Co-designing an Embodied e-Coach With Older Adults: The Tangible Coach Journey

EL KAMALI, Mira <http://orcid.org/0000-0003-2895-6867>, ANGELINI, Leonardo, CAON, Maurizio, DULAKE, Nick <http://orcid.org/0000-0003-1841-5848>, CHAMBERLAIN, Paul <http://orcid.org/0000-0001-6643-3480>, CRAIG, Claire <http://orcid.org/0000-0002-3475-3292>, STANDOLI, Carlo Emilio, ANDREONI, Giuseppe, ABOU KHALED, Omar and MUGELLINI, Elena

Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Co-designing an Embodied e-Coach With Older Adults: The Tangible Coach Journey

Mira El Kamali, Leonardo Angelini, Maurizio Caon, Nick Dulake, Paul Chamberlain, Claire Craig, Carlo Emilio Standoli, Giuseppe Andreoni, Omar Abou Khaled & Elena Mugellini

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2023.2171332

Co-designing an Embodied e-Coach With Older Adults: The Tangible Coach Journey

Mira El Kamali[a], Leonardo Angelini[a,b], Maurizio Caon[b], Nick Dulake[c], Paul Chamberlain[c], Claire Craig[c], Carlo Emilio Standoli[d], Giuseppe Andreoni[d], Omar Abou Khaled[a], and Elena Mugellini[a]

[a]Humantech Institute, University of Applied Sciences and Architecture of Western Switzerland, Fribourg, Switzerland; [b]School of Management, HES-SO, Fribourg, Switzerland; [c]Lab4Living, Sheffield Hallam University, Sheffield, UK; [d]Department of Design, Politecnico di Milano, Milano, Italy

ABSTRACT
This article describes a tangible interface for an e-coach, co-designed in four countries to meet older adults' needs and expectations. The aim of this device is to coach the user by giving recommendations, personalized tasks and to build empathy through vocal, visual, and physical interaction. Through our co-design process, we collected insights that helped identifying requirements for the physical design, the interaction design and the privacy and data control. In the first phase, we collected users' needs and expectations through several workshops. Requirements were then transformed into three design concepts that were rated and commented by our target users. The final design was implemented and tested in three countries. We discussed the results and the open challenges for the design of physical e-coaches for older adults. To encourage further developments in this field, we released the research outputs of this design process in an open-source repository.

1. Introduction

Vocal assistants are entering our homes. Big IT companies such as Amazon, Google and Apple pushed their voice assistant devices with aggressive prices and free services. The widespread adoption of such devices and services and the large amount of voice data collected allow improving further the algorithms for voice recognition, which enables today's unprecedented natural language understanding. As conversational capabilities of vocal assistants keep improving, these devices look promising for the assistance of older adults, providing a convenient and natural interface, eventually an e-coach, for advising and assisting in daily living (Chattaraman et al., 2019; Martin-Hammond et al., 2019). With the increasing population of older people, the advancement of technology and the diffusion of these technologies worldwide, e-coaches (Banos & Nugent, 2018) hold promise for encouraging healthier lifestyles, as recognized by the European Commission, which has funded several research projects in this area (Angelini et al., 2020; Funding & Tender Opportunities, 2022). E-coaches for older adults' wellbeing were recently investigated in a systematic review (El Kamali et al., 2020). Most of the studies reviewed by the authors used a conversational agent to deliver recommendations to the users. While some of the agents were embodied in virtual avatars (e.g., Bickmore et al., 2013), most of them were embodied in robots, which often provided anthropomorphic or zoomorphic cues (e.g., Matilda (Khosla & Chu, 2013), Roberta (Sansen et al., 2016)). While most of the e-coaches found in literature were designed for daily assisted living of fragile older adults, commercial voice assistants and smart speakers were designed for the large public, without involving older adults in the design process. Indeed, Trajkova et al. found a mismatch between users' expectations and needs and the features provided by these devices (Trajkova & Martin-Hammond, 2020).

In this article, we discuss the insights collected through a co-design process with older adults from four European countries (UK, The Netherlands, Italy and Spain), shedding light on the needs and expectations of older adults, to guide future development of devices that could serve as smart assistants for such target group. This co-design process was carried out in the NESTORE EU project. NESTORE is an e-coach that aims to help older adults living independently at home to improve and maintain their well-being. The whole intervention is based on the Health Action Process Approach psychological model (Schwarzer & Luszczynska, 2008) and is fed by a decision support system (Orte et al., 2018) to personalize the e-coach recommendations based on user needs, preferences and habits. The e-coach is multi-domain, i.e., it tackles different domains of well-being being (physical, nutritional, social and cognitive) and multimodal, i.e., it provides recommendations through different interfaces, namely, a mobile app, a chatbot (a text-based application integrated in the same app), and finally a tangible

CONTACT Mira El Kamali mira.elkamali@hes-so.ch Humantech Institute, University of Applied Sciences and Architecture of Western Switzerland, Fribourg, Switzerland.

coach, which is an embodied voice assistant. This tangible coach is co-designed to be the physical companion for promoting healthy lifestyles in older age as part of a more articulated e-coach. The goal of the co-design process was to design an interface for the e-coach able to increase the user's trust, as well to develop an emotional bonding, in order to foster user engagement with the system and, in particular, to encourage her/him to adhere to the well-being coaching plan. We present first the state of the art of embodied conversational agents (CA) for coaching and previous work on CA designs, with a particular focus on three main axes: the physical design, the interaction design, and data privacy and control. We present then the co-design process carried out for the Tangible Coach. Following the UK Design Council's Double Diamond process (Figure 1), we first explored older people's attitudes to technology, investigated the user needs for a virtual coach; successively, we iterated among different design concepts and prototypes, until reaching the definition of a tangible coach, focusing the design process around three main axes. The tangible coach was produced into 70 samples, which then were tested by 33 older adults across Europe. Finally, we discuss the design implications for vocal assistants that might be more accessible and senior-friendly as well as the open technical challenges that require further investigation.

In this article, we are referring to chatbot as a text chatbot on a mobile/desktop app, to vocal assistant as a smart voice assistant embodied in a physical device, such as a smart speaker, and a conversational agent (CA) as an umbrella term for any agent that can engage in written or spoken communication with the user.

The main contributions of this paper are the following:

- Throughout our design process, we shed light on older users' needs and expectations about physical e-coaches for improving their well-being, with a particular focus on vocal assistants. Our contribution will explore three important axes for older adults: physical design (PD), interaction design (ID), and design for privacy and control (DPC). These insights are summarized at the end of this paper as guidelines for designing more accessible and enjoyable embodied vocal assistants.
- We share with the community open source material and guidelines for building a tangible coach, a physical, hardware and software design of a vocal assistant with tangible properties, highlighting the still open technical challenges and open research paths for improving such device.

2. Related work

2.1. Conversational agents for coaching

Several studies have used the e-coach as a tool to ensure, maintain and improve the well-being of users. For example, Project GrowMeUp (Georgiadis et al., 2016) and Robadom (Wu et al., 2013) provide integrated services to the elderly through robotic approaches. Its main goal is to encourage older people to stay active longer. Bickmore et al. (2005) presented pioneering work exploring the acceptance by older people of a conversational agent promoting physical exercise. However, this study involved the use of an agent running in a PC, where users could interact with it through a keyboard and touchscreen. Despite the significant limitations in terms of interaction, the results have already shown interesting potential in the application of such technology to help older people improve their health. After 8 years, the article presenting the results of randomized controlled trials of this conversational agent was published Bickmore et al. (2013). The results demonstrated that this technology was able to get sedentary seniors to adhere to the initial intensive 2-month phase of the intervention, failing over the following 10 months of maintenance. As highlighted by El Kamali et al. (2020), conversational agents have been adopted increasingly in the last years for coaching older adults. In recent years, the most common form of interaction modality for chatbots are text (i.e., chatbots available in messaging apps such as Facebook Messenger and Telegram) and voice (for example, Apple Siri, Google Assistant, Microsoft Cortana). These conversational agents are becoming increasingly popular in commercial use (National Public Media, 2022).

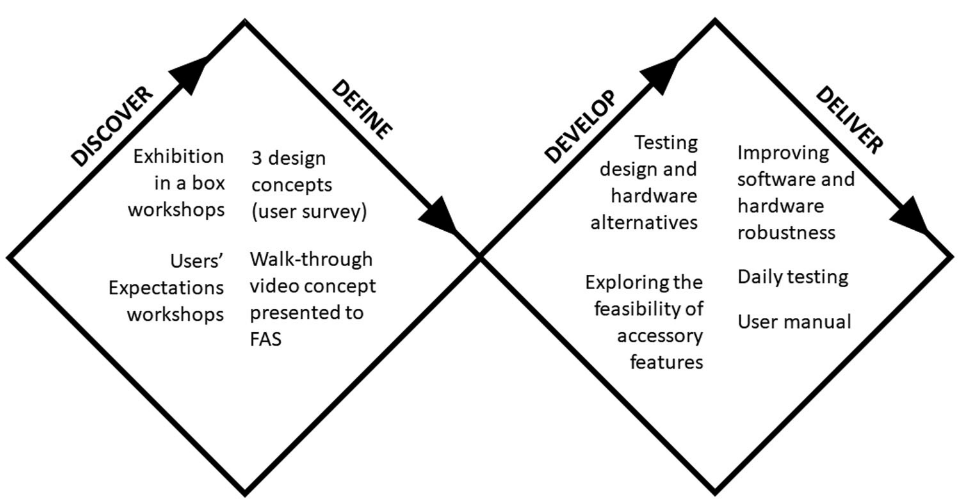

Figure 1. Double Diamond Process and activities carried out.

For this reason, an emerging literature has begun to study the use of these chatbots and has shown that current commercial chatbots and their corresponding smart speakers have a number of limitations in terms of their design.

2.2. Design of conversational agents for older adults

Designing a conversational agent for older adults requires considering several different aspects. Ter Stal et al. (2020) showed the importance of carefully designing the appearance of the agent in establishing user engagement in eHealth applications. Preferences and perception of the agent appearance may also change according to location of the interaction. For instance, Pradhan et al. (2019) indicated that participants move their perception of the CA easily between a "humanlike" or "object-like" agent. This movement is determined by parameters unique to the voice assistant (such as the nature of the conversation) or the user (desire for social engagement or connection), as well as the location and time of interaction. They discovered that the physical position of the gadget and its integration into the home environment are important. Alexa may look human-like while in the same room but not when far away, and the device is more likely to be classified as "human-like" when it facilitates discussions that give participants the impression that there is a presence "about" the house. Instances of companionship through the embodied voice assistant surfaced in their analysis, indicating the technology's potential for facilitating social relationships.

Testing existing commercial conversational agent may be useful to derive guidelines, strategies, implications and recommendations for conversational agents to meet the needs of older adults. For instance, Kim (2021) performed semi-structured interviews with 18 persons aged 74 and up who had never used a smart speaker before, in order to investigate the patterns of use, usability concerns, and viewpoints that older adults have when using a voice assistant (embodied in the Amazon Echo) for the first time. Their findings indicated that older adults had a positive overall first impression of smart speaker-based voice assistants, with healthcare-related questions and music streaming being the top two topics discussed during the first interaction. However, the majority of the subsequent reactions were negative due to the difficulty in constructing a structured sentence for a command; misconceptions about how a voice assistant works; and concerns about privacy, security, and financial burdens. Kim and Choudhury (2021) deployed Google Home devices in the homes of twelve people aged 65 or older up to sixteen weeks. They explained the necessity of enriching user experiences through conversational capabilities. The findings show that the benefits the participants perceived have gradually changed from enjoying simplicity and convenience of operation in the early stages of the study to not worrying about making mistakes and building digital companionship as they became more accustomed to using it. However, they suggested to add conscious responses that were sufficient to compensate for the lack of technology accuracy such as acknowledging the error or the source of information. They finally suggested to support a learning phase at the beginning. Porcheron et al. (2018) collected and studied audio data from month-long deployments of the Amazon Echo in participants' homes, informed by ethnomethodology and conversation analysis. They discovered implications for voice user interface accountability and embedding into conversational settings like family dinners where various simultaneous activities are being achieved. Cultural aspects may emerge when testing commercial devices with diverse groups. Harrington et al. (2022), testing Google Home with 30 black older adults, noticed how participants had to adapt their language to the device capability, highlighting that vocal assistants should be able to understand different languages. Interesting findings may be found when co-designing, instead of merely testing, conversational agents. Kramer et al. (2021) describe a co-design process with older adults that informs both the content to be non-judgmental, warm and friendly and the role of an CA to match the topic. Many factors influencing the design of conversational agents for older adults have still to be explored. Indeed, Sin et al. (2022), comparing popular discourses in mass media with the research findings in academic literature on the design of vocal assistants for older adults, showed that many topics discussed by media lacked of scientific echo in literature. For example, while media suggested that CA design should account for privacy control and for life on-the-go designs instead of stationary designs, such evidence was not reported in the papers analysed by the authors.

In the next sub-sections, we will analyse more in depth the evidence found in literature concerning the physical design, interaction design, and design for privacy and control. In Section 4, we will then compare the results obtained thanks to our co-design process with previous evidence in these three axes.

2.2.1. Physical design (PD)

In terms of physical design, users often may have a limited understanding of the physical form of the coach. In the case of a chatbot, this can be limited to the chatbot's avatar, and in the case of smart speakers, only to gender and pitch of voice. Having a fully embodied conversational agent may help the user increase confidence in the agent and build a stronger emotional connection. For example, the physical enclosure used for smart speakers plays a role in the interaction with users. Cho et al. (2019) showed that thanks to the physical form of conversational agents, users experience a greater sense of presence and are more likely to feel co-located with the intelligent agent that it embodies. These embodied conversational agents are seen as independent objects that are alive and affect the user's expectations, perception and behaviors. The conversational agent embodiment is important in defining user expectations in terms of conversational capabilities. While users would be happy to chat and interact with human-like agents, technology is very often not ready to meet these expectations. In this case, when the embodiment of the agent is too close to humans, but its behavior or form is not close enough to be perceived by the user as such, the user often develops a feeling of rejection, also known as the "Uncanny Valley" phenomenon (Mori, 1970). Therefore, it can be argued that the physical design of an e-coach could help building confidence, understanding the e-coach' goal toward the user, and, ultimately, improving the learning curve of the e-coaching system (Price & Rogers, 2004). This could be achieved enabling users to literally grab data with their hands, unifying representation and control. In fact, with the recent advent of 3 D printing, academics are actively examining the possibilities that result from the materialization of digital information. For example, Wang et al. (2009) combined speech recognition technology with a tangible learning companion/robot to improve English learning conversation for beginners. The results showed that tangibility increased enjoyment and could promote confidence in less proficient learners. Stusak et al. (2014) studied the impact of receiving physical artifacts as a reward for running. The 3D printed physical sculptures that represented running data were very helpful in motivating participants to run and stimulating self-reflection in their 3-week study of 14 people.

2.2.2. Interaction design (ID)

Besides tangible interactions with e-coaches' physical embodiments, conversational interactions have also an important role on the user experience of e-coaches. The literature has also provided suggestions for improving the interaction design of voice assistants. For instance, Sciuto et al. (2018) presented two complementary studies investigating the experience of households living with a conversational agent over an extended period of time. They gathered the history logs of 75 Alexa participants and quantitatively analyzed over 278,000 commands. They performed seven in-home, contextual interviews of Alexa owners focusing on how their household interacts with Alexa. Their findings give a first glimpse of how households integrate Alexa into their lives and they discussed some opportunities to improve the interaction design of these devices. Alexa still lacks the possibility of understanding the user's routines, of traveling with the user, and of locating itself at home in order to change the type of interaction (e.g., being proactive in the living room). Cho et al. (2019) aimed to understand how and why CAs lose their presence in users' everyday lives. For their study, they used the commercial device Alexa, which has a physical form, and analyzed the experience of using it at different stages. Indeed, they showed that users go through a number of stages, including pre-adoption, adoption, adaptation, stagnation, and acceptance, and that each stage is characterized by particular perceptions and usage patterns. They also eventually gave design implications such as having an agent actively approach the user before it is forgotten, by sending him/her suggestions, greeting him/her and asking him/her daily questions and also by increasing the function dependency of the agent toward the user. In fact, proactivity is a characteristic that has been shown in several studies to have a positive impact on older adults: for instance, a study has revealed that proactivity can have an effect on the adoption of such technology by older adults (Ring et al., 2013) because the agent's initiative can better support older users to combat loneliness. Chattaraman et al. (2019) explored that the effectiveness of conversational agents for older adults depends on the style of interaction: when the style of interaction is proactive, it allows higher cognitive and functional results for the users, i.e. older people with poor internet skills. Furthermore, conversational agent functional capability has also been shown to have a positive impact on the interaction design. A study that went in that direction can be found in (Martin-Hammond et al. (2019), where the authors qualitatively explored the beliefs and mental models of the elderly with reference to the use of conversational agents. Their results showed that participants viewed these intelligent assistants as" potentially useful for providing recommendations, facilitating collaboration between themselves and other caregivers, and for critical illness alerts." Melenhorst et al. (2006) claim that adults will only invest time and resources in adopting new technology if they can see a clear benefit for themselves. To sum up, most work focused on identifying that proactivity and functional capabilities are important coach interaction designs. However, besides the content of conversation and the behaviour of the agent, we believe that there is a lack of knowledge about older adults' needs and expectations about the physical interaction with physically embodied vocal assistants.

2.2.3. Design for privacy and control (DPC)

As e-coaches and conversational agents are entering our everyday lives, privacy and security concerns arise, especially when the user's health is involved. In fact, one of the main issues with privacy is that nowadays, many technologies even collect data without user intervention, which makes the user oblivious to the situation (Dubois et al., 2020). Thus, as a greater proportion of seniors go online (Anderson & Perrin, 2017), it is necessary to understand the attitudes and concerns of this group in order to reduce anxiety and ensure better confidentiality. Zeissig et al. (2017) collected data from a survey to assess online privacy concerns for seniors. Their research found that older users differ significantly in their awareness of privacy concerns and protect their data more actively than younger users. Therefore, it was also reported that older adults have privacy concerns and privacy self-efficacy that motivated them to take data protection behaviors. Caine et al. (2012) evaluated older adults' perceptions of privacy and their tendencies to engage in privacy enhancing behaviors by comparing three conditions: a mobile robot monitoring their home to a stationary robot monitoring home and to a wall-mounted camera. They also identified the effect of the physical design, whether embodied and mobile devices make a difference on privacy concerns, privacy awareness and their performance of privacy enhancing behavior from the perspective of older adult users. Their results showed that in both cases, being monitored by cameras or embodied robots created more privacy concerns but the embodied robots led the participants to perform less privacy behavior. Some studies suggested that robots may provide challenges to user privacy because of their embodiment, mobility, and novelty for users (Cho et al., 2019; Luria et al., 2020), while others believe in the opposite (Tonkin et al., 2017). Most of the research was focusing on investigating the characteristics that can play a role in increasing or reducing privacy such as perceptions of anthropomorphism (Ischen et al., 2019), or the capability of an agent to share responses to its own company (Sannon et al., 2020). Some research focused on building an interface for mobile applications to manage privacy (Ataei et al., 2018), or even via a chatbot that manages privacy access (Harkous et al., 2016). For instance, a user can ask the chatbot to restrict some users to see his profile. Another interesting study on data access control in vocal assistants suggested to map out areas where a vocal assistant can and should leverage their access to personal and interpersonal data and areas where they must operate more carefully (Luria et al., 2020). They designed storyboards to explore inchoate social mores around agent actions within a home, including issues of proactivity, interpersonal conflict, and agent prevarication. Besides these previous interesting studies, we could not find previous research that focused on design for privacy and data control for an embodied conversational agent for older adults, as it was also confirmed by the review of Sin et al. (2022).

In summary, little research has combined e-coach design guidelines with proper physical design for older adults. In this frame, few works focused on their use for the promotion of personal well-being, while providing maximum control on their personal data. While research projects mainly focused on costly assistance robots for fragile older adults, studies on commercial smart speakers often highlighted the lack or perceived benefits for older adults. In particular, no previous work was found on design for life on the go or on designs for controlling data privacy in voice assistants. This was confirmed by Sin et al. (2022).Yet, designing a voice based user interface that supports a life-on-the-go life, as opposed to a limited life at home and privacy concerns are important factors for older adults adopting a virtual user interface. To close this gap, the NESTORE project attempted to understand the needs and expectations of healthy older adults about a physical companion that could provide recommendations towards a healthier lifestyle. Plus, it attempts to create design solutions based on older adults' needs. The following section describes the co-design process carried out during the 3.5 years project, shedding light on requirements and features for such devices.

3. Tangible coach co-design

To understand older adults and their needs, to explore creative and innovative design solutions for a tangible coach and to deliver these solutions to the target population, we adopted the UK Design Council's Double Diamond process. The Double Diamond model (Figure 1) that shaped the study design, included two initial phases of exploratory divergent thinking; asking questions, conceptualizing and discovering possibilities rather than offering solutions, then converging findings to define user needs. The third and fourth phase, informed by phase one and two, explored and developed the specific requirements of the product that then converged though the delivery of tangible solutions.

In the next subsections, we present the different phases of each step that we conducted through co-design activities with potential users by focusing on the goal and methods used in each phase, and the results and insights we established after each phase. In each phase where we conduct our research with participants, a consent form has been signed by each participant to conform with ethical protocols.

3.1. Phase 1: Discover

3.1.1. Goals and methodology

We held five workshop activities to build understanding of older user attitudes to technology, identify factors that promote and inhibit its use, explore potential contexts where technologies would be used and elicit older adults' design requirements for creating a virtual coach. Figure 2 shows the flow of these workshops.

In the first part of the process, we carried out a series of workshops (W1). The first part of the series were workshops (W1a) that utilized the "exhibition in the box" methodology (Chamberlain & Craig, 2013), which uses objects and artefacts as methods to stimulate and scaffold thinking, offering valuable vehicles through which the complexities of lives can be understood. 'Exhibition in a box' provides a tangible conduit through which societal assumptions can be made visible, explored and challenged. The goal was to identify users' desires and needs to support well-being through technological devices. These workshops (W1a), that involved 80 target users, were transcribed and the data was analysed in partnership with 10 older people from the UK. Thematic analysis identified the following priorities in relation to the characteristics and functions that potential end users considered to be important. The emergent themes were: empowerment, fits my life, connectedness, robustness, freedom, engaging, ergonomics, reliability, cost, observation, security, keeping active, infrastructure, being social. As follow-up activity to exhibition in a box workshops, further workshops (W1b) were organized in UK (10 people), in Italy (4 people), Spain (9 people) and the Netherlands (17 people) to elicit user needs in order to build a virtual coach for older adults and to to identify if the themes that emerged through the exhibition in a box workshop (W1a) were more broadly relevant, if any were missing and if local history, culture and environments might influence them. Using this methodology, we have explored the user needs and defined the general requirements for the NESTORE e-coach and its technological ecosystem, including the tangible coach, discussed in this article. The co-design of other aspects of the NESTORE e-coach and in particular of the recommended activities suggested by the e-coach are beyond the scope of this article, however can be found in a previous publication (Angelini et al., 2022).

The second workshop (W2) aimed to build an understanding of design features of current digital interfaces that participants find helpful or unhelpful. A digital probe

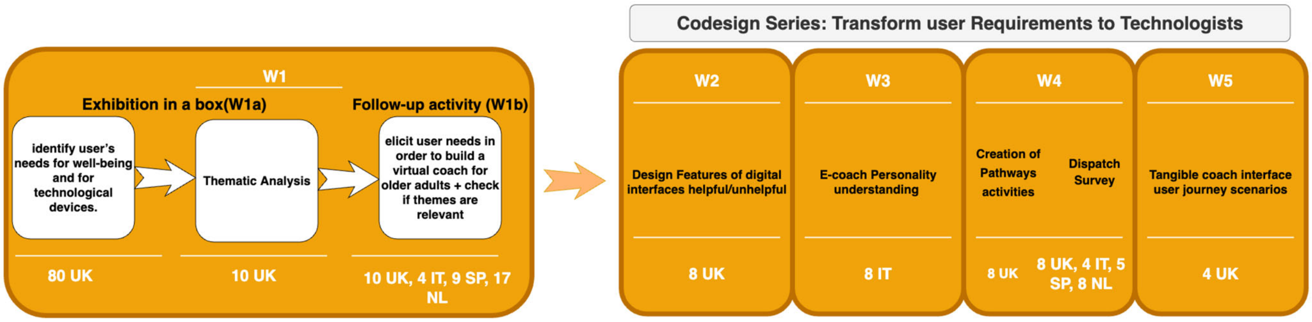

Figure 2. Workshops.

Figure 3. Digital probes.

methodology was used (Koch & Maaß, 2018). Participants were provided with a "probes pack": a disposable film camera, a tick/cross counter disc, and an instruction booklet (Figure 3). Each participant was invited to place counters with either a tick or a cross by particular technology interfaces they found particularly helpful or unhelpful and record with the camera. It was stressed that participants should comment on the interface of technology rather than the activity associated with the technology.

Two further workshops (W3) in Spain (nine people) and Italy (eight people) provided interesting contributions and suggestions on the appearance, personality and modalities of interaction of the tangible coach. In particular the personality of the e-coach was discussed through a bi-directional model exploring possible coach behaviours (Figure 4).

The aim of the fourth workshop (W4) was to map out the potential pathways the system could provide to guide how a user might navigate the system. Participants were presented with a series of cards (Figure 5) as examples of scenarios of use based on the user's profile, their needs/requirements and activities.

The fifth workshop (W5) aimed to provide insights into the interface design of the tangible coach through co-design activity. Participants were asked to map a day in the life. The day would cover a 24 h period to include things that might take place in preparation for the next day and during the night. Participants were encouraged to present an imaginary day based on activities that might in reality happen over a longer period of time. Participants included everyday activities (e.g., washing/grooming, eating, cooking, gardening, cleaning), social activities (visiting friends and family) and interests (hobbies). These were mapped on a template created by the design team. Participants worked in pairs and then shared their daily "stories." Participants were then asked to identify how NESTORE might support these different activities. Materials (e.g. pens, paper, cardboard) were provided to enable the participants to create tangible low-fi prototypes to enact "their day." Consideration was given to the type of activity (e.g., health related, social, work etc), to whether the activity was a social or private activity, to where the activity might take place (e.g., at home, on holiday, "on the move" etc). The workshop reflected on the different forms of interfaces that emerged from the co-design activity and on the features that the NESTORE system should consider in its design (Figure 6)

3.1.2. Results and discussion

The user requirements elicited from these workshops are summarized below. Requirements are associated with three research axes of Physical Design (PD), Interaction Design (ID) and Design for Privacy and Control (DPC) and also showed from which workshops (W#) we established these findings. If all workshops highlighted a certain requirement then no W# is shown next to it. Figure 7 shows the requirements found linked to the themes they fall in.

- R1: Intelligible/controllable system(PD, DPC,W1,W5)
- R2: Reliability and Robustness (PD, ID, W1)
- R3: Easy Charging (PD, W3)
- R4: Affordability (PD, W1)
- R5: Personalisation (ID, PD, W3, W5)

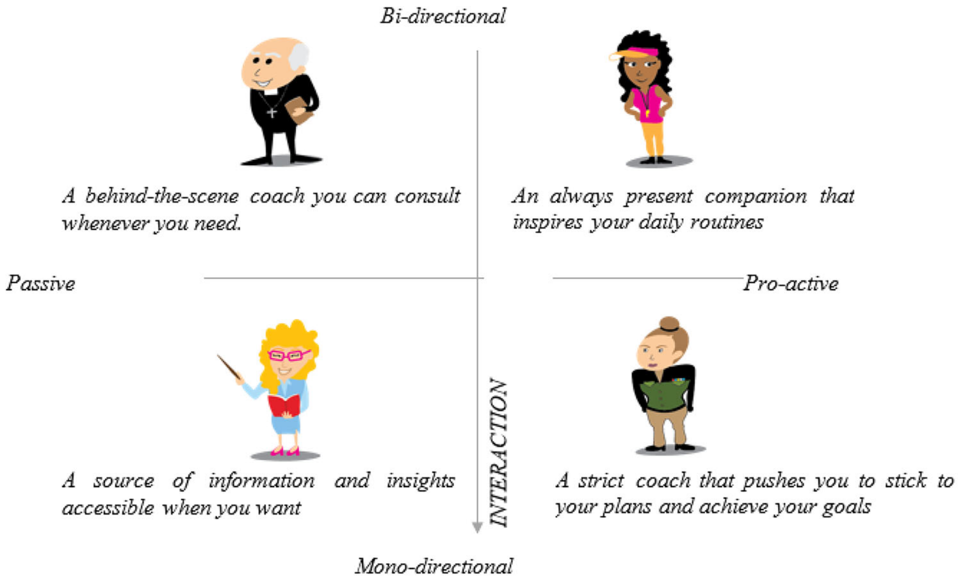

Figure 4. Personalities of the e-coach W3.

Figure 5. Pathways selection workshop W4.

- R6: Very Intuitive and Easy Systems (ID)
- R7: Easy Authentication (ID, PDC, W1)
- R8: Voice Modality (ID):
- R9: Transportable (PD,W1,W5)
- R10: Appearence (PD, W5)
- R11: Context-aware (ID, W5)
- R12: Unobtrusive (DPC,W5)
- R13: Friendly(ID,W3)
- R14: Accessible to all (ID, PD, DPC, W1)

We summarise here the most insightful comments on these requirements.

Feelings from participants such as “the user to decide when to disconnect the system”, “empower the user”, “feel in control” (R1) were found during (W1). Personalisation (R5) is wished from the system-side, but also from the user-side: participants from the UK felt that they should be in control, telling the coach what they want and what to do. Moreover, the attitude of the coach might be different for different activities (W5), which is also aligned with the comments from participants in Spain (W3) starting from the coach attitude and personality, the frequency of interaction, and notifications (R1 and R5). Participants from the UK commented that if they are going to talk to a coach, it would be desirable to have a sensory prompt or a visual point of reference for the interaction (R6), such as lights (W3), with a single sign in from one place (R7, W1). Voice (R8) was preferred by far to typing, in all workshop sites. This is linked to a lack of dexterity for typing on (small) touchscreens but also to some visual impairments. It highlights the need to design interfaces suitable for varied levels of digital literacy. For instance, participants from the UK indicated there was an opportunity for voice activation to overcome slow typing. Participants from Spain consider that the textual input is not very comfortable. In Italy, when older adults were asked how they would like to interact with the tangible object, most users mentioned voice as their preferred method. Voice is regarded as the most natural and easy way to interact with the tangible object, considering many of them have problems with eyesight. Additionally, since the tangible object would be used in the home environment, users mentioned preferring a hands-free type of interaction that would allow them to keep doing their activities while interacting with the coach. Participants from Spain also prefer the use of less intrusive devices (R12) that have audio interaction, especially if they are with friends or doing a group activity (W3). A good looking product (R10) is a requirement highlighted by many workshop participants.

Figure 6. Fifth workshop user journey with tangible coach.

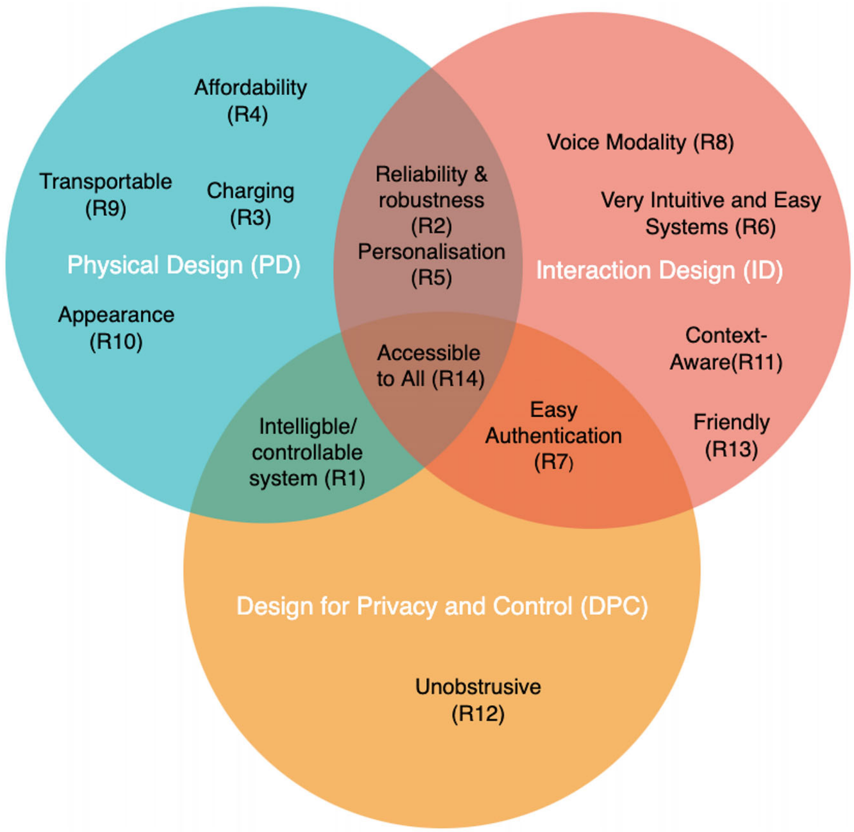

Figure 7. Clustered requirements.

It should be easily blended into the home environment and could be seen as a decorative piece of technology that users might want to show off. Participants might also appreciate a tech look. Concerning the coach embodiment, there was not total agreement on what the tangible coach should resemble. Some users in Spain would prefer a human-looking character, to which they would like to speak as in a videoconference. At the same time, they depicted the object as a "small robot", a compact object that they can put on a table and move in different rooms. In the UK and Italy, participants identified the tangible coach as a geometrical object (W5).

3.2. Phase 2: Define

3.2.1. Goals and methodology

In this step of the process, we have defined some features of the tangible coach that we want to evaluate with users. We have then defined three design concepts: the familiar coach, the growing coach and the compact coach where we highlighted these features and the requirements of Section 3.1.2.

With the help of convergent thinking(double diamond process), we started to define what a physical companion might look like for coaching older adults. Indeed, the requirements and ideas produced by users during the workshops were used to design three different concepts of the tangible coach. Each concept focused on particular characteristics that the tangible coach could boast of. The three concepts were presented to users in order to gather feedback on each feature of each solution, highlighting the points they particularly liked about each design concept and the points they could improve. As a second step, by mixing design and technology, we began to materialize the concepts into believable products that would mimic the features to be tested in a home setting and explore the usage stories more.

We used storytelling as a way to produce a movie format presentation so that we could read them to potential users as a talking point for further development. This video was discussed with the NESTORE Project Forum of Advisory Stakeholders (including representatives of health professionals and associations of older people) and with a smaller group of "experts by experience".

In part (1), we present the features mapped to some of the requirements and in part (2), we present the three design concept descriptions that we showed to users.

1. Features Definition
   In this section, we present the features that were integrated in the three design concepts for each research axis. When relevant, we mapped these features to the requirements (R #) of Section 3.1.2
   - Physical Design
     - Voice Interaction (F1), as discussed in (R8)
     - Interaction with Lights (F2), to provide unobtrusive notifications and information (R12)
     - Interaction with sounds (F3), as an alternative notification modality (R12)
     - Interaction with Shape-changing Interfaces (F4), to convey unobtrusive(R12) notification and to explore the possibility to have a growing interface for the coach embodiment, physically rewarding achievements in daily life as suggested by Stusak et al. (2014)
     - Touch sensitivity (F5), as interaction modality to bond with the coach (R13)
     - Tangible Controls (F6): The attitude of the coach can be changed through tangible manipulation of the design (R1).
     - Animated Facial Expressions (F7), to provide life-like features and a more friendly behaviour.
     - Text/video interaction through other devices (F8). Most smart speakers rely on companion apps to provide additional (often more visual) information that couldn't be processed easily by voice. We integrated this feature with different nuances in each of the three designs.
   - Interaction Design
     - Transportability—Compactness—Easy to plug (F9): In order to answer our requirements (Section 3.2.1), one of our designs is portable and can be moved around the house (R9).
     - Appearance (F10) The appearance of the three design concepts was conceived, on purpose, to explore different levels of animism, to develop a life-like object (Schmitz, 2010) that can engage in long-term interaction without falling into the uncanny valley (Mori, 1970). The familiar coach explored zoomorphism as suggested by Schmitz (2010), the growing coach on a more subtle biomorphic interface emulating a plant, while the compact coach on a rather geometrical object that would rely all the life-like behaviour on the talkative features, therefore in a similar fashion to commercial smart speakers. (R10, R11)
   - Design for Privacy and Data Control
     - Personalized, Context-aware (F11): Personalisation is a key feature of virtual coaches, but it's also a feature that is often appreciated on physical products(R5). Each design concept in the next section explored a different kind of personalisation: The familiar coach has context aware personalisation (R11), whereas the growing coach has lights personalised based on its coaching intervention, and finally the compact coach is based on interaction personalisation (wakeword).
     - Tangible Controls (F6) as described before
2. Design Concepts Definition
   In this section, we describe the three different designs derived from user needs and previous results found during the literature review. They were obtained through a joint effort of all the authors, who reiterated the ideas until they agreed on the following three conceptions. The following sections report the stories and sketches provided to the users for each concept (feature numbers in parenthesis were not provided to participants). The aim of creating these three abovementioned prototypes is to iterate them among participants in order to find the best features in each design and combine them among each other. This will lead to new suggestions to create the final prototype of the e-coach. This method is identified as a way to help users to imagine their ideal e-coach. We sent these three final design concept descriptions to 31 participants (7 in UK, 6 in Spain, 10 in Netherlands and 8 in Italy) and asked them to evaluate and give positive and negative feedback about the features of each concept in order to identify likes and dislikes on a granular level. Participants also had the possibility to suggest improvements, eventually borrowing features from the other concepts. At the end, we asked participants to rank the design prototypes from the most appreciated to to the least appreciated one. We collected the qualitative data and the rating result, analysed them and came up with the final design of the tangible coach.
   - Familiar Coach:

     The appearance of the Familiar Coach(FC) (Figure 8) resembles a knowledgeable bird and has design characteristics that are understated but recognisable (R10, F10). The Familiar Coach is portable and can be moved around the house (R9). It has nests set up in different rooms and can be easily moved from one nest to another (R3, F9). It is also aware of its location and changes its coaching attitude to suit that location (R11), e.g., calm in the bedroom, reduced notifications (R12, F11). The coaching attitude in locations can be personalised through the NESTORE app (R5). The Familiar Coach uses voice interaction (R6, R8, F1) as its primary mode of communication supported by

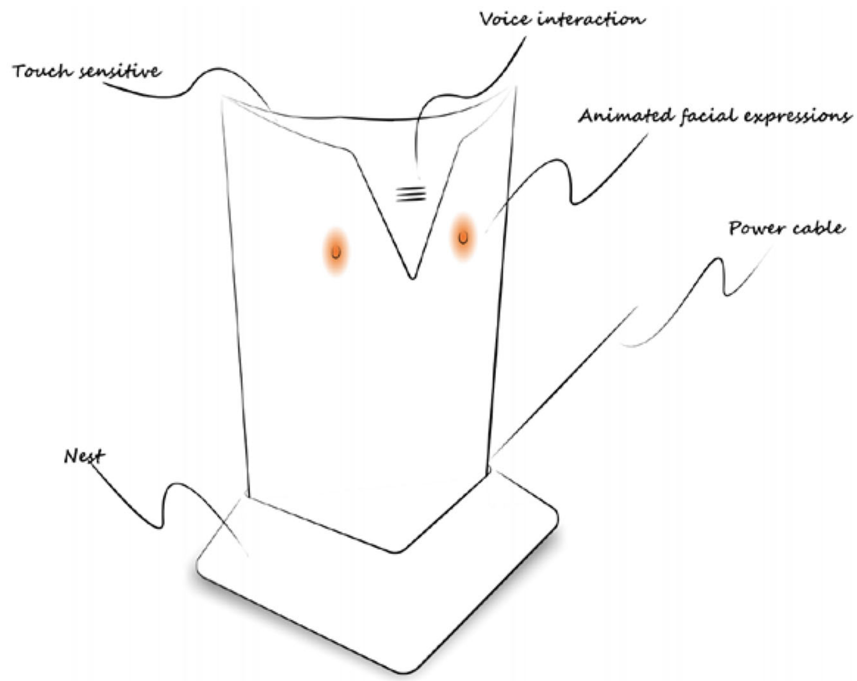

Figure 8. Familiar coach design prototypes.

animated facial expressions on the exterior surfaces of the design as an indication of progress, notifications and a need that it would like to communicate with you (F7). Areas of the Familiar Coach are touch-sensitive. These areas are used to invoke communication, so the user is in control (R1, R12, F5). Users can have a conversation on a new coaching activity or to ask for progress/help on a current activity. If there is a lot of information to be passed on by the Familiar Coach, it will send useful tips and suggestions via text or email (F8).

- Growing Coach:

  The Growing Coach (GC) (Figure 9) would resemble a collection of elegant stems (R10, F10). At the top of the stems are notification sails that unfurl to show progress and achievements throughout coaching activities (R12). The group of notification sails will show overall progress with slow movement and other sails will move quickly to show task achievement (F4). The Growing Coach is designed to be in a fixed social space in the house (R3) and used as a touchpoint throughout the day (R7). It always shows progress notifications, but will wait for your presence to show achievements (R11). It will start by getting your attention using a soft sound before starting the notification sail movement (F3). At the centre of the notification sail there is a light source which illuminates the sail with different colours to indicate the NESTORE coaching activity (F2, F11). This can be personalised through the NESTORE app to aid recognition (R5). With a touch gesture on the stem, the Growing Coach would repeat the last achievement notification (R6, R8, F5). Users can communicate with the Growing Coach via the NESTORE app and chatbot (F8, R14).

- Compact Coach:

  The compact vocal coach (CC) (Figure 10) is of stylish, solid geometric design that blends in with your environment (R10, F10). Elements can be personalised (R5), such as the ability of the outer surface of the product and the wake word that users use to start a conversation can be changed (R8, F1, F11). The compact coach uses light communication to gain your attention when it wants to give you an update on the user's progress (R12). The compact vocal coach would have a light/group of lights at the front as a reference point, so that users know that they are speaking in the correct direction and that the coach is listening (R6, F2). The attitude of the coach can be changed through tangible manipulation of the design and will give a visual reference on the outer surface of the compact coach to show the current coaching attitude (R1). Like a friend (R13), the compact vocal coach will send the user a text from time to time, with some tips or suggestions that are relevant to your activities. This might be a link to a YouTube video on cooking(F8). The Compact Vocal Coach contains a battery, which means (if needed) it could be moved around the house for periods of time (R9) without having to find a plug (R3).

3.2.2. Results and discussion

Participants were asked to comment on the most appreciated and the negative elements of each design concept, providing suggestions for improvement. We conducted a quantitative

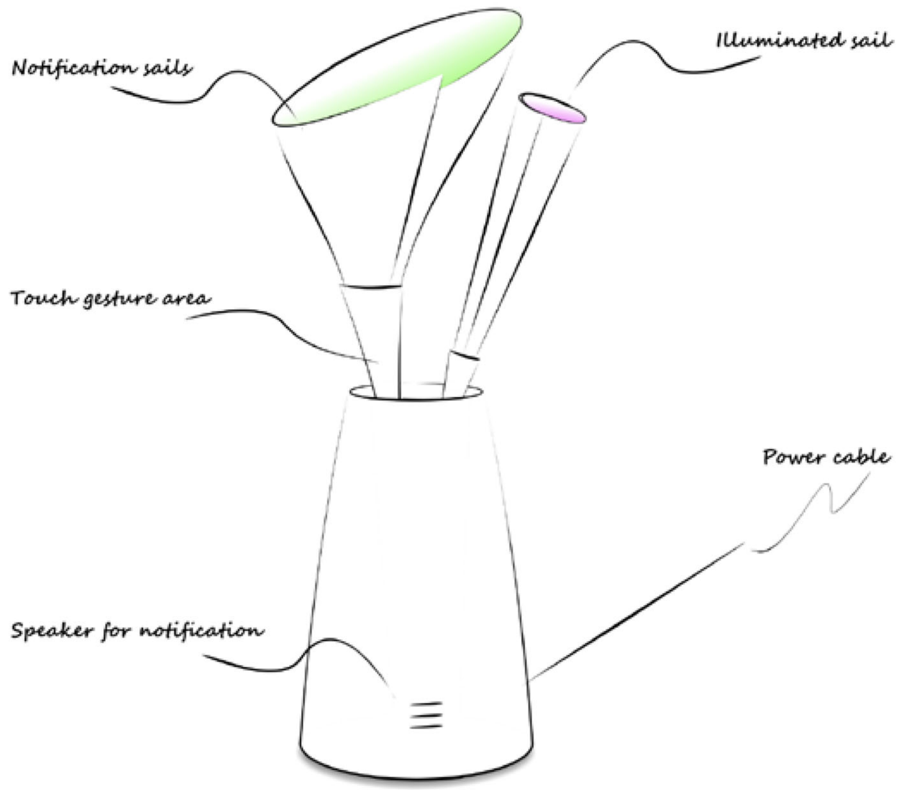

Figure 9. Growing coach design prototypes.

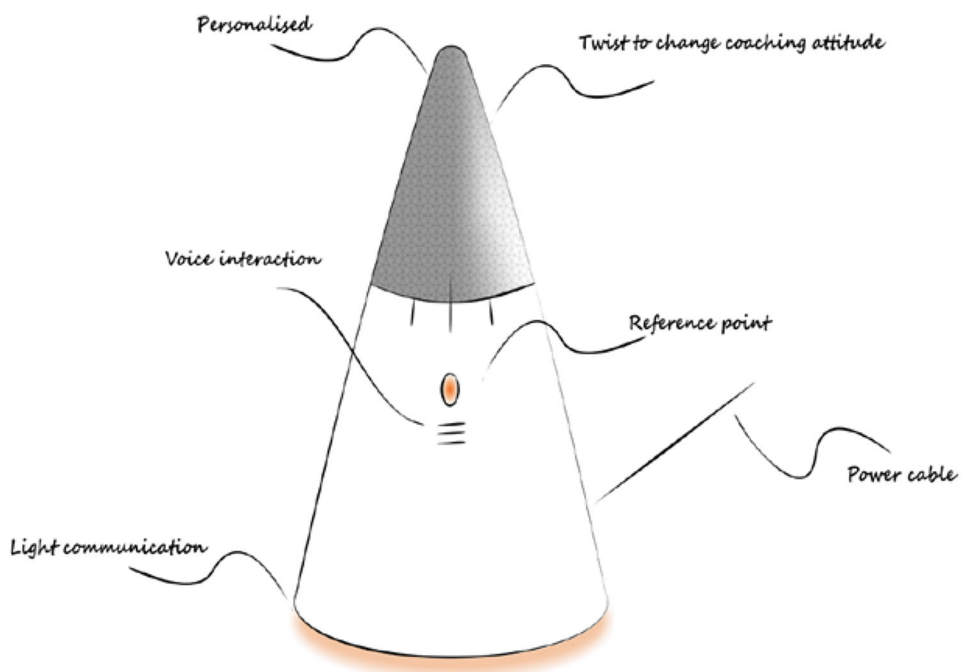

Figure 10. Compact coach design prototypes.

and qualitative analysis of the comments, associating them to each of the features presented in the three different designs. The goal of the survey was to get participants to choose the best features of each design, rather than just voting for the preferred concept. Of the 31 participants, 61.29% ranked the CC as their first choice, whereas only 35.48% ranked the FC and 3.22% ranked the GC as their first choice. 54.84% ranked the FC as their second choice whereas 25.81% and 19.35% ranked the CC and the GC as their second choice respectively. Finally, 67.74% ranked the GC as their last choice, whereas only 19.35% and 12.90% ranked the FC and the CC as their last choice respectively.

1. Rank Analysis

We used the Friedman test to compare the mean ranks between the different design concepts. We can see in the Friedman Test result the median value of each design concept. We can only know at this stage that there is a difference that lies between each design concept. There was a statistically significant difference in design concept preferences between users where

$$\chi^2(2) = 26.387, p < 0.001. \quad (1)$$

We ran post hoc tests to examine where the differences actually occur. We used the Wilcoxon signed-rank tests on the different combinations of related design concepts. Table 4 shows the test statistics. In fact, the growing coach ($M = 2.74$, $SD = 0.51$, $Md = 3$, $n = 31$) was less preferred compared to the familiar coach ($M = 1.74$, $SD = 0.63$, $Md = 2$, $n = 31$), $z = -4.13$, $p < 0.001$, with an effect size $r = 0.43$ and the compact coach ($M = 1.52$, $SD = 0.72$, $Md = 1$, $n = 31$) $z = -4.09$, p¡0.001 with an effect size $r = 0.42$. However, there is no significant difference between the familiar coach and the compact coach, $z = -.95$, $p = 0.34$, $r = 0.1$. One can only observe that the mean rank for the compact coach is the lowest, which makes it the most preferred coach, but the lack of statistically significant difference between the familiar and compact couch led us to focus on the qualitative responses of each design feature proposed in the three design concepts.

2. Features Analysis:

Table 1 summarises the number of positive, negative and" suggestions for change" comments for each design concepts across the four pilot sites toward each feature. Below we summarise the main outcome results of each feature.

- Voice Interaction (F1): Voice interaction was generally commented upon as a positive feature. Most comments appreciated the voice control for the familiar and compact coach, or suggested using voice control for the growing coach. accessibility was also stressed to be important with this feature in terms of the ability to turn the volume up for hearing concerns and adapting the language to the seniors' world. Negative comments were mostly about seniors not accepting to talk to "objects." A personalised voice activation was also suggested for the familiar coach in place of touch activation, or some kind of facial or voice recognition so the coach could operate differently if more than one person is using it, which makes it less seen as a computer.
- Interaction with Lights (F2): Progress notification or the sails or petals that indicate progress were appreciated for the growing coach. Concerns arise for the colour-blind people. customizing color is suggested. Confusion with meaning of colors was also a concern for seniors. 2 participants suggested for the familiar coach lights like the other designs.
- Interaction with sounds (F3): Concerns such as hearing problems were mentioned. Sounds were not very appealing to one participant, it was seen as the coach will have more control than the user with this sounds to attract attention. However, alerting its place of location in the familiar coach design or to gain senior's attention when needed with sounds or even with lights for the compact coach design were suggested by some participants.
- Interaction with Shape-changing Interfaces (F4): Out of 20 comments found on this feature, 12 criticized this feature. The main criticism was related to the robustness of the device and potentiel danger, understandability and explaination of the movements. As shape-changing interfaces go beyond existing mental models that people have about technological devices, this unfamiliarity was also translated into fear by some participants.
- Touch sensitivity (F5): Four participants have commented on the touch feature for the familiar coach as interesting for invoking communication. Three participants appreciated the feature of touching the growing coach to repeat the notification. Three participants have commented this feature as unnecessary, difficult, unimaginable, unlikeable. One participant has suggested touching the holder instead of the stems for the growing coach. It was not flexible for the participants.
- Tangible Controls (F6): Three participants have interest in manipulating an object to change and customize the coach. One participant needs a visual icon to the visual reference for the activity. Overall, tangible controls were interesting to seniors. In the

Table 1. Feature analysis: positive comments (+), negative comments (−), suggestions for change (⟺)

Features	Familiar coach	Growing coach	Compact coach
Voice	6+, 4−, 7 ⟺	2−, 1 ⟺	7+, 1−, 1 ⟺
Interaction with Lights	2 ⟺ , 1 +	9+, 3−, 2 ⟺	4+, 2 ⟺ , 1−
Interaction with sounds	1 ⟺	2+, 3−	1+
Interaction with Shape-changing Interfaces		6+, 11−, 3 ⟺	
Touch sensitivity	4+, 1 ⟺ , 3−	3+, 2− ,1 ⟺	
Tangible Controls	1 ⟺		4+, 1−, 1 ⟺
Animated Facial Expressions	8+, 6−, 1 ⟺		1 ⟺
Text/video interaction through other device	2+, 1−	1+	7+, 1 ⟺ , 1−
Transportability-Compactness-Easy to plug	11+, 5−, 5 ⟺ , 1 −/ ⟺	1+, 4−, 5 ⟺	15+, 1−, 5 ⟺
Appearance	7+, 4−, 3 ⟺	10+, 8−, 4 ⟺	14+, 2−, 1 ⟺ , 1(⟺ , −)
Personalized, Context-aware	7+, 3 ⟺ ,	2+, 1 ⟺	8+, 2 ⟺
Friendly	1−, 2 + 1 ⟺	1 ⟺	
Other stuffs	2 ⟺	1+. 2−	4 ⟺ , 1−

case of the familiar coach, it is also necessary to be able to switch the device off.
- Animated Facial Expressions (F7): Eight participants highly appreciate the animated facial expression in the familiar coach. Criticism was also high: 6 participants did not find usefulness. Six participants did not find usefulness in having facial expressions or they are not keen on the idea of a zoomorphic object.
- Text/video interaction through other devices (F8): 13 comments were found on this feature. Users appreciated the possibility to receive video or images, acknowledging it as a richer communication channel. Sceptical users criticized the difficulty to add more technology to the system.
- Transportability—Compactness—Easy to plug (F9): Transportability and the ability to move the device from one to another was found as a positive feature. One participant also was interested to know if it can be portable outside the house. A participant suggested giving the device a clamp that could be attached to the clothes to avoid losing the object. This feature should be aligned with the shape of the design. Most concerns were about if it's heavy, not easy to handle, and forgetting its place.
- Appearance (F10): 59 comments discussed the appearance of the three concepts, showing the importance of the look of the physical device that would be part of their interior decor. The familiar and growing coach received mixed feelings from positive to negative. A main concern was related also to the stability and robustness of the interface of the growing coach. The Compact Coach received the most positive comments. it was described as simple, with not too many bells and whistles, nice shape and quiet. Hence, the geometric appearance design is mostly recommended by users. Yet, some negative comments were referring to the shape, in particular, to the resemblance with air fresheners.
- Personalized, Context-aware (F11): 23 comments were found on this feature and all of them described appreciation to the possibility to personalize the coach's behaviour or appearance. Users appreciate the location aware feature and the "Knowledge of location with automatic setting of sound level etc., based on the environment." For the compact coach people appreciated "The ability to change the coaching attitude and personalise it." People also wished for the possibility to personalize the shape or color of the devices according to the surroundings and aesthetic sense or to change the voice of the device depending on preference.

3. Summary of results:
Although some features received mixed feelings, others, such as voice interaction with lights patterns reflecting the state of the tangible coach, ease of charging and portability, the notion of sending information to other personal devices, ease of control and personalisation are the key points emerged. These features were synthesised through further design iteration and shared with the user groups for further feedback. Based on the feedback from the expert by experienced users and from participants across the country partners (Spain, Italy), we developed the formal qualities and interactive features of the tangible coach while taking into consideration compatibility with the technology.

4. Contextualisation of design features and user scenarios
Based on the results, the physical design and the interaction design with the tangible coach were modelised into a scenario. We created a video with our final design that is identified as "the ideal prototype," interacting with a user. The two main objectives of the video were to visualise the type of interaction users would like to have with the tangible coach and a way to help us to start the development phase. Sharing the new prototype with our expert user group, we ask them to produce real use case scenarios from their experiences. Participants were individually asked to create a scenario in which they could use the system for a healthy lifestyle. Each participant then played a role in the activity, considering how the NESTORE system and its hardware might provide the physical point of contact and interface for that activity. In creating their us case scenarios, participants reiterated the need for the tangible coach to be portable or at least usable from or in different areas of the home, perhaps adapting as they move to different places. In addition, the mode of operation and the interface with the tangible coach had to adapt to different activities, environments and preferences such as voice and touch activation, audio and visual information. We chose one of the expert group's use case scenarios and produced a film that highlighted the interactions and design so that it could be shown to the larger project group and other groups of expert users we worked with (Figure 11). The film was used as a reflection tool which helped to highlight potential gaps in the process such as set-up, activities, and contexts of use. We also gave them prototypes from the early development tech package to use as a scaffolding for conversation that would help further development.

3.3. Phase 3: Develop

3.3.1. Goals and methodology

Once all the feedback was collected, we began to explore the different functionalities that users wanted with physical designs and hardware designs (Develop), until we converged on a final integrated design that could be produced in 70 samples and that needed to be robust enough to be tested for 14 weeks by 60 users across Europe (Deliver). To make our tangible coach meet all our requirements (from Section 3.1.2) and all our features (Section 3.2.2), a preliminary investigation of possible hardware solutions for the tangible coach has been conducted between the authors concerning each type of design. The compact coach design inspired the shape of the final prototype, but with the additional possibility to have different orientations of the device to unlock different functions

Figure 11. Scenario acting.

Figure 12. Hardware and software final prototype.

(F6). Through several iterations of 3 D printed prototype, we obtained an optimal design, that allowed blocking the selected hardware and was covered with a cozy and soft fabric to enhance the experience during manipulation and to create emotional attachment to the device (F10). Also on the hardware level, we tested several versions of Raspberry Pi boards (Zero, 2B, 3B, 3 A+) to find a good compromise between form factor, processing power, battery consumption and extensibility. We iterated also between different microphones technologies such as Re-Speaker 2 hat, Re-Speaker 4 hat and Matrix Voice (Figure 12). To enhance portability and ease of use, we investigated different possibilities for battery operation and for facilitating the charging process. For this purpose, we first considered the possibility of having a base for wireless charging. This also requires the integration of a Qi microUSB adapter in the tangible coach base surface. Although easier for the user, in our test we faced unreliability problems: if the device is misplaced on the charger or moves suddenly, the device may not charge. To overcome the burden of plugging a USB cable (microUSB for the Pi Juice hat), we explored the possibility of using a magnetic cable adapter.

3.3.2. Results and discussion

Based on the size, cost (R4), the capability of each hardware and to preserve the selected features (F1, F2, F3), we finally chose to use the Raspberry Pi 3 A+ (square smaller model of Raspberry PI) and the Matrix Voice, a 8-microphones, 18 LEDs hat for the Raspberry Pi. To manage the portability (F9) of the tangible coach, we added on the top of the raspberry pi a Pi Juice hat, which enables battery operation and recharging. For the magnetic cable solution, through a cable extender, the magnetic adapter is bonded (with hot glue) to the 3 D printed body, avoiding soldering the cable for minimizing unreliability risks (R2). This turned out to be the cheaper and more reliable solution for our users. Finally, we used a LIS3DH accelerometer to check the tangible coach positions: the sleep position (face-down) and the wake position (on the side or up-right. (F5, F6). If the front base is sat on the table, then the tangible coach is in the sleep position. If the user turns the tangible coach to the side, so that the user can see the front base of the tangible coach, then the tangible coach wakes up and waits to be triggered by its wakeword “NESTORE” (Figure 13b). We used snowboy

(a) Sleep position of the tangible coach

(b) Wake position of the tangible coach

(c) Conversation with the tangible coach (Thinking State)

(d) Charging the tangible coach

Figure 13. Different states. (a) Sleep position of the tangible coach. (b) Wake position of the tangible coach. (c) Conversation with the tangible coach (Thinking State). (d) Charging the tangible coach.

model (Snowboy (n.d.) which is a customisable hotword detection engine that is always listening (even when off-line) and compatible with Raspberry Pi. Figure 12 shows the hardware and software of the final tangible coach prototype.

Figure 14 shows the flow logic of the tangible use of the coach. Users need to change the position of the tangible coach from the sleep position to the awake position in order to talk to the virtual coach. The tangible follows a three-state model: listening state, thinking state and speaking state. Each state has an LED pattern to help the user recognize which state the coach is in. We ended up using a single color (NESTORE theme orange color) to avoid visual problems. The logic of the NESTORE CA is made up of three main components, RASA Core and NLU for intention recognition and Natural Language Understanding (NLU), a Node.js server to manage the conversation logic and an additional server for distribute messages to the different interfaces according to context information and user preferences. The choice of the Speech API as a voice chat agent is motivated by the availability of different languages and their performance to guarantee the user the best user experience. All coaching interventions are provided and orchestrated by the Decision Support System (Orte et al., 2018), which adapts the intervention according to the personal data collected by the system or provided by the user. The tangible coach works for one person in a household based on the authenticated user. Finally, there are two types of conversations: a short conversation where there is only a question-answer conversation and the tangible coach goes back to the wake word entity, waiting to be called again and a long conversation where the tangible coach and the user keeps going until the predefined scenario ends. Examples of questions were about the user's scheduled and recommended activities for the day, user's progress in the different domains: physical, nutritional, cognitive and social. Users can ask the coach for some personal information such as why its name is NESTORE, its main goal, role, job, how it works and what each color pattern means.

3.4. Phase 4: Deliver

3.4.1. Goals and methodology

Once the final design was validated (Figure 13a–d), the tangible coach was produced in 70 samples, the case being 3D printed in SLS and the hardware assembled manually. The coaches were individually tested to ensure they were fully functional. Longer term testing has been conducted to ensure the reliability of the hardware and software, continuously improving the software to fix bugs and to improve battery life. Due to the limited size of PiJuice's default battery (1800 mAh), the coach can last about 2 h unplugged and can engage in continuous conversations for 30 min. Based on the conversations designed for the NESTORE system, we found this to be satisfactory, with no need to install an uncertified battery. During testing, we identified some design flaws, the first related to the PiJuice's battery retention system, which came loose during transport or shocks, the second related to the rear screw compressing the back of

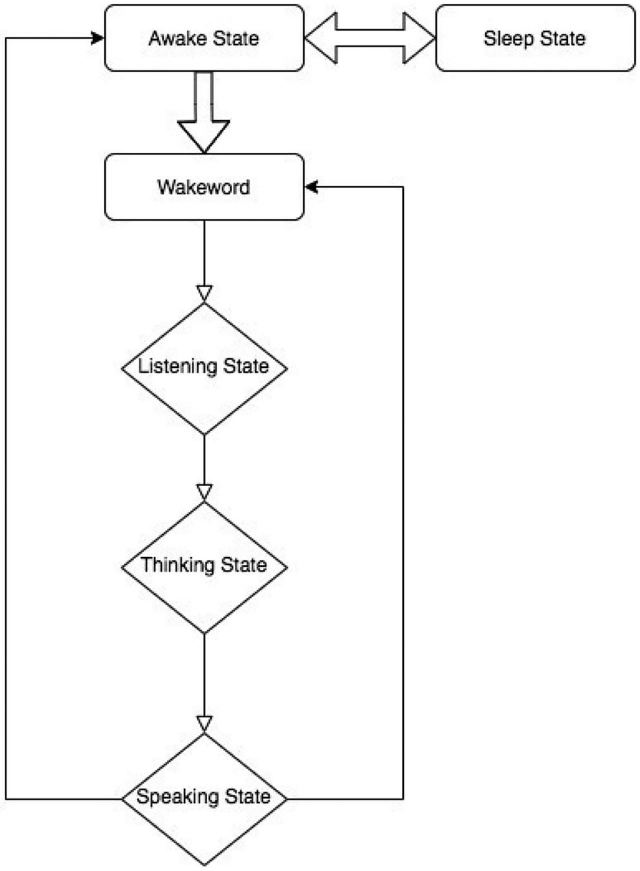

Figure 14. The logic flow of the tangible coach.

the speaker and reducing the ring volume. The two problems were corrected respectively by blocking the battery with bands and partially unscrewing the rear screw.

After the codesign, development and implementation phases, the NESTORE system and its components were tested by final users in their natural and living context. Among the objectives of the project, there was the interest to understand both the use and the impact of the e-coaching system in the daily life of the final users. Then, this testing phase focused on the chance of detecting possible criticalities related to functionality (i.e., the functioning of the system as a whole and its components in the home environment), and to usability, acceptability, and user experience (i.e., if the system is ready to be used by participants, and whether the system was suitable for their capabilities and expectations) of the multi-components, multi-channels, and multi-domains prototyped system. Dealing with the Tangible object, we refer to the most innovative component of the system: a conversational object through which the user and the NESTORE Coach can interact both physically (e.g., users can change the Tangible object's position and orientation, and thereby the NESTORE coach changes its behaviour) and vocally (e.g., there is a bidirectional conversation between user and object). Considering the target users, their characteristics and attributes, and the novelty proposed by the NESTORE system in terms of components, features, and possibilities of interaction, the testing phase is a crucial stage in the development and implementation of our proposal. At the beginning, the pilot testing phase was planned with at least 60 participants, equally divided between Italy, Spain, and the Netherlands; they had to test the NESTORE system for about 12 weeks independently. In addition, there was also a control group, to check the system impact and its use on people's health and well-being (seeing the difference between those who used the system and those who did not). The pilot had different phases; after the first phase of recruitment and evaluation of the users' characteristics (e.g., physical, cognitive, cultural on electronic devices, etc.), the NESTORE system and its components were presented and delivered. After the installation phase, with the support of the pilot sites' staff, the NESTORE system became effective, starting to learn information (either automatically or inserted and selected by the participants) to customise the e-coaching intervention in the different domains and based on the user's habits, preferences, and objectives. After these first two weeks of learning, the intervention period began, in which the system was supposed to propose activities and goals to be achieved. We designed three testing stages: at the beginning, in the middle and at the end of the pilot, to assess the experience and context of the system in use, the positive and negative aspects of using the system, and the needs and barriers to use identified by the participants. These moments were planned to be face-to-face in each country, with the pilot staff partners, adopting a standardised protocol of activities. The evaluation would take place by asking participants to carry out supervised activities, answer standardised questionnaires, and participate in focus groups. In this way, in each of the three evaluation stages, we would collect qualitative and semi-quantitative data on the use and acceptability for implementing the system. During the pilot, we organised three sessions (at the beginning, in the middle and at the end) of face-to-face discussions with the participants from the different countries to gather qualitative (through focus groups) and semi-quantitative (through surveys and tests with users) data, for the development and improvement of the system. Unfortunately, the outbreak of the Covid-19 pandemic forcibly modified the testing and evaluation activities of the NESTORE system, starting from the sample size (reduced from 60 to 24 voluntary participants) to the duration of the pilot (in some cases, it was not possible to do a 12-week pilot), up to the possibility of doing activities linked to the NESTORE intervention domains. This has partially limited the effective action of the e-coaching system. With these changed circumstances, the methods and tools for evaluating the functionality, usability, acceptability, and user experience were updated, selecting both qualitative and quantitative approaches that can be used through different channels, such as online forms, video conferencing systems, and instant messaging apps. The evaluation activities were reduced to a two-stages evaluation, at the beginning and at the end of the pilot. For the management of the activities, the staff of the pilot sites constantly monitored and supported the users, when possible, in presence or using video call and/or chat systems. Table 2 summarises and compares

Table 2. The table presents the activities planned for the pilot testing phases, comparing the methods and tools before and in the pandemic

BEFORE THE PANDEMIC	IN THE PANDEMIC
Usability and Acceptance: SUS and TAM onsite surveys to collect semi-quantitative data, according to usability standard tests.	Usability and Acceptance: SUS and TAM online surveys, to collect semi-quantitative data, according to usability standard tests.
	User Experience: online surveys like the UEQ, the Friendship scale and the Companionship scale, to collect semi-quantitative data and open comments, on the overall experience with the system and its components.
Direct observation and test with users: onsite tests to let the users familarise with the system and make use of the system, and to collect data about tasks and activities (at the installation, the mid-term and the end—different level of knowledge and skills).	Indirect observation and test with users: some tasks asked through online surveys, to check if users have familiarized with the system (at the installation and self-reported during the pilot duration)
Final Interviews and focus groups: onsite and face-to-face interviews, to collect the users' point of view and experience at the end of the pilot with presential focus groups	Final Interviews and focus groups: online interviews to collect the users' point of view and experience at the end of the pilot.

the different methods and tools planned at the beginning of the project and after the Covid-19 outbreak.

Concerning qualitative methods, structured and semi-structured interviews, and focus groups were used; these were mainly used in the final phases to have direct feedback of the user experience with the system. Concerning quantitative methods, standardised questionnaires were used, such as the SUS—System Usability Scale (Brooke, 1996), the TAM—Technology Acceptance Model (Davis, 1989; Davis et al., 1989; Venkatesh & Davis, 2000), the UEQ—User Experience Questionnaire (Laugwitz et al., 2008; Schrepp et al., 2017), the Companionship Scale (Lawson & Chesney, 2007; Luh et al., 2010), and the Friendship Scale (Hawthorne, 2006). These questionnaires investigated the different dimensions of use, interaction, and experience of the system and its components. These tests were selected from the literature, and then considered to be easy for users to read, understand and interpret, lacking direct mediation by the project staff. To ease users' answering, the questionnaires have been translated and sent by email in digital format. Users answered the questionnaires after the first use of the system and at the end of the pilot. We added open-ended questions and self-reported activities to these standardised questionnaires, to gain a more detailed insight into the users' experience. We share open source material and code for building a tangible coach: a physical, hardware and software design of a vocal assistant with tangible properties in the Appendix A.

3.4.2. Results and discussion

Among the devices proposed by NESTORE, the Tangible coach was among those that most aroused the curiosity of potential users. A large majority of the users involved make regular use of smartphones, apps, and wearable devices; therefore, there is a high degree of confidence with these objects. On the other hand, the Tangible presents different modes of interaction and relationship, not commonly known and used by our target users. The use of the voice channel for the interaction between man and object, its two-way nature, and the autonomy in starting a conversation represent something novel for our users. For these reasons, many of them were very curious and had high expectations of the tangible coach. The users' main expectation was to have a natural interaction with the e-coach, as if it were a dialogue between two people: this was partially missed for several reasons. Firstly, because the system was at a prototype level and still in development, some users noticed it (e.g., the 3 D printing, materials and finishing, or the assembly). At a second level, due to the pandemic, it was hard to perform the required activities and actions for the different domains; the lack of collected data has limited the interaction with the e-coach, for obtaining information to propose or stimulate other activities. Finally, frustration was due to misunderstandings in the dialogue between human and tangible (e.g., due to the vocabulary used or functional problems). However, this testing phase aimed at assessing the functionality, usability, and user experience of the NESTORE components—at varying levels of development—and with different degrees of complexity.

Below are some results of qualitative and quantitative results. To evaluate the level of Satisfaction and Acceptance when interacting with NESTORE, a single joint survey has been adopted, integrating the SUS and the TAM surveys. The SUS survey (Brooke, 1996) measures the satisfaction of users while interacting with products, services and systems; it is widespread because, on one hand, it can be customized and used for a wide range of products, services and systems, and, on the other hand, it is easy to manage, that means good reliability and validity measures, and solid benchmarks to aid in the interpretation of scores. In its base version, it is a 10-question survey, to be filled in using a Likert scale measuring the different degrees of satisfaction, from strong disagreement to strong agreement. In the end, it is possible to obtain a score between 0 and 100 for each user; from the literature, from the widespread use of the SUS and thus from comparisons with other product-systems, we know that a score of 68 represents the threshold for a product considered usable by users (Bangor et al., 2008; Brooke, 2013). Even with a small sample of participants, the SUS is a questionnaire that guarantees reliable results (Stetson & Tullis, 2004). Even though the prototypical and experimental level of the system, we decided to use these standardised

Table 3. The table presents the Tangible Score achieved at the initial and final stage of the Pilot.

	INSTALLATION / FIRST USE				FINAL USE			
Tangible	Answers received (Participants)	Average score	Best score	Worst score	Answers received (Participants)	Average score	Best score	Worst score
Global	17 (24)	47.5	77.5	10	9 (22)	32.77	47.5	2.5
ITA	6 (12)	48.75	77.5	10	2 (6)	43.75	47.5	40
SPA	4 (8)	45	75	12.5	3 (7)	31.67	47.5	22.5
NED	4 (4)	48.75	57.5	40	4 (9)	22.9	47.5	2.5

Results are presented as an aggregated score for each Pilot site.

surveys, knowing that the system functionality could have affected the results. At the same time, these surveys are also helpful in assessing the participants' attitudes toward these types of objects and interactions, their experience and the needs and barriers while using these objects.

Table 3 presents the SUS results achieved by the tangible after the first use and at the end of the pilot.

The results are presented both as an aggregate value for the three pilots and for each pilot; the number of answers obtained is indicated in the second and fifth column, with the total sample of respondents in brackets. The participants have indicated whether they had used the component of the system and, in the case of a negative answer, add some comments to explain why they didn't use it. The values obtained by the Tangible were compared with the scores and thresholds presented in the literature for the SUS questionnaire; the score defined as a threshold for a product answering the users' expectations is 54 points. In both evaluations, the tangible is below this value, and there is a worsening between the initial and final phase of the pilot. This negative score can be explained through some participants' comments, related to the product development, its functionality (e.g., technical problems, interaction difficulties, etc.), and the criticality of the experience due to the pandemic situation, such as the difficulty and risk of going out or meeting other people (and thus working on NESTORE's domains), as previously mentioned. Moreover, the worsening could be explained through the high level of expectation that were not met as time passed. During the final interviews and in some comments, the participants continued to show interest and attitude towards the tangible, its functionality, and the expected interaction possibilities despite the operational problems.

At the end of the pilot, to integrate the evaluation framework and to go beyond the concepts of functionality and usability, we used questionnaires specifically designed to measure the user experience, such as the UEQ survey (Laugwitz et al., 2008; Schrepp et al., 2017, p. 1). The UEQ investigates six domains of the user experience, that are: attractiveness (if the system looks attractive, enjoyable, friendly and pleasant), efficiency (if the tasks with the system are performed fast, efficiently and in a pragmatic way), perspicuity (if the system is easy to understand, clear, simple, and easy to learn), dependability (the interaction with the system is predictable, secure and meets user's expectations), stimulation (if the system is perceived as interesting, exciting and motivating), and novelty (if the system is perceived as innovative, inventive and creatively designed). The graph shown in Figure 15 derives from the answers given by the users to the UEQ questionnaire, in which 26 pairs of

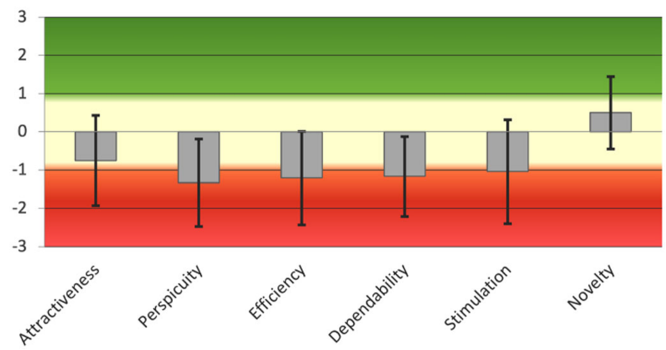

Figure 15. The UEQ graph present the results related to the user experience with the Tangible.

words with opposite meanings are presented, and to which the user must respond by indicating the one that for him/her is the meaning closest to the experience with the NESTORE component. The results are presented in the form of a graph, with a value scale ranging from −3 (extremely negative experience) to +3 (extremely positive experience). Once again, the results for five domains on six are in the neutral-negative threshold—represented by the light-yellow area. Again, these results are influenced by the critical issues in the tangible functioning, related to its prototypical status and to the contextual problems caused by the pandemic. It should be noted that despite these issues, the level of novelty perceived by users has remained very positive (represented in the sixth column in the graph).

Then, it is helpful to report some comments made by the users of the three pilots, here presented in a discursive way; these comments are made by those who had problems with the use and management of the system and therefore were unable to answer the questionnaires. It is also helpful to note these aspects as they constitute the lesson learnt for the project, the pilots, and for future developments of the e-coach and the tangible coach. Firstly, several users noted how Covid-19 affected the use of the system, and some of the functionalities and possibilities offered, for example, concerning monitoring physical activity, socialising, and the chance to choose and carry out some pathways. Concerning the Tangible, several users complained about technical issues, malfunctions, and connection problems with the system. Among those who used it, someone complained that the interaction and dialogue were not as expected, focusing on the quality of the bidirectional conversation and the information exchange. One user has reported some issues related to the Tangible visual interface—the one that uses LED lights. More in general, the functioning was not as easy as users expected. Different issues have affected the use and the interaction with the Tangible: problems with assembly

Table 4. Design recommendations.

Research axes	Theme	Recommendations
Physical Design	Look and feel	Use simple, geometrical designs, as they would appeal most people
	Transportability	Ensure the possibility to use it without a power cord and to grab it with one hand
Interaction Design	Content	Provide meaningful content and suggestions
	Voice	Provide voice conversations that are as natural as possible, while aligning users' expectation to the technological limitations (provide a user manual/instructions/possible intent lists)
	Tangible controls	Provide intuitive tangible controls and data representations. However, keep tangible interfaces simple, stable and robust.
	Lights	Lights should be used to show the state of the system. Prefer light patterns to colors, to account for color blindness.
	Facial expressions	Facial expressions and colors can be used to increase the empathy of the agent
Privacy and data control	Data acquisition control	Provide physical control for disabling voice recognition

had happened due to transportation to pilot sites. Also, issues with the user's slow WiFi negatively affected its functioning. The tangible coach became slower and sometimes failed to send the requests. Finally, participants reported some problems with devices pairing using Bluetooth with the mobile phone.

4. Discussion

4.1. Design recommendations

During the NESTORE project we carried out a co-design process to identify the main features of a physically embodied virtual coach. We discovered that the best embodiment for this virtual coach according to older adults was very similar to commercial smart speakers. Nevertheless, we collected many additional insights about features that such smart speakers should spot. Based on these insights, we propose design recommendations for future generations of senior-friendly smart speakers (Table 4):

4.1.1. Physical design

(1) The shape, look and feel are one of the main considerations to be taken when building an e-coach (R10-F4-F10). Most previous work elicited design guidelines for appearance in terms of role, personality and empathy (Kim, (2021; Kramer et al., 2021; Rheu et al., 2021), however no work was found were they built the e-coach appearance with users. The physical appearance of the e-coach is what attracts the users to buy and start using the product. As we saw in the pilot study, the tangible coach was always the most interesting device compared to the other devices of the system, and triggered the curiosity in older adults. Participants to our study have preferred the geometric design of the device. Something simple, nice and beautiful is enough. It is worth reminding that our target users were older adults that are independent and interested in preserving their well-being. So far, most examples of conversational agents' embodiment were designed for frail older adults, typically for daily assisted living.

(2) Transportability is also important (R9, F9). Users should be able to move the device with them and change its location. The freedom of taking the e-coach wherever they want can also form a sense of companionship at some point. Being able to carry these devices and change their location can affect users' perception of their relationship with them (Pradhan et al., 2019). This also aligns with the conclusions of (Sin et al., 2022) that making these vocal assistant life on-the-go instead of stationary, as suggested in the media, could be an important factor for older adults' adoption of a voice user interfaces. With the NESTORE Tangible Coach, we investigated the transportability of the smart speaker, a feature that was particularly important for our user base, especially in the context of an e-coach for wellbeing, and is often lacking in commercial smart speakers. Although we managed to design a device compact enough to be held with one hand, the duration of the battery could be improved. We believe that future generations of hardware would allow further reducing the power consumption and integrating bigger batteries that could last easily one day without charging.

4.1.2. Interaction design

(3) The embodied vocal assistant must provide meaningful conversations and information (R5, R11, R13, F11), otherwise users could lose interest and will eventually stop using the e-coach. Our pilot study has confirmed that this requirement is a must. In fact, one of our limitations was the fact that our tangible coach was not robust enough for a long period of use. However, the tangible coach was not limited to greeting and topical information, it was capable of giving personalized recommendations, suggestions, and citing the scheduled activities of the user and the score with personalized constructive feedback on his or her progress. This was also shown in literature by Kim, (2021), who explained the necessity of enriching user experiences, not only in content but also through conversational

capabilities. In fact, acknowledging when errors or misinterpretations occur from the agent side (Chen et al., 2021) is recommended. Additional details of the source of information and the reasons of errors can help to compensate for the lack of technology accuracy and foster the digital companionship (Pradhan et al., 2018). Using user's same language and adapting to one's culture's accent (Harrington et al., 2022) is also recommended.

(4) Vocal Interaction is also recommended. Users appreciate the voice modality (R8, F4, F6). Speech has been considered an accessible and useable interaction modality for older adults (Bickmore & Picard, 2005; Rheu et al., 2021). Audio is the preferred modality for people without hearing impairments (Vacher et al., 2015). Many challenges are still open for voice interaction, a feature that is particularly appreciated by older adults, but that also underlines very high expectations in terms of system performance. Indeed, users' mental models for voice interaction are those of human vocal exchanges, which are far distant from the capabilities of conversational interfaces in terms of natural language understanding, general AI to sustain generic conversations, and management of conversational turns. While technology is evolving and improving, it is important to calibrate the user expectations on the system capabilities. It has been also found that older adult users can learn to cope with the speech style required to interact with smart speakers and handling errors, as discovered in a longitudinal study of older adults' perception and use of voice user interfaces (Kim & Choudhury, 2021). Moreover, A short user manual with the typical questions that could be asked to the tangible coach was printed out and given to the users, which helped a lot in receiving appropriate answers from the coach. This helped calibrate users' expectations to the system capabilities.

(5) Tangible controls are also appreciated since it makes the user feel in control of some aspects (R1, F6). An important feature that we investigated for the tangible coach is the possibility to integrate richer physical interactions, based on a shape-changing interface. While many users were sceptical of this feature, we believe that this may be due to a lack of existing devices with such features. At the same time, trading off the robustness of a device with moving parts, especially if activated by servo-motors, is technically not easy. While many design explorations have been done in this direction, it is not easy to find reliable solutions that have been delivered in commercial products. In the NESTORE tangible coach, we chose to limit tangible interaction to the physical manipulation of the orientation and the related coach state. This was easily implemented with an internal accelerometer, robust and reliable enough for being integrated in our design. In fact, academic literature highlights the importance of older adults having control over their interactions with VUIs, particularly the option to switch the systems off, since this plays a critical role in smart speakers and related technology acceptance (Sin et al., 2022). We also investigated the possibility of making the device oscillate for notifications, shifting an internal weight. While this could be eventually implemented in a rather robust way, it also implied to have a hemispherical surface on the base to facilitate the oscillation. As we have seen in the co-design phase, the stability of the device was a main concern, and also this solution was dropped in the final design. Further investigations for the acceptability for older adults of shape changing interfaces should be conducted in our opinion exploring appropriation and routines developed by users in long-term use. The physical control of the tangible coach state allowed us to counteract the unreliability of the customized wakeword, which we could not train on thousands of samples in order to make it robust enough, eliminating many annoying and disturbing false positives.

(6) Lights (R6, F2, F7) can be used to tell the user in what state the device is. During our co-design process, many older adults did not like the idea of having a bot with facial expressions. Hence, in order to help the user to know in what state the user was during the interaction, lights helped us to inform the user about the situation. We also investigated the use of colors in order to depict a certain emotion based on the Plutchik Diagram (Plutchik, 2001). We created several patterns of the six basic emotions (happiness, sadness, surprise, disgust, anger, fear) and implemented them on the 18 circular LEDs of the matrix voice. Although we could not test this feature during the pilot, we would like to keep it for future investigation. Although we used a single color, we believe that personalizing the color of the lights of the tangible coach based on the user's current emotion and visual capabilities can also increase empathy and the sense of companionship.

4.1.3. Privacy and data control

(7) Privacy concerns was always an issue for most older adults (Bonilla & Martin-Hammond, 2020; Kim, 2021; Tsiourti et al., 2014). It has been found that the media also raised many concerns about the lack of privacy control, and their potential impact on technology adoption; however, this aspect was almost unexplored in scientific literature (Sin et al., 2022). Users must be able to control data access and data collection (R5, F11, F6). Since one of the main requirements of older adults when using technology is privacy and data control, it is a must to build a design that focuses on enabling intuitively this requirement. The ability of the tangible coach to have two main states (the sleep state and the awake state) gives the user the full control to choose when he or she wants to be heard. As much as we could not test this feature alone in order to investigate the tangible control impact on older adults' trust, we believe that such feature could provide tangible benefits in many product interface designs.

4.2. Open technical challenges

Besides the common technical problems that arise when building physical hardware devices, the main challenge for the tangible coach was to complete a whole co-design process and obtain at the end of the process a robust device that could be tested in long-term by older adults, with minimal technical support. This was particularly difficult in the short 3-year time-frame of an EU project. Progress in 3D printing and in prototyping hardware allowed us to speed up the physical design and development, obtaining in a relatively short time (less than 1 year) a quite reliable prototype that can be scaled up to 70 samples. Currently available hardware and software solutions all have their own limitations but their continuous evolution can enable new startups to propose alternative designs to the standard propositions of big IT players. We believe that the design of the tangible coach may be used in different opportunities with the aim of having a virtual coach for older adults' wellbeing. Another open challenge concerns the difficulty of building voice conversations that meet older adults' expectations. Although speech services are nowadays getting better in most languages, the path towards a conversational agent that can sustain an open ended conversation is still long. Language models such as GPT-3 (Floridi and Chiriatti (2020) can help in providing general knowledge to a conversational agent, enabling conversations that sound more real. Nevertheless, studies are shown that these models, which are trained over large internet databases, are prone to bias and discrimination. In our project, we wanted to ensure that the information provided by the e-coach would reflect domain expert recommendations, based on data collected from the system, with expressions that were manually translated by humans, to avoid biases.

4.3. Limitations

Our study shed light on older adults' attitudes and their needs in relation to physically embodied e-coaches. The study aimed at gathering user requirements across four EU countries with different cultures and environments. Cultural factors may vary across different continents and different needs and expectations may be elicited for users with different cultures. It is also worth mentioning that we targeted older adults that are still independent and that are willing to maintain their well-being. Needs and expectations may vary for frail older adults. We also expect that as smart speakers get diffused in the market, expectations about such objects may become aligned to technological capabilities of such devices, by means of habit and appropriation of such devices. The insights collected from our target users may therefore vary in the future, as vocal assistants spread in the users' homes and daily routines. The results of our tests were also profoundly affected by the COVID-19 outbreak, both for the logistic of the pilot study (difficulty in recruiting, limited duration of the study, difficulty in providing technical assistance), and for the user experience (limited number of activities that could be performed during COVID, thus limited overall experience for the e-coach and limited data available to ask feedback about it). The hardware and physical design that we produced was meant to rapidly produce a sufficient amount of samples for the pilot study. The design is ready to be reused for producing a limited number of prototypes, but it may still be improved for a larger scale production, especially in terms of robustness and assembly quality. The design was already thought to be shock resistant, however as the different parts of the coach were assembled with aftermarket components, the reliability of connetions could be inferior to a unique board designed from scratch. Indeed, typical technical problems encountered during pilot tests were related to loose connections or detached parts. Considering the current limitations, our goal was to share our results with the community, so that other researchers could take our work as a starting point, either as design recommendations and lessons learnt, or as a development kit for building their own embodied voice-enabled e-coach.

5. Conclusion

E-coaches are the next frontier in digital health. While researchers have developed so far several embodied conversational robots for assisting older adults, a huge gap should be overcome before such devices can enter the market. On the other side, vocal assistants embodied in smart speakers provide affordable technology that could be potentially useful for older adults. However, these devices were not designed with and for older adults. In this article, we presented a 3-year co-design study that allowed us to shed light on older adults' needs and expectations about an embodied e-coach to support healthier lifestyles. A clear need that emerged for such an e-coach was the ability to fit within the complexity of people's lives and engage in natural conversations. At the same time, the high expectations about such a conversational coach may be disappointed by current technology. While proper calibration between user expectations and technology capability should be ensured, other features should be considered to enhance the user experience with such devices. A particular achievement of our co-design process was the tangible control over the device state, which allowed the user to physically manipulate data acquisition and to better stay in control of their data. Within the NESTORE project, we started a co-design journey that allowed us to deepen our understanding of older adults' needs for physically embodied voice assistants, as well as of the solutions that can be designed to address these needs. As many challenges are still open, we would like to call upon other researchers to share this enriching journey. To this purpose, we shared all our findings from the co-design process in this article and the co-design outputs (physical designs, hardware assembly and software) in the annex open source repository.

Acknowledgments

The authors want to thank all the NESTORE Consortium. The opinions expressed in this paper are those of the authors and are not necessarily those of the project partners or the European Commission.

Disclosure statement

The authors hereby certify that, to the best of their knowledge, there are no conflicts of interest to disclose.

Funding

This project is supported by the European Commission under the Horizon 2020 programme, SC1-PM-15-2017—Personalised Medicine topic, through the project grant N.76964.

ORCID

Mira El Kamali http://orcid.org/0000-0003-2895-6867

References

Anderson, M., & Perrin, A. (2017). *Technology use among seniors*. Pew Research Center for Internet & Technology.

Angelini, L., El Kamali, M., Mugellini, E., Abou Khaled, O., Dimitrov, Y., Veleva, V., Gospodinova, Z., Miteva, N., Wheeler, R., Callejas, Z., Griol, D., Benghazi, K., Noguera, M., Bamidis, P., Konstantinidis, E., Petsani, D., Beristain Iraola, A., Fotiadis, D. I., Chollet, G., Torres, I., … Schliete, H. (2020). First workshop on multimodal e-coaches. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (pp. 890–892). Association for Computing Machinery.

Angelini, L., El Kamali, M., Mugellini, E., Abou Khaled, O., Röcke, C., Porcelli, S., Mastropietro, A., Rizzo, G., Boqué, N., del Bas, J. M., Palumbo, F., Girolami, M., Crivello, A., Ziylan, C., Subías-Beltrán, P., Orte, S., Standoli, C. E., Fernandez Maldonado, L., Caon, M., … Andreoni, G. (2022). The nestore e-coach: Designing a multi-domain pathway to well-being in older age. *Technologies*, *10*(2), 50. https://doi.org/10.3390/technologies10020050

Ataei, M., Degbelo, A., & Kray, C. (2018). Privacy theory in practice: Designing a user interface for managing location privacy on mobile devices. *Journal of Location Based Services*, *12*(3–4), 141–178. https://doi.org/10.1080/17489725.2018.1511839

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Banos, O., & Nugent, C. (2018). E-coaching for health. *Computer Magazine*. *51*(3), 12–15. https://doi.org/10.1109/MC.2018.1731070

Bickmore, T. W., Caruso, L., & Clough-Gorr, K. (2005). Acceptance and usability of a relational agent interface by urban older adults. In *CHI'05 extended Abstracts on Human Factors in Computing Systems* (pp. 1212–1215). ACM 1-59593-002-7/05/0004. https://doi.org/10.1145/1056808.1056879

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, *12*(2), 293–327. https://doi.org/10.1145/1067860.1067867

Bickmore, T. W., Silliman, R. A., Nelson, K., Cheng, D. M., Winter, M., Henault, L., & Paasche-Orlow, M. K. (2013). A randomized controlled trial of an automated exercise coach for older adults. *Journal of the American Geriatrics Society*, *61*(10), 1676–1683. https://doi.org/10.1111/jgs.12449

Bonilla, K., & Martin-Hammond, A. (2020, August 9–11). *Older adults' perceptions of intelligent voice assistant privacy, transparency, and online privacy guidelines* [Paper presentation]. In *USENIX Symposium on Usable Privacy and Security (SOUPS),* Boston, MA.

Brooke, J. (2013). Sus: A retrospective. *Journal of Usability Studies*, *8*(2), 29–40.

Brooke, J. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in Industry*, *189*(194), 4–7. https://doi.org/10.1201/9781498710411-35

Caine, K., Šabanovic, S., & Carter, M. (2012). The effect of monitoring by cameras and robots on the privacy enhancing behaviors of older adults. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 343–350). Association for Computing Machinery.

Chamberlain, P., Craig, C. (2013). Engagingdesign–methods for collective creativity. In *International Conference on Human-Computer Interaction* (pp. 22–31). Springer.

Chattaraman, V., Kwon, W.-S., Gilbert, J. E., & Ross, K. (2019). Should AI-based, conversational digital assistants employ social-or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior*, *90*, 315–330. https://doi.org/10.1016/j.chb.2018.08.048

Chen, C., Johnson, J. G., Charles, K., Lee, A., Lifset, E. T., Hogarth, M., Moore, A. A., Farcas, E., & Weibel, N. (2021). Understanding barriers and design opportunities to improve healthcare and QOL for older adults through voice assistants. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 1–16). Association for Computing Machinery.

Cho, M., Lee, S-S., Lee, K.-P. (2019). Once a kind friend is now a thing: Understanding how conversational agents at home are forgotten. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (pp. 1557–1569). Association for Computing Machinery.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Dubois, D. J., Kolcun, R., Mandalari, A. M., Paracha, M. T., Choffnes, D., Haddadi, H. (2020). When speakers are all ears: Characterizing misactivations of I$_O$T smart speakers. *Proceedings on Privacy Enhancing Technologies*, *2020*(4), 1–22. https://doi.org/10.2478/popets-2020-0072

El Kamali, M., Angelini, L., Caon, M., Carrino, F., Rocke, C., Guye, S., Rizzo, G., Mastropietro, A., Sykora, M., Elayan, S., Kniestedt, I., Ziylan, C., Lettieri, E., Khaled, O. A., & Mugellini, E. (2020). Virtual coaches for older adults' wellbeing: A systematic review. *IEEE Access*. *8*, 101884–101902. https://doi.org/10.1109/ACCESS.2020.2996404

Floridi, L., & Chiriatti, M. (2020). Gpt-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Funding and Tender Opportunities. (2022). *Personalised coaching for well-being and care of people as they age*. The Publications Office of the European Union. https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/sc1-pm-15-2017

Georgiadis, D., Christophorou, C., Kleanthous, S., Andreou, P., Santos, L., Christodoulou, E., & Samaras, G. (2016). A robotic cloud ecosystem for elderly care and ageing well: The growmeup approach. In *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016* (pp. 919–924). Springer.

Harkous, H., Fawaz, K., Shin, K. G., & Aberer, K. (2016, June 22–24). *Pribots: Conversational privacy with chatbots* [Paper presentation]. Workshop on the Future of Privacy Notices and Indicators, at the Twelfth Symposium on Usable Privacy and Security, SOUPS 2016, Denver, CO.

Harrington, C. N., Garg, R., Woodward, A., & Williams, D. (2022). "It's kind of like code-switching": Black older adults' experiences with a voice assistant for health information seeking. In *CHI Conference on Human Factors in Computing Systems* (pp. 1–15). Association for Computing Machinery.

Hawthorne, G. (2006). Measuring social isolation in older adults: Development and initial validation of the friendship scale. *Social Indicators Research*, *77*(3), 521–548. https://doi.org/10.1007/s11205-005-7746-y

Ischen, C., Araujo, T., Voorveld, H., van Noort, G., & Smit, E. (2019). Privacy concerns in chatbot interactions. In *International Workshop on Chatbot Research and Design* (pp. 34–48). Springer.

Khosla, R., & Chu, M.-T. (2013). Embodying care in matilda: An affective communication robot for emotional wellbeing of older people in Australian residential care facilities. *ACM Transactions on Management Information Systems*, *4*(4), 1–33. https://doi.org/10.1145/2544104

Kim, S. (2021). Exploring how older adults use a smart speaker–based voice assistant in their first interactions: Qualitative study. *JMIR mHealth and uHealth*, *9*(1), e20427. https://doi.org/10.2196/20427

Kim, S., & Choudhury, A. (2021). Exploring older adults' perception and use of smart speaker-based voice assistants: A longitudinal study. *Computers in Human Behavior*, *124*, 106914. https://doi.org/10.1016/j.chb.2021.106914

Koch, D., & Maaß, S. (2018). Digital probes kit. *i-com*, *17*(2), 169–178. https://doi.org/10.1515/icom-2018-0016

Kramer, L. L., Blok, M., Van Velsen, L., Mulder, B. C., & De Vet, E. (2021). Supporting eating behaviour of community-dwelling older adults: Co-design of an embodied conversational agent. *Design for Health (Abingdon, England)*, *5*(1), 120–139. https://doi.org/10.1080/24735132.2021.1885592

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group* (pp. 63–76). Springer.

Lawson, S., & Chesney, T. (2007). The impact of owner age on companionship with virtual pets. In *European Conference on Information Systems* (pp. 1922–1928).

Luh, D.-B., Li, E. C., & Gao, Y.-R. (2010). The study on companionship scale of electronic pet. In *2010 IEEE 11th International Conference on Computer-Aided Industrial Design & Conceptual Design 1* (Vol. 1, pp. 533–538). IEEE.

Luria, M., Zheng, R., Huffman, B., Huang, S., Zimmerman, J., & Forlizzi, J. (2020). Social boundaries for personal agents in the interpersonal space of the home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). Association for Computing Machinery.

Martin-Hammond, A., Vemireddy, S., & Rao, K. (2019). Exploring older adults' beliefs about the use of intelligent assistants for consumer health information management: A participatory design study. *JMIR aging*, *2*(2), e15381. https://doi.org/10.2196/15381

Melenhorst, A.-S., Rogers, W. A., & Bouwhuis, D. G. (2006). Older adults' motivated choice for technological innovation: Evidence for benefit-driven selectivity. *Psychology and Aging*, *21*(1), 190–195. https://doi.org/10.1037/0882-7974.21.1.190

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, *7*, 33–35.

Orte, S., Subías, P., Maldonado, L. F., Mastropietro, A., Porcelli, S., Rizzo, G., Boqué, N., Guye, S., Röcke, C., Andreoni, G., Crivello, A., & Palumbo, F. (2018). Dynamic decision support system for personalised coaching to support active ageing. In *AI* Aal@ AI* IA* (pp. 16–36). CEUR-WS.org.

Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, *89*(4), 344–350. https://doi.org/10.1511/2001.4.344

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). ACM.

Pradhan, A., Findlater, L., & Lazar, A. (2019). "Phantom friend" or "just a box with information" personification and ontological categorization of smart speaker-based voice assistants by older adults. *Proceedings of the ACM on Human-Computer Interaction*, 3(Cscw): 1–21. https://doi.org/10.1145/3359316

Pradhan, A., Mehta, K., Findlater, L. (2018). "Accessibility came by accident" use of voice-controlled intelligent personal assistants by people with disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). ACM.

Price, S., & Rogers, Y. (2004). Let's get physical: The learning benefits of interacting in digitally augmented physical spaces. *Computers & Education*, *43*(1–2), 137–151. https://doi.org/10.1016/j.compedu.2003.12.009

National Public Media (2022). *The smart audio report*. https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/.

Rheu, M., Shin, J. Y., Peng, W., & Huh-Yoo, J. (2021). Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human–Computer Interaction*, *37*(1), 81–96. https://doi.org/10.1080/10447318.2020.1807710

Ring, L., Barry, B., Totzke, K., Bickmore, T. (2013). Addressing loneliness and isolation in older adults: Proactive affective agents provide better support. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 61–66). IEEE.

Sannon, S., Stoll, B., DiFranzo, D., Jung, M. F., & Bazarova, N. N. (2020). "I just shared your responses" extending communication privacy management theory to interactions with conversational agents. *Proceedings of the ACM on Human-Computer Interaction*, 4(GROUP):1–18.

Sansen, H., Chollet, G., Glackin, C., Jokinen, K., Torres, B. A., Petrovska-Delacretaz, D., Boudy, J., & Schlögl, S. (2016, June 8–10). The roberta IRONSIDE project: A cognitive and physical robot coach for dependent persons. In *Proceedings of Handicap 2016*. Association for the Advancement of Modelling and Simulation Techniques in Enterprises (AMSE).

Schmitz, M. (2010). Concepts for life-like interactive objects. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 157–164). Association for Computing Machinery.

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Construction of a benchmark for the user experience questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*(4), 40–44. https://doi.org/10.9781/ijimai.2017.445

Schwarzer, R., & Luszczynska, A. (2008). How to overcome health-compromising behaviors: The health action process approach. *European Psychologist*, *13*(2), 141–151. https://doi.org/10.1027/1016-9040.13.2.141

Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). "Hey Alexa, what's up?" A mixed-methods studies of in-home conversational agent usage. In *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 857–868). ACM.

Sin, J., Munteanu, C., Chen, D., & Threatt, J. (2022). Avoiding mixed messages: Research-based fact-checking the media portrayals of voice user interfaces for older adults. *Human–Computer Interaction*, 1–24. https://doi.org/10.1080/07370024.2022.2098129

Snowboy (n.d). *Snowboy wakeword engine*. Snowboy. https://github.com/Kitt-AI/snowboy

Stetson, J. N., & Tullis, T. S. (2004). *A comparison of questionnaires for assessing website usability*. UPA Presentation.

Stusak, S., Tabard, A., Sauka, F., Khot, R. A., & Butz, A. (2014). Activity sculptures: Exploring the impact of physical visualizations on running activity. *IEEE transactions on Visualization and Computer Graphics*, *20*(12), 2201–2210. https://doi.org/10.1109/TVCG.2014.2352953

Ter Stal, S., Tabak, M., Op den Akker, H., Beinema, T., & Hermens, H. (2020). Who do you prefer? The effect of age, gender and role on users' first impressions of embodied conversational agents in ehealth. *International Journal of Human–Computer Interaction*, *36*(9), 881–892. https://doi.org/10.1080/10447318.2019.1699744

Tonkin, M., Vitale, J., Ojha, S., Clark, J., Pfeiffer, S., Judge, W., Wang, X., & Williams, M.-A. (2017). Embodiment, privacy and social robots: May I remember you? In *International Conference on Social Robotics* (pp. 506–515). Springer.

Trajkova, M., & Martin-Hammond, A. (2020). "Alexa is a toy": Exploring older adults' reasons for using, limiting, and abandoning echo. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). ACM.

Tsiourti, C., Joly, E., Wings, C., Moussa, M. B., & Wac, K. (2014). 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* (pp. 57–64). ACM.

Vacher, M., Caffiau, S., Portet, F., Meillon, B., Roux, C., Elias, E., Lecouteux, B., & Chahuara, P. (2015). Evaluation of a context-aware voice interface for ambient assisted living: Qualitative user study vs. quantitative system evaluation. *ACM Transactions on Accessible Computing*, *7*(2), 1–36. https://doi.org/10.1145/2738047

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Wang, Y. H., Young, S. S., & Jang, J.-S. R. (2009). Evaluation of tangible learning companion/robot for English language learning. In *2009 Ninth IEEE International Conference on Advanced Learning Technologies* (pp. 322–326). IEEE.

Wu, Y.-H., Wrobel, J., Cristancho-Lacroix, V., Kamali, L., Chetouani, M., Duhaut, D., Le Pevedic, B., Jost, C., Dupourque, V., Ghrissi, M., & Rigaud, A.-S. (2013). Designing an assistive robot for older adults: The robadom project. *IRBM*, *34*(2), 119–123. https://doi.org/10.1016/j.irbm.2013.01.003

Zeissig, E.-M., Lidynia, C., Vervier, L., Gadeib, A., & Ziefle, M. (2017). Online privacy perceptions of older adults. In *International Conference on Human Aspects of IT for the Aged Population* (pp. 181–200). Springer.

About the authors

Mira El Kamali is currently a Ph.D. student in Computer Science at the University of Fribourg, Switzerland. She works in collaboration with the Humantech Institute at the University of Applied Sciences and Arts Western Switzerland. Her research interests include human–computer interaction, conversational agents, focusing on older adults.

Leonardo Angelini has a Ph.D. degree in computer science obtained at the University of Fribourg. He is currently a Professor of digital business at the School of Management of Fribourg. His research interests include human–computer interaction and eHealth, focusing on older adults as target users.

Maurizio Caon is currently full professor at the University of Applied Sciences and Arts Western Switzerland (HES-SO) and Head of the Digital Business Center at the School of Management. His research domains are in the area of human computer interaction, persuasive technology and digital innovation.

Nick Dulake is a research associate at Lab4Living at Sheffield Hallam University and a Senior industrial designer at DesignFutures. His interests are in user centered and co-design methodology to create innovative design solutions. He applies technical and creative design methods and developing interventions with a focus on digital tangible interactions.

Paul Chamberlain is Professor of Design at Sheffield Hallam University, UK, founder and director of the interdisciplinary research group Lab4Living. His interest lies in designing and developing tools and methods to encourage and engender social innovation applied with a focus on healthcare, disability and ageing.

Claire Craig is Professor of Design and Creative Practice in Health and Co-Director of Lab4Living. Her research focuses on the role of design in promoting quality of life and wellbeing, particularly in the context of people experiencing major life transitions.

Carlo Emilio Standoli is Assistant professor at Politecnico di Milano, Department of Design. His research activity focuses on the role of Design Culture in the Healthcare field, addressing the development of physical and digital products, services, and systems, the related interaction and the evaluation of the experience elicited in users.

Giuseppe Andreoni is Full Professor at the Department of Design of Politecnico di Milano where he is the Director of the Laboratory “Technology and Design for Healthcare”. His research activity is focused on the design of products, systems, environments and services for the health and well-being of people.

Omar Abou Khaled received the Ph.D. degrees in computer science from the HEUDIASYC Laboratory, Perception and Automatic Control Group, Université de Technologie de Compiègne. He is a Professor with the University of Applied Sciences of Western Switzerland.

Elena Mugellini hold a Ph.D. degree in Telematics and Information Society from the University of Florence. She is a Professor at the University of Applied Sciences of Western Switzerland, Fribourg. Her current research interests include ambient intelligent, multimodal interaction, natural interface, intelligent data analysis, information, and knowledge management.

Appendix A. Hardware design and software design source code

https://github.com/mirakamali/tangiblecoach.git