



HAL
open science

Multimodal and Mobile Conversational Health and Fitness Companions

Markku Turunen, Jaakko Hakulinen, Olov Ståhl, Björn Gambäck, Preben Hansen, Mari C. Rodríguez Gancedo, Raúl Santos de La Cámara, Cameron Smith, Daniel Charlton, Marc Cavazza

► **To cite this version:**

Markku Turunen, Jaakko Hakulinen, Olov Ståhl, Björn Gambäck, Preben Hansen, et al.. Multimodal and Mobile Conversational Health and Fitness Companions. *Computer Speech and Language*, 2010, 25 (2), pp.192. 10.1016/j.csl.2010.04.004 . hal-00692186

HAL Id: hal-00692186

<https://hal.science/hal-00692186>

Submitted on 29 Apr 2012

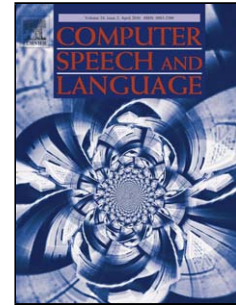
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Title: Multimodal and Mobile Conversational Health and Fitness Companions

Authors: Markku Turunen, Jaakko Hakulinen, Olov Ståhl, Björn Gambäck, Preben Hansen, Mari C. Rodríguez Gancedo, Raúl Santos de la Cámara, Cameron Smith, Daniel Charlton, Marc Cavazza



PII: S0885-2308(10)00035-5
DOI: doi:10.1016/j.csl.2010.04.004
Reference: YCSLA 451

To appear in:

Received date: 30-4-2009
Revised date: 25-3-2010
Accepted date: 1-4-2010

Please cite this article as: Turunen, M., Hakulinen, J., Ståhl, O., Gambäck, B., Hansen, P., Gancedo, M.C.R., Cámara, R.S., Smith, C., Charlton, D., Cavazza, M., Multimodal and Mobile Conversational Health and Fitness Companions, *Computer Speech & Language* (2008), doi:10.1016/j.csl.2010.04.004

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Multimodal and Mobile Conversational Health and Fitness Companions

Markku Turunen^{*,a}, Jaakko Hakulinen^a, Olov Ståhl^b, Björn Gambäck^{b,c},
Preben Hansen^b, Mari C. Rodríguez Gancedo^d, Raúl Santos de la Cámara^d,
Cameron Smith^e, Daniel Charlton^e, Marc Cavazza^e

^a*Department of Computer Sciences, 33014 University of Tampere, Finland*

^b*SICS, Swedish Institute for Computer Science AB, Box 1263, 164 29 Kista, Sweden*

^c*Department of Computer and Information Science, Norwegian University of Science
and Technology, Sem Sælands vei 7-9, 7491 Trondheim, Norway*

^d*Telefonica I+D, C/ Emilio Vargas 6, 28043 Madrid, Spain*

^e*School of Computing, University of Teesside, Middlesbrough TS1 3BA, United Kingdom*

Abstract

Multimodal conversational spoken dialogues using physical and virtual agents provide a potential interface to motivate and support users in the domain of health and fitness. In this paper we present how such multimodal conversational Companions can be implemented to support their owners in various pervasive and mobile settings. We present concrete system architectures, virtual, physical and mobile multimodal interfaces, and interaction management techniques for such companions. In particular, we present how knowledge representation and separation of low-level interaction modelling from high-level reasoning at the domain level makes it possible to implement distributed, but still coherent, interaction with Companions. The distribution is enabled by using a dialogue plan to communicate information from domain level planner to dialogue management and from there to a separate mobile interface. The model enables each part of the system to handle the same information from its own perspective without containing overlapping

*Corresponding author

Email addresses: mturunen@cs.uta.fi (Markku Turunen), jh@cs.uta.fi (Jaakko Hakulinen), olovs@sics.se (Olov Ståhl), gamback@sics.se (Björn Gambäck), preben@sics.se (Preben Hansen), mcrgr@tid.es (Mari C. Rodríguez Gancedo), e.rsai@tid.es (Raúl Santos de la Cámara), c.g.smith@tees.ac.uk (Cameron Smith), d.charlton@tees.ac.uk (Daniel Charlton), m.o.cavazza@tees.ac.uk (Marc Cavazza)

Preprint submitted to Computer Speech and Language

March 25, 2010

1
2
3
4
5
6
7
8
9 logic, and makes it possible to separate task-specific and conversational di-
10
11 dialogue management from each other. In addition to technical descriptions,
12 we present results from the first evaluations of the Companions interfaces.

13
14 *Key words:* Companions, Embodied Conversational Agents,
15 Conversational Spoken Dialogue systems, Mobile interfaces, Cognitive
16 Modelling, Dialogue management
17

18 19 **1. Introduction**

20
21 Most existing spoken dialogue systems provide a single interface to solve a
22 well-defined task, such as booking tickets or providing timetable information.
23 However, there are emerging areas that differ dramatically from task-oriented
24 systems. In domain-oriented dialogues (Dybkjaer et al., 2004) the interaction
25 with the system, typically modelled as a conversation with a virtual human-
26 like character, can be the main motivation for the interaction. These systems
27 are often multimodal, and may be implemented in pervasive computing envi-
28 ronments where various mobile, robotic, and other novel interfaces are used
29 to communicate with the system.
30
31

32
33 In the EC-funded COMPANIONS project (Wilks, 2007) we have devel-
34 oped conversational Companions that build long-lasting relationships with
35 their users via mobile and physical agent interfaces. Such systems have differ-
36 ent motivations for use compared to traditional task-based spoken dialogue
37 systems. Instead of helping with a single, well-defined task, a companion
38 provides long lasting support and companionship on a daily basis.
39

40
41 One of the developed companions, and the topic of this paper, is the
42 conversational Health and Fitness Companion, which helps users to maintain
43 a healthy lifestyle. There are several examples of commercial systems in the
44 domain of health and fitness, in particular in exercise domain. Some of the
45 most well known examples are miCoach from Adidas and NIKE+. ¹ More in
46 line with the present work, MOPET (Buttussi and Chittaro, 2008) is a PDA-
47 based personal trainer system supporting outdoor fitness activities. MOPET
48 is similar to a Companion in that it tries to build a relationship with the user,
49 but there is no real dialogue between the user and the system and it does not
50 support speech input or output. Neither does MPTrain/TripleBeat (Oliver
51 and Flores-Mangas, 2006; de Oliveira and Oliver, 2008), a system that runs
52
53
54
55

56
57 ¹www.micoach.com and www.nike.com/nikeplus
58

1
2
3
4
5
6
7
8
9 on a mobile phone and aims to help users to more easily achieve their exercise
10 goals. The system selects music indicating the desired pace and different ways
11 to enhance user motivation, but without an agent user interface model. InCA
12 (Kadous and Sammut, 2004) is a spoken language-based distributed personal
13 assistant conversational character with a 3D avatar and facial animation.
14 Similar to the Mobile part of Companion, the architecture is made up of a
15 GUI client running on a PDA and a speech server, but the InCA server runs
16 as a back-end system, while the Companion utilizes a stand-alone speech
17 server.
18
19

20 The difference between the Health and Fitness Companion and most of
21 the existing health related systems is that our Companion prototype aims to
22 be a peer rather than an expert system in health-related issues. The ulti-
23 mate goal of the Health and Fitness Companion is to support overall lifestyle
24 changes in the user's daily habits rather than giving detailed advice on any
25 specific health related issues. A lifestyle change often requires great motiva-
26 tion and the Health and Fitness Companion aims to support building and
27 maintaining this motivation. The method the Health and Fitness Compan-
28 ion uses to support the user motivation is to set up a long lasting social and
29 emotional relationship with the user. Since people build relationships mostly
30 in face-to-face conversations, a conversational embodied agent is a potential
31 platform to build such a relationship (Dybkaer et al., 2004).
32
33

34 Both mobile usage and physical and virtual agent interfaces can sup-
35 port the goal of making a spoken dialogue system part of its user's everyday
36 life, and building a meaningful relationship between the system and the user.
37 There are numerous examples of virtual embodied conversational agents, but
38 only a few examples of physical conversational agents. Physical agents, how-
39 ever, can be particularly efficient in applications that try to evolve dialogue
40 systems into being part of people's lives.
41
42

43 While naturalistic human-like physical robots are under development,
44 there is room for a variety of different physical interface agents ranging from
45 completely abstract (e.g., simple devices with lights and sound) to highly so-
46 phisticated anthropomorphic apparatus. For example, Marti and Schmandt
47 (2005) used several toy animals, such as bunnies and squirrels, as physical
48 embodied agents for a conversational system. Another example is an in-
49 door guidance and receptionist application involving a physical human-like
50 interface agent that combines pointing gestures with conversational speech
51 technology (Kainulainen et al., 2005). Some physical agent technology has
52 also been commercialized. Examples include the Paro Therapeutic Robot
53
54
55
56
57
58

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

², and the wireless Nabaztag/tag rabbit ³, which has resulted active user community.

In this work, we have used the Nabaztag/tag device as a multimodal physical interface to create a conversational Health and Fitness Companion. In addition to the physical interface we have used a mobile device interface for outdoor usage, and a graphical, lifelike virtual agent interface in certain dialogues to provide rich multimodal interface when those features are needed.

In these kinds of applications where multiple user interfaces can be used to access parts of the same functionality, e.g., the same physical exercise information, and the system interacts with a user many times over a long time period, modelling the interaction and domain quickly becomes complex. To address this, the system must include a model capable of reasoning about the domain, and learn from the user and his/her actions to provide meaningful interaction, such as to provide reasonable guidance on the user's health and progress as the user's condition alters over time. Such reasoning should be concentrated on one component, instead of duplicating the logic to keep the system maintainable. Still, the information must be communicated over different interfaces and to the components inside them. Therefore, modularization of the system and appropriate knowledge representation become vital.

On the dialogue management level, a common way to take some complexity away from the dialogue manager and limit its tasks more specifically to dialogue management is to separate domain specific processing, such as database queries, into a back-end component. Many researchers have worked with separating generic dialogue management processes from the domain specific processes. Example solutions include shells (Jönsson, 1991) and object oriented programming methods (Salonen et al., 2004; O'Neill et al., 2003). On the other hand, a simple back-end interface, e.g., SQL queries, can be included as configuration parameters (Pellom et al., 2000). Since dialogue management is usually based on some clearly defined model, such as state transition networks or form filling, keeping domain specific processing in the back-end makes it possible to implement dialogue management purely with the selected model without the unnecessary complexity of domain specific

²<http://www.parorobots.com/>

³www.nabaztag.com

1
2
3
4
5
6
7
8
9 information.

10 The Health and Fitness Companion is based on a model where the domain
11 specific module is more than just a simple backend database and includes
12 active processing of domain information, reasoning, learning, and other com-
13 plex processes. We call such a component cognitive model. While the task
14 of a dialogue manager is to maintain and update the dialogue state, the cog-
15 nitive model reasons using domain level knowledge. The separation of the
16 tasks between the different parts is not trivial. For example, managing dia-
17 logue level phenomena, such as error handling and basic input processing, are
18 tasks clearly in the area of a dialogue manager. However, cognitive modelling
19 can help in error handling by spotting input that seems suspicious based on
20 domain level information, and in input parsing by providing information on
21 potential discussion topics. The solution we have devised is to have the cog-
22 nitive model to produce a dialogue plan for the dialogue management in the
23 Home Companion system. The dialogue manager in the Home Companion
24 provides parsed user inputs to the cognitive model and to the Mobile Com-
25 panion. The Mobile Companion provides similar input back to the Home
26 Companion, which communicates it back to the Cognitive Model.
27
28
29
30
31

32 In the rest of this paper, Section 2 introduces the concept of Health and
33 Fitness Companion including the different types of mobile, physical and vir-
34 tual agent interfaces that realize the concept. For the Cooking Companion we
35 present the initial prototype interface developed to demonstrate the concept.
36 For the Home and Mobile Companions, which have been fully implemented,
37 we present the technical solutions to implement such Companions. Section 3
38 details the software architectures used to implement physical and mobile
39 Companions interfaces. Section 4 introduces a novel interaction management
40 solution which separates cognitive modeling from dialogue management and
41 enables their flexible interoperability and distributed but coherent dialogues
42 between different agent interfaces. Together, the solutions presented in these
43 sections result in concrete software components to implement further Com-
44 panion applications. Some of the solutions, such as the PART system used
45 in the Mobile Companion and the Jaspis architecture and the jNabServer
46 software used for the Home Companion have been released as open source
47 software. Finally, Section 5 presents the key results from the early eval-
48 uations of the first Health and Fitness Companion prototypes. Section 6
49 recapitulates the key parts of the discussion.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

2. Companion Interfaces

As discussed previously, the Health and Fitness Companion interacts with its user in three main forms: as a physical Home Companion to provide general support for a healthy lifestyle, as a Mobile Companion to support the user during physical outdoor activities, and as a virtual Cooking Companion to support eating and cooking related lifestyle activities. Together, these interfaces aim to build a long lasting relationship with the user.

2.1. Example Dialogue

The example below show a complete dialogue which a user can have with the three different Health and Fitness Companion interface prototypes. Dialogues 1 and 2 are carried out with the Home Companion interface, Dialogues 3 and 4 are done with the Mobile Companion interface, and Dialogue 5 demonstrates a conversation with the Cooking Companion interface.

(1) During breakfast

C1: Do you have anything arranged for today?

U1: I'm playing Squash

C2: Is that before dinner?

U2: Yes

C3: Okay. Are you going to take the train to work?

U3: Yes

C4: Right. You could take in a packed lunch to work?

U4: Yes, that would be fine

(2) Later on in the evening

C5: Welcome back! How did your game of Squash go?

U5: The game was cancelled, so I went to the gym

C6: Okay. Did you take the train to work?

U6: Yes, I did

C7: Right. Did you take in a packed lunch?

U7: No. I was in a rush so I went to the cafeteria

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

(3) **A physical exercise with the Mobile Companion**

C8: Hello, what kind of exercise would you like to do?

U8: I would like to do some running

C9: Ok, running it is. Select start when you are ready.

(4) **After 12 minutes**

C10: You have been running for 12 minutes. The distance is 2.5 kilometres; the current pace is 5 minutes per kilometre.

U9: Okay, please play some relaxing music

(5) **Cooking with the Cooking Companion**

C11: Good, let's start preparing the ingredients. You should remove the stalk and seeds from the red pepper, then chop it up.

U10: And I should chop up the onion now too?

C12: Yes, that's right, peel and chop the onion. You could peel and crush the garlic too afterwards.

2.2. Home Companion Interface

The Home Companion interface is a physical agent that resides in the home of its user, for example, in the kitchen table, where it is natural to have a conversation before going to work or start a physical activity. This setup is illustrated in Figure 1. The physical agent is implemented with the Nabaztag/tag WLAN rabbit that provides audio output and push-to-talk audio input, RFID-based interaction, moves its ears, and operates four coloured lights to signal its status.

In order to provide meaningful advice to the user on his or her daily activities the Home Companion communicates with the user in two main dialogue phases; the planning phase where the system talks about the coming day with the user (as demonstrated in Dialogue 1), and the reporting phase where the users actual activities are assessed with reference to what was agreed on earlier (Dialogue 2). The covered topics include a range of common situations and activities relevant for health and fitness conversations. In overall, the Home Companion is able to discuss the following topics: travelling to work, getting lunch, activities to be performed before dinner, getting dinner, and activities to be performed after dinner. It knows activities such as playing



Figure 1: The Home Companion interface in a typical usage situation (see <http://www.youtube.com/watch?v=KQSiigSEYhU> for a video presentation)

football, squash, or badminton; going to the gym or shopping; and watching television or reading a book.

2.3. Mobile Companion Interface

The Mobile Companion runs on Windows Mobile devices and can be used during outdoor exercise activities such as walking, jogging or cycling. The Mobile Companion can download the plan of the day which the user has agreed on with the Home Companion. The Mobile Companion will then suggest an exercise, based on the user's current location, time of day and the plans made earlier, or, if there are no suitable exercises, ask the user to define one, as in Dialogue 3. Once an exercise has been agreed upon, the Companion will track the progress (distances travelled, time, pace, and

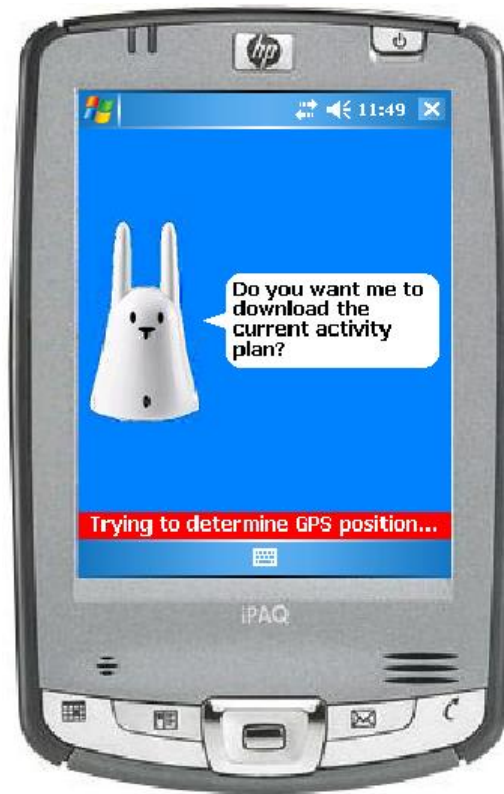


Figure 2: Mobile Companion interface

calories burned) using a GPS receiver. While exercising, the user can ask the Companion to play music or to give reports on how the user is doing. When the exercise is over, the Mobile Companion can upload the result to the home system. The data gathered during an exercise is also stored locally in the device's record store, which allows for access to data from previous exercises (for comparison purposes) even if the mobile internet connection is temporarily unavailable.

The Mobile Companion's graphical user interface, as seen in Figure 2, consists of a single screen showing a static image of the Nabaztag rabbit, along with a speech bubble. The rabbit image is intended to give users a sense of communicating with the same Companion, no matter the interface (Home or Mobile). To further the feeling of persistence and help users associate the two interfaces, the Home and the Mobile Companions use the same



Figure 3: Cooking Companion interface

synthesized voice, while the Cooking Companion uses a different avatar and thus a different text-to-speech engine that produces a voice clearly different to the bunny's, helping the user differentiate between both ECAs.

All spoken messages are also shown as text in the speech bubble. The user can provide input via voice, by pressing hardware buttons on the mobile device, and in some situations, by tapping on a list of selections on the touch screen. Screen-based input is used, for example, when identifying the current exercise route (from a list of route names, defined by the user in previous sessions), but also when ASR errors occur, to perform error correction. In the latter case, the user is presented with a list of input strings that the Companion is able to understand in the current context, and must select one to continue the dialogue.

2.4. Cooking Companion Interface

One of the main aims of the Health and Fitness Companion is to provide tips for a healthier lifestyle, and one of the key aspects is a correct nutrition. The Cooking Companion is a virtual embodiment of a dietary advisor, which:

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
 - 61
 - 62
 - 63
 - 64
 - 65
1. Presents a set of possible recipes to the user, based on availability of ingredients or home delivery food.
 2. Informs the user of the appropriateness of each choice, based on available information such as the user's likes and dislikes, authorised medical and nutritional databases, hers/his current physical condition, special dietary requirements (e.g., allergies) and the physical exercise scheduled by the rest of the Health and Fitness Companion.
 3. Helps in the preparation of the selected recipe using multimedia and/or dialogue.

As an example, the user wants to eat seafood paella. The system then inspects all of the aforementioned parameters and conclude advising the user to prepare vegetable paella, based on his current exercise schedule and high blood fat levels (see Dialogue 5 for an example). Then the system instructs the user on how to prepare the dish using videos and speech.

In Figure 3, the interface for the Cooking Companion prototype is shown. The system augments plain spoken dialogue with the help of an ECA, powered by a third party engine ⁴ and a finger-operated touchscreen interface. The interaction is multimodal, intertwining speech utterances and finger pointing. The ECA, modelled as a photorealistic human, includes some advanced gesturing capabilities that have been demonstrated to make the avatar both more engaging and understandable (López Mencía et al., 2006; López et al., 2007, 2008).

The Cooking Companion interface is presented here as an example of how the Companions paradigm can use different agent interfaces to achieve certain aspects of companionship. Currently, a prototype to demonstrate its functionality has been implemented. In the rest of the paper, we focus on the Home and the Mobile Companions, which have been fully implemented as described in the forthcoming sections. Further information on the Cooking Companion and its relation to the overall Health and Fitness Companion is presented in (Hernández et al., 2008) and (Turunen et al., 2008b).

3. Companions Architectures

In order to construct Health and Fitness Companions, we need concrete system architectures and software components that are flexible enough to

⁴Haptik Player, Haptik Inc.: www.haptik.com

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

implement different interfaces running on different hardware and software platforms. At the same time, we need to maintain shared knowledge for the Health and Fitness Companion. In the following, the Home Companion and the Mobile Companion architectures and their main components are presented. These provide foundations for the interaction management solutions presented in the following sections.

3.1. Home Companion

The Home Companion is implemented on top of Jaspis⁵ (Turunen et al., 2005), a generic open source software architecture designed for adaptive spoken dialogue systems. In order to implement the physical Health and Fitness Companion, the Jaspis architecture was extended to support interaction with physical agents, and the Nabaztag/tag device in particular. For this purpose, the jNabServer software⁶ was created to handle local communication with Nabaztag/tag. This solution offers full control over Nabaztags, including RFID-reading, and makes it possible to use custom programs and technologies to process inputs and outputs, such as the speech recognition and TTS software used in the Health and Fitness Companion.

Figure 4 illustrates the Home Companion architecture. The top-level structure of the system is based on managers, seven of them in total, which are connected to the central Interaction Manager (a HUB/Facilitator style component) using a star topology structure. In addition, the application has an Information Manager (a database/blackboard style component) that is used by all the other components to store and share information. The Information Manager provides a high-level interface to the Information Storage, which contains all the dialogue management and the cognitive model structures (e.g., confirmation pools and activity models). The communication between these two models is based on a dialogue plan, which is also in the Information Storage. These are presented in detail in Section 4. The Information Manager and the Information Storage do not modify this content, they just handle the technical management of data in XML format. The details of this layered information management approach are beyond the scope of this paper, and are discussed in detail in (Turunen, 2004).

All components in the system are stateless, meaning they save all their internal data to the central database when they end processing (or go to

⁵www.cs.uta.fi/hci/spi/Jaspis

⁶Java-based Nabaztag/tag Server: www.cs.uta.fi/hci/spi/jnabserver

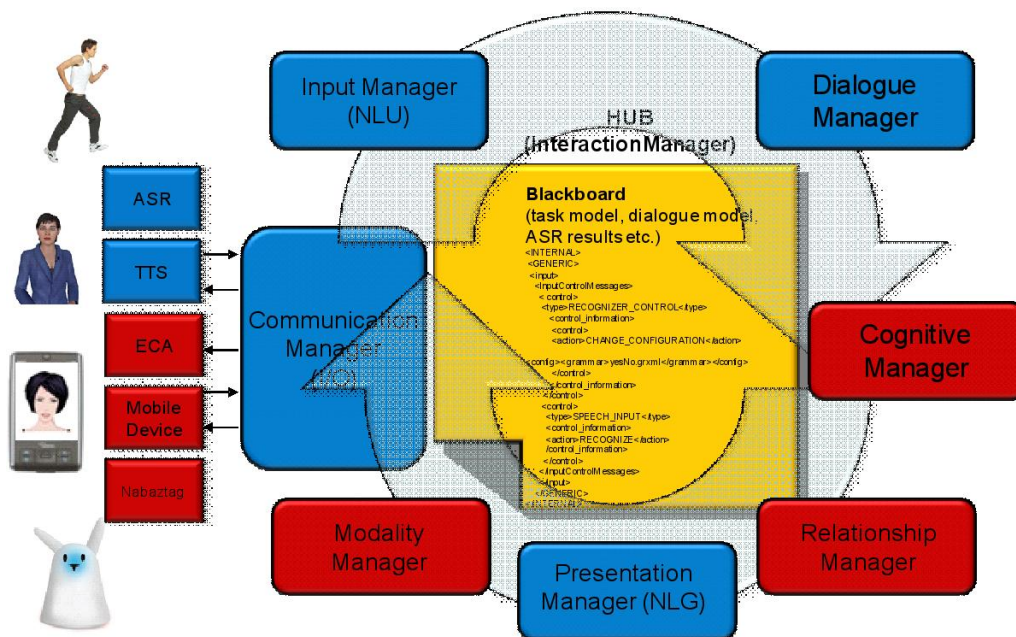


Figure 4: The Home Companion architecture

sleep), and update their internal models when they start processing (wake up) again. This way, the overall system is modular and can be extended easily. We believe this is a requirement needed to construct complex conversational system such as the Health and Fitness Companions, since it separates data from its processing, allowing efficient iterative development, for example, to visualize, debug, and teach complex systems easily (Hakulinen et al., 2007). Most importantly, it makes possible the architecture level modularization and adaption scheme described next.

Modularity is further supported in this architecture by the agents-managers-evaluators paradigm that is used across all system modules. In this approach, all tasks are handled by compact and specialized agents located in modules and coordinated by managers. When one of the agents inside a module, such as the Dialogue Manager, is going to be selected, each evaluator in the module gives a score for every agent in the module. These scores are then multiplied by the local manager, which gives the final score, a suitability factor, for every agent. As an example, the multi-agent architecture of Jaspis (Turunen et al., 2005) is used heavily in dialogue management; in the current

1
2
3
4
5
6
7
8
9 prototype, there are 30 different dialogue agents, some corresponding to the
10 topics found in the dialogue plan, others related to error handling and other
11 generic interaction tasks. These agents are dynamically selected based on the
12 current user inputs and overall dialogue context with using three rule-based
13 reasoning evaluators.
14

15 For speech input and output, Loquendo ASR and TTS components (Lo-
16 quendo, 2008) have been integrated into the Communication Manager. ASR
17 uses grammars in “Speech Recognition Grammar Specification” (W3C) for-
18 mat. In these grammars, semantic tags in “Semantic Interpretation for
19 Speech Recognition (SISR) Version 1.0” (W3C) format are used to provide
20 information for the NLU component (described in detail in the following sec-
21 tions). Domain specific grammars were derived from early informal testing
22 and Wizard-of-Oz sessions conducted during the iterative development of the
23 system. The data collected from these initial studies has been used to build
24 the HFC grammars. There is a set of about 50 grammars and the system
25 dynamically selects these according to the current dialogue state. The gram-
26 mar size for most grammars is about 1400 words, with a total of about 900
27 grammar rules. The vocabulary coverage is balanced across the four relevant
28 domain parts of the activity model: transportation, physical activity, leisure
29 and food. Natural language generation is implemented using a combination
30 of canned utterances and Tree Adjoining Grammar-based generation. The
31 grammar-based generation is used mostly in confirmations, while canned ut-
32 terances with multiple options for each situation are used for most questions
33 about users’ day.
34
35
36
37
38
39
40

41 *3.2. Mobile Companion*

42 The Mobile Companion is realised by two components running on a Win-
43 dows Mobile device, a Java midlet that handles the graphical user interface
44 and the dialogue with the user, and a speech server that performs ASR and
45 TTS operations on request by the midlet (see Figure 5). The Java midlet is
46 built using the PART library,⁷ and uses the Hecl scripting language⁸ for GUI
47 and dialogue management. The speech server uses Loquendo ASR (speaker-
48 independent) and TTS libraries for embedded devices (Loquendo, 2008), and
49 SRGS 1.0 grammars. Pre-compiled grammars are loaded dynamically. The
50
51
52
53

54 ⁷Pervasive Applications RunTime: part.sourceforge.net

55 ⁸www.hecl.org

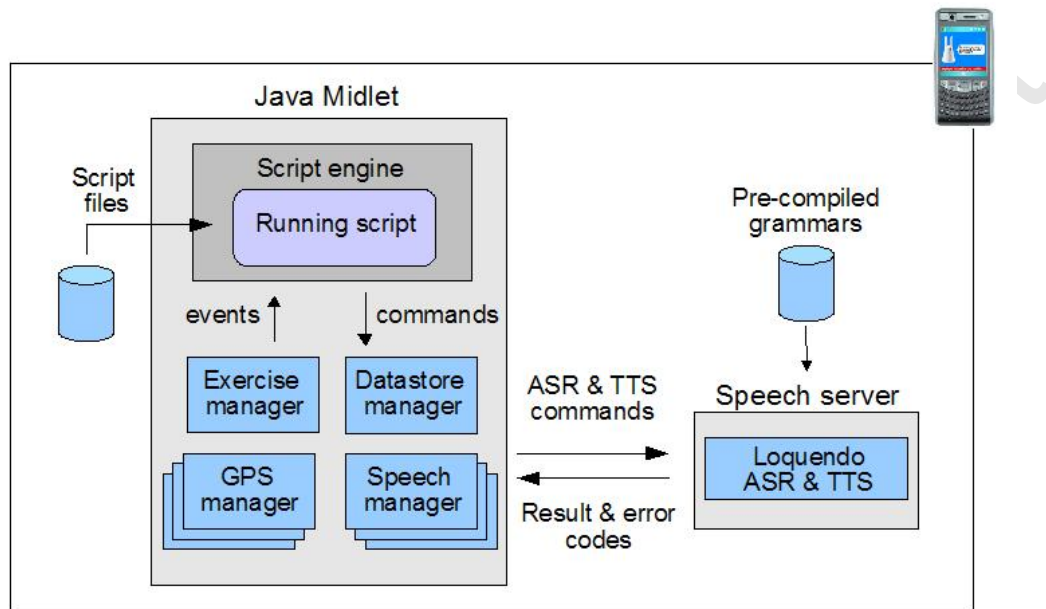


Figure 5: Mobile Companion Architecture

vocabulary and grammar is more restricted on the mobile device than in the Home Companion: its seven different grammars have in total vocabulary size of about 100 words.

While the Mobile Companion itself is running as a stand-alone system, the communication with the Home Companion requires the Mobile Companion to have access to the Internet, for instance via WLAN or 3G/GPRS. The Mobile Companion maintains a persistent data store in device to store user model (name, age, gender, weight, etc.), and the exercise results. Saving the exercise results allows the Companion to compare the progress of an exercise with previous exercises of the same kind. For instance, if the Companion knows that the user is currently cycling from home to work, it can provide feedback on how the user is doing compared to previous sessions. This allows for status messages like “*You are currently 1 minute and 23 seconds behind yesterday’s time.*” The Mobile Companion is described in detail by Ståhl et al. (2008, 2009).

4. Dialogue Management and Cognitive Modelling

As mentioned in Section 1 interaction management easily becomes too complex without clear separation of conversational, task, and domain specific information and components. To address these challenges, we introduce a novel solution for dialogue management and cognitive modelling in this domain.

Traditional architectures for dialogue includes three parts: an input module, which receives user input and parses it into a logical form, dialogue management, which maintains and updates dialogue state based on user input and generates output requests, and an output module, which generates natural language output to user based on the requests. In the case of the Health and Fitness Companion, a special component, Cognitive Model (CM) is introduced as separate component from the Dialogue Manager (DM), as seen in Figure 8. The Cognitive Model contains what can be considered the higher-level cognitive processes of the system. The Dialogue Manager takes care of conversational strategies. It presents questions to a user based on the dialogue plan, maintains a dialogue history tree and a dialogue stack, and communicates facts and user preferences to the Cognitive Model. The Dialogue Manager also handles error management, supports user initiative topic shifts, and takes care of top level interaction management, such as starting and finishing of dialogues. The Dialogue Manager and the Cognitive Model communicate using a dialogue plan.

In the rest of this section we present how the DM of the Home Companion, the CM component, and the Mobile Companion interact to create coherent long-lasting dialogues with the Companion.

4.1. Dialogue Plan

The communication between the Home and the Mobile Companion dialogue managers and the CM is based on a dialogue plan. Various kinds of dialogue plans (Larsson and Traum, 2000; Jullien and Marty, 1989) have been used inside dialogue managers in the past. A plan usually models what the system sees as the optimal route to task completion.

In Health and Fitness Companion, the CM provides a plan on how the current task (planning a day, reporting on a day) could proceed. The plan consists of a list of domain specific predicates. Figure 6 contains two items from the start of a plan for planning the day with the user in the morning. The first plan item (QUERY-PLANNED-ACTIVITY) can be realized as the

question “*Anything interesting planned for today?*” by the NLG component.

During the interaction, new relevant information may become available from the user. The information includes statements on the user’s condition (*tired*), user commitments to the system (*will walk to work*), user preferences (*does not like cafeterias*) and user reports on past activity (*took a taxi to work*), which can be accomplishments or failures of earlier commitments. The DM provides this information to the CM, piece by piece as it becomes available. As the CM receives the information, it updates the dialogue plan as necessary and builds up the day plan or report. Query type items, whose information has been gathered, disappear from the dialogue plan, while new items may appear when the dialogue progress.

As discussed above, the messages sent by the DM to CM can add new information (predicates) to the CM state. The DM can also remove information from the CM if previously entered information is found to be untrue. At the same time, the information is uploaded to a web server, where the Mobile Companion interface can access it any time.

```

30     <plan>
31       <plan-name>Generate-Task-Model-Questions</plan-name>
32       <plan-item>
33         <action>QUERY-PLANNED-ACTIVITY</action>
34       </plan-item>
35       <plan-item>
36         <action>SUGGEST-TRAVEL-METHOD</action>
37         <param>CYCLING-TRAVEL</param>
38         <param>HOME</param>
39         <param>WORK</param>
40       </plan-item>
41     </plan>

```

Figure 6: Start of a plan

The DM in the Home Companion can follow the dialogue plan produced by the CM step by step. Each step usually maps to a single question, but can naturally result in a longer dialogue if the user’s answer is ambiguous, or error management is necessary, or if the DM decides to split a single item into multiple questions. For example, the two dialogue turn pairs seen in Figure 7 are the result of a single plan item (QUERY-PLANNED-ACTIVITY). Since the first user utterance does not result in a complete, unambiguous

1
2
3
4
5
6
7
8
9 predicate, the DM asks a clarification question. A single user utterance can
10 also result in multiple predicates (e.g., *will not take the bus, has preference*
11 *for walking*).
12

13 When the Mobile Companion interface is activated, it downloads the
14 current day plan from the web server and uses it as a basis for the dialogue
15 it has with the user. The exercise which will then take place can be linked
16 to an item in the day plan, or it can be something new. As the exercise is
17 completed (or aborted), information about this is uploaded to the web server.
18 The DM of the Home Companion can download this information from the
19 server. This information is relevant to the DM when the user is reporting on
20 a day. The DM reports the downloaded information back to the CM when
21 the dialogue plan includes related items. The DM may also provide some
22 feedback to the user based on the information, for example, how the exercise
23 relates to the current plan and the overall situation. It is noteworthy that the
24 CM does not need to differentiate in any way, whether the information about
25 the exercise is originating from the Mobile Companion, or was gathered in a
26 dialogue with the Home Companion.
27

28 Similarly, clarifications and confirmations are not directly visible to the
29 CM. The DM can confirm items immediately (for example, when low confi-
30 dence scores are reported by the ASR component) or it can delay confirma-
31 tions to generate a single confirmation for multiple items at an appropriate
32 moment.
33

34 In this model, the DM has freedom to choose how to operate on the base
35 of the CM information. Most importantly, when presenting questions and
36 suggestions to the user, the DM is free to choose any item in the plan, or
37 even to do something not included in the plan at all. When information
38 from the Mobile Companion is available, it can direct where the dialogue
39 starts from. The DM could also decide to do some small-talk to introduce
40 sensitive topics, which can be useful in managing the user-system relationship
41 (Bickmore and Picard, 2005). The communication that is not related to the
42 domain does not reach the CM at any point. In another words, DM takes
43 care of the conversational dialogue topics and meta-communication, while
44 CM is involved in task-related topics.
45

46 In addition to the essential information, such as those presented in Fig-
47 ure 7, the CM can include additional annotation in the plan. First, DM
48 can mark certain information requests as being important. If information is
49 marked important, it is likely, but not certain, that the DM will explicitly
50 confirm it. Second, CM can explicitly request a confirmation by generating
51

```

1
2
3
4
5
6
7
8
9      ( <plan-item>
10     <action>QUERY-PLANNED-ACTIVITY</action>
11     </plan-item>)
12

```

13 **C1:** Good morning. Anything interesting organized for today?

14 **U1:** I'm going jogging.

```

15
16
17
18     (<pred>
19     <action>PLANNED-ACTIVITY</action>
20     <param>ACTIVITY-JOGGING</param>
21     <param>unknownTime</param>
22     </pred> )
23
24

```

25 **C2:** Is that jogging exercise before dinner?

26 **U2:** No, its after.

```

27
28
29
30     ( <pred>
31     <action>PLANNED-ACTIVITY</action>
32     <param>ACTIVITY-JOGGING</param>
33     <param>AFTER-DINNER</param>
34     </pred> )
35
36

```

37
38
39 Figure 7: A dialogue fragment and corresponding plan items and predicates

40
41
42 a separate plan item. For example, if the users report on having run much
43 more than they are likely to be capable of in their condition, the CM can
44 generate a confirmation plan item. It is worth noting, that the DM cannot
45 do reasoning on such a high level (as it would be not meaningful to duplicate
46 such information in DM), and therefore the CM must participate in error
47 handling in such cases.

50 51 4.2. Cognitive Modelling

52
53 In order to make the interaction with the Health and Fitness Companion
54 coherent, we need to provide a shared cognitive model for the Companion.
55 The system architecture provides support for this in the form of the separate
56

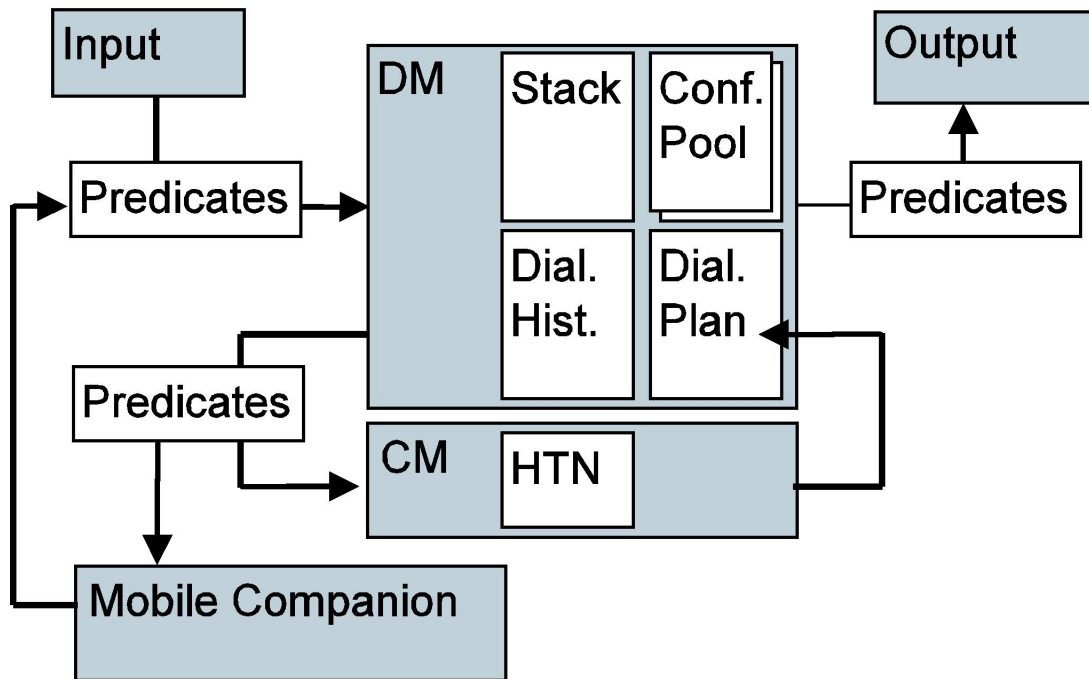


Figure 8: Dialogue management and cognitive modelling components and their information passing

Cognitive Model Manager. The task of the CM is to model the domain, i.e., know what to recommend to the user, what to ask from the user and what kind of feedback to provide.

At the core of the CM is an activity model, which decomposes the day into a series of activities for an office worker during a typical working day. These activities cover transportation to/from work, post-work leisure activities and meals (both in terms of the food consumed and how this food is obtained). A Hierarchical Task Network (HTN) planner is used to generate the activity model in the form of an AND/OR graph.

The planner that implements the activity model includes 16 axioms, 111 methods (enhanced with 42 semantic categories and 113 semantic rules), and 49 operators. The cognitive modelling of the Health and Fitness Companion is presented in detail by Hakulinen et al. (2008); Cavazza et al. (2008); Smith et al. (2008); Turunen et al. (2008a). Here, we focus in its interoperability with the dialogue management.

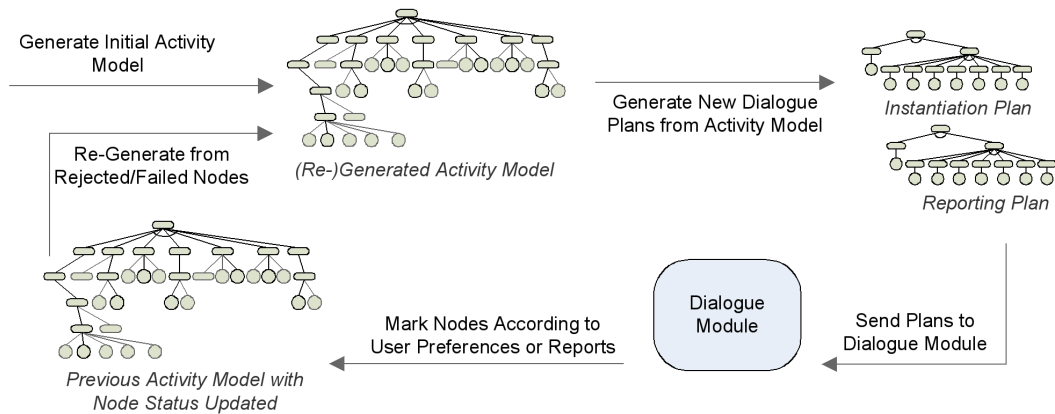


Figure 9: Activity model cycle

The Health and Fitness dialogues have currently two phases. The planning phase involves the system operating through a global interaction cycle, integrating dialogue with planning to construct the user's activity model. This process is illustrated in Figure 9 with a detail of the graph in Figure 10. The first step consists of generating an initial activity model for the user, based on default knowledge, along with any previously captured information on user preferences. Throughout this interaction cycle, each time the Planner generates a candidate activity model, it generates a corresponding dialogue plan, which is then used as a basis for the dialogue to enquire about user preferences and make suggestions in relation to the planned activities.

User responses are used to update the activity model, with the user utterance mapped to the predicates used within the planning domain and semantic categories associated with the domain methods. If a user response validates part of the existing activity model, that part of the model is marked as planned and will not appear in the dialogue plan. If the user response is incompatible with the current activity model, either through explicit rejection or stating of a conflicting preference, the preference is stored so that the relevant items remains unplanned and inactive.

The cycle continues with the activity model being re-generated. Those tasks that have been accepted, that is, marked as planned, are preserved while the unplanned parts are re-planned making use of the latest preferences provided by the user. An updated dialogue plan is then generated and the dialogue with the user continues until the user has agreed on a fully planned

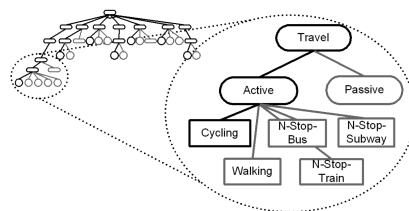


Figure 10: Activity Model Items

activity model.

The reporting phase is accomplished in a similar manner, with a dialogue plan asking the user questions about what he/she did and the user utterances being used to produce an activity model with completed activities.

4.3. Language Understanding and Dialogue Management in the Home Companion

Natural language understanding relies heavily on SISR information (see previous section). The context specific ASR grammars contain tags, which provide logical items to NLU component as part of ASR results. These logical items are either complete predicates, or in most cases, individual parameter values and lists of parameter values. This provides a basis for further input processing, where input is parsed against the current dialogue state to compile full, logical representations compatible with the planning implemented in the Cognitive Model. Input is parsed first against the previous system output. In practise, NLU component tries to unify the logical items into the space of predicates, which match the last system utterance. For example, if the utterance was "Why don't you cycle to work", the possible predicate space for answers includes acceptance and rejection of the suggestion, specification of an alternative travel method, and an expression of preference for or against the suggested travel method. If user input was "yes, that would be fine", the SISR based item is "ACCEPT-TASK", and the unification results in the complete predicate "ACCEPT-TASK TRAVEL-METHOD CYCLING-TRAVEL HOME WORK".

If the first attempt does not result a full parse, then a new predicate space is built from predicates related to the current topic. In the case of the above example, this would be everything related to the different possibilities when traveling to work. The final parsing option is the parsing against the entire space of known predicates. This set includes generic things like help requests

1
2
3
4
5
6
7
8
9 and enabled topic shifts in this domain. If needed, deeper analysis can be
10 made by parsing against the entire dialogue history and the current plan.
11 This way, the information produced by the CM is used in input parsing.
12 The dialogue plan could also be used in dynamic construction of recognition
13 grammars to support this on ASR grammar level.
14

15 In addition to unification of predicates, a reduced set of DAMSL dialogue
16 acts (Core and Allen, 1997) is used to mark functional dialogue acts using
17 rule-based reasoning.
18

19 As said, the task of the DM is to maintain and update a dialogue state.
20 In the Home Companion, the dialogue state includes a dialogue history, a
21 stack of active dialogue topics, and current parsed user input, including ASR
22 confidence scores and N-best lists. In addition, two pools of items that need
23 to be confirmed are stored; one for items to be confirmed individually and
24 another for those that can be confirmed together with a single question.
25

26 The DM receives user inputs as predicates parsed by the NLU compo-
27 nent. The DM is aware of the relations of the predicates on the topics level,
28 i.e., it knows which predicates belong to each topic. This information is used
29 primarily for parsing input in relation to the dialogue context. The DM also
30 has understanding of the semantics of the predicates which relates to interac-
31 tion. Namely, relations such as question answer pairs (suggestion agreement,
32 confirmation acceptance/rejection, etc.) are modelled by the DM.
33

34 If an utterance is successfully parsed and matches the current dialogue
35 plan, the DM does not need to know the the actual meaning of the input.
36 Instead, it takes care of confirmations, when necessary, and provides the
37 information to the CM. Similarly, when generating output requests based on
38 the plan, the DM can be unaware of the specific meaning of the plan items.
39 Overall, the DM does not need to have the deep domain understanding the
40 CM specializes in.
41

42 Dialogue management is implemented as a collection of separate compact
43 dialogue agents, each taking care of a specific task. In each dialogue turn
44 one or more agents are selected by the DM, as presented in Section 3.1. In
45 the current Home Companion, there are over 30 dialogue agents. There is
46 a separate agent for each topic that can occur in the plan. In practice, one
47 topic maps to a single planitem. These agents are all instances of a single
48 class with specific configurations. Each agent handles all situations related
49 to its topic; when the topic is the first item of an active plan, the agent
50 produces related output, and when the user provides input matching to the
51 topic it forwards that information back to the Cognitive Model. In addition,
52
53
54
55
56
57
58

1
2
3
4
5
6
7
8
9 topic specific agents handle explicit topic switch requests from the user (e.g.,
10 “let’s talk about lunch”) and also take turns if the topic is found on top of the
11 dialogue topic stack. A topic ends up on the stack if it has not been finished
12 when a new topic is activated. The Dialogue manager processes stack items
13 before returning to dialogue plan items. Currently, we have not encountered
14 any situations where the dialogue stack needs to be flushed. This might be
15 necessary in more complex dialogues.
16
17

18 The generic, non-topic specific agents include one that generates a con-
19 firmation if the ASR confidence score is too low, one that repeats the last
20 system utterance when the user requests it (“eh, come again?”), and an agent
21 to handle ASR rejection errors. There is no need to model such information
22 in the CM.
23
24

25 4.4. Dialogue Management in the Mobile Companion

26 In the Mobile Companion, the dialogue with the user is managed by
27 the script code running inside the Java midlet, using a finite state machine
28 model. At any point in time there is one active dialogue state. Each state is
29 represented by four script procedures, enter, leave, input, and error.
30
31

32 A typical use of these procedures in the Mobile Companion is to output
33 a question in the **enter** procedure, and then analyse the user’s reply and
34 move on to another state or ask another question in the **input** procedure.
35 The **leave** procedure is used to do state clean-up tasks (if necessary), and
36 the **error** procedure will output an error message and repeat the question.
37
38

39 SISR expressions are used in the Mobile Companion as well. The SISR
40 expressions consist of predicates of the form
41

```
42 #TASK_SELECTED(walking, home, work)
```

43
44 and are used directly by the script code to analyse the input. The user can
45 provide input using voice, by pressing buttons, or tapping on a list on touch
46 screen. However, all user input has the same form, so it does not matter how
47 the input was produced.
48

49 Natural language generation is handled by the state procedures directly.
50 There are some canned utterances and some dynamically constructed output,
51 for instance, based on the content of the activity plan downloaded from the
52 Home Companion.
53
54
55
56
57
58

4.5. *Benefits of the Model for Conversational Distributed Dialogues*

The presented model for interoperability between the Mobile Companion, the DM of the Home Companion, and the CM has provided great flexibility to each component. While the dialogue plan generated by the CM provides a base for dialogue management, which, in most cases, is followed, the DM can deviate from it when feasible. The DM can handle confirmations as it pleases, add small talk, and process the plan items in any order. The model also supports mixed-initiative dialogues; while the DM may follow the plan, the user may discuss any topic.

Most importantly, all this is possible without including domain specific knowledge into DM. All such information is kept exclusive in the CM. Similarly, the CM does not need to know the interaction level properties of the topics, such as recognition grammars and natural language generation details. These are internal to their specific components. The Mobile Companion uses the same knowledge representation as the CM, but the CM does not need to be aware of its existence at all. Similarly, the Mobile Companion can use any part of the information it receives, but is not forced to do anything specific. The DM just feeds all the information to the mobile Companion and lets it decide what to do with the information. When the Mobile Companion provides information back to the Home Companion, the DM handles the access to the information and the CM can ignore completely the fact that different parts of the information it receives were generated using different systems. Similarly, the Mobile Companion does not see any of the internals of the Home Companion.

At the implementation level, the model is independent of the mechanics of either the DM or the CM. The DM can be implemented using state transition networks (a network per plan item), forms (form per item), an agent-based model, like in the case of the Mobile Companion, or any other suitable method. Similarly, the plan does not tie the CM to any specific implementation.

5. **Evaluation of Companions**

Since Companionship in the sense described in the present paper is a fairly new concept, no agreed-on evaluation strategies of it exist, and one of the goals of the Companions project is to develop such evaluation strategies. In order to gain deep understanding on the Companions paradigm, user studies must be conducted both in laboratory and field settings. Thus an initial set

1
2
3
4
5
6
7
8
9 of user studies of the Health and Fitness Companion prototypes has focused
10 on out-of-the-box functionality to set a baseline for further studies. Two
11 mechanisms have been utilized for the evaluation: qualitative and quantita-
12 tive. Qualitative surveys are used to acquire subjective opinions from the
13 users; including Likert-based surveys, focus groups and interviews, while the
14 quantitative measures are based on the following set of metrics:
15
16

17
18 **Speech Metrics**, such as WER (word error rate) and CER (concept error
19 rate);
20

21 **Dialogue Metrics**, e.g., dialogue duration, number of turns, word per turn
22 dialogue structure;
23

24 **Task Metrics**, e.g., task completion; and
25

26 **User Metrics**, e.g., user satisfaction, requirement elicitation.
27
28

29 The metrics mentioned are quite traditional, and they are included in
30 evaluation frameworks such as PARADISE (Walker et al., 1997). One op-
31 tion would have been to use such a complete evaluation paradigm. However,
32 our aim was a lightweight evaluation at this phase of the project instead of
33 running a comprehensive and laborous evaluation. Furthermore, the exist-
34 ing methods are mainly focusing on task-driven dialogues, and while we use
35 here metrics such as Task Completion rate (see definition below), our tasks
36 are rather loose, so we believe it would not be feasible to try to find sim-
37 ilar correlations between objective and subjective metrics as in more more
38 straightforward task-oriented systems.
39
40

41
42 Two evaluation sessions have been carried out so far. The first evaluation
43 session focused on WER and CER metrics for the Home Companion. The
44 CER is based on correctly identifying the semantic result of a user utterance.
45 For example, correctly identifying whether the user accepted or rejected a
46 proposal from the HFC. If a user utterance contained distinct semantic re-
47 sults, these were then counted as separate concepts when counting the score.
48 For example, if in reply to a proposal to have a rest after dinner the user
49 replies "No, I'll play a game of football." the system would have to identify
50 both a rejection of the HFC's proposal and a counter-proposal of playing
51 football. Since we wanted to find out a baseline for further work, we did
52 not include any clarification dialogues to this evaluation. Furthermore, there
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9 were no issues such as grounding taking place in the dialogues. For these
10 reasons, there was no need to model these when calculating CER.

11 Another metric calculated is Task Completion rate, which in the case of
12 Health and Fitness companion is specified as the percentage of the activity
13 model correctly instantiated. That is, it provides the percentage similarity
14 between the activity model as instantiated and the instantiation of an ideal
15 activity model with all utterances optimally understood.
16
17

18 The study involved 20 subjects who interacted with the Home Companion
19 in both planning and reporting phases. They were briefly introduced to the
20 concept of the Companion and the scenario, and provided with a set of slides
21 illustrating (via images of activities and food types) what was known to the
22 system. To avoid bias, the subjects were not shown examples of possible
23 utterances nor allowed to witness experiments with other subjects.
24

25 For the planning dialogues, the WER was 42% and CER 24%. For the
26 reporting dialogues, the WER was 44% and CER 24%. The fact that no
27 user training or speaker adaptation was carried out, along with the realis-
28 tic experimental conditions, explains the high Word Error Rate level. The
29 Concept Error Rate is lower, indicating some resilience to misrecognition of
30 portions of the user utterance. Still, Task Completion was high, with, on
31 average, 80% of the Activity Model being correctly instantiated in planning
32 dialogues and 95% being correctly instantiated in reporting dialogues. This
33 is better than the corresponding per utterance CER as the system was able
34 to eliminate some errors over the course of the entire dialogue.
35
36

37 The significant improvement of CER over WER is due to the reasonable
38 robustness in identifying the semantic meaning of the user's utterances. This
39 is partially due to the semantic results being constrained to a relatively low
40 number of possibilities compared to the reasonably large vocabulary user
41 input to the ASR. For example, in response to the HFC asking "How about
42 bringing a packed lunch to work?" the user responds "Yes, I will make a
43 sandwich". When this is recognised as "yes I love that sandwiches" this
44 results in a WER of 50% but the system is still able to identify the concepts
45 (agreement to the proposal) correctly. There is a similar degree of robustness
46 when instantiating the activity model. For example, the HFC can infer from
47 a counter-proposal that the original proposal was rejected despite potentially
48 missing out on an explicit rejection.
49

50 Also worth noting is that reporting dialogues tended to involve simpler
51 user utterances, such as basic confirmations, than those in the planning
52 phase, which was reflected by smaller average utterance length and higher
53
54
55
56
57
58

1
2
3
4
5
6
7
8
9 task model completion rate. The initial results show that even with relatively
10 high WER acceptable task completion rates can be obtained in this domain.

11
12 In the second evaluation session, the Home Companion and the Mobile
13 Companion were evaluated in more a complex setting. Eight participants
14 completed a study protocol of four distinct tasks: introductory tutorials, us-
15 ing prototypes, on-line surveys, and interviews. Each session began with an
16 introductory tutorial. These presentations (10–16 slides) introduced the pro-
17 totype, established its intentions, its limitations, what the prototype would
18 say and do, how to use the prototype, and gave the user suggestions in how
19 to respond. Participants then used the Home and Mobile Companion for
20 about 10–15 minutes each, completing an on-line questionnaire after each
21 session. Researchers were sitting in the background while the participants
22 interacted with the prototypes, and the participants were video-taped during
23 their interaction with the systems. At the end of the session, each partici-
24 pant was interviewed. In this study, the Home Companion language model
25 was more complex than in the previous session evaluation. The Word Error
26 Rate ranged between 51% and 79% (however, it must be noted that in one
27 case the error rate was over 100% because of massive amounts of rejection
28 errors). While the Word Error Rate was extremely high, the Concept Error
29 Rate (calculated by ignoring the order of recognized concepts) was somewhat
30 lower, ranging between 33% and 65%. These numbers are still high, but most
31 of the time the Concept Error Rate was reasonably good, with the major-
32 ity of the errors being insertions of several concepts in some specific cases.
33 The average length of user utterances varied between participants from 3.0
34 to 8.3 words for the Home Companion, while the average system utterance
35 length was 12 words. Even though the data set is too small for statistical
36 testing, we could see differences in how verbose different people were. The
37 user vocabulary size was surprisingly small, and only varied between 18 and
38 116 words, with an average of 55 words.

39
40
41
42
43
44
45
46 Future plans for the evaluation efforts include to target areas of perfor-
47 mance required in long-term collaborative conversational agents. These are
48 discussed in more detail in (Benyon et al., 2008).
49

50 51 52 **6. Summary**

53
54 In this paper, we have introduced the concept of multimodal Companions
55 that build long-lasting relationships with their owners to support their every-
56 day health and fitness related activities. While traditional spoken dialogue
57
58

1
2
3
4
5
6
7
8
9 systems have been task-based, the Health and Fitness Companion is designed
10 to be part of the user's life for a long time, months, or even years. This re-
11 quires that they are part of the life physically, i.e., that interactions can take
12 place in mobile settings and in home environments outside of traditional,
13 task-based computing devices. With the physical presence of the interface
14 agents, and spoken, conversational dialogue we aim at building social, emo-
15 tional relationships between the user and the Companion. Such relationships
16 should help us in motivating the users towards a healthier lifestyle. The mo-
17 bility of the interface integrates the system into the physical activities it aims
18 at supporting users in.

19
20
21
22 We have shown what kind of architectures such distributed agent inter-
23 faces need, and how they can be realized with proper dialogue and cognitive
24 management techniques. When dialogue systems move beyond limited task-
25 based domains and implement multimodal interfaces in pervasive computing
26 environments, complexity increases rapidly. Dialogue management, which in
27 most cases is handled with well-understood methods such as form filling or
28 state transition networks, tends to grow more complex. Therefore, a model
29 to modularize dialogue management and domain reasoning is needed. At the
30 same time, distributed systems require various kinds of information to be
31 communicated between components and systems.

32
33
34
35 We have presented a novel interaction management model, which sep-
36 arates cognitive modeling from dialogue management and enables flexible
37 interoperability between these two. This model also enables sharing the
38 gathered knowledge to the mobile part of the system and back. This divi-
39 sion, while similar to separation of a back-end from dialogue management,
40 draws the line deeper into the area of interaction management. The pre-
41 sented model has been implemented in the Health and Fitness Companion
42 prototype, and it has enabled the Cognitive Model, the Dialogue Manager,
43 and the Mobile Companion interface to be developed in parallel by different
44 groups using various programming languages.

45
46
47
48 On software architecture level, the solution has enabled the Dialogue
49 Manager to focus only on interaction level phenomena, such as initiative
50 and error management, and other meta-communication while the cognitive
51 model takes care of the domain level processing. The Dialogue manager can
52 also include input from a mobile interface of the system without making this
53 explicit to the cognitive model. One example of flexibility is error manage-
54 ment; while the actual error correction is the task of the Dialogue Manager,
55 domain level knowledge can reveal errors. Using the dialogue plan, the cog-
56
57
58

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

nitive model can provide such information to the Dialogue Manager without knowledge of details of error management. The model also enables user initiative topic shifts, management of user-system relationship, and other such issues relevant in domain-oriented dialogue systems.

In overall, the model presented has enabled a clear division and interoperability of the different components handling separate parts of the interaction. We can recommend the model for other applications, where complex domain modeling is used in a situation where error correction and other dialogue managements task must be handled as well. Using the model, the different component can each focus on their own tasks.

In addition to the interaction model presented, we have produced concrete software components to implement further Companion applications. The PART system used to implement the Mobile Companion and the Jaspis architecture (including the jNabServer software) used for the Homa Companion have been released as open source software. Together, they can be used to construct similar distributed multimodal applications with virtual, physical, and mobile conversational agents.

Finally, we have introduced the first evaluation results of the Companions paradigm. These initial results show that even with relatively low speech recognition accuracy meaningful conversations can be carried out in this domain.

Acknowledgements

This work was funded by the COMPANIONS project ⁹ sponsored by the European Commission as part of the Information Society Technologies (IST) programme under EC grant number IST-FP6-034434.

References

- Benyon, D., Hansen, P., Webb, N., Oct. 2008. Evaluating human-computer conversation in Companions. In: Proceedings of the 4th International Workshop on Human-Computer Conversation. Bellagio, Italy, pp. 1–5.
- Bickmore, T. W., Picard, R. W., Jun. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction* 12 (2), 293–327.

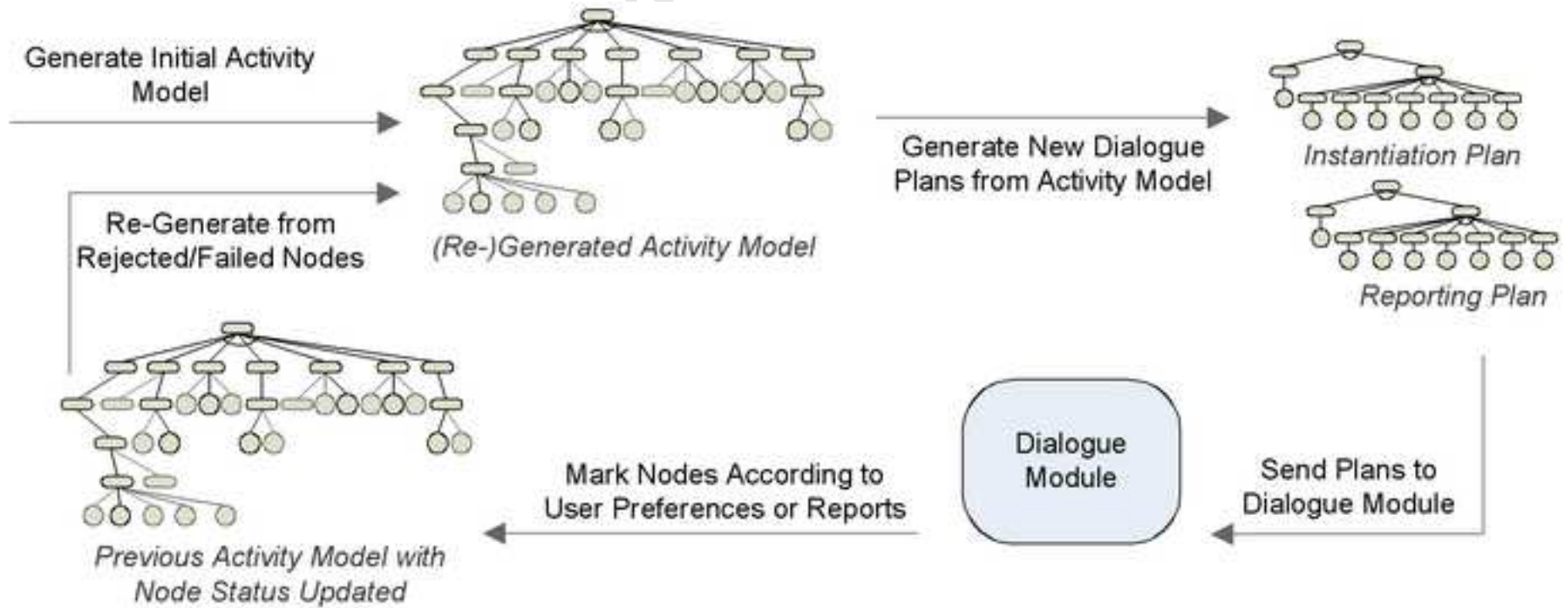
⁹www.companions-project.org

- 1
2
3
4
5
6
7
8
9 Buttussi, F., Chittaro, L., 2008. MOPET: A context-aware and user-adaptive
10 wearable system for fitness training. *Artificial Intelligence in Medicine*
11 42 (2), 153–163.
12
13 Cavazza, M., Smith, C., Charlton, D., Zhang, L., Turunen, M., Hakulinen,
14 J., May 2008. A ‘Companion’ ECA with planning and activity modelling.
15 In: Padgham et al. (2008), pp. 1281–1284.
16
17 Core, M. G., Allen, J. F., Nov. 1997. Coding dialogs with the DAMSL an-
18 notation scheme. In: *AAAI Fall Symposium on Communicative Action in*
19 *Humans and Machines*. Cambridge, Massachusetts, pp. 28–35.
20
21 de Oliveira, R., Oliver, N., Sep. 2008. TripleBeat: Enhancing exercise perfor-
22 mance with persuasion. In: *Proceedings of 10th International Conference*
23 *on Mobile Human-Computer Interaction*. ACM, Amsterdam, the Nether-
24 lands, pp. 255–264.
25
26 Dybkjaer, L., Bernsen, N. O., Minker, W., 2004. Evaluation and usability
27 of multimodal spoken language dialogue systems. *Speech Communication*
28 43 (1-2), 33–54.
29
30 Hakulinen, J., Turunen, M., Salonen, E.-P., 2007. Visualization of Spoken
31 Dialogue Systems for Demonstration, Debugging and Tutoring. In: *Pro-*
32 *ceedings of Interspeech 2005*, pp. 853-856.
33
34 Hakulinen, J., Turunen, M., Smith, C., Cavazza, M., Charlton, D., Oct.
35 2008. A model for flexible interoperability between dialogue management
36 and domain reasoning for conversational spoken dialogue systems. In: *Pro-*
37 *ceedings of the 4th International Workshop on Human-Computer Conver-*
38 *sation*. Bellagio, Italy, pp. 29–34.
39
40 Hernández, A., López, B., Pardo, D., Santos, R., Hernández, L., Relaño,
41 J., Rodríguez, M. C., May 2008. Modular definition of multimodal ECA
42 communication acts to improve dialogue robustness and depth of intention.
43 In: Padgham et al. (2008), pp. 10–17, 1st Functional Markup Language
44 Workshop.
45
46 Jönsson, A., 1991. A natural language shell and tools for customizing the
47 dialogue in natural language interfaces. Research Report LiTH-IDA-R-91-
48 10, Dept. of Computer and Information Science, Linköping University,
49 Linköping, Sweden.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9 Jullien, C., Marty, J.-C., Apr. 1989. Plan revision in person-machine dia-
10 logue. In: Proceedings of the 4th Conference of the European Chapter of
11 the Association for Computational Linguistics. ACL, University of Manch-
12 ester Institute of Science and Technology, Manchester, England, pp. 153–
13 160.
14
15
16 Kadous, M. W., Sammut, C., 2004. InCa: A mobile conversational agent. In:
17 Proceedings of the 8th Pacific Rim International Conference on Artificial
18 Intelligence. Auckland, New Zealand, pp. 644–653.
19
20
21 Kainulainen, A., Turunen, M., Hakulinen, J., Salonen, E.-P., Prusi, P., Helin,
22 L., 2005. A speech-based and auditory ubiquitous office environment. In:
23 Proceedings of 10th International Conference on Speech and Computer.
24 Patras, Greece, pp. 231–234.
25
26
27 Larsson, S., Traum, D., 2000. Information state and dialogue management
28 in the TRINDI dialogue move engine toolkit. Natural Language Engineer-
29 ing Special Issue on Best Practice in Spoken Language Dialogue Systems
30 Engineering, 323–340.
31
32
33 López, B., Hernández, A., Díaz, D., Fernández, R., Hernández, L., Torre,
34 D., Jun. 2007. Design and validation of ECA gestures to improve dialogue
35 system robustness. In: Proceedings of the 45th Annual Meeting of the
36 Association for Computational Linguistics. ACL, Prague, Czech Republic,
37 pp. 67–74, workshop on Embodied Language Processing.
38
39
40 López, B., Hernández, A., Pardo, D., Santos, R., del Carmen Rodríguez,
41 M., Apr. 2008. ECA gesture strategies for robust SLDS. In: AISB Annual
42 Convention: Communication, Interaction and Social Intelligence. Vol. 10.
43 SSAISB, Aberdeen, Scotland, pp. 69–76, Symposium on Multimodal Out-
44 put Generation.
45
46
47 López Mencía, B., Hernández Trapote, A., Díaz Pardo de Vera, D.,
48 Torre Toledano, D., Hernández Gómez, L. A., López Gonzalo, E., Nov.
49 2006. A good gesture: Exploring nonverbal communication for robust
50 SLDSs. In: IV Jornadas en Tecnología del Habla. Zaragoza, Spain, pp.
51 39–44.
52
53
54 Loquendo, 2008. Loquendo embedded technologies: Text to speech and au-
55 tomatic speech recognition.
56 URL www.loquendo.com/en/brochure/Embedded.pdf
57
58

- 1
2
3
4
5
6
7
8
9 Marti, S., Schmandt, C., Oct. 2005. Physical embodiments for mobile communication agents. In: Proceedings of the 8th Annual Symposium on User Interface Software and Technology. ACM, Seattle, Washington, pp. 231–240.
- 10
11
12
13
14
15 Oliver, N., Flores-Mangas, F., Sep. 2006. MPTrain: A mobile, music and physiology-based personal trainer. In: Proceedings of 8th International Conference, on Mobile Human-Computer Interaction. ACM, Espoo, Finland, pp. 21–28.
- 16
17
18
19
20
21 O’Neill, I., Hanna, P., Liu, X., McTear, M., Sep. 2003. The Queen’s Communicator: An object-oriented dialogue manager. In: Proceedings of the 8th European Conference on Speech Communication and Technology. ISCA, Geneva, Switzerland, pp. 593–596.
- 22
23
24
25
26
27 Padgham, Parkes, Müller, Parsons (Eds.), May 2008. Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS, Estoril, Portugal.
- 28
29
30
31
32 Pellom, B., Ward, W., Pradhan, S., Oct. 2000. The CU Communicator: An architecture for dialogue systems. In: Proceedings of the 6th International Conference on Spoken Language Processing. Vol. 2. Beijing, China, pp. 723–726.
- 33
34
35
36
37 Salonen, E.-P., Hartikainen, M., Turunen, M., Hakulinen, J., Funk, A. J., Oct. 2004. Flexible dialogue management using distributed and dynamic dialogue control. In: Proceedings of the 8th International Conference on Spoken Language Processing. Jeju, Korea, pp. 197–200.
- 38
39
40
41
42
43 Smith, C., Cavazza, M., Charlton, D., Zhang, L., Turunen, M., Hakulinen, J., 2008. Integrating planning and dialogue in a lifestyle agent. In: Intelligent Virtual Agents. Springer, Berlin, Germany, pp. 146–153.
- 44
45
46
47
48 Ståhl, O., Gambäck, B., Hansen, P., Turunen, M., Hakulinen, J., Oct. 2008. A mobile fitness Companion. In: Proceedings of the 4th International Workshop on Human-Computer Conversation. Bellagio, Italy, pp. 43–48.
- 49
50
51
52
53 Ståhl, O., Gambäck, B., Turunen, M., Hakulinen, J., Mar. 2009. A mobile health and fitness Companion demonstrator. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. ACL, Athens, Greece, pp. 65–68, demonstrations session.
- 54
55
56
57
58

- 1
2
3
4
5
6
7
8
9 Turunen, M. Jaspis - A Spoken Dialogue Architecture and its Applications.
10 PhD dissertation, University of Tampere, Department of Computer Sci-
11 ences A-2004-2, February 2004.
12
- 13 Turunen, M., Hakulinen, J., Rähkä, K.-J., Salonen, E.-P., Kainulainen, A.,
14 Prusi, P., 2005. An architecture and applications for speech-based accessi-
15 bility systems. *IBM Systems Journal* 44 (3), 485–504.
16
17
- 18 Turunen, M., Hakulinen, J., Smith, C., Charlton, D., Zhang, L., Cavazza,
19 M., Sep. 2008a. Physically embodied conversational agents as health and
20 fitness Companions. In: *Proceedings of the 9th Annual INTERSPEECH*
21 *Conference*. ISCA, Brisbane, Australia, pp. 2466–2469.
22
23
- 24 Turunen, M., Hakulinen, J., Ståhl, O., Gambäck, B., Hansen, P., Gancedo,
25 M. C. R., de la Camara, R. S., Smith, C., Charlton, D., Cavazza, M., Oct.
26 2008b. Multimodal agent interfaces and system architectures for health and
27 fitness Companions. In: *Proceedings of the 4th International Workshop on*
28 *Human-Computer Conversation*. Bellagio, Italy, pp. 49–54.
29
30
- 31 Walker, M., Litman, D., Kamm, C., Abella, A., 1997. PARADISE: a frame-
32 work for evaluating spoken dialogue agents. In: *Proceedings of the eighth*
33 *conference on European chapter of the Association for Computational Lin-*
34 *guistics*, pp. 271 - 280.
35
36
- 37 Wilks, Y., Nov. 2007. Is there progress on talking sensibly to machines?
38 *Science* 318 (9), 927–928.
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58



Script

Paella



Steps:



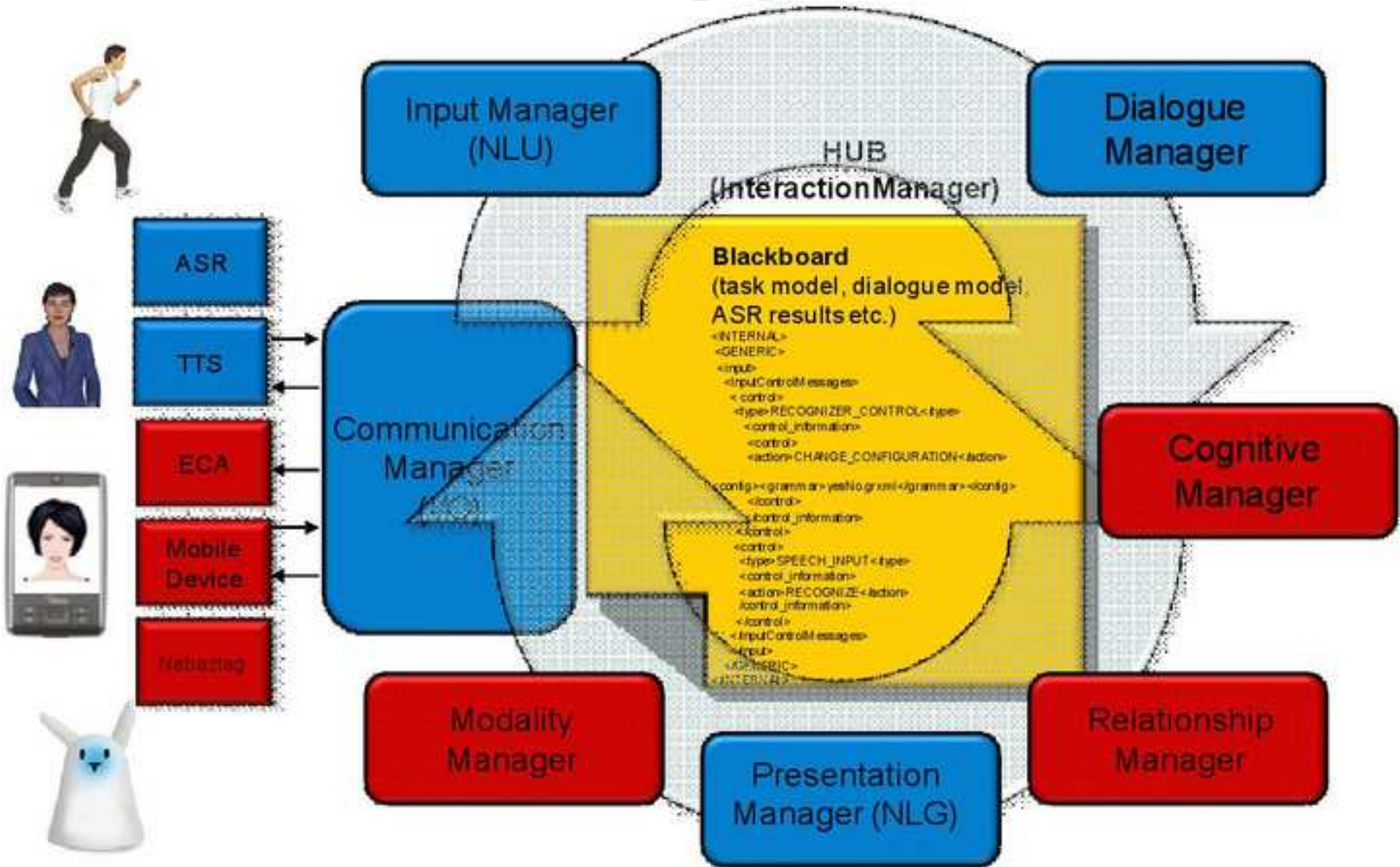
Step 3

Add to this your saffron water, squid, chili and pimento and season with salt and pepper. Give it a good stir...

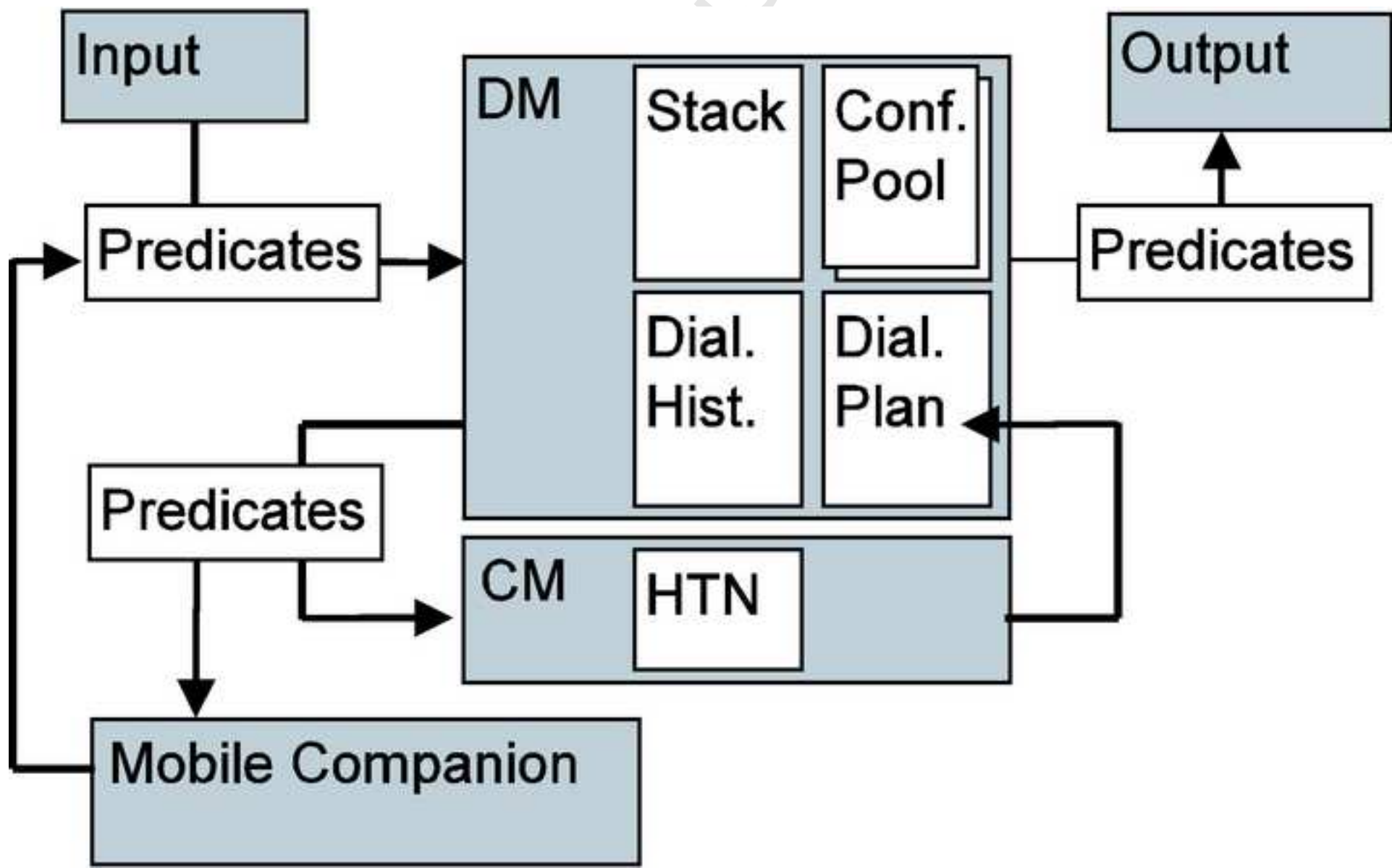


Cooking

Figure



scrip



Script

