Journal of Ambient Intelligence and Smart Environments 1 (2009) 31–36
DOI 10.3233/AIS-2009-0004
IOS Press

The role of spoken language dialogue interaction in intelligent environments

Wolfgang Minker[a,*], Ramón López-Cózar[b] and Michael McTear[c]

[a] *Ulm University, Institute of Information Technology, 89081 Ulm/Donau, Germany*

[b] *Dept. of Languages and Computer Systems, Faculty of Computer Science and Telecommunications, 18071 University of Granada, Spain*

[c] *School of Computing and Mathematics, University of Ulster, Newtownabbey BT37 0QB, Northern Ireland*

Abstract. An Intelligent Environment is a physical space that becomes augmented with computation, communication and digital content, thus transcending the limits of direct human perception. Spoken dialogue is a key factor for user-friendly human-computer interaction. This article details how to integrate Spoken Dialogue Systems into Intelligent Environments. We will outline research areas and future trends including assistive, adaptive and proactive system design, dialogue management and system-environment interaction.

Keywords: Speech, mobile devices, information access, Spoken Dialogue System

1. Introduction

Thanks to breathtaking hardware miniaturization and cost reduction, everyday environments (e.g., home, office, car, etc.) may be populated today with *smart* devices for controlling and automating various tasks in our daily lives. These environments are obviously changing into *intelligent environments*.

Smart devices are mostly used by non-specialists and increasingly frequently by disabled persons [9] without a particular knowledge of complex computer equipment and in their usual context of life. Such systems should therefore be easy to use, non-intrusive and exploit the most natural communication means. Undeniably, enhanced communication and assistive capabilities increase the usability and social acceptability of this kind of system.

Spoken natural dialogue is a key factor for a user-friendly and consistent user-device interaction in intelligent environments. Providing an easy access to these systems, Spoken Dialogue Systems (SDSs) have become an increasingly important interface between humans and computers as they constitute the most natural way of communication.

An overview of an SDS is shown in Fig. 1. After acoustic analysis to clean the speech signal from ambient noise, the input utterance is automatically recognised [19]. The output is then passed to the natural language understanding component, which determines the meaning of the utterance. Human-computer interaction is a matter of interactive and incremental problem solving with both the user and the computer playing active roles in the conversation. At the end, the user utterance needs to be interpreted in the context of the ongoing dialogue, taking into account common sense and task domain knowledge. Meaning representations corresponding to the current utterance are completed using the dialogue history taking into account all the information given by the user earlier in the dialogue. If this information is insufficient, ambiguous or if the application database does not contain the information requested, the dialogue manager may ask the user for clarification and feedback. An application access interface uses the meaning representation to generate a database query. The retrieved information is fi-

[*] Corresponding author. E-mail: wolfgang.minker@uni-ulm.de.

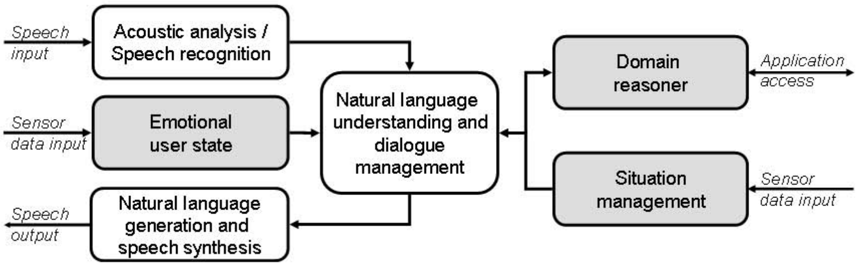
Speech input
Acoustic analysis / Speech recognition
Sensor data input
Emotional user state
Speech output
Natural language generation and speech synthesis
Natural language understanding and dialogue management
Domain reasoner
Application access
Situation management
Sensor data input

Fig. 1. Increase efficiency and user-friendliness of human-computer interfaces. The dialogue management component of an SDS processes input from additional knowledge sources, the user status and the situation of use.

nally presented in the form of speech, text, tables or graphics.

In this article, we will highlight research areas that are in our view essential for integrating SDSs into future Intelligent Environments. Section 2 discusses general systems properties that should make the interaction more flexible and user friendly. Enhancing the assistive capabilities of such interfaces will allow the reduction of the cognitive load of the user during the interaction and make human-computer interfaces behave more like dialogue partners. Section 3 presents a more detailed overview of dialogue management as the central part of SDSs. The development of powerful dialogue management and control represents one of the main challenges at this high interaction level. Section 4 addresses the interaction of the system with its environment. This is managed by middleware. We discuss current approaches to middleware design and show challenges ahead.

2. Assistiveness, adaptivity and proactiveness

International research projects have been focussing on spoken dialogue interaction for more than 20 years. For the lower language levels (such as speech recognition) and reasonably simple dialogue applications (such as information access) the SDS technology has reached a good and even commercializable performance level. However, SDS that need to support a complex and unconstrained dialogue interaction in different conditions of use are still the subject to research.

There is in particular a need to investigate on how to improve the human-computer interaction by endowing SDSs with more intelligence so that such systems are not only able to retrieve information, but also to integrate information from multiple sources and to resolve potential conflicts and problems that may occur if the context of use changes. In our view, an *assistive and empathic SDS* is a competent and sensitive complex multi-functional technical system [11], able to perceive and to interact in a complex and dynamically varying environment. It is able to transform perceptions into a model-based internal representation, to perform reasoning on acquired information and finally, to react accordingly, i.e., to generate and to perform actions based on the information at hand.

In summary, future SDSs should be *communicative and interactive*, as they collaborate with users, systems and the environment in different multimodal ways. They should also be *assistive* as they yield sufficient background knowledge and have the ability to reason about this knowledge. They understand user requests and act accordingly. If required, they are also able to explain their own behaviour and actions. Future SDSs also need to be *proactive*, i.e. behave as autonomous dialogue partners. They track the dialogue and react if required by the conversational situation. Finally, next-generation SDSs should be *adaptive and empathic* as they change the interaction style depending on the user's peculiarities, intentions and emotions. I.e., the interface identifies and understands the feelings, ideas, and (personal) circumstances of its users.

The SDS technologies that are particularly important for intelligent environments include: robust natural language processing (recognition and understanding), adaptive and proactive dialogue modelling, multimodality, intelligent planning, monitoring and plan adaptation, embedded agents, reasoning, inferencing, flexible end device technology and networking. Some of these issues will be addressed in the following.

Assistiveness Nowadays, users benefit from an access to an increasing number of application domains and functionalities in all situations of their lifes, i.e. at home, at work and on the move. It seems evident that state-of-the-art human-computer interfaces need to be able to perform more efficiently. In particular, managing complex and interdependent tasks with these applications requires a high cognitive burden from the user, which may even be dangerous in certain environments (e.g., whilst driving). We think that approaches to represent and process the knowledge involved in the problem solving processes will be a cornerstone in the architecture of more assistive and efficient SDSs [2].

We claim that an assistive SDS for a specific application should be capable of reasoning in its domain, integrate information from different sources, and be able to contribute valuable inferences that lead to the construction of mutually accepted solutions.

Specifically, a domain reasoner inroduced between the dialogue management and the application backend (cf. Fig. 1) provides a common knowledge representation. Such a representation is required to integrate information from the user as well as from various other domain-specific information sources in terms of constraints, i.e, assertions in a logical language. This common representation enables a system of rules to combine the received constraints in order to generate inferences (in particular, conflict detections) that may initiate negotiation dialogues between the user and the system. The domain reasoner also provides additional information for the dialogue manager. This information serves as a basis to engage in explanation and conflict resolution dialogues and to enable hypothetical reasoning by maintaining concurrent partial solutions (possible worlds) for these dialogues.

Proactiveness Mechanisms learned from human-human communication may serve as a basis to further improve the human-computer interaction and may lead to an equivalent partnership between user and system in the communication [17]. Computer systems need to adopt certain features from the way humans communicate and have to learn to respect social rules that prevail in human conversation. Next-generation computer systems will adopt behavioural patterns of the users to act and react appropriately. Furthermore, they will be context aware. Capturing the spacial, temporal, and user specific context of an interaction also paves the way to proactive behaviour. This implies that the system takes initiative of its own accord in contrast to merely responding to the actions performed by the user.

A proactive SDS gets involved in a conversation when required by the situation and not only upon the user's request. In turn, it holds back when it is not needed anymore, for instance if the task has been solved. It proposes solutions to problems, possibly even before the users have become aware of the problem. Proactiveness requires a contextual understanding of the conversational flow at any time. The system therefore needs to understand the users' psychological situation, intentions, and actions. It has to maintain an elaborated dialogue history. Only then is it able to contribute to the conversation in a meaningful way.

Adaptivity The involvement of emotions in assistive human-computer interfaces has emerged as a recent field of research. There are a large number of different approaches to the characterisation of human emotions. There is now a need to handle these emotional parameters, i.e., to enable a more natural and user-friendly human-computer interaction.

The knowledge of the user's current emotional state and the adaptation of the SDS according to that knowledge contribute to an improvement of the interaction, e.g., by shaping the dialogue flow in a more natural way [14]. Reeves and Nass [15] have shown that computers are perceived as partners to whom users apply social norms. An SDS that incorporates these factors needs to be aware of the user's current emotional state and to process this information in the dialogue management component (Fig. 1).

3. Dialogue management

The dialogue manager is the central component of SDSs, accepting interpreted input from the speech recognition and natural language understanding components, interacting with external knowledge sources, producing messages to be output to the user, and generally controlling the dialogue flow (cf. Fig. 1). The dialogue management process can be viewed in terms of two main tasks: dialogue modelling and dialogue control.

Dialogue modelling Dialogue modelling involves keeping track of the state of the dialogue. Dialogue state information may be encoded implicitly in a dialogue graph or in a form consisting of one or more slots to be filled with values elicited in the course of the dialogue. More complex information can be represented, such as the mental states of the participants, their discourse obligations, and their overall and immediate goals and plans, as in Information State Theory [7]. This information is used to support the system's interpretation of the user's utterances and to determine the system's next actions.

While generally dialogue modelling is represented in terms of a single hypothesis about the current dialogue state, an alternative approach is to maintain multiple dialogue state hypotheses in which probabilities are assigned to the different hypotheses [18].

Dialogue control Dialogue control involves deciding what to do next in the context of the current dialogue state. Decisions may include prompting the user for more input, clarifying or grounding the user's previous input, or outputting some information to the user. These decisions may be pre-scripted, with choices based on factors such as the confidence levels associated with the user's input. Alternatively dialogue control may involve decisions taken dynamically based on reasoning about the current dialogue state and using evidence from a combination of different domain and dialogue knowledge sources.

Dialogue control has traditionally required careful design and handcrafting of rules and strategies [3]. However, a major problem that it is difficult to anticipate all the rules that would be required to cover all aspects of dialogue control. An alternative approach is to use a data-driven approach such as reinforcement learning to optimize the dialogue control. In reinforcement learning the system's priorities are specified in a (real-valued) reward function and an optimization algorithm is applied to choose those actions that maximize that function. In the earliest applications of reinforcement learning to spoken dialogue systems, dialogue was formalised as a Markov decision process (MDP) (see, for example, [8]) while more recently POMDPs (Partially Observable Markov Decision Process) have been used as a way of handling the various uncertainties inherent in dialogue interactions (see, for example, [18]).

4. System-environment interaction

One important feature of SDSs designed to work in intelligent environments is that they must interact not only with the user but also with the environment where the user is located (cf. situation management Fig. 1). To do this, they must comply with a number of requirements. One is that they must process data provided by a diversity of sensors placed in the environment which capture information from the user, e.g. microphones, cameras and presence detectors. Another requirement is that they may need to change the status of devices in the environment, for example, switch off a light or turn on TV by operating the corresponding actuators. These requirements impose new challenges to be addressed by the research community, given the wide range of heterogeneous devices, sensors and actuators consituting embedded systems which exist in the market.

In order to ease the implementation of SDSs to be used in ambient intelligence applications, researchers typically employ a *middleware* layer that represents characteristics of the environment, e.g. interfaces and status of devices. Using this layer, the SDS does not need to interact directly with the devices in the physical world, thus getting rid of their specific peculiarities. The communication between the entities in the middleware and the corresponding physical devices can be implemented using standard access and control mechanisms, such as EIB (European Installation Bus) [12,13] or SNMP (Simple Network Management Protocol) [10].

Researchers have implemented this middleware in different ways. For example, Sachetti et al. [16] proposed a middleware called WSAMI (Web Services for Ambient Intelligence) to support seamless access to mobile services for the mobile user, either pedestrian, in a car or in a public transport. This middleware builds on the Web services architecture, whose pervasiveness enables the availability of services in most environments. In addition, the proposal deals with the dynamic composition of applications in a way that integrates services deployed on mobile, wireless and on the Internet. The authors developed three key WSAMI-compliant Web services to offer intelligent-aware multimodal interfaces to mobile users: i) linguistic analysis and dialogue management, ii) multimodality and adaptive speech recognition, and iii) context awareness.

Following a different approach, Montoro et al. [12] implemented a middleware using a blackboard [4] created from the parsing of an XML document that represents ontological information about the environment. To access or change the information in the blackboard, applications and interfaces employ a simple communication mechanism using XML-compliant messages which are delivered via HTTP. An interesting feature of this proposal is that it allows the attachment of linguistic information to each entity in the middleware in order to automatically create a spoken interface associated with the corresponding device. This linguistic information comprises most of the possible ways that a user may employ to interact with the device.

The diversity of middleware proposals leads to a problem of heterogeneity, i.e., an application implemented on a specific middleware cannot interoperate with services developed on another. Hence, research effort must be devoted to overcome this drawback by addressing issues concerned with interoperability, as well as service discovery and access.

Another drawback of current middleware implementations is concerned with software evolution. The middleware must allow the addition of new functionalities and adapt the already available applications to technological changes. An example is to easily adapt existing applications when new communication technologies appear. Aspect Oriented Programming (AOP) [5] and Aspect Oriented Software Development (AOSD) [1] have been proposed in recent years to deal with this problem.

Finally, a third problem to be addressed in the near future is the lack of standard and validated tools that ease the development of applications on the middleware. Among other initiatives, the EU-funded project Hydra [6] has recently aimed at overcoming this drawback by developing a set of tools for application, device and solution developers.

5. Conclusion

In this article we have presented a selection of research areas that we consider to be essential for integrating SDSs into future Intelligent Environments.

The capability to flexibly integrate information from various sources and negotiate solutions between these sources and the human user is another aspect that we consider important for managing the rising complexity of today's and future technical systems. Assistive and proactive dialogue behaviour in particular contributes to this endeavour, since it guarantees that the user is informed in the right way at the right time. Another aspect is the use of appropriate system architectures that allow us to integrate the technology in everyday – even small and less powerful – devices. This seems to be essential so as to make these interfaces accessible for a large public and to increase their usability and acceptability.

The current pace of technological development will continue and produce further challenges for human-computer interfaces. Rather than adapting to each new technology that will appear, we are convinced that natural and assistive SDSs will be an integral part of the future technological environment that we will live in.

Acknowledgements

Parts of the research described in this article will be supported by the Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" funded by the German Research Foundation (DFG).

References

[1] AOSD web site. [Online]. Available: http://www.aosd.net.

[2] D. Bühler and W. Minker, *Reasoning for Information-Seeking and Planning Dialogues*, ser. Text, Speech and Language Technology. Dordrecht (The Netherlands): Springer, 2009, to appear.

[3] M. Cohen, J. Giangola, and J. Balogh, *Voice User Interface Design*. Addison Wesley, Boston, USA, 2004.

[4] R. Engelmore and T. Mogan, *Blackboard systems*. Addison-Wesley, 1988.

[5] L. Fuentes and D. Jiménez, "An aspect-oriented ambient intelligence middleware platform," in *MPAC '05: Proceedings of the 3rd international workshop on Middleware for pervasive and ad-hoc computing*. New York, NY, USA: ACM, 2005, pp. 1–8.

[6] Hydra project web site. [Online]. Available: http://www.hydramiddleware.eu.

[7] S. Larsson and D. Traum, "Information state and dialogue management in the trindi dialogue move engine toolkit," *Natural Language Engineering*, vol. 5, no. 3-4, pp. 323–340, 2000.

[8] E. Levin, R. Pieraccini, and W. Eckert, "A stochastic model of human-machine interaction for learning dialog strategies," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 1, pp. 11–23, 2000.

[9] S. Lockwood and D. Cook, "Computer, light on!" in *The 4th IET International Conference on Intelligent Environments*, Seattle (USA), 2008.

[10] A.E. Martínez, R. Cabello, F.J. Gómez, and J. Martínez, "INTERACT-DM. A solution for the integration of domestic devices on network management platforms," in *Proc. of IFIP/IEEE International Symposium on Integrated Network Management*, Colorado Springs, Colorado, US, 2003, pp. 360–370.

[11] W. Minker, J. Pittermann, A. Pittermann, P.-M. Strauss, and D. Bühler, *Speech Communication at the Leading Edge*. Hauppauge, NY (USA): Nova Science Publishers, Inc., 2008, ch. Intelligent and Empathic Speech Interfaces, pp. 71–105, 2008.

[12] G. Montoro, X. Alamán, and P. Haya, "A plug and play spoken dialogue interface for smart environments," in *Proc. of Fifth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'04)*, Seoul, South Korea, 2004, pp. 360–370.

[13] A.A. Nazari, "A generic UPnP architecture for ambient intelligence meeting rooms and a control point allowing for integrated 2D and 3D interaction," in *Proc. of Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-Aware Services, Usages and Technologies*, 2005, pp. 207–212.

[14] J. Pittermann and A. Pittermann, "Integrating emotion recognition into an adaptive spoken language dialogue system," in *The 2nd IET International Conference on Intelligent Environments*, Athens, Greece, July 2006.

[15] B. Reeves and C. Nass, *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press, 1996.

[16] D. Sachetti, R. Chibout, V. Issarny, C. Cerisara, and F. Landragin, "Seamless access to mobile services for the mobile user," in *Proc. of IEEE Int. Conference on Software Engineering*, Beijing, China, 2004, pp. 801–804.

[17] P.-M. Strauss, H. Hoffmann, W. Minker, H. Neumann, G. Palm, S. Scherer, H. Traue, and U. Weidenbacher, "The PIT Corpus Of German Multi-Party Dialogues," in *6th International Conference on Language Resources and Evaluation*, Marrakech (Morocco), 2008.

[18] J. Williams and S. Young, "Partially observable Markov decision processes for spoken dialog systems," *Computer Speech and Language*, vol. 21, no. 2, pp. 231–422, 2007.

[19] S.J. Young, *HTK V1.4 User, Reference Programmer Manual*, Cambridge University Engineering Dept, Speech Group, Aug. 1992.