Reference Architecture Approach for *Companion*-Systems

Thilo Hörnle[1] and Michael Tornow[2]

[1] Institute of Artificial Intelligence,
Ulm University, D-89081 Ulm, Germany,
thilo.hoernle@uni-ulm.de
[2] Institute of Information Technology and Communications,
Otto-von-Guericke University, D-39016 Magdeburg, Germany,
michael.tornow@ovgu.de

Abstract. This paper focuses on an approach of a reference architecture for a *Companion*-System (CS). It enhances a technical system by a user adaptive interaction and a situation adjustive behavior. To generate a reference architecture existing demonstration systems were analyzed and combined in order to find the essential system components. All aspects of the architecture of a CS will be considered in detail. It is finally enhanced by a system controller which enables a dynamic adaption to the current system state in order to allow future extensions of the architecture.

1 Motivation

Companion-Technology (CT) enables technical systems to be adaptive to the user's emotional state, the current situation, and the current task of the application, as shown in Fig. 1 [5]. Therefore a *Companion*-System (CS) contains additional modules from different research areas. Required and optional modules are defined by the architecture which should ensure compatibility of the system components.

Two demonstration systems have been implemented prototypically in order to test the integration of these modules [1, 3]. The first supports the user while he is setting up a home entertainment system [1]. This system focuses on managing the interaction and the dialog with the user and adjusts the intelligent behavior of the application to the current task.

Adapting to the emotional state of the user and the environmental situation is the research object of the second system [3], which is helping the user to buy tickets at a train station. The user and the environment are observed by measuring multiple modalities. Some of these modalities are explicit inputs to control the application, others are implicit inputs like facial expression or prosody that provide information about the emotional state of the user. As some of these modalities contain redundant information, the certainty of the extracted information or the system's robustness against the failure of a modality is increased in a fusion process.

Both demonstration systems are examples of assistance systems and serve to evaluate parts of the *Companion*-System's capabilities and accordingly certain aspects of the architecture. The adaptivity of the system state and configuration as well as the extendability are important functions an architecture should cover. This paper introduces an architecture to enable the system to extend its functionalities with new modules and support the user for a long term.

2 An Approach to a CS Reference Architecture

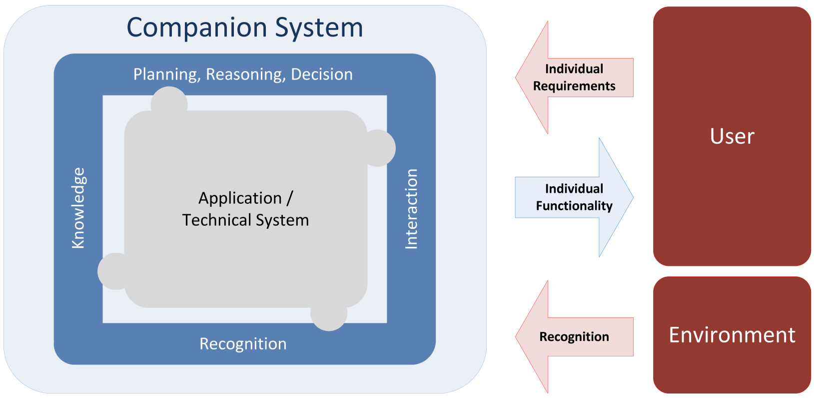

Fig. 1. The figure shows aspects of the CT (planning, reasoning, decision, knowledge representation, recognition, interaction), which are realized by components developed in specific research area. The frame around the technical system symbolizes the CT.

The requirements to the reference architecture can be derived from the *Companion*-System's capabilities as described in Fig. 1 [5]. This leads to the essential components of a CS.

The communication of the system is adaptive to the user. Therefore the systems interface covers specific needs of the current user and presents the information in a user adaptive and comfortable way. The system should interact with the user in a smart way, as it should recognize what the user is doing, e.g., whether the user is interacting with the system or with his environment. In the latter case the system should ignore possible commands from the user.

For a smart interaction with the user a dialog management and an interaction management are required. These modules rely on information, retrieved from a knowledge base, containing knowledge of the task and the user. Information of the current state are observed by sensors. The relevant information is extracted by the signal processing and combined with other sensor data in the data fusion processes. The task of the application is solved by a planner using the information from the knowledge base as well.

The interaction management, the dialog management, knowledge base, and planner are implemented in the first prototype system, but it contains only one sensor giving implicit inputs by detecting the user position (see Sec. 2.1). The sensor data processing and the multi level data fusion are the main focus in the second prototype system, whose aspects are described in Sec. 2.3. The sensors are the connection of the application to the user, the environment and current situation.

The latest developments in embedded devices such as smartphones show that such a device could be the central element of a CS. These devices do not have enough computation power for all the processings required for a CS. Therefore some of the tasks could be swapped to a server. This is described in the Sec. 2.2.

2.1 Architecture for Planning, Reasoning and Interaction for CS

The aspects of smart user interaction and smart application behavior are derived from an assistance system solving the task of setting up a home entertainment system [2, 1]. It was set up as a knowledge-based system, which is focusing on interaction and planning [4].

Analyzing the architecture of the first demonstration system shows that all core components have a connection to the knowledge base, where the data (the different knowledge models for the other components) is stored in the static context knowledge. Furthermore an adaption to the chronological context of the user and the environment is realized in the knowledge base.

The planning components create an action plan for the user. This plan includes all steps, which the assistant system advises the user to execute. Is the user not able to execute a step, the planning components can repair the plan to a functional state. The plan is transparent to the user in all steps. Necessary information of the planning domain and the problem to solve are transfered from the knowledge base to the certain components.

The dialog management (DM) gets the current plan step from the planner. Using the information about the user and the situation, a suitable system reaction and user dialog will be generated by the DM. The dialog steps will be sent to the IM. There the multi-modal fission process distributes those for presentation to the user. The plan explanation works together with the explanation manager of the DM.

The main functionality of the interaction management (IM) is divided into two main parts: the fusion of the explicit user input from the interface and the fission of the different output devices of the user interface. Both parts of the interaction work together via several interaction management components to realize an optimal interaction with the user on multiple devices.

2.2 Smart Sensor Network

The information required by a CS to recognize the user and its emotional state and environment need to be extracted from the raw data obtained by physical

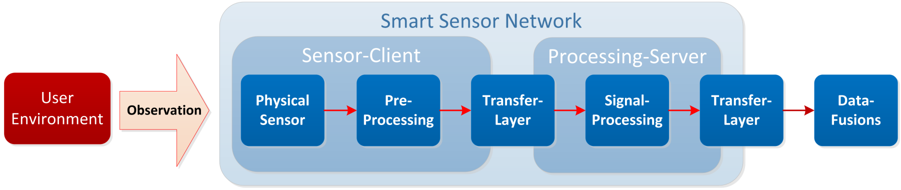

Fig. 2. The figure shows the smallest version of a smart sensor network. The Transfer-Layer can have n input and m outputs. So it is possible to build a complex sensor network as described in Fig. 3.

sensors. Therefore, usually complex signal processing is necessary. The signal processing is commonly a software which can be executed on every processor available. By introducing a client server model for sensor data processing, computationally demanding processing parts could be transferred to a computation center. This would help *Companion*-Systems running on smartphones or tablets.

An example is given in Fig. 2, where the client contains only the physical sensor and a preprocessing unit. A transfer layer is distributing the sensor data to several signal processings if required. Several signal processings can use the data of a certain sensor e.g. prosody extraction and speech recognition. The transfer layer realizes connection with n inputs and m outputs and builds a complex sensor network. The smart sensor network enables the dynamization of the sensor setup.

2.3 Architecture for Sensor Data Fusion

The sensor data fusion is an important issue for *Companion*-Systems, a multimodal sensor array is the window to the real world.

Depending on the measured object or scene usually several sensors are needed to obtain the context. Thus only combining the sensor data in one or more fusion levels enables the system to gain the relevant information with good certainty [3].

The architectural aspects of sensor data processing were covered in a demonstration system which emulates a ticket vendor machine with *Companion*-Capabilities using a sensor phalanx consisting of five modalities e.g audio recording for prosody and speech recognition and RGBD sensors for gestures.

In difference to the system described in [3] in the future systems the data is recorded and distributed by a smart sensor network (see Fig. 3). The results of those smart sensor networks are merged according to certain topics in the data fusion layers. With each layer the data is going to be more abstract.

The perception level fusion is the first layer combining mainly the data of several sensors related to certain topics. Here mainly implicit inputs are processed. The second layer, the knowledge based fusion, includes the system experience given as results of the temporal information stored in the knowledge base. It is used to compare the perception fusion results with the essence of preceding data to gain the temporal change of the user behavior.

This step was not covered in demonstration system 'ticket vendor machine' as a knowledge base is necessary for creating the required information. In Fig. 3 the knowledge based fusion step is realized for every perception level fusion. This is the main connection between both demonstrations systems.

Finally these results are combined in the application level fusion with knowledge to gain application-specific data for the CS, including the explicit inputs.

3 An Approach for a CS Reference Architecture

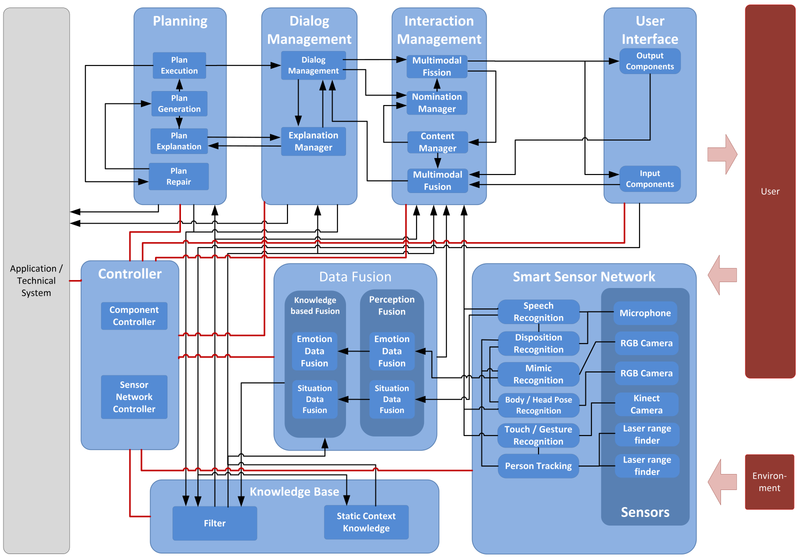

Fig. 3. The figure shows an abstract view of a reference architecture

The concept of a smart sensor network (cf. Sec. 2.2) in combination with the architectural aspects of two prototypical systems, described in 2.1 and 2.3, leads to an approach of a reference system architecture. This architecture is able to cover a fully functional CS (see Fig. 3).

Both demonstration systems overlap in the data fusion part. Especially the knowledge based fusion and the application level fusion are affected by the combination. The architecture combination shows for each fusion topic of perception level fusion a knowledge based fusion is available. These results are merged in the application level fusion which is located in the interaction management.

A central controller is needed to introduce aspects of dynamization to the system architecture. The controller enables a dynamic binding of modules from the CT to a CS. This is realized in real time over multiple devices (like smart phones and a static ticket vendor machine). The concept of the smart sensor networks realizes the use of the sensors from mobile and personal devices, controlled by the sensor network controller. The component controller can temporarily bind e.g. a knowledge base from the personal device to a public system, keeping privacy measures in mind. These functions are the two main features of the controller-components from a CS.

The aspects of dynamization add the system capability to adapt to different sensor setups as well as the extension of the CS structure. Finally it enables the system to react to failure of system components. This increases the availability of the system to the user.

4 Future Work

In this paper a full approach of a reference architecture for a CS was presented, reaching from the sensor data acquisition and processing, over the data fusion in several layers to knowledge base, the planner and the interaction management. The dynamization of the architecture enables *Companion*-Technology to cover most of the possible *Companion*-Systems. It needs to be evaluated and extended to a generalized reference architecture for *Companion*-Systems.

5 Acknowledgements

This paper is based on the work done with the transregional collaborative research center SFB/TRR 62 *Companion*-Technology for Cognitive Technical Systems funded by the German Research Foundation (DFG).

References

1. Bercher, P., Biundo, S., Geier, T., Hoernle, 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 Artificial Intelligence (AAAI). pp. 4264–4265. AAAI Press (2015)
3. Glodek, M., Honold, F., Geier, T., Krell, G., Nothdurft, F., Reuter, S., Schüssel, F., Hörnle, T., Dietmayer, K., Minker, W., Biundo, S., Weber, M., Palm, G., Schwenker, F.: Fusion paradigms in cognitive technical systems for human–computer interaction. Neurocomputing 161(0), 17–37 (2015)
4. Honold, F., Bercher, P., Richter, F., Nothdurft, F., Geier, T., Barth, R., Hoernle, 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 (IE). pp. 378–381. IEEE (2014)
5. Wendemuth, A., Biundo, S.: A companion technology for cognitive technical systems. In: Cognitive Behavioural Systems, Lecture Notes in Computer Science, vol. 7403, pp. 89–103. Springer Berlin Heidelberg (2012)