Advanced User Assistance for Setting Up a Home Theater

Pascal Bercher, Felix Richter, Thilo Hörnle, Thomas Geier, Daniel Höller,
Gregor Behnke, Florian Nielsen, Frank Honold, Felix Schüssel, Stephan Reuter,
Wolfgang Minker, Michael Weber, Klaus Dietmayer, and Susanne Biundo

Abstract In many situations of daily life, such as in educational, work-related, or social contexts, one can observe an increasing demand for intelligent assistance systems. In this chapter, we show how such assistance can be provided in a wide range of application scenarios – based on the integration of user-centered planning with advanced dialog and interaction management capabilities. Our approach is demonstrated by a system that assists a user in the task of setting up a complex home theater. The theater consists of several hi-fi devices that need to be connected with each other using the available cables and adapters. In particular for technically inexperienced users the task is quite challenging due to the high number of different ports of the devices and because the used cables might not be known to the user. Support is provided by presenting a detailed sequence of instructions that solves the task.

1 Introduction

In many situations of daily life, such as in educational, work-related, or social contexts, one can observe an increasing demand for intelligent assistance systems. *Companion*-Technology can be used to design and implement systems that provide intelligent assistance in a wide range of application areas. For an introduction to *Companion*-Technology we refer to Chap. 1 and for a survey to the article by Biundo et al. [5]. Here, we demonstrate a prototypical *Companion*-System that provides in-

---

Pascal Bercher · Felix Richter · Thilo Hörnle · Thomas Geier · Daniel Höller · Gregor Behnke · Florian Nielsen · Frank Honold · Felix Schüssel · Stephan Reuter · Wolfgang Minker · Michael Weber · Klaus Dietmayer · Susanne Biundo
Faculty of Engineering, Computer Science and Psychology,
e-mail: forename.surname@uni-ulm.de (use *oe* instead of *ö* and *ue* instead of *ü*) Please use the address forename.surname@alumni.uni-ulm.de for the authors *Felix Richter* and *Thomas Geier*. For Florian Nielsen, please use florian.nothdurft@alumni.uni-ulm.de.

telligent assistance with the task of setting up a complex home theater [1, 2]. The respective system is also presented in a short video [9].

The home theater consists of several hi-fi devices that need to be connected with each other using the available cables and adapters. In particular for technically inexperienced users the task is quite challenging due to the high number of different ports of the devices, and because the user might not be familiar with the different kinds of available cables. Support is provided by presenting a detailed sequence of individualized instructions that describe how to solve the task. In case of unexpected changes in the environment, for example if a cable turns out to be damaged, the system adapts to the new situation and presents an alternative solution. It can further adapt its multimodal user interface to changes in the context of use (CoU) and is able to provide explanations about the presented instructions. Within the SFB/TRR 62 [6], this *Companion*-Technology has been implemented and demonstrated in a prototype system that integrates advanced user-centered planning capabilities (cf. Chap. 5) with advanced dialog and user interaction capabilities (cf. Chaps. 9 and 10, respectively) to ensure an individual, convenient, and highly flexible interaction with the system.

2 Application Scenario: The Home Assembly Task

Our system is capable of assisting a human user in the task of setting up his or her home theater. The home theater consists of several hi-fi devices which need to be connected with each other using different types of cables and adapters, such that all components correctly work together.

In a fully general assistance system of the depicted application domain, the currently available hi-fi devices and cables and adapters had to be selected by the user. For demonstration purposes, we fixed one specific setting, where the following devices are given: a television for visual output, an audio/video receiver with attached boxes for audio output (video signals can be looped through that receiver, though), a Blu-ray player, and a satellite receiver. For these devices to be fully functional, the television needs to receive the video signals of the Blu-ray player and the satellite receiver, and the audio/video receiver needs to receive their audio signals.

3 System Description

The proposed assistance system uses the technical specification of the available hardware and automatically generates a course of action that solves the assembly task. That is, we modeled the assembly task as a hybrid planning task[1], the solution to which is then presented step-by-step to the user in an adequate way to enable

---

[1] The hybrid planning model is deterministic. A non-deterministic variant of the planning task (a hierarchical POMDP model) is considered in Chap. 6.

a natural interaction with the system. The respective system is a knowledge-based system that implements a generic architecture (cf. Chaps. 22 and 10) that can be used as a basis for the creation of intelligent assistance systems in a wide range of application scenarios [1, 10]. It comprises different modules, which are described in the sequel.

Knowledge Base. The knowledge base serves as a central hub for any domain and user information.[2] It integrates information from various sources, such as data coming from sensors – usually preprocessed through machine learning algorithms – or explicit user input originating from system components responsible for human-computer interaction.

Besides storing information central to the application domain, the main purpose of the knowledge base is the integration of sensory data. Our system can be configured in such a way that – except for touch or sound sensors to gain user input – it works without use of further sensors. However, we can also use laser sensors to track the user's position [7, 18, 19] in order to determine the most appropriate output modality [12]. To cope with uncertainty and to facilitate the concise representation of large graphical models, the knowledge base uses Markov Logic [20] as modeling language. Since the planning components of the system are based on a deterministic model, they require a special treatment. To make the probabilistic view of the knowledge base compatible with this deterministic representation, we consider the most probable world state according to the current beliefs as the actual world state.

Planning, Dialog, and Interaction: An Interactive Cycle. We modeled the assembly task as a *hybrid planning problem*, as it enables to provide advanced user assistance [4]. For a detailed discussion about the user-centered planning capabilities that it integrates and further arguments why it is well-suited for planning with or for humans, we refer to Chap. 5.

Each presented instruction is a single planning action that specifies where the specific ends of the cables and adapters have to be plugged into. Before any instruction is presented to the user, a plan must be found that solves the specified problem. In hybrid planning, plans are partially ordered sequences of actions. Actions have a name (in our example scenario, there are only *plugIn* actions) and a sequence of parameters that specify the involved devices and their ports that are used to connect the devices with each other. For instance,

$$\mathit{plugIn}(\textsc{scart-cable}, \textsc{audio-port}_1, \textsc{av-receiver}, \textsc{audio-port}_2) \tag{1}$$

depicts an action describing that the audio end of the SCART cable should be connected with a specific audio port of the audio/video receiver. Since plans are only partially ordered, but actions need to be presented one after another, one first needs

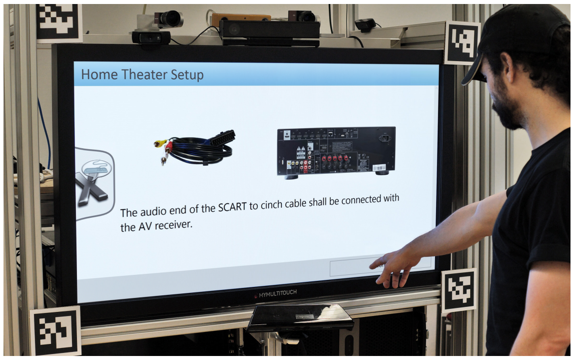
---

[2] The instances of the user and environment model can be loaded from a pre-defined source and altered at runtime by updating their respective properties. These two context model instances can be displayed and manipulated on additional devices, e.g., via a touch-enabled tablet computer or an interactive surface using tangible interaction.

to find an action linearization that is plausible to the user [8]. After such a sequence is found, they need to be communicated step-wise in an adequate way.

For this, actions are passed one-by-one to the dialog management (cf. Chap. 9). Here, every received action is mapped to a dialog model [1, 13] specifying several so-called dialog goals that need to be achieved. Every dialog goal may achieve some or all of the required effects defined by the action. The dialog goals may be structured in a hierarchical way: actions may be decomposed into several user-adapted sequences of so-called dialog steps suitable for the individual user (e.g., novice users receive more detailed instructions distributed to several dialog steps instead of a single one). These dialog steps are modality-independent representations referencing the content to be communicated to the user, where modality means a combination of interaction language plus device component. Verbal intelligence can also be addressed [15]. This allows for user-adaptive selections of the most appropriate dialog steps. The resulting sequence implements those effects of the given action that need to be achieved in a user interaction. Graphical approaches can assist a domain expert in the modeling process of dialog models [3].

Each dialog step is passed on to the fission component (cf. Chap. 10). The fission is responsible for modality arbitration of the respective dialog step [10]. Each of the referenced information items within such a dialog step can be realized with the use of different interaction languages such as video, automatic speech recognition, or text. For instance, depending on the current CoU, it might be inappropriate for the system to read instructions aloud, but to present them using written text or graphics, instead. Further, the fission decides which information fragment is realized via which device component. One example for a possible output of a dialog step is presented in Fig. 1. In this case, the respective action (cf. Eq. (1)) is reflected using only one single dialog step. In an empirical evaluation we studied adaptation rules for the fission to improve user interaction [22].



The audio end of the SCART to cinch cable shall be connected with the AV receiver.

Fig. 1 The user interacts with the system. It displays the instruction representing the action given in Eq. (1). The respective port of the audio/video receiver depicted is flashing in red.

At any point in time during interaction with the system, the system receives user input. This input might be implicit (like the position of the user) or explicit (such as pointing gestures for selection purposes). Since the user often interacts in a multimodal fashion (such as pointing and using speech at the same time), these interactions need to be fused to a coherent system input [23]. Sensor failures or fuzziness can be limited or even resolved when incorporating the interaction history of the user [21]. To resolve remaining conflicts and uncertainty, the system can initiate additional questions to the user [11].

Assuming the instructions are correctly carried out, all components of the home theater will be correctly connected with each other. Note that we do not have sensors to detect execution failures. Those need to be reported by the user manually. So, at any point in time during interaction with the system, the user may state execution errors. For instance, in case one of the cables breaks down, he may state so using either speech or text input (as for example: "The cable is broken!"). In such a case, the system marks the respective cable as unusable and initiates plan repair to find a repaired solution to the problem that does not use that specific cable [1]. The plan repair technique could also handle other unexpected failures (such as unusable ports, for example), but these are currently not handled by the system.

In particular, when execution errors occur and a plan needs to be changed, questions might arise. The user might wonder, for example, what purpose a presented instruction serves (as he or she might be asked to use a different type of cable than before the execution failure, for instance). Such a question, i.e., the demanded justification for a specific instruction, can be answered by the plan explanation component [1, 24]. It derives a sequence of arguments that explains the purpose of the instruction in question. In an empirical evaluation we studied the benefit of such plan explanations in the given scenario and how the assembly task assistance system is perceived in general [1]. In a different scenario, an empirical evaluation was conducted to study how providing justification and transparency explanations may improve the user-experience in unexpected situations (e.g., the change of plans) [17]. In addition, the system can explain declarative facts (such as technical details about a specific cable) and may decide when to present these (declarative) explanations depending on the current user knowledge and the currently presented instruction [14, 15, 16].

4 Conclusion

We presented the integration and interaction of dialog and interaction management, probabilistic reasoning, and planning capabilities in a single system. That system realizes several *Companion* properties: it provides its assistance functionality in an individual manner, adapts its behavior in the light of unforeseen complications, and actively maintains its user's trust by offering various explanation capabilities.

Acknowledgements This work is done within the Transregional Collaborative Research Centre SFB/TRR 62 "*Companion*-Technology for Cognitive Technical Systems" funded by the German Research Foundation (DFG). We also want to thank our former colleagues Gregor Bertrand, Peter Kurzok, Louisa Pragst, Bernd Schattenberg, and Bastian Seegebarth for their work for the demonstrated system. We also acknowledge Roland Barth and Matthias Rau for their contributions in producing the video [9] explaining the system.

References

1. Bercher, P., Biundo, S., Geier, T., Hörnle, T., Nothdurft, F., Richter, F., Schattenberg, B.: Plan, repair, execute, explain - how planning helps to assemble your home theater. In: Proc. of the 24th Int. Conf. on Automated Planning and Scheduling (ICAPS), pp. 386–394. AAAI Press (2014)
2. Bercher, P., Richter, F., Hörnle, T., Geier, T., Höller, D., Behnke, G., Nothdurft, F., Honold, F., Minker, W., Weber, M., Biundo, S.: A planning-based assistance system for setting up a home theater. In: Proc. of the 29th Nat. Conf. on AI (AAAI), pp. 4264–4265. AAAI Press (2015)
3. Bertrand, G., Nothdurft, F., Honold, F., Schüssel, F.: CALIGRAPHI - Creation of Adaptive Dialogues Using a Graphical Interface. In: 35th Annual 2011 IEEE Computer Software and Applications Conference (COMPSAC), pp. 393–400. IEEE (2011)
4. Biundo, S., Bercher, P., Geier, T., Müller, F., Schattenberg, B.: Advanced user assistance based on AI planning. Cognitive Systems Research **12**(3-4), 219–236 (2011). Special Issue on Complex Cognition
5. Biundo, S., Höller, D., Schattenberg, B., Bercher, P.: Companion-technology: An overview. Künstliche Intelligenz **30**(1), 11–20 (2016). DOI 10.1007/s13218-015-0419-3. Special Issue on Companion Technologies
6. Biundo, S., Wendemuth, A.: *Companion*-technology for cognitive technical systems. Künstliche Intelligenz **30**(1), 71–75 (2016). DOI 10.1007/s13218-015-0414-8. Special Issue on Companion Technologies
7. Geier, T., Reuter, S., Dietmayer, K., Biundo, S.: Track-person association using a first-order probabilistic model. In: Proc. of the 24th IEEE Intl. Conf. on Tools with AI (ICTAI 2012), pp. 844–851. IEEE (2012)
8. Höller, D., Bercher, P., Richter, F., Schiller, M., Geier, T., Biundo, S.: Finding user-friendly linearizations of partially ordered plans. In: 28th PuK Workshop "Planen, Scheduling und Konfigurieren, Entwerfen" (2014)
9. Honold, F., Bercher, P., Richter, F., Nothdurft, F., Geier, T., Barth, R., Hörnle, T., Schüssel, F., Reuter, S., Rau, M., Bertrand, G., Seegebarth, B., Kurzok, P., Schattenberg, B., Minker, W., Weber, M., Biundo, S.: Companion-Technology: Towards User- and Situation-Adaptive Functionality of Technical Systems. In: 10th Int. Conf. on Intelligent Environments, pp. 378–381. IEEE (2014). URL http://companion.informatik.uni-ulm.de/ie2014/companion-system.mp4
10. Honold, F., Schüssel, F., Weber, M.: Adaptive probabilistic fission for multimodal systems. In: Proc. of the 24th Australian Computer-Human Interaction Conference, OzCHI '12, pp. 222–231. ACM (2012)
11. Honold, F., Schüssel, F., Weber, M.: The automated interplay of multimodal fission and fusion in adaptive HCI. In: 10th Int. Conf. on Intelligent Environments, pp. 170–177 (2014)
12. Honold, F., Schüssel, F., Weber, M., Nothdurft, F., Bertrand, G., Minker, W.: Context models for adaptive dialogs and multimodal interaction. In: 9th Int. Conf. on Intelligent Environments, pp. 57–64. IEEE (2013)
13. Nothdurft, F., Bertrand, G., Heinroth, T., Minker, W.: GEEDI - guards for emotional and explanatory dialogues. In: 6th Int. Conf. on Intelligent Environments, pp. 90–95. IEEE (2010)
14. Nothdurft, F., Honold, F., Kurzok, P.: Using explanations for runtime dialogue adaptation. In: Proc. of the 14th ACM Int. Conf. on Multimodal Interaction, pp. 63–64. ACM, ACM (2012)
15. Nothdurft, F., Honold, F., Zablotskaya, K., Diab, A., Minker, W.: Application of Verbal Intelligence in Dialog Systems for Multimodal Interaction. In: 10th Int. Conf. on Intelligent Environments, pp. 361–364. IEEE (2014)
16. Nothdurft, F., Minker, W.: Using multimodal resources for explanation approaches in intelligent systems. In: Proc. of the Eight Int. Conf. on Language Resources and Evaluation (LREC), pp. 411–415. European Language Resources Association (ELRA) (2012)
17. Nothdurft, F., Richter, F., Minker, W.: Probabilistic human-computer trust handling. In: Proc. of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), pp. 51–59. Association for Computational Linguistics (2014)

18. Reuter, S., Dietmayer, K.: Pedestrian tracking using random finite sets. In: Proc. of the 14th Int. Conf. on Information Fusion, pp. 1–8. IEEE (2011)
19. Reuter, S., Wilking, B., Wiest, J., Munz, M., Dietmayer, K.: Real-time multi-object tracking using random finite sets. IEEE Transactions on Aerospace and Electronic Systems **49**(4), 2666–2678 (2013)
20. Richardson, M., Domingos, P.: Markov Logic Networks. Machine Learning **62**(1-2), 107–136 (2006)
21. Schüssel, F., Honold, F., Schmidt, M., Bubalo, N., Huckauf, A., Weber, M.: Multimodal interaction history and its use in error detection and recovery. In: Proc. of the 16th Int. Conf. on Multimodal Interaction, ICMI '14, pp. 164–171. ACM (2014)
22. Schüssel, F., Honold, F., Weber, M.: Influencing factors on multimodal interaction during selection tasks. Journal on Multimodal User Interfaces **7**(4), 299–310 (2013)
23. Schüssel, F., Honold, F., Weber, M.: Using the transferable belief model for multimodal input fusion in companion systems. In: F. Schwenker, S. Scherer, L.P. Morency (eds.) Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction, vol. 7742, pp. 100–115. Springer Berlin Heidelberg (2013)
24. Seegebarth, B., Müller, F., Schattenberg, B., Biundo, S.: Making hybrid plans more clear to human users – a formal approach for generating sound explanations. In: Proc. of the 22nd Int. Conf. on Automated Planning and Scheduling (ICAPS), pp. 225–233. AAAI Press (2012)