# Cloud Services for Culture Aware Conversation: Socially Assistive Robots and Virtual Assistants

Carmine Recchuto<sup>1</sup>, Luna Gava<sup>1</sup>, Lucrezia Grassi<sup>1</sup>, Alberto Grillo<sup>1</sup>, Marta Lagomarsino<sup>1</sup>, Davide Lanza<sup>1</sup>, Zijian Liu<sup>1</sup>, Chris Papadopoulos<sup>2</sup>, Irena Papadopoulos<sup>3</sup>, Antonello Scalmato<sup>4</sup>, Antonio Sgorbissa<sup>1</sup>

Abstract—This paper introduces a new Cloud platform providing services for culturally competent interaction, that has been developed to expand the capabilities of Socially Assistive Robots and virtual assistants interacting with older persons. The rationale behind the proposed architecture is discussed, by outlining key principles as well as the functionalities provided, with a specific focus on verbal interaction. Three case studies, the humanoid robot Pepper, a robotic medicine dispenser Pillo, and a custom-built Android-based virtual assistant, are analyzed in detail, by showing how robots and other assistants may easily access culturally competent Cloud services to expand their interaction capabilities. Transcripts of conversations are reported and commented, in order to outline both the positive features and the limitations of the system.

### I. Introduction

International reports show that the market of social and household robots is going to expand in the next years: the most relevant applications include assistive, educational and entertainment robots, with an expected Compound Annual Growth Rate (CAGR) of more than 22% in the period from 2019 to 2024<sup>12</sup>. By focussing on assistive applications and extending the analysis to smart mobile devices, the healthcare virtual assistant market is expected to grow at a CAGR of 34.6% from 2019 to 2024. Finally, the Cloud Robotics market is expected to expand with a CAGR of about 28.1% from 2018 to 2022.

These substantial growth rates are not a surprise: since the possibility of designing small, low-cost robots or home assistants has become a reality, the number of applications and producers is expected to increase rapidly. Hypothesizing a future in which a plethora of ubiquitous robotic assistants will populate our homes is no more a dream, therefore opening unprecedented avenues for research in all fields of Robotics and AI. However, low-cost robots require advanced computational capabilities, which can be reasonably guaranteed only by outsourcing computation: this motivates the need for more advanced Cloud services for human-robot interaction and, more generally, for AI. The AI Cloud services provided by Google<sup>3</sup> and Microsoft<sup>4</sup> (e.g.,

<sup>1</sup>University of Genova, Via Opera Pia 13, 16145 Genova, Italy.

for machine learning, vision, audio, and natural language processing) are just a few examples - but the field is expected to expand in unprecedented directions in the next few years.

This article describes the Cloud services developed in the framework of the CARESSES project<sup>5</sup>, aimed at designing the first Socially Assistive Robots that can assist older people and adapt to the culture of the individual they are taking care of. In CARESSES, the Pepper platform developed by SoftBank Robotics has been programmed to help the users in many ways, including reminding them to take their medication, encouraging them to keep active, helping them keep in touch with family and friends. The unique feature of our approach, which follows results in the field of Transcultural Nursing [1] and Culturally Competent Healthcare [2], is that each action is performed with attention to the older person's customs, cultural practices and individual preferences [3], [4], [5], [6].



Fig. 1: Pepper, Pillo, and an Android virtual assistant

The software architecture that has been used to control Pepper in CARESSES, tested for 8 months with care homes residents in UK and Japan<sup>6</sup>, has been purposely designed to allow for a Cloud-based implementation, where the instructions to achieve a culturally competent behaviour are received from the Cloud, whereas the onboard sensorimotor controllers are required to implement the appropriate behaviours. Thanks to this, Cloud services are not only usable by Pepper: cultural competent behaviour is virtually accessible by any smart device which is equipped with a network connection, and has the onboard functionalities to implement the in-

<sup>&</sup>lt;sup>2</sup>University of Bedfordshire, University Square, Luton, LU1 3JU, UK.

<sup>&</sup>lt;sup>3</sup>Middlesex University, The Burroughs, Hendon, London. NW4 4BT, UK.

<sup>&</sup>lt;sup>4</sup>Pillo Labs S.r.l., Torri Piane, Via Pietro Chiesa, 16149 Genova, Italy. Contact author: antonio.sgorbissa@unige.it

¹https://www.marketsandmarkets.com/Market-Reports/household-robot-market-253781130.html

<sup>&</sup>lt;sup>2</sup>https://www.marketsandmarkets.com/Market-Reports/assistive-robotics-market-37247851.html

<sup>&</sup>lt;sup>3</sup>https://cloud.google.com/

<sup>4</sup>https://azure.microsoft.com/

<sup>&</sup>lt;sup>5</sup>CARESSES stands for Culturally-Aware Robots and Environmental Sensor Systems for Elderly Support, http://caressesrobot.org

<sup>&</sup>lt;sup>6</sup>Results of experiments with care home residents are expected to be published soon, before UR2020.

structions it receives. Figure 1 shows Pepper, the medicine dispenser Pillo<sup>7</sup>, and a virtual assistant implemented as an Android app, three case studies considered in this article.

Cloud services for Robotics have received a increasing attention in the last years. The work described in [7] integrated the Google Goggles image recognition service into a Cloudbased system for robot grasping. The RoboEarth project introduces a "World Wide Web for robots" that provides access to knowledge for different robotic applications ranging from mobile navigation to grasping and manipulation [8]. The DAvinCi project introduces a ROS based Cloud Computing framework for service robots [9], with a specific focus on taking advantage of scalability and parallelism provided by Cloud computing to address mapping and navigation tasks. Collective Robot Learning by using the Cloud has been explored: examples are the Lightning framework, that indexes trajectories from many robots over many tasks, using the Cloud for parallel planning and trajectory adjustment [10], and the RoboBrain database, that learns from publicly available Internet resources, computer simulations and reallife robot trials [11].

A Cloud Robotics solution aimed at Socially Assistive Robots (SAR) for healthy aging has been implemented in [12]: the system is able to provide location-based services exploiting Ambient Assisted Living technologies. In a similar way, an Internet-of-Robotic-Things system architecture has been implemented for controlling a SAR in [13], with the final aim of alleviating behavioral disturbances of people with dementia with a personalized and context-aware approach. In [14], context-aware conversational Cloud services were developed in order to embed a SAR with communication capabilities.

In spite of the existing attempts in Literature, no previous system addressed the problem of providing culturally competent Cloud services. The CARESSES Cloud is the first attempt ever to enable virtually any device equipped with a network interface to access culturally relevant knowledge, which can be exploited to perform tasks in a culturally sensitive way or simply for talking with the user to acquire more knowledge about him/her.

Section II introduces the Cloud-based architecture and the basic principles of culturally competent human-robot interaction. Section III describes in detail the services provided for culturally competent verbal interaction with the user, which are the focus of this article. Section IV discusses three case studies: Pepper, Pillo and an Android virtual assistant. Section V discusses experiments, commenting step by step the transcript of a sample interaction. Conclusions follow.

## II. THE CARESSES ARCHITECTURE

The CARESSES architecture has been designed to increase the modularity of the system and to allow for a Cloud-based implementation. It is made up of three main blocks:

 Culture-Aware Human-Robot Interaction Module, CAHRIM. This is responsible for implementing

<sup>7</sup>Pillo is designed by Pillo Health and commercialized in USA by Black&Decker as Pria, https://pillohealth.com

- low-level sensorimotor behaviours, which may significantly differ depending on the hardware of the platform considered (e.g., Pepper, Pillo, or an Android smartphone). The only constraint is that CAHRIM must be capable of speech acquisition and synthesis, possibly relying on appropriate services for speech-to-text and text-to-speech translation. CAHRIM receives sentences to be translated into speech from the Cultural Knowledge Base (CKB) or sensorimotor actions to be performed from the Culturally Sensitive Planner and Execution Module (CSPEM), and it returns information to update CKB and to notify CSPEM about the status of the robot, the environment and the user, possibly acquired through sensors.
- Cultural Knowledge Base (CKB). This encodes quantitative and qualitative knowledge about different cultural groups, by representing both cultural traits which are shared within a group<sup>8</sup> as well as individual characteristics of the user. To avoid stereotypical representations of people and cultures, over the course of time the system is able to update the individual cultural profile based on the individual preferences, customs, beliefs, behaviours, and needs, acquired from the user during the interaction [4], [6]. The core of the CKB is an OWL2 ontology, which encodes elements that may play a key role in daily life and may vary in different cultures, and a Bayesian network, which allows the CKB to reason about the knowledge acquired during the interaction in a probabilistic way, to avoid drawing rigid conclusions about what the person says. The ontology includes chunks of sentences that can be automatically assembled to talk with the user by exploiting the hierarchical structure of the ontology, as well as algorithms for controlling the flow of the dialogue. CKB sends sentences to CAHRIM to be translated into speech, it receives information from CAHRIM to update the ontology and the Bayesian network, and sends cultural parameters to CSPEM which can be used for planning.
- Culturally Sensitive Planning and Execution Module (CSPEM). Depending on the goal set by the user or

<sup>8</sup>In CARESSES we considered three different cultural groups to populate the ontology: English, Indian and Japanese

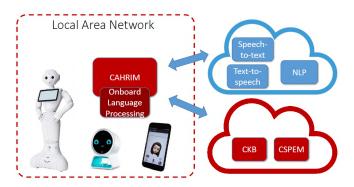


Fig. 2: CARESSES architecture

by the system itself, the CSPEM chooses a sequence of actions to be executed, possibly by taking into account cultural preferences in the way actions should be performed. CSPEM sends actions to CAHRIM and monitors their execution depending on the status of the robot, the environment and the user, and retrieves culturally appropriate goals, operators, actions, and their parameters from the CKB.

Figure 2 shows the relation between the three components above. Whereas CKB and CSPEM are part of the Cloud, CAHRIM is locally executed (onboard the robot/home assistant or somewhere else in the same LAN). The Figure also shows that CAHRIM needs to be capable of speech acquisition and synthesis, but these functionalities shall not necessarily be performed onboard and may be outsourced: in all our implementations, this constraint is met by connecting CAHRIM to dedicated Cloud services.

It should be remarked that this article focusses on the services provided by CKB for culturally competent verbal interaction with the user, and ignores the planning services provided by CSPEM, for which the reader can refer to [5].

Finally, it should be remarked that using CARESSES Cloud services for culturally competent conversation does not exclude the use of other Natural Language Processing systems (e.g., DialogFlow<sup>9</sup> or Watson<sup>10</sup>), but can complement them or even be merged with an onboard commandbased interface that may depend on the specific platform considered (Onboard Language Processing in Figure 2). For instance, the medicine dispenser Pillo is embedded with a vocal interface allowing the user to safely use the main functionalities for which the robot was created in the first place, i.e., dispensing the right pills to the right person at the right time. CARESSES Cloud services are not meant at substituting the conversational capabilities that the robot is already equipped with, but rather provide a backup plan whenever Onboard Language Processing is not able to interpret what the user is saying (which is expected to happen often, in a task-oriented robot as Pillo). In most commercial systems, the backup plan is simply to connect to Wikipedia or other websites; CARESSES Cloud services provide an off-the shelf, culturally competent alternative allowing any robot or home assistant to talk with the user about thousands of different topics. However, as the reader can expect and as clarified in the next Section, the huge number of topics that the system is able to address (more than 2500 in the current prototype) is counterbalanced by poorer performance in understanding the exact meaning of the user's utterance, when compared with more sophisticated systems for conversational NLP (i.e., we are not claiming to have solved the Turing Test).

# III. CULTURALLY COMPETENT VERBAL INTERACTION

The flowchart in Figure 3 shows the basic principle of conversation using Cloud services described in this article, which expands basic ideas previously presented in [15].

The shaded blocks on the right correspond to Onboard Language Processing algorithms, that are responsible to take decisions locally – including the interpretation of task-based commands that the user may issue to require specific, machine-dependent tasks (e.g., "clean the room" for a robotic vacuum cleaner or "give me my medicines" for a robotic pill dispenser).

The blocks on the left correspond to services provided by the Cloud, which are typically machine-independent and are aimed at a culturally competent verbal interaction with the user about a huge number of topics (also referred to as chitchatting), using and updating the information included in the CKB depending on what the user asks and says.

Red and blue arrows outline the fact that the system produces a mixed-initiative verbal interaction: from time to time, both the user (red arrows) and the robot (blue arrows) can take the initiative in leading the conversation. This is aimed at producing dialogues that overcome the typical "command-only barrier" [16] which is the major limitation of home assistants and most robotic systems, typically waiting for a "directive speech act" [17] from the user ("robot, play a song!") which is then reactively executed. On the opposite, the proposed Cloud services – in spite of their limitations in NLP, which is not the focus of this work – are able to produce rich and goal-oriented dialogue patterns, where the robot itself can execute different kinds of speech acts (examples in the following).

Finally, sentences in dark red inside blocks correspond to aspects which are currently subject to improvements, for future expansions of the system.

In order to understand the dynamics of verbal interaction, it is useful to follow the flow of execution from the start. Suppose a robot/home assistant that is able to connect to the Cloud and has been embedded with Onboard Language Processing capabilities to perform a set of machine-dependent tasks, and typically includes a command-based interface able to interpret commands issued by the user to perform one of such tasks. Initially, the system is waiting for the user's input (after introducing itself and its capabilities, if programmers wish so). When the user takes the initiative and says something (red arrow), Onboard Language Processing algorithms may recognize it: in this case, a verbal/not verbal interaction (Block 1) to address the user's requests will start, possibly using cultural parameters in the CKB to suggest the most appropriate options. For instance, if an Italian user asks "I want to listen to music", the developers may program a robot to suggest the most popular Italian songs, by taking this information from the CKB (this is not the focus of this article, and it is up to developers who are in charge of implementing the onboard robot's behaviour).

The Cloud services come into play whenever the Onboard Language Processing is not able to recognize anything relevant in what the user says: in this case, the sentence is sent to the Cloud (Block 2), waiting for suggestions on how this sentence should be handled. As an aside, please remember that a similar backup plan is typically adopted by commercial home assistants, which may connect to Wikipedia or other

<sup>9</sup>https://https://dialogflow.com/

<sup>&</sup>lt;sup>10</sup>https://www.ibm.com/watson

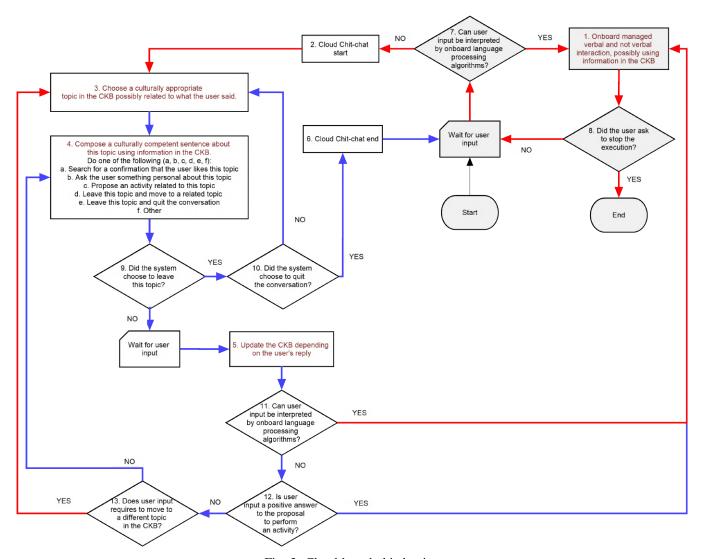


Fig. 3: Cloud-based chitchatting

websites to find information that allows the system to avoid – as much as possible – the most frustrating answer ever: "Sorry, I don't understand that". The CARESSES Cloud service, however, will not simply get a sentence from an online database, but will rather take this an opportunity to start a chitchatting pattern allowing the system to know more about the user and her cultural preferences, and update the CKB accordingly. This enables a more natural response, given that humans seldom respond to questions with the answer "According to Wikipedia...".

Basically, the CARESSES CKB is populated with thousands of concepts related to daily life activities, objects, social relations, etc. which are organized in an OWL2 ontology: instances of the ontology includes, among the other properties, a value which describes the "likeliness" that the user of a given culture is familiar with that concept [6]. Then, depending on the users' utterance, the system selects the topic which is more likely to be appropriate for that user in that context (Block 3): currently, this is simply done by extracting keywords from the sentence, and by selecting the

topic that matches the best with keywords and with the user's cultural profile (this mechanism is subject to improvement).

When the topic has been selected, a loop is started, which may involve different speech acts by the robot and by the user, with the initative taken by the robot (Block 4). For instance, suppose that the selected topic is "Football": the robot may execute Assertive speech acts (e.g., "I know that football is the most popular sport in Italy"), Directives (e.g., searching for a confirmation such as "Do you like football" or proposing an activity such as "Do you want me to set a reminder for the football match"), Expressives (e.g., "I am happy to hear that you enjoy watching football matches"), Commissives (e.g., "I will remember what you told me."). The conversation moves along the same topic (blue arrows) as long as the system itself does not decide to move to something different (answer "NO" in Diamond 9), the user's reply cannot be understood by the onboard command interpreter (answer "NO" in Diamond 11), it does not require to perform an activity (answer "NO" in Diamond 12), and it does not contain cues to move to a different topic

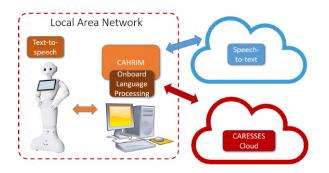


Fig. 4: Pepper and the CARESSES Cloud. Software architecture.

selected by the user (answer "NO" in Diamond 13).

The rationale for picking and composing sentences exploiting the hierarchical structure of the ontology and the sequence in which sentences are chosen is not the purpose of this article. Some more details can be found in [6], where the algorithm for exploring the CKB during chitchatting is described in details, as well as the rationale to update the CKB depending on the user's reply (Block 5), to move to a different topic (answer "YES" in Diamond 9) or to end chitchatting when it is considered appropriate to do so (answer "YES" in Diamond 10).

Very important, during the loop, the user is always allowed to take the initiative by saying something completely different from what the system is expecting (answer "YES" in Diamond 11 or 13), thus forcing the system to deviate from a pre-established dialogue pattern and guaranteeing a dynamism which is unprecedented in off-the-shelf robots and home assistants [16].

# IV. CASE STUDIES: PEPPER, PILLO, VIRTUAL ASSISTANT

## A. Pepper

Pepper (SoftBank Robotics) is a social humanoid robot, with a height of 120 cm and a mobile base endowed with omnidirectional wheels, that has been specifically conceived for interaction with people. It has text-to-speech and speech-to-text abilities, and it is equipped with a 10.1" touch screen display mounted on its chest, that can be used for enhancing the interaction, showing some additional information. Arms and hands are not actually used for grasping and manipulating objects, but they have the main purpose to increase naturalness and pleasantness during the verbal interaction.

In the context of the CARESSES project, the user's utterances were acquired by using Cloud-based speech-to-text services. More in details, Google Speech-to-Text APIs have been integrated with the system, by developing a custom Pepper app for speech recognition. The internal *ALText-ToSpeech* module was used to allow the robot to speak, relying on the Nuance speech engine. Moreover, in order to further ease the interaction with the user, the embedded tablet has been used for acquiring the user's input when the speech-to-text services fail.

In the Pepper implementation, an external PC is required for running a local instance of CAHRIM that connects with the CARESSES Cloud and executes a keyword-based language processing algorithm, as described in Chapter III and shown in Figure 4. If the user's sentence can be interpreted as a task-oriented command, it is handled locally (Block 1 in Figure 3). Otherwise, the sentence is sent to the Cloud where the appropriate conversation topic is started (Block 2 in Figure 3)

#### B. Pillo

The robot Pillo has been purposely designed for reminding the user about medications to be taken, being also able to answer health- and wellness-related questions. However, the commercial version of Pillo cannot engage the user in a proper conversation: thus, the connection to the CARESSES Cloud may be a key factor to extend its current verbal capabilities.

Pillo can be controlled by using REST API [18], a software architectural style that is based on the concept of "Resource", i.e. any object the API can provide information about. Thus, the integration between Pillo and the CARESSES Cloud has required a more complex architecture than Pepper and the Android virtual assistant, shown in Figure 5. With reference to the Figure, the flow of communication is the following:

- 1) When Pillo is in *listen* mode, the user's sentence is locally recorded and sent to the Pillo Server.
- 2) The Pillo servers, relying on external speech-to-text services, convert the audio file in a text file. As for the other two test cases, language processing algorithms detect if the sentence should be directly handled by the device.
- 3) If not, the sentence is sent to an additional REST-based Server, which acts as a bridge between the Pillo Server (based on REST API) and the CARESSES Cloud (based on a TCP protocol for communication).
- 4) The user sentence reaches the CARESSES Cloud, where a culture-aware topic is selected.

Steps 5 to 8 go to the opposite way, transmitting the Cloud

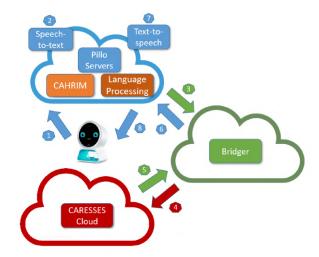


Fig. 5: Pillo and the CARESSES Cloud. Software architecture.

reply back to the Pillo servers, where they are converted into an audio file and sent to the robot.

#### C. Android Virtual Assistant

The Android Virtual Assistant is able to perform speech-to-text conversion, speech understanding, text-to-speech conversion. These capabilities are achieved respectively: connecting to the Google Automatic Speech Recognition (ASR) Cloud using Google APIs for Android for speech-to-text conversion; using Onboard Language Processing as well as the CARESSES Cloud for speech understanding; using Google APIs for Android text-to-speech conversion.

Figure 6 on the bottom depicts the general software architecture, which resembles the one used for the Pepper robot case study, without the need for an external PC. The Figure also shows that a different culture will entail not only different sentences from the Cloud but also Virtual Characters with different ethnicity and cultural heritage.

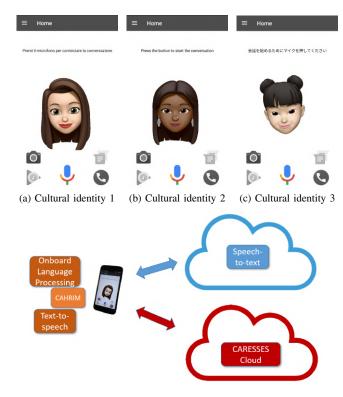


Fig. 6: Android Virtual Assistant and the CARESSES Cloud. Software architecture.

# V. EXPERIMENTS

The system for culturally competent conversation described in this article has been extensively tested from February to September 2019 with care home residents in UK and Japan. During each experimental session, residents were left alone with Pepper for 3 hours, for a total of 6 sessions. Participants were free to interact with Pepper as they preferred (or not to interact at all), by using its functionalities (e.g., to watch videos or listen to news) or simply chitchatting with the robot about a variety of topics (e.g., their childhood, their view about religion and spirituality, their family, etc.).

Evaluation results will be published in the next months. Currently, the system allows the user to talk about more than 2500 different topics in English, Japanese, and Italian, each topic being associated with a number of culturally competent sentences checked by Transcultural Nursing experts, that can be automatically composed exploiting the hierarchical structure of the ontology to achieve a higher degree of variability and reduce the occurrence of repetitions.

Selected software components used during experiments are already available as Open Source [19], and the remaining ones will be released according to the CARESSES release plan. Specifically, starting from the end of September 2020, Cloud services will be freely accessible by any robot or smart devices wishing to do so.

A detailed analysis of the time required by a device to access Cloud services are beyond the scope of this article: benchmarking activities measured a latency of 75ms on average for sending a text to the CARESSES Cloud, processing text, and receiving back an answer from the Cloud. Such latency may have small fluctuations depending on the fact that the CKB needs to be updated or not as a consequence of the interaction. Notice also that, in case of Pillo, the delay that can be observed is significantly higher, but this is not due to the access to the CARESSES Cloud: rather, steps 2 and 7 in Figure 5 requires a longer time for speech-to-text and text-to-speech translation.

The rest of this Section discusses preliminary experiments performed with the Android-based virtual assistant and the Pillo robot, which have been introduced in the previous Sections, in order to give the reader an idea of how verbal interaction may look like when using CARESSES Cloud services. Specifically, in order to produce comparable results, we asked two students to interact with the virtual assistant and with Pillo by giving them a small set of simple instructions: (i) they shall pretend to be Indian persons who identify with the Indian culture, and give answers coherently with this assumption when required; (ii) they shall introduce two topics of conversation in sequence: the "Diwali" festival (i.e., a popular Indian tradition) and "Sport"; (iii) if the system proposes to perform an activity, they shall refuse the offer.

Figure 7 reports transcripts of their conversation with the Android virtual assistant (left column) and with Pillo (right column). The division of each column into different rows correspond to the introduction of the "Diwali" topic (second row) and the "Sport" topic. For the reader's convenience, a video is attached to this article, showing the interaction corresponding to transcripts as well as a similar interaction performed with the Pepper robot<sup>11</sup>.

Red numbers refer to transitions between different states in the flowchart in Figure 3. For instance, in the first column/second row, the sequence "(4-4b) I would like to learn from you about the significance of the Diwali festival" corresponds to the system asking the user to say something personal about the current topic.

 $<sup>^{11}</sup> The$  video is also available online at the address <code>https://youtu.be/Zfr1YUX6BJw</code>

children by their first name, but not adults, especially older know you as much as possible. Because I am here to help you. adults. I find that really interesting. People tell me that it is ok Please let me explain how I can help you. I can tell you the news to call a Lady "auntie" if you don't know her proper name. (4a) and the weather. I can keep you company. I can call your family Can I call you "auntie"? when you ask me to. Yes (5) That's right, I am waiting for Diwali right now. I will call you "auntie". (4-4d-3-4) I only know a little about this. Sorry, I think I haven't understood what you said. (4-4d-3-4) It's People tell me that, in the past, Indian families lived together in interesting to know more about things that you like. I'd love to know more about this. (4a) Do you like traditional dances? a household. Today this practice is not very common. But they continue to hold close relationships. (4a) Can you tell me Yes, but I am waiting for Diwali right now (5) something about your family? No (5) I see. (4-4d-3-4) I'd love to learn more about this. I know that Indian people are very close to their friends, and they may like to talk about the important people in their life. (4a) Can we talk about your friends? No thank you (5) Thank you for answering all my questions. You are very patient. (4-4e-6)I am waiting for the Diwali festival. (2-3-4) I know that Diwali is celebrated in Autumn, and even (3-4-4a) I assume that the Diwali festival is a holiday for you. is though it is more common in the North of India, it is a very that right? popular festival for all Indians around the world. (4a) Do you Yes (5) celebrate the Diwali festival? Diwali is the festival of lights that celebrate a new beginning Yes of course (5) and the triumph of good over evil and light over darkness. (4-People tell me that during Diwali people light small lamps or 4b) I would like to learn from you about the significance of the candles around their house. (4-4b) I would like to learn from you about the significance of the Diwali festival. Yes, as you said Diwali is the festival of lights, and I love it During the Diwali festival I meet with my family and we have because there is a lot of people in the street <...> and I really love it. And I am really looking forward to that. dinner all together. It's nice to learn more about you. (4-4c) Would you like to get I see, very interesting. (4-4c) Would you like to have some some Diyas for Diwali? sweets for Diwali? Yes... actually no, because I would have to do a lot of sport. No thank you, otherwise I have to do sport\* (5) Ok, no problem. (3-4-4a) If you don't mind, I will ask you a Ok, no problem. (3-4) Please tell me I like to know more. I know that many Indian people like cricket. (4a) Do you like sport? question. Do you like sport? Yes, I like some sports (5) Playing and watching sport are really nice hobbies. (4-4b) Please Playing and watching sport are really nice hobbies. (4-4b) Please tell me what you like the most about sport. tell me what you like the most about sport. I really like cricket. I like sport because when I do sports I feel like relieved and... I see, very interesting. (3-4-4a) May I ask you a question? Do even if I am tired I go home and my day is better Oh, interesting. (3-4-4a) Please tell me. I'd like to know more. you like cricket? Do you ever feel tired? Yes (5) Only when I do sports. I know that a cricket bat usually weights between 2 and 3 pounds and is made traditionally from willow wood. (4-4b) Playing and watching sport are really nice hobbies. (3-4-4c) Please, tell me what you like the most about cricket. Would you like me to remind you to watch sport on TV? It is a very hard sport, which requires a lot of concentration. No. thank vou. Oh, interesting. (4-4c) Would you like me to remind you to Ok. (4-4d-3-4-4a) If you don't mind. I will ask you a question. Do watch cricket on TV? vou like cricket? No. thank vou. Yes, of course I like cricket (5) Ok, no problem. (4-4d-3-4-4a) Please tell me, I want to know I know that cricket is a team sport for two teams of eleven more. Do you have any relatives who plays cricket? players each. A game of cricket can last several days. (4-4b) Please, tell me what you like the most about cricket. Yes, my brother (5) Actually is very exciting to see cricket, especially the big match (3-4-4a) Do you have a brother? with areat teams... it's very exciting. <...> and you find a lot of Yes (5) your friends in there, and it's a very nice moment to share. I am very pleased to know that you have a brother. Good to know. (4-4d-3-4-4a) I only know a little about this. Do you ever watch cricket on TV? Yeah, If I am not able to get to the field or the match is in other cities and I can't reach those matches... Usually I watch them on TV. Actually with my friends or... sometimes alone, but I really don't like to watch sports alone (5)

(4) If it doesn't make you uncomfortable. I would love to get to

(2-3-4) I am aware that in the Indian culture it is ok to call little

Fig. 7: Audio transcripts: an Indian person chitchatting with a virtual assistant (left) and Pillo (right) connected to the CARESSES Cloud.

No, thank you.

\* The person, not being an Indian, thinks that "Diyas" are a kind

of food, whereas they are typical candles.

I know that cricket is a team sport for two teams of eleven players each. A game of cricket can last several days. (4-4c)

Would you like me to remind you to watch cricket on TV?

In the second column/third row, the sequence "(3-4-4a) If you don't mind, I will ask you a question. Do you like sport?" corresponds to a change of topic requested by the user, whereas the sequence "(3-4-4c) Would you like me to remind you to watch sport on TV?" corresponds to the system proposing an activity. The transition "(4-4d-3-4-4a) If you don't mind. I will ask you a question. Do you like cricket?" corresponds to the system moving "Sport" from the related topic "Cricket" (which has been chosen by the system – through probabilistic reasoning – based on the cultural profile by the user).

Finally, notice that "(5)" corresponds to an update of the knowledge base depending on the user's reply to a question.

From a first analysis, it is evident that the style of the two users is very different, more verbose in the second column. The readers may explore by themselves the behaviour of the system with the two users, by discovering differences and analogies in the two dialogues, also referring to the videorecording of the experiment attached to this article.

## VI. CONCLUSIONS

This paper described a Cloud platform providing services for culturally competent interaction, that was extensively tested with the robot Pepper to perform experiments with care home residents in UK and Japan, and has now been expanded to allow virtually any robot/home assistant to become culturally competent.

The rationale behind the proposed architecture is discussed, by focussing on the key principles that have been adopted to develop services for culturally competent conversation, which have been tested – in addition to Pepper – on the robotic medicine dispenser Pillo and a custom built android application. In spite of the limitation of the system, transcripts and videos make possible to appreciate the dynamism of the dialogue compared with off-the-shelf robotic systems and home assistants, typically able to understand commands and to reactively perform tasks, but unable to lead a conversation on generic "daily life" topics.

According to the CARESSES release plan, starting from the end of September 2020, CARESSES Cloud services will be accessible to everybody who is interested in trying them [19] (Figure 8 shows our next candidate).

# VII. ACKNOWLEDGEMENTS

This work has been supported by the European Commission Horizon2020 Research and Innovation Programme



Fig. 8: Prof. Einstein by Hanson Robotics.

under Grant Agreement No. 737858 Authors wish to thanks Prof. Alessandro Saffiotti from Örebro University, Prof. Nak Young Chong from JAIST, and SoftBank Robotics for providing support in the development of CARESSES architecture. Dr. Amit Kumar Pandey from Hanson Robotics, partner of University of Genova in the Social Robotics Lab, for his work in CARESSES and for believing in this project.

## REFERENCES

- [1] M.M. Leininger. Leininger's theory of nursing: Cultural care diversity and universality. *Nursing Science Quarterly*, 1(4):152–160, 1988.
- [2] I. Papadopoulos. Transcultural health and social care: development of culturally competent practitioners. Elsevier health Sciences, 2006.
- [3] B. Bruno, N.Y. Chong, H. Kamide, S. Kanoria, J. Lee, Y. Lim, A.K. Pandey, C. Papadopoulos, I. Papadopoulos, F. Pecora, A. Saffiotti, and A. Sgorbissa. Paving the way for culturally competent robots: A position paper. In 26th IEEE Int. Symp. on Robot and Human Interactive Communication, pages 553–560, 2017.
- [4] A. Sgorbissa, I. Papadopoulos, B. Bruno, C. Koulouglioti, and C. Recchiuto. Encoding guidelines for a culturally competent robot for elderly care. In *IEEE Int. Conf. on Intell. Robots and Systems*, pages 1988–1995, 2018. Madrid, Spain.
- [5] A. A. Khaliq, U. Köckemann, F. Pecora, A. Saffiotti, B. Bruno, C. T. Recchiuto, A. Sgorbissa, H. Bui, and N. Y. Chong. Culturally aware planning and execution of robot actions. In 2018 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pages 326–332, Oct 2018.
- [6] B. Bruno, C.T. Recchiuto, I. Papadopoulos, A. Saffiotti, C. Koulouglioti, R. Menicatti, F. Mastrogiovanni, R. Zaccaria, and A. Sgorbissa. Knowledge representation for culturally competent personal robots: Requirements, design principles, implementation, and assessment. *Int. J. of Social Robotics*, 11(3):515–538, 2019.
- [7] B. Kehoe, A. Matsukawa, S. Candido, J. Kuffner, and K. Goldberg. Cloud-based robot grasping with the google object recognition engine. In *IEEE Int. Conf. on Robotics and Autom.*, pages 4263–4270, 2013.
- [8] M. Tenorth, A. Clifford Perzylo, R. Lafrenz, and M. Beetz. The roboearth language: Representing and exchanging knowledge about actions, objects, and environments. In *IEEE Int. Conf. on Robotics* and Autom., pages 1284–1289, 2012.
- [9] R. Arumugam, V.R. Enti, L. Bingbing, W. Xiaojun, K. Baskaran, F.F. Kong, A.S. Kumar, K.D. Meng, and G.W. Kit. Davinci: A cloud computing framework for service robots. In *IEEE Int. Conf. on Robotics and Autom.*, pages 3084–3089, 2010.
- [10] D. Berenson, P. Abbeel, and K. Goldberg. A robot path planning framework that learns from experience. In *IEEE Int. Conf. on Robotics* and Autom., pages 3671–3678, 2012.
- [11] L. Vasiliu, I. Trochidis, C. Bussler, and A. Koumpis. Robobrain: A software architecture mapping the human brain. In *IEEE-RAS Int. Conf. on Humanoid Robots*, pages 160–165, 2015.
- [12] M. Bonaccorsi, L. Fiorini, F. Cavallo, A. Saffiotti, and P. Dario. A cloud robotics solution to improve social assistive robots for active and healthy aging. *Int. J. of Social Robotics*, 8(3):393–408, 2016.
- [13] P. Simoens, C. Mahieu, F. Ongenae, F. D. Backere, S. D. Pestel, J. Nelis, F. D. Turck, S. A. Elprama, K. Kilpi, C. Jewell, and A. Jacobs. Internet of robotic things: Context-aware and personalized interventions of assistive social robots. In 2016 5th IEEE Int. Conf. on Cloud Networking (Cloudnet), pages 204–207, Oct 2016.
- [14] J. Huang, W. Lee, and T. Lin. Developing context-aware dialoguing services for a cloud-based robotic system. *IEEE Access*, 7:44293– 44306, 2019.
- [15] B. Bruno, R. Menicatti, C. T. Recchiuto, E. Lagrue, A. K. Pandey, and A. Sgorbissa. Culturally-competent human-robot verbal interaction. In 15th Int. Conf. on Ubiquitous Robots, pages 388–395, June 2018.
- [16] Nikolaos Mavridis. A review of verbal and non-verbal human-robot interactive communication. Robot. and Auton. Syst., 63:22 – 35, 2015.
- [17] J.R. Searle. Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press, 1969.
- [18] Leonard Richardson and Sam Ruby. RESTful web services. "O'Reilly Media, Inc.", 2008.
- [19] Roberto Menicatti, Carmine Recchiuto, Ha-Duong Bui, Nguyen Tan Viet Tuyen, Maxime Busy, Uwe Köckemann, Ali Abdul Khaliq, and Nicholas Bastos. CARESSES Open Source Software. Zenodo, January 2020.