

THE COMPANIONS PROJECT: RESULTS FROM THE FIRST YEAR

*Enrico Zovato, Morena Danieli
April, 2008*

Loquendo is proud to announce that the FP6-IST COMPANIONS project, coordinated by Prof. Yorick Wilks from the University of Sheffield and with the participation of Loquendo, has successfully completed its first year. The first review of the project was carried out by five scientific experts from academic and industrial institutions nominated by the European Commission, and their conclusions were that the consortium is performing satisfactorily thanks to the validity of the research program and the sound technological base on which the project depends.

The objectives of the COMPANIONS project are to develop “autonomous, persistent, affective, and personal multi-modal interfaces”. In other words, the main goal is to develop personalised conversational interfaces, by means of agents that interact with the user on a long term basis, thus exploiting learning and adaptive strategies. Moreover, an important topic of the project is the analysis of emotional aspects typical of human to human interaction, while the objective is to shift them into the context of human-machine interaction in order to have really “personalised” interfaces. The project goal is to push the state of the art in machine based natural language understanding, knowledge structures, speech recognition, natural language generation and text to speech.

This Integrated Project is organized into two time phases and four Working Areas. Phase one is for literature research and data collection, to compare algorithms and to integrate the available software modules in order to release the preliminary prototypes. In Phase 2, work will concentrate on innovative research, and the final demonstrators (i.e. the personalised conversational agents) will be developed starting from the Phase 1 prototypes. Another important part of this phase will be evaluation tasks.

The working Areas span across both time phases. Area 1 deals with multi-modal architectures and components. Its goal is the design, integration and implementation of the modules involved in the project. Area 2 regards interaction design, evaluation of the performance of the prototypes, Wizard of Oz experiments and data collection. Area 3 covers aspects related to Natural language understanding and generation and to the vocal interfaces. Area 4 concerns self-adaptive multi-modal architecture and provides knowledge representation technologies to be exploited in the COMPANIONS demonstrators.

Besides the University of Sheffield, the partners in the project are the University of Oxford, the University of Teesside, Napier University, Charles University, Swedish Institute of Computer Science, As An Angel, the University of Tampere, the University of West Bohemia, the University of Washington, Telefonica Investigacion y Desarrollo, France Telecom, Telia Sonera and the State University of New York at Albany.

Main Results

After only one year since the project's kick-off, the consortium has designed the basic architecture of the system and has worked on the development of the first two prototypes: the Senior Companion and the Health and Fitness Companion. The first aims to provide access to information and services as well as company for the elderly by chatting, remembering past conversations, and organizing the owner's photographic and image memories or accessing resources on the Internet. The second companion aims to keep track of the user's diet, and monitor and suggest physical exercise. Both systems share the same basic design by which the single modules can be interchanged. In fact, different solutions have been adopted concerning the Dialogue management, the Natural Language Processing module, and the Natural Language Generation part. In any case, all these modules will benefit from the research activities that are the core part of the whole project. Regarding speech technologies, both prototypes exploit Loquendo's state of the art speech synthesis and automatic speech recognition technologies

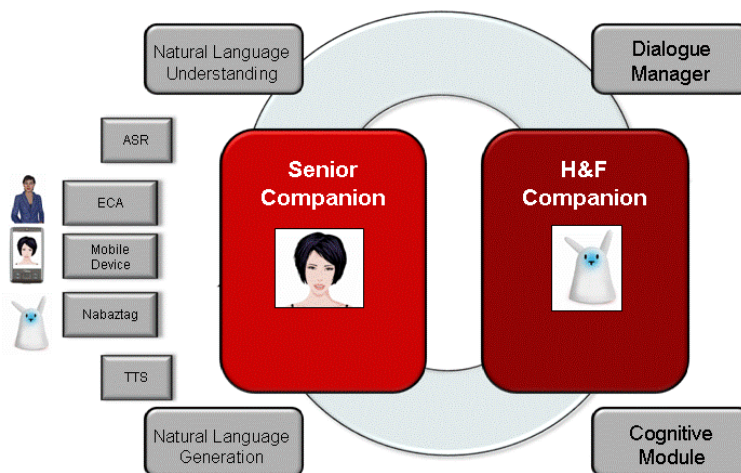


Figure 1- The Companions prototypes.

Communication is mainly based on conversational dialogue, but also on touch screens and sensors. These prototypes run on different platforms like Embodied Conversational Agents (ECA) on servers, mobile devices and the Nabaztag, exploiting different software modules. At the moment, research is mainly focused on Natural Language Processing, Dialogue Management Systems, Speech Technologies and Emotion.

The Interface of the Senior Companion, designed at Napier University, provides the basic functionalities needed for the initial implementation. It consists of a frame containing a single user photo and an avatar that interacts with the user via synthetic speech. Input can be provided by means of a speech understanding interface or by text input ([see the video on YouTube](#)). The Health & Fitness Companion prototype uses a Wi-Fi enabled Nabaztag rabbit to interact with the users. It communicates with the user primarily by understanding and generating speech, but also by using other technologies such as RFID (Radio Frequency Identification) sensors.



Figure 2 - [The H&F prototype on YouTube](#)

Research objectives

In terms of research, a significant area will be to make advances in field dialogue based systems, natural language processing and speech technologies. Regarding the former, a considerable challenge is to investigate a new approach to dialogue management based on machine learning. Machine learning techniques will be also tested and used to improve performance, user adaptation, scalability and portability to new languages and domains.

Research is focused on the improvement of speech and language technologies in order to allow the ECA to display personality, and to use speech characteristics such as politeness and humor in different types of dialogues. To this end, an important role is played by cognitive models and semantic representations that are a core part of the research program. Regarding speech technologies, the Czech language will be developed both for the ASR and TTS systems, and a new English ASR module will be integrated into the prototype. Speech synthesis will be enhanced by the introduction of expressive features, i.e. the ability to pronounce sentences in different styles, characterized by different prosodic contours, depending on the input target.

Loquendo's commitment to the project

Loquendo TTS and ASR technologies, that are successfully integrated into the two prototypes and that benefit from the latest advances developed in our labs, are the core modules of the vocal interfaces of the prototypes based on the English language.

The baseline ASR system for the English language has been provided to the partners involved in the development of the prototypes and in their evaluation. Both language models and context-free grammars have been designed and developed in order to improve performance. In particular, in the Senior Companion, ASR relies on statistical Language Models, trained on specific corpora, while the H&F Companion exploits dynamically loaded hand written grammars.

Regarding speech synthesis, the British English voices could benefit from the improvements carried out during the first year of the project. These improvements regard The Grapheme To Phoneme conversion module that provides lexical stress assignment and phonetic transcription. The adopted solution exploits a language-independent learning scheme and it is trained with revised and incremented lexicons. Moreover, the new solution is more flexible and scalable in terms of the number of rules extracted, and includes the option to manually add transcription rules with higher priority than the learned ones.

The phrasing module has also been improved. Its goal is to identify a set of candidates for possible phrase boundaries and, according to certain rules, to correctly insert variable length pauses in the synthesized speech. To accomplish this task it relies on morpho-syntactic and lexical knowledge. Activities in this field included the revision of syntactic rules, by improving the lexical data base, and a fine-grained classification for prepositional locutions and discourse markers. Moreover, an extended set of POS (Parts of Speech) labels has been considered that includes different sub-classes for verbs.

The goal for Loquendo in the next phase of the project will be to deal with the application-specific emotional aspects, like adding expressivity to the synthetic voices in an effective way.

Conclusions

During the intensive review meeting, all the participating sites presented their activities in meticulous detail, as well as the results achieved after one year since the project's kick-off. Moreover, the two prototypes, and a preliminary demo of the Health and Fitness Companion, were shown to the reviewing commission, which is composed of outstanding representatives from academic and industrial institutions.

The reviewing commission appreciated the technological background, as well as the challenging research topics, on which the project depends. The management of the Integrated Project was also judged positively and the good collaborative spirit between the partners was considered a plus. The Senior Companion was evaluated as a promising application since it is based on original ideas and concepts.

The meeting was also an important opportunity to receive valuable feedback from the reviewers: they suggested further consideration of some crucial aspects of the project, like an accurate user analysis and the emotional/human aspects of the future demonstrators. These aspects are especially important for making further advances and for taking the project forward.

References

COMPANIONS website: <http://www.companions-project.org>

Community Research & Development Information Service:
<http://cordis.europa.eu/en/home.html>

FP6 home page:
<http://cordis.europa.eu/fp6/dc/index.cfm?fuseaction=UserSite.FP6HomePage>

Nabaztag by Violet: <http://www.nabaztag.com>