Conversational Assistant for an Accessible Smart Home*

Proof-of-Concept for Portuguese

Maksym Ketsmur
DETI/IEETA
Universidade de Aveiro
mvk@ua.pt

António Teixeira
DETI/IEETA
Universidade de Aveiro
ajst@ua.pt

Nuno Almeida
DETI/IEETA
Universidade de Aveiro
nunoalmeida@ua.pt

Samuel Silva
DETI/IEETA
Universidade de Aveiro
sss@ua.pt

Mário Rodrigues
ESTGA/IEETA
Universidade de Aveiro
mjfr@ua.pt

ABSTRACT

There is a continued increase in the integration of *intelligent* devices in our homes. Making these new complex ecosystems accessible poses great challenges, stemming from the very nature of the Home (several spaces separated by walls), the diversity of inhabitants (e.g., children, adults, older adults, persons with temporary or permanent impairments), the tasks they perform (e.g., cooking) or their native languages. In this context, an assistant capable of spoken and written *conversation* with the inhabitants can be a valuable help in making a household more accessible for several groups of persons, if not for all. This paper presents information regarding the development and first results of such an assistant targeting an Accessible Smart Home for a Portuguese speaking family with some accessibility needs.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; **Accessibility systems and tools**; *Accessibility technologies*;

KEYWORDS

Smart Homes, Accessibility, Spoken Interaction, Conversational Assistants, Design for All.

ACM Reference Format:

Maksym Ketsmur, António Teixeira, Nuno Almeida, Samuel Silva, and Mário Rodrigues. 2018. Conversational Assistant for an Accessible Smart Home: Proof-of-Concept for Portuguese. In *DSAI 2018: 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, June 20–22, 2018, Thessaloniki, Greece.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3218585.3218594

*In the scope of Project Smart Green Homes (SGH).

DSAI 2018, June 20–22, 2018, Thessaloniki, Greece

ACM ISBN 978-1-4503-6467-6/18/06...$15.00
https://doi.org/10.1145/3218585.3218594

1 INTRODUCTION

Currently, there is a consistent upward trend of integration of *intelligent* devices in our homes, ranging from the smallest internet-of-things (IoT) devices to kitchen appliances and home entertainment systems. Controlling and providing information to these *smart homes* as well as receiving information from them – i.e. interacting – needs to be accessible to all. For example, adopting the common graphical modalities (graphs, buttons to press, text) [4], to provide information to users falls short of adequacy as inhabitants with vision problems will be partially or completely excluded.

As argued in [20], for Assistive Technologies, speech input and output modalities can be the only adequate option, in many situations. For instance, speech output can convey house status information to the visually impaired and spoken input can help a child controlling its environment. "What is a VUI [Voice User Interface] but the ideal, nonvisual experience for the blind and visually impaired?" [16, p. 61].

Besides playing an important role in creating accessible homes for "those who cannot read or type (including very young children and blind)" [20], speech based interaction has additional important advantages, such as: being hands-free, intuitive, fast and fostering empathy [16], which can contribute to more accessible homes for all of us. For example, during common activities at home, such as preparing our meals, the hands-free characteristic of speech ensures a continued access to house interaction. Its naturalness/intuitiveness will also make interaction accessible to young children, even before they learn to read. Simple forms of using speech for interaction, such as the use of simple spoken commands, can be of great value, but the vast potential of spoken interaction can only be exploited if interaction goes beyond that and exploits multi-turn interactions which are characteristic of dialogs/-conversations. As an aftermath of recent technological developments, there has been a rising number of Assistants proposed, with some conversational capabilities, through which the user can interact by voice and/or text in a *natural* dialog. Dialog presents additional advantages, such as context from previous turns, to disambiguate and allow simpler utterances, or the possibility to provide partial information that will be completed in subsequent turns. As a consequence, interaction becomes simpler and more similar to human-human communication.

Three main types of conversational interfaces can be considered: (1) chatbots, in which the interaction is dominantly by text; (2) voice user interfaces (VUIs) based on simple interaction, such as single turn command and control; and (3) conversational, VUIs that start to explore multiple-turn *dialogs* with the user.

Voice user interfaces are widely used in smartphones (e.g., Apple Siri) and, increasingly, in smart homes (e.g., Google Assistant, Amazon Alexa), consisting on *assistants* that allow to do certain tasks [16–18, 22, 24].

Despite all these recent assistants, making homes accessible continues to pose several challenges, due to, among other: (1) the very nature of the Home (e.g., several spaces/rooms separated by walls); (2) the variety of inhabitants that can live in them (e.g., children, adults, older adults, persons with temporary or permanent impairments); (3) the tasks they perform (e.g., cooking); and (4) their native languages (e.g., Portuguese, English, Greek). In this regard, existing systems only support a very limited set of languages and functionalities, not necessarily considering the specificities of all possible house inhabitants, highlighting the need for new assistants providing extended features usable by all and for as much languages as possible.

It is also important to make the development of new assistants easier, considering, as much as possible, existing technologies, to have reduced costs and a prompt answer to specific accessibility needs.

In this paper, we argue that an assistant, capable of spoken and written input and output, and establishing a multi-turn *conversation* with the household occupants, can be a valuable help in making a house more accessible for several groups of persons, if not for all. To support our claim, we present the development and describe the potential of such a Conversational Assistant targeting an Accessible Smart Home for a Portuguese speaking family. The development was made, whenever possible, by use and adaptation of existing tools and resources.

The remainder of this article is organized as follows. The next section summarizes the related work regarding Assistants, some popular existing assistants, and an analysis of tools for their development and support. In section 3, the home occupants, consisting of a family of 5 members, are characterized and is illustrated their possible interaction with the accessible home. In section 4, we present the targeted requirements, the system architecture, and a more detailed description of the processing steps required for the developed assistant. The fifth section shows the system potential to have dialog with the user and the results obtained. Finally, in sixth section we conclude our work, describing the possible future implementations and ideas.

2 BACKGROUND AND RELATED WORK

This section presents background information regarding the concept of dialog, and the relevant related work on conversational assistants, their typical structure, some popular recent home assistants and, finally, tools that make their development possible.

2.1 Dialog and dialog acts

Citing Jurafsky & Martin, "[. . .] conversation or dialog is the most fundamental and specially privileged arena of language" [9, p.418]. It can be seen as a sequence of acts that include [9, 19]: <state>, <assert>, <describe>, <warn>, <remark>, <comment>, <command>, <order>, <request>, <criticize>, <apologize>, <censor>, <approve>, <reject>, <welcome>, <promise>, <express approval> and <express regret>. The use of these types of acts can be found, for example, in Alexandersson et al. [1]. Particularly relevant for a home assistant are the types <command>, <request> and <greet> [14].

2.2 Chatbots, VUIs and Assistants

Currently, chatbots are widely used as a new channel of information and communication that companies, in a variety of businesses (e.g., airlines, hotels, restaurants), use to reach their target audience through different applications (e.g., Facebook).

The other types of conversational interfaces are the single turn command-and-control VUIs (Voice User Interfaces) and the multi-turn VUIs capable of dialog, known as conversational [16] and performing similar interactions to chatbots, but having speech as input and output. Apple's Siri and Microsoft's Cortana are well known examples of conversational assistants available on mobile devices.

But what does *conversational* actually means? It means the ability to go beyond simple one-turn interaction and have a true conversation with the system, i.e., have a dialog in such way that creates the feeling of a *common ground*, remembering the past, from previous conversations, and also from the last turn, supporting contextual awareness, and exchanging appropriate questions [16].

There are several methods of dialog control [9, chap 28], including state-based, frame-based, and agent-based. Systems also differ regarding communication initiative, done just by the user, just by the machine or by both.

An example of a machine initiative system is Personal Health Management Assistant [6], which aims to perform a *daily checkup* by asking questions in natural language using synthesized voice or in written form to collect relevant information about the patient. The assistant takes initiative of the dialog and is able to understand the user's intent and to help him achieve his goals.

In an assistant such as Amazon Alexa, where the initiative is on the side of the user, one of the main tasks is to identify the user's intention, which is related to his dialog act.

Closely related to Assistants are the Q&A systems that allow to answer user questions regarding some domain(s) (see, for example, [10]). Google Assistant and Amazon Alexa have an integrated Q&A system, enabling them to answer questions from web-based sources of information.

2.3 Main blocks of a Conversational Assistant

In order to enable the interaction between the human and the machine, in a dialog format, a typical conversation system must identify user's intention, perform actions and respond using synthesized voice. As such it needs to integrate at least [3, 24] the following modules (Fig. 1): speech recognition, natural language understanding, dialog management, natural language generation, and text-to-speech.

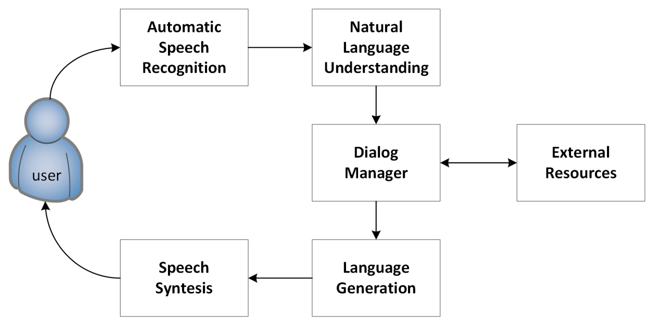

Figure 1: Typical system structure of a conversational assistant.

2.4 Representative (Home) Assistants

With the raising trend for smart homes, many home assistants have been developed to perform *small chores* like setting an alarm or checking the time, as well as more complex tasks such as booking a hotel or calling a taxi. One of the first home assistants was Amazon Alexa, launched in 2014, and later appeared Google Assistant in 2016 [17, 22, 24].

Amazon Alexa is a personal assistant created by Amazon and is being used on multiple devices like Amazon Echo, Echo Dot, and Amazon Tap. This assistant can set alarms, change the house temperature, control the lights, and answer questions about different domains, among other tasks. The interaction is performed by voice and, as these devices (e.g., Amazon Echo) are able to recognize the voice at long distances, it is possible to operate them (almost) anywhere in the house [2, 18].

Google Assistant is used on Google Home device. One of the advantages is its artificial intelligence is to enable it to understand when the user makes a mistake and thus it can try to correct him. Also, it can have the initiative to start the dialog with the user, and it features the possibility to generate contextualized answers, i.e, if the user asks "*How old is Will Smith?*" and then "*What movies did he participate in?*", without referencing the actor, the assistant can identify that the context is provided by the previous interaction [18].

In addition to these two most popular assistants, there is *Siri assistant*, working on the Apple Homepod. This assistant is mainly intended for music playback but it can perform the same tasks as Amazon Alexa and Google Assistant.

The presented home assistants allow to perform a wide variety of things, but it is not possible, for instance, to get relevant data about the smart home such as energy and water consumption and their costs, and also they still do not support some widespread languages such as Portuguese. In countries not covered by the provided languages this prevents its usage by the people that do not master a foreign language, such as youngsters and sometimes the older ones, which are excluded and cannot operate the smart home at its full potential.

Besides these assistants, there are others of lesser impact such as Kuri, Bixby by Samsung, LG ThinQ and, Kevin by Swiss, presented on the CES 2018.

2.5 Tools to Support the Development of Assistants

None of the existing assistant fulfills our needs, particularly for a more advanced interaction with the home, but there are important tools and resources that can be considered for developing one. This section presents a summary of existing tools targeting the development of assistants or parts of them.

Table 1 presents a set of tools that allow developing chatbots. They were evaluated according to the following parameters: access method, costs and, more important, the support for the European Portuguese language. All presented tools provide the same information type, that is, intents and entities.

Table 1: Tools and APIs useful for Chatbots and Conversational Assistants development.

Name	Access method	Costs	Portuguese
API.AI / DialogFlow	HTTP	No	Brazilian
Microsoft LUIS	HTTP/SDK	No	Brazilian
IBM WATSON	HTTP/SDK	No	Brazilian
Amazon LEX	HTTP/SDK	Yes	Yes
WIT.AI	HTTP	No	Yes

DialogFlow (former API.AI) and IBM WATSON were tested in order to choose the best one for our needs. The other were excluded: Amazon LEX because it is not completely free, WIT.AI because it discontinued the possibility of chatbot development, and finally Microsoft Luis was excluded because although there are many chatbots based on it, it does not allow to create a chatbot, being used, in general, as a NLP processing tool.

The DialogFlow tool (API.AI) allows easy chatbot creation and configuration. However, in order to identify the user's intention, the input phrase must match the training phrase. This is a noticeable shortcoming as chatbots aim to identify the user's intention independently of the phrase form. For instance, if the training example is "Turn on the lights!" it will fail to identify the user's intention in the utterance "Lights on!". Despite this, it correctly identifies entities in the input.

IBM Watson allows to perform NLP and chatbot development. It is deployed as a service that uses NLP and machine learning to analyze large amounts of unstructured data, allows to reply to users' questions, extract key information from multiple documents, and identify patterns and relationships on unstructured data [7, 11, 18].

The main concern before using Watson was the lack of support for European Portuguese, despite the support for Brazilian Portuguese. However, some tests confirmed that the platform was able to identify intents and entities using European Portuguese. These features are sufficient to build a smart home assistant.

3 METHOD

A user-centered design and development [21] method was adopted with an emphasis on the definition of a family of Personas and several scenarios.

3.1 Family of Personas

In the scope of project Smart Green Homes, a family of Personas was defined with some accessibility needs consisting of 5 members: mother, father, two children, and a grandmother. Some basic information regarding the family members is presented in Table 2.

Table 2: The Sousa Family of Personas.

Family member	Name	Age	Limitations
Mother	Joana	45	none relevant
Father	Afonso	46	temp. reduced mobility
Grandmother	Maria	70	vision problems
Teenager	Maria	14	none relevant
Youngster	António	4	does not read

3.2 Accessible Scenarios/ Scenarios for All

Several scenarios were defined regarding accessibility challenges faced by the family. Next, several representative examples are presented that are deemed useful for all the inhabitants, but essential for those with *special* needs.

3.2.1 Accessibility for Children: As children have difficulties in reaching light controls and this access can place them at risk, the following scenario was created:

António is playing in the living room at the end of the afternoon and the room is getting too dark. Without needing to bother his parents he says "Please turn on the lights" and the assistant takes care of it.

3.2.2 Accessibility for reduced vision: As decrease in vision is a common effect of age and the family integrates an older adult (Grandmother Maria):

Grandmother Maria is at the bathroom and by speaking she controls the room temperature and the water temperature of her bath. Also, while in the bathroom, she enquires the system on the status of kitchen lights, that she later turns off when informed by the system that she forgot them on.

3.2.3 Accessing house information anywhere: motivated by the difficulties in moving around the House, even temporary:

The father, Afonso, after a small injury during fitness exercises, is siting in the living room, reading a magazine article about the average energy and water consumption in Portuguese homes and wants to understand how their household fits those averages. He asks the assistant how much energy was consumed last month. And after the assistant reply, he adds "And how much it costs?".

4 A PORTUGUESE CONVERSATIONAL ASSISTANT FOR ALL AT HOME

In this section the information on the developed Assistant is presented, starting by the requirements and architecture, describing all the processing steps, and finally some implementation details.

4.1 Requirements

Functional requirements:

- Be capable of answering questions concerning the information provided by sensor and home appliances (i.e. Information available from the smart home ICT infrastructure).
- Capability to separate information by users.
- Capability of handling multiple users.
- Control some appliances, e.g. lights.
- Capability to control several devices with one sentence.

User requirements:

- Process input by text and voice. The input should be in the form of natural language sentences.
- Minimal capability of Multimodal Output.
- Provide output as sentences, lists or graphics, as more adequate for the question, context, and user.
- Handle context of questions, allowing continuation in successive interactions.
- Support European Portuguese
- (if possible) Support for English.
- (if possible) Provide an integrated way of querying House and other knowledge bases (ex: Weather, Wikipedia).

Acessibility Requirements:

- Usable by people with low vision or even by blind people.
- Usable by children with ages as low as 3 years.

4.2 Architecture

Figure 2 represents the architecture of the developed system, which consists of 4 distinct modules allowing users to make requests and commands, processing them and return a response. Although modules are different, they are interconnected, since they share the output information of each other. In the same figure it is possible to see 4 distinct areas: (1) the first area represents the system's modules; (2) the second one represents the output data produced by each module, (3) the third represents the internal resources, and (4) the last area shows the external services and resources.

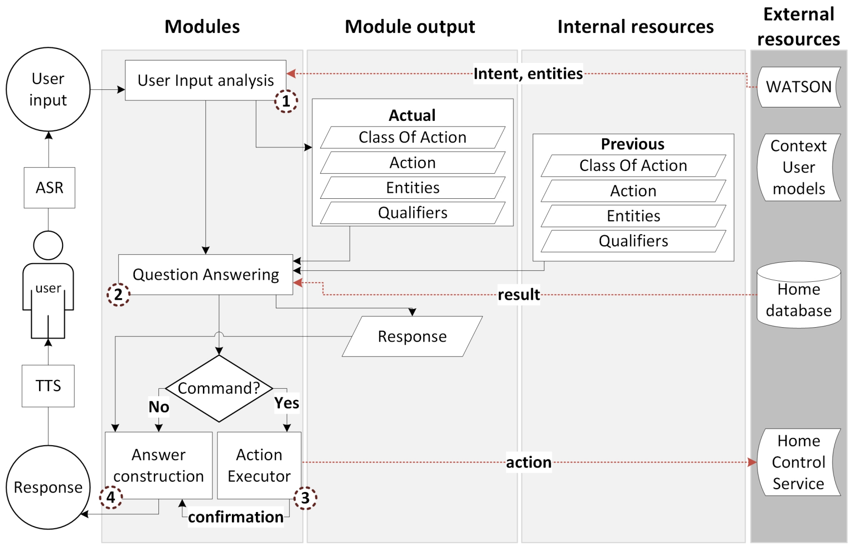

Figure 2: Overall Architecture of the developed Conversational Assistant System.

4.3 Assistant in action

The assistant operation can be seen as a sequence of steps as follows:

First step: The first phase of processing is carried out by the `User input analysis` module, proposed to identify the user's intent, what actions to take, and the involved entities, in the Watson output. It is one of the most important modules, since it allows to manage and maintain the dialog with the user.

Due to the application domain, the system was designed to identify three types of intentions (dialog acts): (1) greetings used to *wake up* the system; (2) requests used to consult consumption and state of home appliances; and (3) commands in positive and negative form to appliances control.

As aforementioned, the identification of the intention allows to maintain a more natural dialog with the user but the actions and entities and respective qualifiers must be identified too. The actions differ from the intentions and can be found in commands and in requests, for example, actions such as *turning on/off* are frequent in commands, while an action like *consult* is used in requests. Entities can represent resources (e.g., electricity, water, gas), house smart devices or appliances (e.g., lights, plugs, oven, tv), and qualifiers represent home divisions (e.g., bedroom, kitchen, living room), and dates.

To illustrate the process, we consider an example of a possible user request, *"What was the water consumption this month in the kitchen?"*. First the user's input is analyzed by Watson, identifying the intent and the entities present in the request. Then our system organizes the identified entities in actions, entities, and qualifiers.

One of the main advantages of Watson is the off-the-shelf identification of dates and numbers, that is, for *"this month"* Watson returns two dates, the first and the last day of current month. The same result can be achieved for *"this week"*, *"last month"*, *"next year"*.

Second step: The `Question Answering` module, aims to construct SPARQL queries using obtained data from the previous module (actual data) and the stored information from previous interactions. For the system database was used the Apache Jena Semantic Web framework, allowing to query the semantic triples, which represent the hourly appliances' consumption.

Given the presented example and the previous module output, the resulting SPARQL query can be read as: *sum all the registered water consumptions by appliances located in the kitchen* (Fig. 3).

```
SELECT (SUM(distinct ?consumptionAmount) as ?totalAmount) WHERE {
  ?subject :hasConsumptionType ?consumption .
  ?subject :hasAmount ?consumptionAmount .
  ?consumption rdf:type ?consumptionType .
  ?device :hasConsumed ?subject .
  ?device :isLocatedin ?deviceLocation .
  ?deviceLocation rdf:type ?deviceLocationType .
  # ---------------- FILTER BY Resource --------------------
  FILTER (?consumptionType = :Water) .
  # ---------------- FILTER RY HOME PARTITION --------------
  FILTER (?deviceLocationType = :Kitchen) .
}
```

Figure 3: SPARQL example for consumption consulting

Third step: The `Action Executor` module aims to call the commands from the Home Control Service, controlling appliances, lights, and plugs in all house divisions.

Fourth step: The `Answer Construction` module generates the final answer, gathering the resulting data from previous modules. The result of the first module allows to construct the response in a more natural way, combining with the result of the second or third modules, depending on the identified intent type. The final answer can be presented to users in two distinct forms, synthesized voice or text.

4.4 Some Implementation details

Using Watson

As aforementioned, Watson allows to analyze text and identify the entities and their context, and it must be trained in order to obtain the best results. The basic configuration allows to extract currency values (both the amount and units), date mentions (Friday), numbers mentioned from user input as digits or written numbers and amounts, and the % signs (15%). Extracting date mentions is a very useful feature for the developed assistant, since the user often uses expressions like "*last week*", "*previous month*", "*yesterday*" in his requests.

To develop a conversational home assistant, Watson must provide some basic information: the current dialog act, the action to perform, and the involved entities.

First, the dialog acts – called intents in Watson – were defined. Accordingly, the following Watson main intents were considered: request and command in positive and negative form.

To improve the correct identification of intents (dialog acts), Watson was trained with a set of sentences of each type. For example, for requests (consult consumption) we considered sentences such as *"What was the consumption of ...?"*, *"How much did I spend from ...?"*, *"Gas consumption was ..."*, among other. This makes Watson use the sentence form and its structure instead of specific words to identify intentions.

Another Watson output concerns entities (keywords identified in the input sentence). Considering the smart home domain, several entities have been defined, including: action, device, home partition, resource, and date.

Action: allows identifying the action to be taken. In the requests, the most common action is consumption consulting, while in the commands, actions such as "*on*" and "*off*" are most common.

Device: refers to home appliances, lights, plugs or any other smart system.

Home partition: represents the location based on the different rooms in the house.

Resource: relates to resources used such as water, electricity, and gas.

Date: allows the identification of more complex dates. Although Watson can identify dates, it shows some problems when input was like "*January 15, 2016*", where it could return more than one result (15.01.2016, 03.2018, 15.03.2018). To solve this issue, the system has been trained to correctly classify this type of entries.

All information can be easily added, changed or deleted using both the web interface and the application, which makes Watson a flexible tool to any domain chatbot development.

Ontology

Analyzing some of the related systems [8, 12, 15, 23], a base ontology was developed that allows describing the data concerning appliances' consumptions, the home environmental information, and the relation between the inhabitants and the appliances.

The system's ontology consist of 5 main classes (Device, Environment, Person, Home and Resource) and some relations. The `Device` class has two subclasses, `Appliance` and `Sensor`, where the first represents the house appliances and the second the sensors that allow to measure environmental characteristics. The subclasses of the class `Part` represent the home divisions. The `Resource` class represents the resources consumed by appliances, being electricity, gas, and water the most relevant for our scenarios.

House Control Service

The house control service aggregates the different APIs to control all smart home devices. This way, using only this REST service it is possible to control all installed devices, such as smart lights, plugs and appliances, by considering a common interface, which internally adapts to the specificities of each device.

The service can be called by referencing the device, action to perform and the room where the device is located. For instance, if the assistant needs to *turn on the lights in the living room* it requests the service, using HTTP GET, to the address where the service is installed and adding "`/lightOn/livingroom`".

5 RESULTS

This section describes the simulated smart home and some of the currently possible dialogs are presented with the purpose of showing the system's conversational ability.

5.1 A Simulated House

With the goal of having the most realistically possible scenario for the system assessment, to support development, a smart home was simulated, consisting of 7 rooms: living room, kitchen, bathroom, garage, bedroom, laundry, and office. Each room contains a set of appropriate appliances, lights and plugs, in a total of 38 devices. Appliances consume resources such as water, electricity, and gas, and some of them consume more than one resource (e.g., a washing machine consumes water and electricity).

In order to set the appliances' consumption in closer relation to reality, the simulated consumption values are the same reported in [13], identifying the minimum, maximum and standby consumption for each of the appliances, and also the home environmental data [5].

In this way, the hourly consumption of each appliance was simulated during the period of one year, where the registered consumption consists in two main values: the consumption amount [1] and the type of consumed resource for each period of one hour.

5.2 Sample dialogs

This section aims to show what kind of interaction is already supported by the system, what type of response it returns, and how it manages and maintains the dialog with the user.

[1]Random value between the minimum and maximum or standby consumption.

As already mentioned, in the development section, the system is prepared to receive consumption requests and commands. In addition, the system was developed with the objective of maintaining the dialog with the user in a more natural way, which requires user's intention identification.

The simple interactions are illustrated first, then the more complex, also depicting parts of the interaction where the system shows its ability to maintain the dialog.

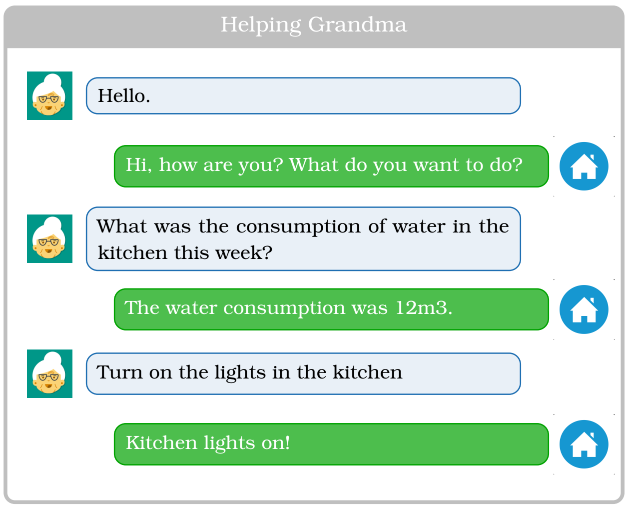

Figure 4: Example of use 1 - Grandmother getting information regarding water consumption and turning off lights at a different room.

In the example depicted in Fig. 4 is presented a simple interaction with the system. In the first interaction, the user (Grandma) greets the system and receives, as reply, a question inquiring what she wants to do. The Grandma makes a request regarding the amount of consumed water in the kitchen this week and the system replies in a succinct form. Then, the user's intention changes, that is, the user now wants to execute a command, "*turn on the lights in the kitchen*", which is confirmed using synthesized speech after correct execution of the command.

Figure 5 presents more complex interactions and the dialog follow-up techniques between the system and the user. First, the user requests information regarding electricity consumption in the bedroom. After getting the answer, he wants to know how much it costs and, based on the defined energy costs, the system responds with total costs. Then, the user asks "*and in the bathroom?*". This time, the system understands that the context is the same and the user still wants to know the cost of electricity consumption, but for another room.

In the meantime, Grandma issues a command to turn on the lights in the bathroom and after says "*turn off*". As in the previous example, the system identifies that the user's intent is the same, but now the action is to turn on the lights.

In the end, the user makes a more complex command, involving various actions and entities, "*turn on the lights in the room and turn off in the kitchen*", requiring that the system understands the sequence of words in order to execute the correct actions.

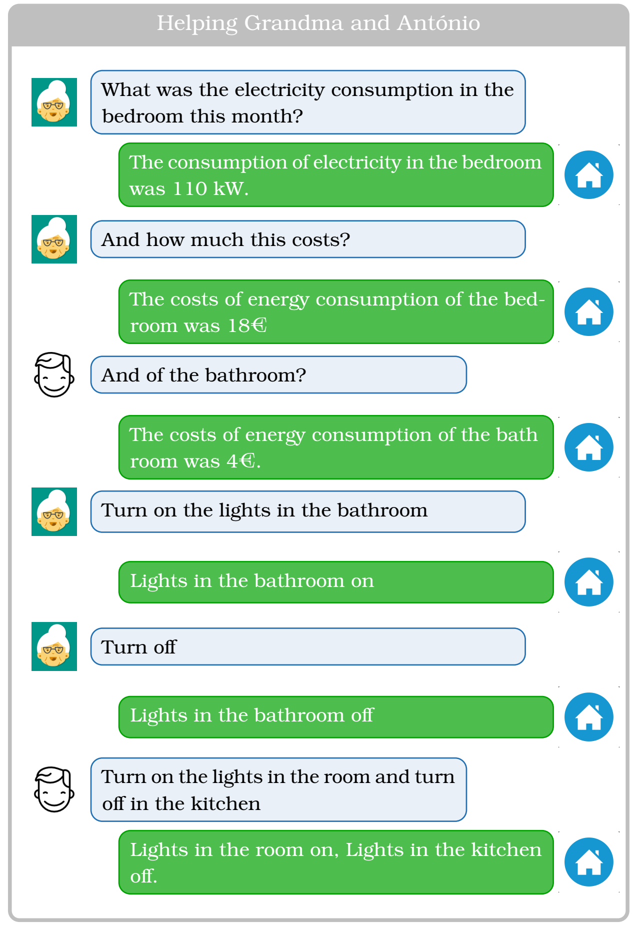

Figure 5: Example of use 2 - Assistant in use by Grandma and young António, demonstrating the use of context of discourse to simplify interaction.

Another example, (Fig. 6), illustrates the system's feedback when the user does not provide enough information to execute a command.

Overall, these examples show that the system is able to receive user's requests and commands, which can either be simple or more complex, and is also capable of memorizing the user's intention, enabling an interaction closer to a real dialog.

6 CONCLUSION

Targeting the needed accessibility of emerging Smart Homes, in this paper is presented a proof of concept for a Conversational Assistant. In order to make the creation of the prototype fast, the system was developed using available services, being selected IBM Watson and modules for Speech-to-Text and

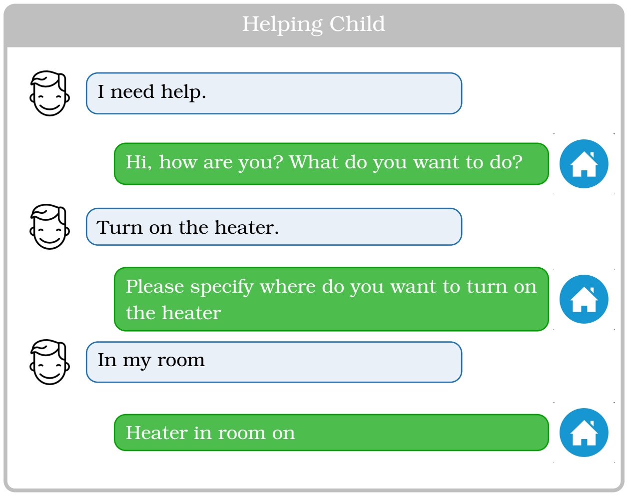

Figure 6: Example of use 3 - The assistant asking for missing information to complete a command.

Text-to-Speech previously developed by the authors. The developed Assistant can provide information on the house (e.g., consumptions) and act as an assistant to control home devices such as lights, being able to keep and make use of some context of the dialog and user location. The system potential is demonstrated by examples of dialogs.

6.1 Future work

The system is, at the moment, a functional proof-of-concept and further development and evaluation is needed.

During the development phase, some problems were encountered when training Watson, an issue of relevance since Watson provides the basic information for dialog management. One of the biggest problems is the Watson synonyms training, needing further research. Eventually Watson needs to be complemented with additional NLU processing.

Already in progress is the connection of the system to context and user models, running in the house, which are capable of providing information on location of the user (if the user authorizes). With this connection, saying "*Turn on the lights*" will be enough to act upon the lights of the current location of the user. It will also allow the system to refuse to perform an action if not adequate to the user (e.g., young children with limited access to house controls).

Another highly important future work is the evaluation by users, particularly users with real accessibility needs. There is also need for work on how to deploy in real houses. For now, the system uses a mobile device for speech input and output, to record and to play, but other solutions need to investigated.

ACKNOWLEDGMENTS

Research funded by project Smart Green Homes (POCI-01-0247-FEDER-007678), a co-promotion between Bosch Termotecnologia S.A. and the University of Aveiro, financed by Portugal 2020 under the Competitiveness and Internationalization Operational Program, and by the European Regional Development Fund. Additional support from IEETA Research Unit (UID/CEC/00127/2013). Samuel Silva funded by SOCA – Smart Open Campus (CENTRO-01-0145-FEDER-000010).

REFERENCES

[1] Jan Alexandersson, Bianka Buschbeck-Wolf, Tsutomu Fujinami, Michael Kipp, Stephan Koch, Elisabeth Maier, Norbert Reithinger, Birte Schmitz, and Melanie Siegel. 1998. *Dialogue acts in Verbmobil 2*. DFKI Saarbrücken.
[2] Amazon Alexa 2018. (2018). https://developer.amazon.com/alexa Accessed: 2018-02-27.
[3] Suket Arora, Kamaljeet Batra, and Sarabjit Singh. 2013. Dialogue System: A Brief Review. *arXiv preprint arXiv:1306.4134* (2013).
[4] Niels Ole Bernsen and Laila Dybkjær. 2009. *Multimodal usability*. Springer.
[5] Luis M Candanedo, Véronique Feldheim, and Dominique Deramaix. 2017. Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings* 140 (2017), 81–97.
[6] George Ferguson, Jill Quinn, Cecilia Horwitz, M Swift, J Allen, and Lucian Galescu. 2010. Towards a personal health management assistant. *Journal of biomedical informatics* 43, 5 (2010), S13–S16.
[7] David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A Kalyanpur, Adam Lally, J William Murdock, Eric Nyberg, John Prager, and others. 2010. Building Watson: An overview of the DeepQA project. *AI magazine* 31, 3 (2010), 59–79.
[8] Inbong Joo, Jongyoul Park, and Euihyun Paik. 2007. Developing ontology for intelligent home service framework. In *Consumer Electronics, 2007. ISCE 2007. IEEE International Symposium on*. IEEE, 1–6.
[9] Dan Jurafsky and James H Martin. 2017. *Speech and language processing* (third edition draft ed.).
[10] Maksym Ketsmur, Mário Rodrigues, and António Teixeira. 2017. DBpedia based factual questions answering system. *IADIS International Journal on WWW/Internet* 15, 1 (2017).
[11] Jangho Lee, Gyuwan Kim, Jaeyoon Yoo, Changwoo Jung, Minseok Kim, and Sungroh Yoon. 2016. Training IBM Watson using Automatically Generated Question-Answer Pairs. *arXiv preprint arXiv:1611.03932* (2016).
[12] Achraf Lyazidi and Salma Mouline. 2015. ONDAR: An ontology for home automation. In *Intelligent Systems Design and Applications (ISDA), 2015 15th International Conference on*. IEEE, 260–265.
[13] Stephen Makonin, Bradley Ellert, Ivan V Bajić, and Fred Popowich. 2016. Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014. *Scientific data* 3 (2016), 160037.
[14] Marion Mast, Heinrich Niemann, Elmar Nöth, and Ernst Günter Schukat-Talamazzini. 1995. Automatic classification of dialog acts with semantic classification trees and polygrams. In *International Joint Conference on Artificial Intelligence*. Springer, 217–229.
[15] David Milward and Martin Beveridge. 2003. Ontology-based dialogue systems. In *Proc. 3rd Workshop on Knowledge and reasoning in practical dialogue systems (IJCAI03)*. 9–18.
[16] Cathy Pearl. 2016. *Designing Voice User Interfaces: Principles of Conversational Experiences*. O'Reilly.
[17] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proc. CHI'18*. ACM.
[18] João Rato. 2016. *Conversação homem-máquina. Caracterização e avaliação do estado actual das soluções de speech recognition, speech synthesis e sistemas de conversação homem-máquina*. Master's thesis.
[19] John Searle. 1965. What is a speech act? (1965).
[20] António Teixeira, Daniela Braga, Luís Coelho, A Fonseca, Joaquim Alvarelhão, Inácio Martín, Alexandra Queirós, Nelson Rocha, António Calado, and Miguel Dias. 2009. Speech as the basic interface for assistive technology. In *Proc. DSAI'09*.
[21] António Teixeira, Flávio Ferreira, Nuno Almeida, Samuel Silva, Ana Filipa Rosa, José Casimiro Pereira, and Diogo Vieira. 2017. Design and development of Medication Assistant: older adults centred design to go beyond simple medication reminders. *Universal Access in the Information Society* 16, 3 (2017), 545–560.
[22] Alexandra Vtyurina and Adam Fourney. 2018. Exploring the Role of Conversational Cues in Guided Task Support with Virtual Assistants. (2018).
[23] Jingjing Xu, Yann-Hang Lee, Wei-Tek Tsai, Wu Li, Young-Sung Son, Jun-Hee Park, and Kyung-Duk Moon. 2009. Ontology-based smart home solution and service composition. In *Proc. IEEE Int. Conf. Embedded Software and Systems, 2009 (ICESS'09)*. 297–304.
[24] Steve Young, Milica Gašić, Blaise Thomson, and Jason D Williams. 2013. Pomdp-based statistical spoken dialog systems: A review. *Proc. IEEE* 101, 5 (2013), 1160–1179.