Talk and Tools: The Best of Both Worlds in Mobile User Interfaces for e-Coaching

Robbert Jan Beun, Siska Fitrianie, Fiemke Griffioen-Both, Sandor Spruit, Corine Horsch, Jaap Lancee, Willem Paul Brinkman

Technical Report UU-CS-2016-008
November 2016

Department of Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
www.cs.uu.nl

ISSN: 0924-3275

Department of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
3508 TB Utrecht
The Netherlands

TAK AND TOOLS: THE BEST OF BOTH WORLDS IN MOBILE USER INTERFACES FOR ECOACHING

Robbert Jan Beun, Siska Fitrianie, Fiemke Griffioen-Both
Infromation and Computing Sciences, Utrecht University, The Netherlands
{r.j.beun, s.fitrianie, f.griffioen-both, a.g.l.spruit}@uu.nl

Corine Horsch, Willem-Paul Brinkman
Interactive Intelligence, Delft University of Technology, Delft, The Netherlands
{c.h.g.horsch, w.p.brinkman}@tudelft.nl

Jaap Lancee
Dept. of Clinical Psychology, University of Amsterdam, Amsterdam, Netherlands
j.lancee@uva.nl

Abstract. *In this paper a user interface metaphor, called Talk-and-Tools, is presented for automated e-coaching. The metaphor is based on the idea that people interact in two ways with their environment: symbolically and physically. As a proof of concept, an e-coaching system is implemented that supports an insomnia therapy on a smartphone. The role of a human coach was substituted with a cooperative virtual coach that is able to interact with a human coachee. In the interface of the system, we distinguish between a set of personalized conversations ('Talk') and specialized modules that form a coherent structure of input and output facilities ('Tools'). Conversations contained a minimum of variation to exclude unpredictable behavior, but included the necessary mechanisms for variation to offer personalized consults and support. A variety of system and user tests was conducted to validate the use of the system. During a six-week therapy, some users spontaneously reported the experience of building a relationship with the e-coach.*

Keywords: *user interface metaphors, talk-and-tools, automated e-coaching, conversation, tools, persuasive strategies, cognitive behaviour therapy, insomnia*

1 INTRODUCTION

In the interaction with the world that surrounds us, two types of information flow can be distinguished: a symbolic and a physical one [19, 1]. On the one hand, human beings can interact symbolically with other individuals using verbal and nonverbal signs to express their thoughts or ideas (e.g. asking, providing information or giving commands). On the other hand, humans can interact with the world physically by directly manipulating the domain (e.g., pushing, moving, lifting) and through sensing (e.g., hearing, seeing, touching). The essential difference between the two interaction types is that actions of the first type need an interpreter who can bridge the gap between the symbols and their actual meaning and purpose, while actions of the second type are related in a more direct manner to human perception and action [18, 33].

Both types of interactions have given birth to two major existing metaphors for designing user interfaces of computerized systems: a. the so-called conversational or dialogue interfaces that mimic the conversational behavior between two social actors [22] and b. the so-called model world or direct manipulation interfaces that involve continues representations of objects and real-time feedback of their behavior or manipulations [19]. With the advent of mobile computer systems and other technological progress, there is a need to re-conceptualize the two metaphors in the light of these changes. Users can 'Talk' via or with these systems with the aid of a conversational interface and they can work with and through these systems using 'Tools' that are based on direct manipulation interfaces. Here, the Talk pertains to spoken or written dialogues between human users and/or a computer system that behaves as a social actor, while the Tools enable the observation, analysis and manipulation of task objects in a physical or virtual domain. We will, therefore, refer to this concept as the Talk-and-Tools metaphor and assume that the new metaphor is the integration of the conversation and model world metaphors of user interaction with computer systems that play a role as a social actor.

In this paper, we apply the Talk-and-Tools metaphor in the domain of e-coaching for behavior change, in particular in the domain of insomnia therapy. The human coach will be replaced by an e-coach – a virtual social agent that mimics the role of a personal coach. To make the discussed concepts and their relations more concrete, the metaphor is realized in a mobile application called SleepCare – a system that supports an individual human client (the so-called 'coachee') with a tailored insomnia therapy on a smartphone. The interaction provides Talk reflecting dialogue and periodic consults with the e-coach, and Tools for data acquisition, visualization and analysis.

2 THEORETICAL BACKGROUND OF THE TALK-AND-TOOLS METAPHOR

The rationale behind the Talk-and-Tools metaphor is depicted in Fig. 1, where the distinctive interaction channels are represented in a triangular structure. The corners represent the external world and two social actors, and the arrows represent the information that flows between them. The external world is often called the domain of discourse and the social actors may embody the participants in a conversation. The Tools enable the participants to have different views on the external world or different ways to manipulate objects and their features [3].

The domain of discourse generally refers to a collection of entities being discussed between the two participants and can be represented by an object model, i.e. a set of concepts representing the entities, their properties and relations. For example, in the insomnia therapy domain presented below, key entities are 'bedtime', 'adherence', 'sleep efficiency', and 'exercise'. Not all entities discussed in the conversations between the two agents are represented in the domain of discourse – greetings for example, although vital for the human interaction, are not.

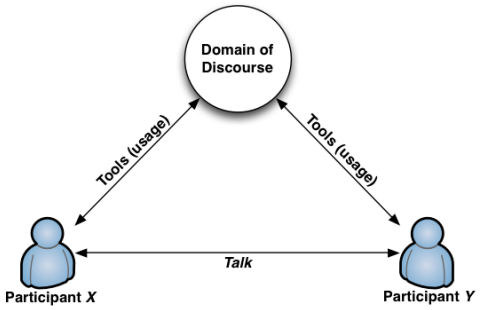

Fig. 1. The Talk-and-Tools metaphor: an external world and two social actors who participate in interaction; the arrows show the flow of information between entities

The domain of discourse can be replaced by a computer system, containing a database, an ontology, a simulation program or a digital game, and may even be in connection with a sensor or effector system to connect to the physical world. In the latter case, the system not only enables the measurement or manipulation of the physical conditions, such as light and location, but also conditions of the human body of a human participant (e.g., skin temperature, heartbeat). Moreover, one of the participants (e.g., Participant *X*) may be simulated by a computer system that behaves as a social agent and that has the ability to access the domain of discourse. To the human user (e.g., Participant *Y*) the appearance of the computer agent may take different forms, ranging from a simple text field to a human robot or a sophisticated embodied character that interprets and generates speech and shows nonverbal expressions.

The metaphor encourages a human user to view interaction with computer systems as: (i) a conversation with some intermediary; (ii) a sequence of direct manipulations and observations of some virtual world; or (iii) a combination of the two. The third type of interaction was applied in the design of, for instance, SHRDLU [39], Collagen [28] and the DenK-system [1]. In SHRDLU the concrete domain was a blocks world, in Collagen a virtual videocassette recorder and in DenK a simulated electron microscope.

To be able to ‘talk’, the computer agent employs an intermediate structure or discourse model for interpreting messages and constructing the response [31]. A simple approach to initiate a response from the agent is to develop a structure that uses a combination of word categories and pattern rules as in Eliza [38] and Alicebot [36]. Others use logic-oriented approaches that include contextual information and the ability to reason about the domain of discourse or about the mental state (e.g. beliefs and goals) of the other communication partner. For this purpose, the agent usually contains information structures about the dialogue, the domain, the tasks and the users. Examples of such structures are dialogue grammars or finite state machine to model the dialogue [20], Type-Theory to represent knowledge structures and discourse [1] (Ahn, et al., 1995), and Beliefs-Desires-Intentions architecture to reason about, for instance, goals and plans [2, 27, 37].

In the present work, we will introduce an activity schedule and a constraint-based model to trigger the system’s communication activities. Participant *X* will be replaced by an intermediate structure that mimics the conversational behavior of an e-coach in the insomnia therapy domain; the

direct interaction flow will be represented by Tools that form a coherent structure of visual and auditory interaction facilities.

3 COACHING IN THE DOMAIN OF INSOMNIA

An essential question for the development of a virtual participant in the Talk-and-Tools metaphor is what should be communicated by that participant at a particular moment in time, in particular when this participant plays the role of a coach. Here, we assume that the interactive behavior of the coach is determined by three elements: the typical behavior of a coach (e.g. goals, knowledge and responsibilities), the constraints of the therapy and the behavior of the other participant. Let us start with the typical behavior of a coach.

3.1 Coaching

[14] (p. 254) defines coaching as a result-oriented, systematic process in which the coach facilities the enhancement of life experience and goal-attainment in the personal and/or professional lives of normal, non-clinical clients. In the coaching process, two learning dimensions are distinguished that strikingly correspond with the Talk-and-Tools metaphor [32]: (1) learning as a social and collaborative practice and (2) learning through individual subjective experiences. The first dimension refers to the idea of a collaborative dialogue that unfolds between the coaching parties and where the coachee learns in interaction with the coach (‘Talk’); the second refers to the experiential and action oriented process of the coachee (‘Tools’). Through the dialogue, the coach and coachee exchange information to achieve alignment, set goals and discuss the results of the individual activities and exercises [15].

In contrast to therapy and teaching, coaching involves a more collaborative approach and respect of the coachee’s autonomy. Current society and daily lives of people are highly contextual and characterized by a growing degree of uncertainty. What counts as a solution for one person, not necessarily matters as a solution for another. Consequently, coaches will be cautious in offering solutions and help coachees to learn rather than to teach them. Coaching refers to a more explorative interaction style, a highly personalized and contextualized process of deliberation characterized by a continuous cycle of questioning, advising, agreement, observation, feedback and adaptation. Since computer systems contain only a fraction of the knowledge necessary to offer an adequate tailored therapy, the role of a coach seems an excellent candidate for a digital agent that supports automated self-help therapies; we will refer to these systems as e-coaches.

3.2 E-Coaching

E-coaching systems exist in many forms and functions, may include sophisticated avatars and may be targeted at a variety of domains. Here, we focus on fully automated mobile systems (i.e. without human interference) that aim at some sort of sustainable behavior change in the health domain. Health coaching dialogue systems have been developed on the basis of research methods from persuasive technology (e.g. [12]) and behavior medicine (e.g. [30]), and cover a wide range of health domains,

ranging from the treatment of obesity to sleep disorders and revalidation. In [7], for example, an e-coach was developed to support overweight people improving their lifestyle. The e-coach was able to help motivated participants to adhere to the program and lose weight. [8] described an e-coach that offers a fully automated treatment for depression, based on behavioral activation, a form of psychotherapy. [9] described an e-coach that offers an insomnia treatment in six weekly sessions. Also, the research by, for instance, [34, 29, 35] showed promising effects from this type of approaches in the health domain.

While many studies mainly focus on one specific domain, some projects aim at developing a reusable framework for e-coaching. Examples are the work by [6, 10, 23, 26]. In particular, e-coaches of [6, 10] provided fully automated Talk-based communication, albeit the latter used only text messages instead of a dialogue interface. In [6], various theory-driven computational models are introduced to develop a knowledge representation for behavior change counseling and focus on modeling counseling knowledge from which dialogue actions can be inferred.

In line with [6], we aim at developing a reusable framework based on generic coaching principles, but our perspective starts from interaction, communication and cybernetics theory. We will incorporate the combination of 'Talk' and 'Tools' as the central interaction concept in coaching. Before we discuss the underlying principles of our e-coaching model, we first briefly present the insomnia domain that will be used as a proof of concept.

3.3 Insomnia and Insomnia Therapy

Insomnia is a sleep disorder with a high prevalence (about 10% of the population) that can have severe individual and societal consequences (e.g., concentration problems, increased risks of accidents, depression, reduced productivity, increased absenteeism); people with insomnia have difficulty initiating and/or maintaining sleep. Today, it is widely accepted that Cognitive Behavior Therapy (CBT) produces sustainable positive changes in the condition of insomnia [24].

CBT for insomnia (CBT-I) is designed to change dysfunctional beliefs, attitudes, and behavior that support sleep-disruptive habits, thoughts, and emotions, and usually rests on two pillars: (1) a behavioral component to unlearn maladaptive habits and to learn sleep improving behavior and (2) a cognitive component to change a person's dysfunctional cognition, such as negative thinking or unrealistic expectations with respect to poor sleep. A CBT-I protocol usually takes between 6 to 10 weeks coaching sessions and offers a variety of exercise types that differ in aim and properties [25]: sleep restriction, stimulus control, relaxation, cognitive therapy and sleep hygiene. Sleep restriction involves curtailing the time spent in bed to stabilize the sleep pattern and lengthening sleep time as sleep efficiency improves. Stimulus control aims at restoring the coachee's association of the bed and the bedroom with sleep. Relaxation training involves methods aimed at reducing somatic tension. Cognitive therapy aims at the dysfunctional beliefs and attitudes. Sleep hygiene and education aims to make the person aware of practices and environmental factors that may either be detrimental or beneficial for sleep. These activities require the coachee to maintain multiple interactions with the coach over an extended period of time. The interaction frequencies can range from multiple times a

day (e.g. for physical exercises), to once a day (e.g. journaling), or to one or more times per week (e.g. evaluation of progress). The actual intervention is preceded by a one- or two-week baseline sleep-diary monitoring period.

4 THE SLEEPCARE E-COACHING SYSTEM

The SleepCare e-coaching system is responsible for an acceptable coaching process that conforms to the proper standards of CBT-I. Basically, it gives tailored advice to the coachee to do certain exercises, makes appointments for consultation and gives feedback with respect to the current state of the coachee and progress in the therapy. In addition, it includes a variety of persuasive strategies that are targeted at improving the coachee's adherence to the therapy [4]. These strategies usually relate to increasing ability and/or motivation of the coachee [13]. For that, a distinction is made between two types of CBT-I related activities: a. main activities that have to be performed and experienced by the coachee, for example, sleep restriction and b. supporting activities, such as activity scheduling and calculating sleep efficiency. To improve ability, supporting activities can be simplified by Tools; for instance, the paper and pencil sleep diary in CBT-I is replaced by an electronic version in SleepCare. An additional benefit of built-in Tools is that interaction with these Tools may give valuable information to personalize the intervention. So, for instance, sleep efficiency can be automatically calculated on the basis of the electronic sleep diary and integrated in a proposal for bedtimes. Motivational strategies, such as shared decision-making, are usually related to main activities.

4.1 The Coaching Process

The therapy offered by the SleepCare system takes about six weeks and includes a number of assignments and consultation dialogues based on exercise types from CBT-I. The coaching process consists of three phases [5]: the opening phase, the intervention phase and the closure phase (see Fig. 2). In the opening phase, the e-coach explains the details of the therapy, acquires relevant information to personalize the exercises, and establishes a shared commitment to the therapy. In the intervention phase, the actual therapy is conducted and the coachee is supposed to carry out the assignments. The closure phase starts when all assignments have been performed or when the coachee indicates the desire to withdraw; in this phase the e-coach and coachee evaluate the offered therapy and say goodbye.

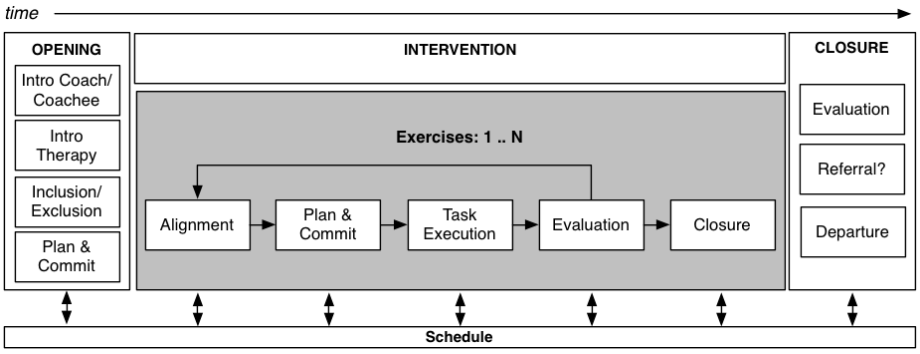

Fig. 2. The basic interaction model of the coaching process [5]. See text for an explanation.

The exercises in the intervention phase can be very different in nature (e.g. in duration and intensity), but always follow the same pattern (see Fig. 2). They start with an introductory conversation (Alignment) followed by the establishment of a shared commitment about the assignments (Plan & Commit). Subsequently, the coachee performs the assignments (Task Execution), which are usually evaluated after one week (Evaluation). After the evaluation stage and depending on the results of the previous week, an exercise may be re-established or ended (Closure).

4.2 The Behavior of the SleepCare System

The SleepCare-system behaves as an active partner, i.e. it does not only respond to interactional activities of the coachee, but also operates proactive and takes the initiative to interact with the coachee. For that, an activity schedule and a constraint mechanism are included. On the basis of these two mechanisms, notifications and reminders are sent to trigger the coachee to start a particular activity, for instance, a conversation or a therapy related exercise. The essential difference between the two mechanisms is that planned activities are triggered by a clock-event in the schedule, while unplanned activities are triggered by interrupting events that may be detected during the therapy.
The primary task of the activity schedule is to keep track of the various assignments that have been or should be carried out during the therapy, such as consultation dialogues or relaxation exercises. All scheduled assignments result from an agreed contract between coach and coachee (see Fig. 2). The end time of the activities is usually unknown, but in practice the time interval of assignments is relatively short (e.g., between 1 and 16 min).

The constraint mechanism detects violations of pre-defined rules (the so-called ‘constraints’) such as in cases of non-adherence by the coachee to the agreed assignments (see [5]). Whenever a violated constraint is detected, it has to be repaired by the e-coach by triggering a conversation. The status of constraints (i.e. violated or not) is verified by a constraint checker at least once a day or after a registered event, such as the completion of a conversation or an exercise.

To enable the functionality of the e-coach, various knowledge repositories can be consulted by the system. We distinguish, for instance, static and dynamic knowledge. Static knowledge covers general and timeless background information about the coaching process and is the same for all coachees. Examples of this type of knowledge are a representation of the constraints, the notifications and the structure and content of the conversations. The actual realization is based on dynamic knowledge about the coachee, such as subjective opinions about sleep quality, recorded activities and the current stage of the therapy. The dynamic knowledge may be considered as a long-term repository reflecting the formal knowledge that is built up during the coaching process.

4.3 The Talk-and-Tools Interfaces

The system provides one ‘Talk’-interface and a set of ‘Tool’-interfaces. The Talk-interface provides the functionality for the conversational component in a WhatsApp kind of interface structure; while the Tool-interfaces provide direct manipulation and observation of the insomnia domain.

Conversations in SleepCare consist of a sequence of natural language moves between coach and coachee, and are inspired by the periodic consults with a human coach. Tools are relatively independent modules that form a coherent structure of input and output facilities that enable the coachee to get access to the domain of discourse (observation and/or manipulation).

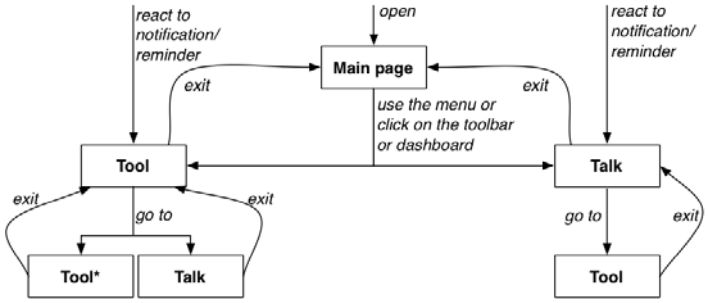

Fig. 3. Navigation from Talk to Tools and vice versa. Tool* refers to a different Tool from (the previous) Tool.

To optimize the interaction, the two interface types are constantly intertwined and, depending on the type of information exchange, Tools can be accessed from the Talk-interface and vice versa. Furthermore, the coachee can access any of the Talk-and-Tools interfaces from the main page of the system (i.e. using the provided menu, toolbar or dashboard after activating the e-coach; see Fig. 6, right picture) or simply by reacting on a reminder/notification. Fig. 3 shows the navigation through the both interface types. We will now discuss the two interface types separately.

4.3.1 ***Talk: Natural Language Conversations between Coach and Coachee.***

In the SleepCare therapy process, conversations may contribute to the following persuasive elements:

a. *Tunneling*: Conversation enables the e-coach to tell which activities should be performed and which should be refrained from. This is what we call ‘tunneling’. Tunneling may consist of introducing or discussing activities in the coachee’s daily life, such as filling in a sleep diary or changing bedtimes; in other cases activities and habits such as drinking coffee or alcohol just before bedtime are discouraged. Hence, conversation improves the coachee’s ability by tunneling therapy related activities.

b. *Improving transparency*: The use of natural language substantially increases the expressivity of the coaching system. Consequently, conversation also enables the e-coach to manifest a variety of speech acts that pertain to the explanation and clarification of the therapy and the use of various Tools. In other words, natural language conversation explains why, how and when activities should be performed.

c. *Creating awareness*: An important component of the therapy is the creation of the coachee’s awareness about their behavior, thinking and feelings. For that, conversations are included that

discuss results, progress and actual behavior of the coachee over a past period. These conversations are the stepping-stone to further personalization of the therapy.

d. *Tailoring the therapy*: Conversation enables the adaptation of exercises and communication to the circumstances and the characteristics of the coachee. To align e-coach and coachee, the therapy starts with the general introductory conversation where both e-coach and coachee get acquainted to each other. Also, frequent feedback loops are included where e-coach and coachee reshape the offered techniques, and where shared decision making facilitates the personalization of exercises. Conversational turns enable the user to adapt the amount of information and the presented discourse content: large chunks of text can be broken down into smaller pieces; options for content choices facilitates to present text to the interest of the coachee.

e. *Building a relationship and improving engagement*: The use of conversation introduces a wide range of social elements that creates a feeling of engagement and the presence of a social partner that collaborates, judges and gives motivational support. In particular personalization, transparency and speech acts such as ‘welcome’, ‘praise’, and ‘promise’ considerably contribute to the establishment of a relationship between e-coach and coachee that contributes to a feeling of trust and commitment.

In the design of the conversations, we need a balance between flexibility and rigorousness. Impracticable exercises, inaccurate or false information, and irrelevant discourse not only contribute to a feeling of irritation and mistrust, but also may even cause dangerous situations in cases of, for instance, car driving or operating complex machinery. Therefore, conversations should be carefully designed with a minimum of variation to rule out unpredictable behavior. On the other hand, conversations should include the necessary mechanisms for variation to include personalization and to improve the coachee’s engagement.

To meet these requirements, the dialogue component that generates the conversation uses a recursive tree-like specification called the E-Coach Markup Language (ECML) described in [11]. This format provides predefined templates to describe the interaction moves and the mechanism to collect knowledge from various sources and to update the coach’s knowledge base (the domain of discourse). Conversations have some flexibility with respect to the information content, the surface structure of the expressions and their timing. The flexibility of the content and surface structure of the moves is determined by branches in the tree and the variables in the templates. A particular branch in the tree may be selected on the basis of the coachee’s choice in a number of predefined options or on the basis of contextual information, such as the coachee’s sleep efficiency. Also, variables in the templates are assigned a particular value based on contextual information, such as time of day (e.g. ‘*Good morning*’, ‘*Good evening*’) or coachee information (e.g., ‘Your sleep efficiency is *79*%’). This idea is illustrated in the evaluation dialogues below where the values of variables are represented in italics and the coachee’s selections are in bold (adapted versions of the original Dutch dialogues; E: e-coach, C: coachee):

Dialogue I
E1: *Good morning*. Let's evaluate filling in the sleep diary. How did it go last week?
C: **Very easy** – Sometimes it's quite difficult – It is very hard to do
E2: Nice to read. I can see that you filled it in *7* out of *7* days. That's perfect!

Dialogue II
E1: *Good evening*. Let's evaluate filling in the sleep diary. How did it go last week?
C: Very easy – **Sometimes it's quite difficult** – It is very hard to do
E2: That is a pity. I can see that you filled it in *3* out of *7* days. That's not bad, but you can do better.

In general, the flexibility in the coachee's response is limited to the selection of one or more options (three in Dialogue I and II). Depending on the presented options, a selection may have different results:

a. Updating the knowledge base of the e-coach, for instance, the coachee's opinion about the therapy.
b. Activation of a specific tool, for instance, to fill in the diary, to start a relaxation exercise, or to edit the coachee's personal profile.
c. Continuation of the conversation; in case there is only one option to continue, the option is included to break down large chunks of text into smaller ones.
d. Closing the conversation; the status of the conversation changes into 'completed'.

The status 'completed' is included for different reasons. First, the order of conversations plays a crucial role. For instance, a scheduled evaluation about a particular exercise type may not be generated before the introduction is completed. In fact, there is only one active conversation and as long as a conversation is active, other 'ready-to-start' conversations are listed and opened only when the active conversation is completed (first in, first out). Also, completed conversations do not generate reminders. Another reason for the status is that it is used as a metric for the coachee's adherence.

With respect to the timing, two types of conversation may be distinguished: planned conversations that are triggered by the schedule (e.g., introductions and evaluations of exercises) and unplanned conversations that are triggered by the violation of a particular constraint (e.g., detection of non-adherence). Some, but not all, planned conversations are scheduled during the general introductory conversation; new conversations may also be scheduled as a consequence of performing particular conversations. In general, and depending on the behavior of the coachee, the frequency of interactions is around two or three conversations a week.

Each time a conversation is triggered, either by the schedule or by a constraint, a notification is sent to the coachee (e.g., 'Let's start the sleep training', 'Let's evaluate a relaxation exercise'); the actual realization of the notification is not part of the dialogue with the coachee, but is presented in the

notification area of the smartphone. In case a conversation is ignored, the coachee will be reminded periodically (usually the next day) until the conversation is completed.

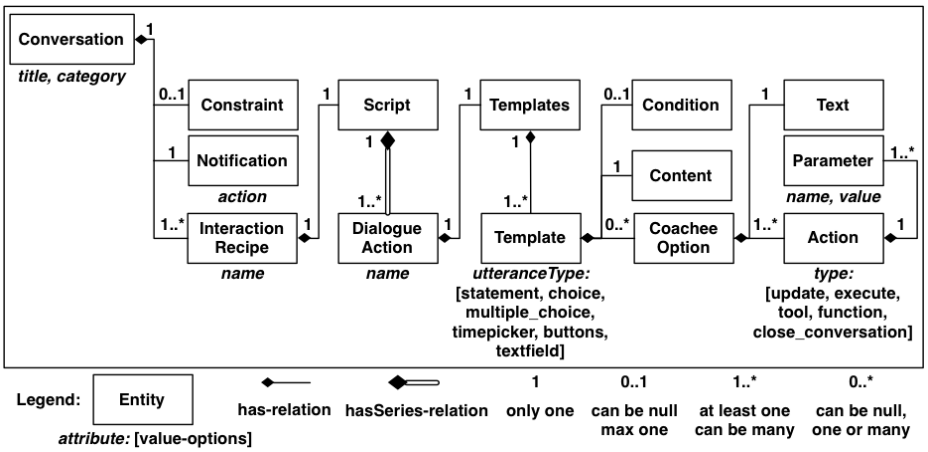

Fig. 4. Class diagram of a Conversation

In Fig. 4, the type `Conversation` is specified in a class diagram. Planned conversations contain two elements: a `Notification` and a nonempty set of so-called `Interaction Recipe`s. Unplanned conversations also include a representation of the `Constraint` that triggers the conversation.

`Interaction Recipe`s are the central entities in the specification of a conversation. They represent one e-coach turn in the conversation and end with one or more options indicating the coachee's turn. To some extent, they may be compared with the notion of 'adjacency pairs' in Conversation Analysis (see e.g., [21]). `Interaction Recipe`s can be called at each stage in the conversation to prepare the interaction with the coachee based on a `Script` containing a set of `Dialogue Action`s. `Dialogue Action`s consist of one or more `Template`s that can be selected on the basis of a `Condition`. The actual text of the conversation is embedded in the `Template`s. A `Condition` refers to expressions that refer to concepts in the domain of discourse and so define the current context of the coachee.

In Dialogue I and II, turn E1 and C are the result of the execution of one `Interaction Recipe`. Each sentence (or paragraph) in E1 is represented by one `Template`. In E1 the same `Template` s were included, but the assigned value of the variable in the greeting is different. In E2, the content of the first sentence is based on the coachee's selection, but the last sentence is based on the value of contextual information represented in the knowledge base. In other words, the information in the knowledge base not only influences the value of the variables of the `Template`s, but also the choice of a particular discourse fragment.

In Table 1, a more elaborate conversation is given with the actual interface of the smartphone. The fragment illustrates the first evaluation of the bedtime restriction exercise.

Table 1. Left: an example of a SleepCare conversation (translated from Dutch); right: the corresponding screenshot of the first part of the original conversation. The italics show the values that are based on the e-coach's knowledge base; in bold the choice of the coachee.

E1:	*Good evening Emily*! / Let's see how the exercise bedtime restriction went.
C1:	**Next**
E2:	What is your opinion about it? Is it easy or difficult to keep the new bedtimes?
C2:	Very easy – Easy – **Difficult** – Very difficult
E3:	Pity you think it is difficult. / Your mean bedtime in *7* hours, our agreement was *6* hours. Your mean sleeping time is *5.75* hours. / In your opinion it didn't go very well, but I think that you kept yourself quite well to our agreement. / You may give yourself a reward. Think of taking a bath, buying a magazine or a book, or spending more time on your hobby.
C3:	**Next**
E4:	Let's take a look at your sleep efficiency. When you started the bedtime restriction exercise, it was *71*%, now it is *81*%.
C4:	**Next**
E5:	…

To cover the main interaction needs for a six-week CBT-I training program, 19 different conversational types were included (e.g. introduction, planning and commitment, and evaluation). Ten of them are planned conversations and the others are unplanned. We believe that the format gives us the flexibility to personalize and engage, but has enough control to avoid the danger of undesired advice.

*4.3.2 **Tools.***

In SleepCare, Tools are represented by interactive visual and auditory interface structures that may contain words, pictures and sounds. Tools were included for:

a. *Registration of sleep data*: an electronic sleep diary functions as a Tool to update sleep related data, such as time in bed and sleep quality. The information was based on the most relevant items of the consensus sleep diary and data was stored to calculate at least the coachee's sleep efficiency. The interface of the tool consists of four pages: the first two pages ask for the time the coachee went to bed the previous evening and got out of bed in the morning; these pages

function as a delimiter of the offered input possibilities in the next page. This next page offers a structure of blocks that each represents a quarter of an hour where the coachee can register three types of information: in_bed/awake, in_bed/sleep, out_of_bed (see Fig. 5, left picture). In future versions of the system, the sleep diary can in principle be complemented or replaced by sensor information that updates the sleep data.

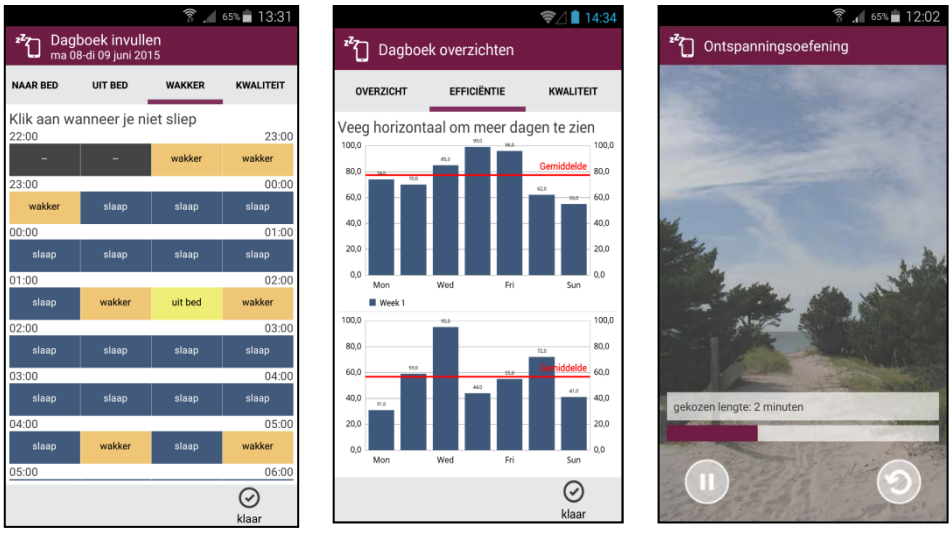

Fig. 5. Screen shots of the diary, results and relaxation tool. The left picture shows the block structure to register sleep data; the middle picture depicts the coachee's sleep efficiency; the right picture shows a 2 minute relaxation exercise.

b. *Observation of sleep data*: an overview is given about sleep data since the start of the program. The menu offers an entrance 'overview' (Dutch: 'overzicht') where three types of data are presented: a. sleeping and waking times, b. graphs of sleep efficiency (see Fig. 5, middle picture) and c. sleep quality and remarks. This enables the coachee to observe progress and other relevant information with respect to the program, such as past bedtime behavior.

c. *Relaxation exercises*: the relaxation tool offers a spoken progressive muscle relaxation exercise of various lengths (1, 2, 4, 8 and 16 minutes). The interface offers a picture of a relaxing environment, the option to choose a particular duration of the exercise (see Fig. 5, right picture) and a number of spoken utterances that guide the exercise. The utterances were spoken by a professional female voice actor and recorded by Teejay productions. We did not include music.

d. *Providing and changing general information*: various content structures are offered via the menu that provide sleep education and sleep hygiene. This includes non-personalized information about the sleep domain, sleep restriction and sleep hygiene (e.g., 'limit the amount of alcohol', 'during sleep our muscles relax and the waves of our brain slow down'; see Fig. 6, left picture). Moreover, the structures provide a disclaimer and enable the exchange of information about the system's stakeholders and settings, and the coachee's profile.

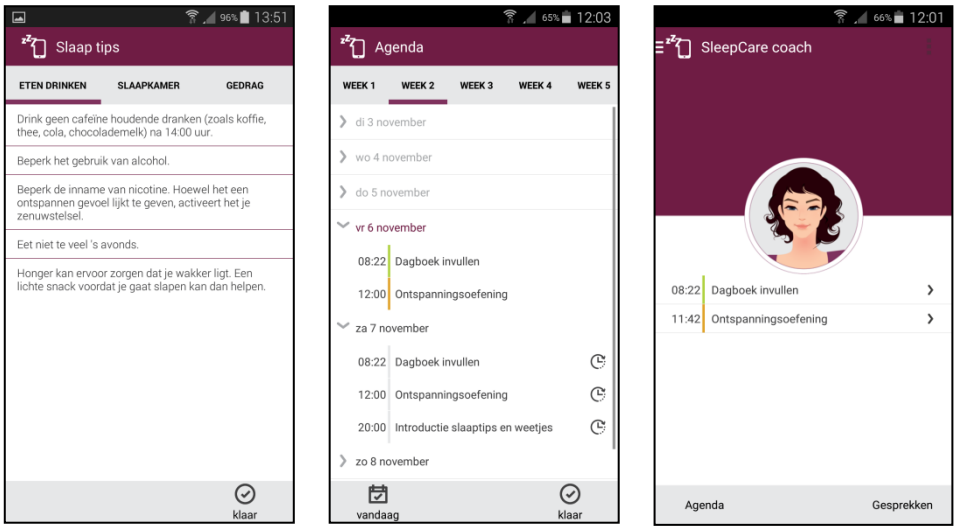

Fig. 6. The left picture shows a screen shot of a document that provides non-personalized behavior advice to improve the quality of sleep (in this case food and drink behavior). The middle and right picture show parts of the schedule interface.

e. *Scheduling*: to observe the properties of the scheduled assignments, such as starting-time and performance status, a corresponding interaction tool exists for each type of assignments. The most important function of the tool is to display the coachee's commitments and adherence to the assignments. For that, the interaction tool not only displays the scheduled starting time and date in the past, present or future, but also the performance status of an assignment by a colored line in front: red means missed, orange means nearly-missed (i.e. still possible to perform), green means completed and grey means planned. Scheduled assignments result from an agreed contract between coach and coachee that is established in a conversational setting and may trigger the generation of notifications and reminders. The middle picture in Fig. 6 shows the agenda that contains information about past and future commitments; the right picture depicts the planned activities of the day.

5 TESTING AND USER EVALUATION

The primary goal of the testing and user evaluation was to assess the quality of the SleepCare system to prepare it for a Randomized Controlled Trial (RCT) with insomnia patients. We aimed in particular at *defect detection* and *reliability estimation*. Both testing goals require elaborate testing and the possible input cases of the conversations and the Tools. For that, various activities were conducted, such as testing for error code, expert analysis and user evaluation.

5.1 Automated Testing

We tested the system for error code using a test automation approach running on the Android environment [16]. The approach met the challenges posed by a mobile CBT e-coach, such as the lead-time of the system, the contextual knowledge build-up during the lead-time, and unpredictable user behavior and absence of behavior. Our approach to the lead-time challenge of around six weeks was to simulate (parts of) the program in accelerated time within a few minutes.

To test the build-up of knowledge during the lead time, 32 realistic scenarios were written in collaboration with a sleep psychologist. The scenario's covered the most common sleeping patterns for general testing; more importantly, extreme sleeping cases were covered to make sure that the app acts responsibly to potentially dangerous situations (e.g. sleeping less than five hours). Each scenario was divided into three elements: activities, dialogue paths and sleep data. The first describes the activities of one day, the second contains the details of each conversation performed in a dialogue activity and the third consists of sleep data used for filling in the diary tool, which forms the actual basis of each scenario. These three elements are very much interconnected: sleep data and missed activities directly influence the content and occurrence of the dialogues and the possible paths in a conversation.

Since each conversation relates to a specific exercise type, we divided the testing process of the app in parts describing one type, or one week at a time. Each part could be tested separately using several scenarios, limiting the amount of possible courses through the six-week SleepCare program. Consequently, failures could be identified effectively and error propagation could be investigated in a smaller behavioral space. Three levels of aggregation were identified on which the issues could occur:

1. *Tools level*: unprocessed data from the various Tools.
2. *Aggregated data level*: the result of raw data processing, such as sleep efficiency.
3. *Conversational level*: constraints and dialogues.

Most difficult detected errors were caused by error propagation. For example, we found an error in the report of the average sleep time in a conversation; it turned out that the error originated from the input processing in the sleep diary tool. Such errors could only be found by studying the interaction between the e-coach and coachee spanning weeks of insomnia training.

5.2 Expert Evaluation

Because conversations play a decisive role in the therapy, all conversations (including their branches) were evaluated by a sleep therapist and two experts in Dutch communication. The experts were asked to analyze the system specifically on safety and quality of the advice, and on coherence and relevance of the conversational units. On the basis of the 32 scenarios, we recorded simulations that displayed the user interaction with the Talk-and-Tools interface under different conditions and time-scales. Using the test automation environment, the approach allowed the experts to playback prerecorded and predefined actions to compare the results with the desired behavior and to detect any errors.

Several issues were observed during these expert tests. For instance, the e-coach complimented the coachee in an evaluation conversation while the sleep data indicated that the coachee didn't sleep well.

After studying the recordings, it was found that the e-coach presented sleep parameters that did not match the user data from the scenario. These sleep parameters itself are calculated on the aggregation level using raw data from the first level. It turned out that there were issues with the storage of sleep diary data (level 1), and with the queries executed to gather data (level 2). In addition, there was a small issue with the rounding of variables resulting in noticeable mistakes after accumulating sleep data for a week (level 2). Thus, the observed bug appeared to be a conversational issue, but turned out to have several origins on lower levels of aggregation [16]. Other examples were errors that occurred when a conversation was triggered too soon, because the constraints were not restrictive enough. In addition to these tests, the communication experts improved the language used in the conversations considerably in terms of terminology used, length of the sentences and paragraphs, and grammar (e.g. avoid passive sentences as much as possible).

5.3 Evaluation of the Tools

In addition to the previous tests, three evaluation methods were applied to test the five Tools of the SleepCare system: a. heuristic evaluation, b. cognitive walkthrough, and c. user experience evaluation based on the thinking aloud method. All tests were performed in the laboratory. With respect to the heuristic evaluation 58 guidelines were applied, ranging from visibility of the system status to the protection of the user's privacy. In the cognitive walkthrough, six activities were evaluated: filling in the sleep diary, reading sleep information, starting a relaxation exercise, finding information about sleep problems, switching of an exercise reminder and rescheduling an appointment. In the user experience evaluation, ten subjects participated (five men, five women; age between 20 and 80), of which three reported having a sleeping problem. None of them had used a sleeping app before. Starting from the introductory screen (Fig. 6 right), subjects had to perform six tasks in line with the activities of cognitive walkthrough. Afterwards they had to fill in several questionnaires concerning usability of the interfaces.

The three evaluation methods resulted in 49 suggestions for improvement, ranging from visibility of the diary status to the addition of a clear privacy statement and the amount of text on the screen. These suggestion were classified in four categories that indicated the seriousness of the problem: 3: high priority, mandatory to solve this in a new release (1), 2: average priority, if possible solve this problem (9), 1: low priority: can be done, but not necessary (16), 0: cosmetic, change only when enough time left (13). The only high priority problem pertained to the ambiguous meaning of the 'ready' ('klaar') button. In general, participants had become familiar with the interfaces of the Tools and managed to perform successfully the tasks; all participants were satisfied with the system and found it easy to learn. An important issue was that the perceived usefulness reported by participants that had a sleeping problem was considerably higher than those not having a sleeping problem (one of participants stated: 'I don't have a sleeping problem, so I don't see how this will be of any help to me').

In the final preparation phase for the RCT, a pilot experiment was conducted to test the SleepCare-system with intended users. 24 people asynchronously started using the app in their daily life and

focused in particular on usability issues, the interaction with the e-coach and the circumstances of a possible system break down. Main success criteria for continuation were: a. no decrease in sleep efficiency and sleep quality, and b. no technical failure. The results gave us enough confidence in the system and the training program to execute a randomized controlled trial (RCT) as a next step in the research process.

5.4 First Remarks from the RCT

In the RCT, 74 participants were allocated to the SleepCare system and 77 to the waiting list condition. Since the RCT aims at the therapy and the results with respect to sleep improvement, the RCT will be discussed elsewhere [17]. Here we briefly discuss some of the comments that were reported by email by the participants during the experiment.
During the RCT, we received 942 emails that could be organized into three categories with respect to their content: the experiment, the therapy and the usability of the app. Participants asked questions, reported problems, provided suggestions for improvement and gave their judgement about various aspects of these categories. Of particular importance are remarks that refer to the relational aspects between app and subjects. One participant reported (translated from Dutch):

'May be it sounds strange, but I had the feeling of building a bond with my coach, although I am aware that it is just an algorithm. That is why I was shocked somehow by the way the therapy was closed. [..] I expected a closing conversation where my results were evaluated and where the coach wishes me success. It only indicates how realistic the conversations with the coach felt.'

Other participants reported:

'Bedtime restriction was a revelation. I had read about it, but it is so beautiful to see that the app helped me to give a personal and tailored advice.'

'It felt so nice that someone was so 'involved' in my sleeping problem'

Not all remarks were so positive, however. Another participant reported:

'I don't use the relaxation exercise. It makes me feel itchy.'

What is important here is that some of the participants reported having a bond with the system and had the feeling that the advice was tailored to their situation.

6 CONCLUDING REMARKS

In this paper we presented the so-called ‘Talk-and-Tools’ metaphor for human computer interaction and showed how both interaction types can be applied and integrated in the interface of mobile e-coaching systems. Let us summarize our main findings and add some considerations for the future.

As stated in the introduction, the metaphor was based on the insight that the interaction of humans with the environment can be separated into two different information channels, a symbolic and a physical one. In practice, the channels can be considerably interrelated, as in sentences such as ‘Put that there’ where the interpreter of the demonstratives (‘that’ and ‘there’) has to detect visual objects in the domain of discourse to disambiguate their meaning. Note that in practice ‘Talk’ also changes aspects of the physical world, e.g., in spoken interaction the wave of air that moves the eardrums and that is transformed into electrical signals that stimulates our brain; we have abstracted from these layered interpretation processes in our model. What is important, though, is that the symbols used in the conversation have a shared meaning and that the semantics of a relevant part of these symbols is related to the properties of the domain of discourse.

We have shown that the metaphor can be implemented in mobile applications such as smartphones. Compared to a personal computer, smartphones have small screens and limited capabilities with respect to natural language dialogue. These limitations force us to carefully consider the language used in the conversations: well-known terminology, short sentences and paragraphs that can be offered in chunks tailored to the user’s pace of information processing. Also, by offering the user only limited possibilities to respond, we were able to avoid sophisticated natural language interpretation as presented in, for instance, [1, 2].

In the design of the conversations, we balanced between flexibility and rigorousness. Flexibility improves the personalization of the e-coach’s contribution to the characteristics of the coachee and the results of the exercises. On the other hand, dangerous advice by the e-coach should be avoided at all costs. Therefore, conversations were designed with a minimum of variation to preclude unpredictable behavior. We believe that the formalism introduced in Section 4 and in [11] gives us the necessary stability to guarantee a safe advice and the variability to avoid the boredom of impersonalized texts. All this requires a *profound analysis of the coaching process*, however.

Including a symbolic level in the interaction in terms of natural language texts and dialogue opens a way to include a variety of powerful persuasive strategies to improve exercise adherence. Designing a system as a social actor [12] with natural language capabilities not only dramatically increases the expressive power of the system, but also enables us to include meta-level strategies based on social influence, such as commitment and shared decision-making. Dialogue offers coherence in the activities and reference to the future, the past and the various objects in the object domain. In that respect, *Talk* may be considered as the glue in a coaching program that sticks together the various coaching elements, such as individual exercises or homework, and the tools that support these elements.

Tools, on the other hand, emphasize the separation of the functionality of a system into independent, interchangeable modules that may adapt their interface to the situation at hand. Each Tool contains the functionality and an optimized set of interface constructs towards the domain of discourse (the API) and the coachee to focus one aspect of the desired functionality. The coachee can focus on the assigned task at hand, which is more manageable and can be done in a relatively short time. Furthermore, by inheriting features from the direct manipulation concept, Tools present the coachee with an easy to use, familiar method of interaction. In other words, Tools increase the coachee's ability and motivation to support the process of behavior change, because the approach can reduce the amount of effort that should be put into the therapy. To improve motivation even further, Tools may be included in the future with gaming elements or even replaced by a complete game.

The architecture of the system enables us to connect the domain of discourse to the physical world, which opens a way to include sensors that measure all types of information relevant to the therapy and to improve the coachee's ability (e.g. automated sleep detection instead of filling in a diary). We should be careful, though: in personal communication with one of the researchers, a participant reported afterwards that replacing a sleep diary by sensors would keep her away from interaction with the system and, consequently, would impede her relationship with it.

In conclusion, the Talk-and-Tools metaphor offers a generic approach to the design of automated systems for e-coaching. It has been implemented in the domain of insomnia therapy and evaluated in a Randomized Control Trial [17]. Future extensions are the inclusion of sensors (e.g. heart beat, skin conduction, location), the application in different domains of behavior change (e.g. obesity, depression, sustainability), users (e.g. shift workers instead of insomnia patients), stages of change (e.g. prevention vs. therapy) and the possibility to communicate with other human or non-human agents.

In the future, e-coaching may offer important advantages compared to traditional human-human coaching. It permits the assessment of relevant momentary information and the delivery of fully automated feedback tailored to the current context. Non-obtrusive sensor measurements enable the obtainment of objective data and integration of the data into the treatment protocol. And a communication platform facilitates information exchange between various user groups, i.e., peers, human therapists, and medical institutions. As such, the e-coach can be part of a (stepped) health-care structure with human care providers that discuss results and only take over when problems are detected.

Acknowledgements

We would like to thank Rene Ahn, Ronald Wondolleck, Coen Pijpers and the reviewers for their valuable contributions. This research is supported by Philips and Technology Foundation STW, Nationaal Initiatief Hersenen en Cognitie NIHC under the Partnership program Healthy Lifestyle Solutions.

References

1. Ahn, R., Beun R.J., Borghuis, T., Bunt, H.C., van Overveld, C.W.A.M.: The DenK-architecture: A fundamental approach to user-interfaces. AI Review, 8: 431-445 (1995).
2. Allen, J.F.: Natural Language Processing. 2nd ed. Redwood, CA: The Benjamin / Cummings Publishing Company Inc. (1995).
3. Beun, R.J.: On the generation of coherent dialogue: a computational approach. Pragmatics & Cognition, 9(1): 37-68, (2001).
4. Beun, R.J., Brinkman, W.P., Fitrianie, S., Griffioen-Both, F., Horsch, C., Lancee, J., Spruit, A.G.L.: Adherence in Automated e-Coaching: a Case from Insomnia Therapy. In: Persuasive Technology. In Proc. of the 11th Int. Conf. PERSUASIVE 2016, Springer International Publishing, LNCS 9638: 276-287 (2016).
5. Beun, R.J., Griffioen-Both, F., Ahn, R., Fitrianie, S., Lancee, J.: Modeling Interaction in Automated e-Coaching: A Case from Insomnia Therapy. In Proc of 6th Int. Conf. on Adv. Cognitive Technologies and Applications (2014).
6. Bickmore, T.W., Schulman, D., Sidner, C.L.: A reusable framework for health counseling dialogue systems based on behavioral medicine ontology. J. of Biomedical Informatics 44, 183-197 (2011).
7. Blanson Henkemans, O.A., van der Boog, P.J.M., Lindenberg, J., van der Mast, C.A.P.G., Neerincx, M.A., Zwetsloot-Schonk, B.J.H.M.: An online lifestyle diary with a persuasive computer assistant providing feedback on self-management, Technology & Health Care, 17(3): 253-257 (2009).
8. Both, F., Cuijpers, P., Hoogendoorn, M., Klein, M.: Towards Fully Automated Psychotherapy for Adults: BAS - Behavioral Activation Scheduling via web and mobile phone. In: 3d Int. Conf. on Health Informatics 2010, 375-380 (2010).
9. Espie, C.A., Kyle, S.D., Williams, C., Ong, J.C., Douglas, N.J., Hames, P., Brown, J.S.L.: A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application, Sleep, 35(6): 769-781 (2012).
10. Filler, A., Kowatsch, T., Haug, S., Wahle, F., Staake, T., Fleisch, E.: MobileCoach - A Novel Open Source Platform for the Design of Evidence-based, Scalable and Low-Cost Behavioral Health Interventions – Overview and Preliminary Evaluation in the Public Health Context. In Proc. of WTS. 1-6 (2015).
11. Fitrianie, S., Griffioen-Both, F., Spruit, S., Lancee, J., Beun, R.J.: Automated Dialogue Generation for Behavior Intervention on Mobile Devices. Procedia Computer Science 63, 236-243 (2015).
12. Fogg, B.J.: Persuasive Technology: Using computers to change what we think and do. San Francisco: Morgan Kaufmann (2003).
13. Fogg, B.J.: A behavior model for persuasive design. In Proc. of Persuasive '09. ACM NY, USA, Article no. 40, (2009).
14. Grant, A. M.: The impact of life coaching on goal-attainment, metacognition and mental health. Social Behavior and Personality, 31: 253-264 (2003).
15. Greene, J., Grant, A.M.: Solution-focused coaching: Managing people in a complex world. London: Momentum Press (2003).
16. Griffioen-Both, F., Spruit, A.G.L., Fitrianie, S., Lancee, J., Beun, R.J.: Testing for Mobile E-Health Intervention. In Proc. of SAC 2016. ACM digital library, 137-142 (2016).
17. Horsch, C.H.G.: A virtual sleepcoach for people suffering from insomnia. Doctoral Dissertation, Delft University of Technology, (2016).
18. Hutchins, E.L.: Metaphors for Interface Design. In Taylor, M. M., Néel, F. & Bouwhuis, D.G. (eds.) The Structure of Multimodal Dialogue, North-Holland: Amsterdam, 11-28. (1989).
19. Hutchins, E.L., Hollan, J.D., Norman, D.A.: Direct Manipulation Interfaces. Human Computer Interaction, 1(4): 311-338 (1985).
20. McTear, M.F.: Spoken dialogue technology: Enabling the conversational user interface. ACM Computing Survey. 34(1): 90-169 (2002).
21. McTear, M.F.: Spoken dialogue technology: Toward the conversational user interface. Springer-Verlag, Berlin (2004).
22. McTear, M., Callejas, Z., Griol, D.: The Conversational Interface: Talking to smart devices, Springer, (2016).
23. Mehrotra, A., Pejovic, V., Musolesi, M.: SenSocial - A Middleware for Integrating Online Social Networks and Mobile Sensing Data Streams, In Proc. of Middleware, ACM. 205-216 (2014).
24. Morin, C.M., Bootzin, R.R., Buysse, D.J., Edinger, J.D., Espie, C.A., Lichstein, K.L.: Psychological and behavioral treatment of insomnia: Update of the recent evidence (1998- 2004), Sleep 29, 1398-1414 (2006).
25. Morin, C.M., Espie, C.A.: Insomnia. A clinical guide to assessment and treatment. New York: Springer (2003).
26. Moturu, S., Khayal, I., Aharony, N., Pan, W. Pentland, A.: Sleep, Mood and Sociability in a Healthy Population, In Proc. of IEEE EMBS, 5267-5270 (2011).
27. Rao, A., Georgeff, M.: BDI agents: From theory to practice. In Proc. of ICMAS 95, San Francisco, USA, 312-319 (1995).
28. Rich, C., Sidner, C.L., Lesh, N.: Collagen: applying collaborative discourse theory to human-computer interaction. AI Magazine. 22(4): 15-25 (Oct, 2001).

29. Ritterband, L.M., Thorndike, F.P., Gonder-Frederick, L.A., Magree, J.C., Bailey, E.T., Saylor, D.K.: Efficacy of an internet-based behavioral intervention for adults with insomnia. Arch Gen Psychiatry, 66(7): 692–698 (2009).
30. Sallis, J.F., Owen, N.: Physical Activity and Behavioral Medicine. Thousand Oaks, CA: Sage, (1998).
31. Simmons, R.F.: Computational Linguistic, Natural Language Question Answering System: 1969, Communications of the ACM, 8(1): 15-29 (1970).
32. Stelter, R.: A guide to third generation coaching: Narrative-collaborative theory and practice. Dordrecht: Springer Science & Business Media, (2012).
33. De Souza, C.S.: The Semiotic Engineering of User Interface Languages. International Journal of Man-Machine Studies 39: 753-774 (1993).
34. Van Straten, A., Cuijpers, P., Smit, F., Spermon, M., Verbeek, I.: Self-help treatment for insomnia through television and book: a randomized trial. Patient Education Counseling. 74(1), 29–34 (2009).
35. Vincent, N., Lewycky, S.: Logging on for better sleep - RCT on the effectiveness of online treatment for insomnia. Sleep 32(6): 807–815 (2009).
36. Wallace, R.S.: Alicebot. Retrieved 20-02-2016 from http://www.alicebot.org/ (1995).
37. Wallis, P.: Believable conversational agents: Introducing the intention map. In: Pelachaud, C., Andre, E., Kopp, S., Ruttkay, Z. (eds.) Creating Bonds with Humanoids. In Proc. of the Workshop at AAMAS (2005).
38. Weizenbaum, J.: ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine, Comm. of the ACM 9 (1): 36–45 (1966).
39. Winograd, T.: Procedures as a Representation for Data in a Computer Program for Understanding Natural Language, MAC-TR-84, MIT Project MAC (1971).