A Virtual Assistive Companion for Older Adults: Design Implications for a Real-World Application

Christiana Tsiourti[1(✉)], Maher Ben Moussa[1], João Quintas[2], Ben Loke[3], Inge Jochem[4], Joana Albuquerque Lopes[5], and Dimitri Konstantas[1]

[1] Institute of Services Science, University of Geneva, Geneva, Switzerland
{Christiana.Tsiourti, maher.benmoussa, dimitri.konstantas}@unige.ch
[2] Laboratory of Automatics and Systems, Instituto Pedro Nunes, Coimbra, Portugal
jquintas@ipn.pt
[3] Noldus Information Technology, Wageningen, The Netherlands
b.loke@noldus.nl
[4] Zuyderland Medisch Centrum, Sittard, The Netherlands
i.jochem@orbisconcern.nl
[5] Association Valorisation Vieillissement (VIVA), Petit-Lancy, Switzerland
joana.albuquerquelopes@gmail.com

Abstract. Socially intelligent virtual agents are a promising solution to the increasing challenges of eldercare. This article describes an autonomous conversational agent system, simulating human-like affective behaviour to act as a daily life companion for older adults living at home. We present results from a user-centred design study that informed the design of the companion. The interaction with the user is based on a multimodal interface including automatic speech recognition, text-to-speech, and a graphical touch-based user interface. The companion offers support to older adults by locating objects, offering reminders, and guidance with household activities. It also supports emergency detection and handling. We present the companion's architecture and key components as well as the affective interaction paradigm and its implementation in the multimodal user interface. We conducted an exploratory evaluation study where the companion was introduced in 20 homes of older adults (aged 65+) in two EU countries. For a total duration of 12 weeks, various scenarios were tested, covering different types of interactions and tasks that occur in daily life. Qualitative and quantitative data on acceptance, perceived usability, and usefulness yield rich information on how agents perform in real-world settings and which factors influence their success in this context. We reflect on the results of the evaluation study in terms of lessons learned and discuss future opportunities for fellow researchers who are striving to bring virtual agents out of the laboratories into successful real world applications.

Keywords: Embodied conversational agents · Human-agent interaction · Agent architecture · Multimodal affective interaction · Real world user studies

Y. Bi et al. (eds.), *Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016*, Lecture Notes in Networks and Systems 15, DOI 10.1007/978-3-319-56994-9_69

1 Introduction

The worldwide population of adults aged 65 and above is expected to triple to 1.5 billion in 2050, according to the World Health Organization. The increasing demand for healthcare and quality of life services to support the ageing population has inspired researchers worldwide to explore the applicability of new intelligent technologies to support older adults to cope with the challenges of ageing and live independently for longer periods of time.

Embodied Conversational Agents (ECAs) are computer-animated characters exhibiting a certain level of intelligence and autonomy as well as social skills to simulate human face-to-face conversation, and the ability to sense and respond to user affect. A number of ECA systems have been successfully developed for various target applications to monitor, encourage, and assist older adults. Based on recent research findings [1–4], it is anticipated that this is a promising technology that can play an important role in maintaining the health, wellness, and independence of older adults in the future, either by complementing human care or acting as an alternative for those who cannot receive it due to high cost or low availability of care personnel.

A number of researchers have explored ECAs that interact with users over multiple conversations, ranging from a handful of interactions to interactions spanning long-term periods [5–7]. Nonetheless, most of the developed ECA systems are designed for specific controlled environments and have rarely made the step out of the laboratory as autonomous applications in real-world settings. As a consequence, there is still little information available about how autonomous ECAs perform and which factors influence their acceptance and success in contexts such as private households. To achieve useful and successful virtual agents, that maintain their users engaged in beneficial long-term relationships, we need to integrate these systems seamlessly in real-world environments and make them capable of interacting with humans autonomously, in an intuitive, natural and trouble-free way in everyday situations.

Towards this goal, this paper describes the design and implementation of a fully autonomous assistive companion for older adults and its exploratory evaluation in real-life settings for a total duration of 12 weeks. The companion, named Mary, operates on a static screen installed in the home environment of older adults. Mary's primary task is to identify and respond to her user's needs, and assist them to organize and accomplish their daily routine and self-manage their care. Mary engages in daily face-to-face interactions with older adults in which she provides comprehensible and simple information about their daily schedule. She provides reminders and information about social activities and helps older adults to connect with others by exchanging messages. Mary is equipped with camera-based visual perception and can locate lost objects when requested by the user as well as monitor the environment for emergency detection (e.g. falls) and handling. To maintain a beneficial, long-lasting relationship with older adults, Mary should be as natural and believable as possible, simulating a real-life human companion. She supports natural speech interaction, is capable of recognizing her user's emotional state and delivers appropriately tailored empathetic feedback, including emotional facial expressions.

After a brief review of related work, we describe the methodology and results of our user-centred design studies, conducted to understand the requirements of older adults regarding an assistive companion. In Sect. 4 we present the overall architecture which gives rise to the companion's human-like abilities. We focus on how Mary's interaction capabilities are achieved and describe the underlying psychology-inspired motivational and affective system, and the multimodal interface. Next, we present an evaluation study exploring how the system can enter into private households and how the target user group perceives and interacts with the companion in daily life settings. We present results on usability, usefulness and acceptance, and we share with fellow researchers lessons learned from this experience, towards successful and acceptable ECA systems for real world applications.

2 Related Work

ECAs are anthropomorphic or zoomorphic software characters, that are specifically lifelike and believable in the way they appear and behave. ECAs engage in human-like face-to-face conversations with their users; they recognize verbal and non-verbal input, and convey information using verbal and non-verbal output (i.e., intonation, facial expressions, gaze, head movements, hand gestures) [8]. Due to these characteristics, ECAs provide natural conversational interfaces, in contrast with traditional computer graphical user interfaces. ECAs are especially useful for building intuitive and engaging systems, for users suffering from age-related or other cognitive impairments [9, 10]. We briefly review prior work on ECAs and agent-based systems designed to address the needs of older adults.

2.1 Embodied Conversational Agents for Older Adults

There has been a number of studies on the use of ECAs to provide social and emotional assistance to older adults [14–16] and to address daily needs for an autonomous living [10]. Agents have also been used as coaches and wellness counselors in health behaviour change interventions for older adults [17, 18]. Wizard of Oz [13] experiments showed high acceptance and positive attitude towards ECAs as companions for older adults [10, 14]. Overall, these studies suggest that ECAs are particularly capable of capturing the attention and engaging older adults in active tasks [11] and have the potential to fulfil the need of humans to build up social relationships [12].

Bickmore introduced and explored relational agents [7], namely ECAs that build and maintain long-term, social and emotional relationships with human users. Building a human-computer relationship can be advantageous for the users in a number of applications; for example when building a daily life companion to assist humans through conversations. Relational agents have underlying computational models of affect and maintain memory of specific interactions. By recalling and referring to past interactions, relational agents imitate the way people incrementally get to know each other and form trusted relationships with their users [18, 19]. Kasap et al. [20, 21] also discuss a virtual agent designed for long-term interaction. The agent's behaviour in

driven by an underlying relationship model of the user which is updated based on the content of each interaction.

A handful of studies were conducted in which autonomous ECAs were installed for prolonged periods in the daily living environments of older adults. In an exploratory pilot study by Ring et al. [22], an ECA designed to provide longitudinal social support to isolated older adults using empathic feedback was placed in the homes of 14 older adults for a week. Results demonstrated significant reductions in loneliness based on self-reported affective state. In a randomized controlled trial by Bickmore et al. [17], a virtual exercise coach designed to encourage sedentary older adults to walk more was installed in homes for two months, followed by another ten months where participants had the opportunity to continue the interaction in a kiosk in their clinic waiting room. Participants in the intervention group walked significantly more on average than participants from the control group.

3 Design Methodology

3.1 Understanding the Role of a Daily Life Companion

In order to identify "useful" functionality and to decide which "social skills" are required for an assistive companion, we had to construct a holistic view of the multifaceted daily life routine of older adults and to explore the circumstances of their care at home and in assisted living environments. We conducted two user-centred design studies, based on focus groups and individual interviews with older adults aged 65+ (N = 20), professional caregivers (N = 12) and psychologists specialized in the aging process (N = 2). The methodology and qualitative results of the studies are described in detail in previous work [23]. The key user requirements are outlined here.

3.2 Requirements for an Assistive Companion Agent

Appearance: Based on the focus group interviews, the central requirements for the companion's appearance were realism and a friendly face and voice that makes older adults feel comfortable. Participants ranked different virtual characters, covering a range of different possible looks: different gender and age, realistic human-like look versus cartoonlike look and formal versus informal look. The most wanted character was a female with an informal look.

Verbal and Non-verbal Abilities: The study participants emphasized the importance of interacting with a companion that has "smooth" communicative skills, such as the abilities to perceive and interpret verbal and non-verbal input (e.g., spoken language and facial expressions) and to express verbal and non-verbal behaviour. Speech was the users' preferred interaction modality since it is more intuitive and can be used when no free hands are available, or the user is at a certain distance from the companion.

Decision-Making and Personality: Just as happens with humans, the companion should express coherent behavioural and emotional responses that can be interpreted

by the users as indicative of a natural social personality. Participants identified several social skills such as being non-intrusive, considerate and proactive as important to engage in favorable long-term interactions.

Assistance and Care Provision: Caregivers provided valuable insights into the kind of assistance that is required by older adults on a daily basis and served as human role models, helping us to understand how older adults might interact with a virtual companion while carrying out specific tasks at home. The main needs and requirements of older adults are related to aspects of communication and socialization, guidance with household activities, the organization of daily activities, safety, wellness, and leisure.

Perception and Awareness: The companion should perceive the environment where it operates and be aware of the current situation of the user. Older adults envisioned a companion that can interpret their emotional state (e.g., joy, sadness) and activity (e.g., idle, working) and use this information to adapt its behaviour to offer support. Further, they expected that the companion can identify emergency situations (e.g., falls) and help them detect various objects in the household.

4 System Architecture Design

To develop an assistive companion with the abovementioned characteristics, we developed an overall system architecture (Fig. 1) that resembles what has been proposed as a reference architecture for ECAs [24]. The well-orchestrated interplay of the components of perception, decision making, and synthesis of verbal and non-verbal output gives rise to the human-like capabilities of the assistive companion.

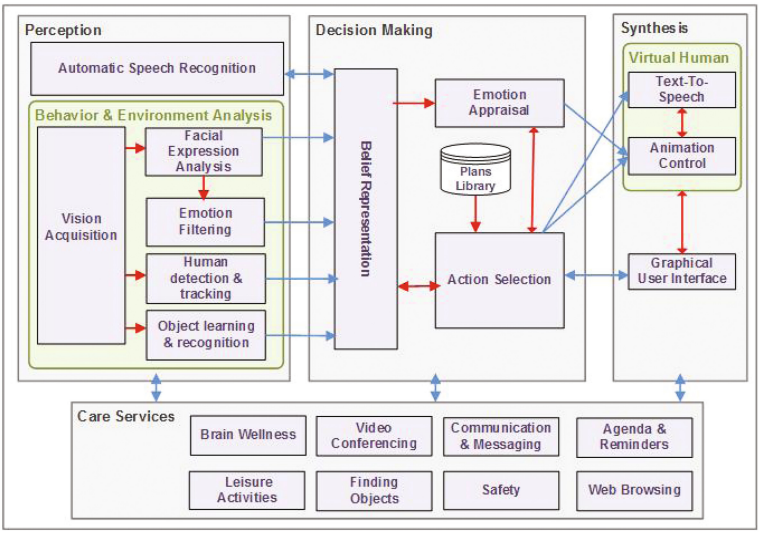

Fig. 1. Overview of the system architecture design

4.1 Hardware Platform

The companion operates on an all-in-one stationary computer (Lenovo ThinkCentre Edge 93z All-in-One), mounted on the wall or placed on top of furniture, at a height ranging from 80 to 120 cm, allowing the user to interact from a standing or sitting position. A built-in microphone is used to enable speech as an input modality, and high definition speakers are used for producing audio output. A built-in webcam (1920 × 1080) located in the front of the computer and a Microsoft Kinect camera mounted on a pan–tilt unit, are used to collect colour (RGB) and depth (RGB-D) visual data. The Kinect camera was mounted at a height of approximately 130 cm, with the viewing angle approximately 60° vertical by 70° horizontal, ensuring a suitable field of view for detecting standing persons and seeing objects at heights of up to 90 cm (e.g., kitchen counter) and also on the ground. A light version of the system runs on a portable tablet (Microsoft Surface Pro 3) with a built-in webcam (1920 × 1080), microphone and high definition speakers. Compared to the stationary computer, the tablet has a smaller screen size and lower processing capacity, which is somewhat reflected in the responsiveness of the companion. A depth camera is not included in this version. This design was requested by a sub-group of the target group, consisting of older adults who maintained an active lifestyle outside their home and preferred to interact with the companion also while on the go.

4.2 Perception Components

The Perception components collect input from the speech (Automatic Speech Recognition) and visual modalities (Vison Acquisition) and send information, such as the user's speech commands and perceived emotion, to the Decision Making components for further processing.

(1) *Automatic Speech Recognition (ASR)*

We use the Microsoft Kinect for Windows SDK, which provides the tools to acquire the sound signal and provides the Speech Platform SDK to perform speech recognition. The companion supports three languages: English, Dutch, and French. For the English and French language, Microsoft provides models as part of the Kinect for Windows SDK 2.0 Language Packs, which is up-to-date and works reasonably well. For Dutch, only an older and less accurate version of the acoustic model is provided by Microsoft as part of the Microsoft Speech Platform 11 Runtime Languages. For each of the three languages, we generated custom grammar files (i.e., language models) which indicate which utterances are recognized and contain the probabilities of sequences of words to guide and constrain the search among alternative word hypotheses during speech recognition. To do this, we collected words and phrases frequently used by older adults in the context of their daily social interactions. Subsequently, we defined appropriate response rules for each utterance.

(2) *Human Behaviour and Environment Analysis*

We take advantage of recent advancements of computer vision and the availability of low-cost depth-aware sensors, to equip the companion with vision-assisted competencies for full-scale human body detection and tracking as well as automatic facial expression analysis.

- **Human Detection and Tracking**
  We developed a component for full-scale human detection and tracking based on [25]. We follow a probabilistic approach to classify human behaviour based on movement features using visual data (RGB-D) captured by a Microsoft Kinect camera. Each acquired depth frame is initially processed to extract information about the scene background and foreground. Subsequently, the full 3D human skeleton is detected in each frame, segmented and tracked in the environment. For each frame, the detected depth-based pixels assigned to a human skeleton are fed to a Bayesian Network that estimates the pose of the human body by fitting a 3D skeleton model including 3D positions/orientations of the basic body limbs. This 3D skeletal tracking is performed across frames and using the related length, velocity, and acceleration information the Bayesian Network can recognize abrupt body motions that indicate instability or a sudden fall.
- **Facial Expression Analysis**
  The Facial Expression Analysis module is based on and extends the kernel of the commercial product FaceReader [26, 27], developed and validated for emotion recognition in laboratory settings. It receives as input raw RGB images and classifies the six basic emotions described by Ekman [28] (happy, sad, angry, surprised, scared, disgusted) and a neutral state. Each expression has a value between 0 and 1, indicating its intensity (0 = expression absent, 1 = expression fully present). As facial expressions are often caused by a mixture of emotions, it is likely that two (or even more) expressions occur simultaneously with high intensity. The sum of the intensity values of the seven expressions at a particular point in time is, therefore, normally not equal to 1.
- **Object Learning and Recognition**
  The Object Learning and Recognition module is based on the novel Global Hypothesis Verification approach proposed by [29] for verifying model hypotheses in cluttered and heavily occluded 3D scenes. The 3D shape descriptors are calculated from depth data corresponding to different views of the object, acquired from the RGB-D sensor. Thus, object recognition is done based on point cloud data (PCD) using specifically the Point Cloud Library (PCL) framework [30]. The PCD is fed through a 3D descriptor matching stage, which then feeds its correspondences to the correspondence grouping algorithms, resulting in the identification of a known object. This approach has the inherent ability to detect significantly occluded objects without increasing the number of false positives. Thus, it is especially useful in the uncontrolled environment of a household where the companion operates.

- **Emotion Filtering**
  The Emotion Filtering module receives inputs from the Automatic Facial Expression Analysis module and determines the user's short-term and long-term affective state. This information is sent to the motivational system, where two types of responses are generated. In the case of short-term emotions, the companion instantly mimics the user, since earlier research suggests that mimicry increases liking and trust, both in other people and in virtual agents [31]. A time window is defined within which all emotion intensities are summed over all samples captured. The emotion with the highest value is selected as the most dominant for that specific time window. When a persistent emotion is detected, the companion initiates appropriate dialogue or triggers a care service. For each output of the Facial Analysis Module, the dominant emotion is computed and stored in a map that keeps track of which emotion was dominant at which timestamp. Simultaneously, the Emotion Filtering module checks within a predefined time window, how long each emotion was present. If a certain emotion is present for more than a predefined threshold value, this emotion is indicated as active.

4.3 Decision-Making Components

The decision-making components manage the flow of interaction, reason on events and context, and simulate the companion's human-like behaviour. The logic is based on several high-level goals and motives [32], that influence the emotional and adaptive behaviour of companion and drive it to take appropriate decisions. These goals/motives, maintained in a plan library, are related to the user's well-being as well as the user's satisfaction with the system.

- **Belief Representation**
  The system maintains knowledge in a structured form, in a belief system, which drives decision-making and action selection. This includes a mental representation of the perceived events, the conversation status and a mental model of the user. The implementation of the belief system draws similarities with David Traum's information state approach and other structured knowledge base approaches [33].
- **Emotion Appraisal**
  The Emotional Behaviour System appraises and simulates the companion's emotional reactions at each state of the interaction based on an evaluation of its internal goals and motives and in response to incoming events which, may originate either externally (e.g., user input) or internally (e.g., assertion of a new goal) [34]. This evaluation takes into account several appraisal variables (e.g., desirability, expectedness, agency, and control potential) to assess an event and determine the appropriate emotion. The implementation is based on Roseman's [35] psychological model as well as partly on the models of Scherer [36] and Ortony and colleagues [37].

- **Action Selection**
  The action selection component is responsible for selecting and executing update-rules (also known as AI Plans) on the belief representation system and is based on the basic Belief-Desire-Intention (BDI) concept from Artificial Intelligence. Based on the current state of the interaction, the action selection component selects a speech output action or the emotion of the agent, which is in turn, communicated to the synthesis components. The companion then generates the appropriate animation related to the received action command.

4.4 Assistive Care Services

A set of care services are developed to solve various daily life problems, provide information, organize the user's social activities and enable the user to participate in society:

Self-management of Daily Activities: To help elderly manage their daily activities the companion maintains a digital version of their personal agenda, assists to keep it up to date (e.g., enter or modify social activities, medical appointments) and issues appropriate reminders during the day.

Safety: The companion monitors the elderly and recognizes abrupt human body motions that indicate instability or a sudden fall or a call for help, based on a predefined voice command. In the case of an emergency, the companion initiates a dialog to acknowledge the detection and reassure the user and if necessary, automatically dispatches a phone call to a designated caregiver.

Guidance for Household Activities: Since elderly are prone to forgetfulness, the companion offers assistance for locating objects around the house employing real-time vision-based detection of previously learned objects, whenever possible (i.e., when an object is in the field of view of the camera).

Wellness and Leisure: The companion knows the user's personal interests and proposes leisure activities or guides users through brain wellness activities (i.e., guided meditation and relaxation exercises) and a program designed to teach concrete strategies to improve the performance of prospective memory to apply these skills in everyday life.

Communication and Socialization: To facilitate communication with friends and caregivers the companion guides the elderly through the use of a private message exchange system as well as Skype communication and Internet navigation.

4.5 Synthesis Components

The synthesis components generate the appropriate verbal and non-verbal behaviour of the companion and deliver relevant information to the user via the graphical user interface.

(1) *Virtual Human*

The 3D virtual human (Fig. 2) closely simulates human conversational behaviour through the use of synthesized voice and synchronized non-verbal behaviour such as head nods and facial expressions. It is built on top of the Behavioural Markup Language (BML) realization engine SmartBody [38], which is an XML description language for controlling the verbal and nonverbal behaviour of humanoid ECAs. Each BML block describes the physical realization of behaviours such as speech and facial expressions and the synchronization constraints between these behaviours. SmartBody enables easy control of the facial expressions of the virtual human using the Facial Action Coding System (FACS). This lets us individually control each action unit (AU) to achieve particular emotions. SmartBody utilizes Microsoft's native text-to-speech engine, to convert text into natural spoken speech. To control the virtual human's mouth and to synchronize the mouth shapes with the speech, speech is analyzed, and the visemes (the shape of the mouth when making a certain utterance) are extracted and used to move the virtual human's mouth.

Fig. 2. The companion and the GUI

5 Multimodal Interaction Paradigm

The users interact with the companion using a multimodal interface including automatic speech recognition and a touch-based graphical user interface (GUI). The text on the GUI buttons and the speech commands of the speech interface are consistent with respect to each other to simplify the learning process.

5.1 Natural Language Interface

The natural speech interaction interface is based on human-to-human dialogues from daily life interaction scenarios. The dialogues are highly situation and task dependent. We have worked with older adults and caregivers to design a set of dialogue scripts corresponding to each of the assistive care services mentioned above. In most of the scenarios, the user initiates the conversation with the companion, and the interaction

follows a question-answer sequence. At each step of the interaction, the speech recognition is updated with a new speech dictionary that is related to the phase where the dialogue is at that moment. According to the dictionary, the users can issue commands in an imperative, sentence-like structure. When one of the expected commands is detected by the speech recognition, the appropriate response is selected based on the current internal beliefs and goals of the motivational system. This information is sent to the text-to-speech component, which in turn dynamically generates the companion's natural speech output.

A natural speech interaction scenario is shown below (Fig. 3). The dialogue begins when the user requests information about social activities taking place in the city. The companion then announces a list of activities and their corresponding details. Subsequently, the user requests to participate in an activity. The companion acknowledges that she received the user's request and reformulates it as a question. This, in turn, is accepted by the user and his/her personal agenda is updated with the details of the new activity. Additional non-verbal feedback is provided to make the interaction more natural and human-like. A set of facial expressions (e.g., joy, sadness, concerned) has been designed and integrated with the speech output to make the companion seem more believable and life-like.

User: Please show me some activities for this week.
Companion: On [*activityDate*], from [*activityStartTime*] to [*activityEndTime*], the activity [activityName] will take place in the location [*activityLocation*]. It will cost [activityCost] euro.
User: I want to register for [*activityName*] on [*activityDate*]
Companion: Ok, should I add the [*activityName*] in your agenda?
User: Yes, please.
Companion (Facial expression [*Joy*]): Ok, you have registered! The activity [*activityName*] has been added to your agenda.

Fig. 3. Sample dialogue between the user and the companion

5.2 Graphical User Interface (GUI)

In addition to the natural speech interaction, the users can issue commands to the companion using buttons and elements on the GUI. The GUI follows standards of User Interface design and is the result of in iterative design process taking into account the requirements and feedback from older adults. The final design of the GUI (Fig. 1) is restricted to a minimum of necessary functions and provides clear feedback on the companion's status. This design ensures that the users can interact with the touch-based interface intuitively after a short introduction. The GUI is structured into three main areas: (1) on the right-hand side there is a menu with big clearly spaced buttons which provide access to the care services; (2) on the top, there is a status bar including information like the current time, date, and weather, as well as notifications. On the right-hand side of the status bar, icons are animated to inform the user about the companion's status (e.g. enabled/disabled, currently talking/listening for commands). Additional buttons are provided to enable the on-screen keyboard and the help menu

and to exit the system; (3) the main content area is reserved for displaying the companion. Occasionally, information about the different care services needs to be displayed on the screen. Thus, the companion is temporarily minimized on the left-hand side of the screen.

6 Evaluation Study

To assess the effects of the companion, we conducted a longitudinal evaluation study where the companion was introduced in private households of 20 older adults for a total duration of 12 weeks. The aim of the study was twofold: (1) to examine empirically how our target population interacts with an ECA-based companion integrated in their private household; (2) to explore how our target population subjectively assesses the companion in terms of acceptance, perceived usability and usefulness in realistic scenarios covering various types of interaction and tasks that occur in daily life. Quantitative and Qualitative data were collected at the beginning, middle and completion of the study (questionnaires, focus groups) and during each interaction with the companion throughout the study (log files, think aloud protocol, diary).

A. *Setting and Procedure*

The study took place in two test-beds: in Switzerland, with 13 elderly living alone in apartments of an assisted living complex, and the Netherlands with 11 participants living alone in independent apartments. From the second group, 4 participants dropped out for health reasons (N = 2) and loss of engagement (N = 2). In total, 20 older adults of average age 77.91 completed the study. Some of those had participated in the design studies, and the rest were recruited voluntarily.

The evaluation study consisted of three parts. In the introduction phase (T0) the system was installed in the older adults' apartments, a demonstration took place featuring all the capabilities of the companion and instructions about how to interact with it. Participants received a help manual with further details and were instructed to interact freely with the companion during their daily activities. 16 participants chose the version running on the all-in-one computer, which was installed in the living room, and 4 participants preferred the tablet version. Baseline measurements were taken, including demographics, the World Health Organization Quality of Life (WHOQOL) questionnaire [39] and semi-structured interviews. An intermediate evaluation (T1) was conducted after one month of interaction with the companion, and the System Usability Score (SUS) questionnaire [40] was administered. The final evaluation (T2) was conducted after the third month of interaction and included the SUS, WHOQOL and debriefing questionnaires and semi-structured interviews.

The companion used in the study was completely automated and constantly available (24 h/7 days a week). Although the procedure was designed to evaluate the independent and prolonged interaction with older adults in their environment, some participants, especially those with limited experience with technology, faced technical difficulties that prevented them from using the system autonomously in a smooth and trouble-free way. In the country 2 test-bed, two trained staff members visited the 13 participants 2–3 times a week, for at least 1 h and used the system together. At the

second test-bed, the 7 participants used the system by themselves and relied on the help of staff members on demand, when they encountered a problem.

6.1 Instruments and Measures

(1) *Qualitative*

Focus Group Interviews: We conducted semi-structured interviews with open questions about the participants' level of autonomy and quality of life before (T0) and after using the system (T1, T2). Also, we collected their expectations (T0) and their conclusions (T2) about the system's impact on their daily life, regarding autonomy, daily life organization, activity, and memory.

Thinking aloud protocol [41]: Participants were invited to say what came into their mind, what they were looking at, thinking, doing, and feeling as they performed tasks with the companion (T0, T1, T2).

Diary: Participants were invited to keep a diary where they reported remarks, questions, and ideas throughout the study.

(2) *Quantitative*

Questionnaires: Participants answered a set of questionnaires with the help of a facilitator who ensured that the questions were correctly understood.

- Demographics: age, gender, years of study and autonomous living, experience participants with technology and ECAs.
- WHO Quality of Life-BREF (WHOQOL-BREF) [39]: a 26 item questionnaire which measures the domains of physical health, psychological health, social relationships, and the environment.
- System Usability Score (SUS) [40]: a 10 item questionnaire (5 point Likert scale) giving a global view of subjective assessments of usability in terms of effectiveness (e.g., can users successfully achieve their objectives?), efficiency (e.g., how much effort and resource is expended in achieving those objectives?), satisfaction (e.g., was the experience satisfactory?).
- Debriefing: a 22 item questionnaire, including 17 questions based on a 5 point Likert scale and five open-end questions, designed to examine how participants subjectively assess the different attributes of the companion, the GUI, and care services.

Log Files: during each interaction with the companion, information was automatically logged, including the user's verbal and nonverbal input, the companion's verbal and nonverbal output, GUI interaction and the use of different care services.

7 Results and Lessons Learned

Learning from our own and our participants' experiences, we present our findings and discuss implications of deploying long-term studies with ECAs in real world settings.

7.1 How Older Adults Perceive and Interact with a Companion

(1) *Acceptance and Interaction*

Regarding the overall acceptance, the companion was well received in the living environments of older adults during the 12-week duration of the study. Feedback about Mary's appearance was variable. In Switzerland, Mary's age, gender, and look were rated the most positive ($M = 2.73$, $SD = 0.59$), whereas her facial expressions ($M = 3.5$, $SD = 0.65$) were rated with the lowest values. According to qualitative results, participants from the Netherlands evaluated Mary as a spontaneous, charming and friendly lady. A particularly appreciated feature in both test-beds was Mary's ability to mirror the emotional facial expressions of users. Although the implementation of mimicry was relatively simple, participants perceived the companion as persuasive and likable. These findings demonstrate the potential of a largely unutilized non-verbal ECA skill that can endorse interpersonal rapport and empathy, and consequently lead to smoother interactions.

According to the final evaluation, the companion did not fully reach the original expectations of the users. This was primarily attributed to mismatched expectations about the companion's verbal communication capabilities. In fact, the expected scope of the companion's communication capabilities was so complex that an unconstrained spoken interaction was technically not feasible, although ASR systems have advanced remarkably in the last years. More than half of participants reported that the speech commands were rigid, and required a lot of memorization. Misunderstanding or unexpected verbal behaviour by the companion rapidly increased the participant's stress levels and generated feelings of frustration. Training the interaction with the companion is a complex task for older adults, and a lot of repetitions are required to establish sustainable routine interactions. An ideal solution would be to have more flexibility and variety in the speech commands.

Based on these results, for empirical ECA studies targeting older adults, it is essential that all the interaction components run as robustly as possible, are fault-tolerant, and support repair mechanisms. Furthermore, we recommend establishing a longer initialization phase in which the participants get acquainted with the capabilities and limitations of the agent. This can reduce the novelty effect bias and minimize the discrepancy with mismatched expectations.

(2) *Usability*

Regarding usability, the average SUS scores of the two test-beds were 62.2 for Switzerland and 52 for the Netherlands. The combined average score was

58.8 (range 37.5–75), indicating that the interaction with the companion was perceived as "above average." In both testbeds, the mean score increased between T0 and the T1, because participants felt more comfortable after various interactions with the companion. Results from Switzerland indicate a high variability of usability ratings (SD = 12.41). This finding is quite informative since it indicates that usability may be "correct" to "good" for older adults who are already familiar with technology, whereas it seems to be between "poor" and "OK" for participants who are not familiar with technology. The fact that the Netherlands participants did not use the system autonomously, but with the guidance of care staff members, led to lower usability ratings. Qualitative results indicated that overall these participants liked the idea behind the system; however, they consider that they need more training and assistance before they can interact with the companion autonomously in their daily life.

(3) *Usefulness*

Participants were asked to rank the assistive care services offered by the companion based on their usefulness. The qualitative and quantitative results revealed significant variability, between the two test-beds and between participants. For county1, the preferred services were the memory training, the agenda (with requests for improvements and additional features) and the Internet navigation and Skype communication. For the Netherlands, the preferred services were the agenda, the social activities, and message exchange. Participants from Switzerland reported that the system improved their daily life in general (M = 2.33; SD = 0.9), helped them to be more motivated to perform daily life activities (M = 2.36; SD = 1.01) and to be more active (M = 2.53; SD = 0.92). They also considered that the brain wellness exercises they performed with the companion could help them improve their memory (M = 1.87; SD = 0.52). In summary, these findings reveal that the aspect of usefulness is tightly coupled with the older adults' care needs and lifestyle choices. To remain useful in the long-run, a companion should learn and adapt to its users' changing needs and social and physical context.

7.2 A Companion in Real-World Settings

Applying an autonomous ECA system to an everyday life household environment poses several challenges. Firstly, it requires harmonizing the requirements and daily lifestyle choices of older adults with the technical goals of the system. Secondly, various components developed and validated primarily in laboratory or controlled settings have to cope with a variety of issues which occur in the uncontrolled and sometimes unpredictable context of daily life.

(1) *Ethical and Privacy Considerations*

Several ethical issues must be addressed when using human-like socially intelligent ECAs for the sensitive population of older adults. First and foremost, a companion should not be obtrusive or stigmatizing for the users, nor restrict their privacy. During the study, some older adults expressed worries related the use of cameras and uncertainty about whether or not their interactions with the companion would be recorded.

Further restrictions were imposed by the household environment layout, for example, insufficient space or inappropriate location for mounting the computer and Kinect camera at the optimal positions (functional for the system vs. comfortable for the user). Introducing hardware in the household is inevitable, but ideally, no recorded data should be stored, and all the devices should be integrated into the existing furnishings. In addition, it is essential that older adults are not deceived into thinking they are interacting with a companion that is capable of doing things that only humans can do in an interaction. It must be clear that the companion is merely a computer-driven character with limited capabilities.

(2) *Technical Aspects*

- **Automatic Speech Recognition**
  The companion is equipped with ASR, which gives the ability to match natural voice patterns against a determined carefully-spoken vocabulary of words and phrases. Despite the wide selection of intuitive vocabulary, many participants found it difficult to adapt, remember and use the pre-defined speech commands. A possible way to tackle this problem is to use a more sophisticated ASR language model which accepts and recognizes free natural speech for large vocabularies in real time. The system could also incorporate real-time definition, and configuration of new speech commands and phrases that the user may desire to introduce and assign to any of the existing commands. Therefore, the speech interface will adapt to match the personal preferences, daily habits and cultural background of the users. Another ASR-related challenge was the presence of additional people and/or ambient noise in the household and what this meant for the companion's behaviour. Ongoing conversations (e.g., visitors, phone calls, radio) were often recognized as speech commands and triggered the unexpected reaction of the companion. To cope with this situation, we devised the "privacy" mode of the system, which was enabled when a user wanted to interact with additional people or did not want to be interrupted by the companion. In this mode, the cameras and microphone remained in standby mode. Thus, video data or voice commands were not captured, and audio was not emitted by the system. For ASR systems, silence is almost as important as what is spoken, because it delineates the start and end of an utterance. Although the use of a noise-cancelling headset can significantly increase our speech engine's performance, to respect the user requirements for unobtrusiveness we did not impose the use of a headset but relied only on the computer's built-in microphone. After tuning the audio components and the speech engine, acceptable error rates were achieved, for distances up to 3 m. Beyond this distance, ASR engines, in general, have a low performance, due to different types of sound distortion (e.g., background noise, echo, and reverberation, the orientation of the speaker's head) [42].

- **Human Detection and Tracking**
  The performance of 3D human detection and tracking while older adults were executing daily activities was challenging. Firstly, humans were only monitored within the field of view of the RGB-D sensor. The performance of the human body detection and tracking for sitting users was frequently deteriorated due to occlusion (e.g., by furniture present in the room). Furthermore, the Microsoft Kinect sensor was often considered as invasive due to bulky and fixed installation, as it had to be mounted in a particular location on a pan–tilt unit and with a fixed orientation. To tackle these challenges, it is required to use sensor networks to broaden the monitoring coverage and to promote the adoption of less invasive methods of installation (e.g. sensors built into furniture, Internet of Things). Nevertheless, our setup allowed the implementation of an accurate vision-based fall detection approach.
- **Object Learning and Recognition**
  In the current state of the art, recognition of objects by PCL has some limitations: it is invariable to scaling; recognizing all poses of an object requires a 3D model; recognition does not work well with occlusions. To make this a feasible option, given the high number of objects that the users wanted the system to recognize, objects were learned just with one surface instead of a complete 3D model. As a result, recognition worked only for the learned pose. Because of the limitations of scaling, the object to be recognized had to be within close range of the distance at which it was trained (88 cm). Moreover, recognition was not successful if the object was rotated between 210° and 315° relative to the learned model. Therefore, using current vision-based technology, it is not sufficient to achieve satisfactory results regarding object recognition at long range, which was what users expected. To tackle this challenge, we plan to integrate complementary technologies to detect, identify and track objects (e.g. RFID).
- **Automatic Facial Expression Analysis**
  The household environment conditions were challenging for automatic facial expression analysis especially with respect to dealing with variance in the visibility, pose, and orientation of the face as well as suboptimal lighting conditions. The performance of the facial expression classification was hindered when the face was partially hidden, for instance by glasses, a hat or hefty facial hair. Especially glasses with thick and dark frames, often used by older adults, reduced the performance significantly. A polarization filter on the camera can help to avoid reflections in the glasses improving thus the classification results. Further, it was difficult to classify facial expressions during daily life activities, for example, while a person is eating, because the hand covers part of the face while putting food in the mouth and the muscles in the face move. For optimal facial expression analysis the person should ideally stand or sit looking frontally into the camera (angle <40°). Finally, ambient light conditions influenced the classification. Diffuse frontal lighting is desirable, and strong shadows or reflections should be avoided, thus finding optimal lighting conditions imposed restrictions on the placement of the camera (functional for the system vs. comfortable for the user). As the commercial kernel of our Automatic Facial Expression Analysis module is being trained with more natural poses from people, a wider variety of people and more natural circumstances, we expect these restrictions will be less strict in the future.

8 Conclusion and Future Work

ECAs are a promising solution to the challenges faced by the ageing population who wish to live independently for longer. We designed and developed an autonomous human-like companion that provides support to older adults in private households. We conducted a 12-week evaluation study where the companion interacted with 20 potential end-users in daily-life scenarios. We followed a qualitative methodology and looked at the users interacting in daily-life scenarios with the companion. The target group perceived the companion as appropriate and useful and as a whole the results of our study yield promising evidence that ECAs offer a user interface worth pursuing in the context of daily life assistance in independent living environments.

We present our findings as a starting place for future researchers interested in conducting longitudinal studies with ECAs in real-world settings. More specifically, we discuss design implications for ECAs deployed in residential environments of older adults: first, the way an ECA is introduced in the daily living environment of older adults requires caution; second, managing the expectations of the users towards the ECA's natural interaction capabilities is crucial; and third, residential environments pose specific restrictions, and domestic ECAs must be adaptive and fault-tolerant to uncontrollable factors.

Our system and evaluation study had several limitations, and our results provide clues to how our companion design can be enhanced. A particularly challenging issue which emerged from our study is the need for intuitive interaction and clear and explicit feedback to the user, both in terms of the companion's current state and to confirm user requests. Our intention is to improve further our multimodal affective interaction paradigm to make the dialogues more natural, robust and fault-tolerant. We are also improving the companion's human-like social behaviour by expanding the internal motivational and emotional behaviour system with features like small-talk dialogues and episodic memory [43]. In future studies with an enhanced version of our system, we will collect further empirical evidence from fully autonomous long-term interactions, to explore how a companion can enhance the performance of daily tasks and which additional factors influence the perception and acceptance of ECAs in daily life scenarios. This will enable us to improve the design and development of successful assistive companions to promote a better quality of life and long-term independent living for older adults.

Acknowledgment. This work was supported by the European research projects Miraculous Life (Grant No. 616421) and CaMeLi (Grant No. 010000-2012-16). We express our gratitude to all the study participants.

References

1. Bickmore, T.W., Schulman, D., Sidner, C.L.: A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. J. Biomed. Inform. **44**(2), 183–197 (2011)

2. Bickmore, T., Gruber, A., Picard, R.W.: Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ. Couns. **59**(1), 21–30 (2005)
3. Lisetti, C., Yasavur, U., de Leon, C., Amini, R., Rishe, N., Visser, U.: Building an on-demand avatar-based health intervention for behavior change. In: Proceedings of the Twenty-Fifth International Florida Intelligence Research Society Conference, pp. 175–180 (2012)
4. de Rosis, F., Novielli, N., Carofiglio, V., Cavalluzzi, A., De Carolis, B.: User modeling and adaptation in health promotion dialogs with an animated character. J. Biomed. Inform. **39**(5), 514–531 (2006)
5. Schulman, D.: Embodied Agents for Long-Term Interaction. Northeastern University, Boston (2013)
6. Bickmore, T.W., Picard, R.W.: Establishing and maintaining long-term human-computer relationships. ACM Trans. Comput. Hum. Interact. **12**, 293–327 (2005)
7. Bickmore, T.W.: Relational agents: effecting change through human-computer relationships. Massachusetts Institute of Technology (2003)
8. Cassell, J.: More than just another pretty face: embodied conversational interface agents. Commun. ACM **43**, 70–78 (2000)
9. Kramer, M., Yaghoubzadeh, R., Kopp, S., Pitsch, K.: A conversational virtual human as autonomous assistant for elderly and cognitively impaired users? Social acceptability and design considerations. In: Lecture Notes in Informatics (2013)
10. Yaghoubzadeh, R., Kramer, M., Pitsch, K., Kopp, S.: Virtual agents as daily assistants for elderly or cognitively impaired people. In: Proceedings of the 13th International Conference Intelligent Virtual Agents, vol. 8108, p. 91 (2013)
11. Klein, J., Moon, Y., Picard, R.W.: This computer responds to user frustration: theory, design, results and implications. In: Proceedings of CHI 1999 Extended Abstracts on Human Factors in Computing Systems, p. 242 (1999)
12. Nijholt, A.: Disappearing computers, social actors and embodied agents. In: Proceedings of the International Conference on Cyberworlds, pp. 128–134 (2003)
13. Kelley, J.F.: An iterative design methodology for user-friendly natural language office information applications. ACM Trans. Inf. Syst. **2**(1), 26–41 (1984)
14. Vardoulakis, L.P., Ring, L., Barry, B., Sidner, C.L., Bickmore, T.: Designing relational agents as long term social companions for older adults. In: Proceedings of the 12th International Conference on Intelligent Virtual Agents, vol. 7502, pp. 289–302 (2012)
15. Sakai, Y., Nonaka, Y., Yasuda, K., Nakano, Y.I.: Listener agent for elderly people with dementia. In: 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 199–200 (2012)
16. Ring, L., Barry, B., Totzke, K., Bickmore, T.: Addressing loneliness and isolation in older adults. In: Affective Computing and Intelligent Intreraction (2013)
17. Bickmore, T.W., Silliman, R.A., Nelson, K., Cheng, D.M., Winter, M., Henault, L., Paasche-Orlow, M.K.: A randomized controlled trial of an automated exercise coach for older adults. J. Am. Geriatr. Soc. **61**(10), 1676–1683 (2013)
18. Bickmore, T.W., Caruso, L., Clough-Gorr, K., Heeren, T.: 'It's just like you talk to a friend' relational agents for older adults. Interact. Comput. **17**(6), 711–735 (2005)
19. Campbell, R.H., Grimshaw, M.N., Green, G.M.: Relational agents: a critical review. Open Virtual Real. J. **1**(1), 1–7 (2009)
20. Kasap, Z., Ben Moussa, M., Chaudhuri, P., Magnenat-Thalmann, N.: Making them remember—emotional virtual characters with memory. IEEE Comput. Graph. Appl. **29**(2), 20–29 (2009)

21. Kasap, Z., Magnenat-Thalmann, N.: Building long-term relationships with virtual and robotic characters: the role of remembering. Vis. Comput. **28**(1), 87–97 (2012)
22. Ring, L., Shi, L., Totzke, K., Bickmore, T.: Social support agents for older adults: longitudinal affective computing in the home. J. Multimodal User Interfaces **9**, 79–88 (2015)
23. Tsiourti, C., Ben Moussa, M., Joly, E., Wings, C., Wac, K.: Virtual assistive companions for older adults: qualitative field study and design implications. In: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) (2014)
24. Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., Yan, H.: Human conversation as a system framework: designing embodied conversational agents, pp. 29–63 (2001)
25. Khoshhal, K., Aliakbarpour, H., Quintas, J., Drews, P., Dias, J.: Probabilistic LMA-based classification of human behaviour understanding using power spectrum technique. In: Proceedings of the 2010 13th International Conference Information Fusion, pp. 1–7 (2010)
26. Den Uyl, M.J., Van Kuilenburg, H.: The FaceReader: Online facial expression recognition. In: Proceedings of measuring behavior. pp. 589–590 (2005)
27. Van Kuilenburg, H., Wiering, M., Den Uyl, M.: A model based method for automatic facial expression recognition. In: Proceedings of the 16th European Conference on Machine Learning. pp. 194–205. Springer, Heidelberg (2005)
28. Ekman, P., Keltner, D.: Universal facial expressions of emotion. Calif. Ment. Health Res. Dig. **8**(4), 151–158 (1970)
29. Aldoma, A., Tombari, F., Di Stefano, L., Vincze, M.: A global hypotheses verification method for 3D object recognition. In: ECCV, pp. 511–524 (2012)
30. Rusu, R.B., Cousins, S.: 3D is here: point cloud library. In: IEEE International Conference on Robotics and Automation, pp. 1–4 (2011)
31. Verberne, F.M.F., Ham, J., Ponnada, A., Midden, C.J.H.: Trusting digital chameleons: the effect of mimicry by a virtual social agent on user trust. In: Berkovsky, S., Freyne, J. (eds.) Persuasive Technology, vol. 7822, pp. 234–245. Springer, Berlin (2013)
32. Austin, J.T., Vancouver, J.B.: Goal constructs in psychology: structure, process, and content. Psychol. Bull. **120**(3), 338–375 (1996)
33. Traum, D., Larsson, S.: The information state approach to dialogue management. In: Current and New Directions in Discourse and Dialogue, pp. 325–353 (2003)
34. Ben Moussa, M., Magnenat-Thalmann, N.: Toward socially responsible agents: integrating attachment and learning in emotional decision-making. Comput. Anim. Virtual Worlds **24** (3–4), 327–334 (2013)
35. Roseman, I.J.: Appraisal in the emotion system: coherence in strategies for coping. Emot. Rev. **5**(2), 141–149 (2013)
36. Scherer, K.R.: Emotions are emergent processes: they require a dynamic computational architecture. Philos. Trans. R. Soc. Lond. B Biol. Sci. **364**(1535), 3459–3474 (2009)
37. Colby, B.N., Ortony, A., Clore, G.L., Collins, A.: The cognitive structure of emotions. Contemp. Sociol. **18**(6), 957 (1989)
38. SmartBody. http://smartbody.ict.usc.edu/
39. T.W. Group: Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol. Med. **28**(3), 551–558 (1998)
40. Brooke, J.: SUS: a quick and dirty usability scale. Usability Eval. Ind. **189**(194), 4–7 (1996)
41. Lewis, C.: Using the "Thinking Aloud" Method in Cognitive Interface Design. IBM T. J. Watson Research Center, Yorktown Heights (1982)
42. Wölfel, M., McDonough, J.: Distant Speech Recognition. Wiley, Hoboken (2009)
43. Lim, M.Y.: Memory models for intelligent social companions. In: Human-Computer Interaction: The Agency Perspective, pp. 241–262. Springer (2012)