have been considered. Fogg [26] devotes a chapter to it. Heckman and Wobbrock [31] point at the fact that "anthropomorphic agents provide electronic commerce consumers with terrific benefits yet also expose them to serious dangers". They conclude that developers must focus agent design on consumer welfare, not technical virtuosity, if legal and ethical perils are to be avoided. In a special issue of the Communications of the ACM [7] present ethical principles of persuasive system design in a framework in which they model persuasive technology as an instrument between the persuader and the persuadee. Based on new insights in the phenomenology of technology [66] considers persuasive technology as an example of technological mediation. New technologies reshape human practices and views on reality. People don't need cars or internet when cars or internet have not been invented. Ethical concerns are not restricted to "intended persuasions" and "unintended outcomes" of the use of technology. For example, persuasive technology made it possible to frame the act of Jones having a coffee within a framework of healthy living and makes him aware of the fact that the system is looking over his shoulder; maybe more as a big brother than as a caring sister. Being a stakeholder in technology development researchers and designers of new technologies have special responsibilities for the mediating implications of these technologies.

Advances in computer graphics made it possible to build *animated graphical interface agents*. Such agents appear as embodied characters that exhibit lifelike human behaviors and that convey affective state and communicative intentions to the user by gesturing and gaze behavior [24]. The animated embodiment is often just the visual appearance of an *embodied conversational agent* (ECA) [21]. Under the hood the ECA is a more or less sophisticated (spoken or multi-modal) dialogue system. Cassell [19] advocates the use of ECA interfaces and to represent a system as a human in those cases where social collaborative behavior is key. The purpose of the ECA is "to leverage users' knowledge of how to conduct face-to-face conversation" and to "leverage users' natural tendencies to attribute humanness to the interface." (p. 82).

The terms "virtual human", "embodied conversational agent" or "intelligent virtual agent" for "computer models resembling humans in their bodily look and their communication capabilities" [59] are often used interchangeably. Virtual humans can however play quite different roles in research and in systems. Depending on the interest of the researcher virtual humans are built and studied as simulation of real humans with a certain personality (e.g., [13]) or as animated interface agents. The term *interface* agent suggests that the function of the virtual human is that of communication channel between a user and the system. When a virtual human plays—for example—the role of a suspect in a serious game for interrogative interview training the social interaction itself with the simulated suspect is what the system is made for. Hence, it depends on the specific function that the virtual human has what graphical, behavioral, cognitive and affective features need to be implemented.

Technology aided coaching on healthy behavior is widely regarded as a promising paradigm to aid in the prevention of chronic diseases, in the adherence to medical treatment and in the process of healthy ageing in general [33]. In order to encourage physical activity in patients suffering from chronic diseases, as well as healthy adults, many different coaching systems have been developed. Typically these consist of an activity sensor and some form of coaching application delivered either through a web portal, smartphone or through the sensor itself [53]. These personal coaching systems more and more try to make use of the progress made in the field of natural interfaces, spoken dialogue systems and embodied conversational agents. Since the human coach is paradigmatic for the personal digital coach it lies at hand to present the digital coach by an (animated) virtual human, a talking head or an ECA [36].

In this chapter we look at the use of these embodied conversational agents as a natural interface of persuasive systems that care for their user's well being and health. Does the use of an ECA add to the acceptance, the usability, the efficacy, the persuasiveness, the attractiveness of these type of systems? And if they do, is it just because of the (animated) graphical appearance of a face or is it because of the naturalness of the interaction made possible by the speech interface? Or, are other factors into play that influence these outcome measures?

An important issue is the method for evaluating the effects of the interface design on the efficacy of behavior change support systems. Experience with therapies for patients with chronic conditions (low back pain, COPD, diabetes) show that adherence to the therapy in the first weeks is no guarantee for adherence on the longer

term. Thus to measure the efficacy of coaching systems long-term experiments seem to be required. In research environments most user evaluation studies are performed with early stage technologies and through short-term lab experiments. Research has shown that the use of animated graphical agents often raises the entertainment value of a system. Dehn and van Mulken [24] point at the fact that short-term user studies might not be appropriate to measure the long-term effect on utilitarian performance measures. Although the entertainment value of the interface design may motivate the user to seek interaction with the system, in the long run usefulness is fundamental for long-term user engagement.

Klasnja et al. [39] argue that "behavior change in the traditional clinical sense is not the right metric for evaluating early stage technologies that are developed in the context of HCI research." They argue that because of the complexity of behavior change support systems randomized control trials are not suitable to answer the question *what* exactly causes the outcome of such trials. Instead HCI evaluation methods should be developed that focus on the efficacy of *specific persuasive strategies* implemented in the persuasive system, such as for example self-monitoring, or social dialog. The question we focus on here is how the use of a virtual human in the interface impacts the effect of certain persuasive strategies used.

To find an answer to this question we will first in Sect. 8.2 review what type of persuasive features and strategies are included in persuasive systems. In Sect. 8.3 we will review research findings concerning the effects of virtual humans as interface agents of systems in which perceived credibility and persuasiveness play a role. In Sect. 8.4 we will present Kristina, a digital coaching system that supports and motivates people to live a balanced and healthy lifestyle by providing feedback about their level of physical activity [52]. Kristina is designed as a multi-device (PC, Mobile, TV) system that is able to monitor physical activity and medication intake of its users.[2] It implements a number of persuasive strategies and the virtual human used has some features to improve persuasiveness. In Sect. 8.5 we present a six week user evaluation study of the Kristina coaching system in which we compare two different ways of presenting coaching messages, with or without the use of an ECA in the interface. In Sect. 8.6 we present lessons learned regarding the use of virtual humans as user interface in this type of persuasive systems.

8.2 Persuasive Communication and Persuasive Technology

Will Jones follow the advice of the coffee machine and take a break instead of having another coffee? A change in someone's opinion, attitude or behavior is often the result of persuasion, in which someone, who has the role of *persuader*, tries to influence the *persuadee* through a process of communication. In this section we

[2]The system is built in the European ARTEMIS project on Smart Composite Human-Computer Interfaces (Smarcos) that focused on inter-usability aspects of multi-devise/multi-sensor service systems (2010–2012). https://artemis-ia.eu/project/24-smarcos.html.

review the most important theories of persuasive communication and persuasive technology. We do that in order to see in the next section in what way the use of virtual humans in the interface of a persuasive system could impact the persuasiveness of the communication.

8.2.1 Persuasive Communication

The process of persuasive communication has four elements each with a number of aspects [9]:

1. Message: timing, the argument, the emotional value
2. Source: expertise, trustworthiness, credibility, attractiveness (liking, similarity), role model, power
3. Channel: media types (newspaper, tv, internet), devices, modalities (visualisation, spoken language)
4. Recipient: intelligence, self-esteem, state of change, motivation, needs, personality

Petty and Cacioppo [56] developed a model of how recipients elaborate on the message they receive. According to this Elaboration Likelihood Model there are two routes for persuasion to behavior and attitude change: the central route, where the receiver elaborates on the (argumentative or affective) content of the message, and the peripheral route where the receiver responses to peripheral cues. The way the receiver elaborates on the message and involves himself in the content can vary. Internal processes on the side of the recipient influence the persuasiveness of the communication (see Fig. 8.1): attention, comprehension, yielding, and retention.

Regarding the message content: does it matter for the persuasiveness of the argument if the message brought by the persuader contains evidence -statements referring to others than the persuader to support his claim? In a study with students as audience of two speakers on two different topics [47] found that although good use of evidence increases perceived authoritativeness, it does not appear to have an effect on perceived character. Topic and persuader's gender may however mediate the effect of evidence use on the perceived character. Moreover, speeches which include good use of evidence produce a greater attitude shift than those which do not.

But often receivers of a persuasive message are not motivated to elaborate on the content of the message. Instead, they evaluate and respond to it by looking at some peripheral cues. Common used cues are related to *source credibility* (we follow a suggestion because we believe it comes from a credible source), to *liking* (the persuader is an attractive person) and to *social consensus* (we believe, do and need what our social peers believe, do and need).

In order to be successful in bringing about a change in the persuadee's behavior receiver and persuasive communication must satisfy some requirements. The Transtheoretical Model of Behavior [57] is a popular model that describes the process people go through when changing their behavior. According to Prochaska change in

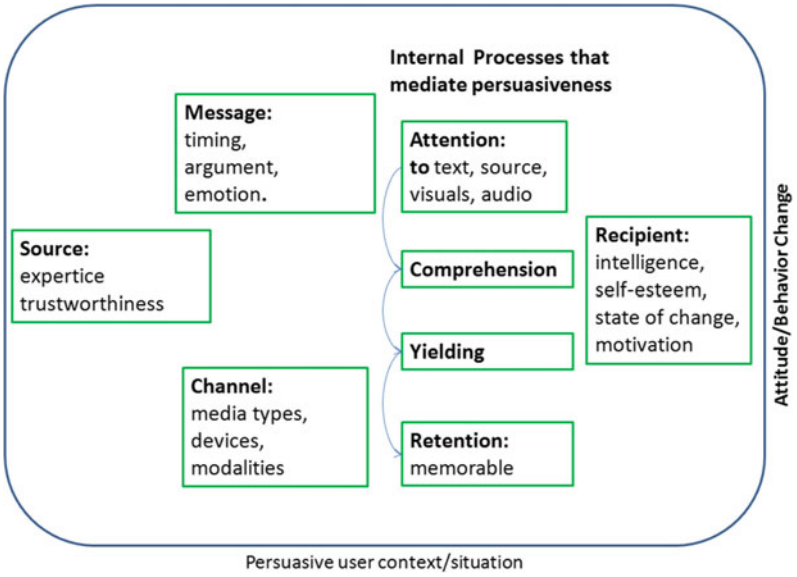

Fig. 8.1 A model of persuasive communication (based on [55])

behavior can be seen as a process involving different stages: (1) pre-contemplation, (2) contemplation, (3) preparation (or determination), (4) action and (5) maintenance. A persuader, no matter whether it is a human or a persuasive system, should adept it's intervention strategy to the persuadee's state of change. When Jones is already motivated to be more physical active a motivational interview may not be the effective way to intervene. The self-efficacy theory by Bandura [2] is integrated into the Transtheoretical Model. Self-efficacy is the confidence people have in their own ability to deal with specific situations without returning to old negative behaviors. Self-efficacy is strengthened naturally through success. A behavior change support system can make users explicitly aware of their successes by showing them monitored data and by giving them positive feedback.

The persuasive behavior model of Fogg [27] provides three factors that determine if people can be persuaded to change their behavior: (1) Motivation, (2) Abilities, and (3) Triggers. The model asserts that for a target behavior to happen, a person must have sufficient motivation, sufficient ability to perform the intended behavior, and an effective trigger. All these three factors should be present for the targeted behavior to occur. The motivation of a user and ability to perform the targeted behavior define the chance that the user will actually perform the targeted behavior. The timing of the triggers is an important issue. According to the principle of Kairos a trigger such as a suggestion to do some action should be presented at the most opportune moment. Jones' coffee machine works according this principle when it suggests the user to take a break instead of having another coffee at the moment he is ordering a coffee. The principle of Kairos is one of the principles for message tailoring [54]. Fogg's

three factors also occur in the Elaboration Likelihood Model in Petty and Cacioppo [56]. Dependent on the level of attention and the motivation of the receiver a trigger can work through the central route or through peripheral cues.

Burgoon et al. [16] studied the relation between non-verbal behaviors, several dimensions of source credibility and the persuasiveness of a public speaker. They found that speaker fluency (sign of competence) and pitch variety (sign of engagement), eye contact, smiling and facial pleasantness (signs of sociability) and expressiveness as well as the use of gestures, in particular illustrators (i.e. gestures timed with speech, such as deictic movements, head nods, eye blinks as sign of sociability) had a significant positive correlation with perceived credibility of the speaker. These are the typical "behaviors" that virtual humans exhibit when they are used in the interface of persuasive systems.

8.2.2 Persuasive Technology

Fogg [26] defined persuasive technology as technology that is designed to change attitudes or behaviors of the users through persuasion and social influence, but not through coercion. According to Fogg there are different strategies that can be applied in persuasive technology:

- Reduction—Using computer technology to reduce complex behavior into simple tasks. Reduction makes it easier to perform a behavior and will increase the benefit/cost ratio of the behavior. A better benefit/cost ratio increases the motivation of a user to engage in the behavior.
- Tunneling—Using computer technology to guide users through the process of behavior change. During this process the technology will create opportunities to persuade their user.
- Tailoring—Using computer technology to present (persuasive) information tailored to the needs, interests, personality and context of the user.
- Suggestion—Using computer technology to present suggestions to the user to act in a certain way. Suggestions presented at the right moment in time will be more effective.
- Self-monitoring—Using computer technology to automate the tracking of behavior of the user. Self-monitoring helps people to achieve predetermined goals or outcomes.
- Surveillance—Using computer technology to monitor the behavior of other users in order to change your own behavior.
- Conditioning—Using computer technology for positive reinforcement to change existing behavior into habits or to shape complex behaviors.

Although Fogg's model of behavior does suggest design principles for persuasive technology, it does not explain how these principles should be transformed into software requirements and system features. Based on the principles of persuasive communication [51] provided a Persuasive System Design (PSD) framework for

designing and evaluating persuasive systems. PSD presents four categories of system features; (1) primary task support, (2) dialogue, (3) system credibility and (4) social support.

Primary task support features correspond with the above mentioned persuasive strategies of Fogg. The design principles related to the dialogue category are about the user interaction with the system. Computer-human dialogue includes praise, rewards, reminders and suggestions. The third category of system features is related to system credibility. The system must provide information that is truthful, fair and unbiased and it must incorporate **expertise** by showing knowledge, experience and competence to increase persuasiveness. The fourth and last category is related to social support. Design principles in this category are based on the idea of designing a system that motivates users by leveraging social influence. Design principles in this category include **social learning**, people will be more motivated to perform a targeted behavior when he or she can observe the behavior of others who are using the system or when they can compare their own results with other users.

8.2.3 The Computer as Persuasive Social Actor

The metaphor of the computer as social actor is supported by experimental research that shows that users treat the computer they work with in a way they treat other human subjects [58]. This works on a subconscious level. If we ask people they will say they do not consider the computer as a social agent: the computer is just a machine. But instinctively people behave as if they are interacting with a human. Social cues can be divided into five primary types:

- Physical—such as having a face, body, eyes or movement
- Psychological—such as showing personality, feelings or empathy
- Language—such as interactive language use or spoken language
- Social dynamics—such as turn taking, cooperation, praise for good work or answering questions
- Social roles—such as doctor, coach, team mate, opponent, teacher or pet

Embodied conversational agents offer many opportunities to support the presentation of a persuasive system as a social actor. The working of the interface design is largely based on social psychology. By simulating natural signs that we know from human-human social interaction the interface designer tries to elicit user experiences that make him behave in a similar way he does in the real case. HCI designers do not do that because they want to deceive the user (they had to deceive themselves then) or for artistic reasons (although aesthetics matters), but primarily because simulation works; it is effective. A very commonly known user experience occurs when we read or hear a text. When Jones' coffee machine "says" "Please show your card" Jones immediately experiences this as a request because it reminds him of similar signals in similar situations. In the following section we will review experimental

research that aims to identify those socio-psychological mechanisms in the use of virtual humans that impact the various factors that influence the efficacy of persuasive communication by means of persuasive technology.

8.3 The Persuasiveness of Virtual Humans

Since the introduction of the idea of using virtual humans or embodied conversational agents (ECA) as interface agent in the field of HCI, the question recurs "whether software agents should be personified in the interface" [20, 40]. In 1996 Koda and Maes state that "the goal of HCI work should be to understand when a personified interface is appropriate". Many researchers examined the presence and the type of embodied agents but there is little consensus as to whether or not the presence of visual agents improves a users experience with an interface, and if so by what degree [70].

Does the use of a virtual human in the interface affect the persuasiveness of personal coaching systems? Or, rather: what kinds of animated interface agent and which kind of behavioral and expressive features in what task domain and use context have a positive impact on the persuasiveness of the system? The first question is the question of *presence*: in what cases is it a good idea to have a graphical representation of the interface agent and if so how should it be presented; e.g. how realistic: a 2D or 3D caricature, photographic, by means of a talking head or full-body, animated, or as a physical robot? Then the question is how it should look like and what types of behaviors impact what factors that influence persuasiveness.

We have seen that source credibility (believability, trustworthiness) is an important factor for persuasiveness. Research in human communication has shown that eye contact rates influences credibility. To be persuasive a speaker should engage in more eye contact. But does this hold for ECAs as well? There are several reasons why we must be cautious before we jump to this conclusion. The communication channel has effect on the communication. It is not necessarily so that gaze of an ECA in the direction of the user is perceived similar as eye contact is perceived by a human persuader in a face-to-face encounter. Moreover, many variables describing behavior of the persuader jointly affect credibility and -indirectly- persuasiveness. Many studies in social psychology and HCI try to isolate the influence of one or a pair of variables in a controlled experimental setting. But this does not mean that the same effect happens in a different setting. The relation between the complex structure of variables on the one hand and the perceived credibility of the source on the other hand is non-linear and unstable so that a small change in the situation may make the difference in how the recipient perceives the source's credibility. (Compare the relation between a scribble and the word or picture that the reader recognizes in it: a small change may result in a completely different perceived word or picture.) That being said we will present in this section results from studies that provide at least indications of factors that influence the persuasiveness of a virtual human in the interface.

8.3.1 Naturalness and Credibility

Yee et al. [70] examined a number of theoretical questions related to interfaces: the effect of the presence of an embodied agent, how **realistic** the agent appears through animations and behaviors, and the type of response which was measured (subjective or performance). Do people react differently to interfaces with (1) no visual representation, (2) a human-like representation with low realism (e.g., cartoon figure), and (3) a human-like representation with high realism (e.g., 3D model animated with gestures)? Realism is not the same as anthropomorphism, which is essentially the perception of humanlike characteristics in either real or imagined nonhuman agents or objects [25]. There are many dimensions on which an interface agent can be considered real. It can behave realistically through animations, it can be highly photographically realistic via computer graphics, or alternatively it can be highly human like (i.e., anthropomorphic) [13]. Results by Guadagno et al. [29] supported the model of social influence within immersive virtual environments by Blascovich et al. [13]. Specifically, the prediction that virtual humans high in **behavioral realism** would be more influential than those low in behavioral realism was supported, but this effect was moderated by the gender of the virtual human and the research participant.

Baylor [5] points at the impact that anthropomorphic interface agents have on the motivation of learners. "Together with motivational messages and dialogue, the agentś appearance is the most important design feature as it dictates the learnerś perception of the agent as a virtual social model, in the Bandurian sense. The message delivery, through a human-like voice with appropriate and relevant emotional expressions, is also a key motivational design feature."

Effects of animated agents in the user interface with the intention to simulate a natural and believable way of interaction have been evaluated in many studies [24, 43, 49, 50, 62]. In 1997 Lester et al. [43] posed the *persona effect*. The idea behind the persona effect is that a lifelike agent in a learning environment has a strong positive effect on the user's perception of their learning experience, and conclusions from this research showed improved learning performances. These results were questioned in further research [24]. Adding a virtual human to the user interface can be a source of distraction to the user and may disturb the human computer interaction [8]. Mazzotta et al. [45] showed that feedback messages in the domain of healthy eating were better evaluated when they were presented by a virtual human compared to feedback messages presented in text. Text messages were easier to understand, but messages presented by the virtual human were perceived as more persuasive and reliable. Mazotta concludes by hypothesing that text messages are better suited for simple information given tasks, while more persuasive messages (reflecting the social and emotional intelligence of the virtual human) could be presented by a virtual human to increase the effectiveness of the persuasive strategies. Several studies by Murano et al. [50] showed seemingly inconsistent results in relation to using anthropomorphic user interfaces to present feedback in computer systems. The study shows that acceptance and effectiveness of anthropomorphic interfaces depends on the specific task. Catrambone et al. [22] also suggested that the types of task influences

the results of these type of experiments. In the long-term users may assign affective states to animated interface agents that are seen as companions [10]. Turunen et al. [63] developed a multi-modal companion system that built relationship with the user of the system to support everyday health and fitness related activities. User interface agents and support for conversational dialogues should help to build social relationships between the user and the system, and these relationships should motivate towards a healthier lifestyle.

Using a meta-analysis of research in 25 papers reporting empirical studies on the effect of presentation and realism on task performance [70] produced some consistent findings. First, the presence of a representation produced more positive social interactions than not having a representation. This effect was found in studies that used both subjective and behavioral measures. Subjective measures are taken by surveys, interviews and questionnaires. Secondly, **human-like** representations with **higher realism** produced more positive social interactions than representations with lower realism; however, this effect was only found when subjective measures were used. Behavioral measures did not reveal a significant difference between representations of low and high realism. A recent survey of 33 studies by Li [44] compares persuasiveness of physical robots with robots on a screen and with virtual agents. Physical robots score higher in persuasiveness than robots and virtual characters on the screen. There was no difference in user response between the robot on the screen and the virtual character on the screen. This indicates that physical presence more than appearance of the virtual agent affects persuasiveness. Long-term field studies in social robotics are needed to see if these effects persist over time.

Perceived expressions of emotional involvement: signs of caring, empathy and understanding has impact on sociability and thus on credibility. Does synthesized expression of emotion either verbally or non-verbally impact persuasiveness? In a literature review [6] focuses on the question "which kind of emotional expressions in which way influences which elements of the user's perceptions and behaviours?" (p. 757). Berry et al. [8] empirically evaluated an ECA named Greta to investigate the effects of emotion expression in a system that promotes healthy eating habits. The messages from the system could be presented by Greta, by a human actor, only by the voice of Greta or via a text message. Except for the human actor, presenting the message by an ECA received the highest ratings for helpfulness and likability. They found that persuasive health messages appeared to be more effective (measured by subjective measures) when presented with an emotionally neutral expression compared to when messages were presented with an emotion consistent with the content of the message.

8.3.2 Virtual Humans and Behavior Change Support Systems

Results from studies by Henkemans et al. [32], Schulman and Bickmore [60], Berry et al. [8] indicate that the use of virtual characters can have a positive effect on the likability, helpfulness, ease of use and motivation to use computer systems.

Bickmore conducted several experiments with the MIT FitTrack system [10, 11, 60]. By using the FitTrack system users can enter their daily steps counted by a pedometer and estimated time of physical activity. Schulman and Bickmore [60] studied the effects of different versions of computer agents on users' attitudes towards regular exercises. This experiment compared two situations, an ECA versus a text agent and whether the agent attempted to build a user-agent relationship through social dialogue in a wizard-of-oz set-up. Participants following the persuasive dialogue showed a significant increase in positive attitude towards the exercises. This change was significantly smaller when the agent used social dialogue. Users' perceptions of the dialogue were most positive for an ECA with social dialogue or a text-only agent without social dialogue.

In a long-term (2 and 12 months) study Bickmore et al. [12] tested the efficacy of an animated virtual exercise coach designed for sedentary older adults with low health literacy. Positive effects of the ECA enhanced system compared to a system that uses only a pedometer measured after 2 month waned after 12 month. This might be caused by the fact that in the ten months period after the first two months interaction was not through a tablet anymore but through a screen in a waiting room kiosk of a clinic.

The MOPET system of Buttussi and Chittaro [17] is a *mobile* personal trainer system. MOPET is designed to support the user throughout exercise sessions, by guiding the user through fitness trails that alternate running with physical exercises. It tracks the user's position on the trail and shows the user's speed, and also tries to motivate the user through messages. It uses an external sensor device that collects heart rate and accelerometer data. When the user comes to an exercise point along the route, the system recognizes this and demonstrates the exercise to the user. The virtual human is presented as a full-bodied animated 3D character that is rendered in real-time. While this system can make exercise more effective and more enjoyable for users, it does not actually motivate users to start exercising. The MOPET system was evaluated with 12 participants. During the user evaluation the MOPET system was compared with guiding users through fitness trails by written instructions in text. Results showed that the MOPET system was rated as more useful compared to written instructions and participants made fewer mistakes using the MOPET system [18]. One of the functions of the full body animated ECA in MOPET is to demonstrate the physical exercises, a function that should be clearly distinguished from its conversational functions.

There is a complexity of mediating factors that influences the user's experience of systems that use virtual characters to personify software agents. Experiments by Koda and Maes [40] already demonstrated that the perceived intelligence of a face is determined not by the agent's appearance but by its competence. There is a dichotomy between user groups which have opposite opinions about personification. Not all people are evenly inclined to anthropomorphize. Epley et al. [25] studies personal and situational factors that determine the need for socializing and the need to understand and control, the two factors that influence if someone anthropomorphizes. "Thus, agent-based interfaces should be flexible to support the diversity of users' preferences and the nature of tasks." [40]. A positive impact of features of an ECA

on the perception of the user and on perceived credibility (by subjective measures) does not always come with a positive impact on long-term behavior change as the work of Bickmore and Picard [10] and Bickmore et al. [12] has shown.

In the next section we present the results of our own comparative field trials that contribute to more insight into the effects of the use of virtual humans as interface agents for personal coaching systems.

8.4 Kristina—Virtual Coach for Physical Activity

Office workers and many other adults in Western countries live a sedentary lifestyle. They sit too much and they don't reach the thirty minutes daily moderate intensive physical activity recommended by the World Health Organisation. Awareness of the risks involved and motivation to change one's lifestyle are the main conditions for a positive change. Wearable unobtrusive sensor technology has become cheaper and smaller over the last years. They can help people to at least monitor their physical activity level. Changing one's lifestyle is however easier said than done. A personal digital coach may help the user.

Kristina is a personal digital coaching system built to support and motivate users to live a balanced and healthy lifestyle. The system can target better adherence to medication intake and physical activity. To measure the amount of physical activity the system makes use of an activity monitor. To keep track of medication intake the system uses a digital pillbox. The coaching rules and the sensor data are stored on a central server that generates the feedback messages and send them to the device in use. Each output device has a set of available presentation formats and it is up to the user to select the format that he wants to use. On a smartphone a text message can be presented by simple text, it can be spoken or presented in a multi-modal way by an animated virtual human.

Kristina will motivate and support the users by providing information on the current behavior of the user (self-monitoring), by presenting feedback about the behavior of the users, such as the progress towards their personal goal. Hints how to reach these goal(s) will be presented to the users when appropriate. The messages and the devices that can be used to present feedback messages are based on user studies presented in Klaassen et al. [37]. Users have a preference for receiving feedback on the smartphone, computer, or television. Examples of situations in which users would mainly like to receive feedback in, is while relaxing in front of the television or while having a (lunch) break. As for the type of messages that can be sent, users seem to have a preference for reports about their progress, rather than advice or learning messages.

A coaching system can be real-time or off-line. In a real-time coaching system the sensor data is continuously available from the user for the coach who can decide to act anytime by sending feedback to the user. In some of these systems the data is sent through a Bluetooth connection to a mobile platform. Nowadays smartphones have built in activity sensors. op den Akker et al. [53] provides an overview of real-time

Fig. 8.2 Overview of physical activity presented on PC and smartphone

physical activity coaching systems. Kristina is not a real-time coaching system. The user has to connect the activity sensor physically to a PC in order to upload the data to the server. The user can only get feedback and a data overview on the data that was uploaded. This implies that in order for the coach to give timely feedback the user has to upload his data regularly, that is at least once every day. This is a disadvantage compared to real-time systems. In a real-time system the coach can send a motivational message saying "*It is time for a walk. You haven't been very active up to now today*." Actually one of the tasks of Kristina is to remind the user to regularly upload his sensor data. Kristina uses strategies such as "tailoring", "suggestions" and "self-monitoring" and it targets users that are in the "action" state of the Transtheoretical Model (see Sect. 8.2.1 for more details about the Transtheoretical Model), i.e., they are motivated to change their lifestyle. Figure 8.2 gives an example of how data is presented to the user for self-monitoring.

8.4.1 Articulated Social Agent Platform

To take the full advantage of the power of embodied conversational agents in service systems it is important to support real-time, online and responsive interaction with the system through the embodied conversational agent. The Articulated Social Agent Platform (ASAP) is a model-based platform for the specification and animation of synchronised multi-modal responsive animated agents developed at the Universities of Twente and Bielefeld.[3] This is a realizer for the interpretation of embodied agent behavior specified in the Behavior Markup Language (BML) [67]. The back-end animator can be a full 3D graphical avatar and is running on desktop PC or a physical robot. Klaassen et al. [38] presents a light-weight PictureEngine as back-end engine which allows to run the platform in mobile Android applications. This makes it possible to generate real-time multi-modal behavior of an animated cartoon-like virtual human showing gestures and expressions, such as smiles and eye blink, synchronized

[3] http://asap-project.ewi.utwente.nl/ [64].

Fig. 8.3 Examples of different expressions of the virtual human used in the Kristina system. In this case dressed like a doctor

with synthetic speech. The ASAP platform is used in the Kristina personal coaching system. The user can choose how he will receive the messages of the coaching system through speech by an animated graphical character or by a text displayed on the screen. Figure 8.3 gives examples of the animated character. Here, Kristina is represented as a medical doctor. For speech we used the SVOX text-to-speech engine in English and Dutch for Android.

In the next section we will describe a user evaluation with the the mobile Kristina coaching in which we compare two versions one with and one without an animated graphical interface.

8.5 Coaching by Means of Text Messages or by a Virtual Human

In a long-term field study we investigated the possible effects of presenting Kristina on a smart phone by means of an animated virtual character. We measured effects on usability, user experience, likability, the quality of coaching, anthropomorphism, the credibility of the Kristina system and the usage of the system as well as the performance in terms of the user's physical activity. The three user groups that participated in this study are listed below. The evaluation was held in the period of January to August 2013.

- Text group—A group that used the coaching application on their smartphone and computer and received feedback messages in text.
- Virtual Human group—A group that used the coaching application on their smartphone and computer and received feedback presented by a virtual human.
- Control group—A group that used the coaching application on their computer and received feedback in text on the website, not on their smartphone.

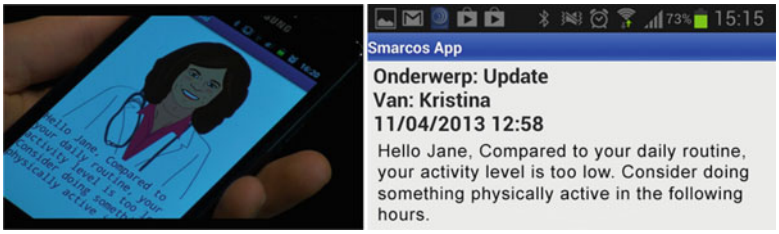

Fig. 8.4 A feedback message presented in text and presented by a virtual human on a smartphone

8.5.1 Methodology

We used a between subjects design to evaluate the coaching system. The independent variables are the way in which feedback messages were presented to the users. Feedback messages could be presented in text or via a virtual human (see Fig. 8.4). Participants of the control group did not receive feedback messages actively and they did not receive notifications about new feedback messages. The dependent variables of this evaluation were the physical activity levels (PAL), the person's total energy expenditure (TEE) in a 24-hour period, divided by his or her basal metabolic rate [69], the number of uploads of physical activity data and the user perception. The dependent variables were measured by log data, questionnaires, and interviews.

Participants were asked to complete three questionnaires. The intake questionnaire was meant to collect background information like age, weight, height, gender and experiences with smartphones/PDAs, virtual humans and digital coaching systems. The second and third questionnaires (halfway through and at the end of the evaluation) were meant to evaluate the participants' experiences with the different feedback versions. We used questionnaires to measure the usability (System Usability Scale [15]), user experience (AttrakDiff2 [30]), quality of coaching of the coaching system (an adapted version of the Coaching Behaviour Scale for Sport (CBS-S) [23]), credibility (Source Credibility Scale [46]) and acceptance (UTAUT [65] and Godspeed [4]).

At the end of the user evaluation participants were invited to discuss their experiences in a semi-structured interview.

8.5.1.1 Subjects

Participants were asked to join the user evaluation by email, social media and face-to-face communication. Participants had to be office workers (sedentary profession) and should own their own Android smartphone with Android version 2.3 or higher. Sixty office workers indicated they were willing to join the experiment and were divided into the three groups. The distribution over the three groups was random. Participants in the virtual human group had to install the Dutch text-to-speech engine from the Google Play store.

8.5.1.2 Procedure

The duration of the complete evaluation was seven weeks. This included one assessment week at the start of the user evaluation. Before the start of the evaluation participants received the activity sensor and an information sheet with details about the user evaluation including an informed consent to be signed. Software, applications for the smartphone, user manuals about the installation of software and links to questionnaires were provided via email.

The first week was an assessment week in order to establish normal activity level for tailoring the system and to set a personal goal. During this assessment week no feedback messages were given by the system. After the assessment week participants received their personal goals and used the system for six weeks. In these six weeks feedback (updates, requests, reminders and overviews) about their progress was provided. Participants were asked to upload their activity data at least one time per day by connecting their activity monitor to their computer. Participants who over or under performed in the first three weeks were offered a new (higher or lower) goal by the system. Participants were free to accept this new goal or not.

Questionnaires were available online and participants received an email when it was time to fill in a questionnaire. Shortly after the evaluation the participant was visited again to collect the materials. During this visit participants were invited for a short final interview to discuss experiences.

8.5.2 Results

Forty-three participants completed the user evaluation by finishing the assessment week, using the system for at least six weeks and completing all the questionnaires. Participants were between 21 and 57 years old (average $= 36.5 \pm 10.4$), and worked 36.1 hours per week on average. All participants except two owned an Android smartphone. Those two participants were included in the control group. All the participants were familiar with mobile Internet. Most of the participants did sports, forty indicated doing sports every week.

Seventeen participants did not finish the user evaluation due to several personal reasons. Two participants (from the text group) indicated having difficulties in answering some of the questions from the questionnaires. They were excluded from the results of the questionnaires. Twenty-one participants participated in the interview at the end of the user evaluation study.

8.5.2.1 Usability

The System Usability Scale from Brooke [15] was used to measure the usability of the system. The system usability score can range from 0 to 100. We analysed the results of the questionnaires to investigate differences between the text group and the virtual

human group by using an independent samples t-test. To investigate differences between the halfway and end questionnaires we used a paired samples t-test. The system usability score of the Kristina system was above 62.2, which can be seen as acceptable. The System Usability Scale was reliable (Cronbach's alpha = 0.716). From the results of the halfway questionnaire we found one significant difference on the statement "*I thought there was too much inconsistency in this system*"between the text group and the virtual human group (Group Text: $M = 3.65$, $SD = 0.70$, Group VH: $M = 2.93$, $SD = 1.16$, $t(30) = -2.131$, $p = 0.041$). From these results we can conclude that the virtual character of the Kristina coaching system made the system more inconsistent compared to the Kristina coaching system that was using text messages.

We asked the participants about their general impression of the coaching system. Four people mentioned explicitly that they liked obtaining awareness about the amount of physical activity. Participants tried to reach their daily goal and mentioned that they used the system in a different way over time. In the beginning they were more enthusiastic about the system compared to the end of the user evaluation.

We asked the participants about their experiences related to the feedback and overviews presented by the system. Remarks from the participants were about the timing and the content of the feedback messages. Feedback messages were seen as a kind of standard. Participants wanted to receive more detailed feedback. Participants stated that the system should become more intelligent or more exciting. Some suggest to use history data in feedback messages. The timing and content of the feedback messages were seen as predictable. Two participants mentioned that receiving feedback and/or tips was not always useful, because users do not always have the ability to follow the tips or reminders from the system. Three participants stated that feedback about physical activity was nice to have.

Participants were asked to give their opinion about the question whether such a system is useful to live a healthier life. From the text group eight participants answered this question with "Yes" and two participants answered with "No". Participants who answered "Yes" stated that the system created awareness, but feedback messages needed to be further developed to become more concrete. Participants who answered "No" stated that the system was only useful to live a healthier life for people who were already motivated to use such a system or the system should be advised by a doctor or health provider. Eight participants from the virtual human group also answered this question with "Yes" and two participants answered with "No". Participants who answered "Yes" stated that the system created awareness about physical activity, but it was important that the user was motivated to use the system and change their behavior. Participants who answered "No" stated that the timing and content of the feedback should be improved and the fact that the system will only help to lead a healthier lifestyle in the short term.

Table 8.1 Differences between the text and virtual human group on word pairs of the Attrakdiff questionnaire halfway through and at the end of the user evaluation on a 5 point Likert scale

Word pair	Text ($M(SD)$)	VH ($M(SD)$)	Independent samples test
Halfway			
"Unprofessional–professional"	4.94(1.09)	3.93(1.22)	$t(30) = 2.468$, $p = 0.020$
"Cautious–bold"	4.13(0.49)	3.67(0.90)	$t(30) = 1.795$, $p = 0.083$
End			
"Unruly–manageable"	5.36(1.17)	4.40(1.35)	$t(30) = -2.138$, $p = 0.041$)

8.5.2.2 User Experience

User experience of the coaching system was measured using the AttrakDiff2 questionnaires. We analysed the results of the AttrakDiff2 questionnaires to investigate differences between the text group and the virtual human group by using independent samples t-test. To investigate differences between the halfway and end questionnaires we used a paired samples t-test. The reliability of all four dimensions was good (Cronbach's alpha > 0.6). Table 8.1 presents the significant results of the AttrakDiff2 questionnaires. From these results we can conclude that feedback messages presented by a virtual character were seen as less professional and less manageable compared to the same feedback messages presented in text.

8.5.2.3 Likeability

The acceptance of the coaching application was measured using a questionnaire based on the UTAUT questionnaire. We analysed the results of the questionnaires to investigate differences between the text group and the virtual human group by using an independent samples t-test. All constructs were reliable (Cronbach's alpha > 0.6) except for the anxiety (Anx) construct. From the results of the halfway questionnaire we found one marginally significant difference between the text group and the virtual human group on the statement "*The system would make my life more interesting*" (Group Text: $M = 1.71$, $SD = 1.10$, Group VH: $M = 2.47$, $SD = 0.52$, $t(30) = -1.927$, $p = 0.063$). This difference could not be found at the end of the user evaluation.

8.5.2.4 Anthropomorphism

The Godspeed questionnaire of Bartneck et al. [3] was used to measure the acceptance of the virtual coach. We analysed the results of the questionnaires to investigate

Table 8.2 Differences between the text and virtual human group on word pairs of the Godspeed questionnaire halfway through the user evaluation on a 5 point Likert scale

Word pair	Text ($M(SD)$)	VH ($M(SD)$)	Independent samples test
"Fake–natural"	2.34(0.86)	1.87(0.64)	$t(30) = 1.791$, $p = 0.083$
"Dead–alive"	2.71(0.69)	2.20(0.94)	$t(30) = 1.752$, $p = 0.090$
"Anxious–relaxed"	3.12(0.33)	3.47(0.74)	$t(30) = -1.751$, $p = 0.090$
"Still–surprised"	2.53(0.72)	2.07(0.80)	$t(30) = 1.727$, $p = 0.094$

Table 8.3 Differences between the text and virtual human group on word pairs of the Godspeed questionnaire at the end of the user evaluation on a 5 point Likert scale

Word pair	Text ($M(SD)$)	VH ($M(SD)$)	Independent samples test
"Unconscious–conscious"	2.71(0.69)	1.87(0.64)	$t(30) = 3.563$, $p = 0.001$
"Artificial–lifelike"	2.59(0.71)	1.80(0.68)	$t(30) = 3.199$, $p = 0.003$
"Dead–alive"	3.00(0.71)	2.33(0.82)	$t(30) = 2.476$, $p = 0.019$
"Unintelligent–intelligent"	3.012(0.49)	2.40(1.12)	$t(30) = 2.401$, $p = 0.023$
"Fake–natural"	2.65(0.79)	2.13(0.83)	$t(30) = 1.793$, $p = 0.083$
"Machinelike–humanlike"	2.59(0.94)	2.00(0.93)	$t(30) = 1.789$, $p = 0.085$
"Still–surprised"	2.59(0.62)	2.13(0.74)	$t(30) = 1.890$, $p = 0.068$

differences between the text group and the virtual human group by using an independent samples t-test. All constructs were reliable (Cronbach's alpha > 0.6) except for the perceived safety construct. Differences between the text group and the virtual human group halfway the user evaluation can be found in Table 8.2. Differences between these groups at the end of the user evaluation can be found in Table 8.3.

The coaching system that was using the virtual character to present feedback messages was seen as more artificial, fake, dead, and machine-like. The system using the virtual character was also seen as less intelligent compared to the system that was using text messages.

Participants were asked for their opinion on the two presentation forms on the smartphone. Before that we gave a short explanation and demonstration of both the virtual human coach and the text coach on the smartphone of the researcher. Three

participants from the text group stated that a virtual coach represented by text was fine for them because they were afraid that a virtual human would become annoying or they did not like the idea of having a virtual human as an user interface telling them what to do. Four participants from the text group stated that it would not make any difference in this coaching system. Participants stated that the system would stay mechanical and predictable because it would not change the behavior of the complete system. Some stated that the motivation to use such a system was more important than the way that a coach is represented. Three participants from the text group had the impression that a coach represented by a virtual human was more fun to use, more human-like or closer to the way a normal coach would provide feedback. Eight of the participants from the virtual human group stated that they preferred feedback messages in text. They had the impression that the virtual human did not add anything extra compared to a text message for the kind of feedback messages presented in this coaching system. Five of these participants stated that the virtual human can be useful when more detailed feedback will be presented, when it is possible to have interaction with the coach or when users need extra motivation. Two participants preferred feedback presented by the virtual human. In their opinion the virtual human added some extra dimension to the coaching system, it was fun to get feedback presented by a virtual human and it added some social pressure to the feedback message.

8.5.2.5 Quality of Coaching

We used an adapted version of the CBS-S to measure the quality of coaching by the system in the different conditions. We analysed the results of the questionnaires to investigate differences between the text group and the virtual human group by using independent samples t-test. The reliability is good (Cronbach's alpha > 0.958). There was no difference between virtual human and text group in the perceived quality of coaching.

8.5.2.6 Credibility

The credibility of the feedback messages and the coaching system was measured using the Source Credibility Scale of McCroskey and Young [48]. To investigate differences between the halfway and end questionnaires we used a paired samples t-test. The reliability was good (Cronbach's alpha > 0.88). Results of the Source Credibility Scale can be found in Table 8.4.

From the results we can conclude that the Kristina coaching system that was using the virtual character was seen as less intelligent compared to the system that was using text messages. Feedback messages presented by the virtual character were seen as more honest.

Table 8.4 Differences between the text and virtual human group on the Source Credibility Scale halfway through and at the end of the user evaluation on a 5 point Likert scale

Word pair	Text ($M(SD)$)	VH ($M(SD)$)	Independent samples test
Halfway			
"Unintelligent–intelligent"	4.35(1.11)	3.60(1.12)	$t(30) = -1.902$, $p = 0.084$
End			
"Dishonest–honest"	4.53(1.23)	5.33(1.11)	$t(30) = 1.928$, $p = 0.063$
"Unselfish–selfish"	4.47(0.87)	5.07(1.10)	$t(30) = 1.702$, $p = 0.098$
"Sinful–virtuous"	4.06(0.43)	4.60(0.99)	$t(30) = 2.057$, $p = 0.048$

8.5.3 Log Data

During the user evaluation data about the physical activity levels (PAL), the number of uploads of activity data and the number of feedback messages sent by the coach were collected.

Activity data

Figure 8.5 presents the physical activity level of all the participants per group. We analysed the physical activity levels of each participant and used a Mixed Models

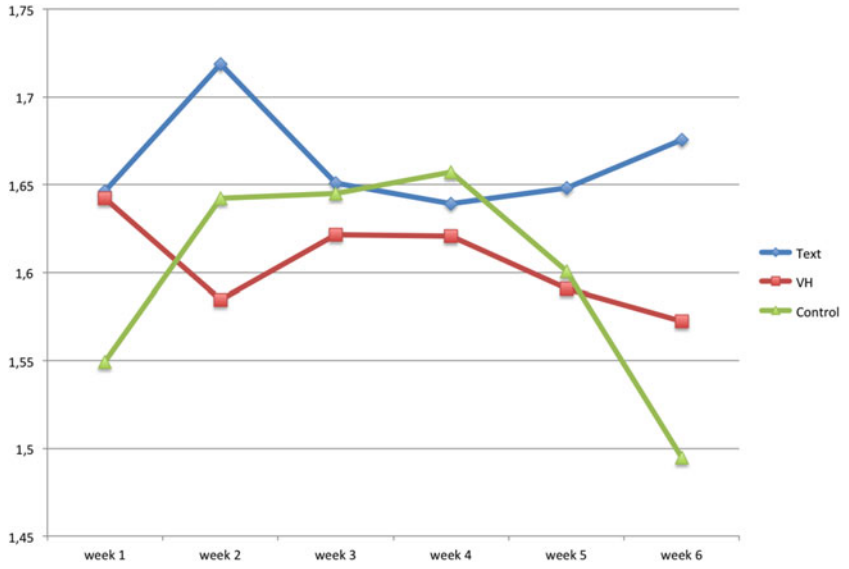

Fig. 8.5 Physical activity levels per week. The lines represent the average PAL per group

approach to investigate whether time (in weeks) or group (condition) was a predictor for the level of physical activity. From this test we can conclude that there was no relationship between time or condition and the physical activity level. We used a Mann-Whitney test to compare the average values of all participants of the groups to investigate whether there is a difference between the groups. At the end of week 6 the PAL level of the control group was significantly less than the text group ($Mdn = 1.61$, $U = 36.00$, $r = 0.460$, $p = 0.014$).

Number of uploads and feedback messages

Participants were asked to upload their physical activity data on a regular basis, preferable once a day. During the six weeks (42 days) user evaluation participants uploaded their activity data 32 times on average. We used a Mann-Whitney test to compare the results between the different groups. The number of uploads of physical activity data in the control group was significantly less than in the Text group during all six weeks ($Mdn = 3.00$, $U = 51.00$, $p = 0.045$, $r = 0.378$). No difference was found between the text group and virtual human group. Over the whole period on average 74 feedback messages were sent per participant.

8.5.4 Conclusion

We can conclude that presenting feedback messages in text was a better approach compared to presenting feedback messages by a virtual human. In general the feedback messages were seen as a kind of standard and participants stated that there was too much repetition. Behavioral measures (log data from sensor) showed no differences between the text group and the virtual human group, but the physical activity level of the Text group was increasing in week 6 of the user evaluation, while the physical activity levels of the virtual human and control group were decreasing. We found a significant difference between the Text group and the Control group at the end of the user evaluation when we look at the physical activity level and the number of uploads of data. Regular feedback and reminders help but use of a virtual human does not have a positive effect compared to a simple text message.

Subjective measures obtained by the questionnaires were in favor of the text coach. This is supported by the interviews. From the interviews we have indications that the virtual human coach could improve users' perception if the feedback messages become more intelligent, if interaction with the virtual coach was possible or when users are motivated more to use a physical activity coaching system.

Perceived quality of a system is a subjective notion. It is influenced by desires as well as by the expectations users have [68]. Participants from the two groups that received feedback messages had different reference points to which they compared the messages from the coaching system. Using a virtual human in a coaching application that makes use of verbal and non-verbal communication creates high expectations by the users. Using speech output may create the expectation that the virtual human is able to recognise verbal input, a feature not implemented in Kristina.

The context of use influences the acceptance of the verbal animated presentation of feedback messages and hints by the virtual coach. Short attention spans are typical in mobile use [34]. In the mobile context, many other tasks and other applications on smartphones ask for the users' attention. Text messages are easier and quicker to read and understand compared to a message presented by a virtual human that uses text to speech thereby controlling the timing and duration of the presentation of the message. Glanceablility of the interface affects the perceived ease of use and usefulness of the system and thereby the efficacy of the coaching [14].

8.6 Guidelines for Use of Virtual Coaches

We have seen that a combination of properties of the user (motivation, needs and expectations), the (mobile) use context, the task of the virtual coach (to give short hints and feedback or to motivate) and the appearance of a virtual coach determine the way users experience and react to personal coaching systems.

Based on lessons learned from our experiences with the Kristina system and the literature reviews we can now present some guidelines and considerations for deciding when to use an embodied conversational agent in the interface of a personal coaching system.

1. Utilitarian functions come first, hedonic aspects are important but come second place.
2. Context of use is of prime importance and impacts attention spans and communication efficiency.
3. Dependent on the task of the system and the target user group it is sometimes better not to use a virtual human that represents an authority like a personal trainer. An animated emphatic character should be considered.
4. The functions of the (animated) graphical interface of a persuasive system should be reflected in a coherent way in appearance, expressions and behaviors and be related to the utilitarian functions of the system.
5. Do not use verbose text messages when simple signals can transmit the same message.
6. The user must have control over the system and the way user data is handled.
7. Given the state of the art of spoken dialog systems use of a spoken dialog may have negative impact on the user acceptance if not handled with care.
8. Do not let the system give information or suggest or advice actions based on uncertain information.
9. Let the system be clear about the intention and the cognitive status of information it provides to the user.
10. Make your persuasive actor so that it is aware of the fact that some users don't like to be bothered by calls for attention they didn't ask for.
11. Be aware of the fact that as designer you are responsible for making the system so that it allows the user to act in a responsible way.

We believe there is potential in the use of virtual humans for personal coaching systems when used in the appropriate way. The quality of automatic speech recognition and speech generation systems has improved so that interaction between the user and the system by means of a natural conversation comes within reach. Speech interfaces have arrived in every day use. In 2011 Apple introduced their speech enabled personal assistant Siri. Android mobile users can now choose from a wealth of similar personal assistants: Google Now, Voice Actions, Indigo, and many more. These commercial speech interfaces do not allow real dialogues. The interaction consists of isolated question response pairs. Spoken dialogue systems is an active research field adding to the development of embodied conversation agents that allow a conversational style of human computer interaction [42]. Progress in the quality of speech recognition and the use of more sophisticated spoken dialog systems designed for this type of application domains will improve the quality of the interaction and create new potentials for interaction between users and virtual coaches through the use of embodied conversational agents. Dialogs can either be used to collect information from the user that is hard or not to obtain through sensors, for motivational interviewing, or for persuasive argumentation [1, 28].

Future will show what role personal coaching systems can play in cooperation with patients and medical experts in clinical settings and how we can best exploit the strong points of computer systems on the one hand and those of the human on the other hand. Care for our selves, our environment and others cannot be transmitted to technical systems. Despite the fact that we sometimes perceive them as caring social actors.

Acknowledgments Thanks to all Smarcovians who contributed in one way or another to the work reported here. This work was funded by the European Commission, within the framework of the ARTEMIS JU SP8 SMARCOS project 100249

References

1. Andrews P, Manandhar S, Boni MD (2008) Argumentative human computer dialogue for automated persuasion. In: Proceedings of 9th SIGdial workshop on discourse and dialogue, association for computational linguistics, pp 138–147
2. Bandura A (1977) Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 84(2):191–215
3. Bartneck C, Croft E, Kulic D (2008) Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots. In: Metrics for HRI workshop, technical report, vol 471. Citeseer pp 37–44
4. Bartneck C, Kuli D, Croft E, Zoghbi S (2009) Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. Int J Soc Robot 1:71–81
5. Baylor AL (2011) The design of motivational agents and avatars. Educ Technol Res Dev 59(2):291–300
6. Beale R, Creed C (2009) Affective interaction: how emotional agents affect users. Int J Hum-Comput Stud 67(9):755–776

7. Berdichevsky D, Neuenschwander E (1999) Toward an ethics of persuasive technology. Commun ACM 42(5):51–58
8. Berry DC, Butler LT, de Rosis F (2005) Evaluating a realistic agent in an advice-giving task. Int J Hum-Comput Stud 63(3):304–327
9. Bettinghaus EP, Cody MJ (1994) Persuasive communication, 5th edn. Wadsworth, Ted Buchholz
10. Bickmore T, Picard R (2005) Establishing and maintaining long-term human computer relationships. ACM Trans Comput-Hum Interact 12:293–327
11. Bickmore T, Mauer D, Crespo F, Brown T (2007) Persuasion, task interruption and health regimen adherence. Proceedings of the 2nd international conference on persuasive technology (PERSUASIVE'07). Springer, Berlin, pp 1–11
12. Bickmore TW, Silliman RA, Nelson K, Cheng DM, Winter M, Henault L, Paasche-Orlow MK (2013) A randomized controlled trial of an automated exercise coach for older adults. J Am Geriatr Soc 61(10):1676–1683. doi:10.1111/jgs.12449
13. Blascovich J, Loomis J, Beall AC, Swinth KR, Hoyt CL, Bailenson JN (2002) Immersive virtual environment technology as a methodological tool for social psychology. Psychol Inq 13(2):103124
14. Boerema S, Klaassen R, op den Akker HJA, Hermens HJ (2012) Glanceability evaluation of a physical activity feedback system for office workers. In: Proceedings of EHST 2012: the 6th International symposium on ehealth services and technologies, SciTePress–science and technology publications, pp 52–57
15. Brooke J (1996) Sus-a quick and dirty usability scale. Usability Eval Ind 189:194
16. Burgoon JK, Birk T, Pfau M (1990) Nonverbal behaviors, persuasion, and credibility. Hum Commun Res 17(1):140–169
17. Buttussi F, Chittaro L (2008) Mopet: a context-aware and user-adaptive wearable system for fitness training. Artif Intell Med 42(2):153–163
18. Buttussi F, Chittaro L, Nadalutti D (2006) Bringing mobile guides and fitness activities together: a solution based on an embodied virtual trainer. Proceedings of the 8th conference on human-computer interaction with mobile devices and services (MobileHCI '06). ACM, New York, pp 29–36
19. Cassell J (2001) Embodied conversational agents: representation and intelligence in user interfaces. AI Mag 22(4):67–83
20. Cassell J, Pelachaud C, Badler N, Steedman M, Achorn B, Becket T, Douville B, Prevost S, Stone M (1994) Animated conversation: rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. Proceedings of the 21st annual conference on Computer graphics and interactive techniques (SIGGRAPH '94). ACM, New York, pp 413–420
21. Cassell J, Sullivan J, Prevost S, Churchill EF (eds) (2000) Embodied conversational agents. MIT Press, Cambridge
22. Catrambone R, Stasko J, Xiao J (2002) Anthropomorphic agents as a user interface paradigm: Experimental findings and a framework for research. In: Proceedings of the 24th annual conference of the cognitive science society, pp 166–171
23. Cote J, Yardley J, Hay J, Sedgwick W, Baker J (1999) An exploratory examination of the coaching behaviour scale for sport. AVANTE 5:82–92
24. Dehn D, van Mulken S (2000) The impact of animated interface agents: a review of empirical research. Int J Hum-Comput Stud 52(1):1–22
25. Epley N, Waytz A, Akalis S, Cacioppo J (2008) When we need a human: motivational determinants of anthropomorphism. Soc Cognit 26(2):143–155
26. Fogg B (2002) Persuasive technology: using computers to change what we think and do. Morgan Kaufmann Publishers, San Francisco
27. Fogg B (2009) A behavior model for persuasive design. In: Proceedings of the 4th international conference on persuasive technology (Persuasive '09). ACM, New York, pp 40:1–40:7
28. Grasso F (2003) Rhetorical coding of health promotion dialogues. In: Dojat M, Keravnou E, Barahona P (eds) Artificial intelligence in medicine, vol 2780. Lecture notes in computer science. Springer, Berlin, pp 179–188

29. Guadagno RE, Blascovich J, Bailenson JN, McCall C (2007) Virtual humans and persuasion: the effects of agency and behavioral realism. Media Psychol 10:122
30. Hassenzahl M (2008) User experience (UX): towards an experiential perspective on product quality. Proceedings of the 20th international conference of the association francophone d'Interaction homme-machine (IHM '08). ACM, New York, pp 11–15
31. Heckman CE, Wobbrock JO (2000) Put your best face forward: anthropomorphic agents, e-commerce consumers, and the law. Proceedings of the fourth international conference on autonomous agents (AGENTS '00). ACM, New York, pp 435–442
32. Henkemans OAB, van der Boog P, Lindenberg J, van der Mast C, Neerincx M, Zwetsloot-Schonk BJHM (2009) An online lifestyle diary with a persuasive computer assistant providing feedback on self-management. Technol Health Care 17:253–257
33. Hermens H, op den Akker H, Tabak M, Wijsman J, Vollenbroek M (2014) Personalized coaching systems to support healthy behavior in people with chronic conditions. J Electromyogr Kinesiol 24(6):815–826
34. Kaasinen E, Roto V, Roloff K, Väänänen-Vainio-Mattila K, Vainio T, Maehr W, Joshi D, Shrestha S (2009) User experience of mobile internet analysis and recommendations. J Mob HCI Spec Issue Mob Internet User exp 1(4):4–23
35. Kay A (1990) User interface: a personal view. In: Laurel B (ed) The art of human-computer interface design. Addison-Wesley, Reading, pp 191–208
36. Klaassen R, op den Akker R, op den Akker H (2013) Feedback presentation for mobile personalised digital physical activity coaching platforms. In: Proceedings of the 6th international conference on pervasive technologies related to assistive environments (PETRA 2013). ACM, New York, p 8
37. Klaassen R, op den Akker R, Lavrysen T, Wissen S (2013) User preferences for multi-device context-aware feedback in a digital coaching system. J Multimodal User Interfaces 21:1–21
38. Klaassen R, Hendrix J, Reidsma D, op den Akker R, van Dijk B, op de Akker H (2013) Elckerlyc goes mobile–enabling natural interaction in mobile user interfaces. Int J Adv Telecommun 6(1–2):45–56
39. Klasnja P, Consolvo S, Pratt W (2011) How to evaluate technologies for health behavior change in HCI research. Proceedings of the SIGCHI conference on human factors in computing systems (CHI '11). ACM, New York, pp 3063–3072
40. Koda T, Maes P (1996) Agents with faces: the effect of personification. In: Proceedings of 5th IEEE international workshop on robot and human communication, pp 189–194
41. Kozierok R, Maes P (1993) A learning interface. Proc AAAI 1993:459–465
42. Lemon O (2012) Conversational interfaces. In: Lemon O, Pietquin O (eds) Data-driven methods for adaptive spoken dialogue systems. Springer, Berlin, pp 1–4
43. Lester JC, Converse SA, Kahler SE, Barlow ST, Stone BA, Bhogal RS (1997) The persona effect: affective impact of animated pedagogical agents. Proceedings of the SIGCHI conference on human factors in computing systems (CHI '97). ACM, New York, pp 359–366
44. Li J (2015) The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. Int J Hum-Comput Stud 77:23–37
45. Mazzotta I, Novielli N, De Carolis B (2009) Are ECAs more persuasive than textual messages? In: Ruttkay Z, Kipp M, Nijholt A, Vilhjálmsson HH (eds) Intelligent virtual agents, vol 5773. Lecture notes in computer science. Springer, Berlin, pp 527–528
46. McCroskey J (1966) Scales for the measurement of ethos. Speech Monogr 33(1):65–72
47. McCroskey J (1967) The effects of evidence in persuasive communication. West Speech 31:189–199
48. McCroskey JC, Young TJ (1981) Ethos and credibility: the construct and its measurement after three decades. Commun Stud 32(1):24–34
49. Murano P (2006) Why anthropomorphic user interface feedback can be effective and preferred by users. In: Chen CS, Filipe J, Seruca I, Cordeiro J (eds) Enterp Inf Syst VII. Springer, Netherlands, pp 241–248
50. Murano P, Gee A, Holt PO (2011) Evaluation of an anthropomorphic user interface in a travel reservation context and affordances. J Comput 3:8

51. Oinas-Kukkonen H, Harjumaa M (2008) A systematic framework for designing and evaluating persuasive systems. Proceedings of the 3rd international conference on persuasive technology (PERSUASIVE '08). Springer, Berlin, pp 164–176
52. op den Akker R, Klaassen R, Lavrysen T, Geleijnse G, van Halteren A, Schwietert H, van der Hout M (2011) A personal context-aware multi-device coaching service that supports a healthy lifestyle. Proceedings of the 25th BCS conference on human-computer interaction (BCS-HCI '11). British Computer Society, Swinton, pp 443–448
53. op den Akker H, Jones VM, Hermens HJ (2014) Tailoring real-time physical activity coaching systems: a literature survey and model. User Model User-Adapt Interact 24:351–392
54. op den Akker H, Cabrita M, op den Akker R, Jones VM, Hermens HJ (2015) Tailored motivational message generation: a model and practical framework for real-time physical activity coaching. J Biomed Inform 55:104–115
55. Petty R, Cacioppo J (1981) Attitudes and persuasion: classic and contemporary approaches. William C, Brown, Dubugue
56. Petty R, Cacioppo J (1986) Communication and persuasion: central and peripheral routes to attitude change. Springer, Berlin
57. Prochaska JO, Velicer WF, Rossi JS, Goldstein MG, Marcus BH, Rakowski W, Fiore C, Harlow LL, Redding CA, Rosenbloom D, Rossi SR (1994) Stages of change and decisional balance for twelve problem behaviors. Health Psychol 13:39–46
58. Reeves B, Nass C (1996) The media equation: how people treat computers, television, and new media like real people and places. Cambridge University Press, Cambridge
59. Reidsma D, Ruttkay Z, Nijholt A (2007) Challenges for virtual humans in human computing. In: AI for human computing. Lecture notes in artificial intelligence, vol 4451. Springer, Berlin, pp 316–338
60. Schulman D, Bickmore TW (2009) Persuading users through counseling dialogue with a conversational agent. Persuasive '09: proceedings of the 4th international conference on persuasive technology. ACM, New York, pp 1–8
61. Shneiderman B, Maes P (1997) Direct manipulation versus interface agents. Interactions 4(6):42–61
62. Sproull L, Subramani M, Kiesler S, Walker JH, Waters K (1996) When the interface is a face. Hum Comput Interact 11(2):97–124
63. Turunen M, Hakulinen J, Stahl O, Gamback B, Hansen P, Gancedo MCR, de la Cmara RS, Smith C, Charlton D, Cavazza M (2011) Multimodal and mobile conversational health and fitness companions. Comput Speech Lang 25(2):192–209
64. van Welbergen H, Yaghoubzadeh R, Kopp S (2014) AsapRealizer 2.0: the next steps in fluent behavior realization for ECAs. In: Intelligent virtual agents. LNCS, vol 8637. Springer, Berlin, pp 449–462
65. Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: toward a unified view. MIS Q 27:425–478
66. Verbeek PP (2006) Persuasive technology and moral responsibility: toward an ethical framework for persuasive technologies. Persuasive 6:1–15
67. Vilhjalmsson H, Cantelmo N, Cassell J, Chafai NE, Kipp M, Kopp S, Mancini M, Marsella S, Marshall AN, Pelachaud C, Ruttkay Z, Thorisson KR, van Welbergen H, van der Werf RJ (2007) The behavior markup language: recent developments and challenges. In: Pelachaud C, Martin JC, Andre E, Collet G, Karpouzis K, Pele D (eds) Intelligent virtual agents. Lecture notes in computer science, vol 4722. Springer, Berlin, pp 99–111
68. Wechsung I (2013) An evaluation framework for multimodal interaction: determining quality aspects and modality choice. Springer, Berlin
69. Westerterp KR (1999) Obesity and physical activity. Int J Obes 23(1):59–64
70. Yee N, Bailenson J, Rickertsen K (2007) A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. Proceedings of the SIGCHI conference on human factors in computing systems (CHI '07). ACM, New York, pp 1–10