Improving Domotic Services Combining a Dialog System and a Resident Tracking System

David Antunes
Institute for Systems and Robotics,
L^2F - INESC-ID
Email: davidmiguel [at] antunes.net

David Martins de Matos
L^2F - INESC-ID
R. Alves Redol 9, Lisboa, Portugal
Email: david.matos [at] l2f.inesc-id.pt

José Gaspar
Institute for Systems and Robotics,
IST/UTL,
Av. Rovisco Pais 1, Lisboa, Portugal
Email: jag [at] isr.ist.utl.pt

***Abstract*—In this paper we propose the combination of a dialog system with a resident tracking system as a way to improve the available services, usefulness, and functionality of the smart home. Several services which are made possible by this combination are highlighted. The need of a resident tracking system and its advantages are explained. Several implications of the proposed approach are discussed.**

I. Introduction

There is plenty of room for improvement in domotics as they exist today. A recently built house is not fundamentally different from one built decades ago. Some of the systems which equip a house nowadays include: alarm systems, HVAC systems (Heating, Ventilation, and Air Conditioning), and domotic systems. These typically perform very simple functions, or require extensive configuration and customization to perform advanced functions (especially the available domotic systems). Better domotic systems should provide useful, advanced functionality with reduced one-time customization and reduced need for constant attention and effort to use the system by the user.

The ultimate goal of any domotic system is to provide useful services to the resident, preferably without requiring much input from him. For example, consider a lights remote control: if the resident has a wireless remote control with which he can control the house lights (on/off/dim/bright), has to carry it around the house all the time and press it four times to turn a light on, then it may just be better to use the light switch.

Dialog systems already provide a set of interesting domotic services nowadays, but they are often slow and inefficient because they may require a significant number of interactions to complete a task that would be very fast otherwise, for example to turn on a light. In this paper the combination of a dialog system and a tracking system is proposed as a way of providing new kinds of useful services in the smart home, many of which are not currently available. In the proposed services the dialog system takes the initiative of interacting with the resident, providing him for example with useful information or alerts. The paper will focus on the discussion of the implications of these systems combination, without detriment of the analysis of some implications specific to each system.

Soronen et al. [1] analyzed the attitudes and opinions of 1004 Finnish consumers toward voice interaction in the home environment. They found that most consumers dislike speaking to a device, and most (67%) feel that controlling a device with speech is unreliable. On the other hand, the respondents are much more accepting towards devices with voice feedback, which they find more acceptable and natural, and clearly prefer speech over beeps and blinking led's.

The organization of the paper is the following. First, the concept of tracking system will be presented. What it requires, and what it produces is specified. Then, the envisioned impact of its combination with a dialog system is presented, namely, the new services which can be provided by the dialog system. Finally, several implications of the proposed approach are discussed.

II. Tracking System

A resident tracking system operates in a house with several rooms and multiple residents. The system tracks the residents across a number of zones. A zone may correspond to one room, more than one room, or part of a room. It depends on the number and types of sensors available. The zones representation can be considered a graph where the zones are the nodes and their connections the edges. A detailed description of this representation is provided by Songhwai et al. [2]. The house has multiple sensors, one or more per zone, which may be of multiple types such as RFID, movement sensors, contact sensors, pressure mats and face recognition (cameras). The sensors produce triggers which are received and processed by the tracking system. The tracker integrates the information from the different sensors, and keeps a number of possible exclusive hypotheses on the positions of the residents in the house (they may be implemented as particles if a Monte-Carlo approach is used [3]). When a sensor trigger is received it is considered in each of the hypotheses and new hypotheses are generated containing the new information. The integration of multiple sensors over time allows the tracker to produce a correct output, even in the event that incorrect information is sometimes provided by the sensors. The tracker provides the prediction of the position of each of the residents.

Figure 1 presents an example of a tracking system. In the example, the tracker receives input from several different sensors: a movement sensor detects someone in the kitchen (this sensor does not provide the identity of the resident but the tracker may infer it using past information); an RFID tag

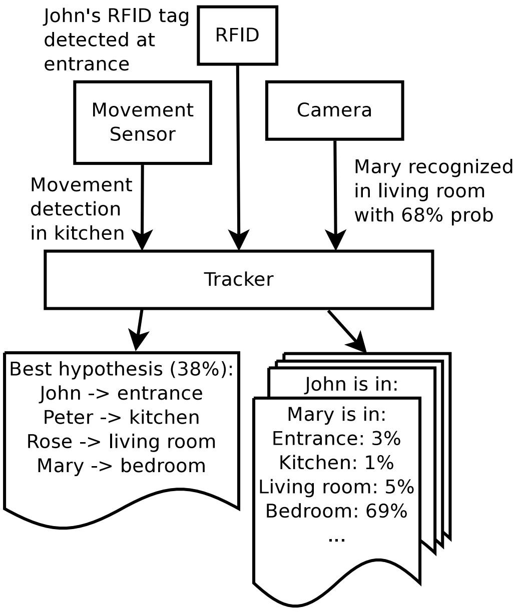

Fig. 1. Sensors and tracking system

reader installed at the entrance of the house reads John's RFID tag; and, a camera at the living room recognizes Rose with a confidence of 68%.

The tracker has to deal with sensor false alarms due to several factors, for example: light variations trigger the movement sensor, Rose is using John's door keys today and the RFID tag is attached to the keychain, or the living room was rather dark and the face recognition system failed to recognize correctly the person in the image, which was Peter and not Rose.

The tracker integrates sensor information over time, and uses it to resist to false alarms and to calculate probabilistically the position of each resident. In concrete terms, the tracking system may produce two different outputs. The first type of output, represented left in figure 1, is the best prediction of the system, it corresponds to a hypothesis on the state of the world (in this case the positions of the residents in the house). This hypothesis is consistent in terms of the position of each and every resident, that is, according to the hypothesis every resident is in some zone, and no resident is in two distinct zones. The hypothesis has an associated probability.

There is a second type of output that can be produced by the tracker. Consider the system is taking into account dozens or hundreds of different hypotheses. In the example scenario, the best hypothesis has a probability of 38% and, according to it, Mary is in the bedroom. If all that needs to be know is: "where is Mary?" then it can only be asserted that she is in the bedroom with *at least* 38% of certainty. To know exactly the probability of Mary being in the bedroom all hypotheses need to be taken into account, instead of using only the best one. Note that each hypothesis has a probability, all the hypothesis are competing and mutually exclusive, and only one of them is right (at most). The analysis of Mary's position in the other hypotheses may reveal that there is 100% certainty that Mary is in the bedroom (because Mary is there according to all other hypotheses), or even that all the other hypotheses place Mary in the living room (so there would be 62% probability that Mary is in the living room). Therefore, the tracking system produces an output with the probability of the location of each resident alone, taking in account all the hypotheses (represented in figure 1 by the output on the right). The system that uses the tracker, in this case the dialog system, must choose which of the two distinct outputs it will use (or use both).

Several resident tracking systems implementations have been proposed. Lee et al. [4] proposed an array of PIR (Pyroelectric InfraRed) sensors to be mounted at all the room's ceilings. Srinivasan et al. [5] uses height sensors in every doorway, obtaining the residents' identity reliably with a small number of sensors, and a non-intrusive approach. Wilson et al. [3] uses multiple types of anonymous binary sensors and learns the behavior of the various residents to obtain their identities.

III. Dialog System

The dialog system interacts with the residents. A well known dialog system architecture, proposed by Allen et al. [6] is composed of three main components: an interpretation manager interprets user input; a behavioral agent dictates the system behaviour based on its goals and obligations; and a generation manager produces utterances or updates a display. Interpretation and generation are coordinated using a shared state (the Discourse Manager).

A. Services

The dialog system interacts with the house residents. Through its combination with the tracking system it is possible to implement several services, for instance:

- Notifications of someone arriving home
- You received an email notification
- A friend posted on your facebook wall notification
- Read aloud an email, a facebook post
- Integration with a calendar (e.g. google calendar)
- Weather information when the resident is getting dressed in the morning
- Warn the resident that he is late for work in the morning
- Play ambient music (or radio)
- Provide appliances-related information, e.g. washing machine or dishwasher stopped
- "Nany" service: e.g. sending children to bed if parents are not home and it is past 10:00pm
- Provide evacuation instructions in case of emergencies (e.g. fire)
- Deterring intruders in case of an intrusion

Most of the proposed services do not require the resident to talk to the system. Therefore, to implement them, speech recognition would not be necessary. This has several advantages, such as reducing the cost of the system. Furthermore, speech recognition is still a very difficult and error-prone task, especially in the home environment. Oulasvirta et al. [7] analyzed communication failures in speech-based control in a smart home. They found that 26% of the exchanges of information between the system and the users involved one or more errors. But, even though there was a significant error rate, the users where 99.2% successful in completing the given tasks. Therefore, the problem is not effectiveness, but efficiency. With an average of between three and four commands per task and a high error rate, it may be much simpler for the resident to perform the task the usual way than by using the dialog system. Thus, the usefulness of such system would be questionable.

Recognition errors reduce the system usability. Aldrich [8] identified the insufficient attention paid to usability as one of the obstacles to consumer take-up of smart homes.

In the case of services involving only communication from the system to the resident, there are few possible errors and, if configured correctly to deliver useful information, the resident is always better off with the system than without it. That is, in a scenario where a resident has to request information from the system, there is the possibility that the effort required to use the system due to communication errors and multiple interactions to accomplish a single task greatly reduces the benefits of using the system. On the contrary, when the system takes the initiative of delivering useful information to the resident, the resident does not need to invest his time and effort trying to interact with it, therefore, the resident is always better off with the system than without it. Most of the presented services need no interaction from the resident to the system (the only exceptions are reading the full text of a received email or facebook post and playing music, where the user should request the system to start the service).

Therefore, two different kinds of interactions between the dialog system and the user can be distinguished. In the usual type of interaction, it is the resident who starts by addressing the system with a request, and there is a somewhat extensive question-and-answer interaction between the two, which may be error-prone and slow. In the proposed type of interaction, the system asynchronously interacts with the resident, providing him for example with useful information. As a positive aspect it presents little room for errors. However, it is not applicable to every kind of interactions (it is not applicable, for example, if the resident wants turn a light on).

Aldrich [8] addressed the question of “What is a smart home?”. To answer this question he proposed five hierarchical classes of smart homes, ordered in terms of both increased functionality and increased technical difficulty. The highest class of smart homes proposed by Aldrich are the “Attentive Homes” which are similar to the ones proposed in this paper. In the attentive home the activity and location of the residents is constantly registered, and it is used to control technology in a way that anticipates the residents needs. Aldrich suggests that the implementation of this type of homes is susceptible of introducing a pragmatic shift in the way residents live with domestic technology, and may offer a home environment qualitatively different from the existing ones.

B. Relationship with the tracking system

When trying to implement the proposed behavior with the existing dialog systems there will be several problems. For example, in a 15-division house equipped with a dialog system there is, presumably, a speaker/microphone in each division. When the system needs to send a notification to a resident, it must use all speakers in all rooms, because the information on the location of the resident is usually not available. This is very inconvenient. Furthermore, if there are four residents in the house and John receives an email, then all the residents will receive the notification of John’s email.

With a tracking system, the dialog system sends the notification directly to the appropriate resident. Thus, a tracking system is essential to implement an asynchronous system-to-user interaction behavior in which the initiative belongs to the system.

IV. Discussion

One important aspect to take into consideration when implementing the proposed approach is privacy. A resident may want to keep some kind of notifications private from the other residents (for example a new email notification, especially if part of the email subject is included in the notification). In this case, the system should allow the resident to classify some notifications as private and, in that case, if the resident is accompanied by other residents, it should save the notification and notify the resident only when he is alone. For even better privacy, the system should use a map to associate a speaker with the zones in which it can be heard. There is a trade off between losing important notifications and keeping privacy: if the resident classifies too many notifications as private, he may lose important ones (they may be delivered too late). On the other hand, a permissive privacy policy may lead to inconvenient situations. Without a tracking system, this privacy could not be ensured as the dialog system would have to send the notification to all zones in the house (as the residents’ location is unknown).

Usually, each interaction between a user and the dialog system is done with one (or more) microphones and one (or more) speakers in a room. But, with the introduction of a tracking system, it is possible to change the microphone and speaker being used, as the resident walks through the house. This could conceivably be done without tracking, but only for one resident. With a tracking system, multiple residents may be walking through the house and interacting with the system at the same time. This does not solve the problem of multiple residents in the same zone, which would require the dialog system to support speaker identification. The services that would take clear advantage of this feature are the ones in which there is a long interaction, for example reading a long email to the resident or playing ambient music: as the resident walks through the house he keeps listening to his email or radio station.

The difference between the proposed dialog systems approach and the existing one is similar to the difference between a servant and a steward. A servant needs to be told what to do, because he does not know his master's intentions. A steward anticipates the needs of the master and promptly gives him the information he needs at the right place and at the right time. A steward is commonly associated with a greater quality of service, greater usefulness, and higher cost. The proposed approach is similar to a steward while the existing approaches are closer to a servant.

There are also advantages for the tracking system in being combined with a dialog system. The dialog system may act as an additional sensor for the tracking system, using its microphones. In a simplistic manner, they can be used as a binary sensor which detects the presence of a resident by detecting sound. This is a fragile approach since TV sets, radios, washing machines, among others, may produce noises without the resident being present. A more reliable approach can be used if the dialog system is able to perform speaker identification. In this case, there is higher resilience to ambient noises, and it can even provide the identity of the resident to the tracker. Even if this identity is sometimes wrong, the tracker can deal with these mistakes as it integrates information from other sensors.

Some of the proposed services require some level of activity recognition. This has been addressed by Wilson et al. [3] who worked in simultaneous tracking and activity recognition. Furthermore, the existence of a tracking system allows a simpler recognition of many activities.

Although most examples given are related with notifications to the residents which supply them information, there are several other types of services which benefit from a tracking system (as discussed in section III-A). Those are services such as sending children to bed, directing residents away from danger in case of a fire, or intimidate and/or frighten an intruder while evacuating the residents from the house in case of an intrusion. These can only be implemented if the initiative of starting the interaction comes from the system and not from the person and, for these services, the location and identity of the residents (or intruders) is essential.

V. Conclusions

In this paper several advantages of combining a dialog system with a tracking system in the smart home environment have been presented. The proposed approach has been motivated by several problems of the existing dialog systems when applied to the smart home. The approach is based on the combination of a resident tracking system with a dialog system, and enables asynchronous communication from the system to the resident providing him with useful information, notifications, or other kinds of services. With this approach the need for residents to engage the system is greatly reduced, reducing recognition errors and the problem of extensive interactions to complete even simple tasks. Several scenarios which become possible through the combination of the two types of systems were identified. Future work includes the development of a prototype to evaluate quantitatively the improvements obtained with the proposed combination of systems.

Acknowledgments

This work was partially supported by the European Commission (EC) and is currently funded by the EU FP7 ICT-215554 project LIREC (Living with Robots and Interactive Companions). The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

References

[1] H. Soronen, M. Turunen, and J. Hakulinen, "Voice commands in home environment - a consumer survey," in *Proceedings of Interspeech 2008*, 2008, pp. 2078–2081.

[2] S. Oh and S. Sastry, "Tracking on a graph," in *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks*. Piscataway, NJ, USA: IEEE Press, 2005, p. 26.

[3] D. Wilson and C. Atkeson, "Simultaneous tracking & activity recognition (star) using many anonymous, binary sensors," in *In The Third International Conference on Pervasive Computing*. Springer-Verlag, 2005.

[4] S. Lee, K. N. Ha, and K. C. Lee, "A pyroelectric infrared sensor-based indoor location-aware system for the smart home," *Consumer Electronics, IEEE Transactions on*, vol. 52 Issue 4, pp. 1311–1317, 2006.

[5] V. Srinivasan, J. A. Stankovic, and K. Whitehouse, "Using height sensors for biometric identification in multi-resident homes," in *Pervasive*, 2010, pp. 337–354.

[6] J. Allen, G. Ferguson, and A. Stent, "An architecture for more realistic conversational systems," in *Proceedings of the 6th international conference on Intelligent user interfaces*, ser. IUI '01. New York, NY, USA: ACM, 2001, pp. 1–8.

[7] A. Oulasvirta, K. P. Engelbrecht, A. Jameson, and S. Moller, "Communication failures in the speech-based control of smart home systems," in *Intelligent Environments, 2007. IE 07. 3rd IET International Conference on*, Sept. 2007, pp. 135–143.

[8] F. Aldrich, "Smart homes: Past, present and future," in *Inside the Smart Home*, R. Harper, Ed. Springer London, 2003, pp. 17–39.