

A Mobile Health and Fitness Companion Demonstrator*

Olov Ståhl¹ Björn Gambäck^{1,2} Markku Turunen³ Jaakko Hakulinen³

¹ICE / Userware
Swedish Inst. of Computer Science
Kista, Sweden
{olovs,gamback}@sics.se

²Dpt. Computer & Information Science
Norwegian Univ. of Science and Technology
Trondheim, Norway
gamback@idi.ntnu.no

³Dpt. Computer Sciences
Univ. of Tampere
Tampere, Finland
{mturunen,jh}@cs.uta.fi

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. The paper presents a multimodal conversational Companion system focused on health and fitness, which has both a stationary and a mobile component.

1 Introduction

Spoken dialogue systems have traditionally focused on task-oriented dialogues, such as making flight bookings or providing public transport timetables. In emerging areas, such as domain-oriented dialogues (Dybkaer et al., 2004), the interaction with the system, typically modelled as a conversation with a virtual anthropomorphic character, can be the main motivation for the interaction. Recent research has coined the term “Companions” to describe embodied multimodal conversational agents having a long lasting interaction history with their users (Wilks, 2007).

Such a conversational Companion within the Health and Fitness (H&F) domain helps its users to a healthier lifestyle. An H&F Companion has quite different motivations for use than traditional task-based spoken dialogue systems. Instead of helping with a single, well-defined task, it truly aims to be a Companion to the user, providing social support in everyday activities. The system should thus be a peer rather than act as an expert system in health-related issues. It is important to stress that it is the Companion concept which is central, rather than the fitness area as such. Thus it is not of vital importance that the system should be a first-rate fitness coach, but it is essential that it

*The work was funded by the European Commission’s IST priority through the project COMPANIONS (www.companions-project.org).

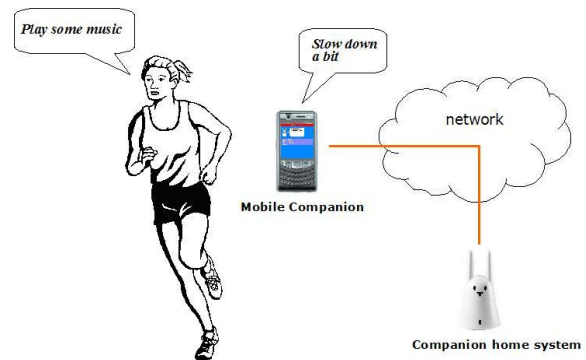


Figure 1: H&F Companion Architecture

should be able to take a persistent part in the user’s life, that is, that it should be able to follow the user in all the user’s activities. This means that the Companion must have mobile capabilities. Not necessarily self-mobile (as a robot), but allowing the user to bring the system with her, like a hand-bag or a pair of shoes — or as a mobile phone.

The paper describes such a Health and Fitness Companion. It has a stationary (“home”) component accounting for the main part of the user interaction and a mobile component which follows the users in actual exercise activities. Section 2 outlines the overall system and its two basic components, and Section 3 details the implementation. Section 4 discusses some related work, while Section 5 describes the demonstrator set-up and plans for future work.

2 The Health and Fitness Companion

The overall system architecture of the Health and Fitness Companion is shown in Figure 1. The system components communicate with each other over a regular mobile phone network. The home system provides an exercise plan to the mobile part and in return gets the results of the performed exercises from the mobile component.



Figure 2: Home Companion interface

2.1 The Home H&F Companion

The home part of the H&F Companion gathers information from the user and makes suggestions targeted for a healthy lifestyle on a daily basis. The Companion communicates with the user in two main dialogue phases: a planning phase where the system talks about the coming day with the user and a reporting phase where the user's actual activities are assessed with reference to what was agreed on earlier. The Companion can 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 football, squash, or badminton; going to the gym or shopping; and watching television or reading a book.

The Home Companion interface, illustrated in Figure 2, uses a Nabaztag/tag WLAN rabbit (www.nabaztag.com) as the physical agent interface. The Nabaztag rabbit provides audio output and push-to-talk speech input, moves its ears, and operates four coloured lights to signal its status. In the current implementation, the user needs to push the button located on top of the Nabaztag/tag device in order to speak to it. As an alternative, the system supports external micro-phones and loudspeakers. The user is also able to pick a topic by showing RFID tags (embedded in picture cards or other objects).

The following dialogue example demonstrates the interaction with the Companion.

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 get the train to work?
 U3: Yes
 C4: Right. Could you take in a packed lunch to work?
 U4: Yes, that would be fine

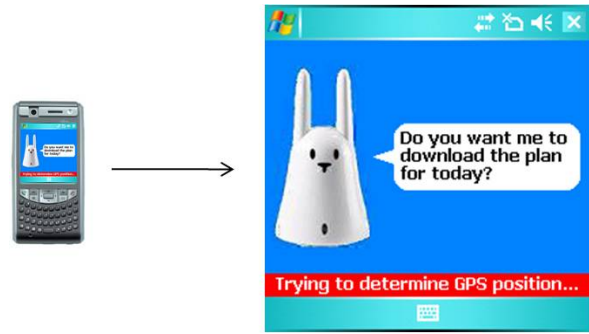


Figure 3: The Mobile Companion GUI

(Later on in the evening)

C6: Welcome back! How did your game of Squash go?
 U5: The game was cancelled, so I went to the gym
 C7: Okay. Did you take the train to work?
 U6: Yes, I did
 C8: Right. Did you take in a packed lunch?
 U7: No. I was in a rush so I went to the cafeteria

2.2 The Mobile H&F Companion

The mobile part of the H&F Companion runs on a mobile handset (e.g., a PDA), and is used during physical exercise (e.g., while running or walking) to track the distance, pace, duration, and calories burned. The data gathered during an exercise is stored in the device's record store, and can be used to compare the results to previous runs.

The user interface of the Mobile Companion consists of a single screen showing an image of a Nabaztag rabbit along with some text areas where various exercise and device status information is displayed (Figure 3). The rabbit image is intended to give users a sense of communicating with the same Companion, no matter if they are using the home or mobile system. To further the feeling of persistence, the home and mobile parts of the H&F Companion also use the same TTS voice.

When the mobile Companion is started, it asks the user whether it should connect to the home system and download the current plan. Such a plan consists of various tasks (e.g., shopping or exercise tasks) that the user should try to achieve during the day, and is generated by the home system during a session with the user. If the user chooses to download the plan the Companion summarizes the content of the plan for the user, excluding all tasks that do not involve some kind of exercise activity. The Companion then suggests a suitable task based on time of day and the user's current location. If the user chooses not to download the plan, or rejects the suggested exercise(s), the Companion instead asks the user to suggest an exercise.

Once an exercise has been agreed upon, the Companion asks the user to start the exercise and will then track the progress (distances travelled, time, pace and calories burned) using a built-in GPS receiver. While exercising, the user can ask the Companion to play music or to give reports on how the user is doing. After the exercise, the Companion will summarize the result and up-load it to the Home system so it can be referred to later on.

3 H&F Companion Implementation

This section details the actual implementation of the Health and Fitness Companion, in terms of its two components (the home and mobile parts).

3.1 Home Companion Implementation

The Home Companion is implemented on top of Jaspis, a generic agent-based architecture designed for adaptive spoken dialogue systems (Turunen et al., 2005). The base architecture is extended to support interaction with virtual and physical Companions, in particular with the Nabaztag/tag device.

For speech inputs and outputs, the Home Companion uses Loquendo™ ASR and TTS components. ASR grammars are in “Speech Recognition Grammar Specification” (W3C) format and include semantic tags in “Semantic Interpretation for Speech Recognition (SISR) Version 1.0” (W3C) format. Domain specific grammars were derived from a WoZ corpus. The grammars are dynamically selected according to the current dialogue state. Grammars can be precompiled for efficiency or compiled at run-time when dynamic grammar generation takes place in certain situations. The current system vocabulary consists of about 1400 words and a total of 900 CFG grammar rules in 60 grammars. Statistical language models for the system are presently being implemented.

Language understanding relies heavily on SISR information: given the current dialogue state, the input is parsed into a logical notation compatible with the planning implemented in a Cognitive Model. Additionally, a reduced set of DAMSL (Core and Allen, 1997) tags is used to mark functional dialogue acts using rule-based reasoning.

Language generation is implemented as a combination of canned utterances and tree adjoining grammar-based structures. The starting point for generation is predicate-form descriptions provided by the dialogue manager. Further details and

contextual information are retrieved from the dialogue history and the user model. Finally, SSML (Speech Synthesis Markup Language) 1.0 tags are used for controlling the Loquendo synthesizer.

Dialogue management is based on close-cooperation of the Dialogue Manager and the Cognitive Manager. The Cognitive Manager models the domain, i.e., knows what to recommend to the user, what to ask from the user, and what kind of feedback to provide on domain level issues. In contrast, the Dialogue Manager focuses on interaction level phenomena, such as confirmations, turn taking, and initiative management.

The physical agent interface is implemented in jNabServer software to handle communication with Nabaztag/tags, that is, Wi-Fi enabled robotic rabbits. A Nabaztag/tag device can handle various forms of interaction, from voice to touch (button press), and from RFID ‘sniffing’ to ear movements. It can respond by moving its ears, or by displaying or changing the colour of its four LED lights. The rabbit can also play sounds such as music, synthesized speech, and other audio.

3.2 Mobile Companion Implementation

The Mobile Companion runs on Windows Mobile-based devices, such as the Fujitsu Siemens Pocket LOOX T830. The system is made up of two programs, both running on the mobile device: a Java midlet controls the main application logic (exercise tracking, dialogue management, etc.) as well as the graphical user interface; and a C++-based speech server that performs TTS and ASR functions on request by the Java midlet, such as loading grammar files or voices.

The midlet is made up of Java manager classes that provide basic services (event dispatching, GPS input, audio play-back, TTS and ASR, etc.). However, the main application logic and the GUI are implemented using scripts in the Hecl scripting language (www.hecl.org). The script files are read from the device’s file system and evaluated in a script interpreter created by the midlet when started. The scripts have access to a number of commands, allowing them to initiate TTS and ASR operations, etc. Furthermore, events produced by the Java code are dispatched to the scripts, such as the user’s current GPS position, GUI interactions (e.g., stylus interaction and button presses), and voice input. Scripts are also used to control the dialogue with the user.

The speech server is based on the Loquendo Embedded ASR (speaker-independent) and TTS software.¹ The Mobile Companion uses SRGS 1.0 grammars that are pre-compiled before being installed on the mobile device. The current system vocabulary consists of about 100 words in 10 dynamically selected grammars.

4 Related Work

As pointed out in the introduction, it is not the aim of the Health and Fitness Companion system to be a full-fledged fitness coach. There are several examples of commercial systems that aim to do that, e.g., miCoach (www.micoach.com) from Adidas and NIKE+ (www.nike.com/nikeplus).

MOPET (Buttussi and Chittaro, 2008) is a PDA-based personal trainer system supporting outdoor fitness activities. MOPET is similar to a Companion in that it tries to build a relationship with the user, but there is no real dialogue between the user and the system and it does not support speech input or output. Neither does MPTrain/TripleBeat (Oliver and Flores-Mangas, 2006; de Oliveira and Oliver, 2008), a system that runs on a mobile phone and aims to help users to more easily achieve their exercise goals. This is done by selecting music indicating the desired pace and different ways to enhance user motivation, but without an agent user interface model.

InCA (Kadous and Sammut, 2004) is a spoken language-based distributed personal assistant conversational character with a 3D avatar and facial animation. Similar to the Mobile Companion, the architecture is made up of a GUI client running on a PDA and a speech server, but the InCA server runs as a back-end system, while the Companion utilizes a stand-alone speech server.

5 Demonstration and Future Work

The demonstration will consist of two sequential interactions with the H&F Companion. First, the user and the home system will agree on a plan, consisting of various tasks that the user should try to achieve during the day. Then the mobile system will download the plan, and the user will have a dialogue with the Companion, concerning the selection of a suitable exercise activity, which the user will pretend to carry out.

¹As described in “Loquendo embedded technologies: Text to speech and automatic speech recognition.” www.loquendo.com/en/brochure/Embedded.pdf

Plans for future work include extending the mobile platform with various sensors, for example, a pulse sensor that gives the Companion information about the user’s pulse while exercising, which can be used to provide feedback such as telling the user to speed up or slow down. We are also interested in using sensors to allow users to provide gesture-like input, in addition to the voice and button/screen click input available today.

Another modification we are considering is to unify the two dialogue management solutions currently used by the home and the mobile components into one. This would cause the Companion to “behave” more consistently in its two shapes, and make future extensions of the dialogue and the Companion behaviour easier to manage.

References

- Fabio Buttussi and Luca Chittaro. 2008. MOPET: A context-aware and user-adaptive wearable system for fitness training. *Artificial Intelligence in Medicine*, 42(2):153–163.
- Mark G. Core and James F. Allen. 1997. Coding dialogs with the DAMSL annotation scheme. In *AAAI Fall Symposium on Communicative Action in Humans and Machines*, pages 28–35, Cambridge, Massachusetts.
- Laila Dybkjaer, Niels Ole Bernsen, and Wolfgang Minker. 2004. Evaluation and usability of multimodal spoken language dialogue systems. *Speech Communication*, 43(1-2):33–54.
- Mohammed Waleed Kadous and Claude Sammut. 2004. InCa: A mobile conversational agent. In *Proceedings of the 8th Pacific Rim International Conference on Artificial Intelligence*, pages 644–653, Auckland, New Zealand.
- Rodrigo de Oliveira and Nuria Oliver. 2008. TripleBeat: Enhancing exercise performance with persuasion. In *Proceedings of 10th International Conference, on Mobile Human-Computer Interaction*, pages 255–264, Amsterdam, the Netherlands. ACM.
- Nuria Oliver and Fernando Flores-Mangas. 2006. MPTrain: A mobile, music and physiology-based personal trainer. In *Proceedings of 8th International Conference, on Mobile Human-Computer Interaction*, pages 21–28, Espoo, Finland. ACM.
- Markku Turunen, Jaakko Hakulinen, Kari-Jouko Riih , Esa-Pekka Salonen, Anssi Kainulainen, and Perttu Prusi. 2005. An architecture and applications for speech-based accessibility systems. *IBM Systems Journal*, 44(3):485–504.
- Yorick Wilks. 2007. Is there progress on talking sensibly to machines? *Science*, 318(9):927–928.