The EMPATHIC Project: Mid-term Achievements

M. I. TORRES, J. M. OLASO, C. MONTENEGRO, R. SANTANA, A. VÁZQUEZ, R. JUSTO, and J. A. LOZANO, Universidad del País Vasco UPV/EHU, Spain

S. SCHLÖGL, MCI Management Center Innsbruck, Austria

G. CHOLLET, N. DUGAN, M. IRVINE and N. GLACKIN, C. PICKARD, Intelligent Voice Ltd, UK

A. ESPOSITO, G. CORDASCO and A. TRONCONE, Università degli Studi della Campania, Luigi Vinvitelli, Italy

D. PETROVSKA-DELACRETAZ and A. MTIBAA, M. A. HMANI, Institut Mines-Telecom, France

M. S. KORSNES and L. J. MARTINUSSEN, Department of Old Age Psychiatry, Oslo University Hospital, Norway

S. ESCALERA and C. PALMERO CANTARIÑO, Universitat de Barcelona and Computer Vision Center, Spain

O. DEROO and O. GORDEEVA, Acapela Group, Belgium

J. TENORIO-LARANGA, E. GONZALEZ-FRAILE, B. FERNANDEZ-RUANOVA, and A. GONZALEZ-PINTO, Osatek/ Osakidetza, Spain

The goal of active aging is to promote changes in the elderly community so as to maintain an active, independent and socially-engaged lifestyle. Technological advancements currently provide the necessary tools to foster and monitor such processes. This paper reports on mid-term achievements of the European H2020 EMPATHIC project, which aims to research, innovate, explore and validate new interaction paradigms and platforms for future generations of personalized virtual coaches to assist the elderly and their carers to reach the active aging goal, in the vicinity of their home. The project focuses on evidence-based, user-validated research and integration of intelligent technology, and context sensing methods through automatic voice, eye and facial analysis, integrated with visual and spoken dialogue system capabilities. In this paper, we describe the current status of the system, with a special emphasis on its components and their integration, the creation of a Wizard of Oz platform, and findings gained from user interaction studies conducted throughout the first 18 months of the project.

CCS Concepts: • **Human-centered computing** → **Interactive systems and tools**; **Empirical studies in HCI**; *Interaction techniques*; • **Applied computing** → *Health informatics*.

Additional Key Words and Phrases: Emotional Artificial Agents, Assisted Living, Coaching, Spoken Dialogue Systems

Authors' addresses: M. I. Torres, J. M. Olaso; C. Montenegro, R. Santana; A. Vázquez, R. Justo; J. A. Lozano, Universidad del País Vasco UPV/EHU, Bilbao, Spain, manes.torres@ehu.eus; S. Schlögl, MCI Management Center Innsbruck, Innsbruck, Austria, stephan.schloegl@mci.edu; G. Chollet, N. Dugan, M. Irvine; N. Glackin, C. Pickard, Intelligent Voice Ltd, London, UK, gerard.chollet@intelligentvoice.com; A. Esposito, G. Cordasco; A. Troncone, Università degli Studi della Campania, Luigi Vinvitelli, Caserta, Italy, iiass.annaesp@tin.it; D. Petrovska-Delacretaz; A. Mtibaa, M. A. Hmani, Institut Mines-Telecom, Evry, France, dijana.petrovska@telecom-sudparis.eu; M. S. Korsnes; L. J. Martinussen, Department of Old Age Psychiatry, Oslo University Hospital, Oslo, Norway, m.s.korsnes@psykologi.uio.no>; S. Escalera; C. Palmero Cantariño, Universitat de Barcelona and Computer Vision Center, Barcelona, Spain, sergio@maia.ub.es; O. Deroo; O. Gordeeva, Acapela Group, Mons, Belgium, olivier.deroo@acapela-group.com; J. Tenorio-Laranga; E. Gonzalez-Fraile; B. Fernandez-Ruanova; A. Gonzalez-Pinto, Osatek/ Osakidetza, Bilbao, Spain, jtenorio@osatek.eus.

1 INTRODUCTION

Despite advances in health care and technology, most of the elder care is still provided by informal caregivers, i.e. friends and family members. According to predictions, however, this type of care will decrease in the future, for which studies encourage society to concentrate on improving the elderly's lifestyle, helping them to remain independent for a longer period of time [39]. In particular, the focus should be on external and internal difficulties of the elderly, offering arrangements and facilities to support active aging. Indeed, socio-behavioral and environmental conditions are a crucial factor affecting longevity [14], which to some extent explains variations found in the aging process, ranging from active and positive to feeble and dependent. We believe that four principles promote active aging, namely dignity, autonomy, participation, and joint responsibility. Information and Communication Technologies (ICT) are expected to make such principles possible, allowing the elderly to stay active members of the societal community while helping them remain independent and self-sufficient [2].

Consequently, the EMPATHIC (*Empathic, Expressive, Advanced Virtual Coach to Improve Independent Healthy-Life-Years of the Elderly*) project[1] aims to contribute to technological progress in this area by researching, innovating and validating new interaction paradigms and platforms for future generations of personalized Virtual Coaches (VC) to promote active aging. It is centred around the development of the EMPATHIC-VC, a non-obtrusive, emotionally-expressive virtual coach whose aim is to engage senior users in enjoying a healthier lifestyle concerning diet, physical activity, and social interactions. This way, they actively minimize their risk of potentially chronic diseases, which contributes to their ability to maintain a pleasant and autonomous life, while in turn it helps their carers. The main goal of the VC is to create a link between one's body and emotional well-being. To do so, it will perceive and identify users' social and emotional states by means of multi-modal face, eye gaze and speech analytics modules. Furthermore, it will learn and understand users' requirements and expectations, and adaptively respond to their needs through novel spoken dialogue systems and intelligent computational models. Such a combination of modules will allow for user-coach real-time interaction, thus promoting empathy in the user.

In this paper, we describe our mid-term achievements, explaining where we currently stand with these goals, 18 months into the project. Section 2 describes the current status of the system components, with particular emphasis on the most robust modules up to date. The integration of these modules is presented in Section 3. Lessons learned from the preliminary human-coach interaction studies are explained in Section 4, while Section 5 concludes the paper.

2 STATUS: SYSTEM COMPONENTS

The EMPATHIC VC is based on the following system components, each of which is researched, built and evaluated independently.

2.1 Automatic Speech Recognition

The Automatic Speech Recognition (ASR) component turns speech from a continuous stream of audio into structured data, containing likely words and their alternatives, labelled with the confidence level for each, start time (as an offset from the stream start), and duration. So far, our main achievements with this component are:

- The development ASR.online, a new method of reading data into the ASR engine which uses the opensource GStreamer framework[2] to continuously stream the audio through the ASR executable. Compared with our

---

[1] http://www.empathic-project.eu/
[2] https://github.com/GStreamer/gstreamer

Table 1. The relative performance of the original (ASR) and new (ASR.online) approaches. All values are in seconds.

Audio Duration (in sec.)	ASR CPU (in sec.)	ASR GPU (in sec.)	ASR.online CPU (in sec.)	ASR.online GPU (in sec.)
2.0	6.0	6.0	2.5	1.5
5.0	7.0	6.0	3.0	1.5
10.0	7.0	6.0	4.6	2.0
15.0	10.0	6.0	6.5	2.5
20.0	13.0	6.0	8.0	4.0
25.0	15.0	6.0	9.5	4.5
40.0	22.0	8.0	17.0	5.5
70.0	35.0	9.0	26.0	5.5
100.0	47.0	10.0	38.5	7.0
130.0	63.0	12.0	45.0	7.5
600.0	279.0	25.0	210.0	16.5
1200.0	550.0	64.0	443.0	30.0

Table 2. SCLITE benchmarking for supported languages.

Language	COR	SUB	DEL	INS	WER
English	87.7	9.2	3.1	4.0	16.3
French	73.9	22.7	3.4	10.1	36.3
Spanish	78.1	12.7	9.3	4.4	26.3
Norwegian	55.9	19.8	24.4	7.5	51.5

previous approach of buffering audio and processing small chunks, ASR.online reduces the latency between the speech and the transcription. The average latency has been reduced to approximately 500 milliseconds, as a result of both overall performance improvements as shown in Table 1, and the immediate transmission of transcription data.

- The training of acoustic models for French, Spanish and Norwegian, the languages which will be used in the EMPATHIC field trials. The acoustic models of the ASR component were trained using the Kaldi ASpIRE recipe[3]. The training data consists of 1067 hours of Spanish speech, 271 hours of French and 228 hours of Norwegian, augmented at the DNN-HMM training stage by the addition of noise from RWCP, AIR and Reverb2014 databases using the reverberation algorithm implemented in the Kaldi framework. The training process took approximately two weeks, and used two systems with 32 CPU cores and 64GB RAM each, with a total of 6 NVIDIA Pascal architecture GPUs. 3-gram language models were adapted using transcription data from the training set. These models contain a vocabulary of approximately 67,000 words in the Spanish model, 60,000 words in French and 48,000 in Norwegian. The Norwegian model uses Bokmål, one of two official written standards of Norwegian. The models were tested using the NIST SCLITE utility[4], which scores the best-path output against the ground-truth for correct words, substitutions, deletions, insertions and an overall word error rate. The results are shown in Table 2

---

[3]https://github.com/kaldi-asr/kaldi/tree/master/egs/aspire
[4]https://github.com/usnistgov/SCTK

2.2   Natural Language Understanding

The Natural Language Understanding (NLU) component translates the output of the ASR into semantic units to be processed by the Dialog Management (DM) component. The main mid-term achievements in the conception of the NLU component concern the development of two multi-lingual methods for topic classification.

First we had to detect the user's end-of-turn pauses. Following previous approaches (e.g. [23, 31]), we addressed this question as a classification problem. As features for classification, we use the temporal profile of speech, and the syntactic and semantic information encoded in the utterances. As classifiers, we used different variants of deep neural models. A distinguished feature of our approach is that we implemented an ASR simulator that allows us to evaluate the sensitivity of the End-of-Turn-Detection (EOTD) with respect to particular characteristics of the speaker (speech profiles), and to errors in the ASR output. This validation step is essential, since in the implementation of dialogue systems an early mistake in any of the system components can produce a negative cascading effect in the performance of the subsequent elements of the pipeline.

Our NLU component is expected to treat user utterances in three + one different languages, i.e. Spanish, French and Norwegian + English. In the literature, such systems are usually referred to as multi-lingual models, and different strategies have been proposed to develop them. Learning becomes a challenging task for a multi-lingual system since languages are diverse in their grammar. Furthermore, the availability and quality of the corpora with which machine learning models are usually trained is not equal for the languages. Thus, we focused on topic classification. Our approach was to create a modular system where information available in one language is transferred or exploited while learning models for the other languages. We implemented two strategies: the first based on the use of the Wordnet semantic network and synsets [18], the second based on parallel corpora.

Wordnet semantic synsets encapsulate information about different senses of commonly used words. This information is language independent, and can thus be used not only to obtain the equivalent word in another language, but also to obtain a set of sense-connected synsets, by computing the closure set that includes the hyperonymies. Thus, once we have a group of sense-connected synsets, we can generate the set of words that represent them for each language.

Our second strategy focuses on training one model for one particular language for which we have quality labeled data. By means of parallel corpora, we then extrapolate the obtained labels to a second language, so as to later train a specific model. Labeling is a tedious work, yet thanks to this strategy, we only have to label in one language while still obtaining a corpus for each of the languages.

2.3   Dialogue Management

The Dialogue Management (DM) component maintains the state and manages the flow of the conversation or, in other words, determines the action a system has to perform at each step.

For the EMPATHIC VC we used a DM providing an advanced management structure based on distributed software agents. It enforces a clear separation between the domain-dependent and the domain-independent aspects of the dialogue control logic. The domain-specific aspects are defined by a dialogue task specification, and a domain-independent dialogue engine executes the given dialogue task to manage the dialogue. The dialogue task specification is defined by a tree of dialogue agents, where each agent is responsible for managing a sub-part of the dialogue. For instance, Figure 1 shows the high level dialogue task structure for the EMPATHIC VC where the "Introduction" agent is responsible for handling the introductory dialogue of the EMPATHIC VC, the "Nutrition" agent is responsible for handling the nutrition dialogues, etc.
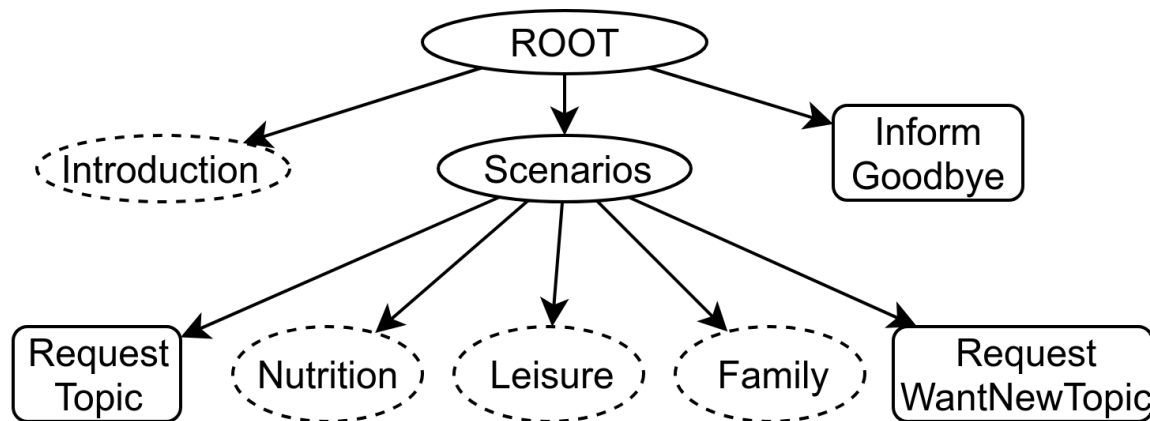
Fig. 1. High level dialogue structure employed by the DM of the EMPATHIC VC.

The aim of the system is to have a virtual coach that helps users in certain aspects of their lives. To this end, we have emulated a GROW (Goal - Reality - Obstacle - Will) [28] coaching model into the DM's dialogue strategy. The GROW model is a structured method based on problem solving, goal setting and goal-orientation. The Model is divided into four phases that propose four questions to guide the user towards obtaining and achieving a goal. These questions are asked in a pre-established order. In the first session, this order must be respected to facilitate the user to follow the thread and be able to explore its goal and the necessary steps to achieve it. In the following sessions, the order can be changed or specific phases be chosen. Using the GROW dialogue model, the following dialogue topics have been implemented:

- **Introduction** dialogues to make the users feel comfortable with the system and to obtain some basic information. This dialogue is carried out the very first time a user dialogues with the system.
- **Sport and Leisure** [25] [27] dialogues based on users' leisure time activities. The aim of these dialogues is to explore users' leisure time activities, and if necessary, nudge them towards a more active lifestyle.
- **Nutrition** [26] dialogues focused on users' nutritional habits. The goal of these dialogues is to explore the nutritional routines of the users and, if necessary, try to make the users vary those routines in order to form nutritional habits that are potentially more healthy.

2.4 Natural Language Generation

The Natural Language Generation (NLG) component maps the abstract dialogue acts provided by the DM to natural language constructs (in Spanish, French or Norwegian), written in orthographic form. So far, the NLG component has been developed from a reduced database of coaching turns. The data have been extracted from video recordings of real user sessions with a professional coach, and some handmade dialogues created by a professional coach. In the process of labeling, two types of labels have been used: (1) based on the GROW coaching model [1], and (2) based on linguistic features needed to construct the text. Considering the restrictions in the amount of training data, the NLG currently represents a rule-based system, using a template-based approach [21]. Future work, however, aims to design a new NLG component based on a seq2seq neural network [6] (note: this plan of implementing a seq2seq neural model only concerns the NLG part of the EMPATHIC VC and will not affect the other components).

2.5 Text-to-Speech Synthesis

A Text-To-Speech (TTS) component converts any text into a spoken message. For EMPATHIC, TTS speaking styles should be fully compatible with the role of the VC communicating with elderly users. The Acapela TTS system employs a range of internally developed technologies, such as unit selection or parametric synthesizers based on Hidden Markov Models (HMM's) or Deep Neural Networks (DNN's). During the project, all of these technologies are adapted for the EMPATHIC communication and compared during evaluation with end-users. Professional speakers will be recorded to capture the communicative role of the VC with the aim to reflect the expressive possibilities of the dialogue system by coherent audio responses to the user's emotional state so as to support the credibility, naturalness and adaptability of the full dialogue chain. So far, Acapela already recorded a Spanish professional speaker enacting a VC for the elderly and trained a snit selection, and initial DNN synthesis systems on this corpus. We run evaluations of this elderly coach style TTS in terms of naturalness and intelligibility. A Spanish coaching TTS voice is already available and integrated in the mid-term version of the EMPATHIC VC. Upon the evaluation of the Spanish system, Acapela will fine-tune the process and develop the French and Norwegian voices.

2.6 Emotional Agents

For initial prototyping purposes (cf. Section 4.4), we used a multi-step creation process (cf. Figure 2) to build five virtual agent coaches (3 female and 2 male) named Natalie, Alice, Lena, Christian and Adam. The first of the steps depended on the origin of the 3D model. The coaches Alice and Adam were created based on 2D images. For this, we had to create their 3D model from a 2D image using CrazyTalk[5] and then exported this to the RLhead format. For the coaches Christian, Natalie and Lena, who were created based on 3D models predefined by iClone[6], we skipped the CrazyTalk step. Second, we imported the RLhead to the Character creator. This tool helped us create realistic-looking 3D human models. We fixed the 3D model design, imported 3D clothing designs and generated humanoid animations with extensive customization tools. After that, a model was imported into iClone. At this step, we used iClone to blend character creation, animation and scene design into a real-time engine, and to edit them in 3DXchange. Next we exported the model and animation using 3Dxchange into the FBX format. Then, we imported the FBX file as a new resource into Unity3D[7]. The transition from iClone to unity degrades the appearance of the model. To solve this, we had to correct the texture and optimize the shader of the model. Still in this step, we added the lip-synchronization using the SALSA plugin. The audio was processed so as to automate four basic mouth positions, which are the basis for lip-sync approximation. Instead of pre-processing or mapping shapes to audio markers, mouth movements were procedurally applied to a minimal set of mouth shapes to provide variation. It uses a combination of waveform analysis and four mouth shapes to produce high-quality lipsync approximation. The last step was to build the WebGL format within the Unity environment.

3 STATUS: TECHNOLOGY INTEGRATION

To integrate the different software components of the EMPATHIC-VC while meeting the challenges of a low-latency interactive system with the additional security, confidentiality and privacy requirements for health information, we took a multi-stage approach. First we defined up front a model for development which uses fully separated containers for each component, with communication between components over sockets and messaging via a global message

---

[5]https://www.reallusion.com/crazytalk/
[6]https://www.reallusion.com/iclone/
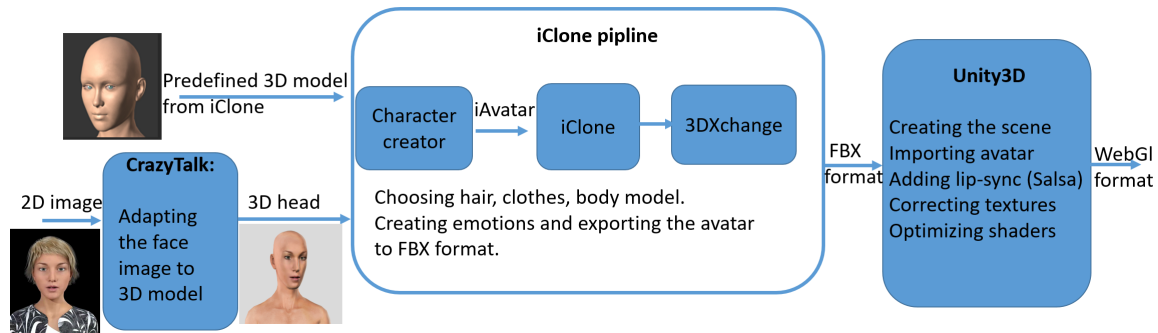[7]https://unity3d.com/fr

Fig. 2. Workflow to generate 3D virtual coaches.

queue. Next, a review process was conducted for each of the components, covering the component container layout and requirements, capabilities and testing approach. Subsequently, a dedicated integration environment was set up with the supporting tools such as the container orchestration and message queue. Finally, each component was tested independently in the integration environment, both initial smoke testing and then testing component inputs and outputs.

Once all components are validated in the integration environment, according to the above described procedure, we will be ready to test the entire system. For the purpose of testing we have split the system into four sub-systems:

- An "inbound" system, using pre-recorded input, joining ASR, NLU and DM.
- An "outbound" system, using pre-generated DM output, joining NLG, TTS and the virtual agents.
- A "user interaction" system, joining the Web UI, Web A/V proxy and also test recordings.
- A "human sensing" system, joining the emotion detection for speech, text, face and gaze, and the biometric authentication.

In addition to the integration of the VC components we started to work on the provision of secure cloud connectivity. For this we have designed a review process covering:

- the secure connection over WebRTC between participant devices and the EMPATHIC-VC system;
- the secure configuration of host servers;
- secure remote administration;
- secure software development practices;
- secure storage and encryption of participant data;
- and physical security at field trial hosting sites.

4 STATUS: USER INTERACTION

In order to ensure acceptance of the EMPATHIC VC, we have to show an added value for the end-user. The creation of this value proposition is an iterative and contrasted process that involves interdisciplinary collaboration [22]. To start this, we analyzed relevant target groups in Spain, France and Norway. From those studies we were able to extract general as well as country-specific end-user traits, leading to an archetypical end-user definition with some specific factors that may vary from country to country.

4.1 Defining the Target Population

As a common definition, we can consider our target population as "Young Olds", aged 65 to 79 years, with a healthy and active life, characterized by the continuation of their former lifestyle after retirement, yet focused more on enjoyment and leisure activities. It is important to highlight the two main indicators used for this definition. The first indicator concerns the healthy life years at birth; this indicator shows the average of life years sans diseases. The second indicator concerns the healthy life expectancy based on self-perceived health; this indicator shows the population's self-perception of health. Thus, it include factors such as economic status, emotional problems and social relations. This is a subjective indicator based on surveys, perception records and/or self-assessments[8].

Therefore, the term "Young Olds" refers to people older than 65, who perceive their health as good or very good. Yet although, their self-perception is good, "Young Olds" may already have some type of disease or sickness[9].

4.2 Understanding the "Young Olds"

Before describing the priorities defined by our analysis, it is worth to highlight four basic recommendations to promote user acceptance of ICT by seniors, defined by the European Active and Assisted Living (AAL) programme. Those recommendations are [10]:

- Provide clear additional value and benefit of the solution;
- Balance between supporting the users and activating them;
- Maintain simplicity on the interaction user-solution;
- Provide joyful experiences;

Those recommendations act as a starting point for integrating priorities into a solution design.

Regarding our analysis, family has been revealed as the main priority across different countries. On average, more than two thirds of seniors with children see them several times per month. This ratio is even higher when children live close by. In addition, the use of ICT tools is significantly higher when it is used for communication purposes with family members (especially children and grandchildren). Therefore, the involvement of family members as part of the solution will strongly increase the acceptance and usability of the solution.

Another key point to promote the ICT use among elderly is to show the clear benefit of a solution. For example, seniors are willing to learn to use ICT tools if this provides more interaction with their younger family members.

As mentioned before, more than half of our target population perceives their health as "good" or "very good", although a common fear is the physical decline associate to ageing. This could be one strong reason why, independent of the country, more than half of the people we talked to perform some type of physical activity on a weekly basis (note: most commonly walking). Supporting those activities, providing motivational strategies, gamification tools and/or professional advice may thus be seen an opportunity to increase the acceptability of ICT. Even better, it would not only support and enrich a leisure activity that is already common among the target population, but also promote "healthy habits" and "well-being".

Following a similar approach, nutrition is an area that combines "joyful" experiences and "healthy habits". Cooking, planning meals, going out to restaurants, increasing expenditure on food etc. are related activities that are more promoted with representatives of the target population than with younger people. In that sense, changes in the nutrition

---

[8]https://ec.europa.eu/eurostat/data/database
[9]Encuesta Europea de Salud en Espana 2017, Pub. Pub. Instituto Nacional de estadística (INE)
[10]http://www.aal-europe.eu/wp-content/uploads/2015/02/AALA_Knowledge-Base_YOUSE_online.pdf

habits can be seen as an indicator for physical or physiological changes. On the other hand, promoting and motivating a healthier nutrition habit can positively influence a human's physical and emotional status.

While motivational factors and physical as well nutritional habits were similar with our target populations in Spain, France and Norway, we did also find differences. Those mainly relate to the familiarity and use of the Internet. Here it was shown that Seniors from Norway have a significantly higher percentage of people that use the Internet on a regular basis, in particularly when compared to Spain (note: France lies in between). This information is relevant, as personal experience with technology is a relevant factor influencing technology acceptance.

4.3 Aspects of User Acceptance

One aspect to focus on in human-agent interaction is the user's level of technology acceptance. The concept was introduced by Davis [4] in the attempt to explain people's acceptance (or not acceptance) of an interactive system. It led to the development of the Technology Acceptance Model (TAM), a questionnaire where acceptance is assessed in terms of a user's perceived usefulness, and perceived ease of use [4] of the system. TAM was extended into TAM2 in 2000 [35], adding two theoretical constructs that accounted for a user's social influence and for how well a user's work goals are supported by the interactive system. TAM2 evolved into the Unified Theory of Acceptance and Use of Technology (UTAUT) [36] and later into UTAUT2, where hedonic motivations (the fun or pleasure derived from using a technology), price values (trade-off between perceived benefits and monetary costs), and habits were added to the original questionnaire, as further determinants theorized to affect user's behavioral intentions and use behavior [37]. Finally, the Almere questionnaire was developed as a further evolvement of UTAUT2 objecting that the latter was developed without accounting for variables that relate to social interaction with robots or virtual agents and without considering seniors as potential users [12, 34]. Following the same reasoning, Hassenzahl developed the AttrakDiff questionnaire[11] [10, 11], a four cluster test, where each cluster, composed of 7 items, was assessing a desired user requirement.

It must be noted that, currently, all the theoretical formulations of questionnaires aiming to assess the user experience of an interactive system are, to a certain extent, dated, since those systems are increasingly more complex, showing humanoid appearance, and human features. Thus, even though the theory for defining user experience is still valid, new concepts have to be accounted for, in order to have a fair assessment of modern interactive systems. This is why, to date, there are no systematic investigations devoted to assessing the role of virtual agents' features exploiting the above mentioned questionnaires. Furthermore, seniors have only been involved in a very limited number of studies on virtual agent's. In the few they have been involved in, it has been shown that they clearly enjoy interacting with a speaking synthetic voice produced by a static female agent (note: these were 65+ aged seniors in good health [3]), and that such seniors are less enthusiastic than impaired people in recognizing the agent's usefulness [40]. The only comparison among user's age we know about, was conducted by Straßmann & Krämer [33]. The study was *"a qualitative interview study with five seniors and six students"* and showed that senior users prefer embodied human like agents over machine or animal-like ones. No information were obtained on the gender of the agents as well as their pragmatic and hedonic features, as advocated by the TAM, UTAUT, and AttrakDiff questionnaires discussed above.

Therefore, the Empathic project aims to *"develop causal models of [agent] coach-user interactional exchanges that engage elders in emotionally believable interactions keeping off loneliness, sustaining health status, enhancing quality of life and simplifying access to future telecare services"*. To this end, an initial research step was the development of

---

[11]AttrakDiff(tm)Internet Resource – http://www.attrakdiff.de.

an ad-hoc questionnaire to assess the pragmatic and hedonic features of the to be developed EMPATHIC VC. The questionnaire was developed through an iterative process that involved several experiments and the exploitation of the theoretical concepts already advocated by the authors of TAM 2, UTAUT2, and AttrakDiff, and the inclusion of new theoretical considerations regarding our users' age. The goal of the questionnaire was also to provide information on the user's preferences regarding the agent's physical and social features, including its face, voice, hairdo, age, gender, eyes, dressing mode, attractiveness, and personality.

The first experiment with the aim to build and evaluate such a questionnaire (and start collecting said data) was conducted in Italy and involved 45 healthy seniors (50% female), aged 65+ [8]. As this study was conducted before any of the agents described in Section 2.6 were built, our stimuli were based upon the four conversational agents proposed by the Semaine project[12], each possessing different personality features able to arise user specific emotional states i.e., Poppy (female, expressing optimism), Obadiah (male, expressing pessimism), Spike (male, expressing aggression) and Prudence (female, expressing a high degree of pragmatism) [20]. For each agent, a video-clip was extracted from the videos available on the Semaine website. In order to contextualize them to the Italian culture they were renamed, using names very popular in the local area (i.e. Serena, Gerado, Pasquale, and Francesca). Agent's names and video durations were carefully assessed by four people. The final set of stimuli consisted of 4 video clips, each 10 secs long showing the agent's half torso, all of the same dimensions, acting as if they were speaking while the audio was mute.

The preferences the target group had towards each of the proposed agents were assessed through a first version of our questionnaire, structured in 4 clusters each containing 7 items devoted to assess the practicality, pleasure, feelings, and attractiveness experienced by participants while watching the agent video-clips. The items proposed in each cluster exploited the theoretical foundation inherent to the UTAUT2, Almere, and AttrakDiff questionnaires. Results showed seniors' positive tendency to initiate an interaction with the agent, with a strong preference towards agents with a positive personality. That is, Francesca and Serena scored always significantly higher than Pasquale and Gerardo for the pragmatic, hedonic and attractiveness features. Although seniors had not been informed of the agents' personality, somehow they had perceived negativity or positivity from the dynamics of their facial expressions, which triggered a behavior of acceptance/rejection, suggesting that participants have preferences for positive facial dynamics. These results, however, required deeper investigations as all agents showing positive facial dynamics were females, hinting towards a potential gender influence on the processing of emotional facial expressions (cf. [17, 24]). Furthermore it must be noted that in this first study agents were dynamically moving their lips as if they would be speaking. Their voice was, however, muted which might have had an effect on people's perceptions.

These aspects were accounted for in a subsequent experiment, conducted again in Italy [7], but this time with agents created with BOTLIBRE[13]. Also these agents, which were selected by 3 experts, showed half their torso, with definite clothing. Again, so as to contextualize, we named them in accordance with typical names for the region, i.e. Michele, Edoardo, Giulia and Clara. Each agent was provided with a different synthetic voice, producing the following sentence pronounced in Italian: *"Hi, my name is Clara / Edoardo / Giulia / Michele. If you want, I would like to assist you with your daily activities!"*. The synthetic voice was created through the website Natural Reader[14]. The voices (recorded using the software Audacity[15]) were embedded into each agent's video-clip which had an average duration of approx. 6 seconds. The proposed agents did not show a particular personality. Care was taken to ensure that no emotion was depicted by their faces and they were shown on the same background, avoiding exaggerated cloths colours or facial features. This

---

[12]http://www.semaine-project.eu/
[13]http://www.botlibre.com
[14]http://www.naturalreaders.com
[15]https://www.audacityteam.org/

second experiment, again devoted to assessing senior's preferences towards each of the proposed agents, exploited a new version of our questionnaire. It had 6 clusters:

- **Cluster 1:** 10 items focusing on the usefulness, usability, and accomplishment of the tasks of the proposed system, i.e. the system's pragmatic qualities (PQ). High scores in the PQ dimensions indicate that the users perceive the systems as well structured, clear, controllable, efficient and practical.
- **Cluster 2:** 10 items focusing on motivations, i.e. the reason why a user should own and use such an interactive system, i.e., the system's stimulating hedonic qualities (HQS). A system receiving high scores in the HQS dimensions is meant to be original, creative, captivating.
- **Cluster 3:** 10 items focusing on how captivating, and, of good taste the system appears, i.e. the system's hedonic qualities of feeling (HQF). A system receiving high scores in the HQF dimensions, is considered presentable, professional, of good taste, and bringing users close to each other.
- **Cluster 4:** 10 items on the subjective perception of the system's attractiveness (ATT), which is the hedonic dimension that gives rise to behaviors as increased use, or dissent, as well as, emotions as happiness, engagement, or frustration.
- **Cluster 5:** 4 items assessing the type of professions seniors would endorse to the proposed agents, among which were welfare, housework, security, and front desk jobs.
- **Cluster 6:** 3 items assessing agent's age range preferences.

Each questionnaire item required a response given on a 5-point Likert scale ranging from 1=strongly agree to 5=strongly disagree (3=I don't know). Since all clusters contained positive and negative items evaluated on a 5-point Likert scale, scores from negative items were corrected in a reverse way. This implies that low scores summon to positive evaluations, whereas high scores to negative ones. Each participant was first asked to provide answers to the items related to demographic information and user technology savviness, then they were asked to watch each agent's video-clip and immediately after to complete the items from the remaining 6 clusters.

Final Results. Our results show that seniors clearly expressed their preference to interact with the female rather than the male agents. This was true for the pragmatic, hedonic stimulation, hedonic feeling and the attractiveness dimensions, and independent from their gender and technology savviness (note: between the two proposed female agents, senior's preferences for Giulia scored statistically significantly higher than those attributed to Clara). Consequently, the data suggests, that seniors' willingness to be assisted by a virtual agent is strongly affected by the gender of the proposed agent, and that, up to now, female agents seem to be the preferred choice. In addition, these two preliminary experiments allowed the definition of a questionnaire on virtual agent acceptance, which also contains a demographic information section and a section on user technology savviness. We call this the Virtual Agent Acceptance Questionnaire (VAAQ). So far, it has been translated from English into French, Norwegian, Spanish, Italian, and German.

4.4 A Simulated Virtual Coach

Various modules of the EMPATHIC VC will require ongoing development and improvement, e.g. to advance speech recognition and language understanding, or foster user experience and acceptance. In order to simulate such modules while they are being developed, the goal was to build and consequently use a simulated Wizard of Oz (WOZ) component. WOZ constitutes a prototyping method that uses a human operator (i.e., the so-called wizard) to simulate non- or only partly-existing system functions. In language-based interaction scenarios, like the ones envisioned by EMPATHIC, WOZ is usually used to explore user responses and the consequent handling of the dialogue, to test different dialogue

strategies or simply to collect language resources (i.e., corpora) needed to train technology components. In EMPATHIC, however, the goal is to use WOZ beyond this traditional prototyping stage, and make it a fallback safety net for situations in which the automated coach may be unable to respond. That is, the goal is to develop a system component, which initially serves as a prototyping tool supporting the research on language-based interaction and dialogue policies, but then becomes an always-on backup channel dealing with those user requests the system is incapable of handling by itself. A first version of this tool has been built and consequently used in several user studies.

4.4.1 Technical Setup. As the specification and integration of the the EMPATHIC VC is still ongoing (cf. Section 3), yet user feedback is urgently needed to inform the design of technology components, we decided to work on two separate WOZ systems. The first one, referred to as the EMPATHIC WOZ Platform, acts as a stand-alone tool, which is being developed independently from the EMPATHIC VC architecture, despite being a testbed for feasibility evaluations regarding the different technologies to be used. In doing so, we were able to perform initial user studies before final decisions on the EMPATHIC architecture were taken. On the contrary, the second WOZ system, referred to as the EMPATHIC WOZ Component, is meant to become a component of the final EMPATHIC coach for which it is implemented and adapted in accordance with the overall EMPATHIC system architecture. This component is currently in development and thus not yet used for user studies.

4.4.2 The EMPATHIC WOZ Platform. Several researchers have worked on WOZ tools before (e.g., [5, 9, 13, 16, 19, 32, 38]), but only few of those tools are openly available for implementation and adaption. One such tool is the WebWOZ Wizard of Oz Prototyping Platform [29][16], which has already been employed by a number of previous projects (e.g., vAssist [30], Roberta Ironside [15]). Building upon the experience of these projects, we used the platform as a core system and implemented the following EMPATHIC-specific adaptations:

- **Audio Transmission and Recording:** Previous versions of WebWOZ required a separate audio channel to transfer a study participant's voice input to the wizard. Usually, such was achieved using some sort of Voice-over-IP tool such as Skype or Google Talk. To avoid 3rd party tool usage we directly integrated audio transmission between a study participant and a wizard via a connection channel based on the WebRTC standard. In our adapted WebWOZ Platform, WebRTC does not only serve as a communication channel but also handles the recording of sessions, allowing for the collection of relevant language resources needed to inform the design of future EMPATHIC language components (e.g., ASR, NLU, DM, etc.). In addition, the technology will be used as a core technology for the final EMPATHIC VC architecture, for which its implementation into the WebWOZ Platform acted as a test case evaluating its feasibility.
- **Video Transmission and Recording:** In addition to the audio transmission channel described above, we integrated video transmission and recording between participant and wizard. Also here, WebRTC acted as the core technology. The video link, which was added to the wizard interface, allows the wizard to see the participant's face, providing important contextual information and thus supporting the cognitively rather demanding task of simulating machine behaviour (cf. Figure 3).
- **Flow-chart in Wizard Interface:** In order to help the wizard follow a systematic interaction path, we further added a flow-chart graph to the wizard interface (see Figure 3). This graph shows the optimal flow of dialogue steps and thus acts as an interaction manual supporting the selection of pre-defined utterances to be sent to a study participant.

---

[16]https://github.com/stephanschloegl/WebWOZ

Fig. 3. EMPATHIC WOZ Platform wizard interface incl. live video feed and flow-chart graph.

- **Web-based Scenario Upload Mechanism:** The definition of pre-defined utterances to be sent to a study participant counts as a key feature of a language-based WOZ tool. In order to speed up the definition of these utterances and at the same time allow developers and interaction designers to work with standard tools, we integrated an Excel (.csv) import feature. A similar feature was already available in earlier implementations of WebWOZ [30], yet such needed root access to the server infrastructure where the platform was hosted. We extended this feature by a simple web-based upload mechanism, so that respective user rights are no longer required.

- **Agent Interface incl. Text-to-Speech Synthesis:** The client interface of the original WebWOZ Prototyping Platform does not offer any agent or avatar feature. Hence, in order to obtain feedback on our EMPATHIC virtual agent designs, we implemented five different agent prototypes (3 female, 2 male) and connected them to the wizard interface (cf. Section 2.6). All agents use similar core features (size, background, facial features, etc.) and integrate, depending on the environment setup, Spanish, French and Norwegian (as well as German, Italian and English) text-to-speech synthesis provided by Acapela (note: female and male agents use different voices).

Fig. 4. EMPATHIC WOZ Platform client interface featuring five different agents to interact with.

With this setup, a human wizard is currently able to remotely control a virtual agent in any of these languages. Lessons learned from first tests using this setup are reported next.

4.5   Lessons Learned from User Studies

A total of 176 WOZ user studies (à 2 sessions) have so far been conducted (i.e., 68 in Spanish, 54 in French, and 54 in Norwegian). The following insights are based on feedback provided by wizards, who simulated the EMPATHIC VC, as well as study participants.

4.5.1   Insights Regarding the Study Setup. Experience has shown that at least two people are required to realistically conduct a WOZ user study – one who acts as a human simulator, i.e. the wizard (usually sitting in a different location), and one who acts as a facilitator, greeting study participants, introducing them to the study purpose, administrating questionnaires, and helping the participants in cases of confusion. A setup without facilitator, i.e. without a second person, did in our case not seem feasible. From a procedural point of view, we further found it imperative that, once the interaction with the simulated agent has started, the facilitator needs to leave the room, because otherwise the participant tends to look at and talk to the facilitator instead of conversing with the actual agent. This behaviour may

be explained by a participant's lack of reassurance when interacting with a novel technology. Taking the additional person out of the room helps eliminate this potential distraction, yet it may also increase a participant's level of anxiety.

4.5.2 Insights Concerning Study Participants. In general, we found that the concept of a virtual agent seems rather frightening to many people of the targeted age group (i.e. aged 65 or older). While we did use face-to-face meetings to overcome this fear as much as possible, it should be noted that for this type of technology anxiety poses a significant challenge, particularly when it comes to the recruitment of study participants. Consequently, recruitment via flyers/posters was difficult (even when conducted in senior centres or elderly homes). However, we found that recommendations coming from other participants who had already taken part and enjoyed the study, helped mitigate the problem. Still, a lot of personal coaching was usually required to make people feel comfortable. Here, our experience has shown that participants needed approx. 10 minutes to 'lose their fear' regarding the technology – in particular, when studies took place somewhere away from peoples' homes or familiar living environments. A technical setup, which would allow studies to be mobile and, thus, be brought to potential study participants, may therefore help circumvent some of this felt uneasiness. With respect to the study inclusion criteria, the studies have shown that elderly people are rather pessimistic when evaluating their personal health status. That is, while initially we were searching for 'healthy' participants aged 65 or older, we had to realize that most representatives of this group would not include themselves due to minor health issues they perceived preclusive (e.g. minor hearing problems). A slight change in wording has helped tackle this problem. What remained difficult, however, was the recruitment of people with depression. As for the interaction, it seemed important that participants thought they would interact with a prototypical system. This helped keep the expectations regarding speed and accuracy low. In this context, the speed with which a simulated system responds may be seen a particular challenge. Especially in cases where the wizard could not use a pre-defined utterance and, thus, had to type a response. An additional challenge with this generation of on-the-fly utterances concerns the great potential for typo's and other mistakes, which are forwarded to the TTS and, consequently, spoken out loud to a study participant. However, being aware of the prototypical status of the system, study participants were rather tolerant towards these types of issues.

4.5.3 Insights Concerning the Dialogue. With respect to the scenarios, participants were usually pre-informed about some of the content to be addressed by the coach so that they could think about relevant topics in advance (e.g. they were told to think about certain health goals they would like to achieve before starting the conversation). Such was necessary to keep the interaction going and reduce the number of "yes/no" answers. Still, in particular with respect to the the nutrition scenario, it was difficult to keep the conversation flowing, as the scenario was looking for personal goals, yet people were often satisfied with their status-quo and, thus, did not find much to talk about. This somewhat restricted the number of available interaction turns, for which we had to slightly shift participants' foci to other nutrition-relevant topics so as to keep the conversation alive. To this end, the pre-defined utterances that were prepared for the wizard seemed rather limited in scope as well as in variation, which significantly increased the use of the (arguably much slower and more error-prone) free-text feature of the platform. A way of making the conversation more flexible, may be found in a speech input interface for the wizard. Yet such a feature, while in consideration for the EMPATHIC WOZ Component, has so far not been integrated. Changing the conversational focus due to missing participant goals also caused some side effects. Particularly in France, participants often felt insufficiently 'coached'; i.e. they had the impression that the virtual coach wanted their information, yet did not provide them with any advice on what they should do in order to change their habits. Finally, from a conversational point of view, we found that different types of back-channelling (i.e. approving a participant's input) had a significant influence on the 'smoothness'

of the conversation. That is, while rather basic approval utterances such as "interesting" or "good" seemed to distort the conversation, other strategies which re-used participants' words or sentence structures (e.g. Participant: "I like to walk 2 hours every day"; Agent: "You walk 2 hours every day?") helped in keeping participants engaged and consequently the conversation flowing.

4.5.4 Insights Regarding Administered Questionnaires. Regarding the study closure and debriefing phase, study participants perceived the number and length of the administered questionnaires as too much (note: this also included our own questionnaire whose development was presented in Section 4.3). In addition, some of the questions were hard to understand or potentially ambiguous and, thus, study participants found them to be rather difficult to answer (note: this led to a high number of "I don't know" answers). In particular, questions concerning more abstract concepts or terms as well as feelings not usually connected to a human-agent dialogue (e.g. hedonic features) seemed to pose problems. Often it was the wording, which played a significant role here and might, thus, be adapted in future studies. Finally, unexpectedly high depression scores found with healthy study participants caused some doubts regarding the used scales, which will require further investigation.

4.5.5 Insights Regarding Technical Issues. From a technical point of view, a re-occurring problem seemed to be connected to the session management, which influenced the connection between the wizard and respective client. The issue is currently under investigation, with a viable solution expected to be implemented in the upcoming weeks. A second challenge concerned the technical support during user studies. Given that the entire EMPATHIC WOZ Platform is currently hosted in a location (i.e. the UK) different from where studies are conducted, and technical support provided from yet another location (i.e. Spain), technical problems often caused long delays. Local setups incl. respective technical support – as it is planned for the final EMPATHIC VC – may thus be considered in future studies.

5 CONCLUSION

In this paper we reported on the mid-term achievements of the H2020 EMPATHIC (*Empathic, Expressive, Advanced Virtual Coach to Improve Independent Healthy-Life-Years of the Elderly*) project. Those achievements include, on the one hand, significant efforts put into the development and integration of various technical components required to run a modern, virtual agent based dialogue system geared towards supporting elderly people in their daily activities; on the other hand, a number of user studies aimed at understanding said user group (i.e., healthy seniors aged 65+) and their preferences with respect to agent acceptance. Results of these efforts are manifested in a working WOZ prototyping tool for simulating human-agent interaction, a multi-lingual (i.e. Spanish, Norwegian, French, Italian, German and English) questionnaire assessing virtual agent acceptance, and a better understanding of particular technical challenges inherent to the provision of a web-based, secure, responsive and reliable agent platform.

6 ACKNOWLEDGMENTS

REFERENCES

[1] Graham Alexander. 2010. Behavioural coaching–the GROW model. *Excellence in coaching: The industry guide* (2010), 83–93.

[2] Luisa Brinkschulte, Natascha Mariacher, Stephan Schlögl, Maria Inès Torres, Raquel Justo, Javier Mikel Olaso, Anna Esposito, Gennaro Cordasco, Gérard Chollet, Cornelius Glackin, et al. 2018. The EMPATHIC project: building an expressive, advanced virtual coach to improve independent healthy-life-years of the elderly. In *SMARTER LIVES 2018: digitalisation and quality of life in the ageing society*. Universität Inssbrück, 36–52.

[3] Gennaro Cordasco, Marilena Esposito, Francesco Masucci, Maria Teresa Riviello, Anna Esposito, Gérard Chollet, Stephan Schlögl, Pierrick Milhorat, and Gianni Pelosi. 2014. Assessing voice user interfaces: the vAssist system prototype. In *2014 5th IEEE Conference on Cognitive Infocommunications*. IEEE, 91–96.

[4] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* (1989), 319–340.

[5] Richard C Davis, T Scott Saponas, Michael Shilman, and James A Landay. 2007. SketchWizard: Wizard of Oz prototyping of pen-based user interfaces. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*. ACM, 119–128.

[6] Ondřej Dušek and Filip Jurcicek. 2015. Training a natural language generator from unaligned data. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Vol. 1. 451–461.

[7] Anna Esposito, Terry Amorese, Marialucia Cuciniello, Antonietta M Esposito, Alda Troncone, Maria Inés Torres, Stephan Schlögl, and Gennaro Cordasco. 2018. Seniors' Acceptance of Virtual Humanoid Agents. In *Italian Forum of Ambient Assisted Living*. Springer, 429–443.

[8] Anna Esposito, Stephan Schlögl, Terry Amorese, Antonietta Esposito, Maria Inés Torres, Francesco Masucci, and Gennaro Cordasco. 2018. Seniors' sensing of agents' personality from facial expressions. In *International Conference on Computers Helping People with Special Needs*. Springer, 438–442.

[9] Armin Fiedler and Malte Gabsdil. 2002. Supporting progressive refinement of Wizard-of-Oz experiments. In *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems Workshop*, Vol. 6.

[10] Marc Hassenzahl. 2004. The interplay of beauty, goodness, and usability in interactive products. *Human-computer interaction* 19, 4 (2004), 319–349.

[11] Marc Hassenzahl. 2018. The thing and I: understanding the relationship between user and product. In *Funology 2*. Springer, 301–313.

[12] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2010. Assessing acceptance of assistive social agent technology by older adults: the almere model. *International journal of social robotics* 2, 4 (2010), 361–375.

[13] Christopher D Hundhausen, Anzor Balkar, Mohamed Nuur, and Stephen Trent. 2007. WOZ pro: a pen-based low fidelity prototyping environment to support wizard of oz studies. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2453–2458.

[14] Thomas B.L. Kirkwood. 2005. Understanding the odd science of aging. *Cell* 120:4 (2005), 437–447.

[15] Minha Lee, Stephan Schlögl, Seth Montenegro, Asier López, Ahmed Ratni, Trung Ngo Trong, JM Olaso, F Haider, G Chollet, K Jokinen, et al. 2017. First time encounters with Roberta: a humanoid assistant for conversational autobiography creation. In *eNTERFACE'16, July 18th-August 12th 2016, Enschede, the Netherlands*.

[16] David V Lu and William D Smart. 2011. Polonius: A Wizard of Oz interface for HRI experiments. In *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 197–198.

[17] Abigail A Marsh, Nalini Ambady, and Robert E Kleck. 2005. The effects of fear and anger facial expressions on approach-and avoidance-related behaviors. *Emotion* 5, 1 (2005), 119.

[18] George A Miller. 1995. WordNet: a lexical database for English. *Commun. ACM* 38, 11 (1995), 39–41.

[19] Cosmin Munteanu and Marian Boldea. 2000. MDWOZ: A Wizard of Oz Environment for Dialog Systems Development.. In *LREC*. Citeseer.

[20] Magalie Ochs, Radosław Niewiadomski, and Catherine Pelachaud. 2010. How a virtual agent should smile?. In *International Conference on Intelligent Virtual Agents*. Springer, 427–440.

[21] Alice H Oh and Alexander I Rudnicky. 2000. Stochastic language generation for spoken dialogue systems. In *ANLP-NAACL 2000 Workshop: Conversational Systems*. 27–32.

[22] Claudia Pagliari. 2007. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. *J. of medical Internet research* 9, 2 (2007).

[23] Matthew Roddy, Gabriel Skantze, and Naomi Harte. 2018. Investigating Speech Features for Continuous Turn-Taking Prediction Using LSTMs. *arXiv preprint arXiv:1806.11461* (2018).

[24] Mark Rotteveel and R Hans Phaf. 2004. Automatic affective evaluation does not automatically predispose for arm flexion and extension. *Emotion* 4, 2 (2004), 156.

[25] Susana Sayas. 2018. *Dialogues on Leisure and Free Time*. Technical Report DP3. Sayasalud and Empathic project.

[26] Susana Sayas. 2018. *Dialogues on Nutrition*. Technical Report DP1. Sayasalud and Empathic project.

[27] Susana Sayas. 2018. *Dialogues on Physical Exercise*. Technical Report DP2. Sayasalud and Empathic project.

[28] Susana Sayas. 2018. *Rationale and basis for the structure of the dialogue between the user and the virtual coach. To establish the concept of goal setting in coaching: SMART*. Technical Report D.R1_2. Sayasalud and Empathic project.

[29] Stephan Schlögl, Saturnino Luz, and Gavin Doherty. 2013. WebWOZ: A Platform for Designing and Conducting Web-based Wizard of Oz Experiments. In *Proceedings of the SIGDIAL 2013 Conference*. 160–162.

[30] Stephan Schlögl, Pierrick Milhorat, Gérard Chollet, and Jérôme Boudy. 2014. Designing language technology applications: a Wizard of Oz driven prototyping framework. In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*. 85–88.

[31] Matt Shannon, Gabor Simko, S yiin Chang, and Carolina Parada. 2017. Improved end-of-query detection for streaming speech recognition. In *Proceedings of Interspeech 2017*. 1909–1913.

[32] Jan Smeddinck, Kamila Wajda, Adeel Naveed, Leen Touma, Yuting Chen, Muhammad Abu Hasan, Muhammad Waqas Latif, and Robert Porzel. 2010. QuickWoZ: a multi-purpose wizard-of-oz framework for experiments with embodied conversational agents. In *Proceedings of the 15th international conference on Intelligent user interfaces*. ACM, 427–428.

[33] Carolin Straßmann and Nicole C Krämer. 2017. A categorization of virtual agent appearances and a qualitative study on age-related user preferences. In *International Conference on Intelligent Virtual Agents*. Springer, 413–422.

[34] Christiana Tsiourti, Emilie Joly, Cindy Wings, Maher Ben Moussa, and Katarzyna Wac. 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*. ICST (Institute for Computer Sciences, Social-Informatics and âĂę, 57–64.

[35] Viswanath Venkatesh and Fred D Davis. 2000. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science* 46, 2 (2000), 186–204.

[36] Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS quarterly* (2003), 425–478.

[37] Viswanath Venkatesh, James YL Thong, and Xin Xu. 2012. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS quarterly* 36, 1 (2012), 157–178.

[38] Michael Villano, Charles R Crowell, Kristin Wier, Karen Tang, Brynn Thomas, Nicole Shea, Lauren M Schmitt, and Joshua J Diehl. 2011. DOMER: A Wizard of Oz interface for using interactive robots to scaffold social skills for children with Autism Spectrum Disorders. In *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 279–280.

[39] Donald Craig Willcox, Giovanni Scapagnini, and Bradley J. Willcox. 2014. Healthy aging diets other than the Mediterranean: A focus on the Okinawan diet. *Mechanisms of Ageing and Development* 136-137 (2014), 148–162.

[40] Ramin Yaghoubzadeh, Marcel Kramer, Karola Pitsch, and Stefan Kopp. 2013. Virtual agents as daily assistants for elderly or cognitively impaired people. In *International Workshop on Intelligent Virtual Agents*. Springer, 79–91.