1 Publishable summary

1.1 Introduction

The DIRHA (Distant-speech Interaction for Robust Home Applications) project was launched as STREP project FP7-288121 in the Commission's Seventh Framework Programme on January 1st, 2012, with a duration of 36 months. The Action Line of the related call was FP7-ICT-2011-7 Objective 4.2 – Language Technologies.

1.1.1 General goals

The DIRHA project addresses the challenge of natural spontaneous speech interaction with distant microphones in a home environment.

The main fields on which research is conducted, and for which suitable solutions will be identified and embedded in real-time prototypes, are: multichannel acoustic processing, distant speech recognition and understanding, speaker identification/verification, and spoken dialogue management.

The project also aims to investigate the use of a new type of acquisition device consisting of MEMS (Micro Electrical-Mechanical System) digital microphone arrays.

The project addresses four languages: Italian, Greek, Portuguese and German.

An intermediate prototype is foreseen during the second year, which will be installed in a real automated home, located in Trento, and will operate in the Italian language.

The final prototype will be able to interact in the given four languages, it will be integrated and evaluated by real end-users in their automated homes.

1.1.2 Facts and Figures

Project details	
Project Reference: 288121	**Contract Type:** Specific Targeted Research Project
Start Date: 2012-01-01	**End Date:** 2014-12-31
Duration: 36 months	**Project Status:** Execution
Project Cost: 4.85 million euro	**Project Funding:** 3.45 million euro

Coordinator	
Contact Person: Name: OMOLOGO MAURIZIO Tel: +39-0461-314563 Fax: +39-0461-314591 Email: omologo@fbk.eu	**Organisation:** Fondazione Bruno Kessler (FBK) VIA SANTA CROCE 77 38100 – TRENTO, ITALY

Other Beneficiaries	**Country**
ATHENA RC-IAMU	GREECE
DOMOTICAREA	ITALY
INESC-ID	PORTUGAL
NEWAMUSER	ITALY
ST Italy	ITALY
TU Graz	AUSTRIA

1.2 Scientific, technical, and application oriented objectives

Several scientific/technological challenges are tackled within DIRHA, most of them related with the problem of modifications encountered by acoustic waves propagating from a sound source (e.g. the speaker) to a set of far-field microphones, and to the consequent effect of such modifications on performance of a speech recognition system or, more generally, of a system for acoustic event interpretation.

One of the most challenging and innovative aspects of the project is the development of a distant speech interaction system, robust to speaker position, even in a noisy and reverberant environment and eventually in a multi-speaker context. Other projects have recently addressed this concept and tried to realize some early solutions. However, DIRHA investigates on a novel approach and on techniques for distant-speech interaction in a multi-room environment and possibly with multiple users.

Among the most relevant innovative aspects, it deserves to be mentioned that acoustic scene analysis is performed in an "always listening" mode (i.e., without the need of any push-to-talk button), with the goal of interpreting acoustic/speech activities concurring in the given environment, and eventually delivering speech chunks to the recognition and understanding components. To this end, one needs to realize robust technologies able to tackle unforeseen acoustic environments and noisy conditions. Such goals are new and far beyond the state of the art, not only for an application in the home scenario but also for other domains.

The targeted application includes voice-enabled interaction with appliances and other automatic services available in a household. Although in some cases users could simply try to speak close to the microphone and in a rather controlled way, the expectation is that in the future they would require being able to interact at four-five meters from microphones in a crowded room, with music playing, and other possible active sound sources. For some individuals (e.g. motor impaired), this is a strong immediate requirement, which is the main reason for addressing firstly this category of users under the DIRHA project. To this purpose, a group of possible end-users was involved from the beginning of the project, in order to define concrete and realistic user requirements. It is foreseen that the most advanced technologies resulting from the project will be integrated in a real-time prototype installed in automated homes, and daily used by the end-users for evaluation purposes.

1.3 Expected Results

The DIRHA project aims both to make advances at research level in the given scientific fields and to progress at technological level, with the development of a proof-of-concept system which can represent the starting point for a next exploitation action to be addressed by the involved industrial partners.

Research activities also include the creation of experimental tasks and corpora, which might also enable future initiatives of dissemination and benchmarking at international level.

As for the final prototype, it will run based on microphone devices installed in different rooms in order to monitor selectively acoustic and speech activities observable inside any space of the household. In the targeted scenario, the user can speak from any position in space, i.e. any point in any room of a house given any background noise and acoustic conditions typical of a household, and no matter of where the closest microphones are. A spoken dialogue session can be activated

based on a user request, for instance in order to switch on the light, open a door, have access to appliances and devices, or to services regarding emergency situations.

1.4 Impact

The final objective of the project targets application of automatic speech recognition in four languages with common multi-microphone front-end, spoken dialogue management, and user interface. This will have a relevant impact in terms of synergetic approach to the development of spoken language interaction systems and to the immediate evidence of a possibly easy portability to other languages. The project can also represent a milestone for developers and integrators of home automation systems, since the targeted prototype can be a first proof-of-concept realization in a real world context, based on concrete and realistic user requirements and operational constraints.

The DIRHA consortium aims to examine the impact of its novel technologies primarily with collaborative users. In other words, the DIRHA system is conceived for subjects who have, in principle, no difficulty in understanding the way to access the system in order to obtain the highest satisfaction (e.g., based on a high completion rate in the proposed tasks) and who have a very good attitude towards this experimentation. Once the basic technology has been established and evaluated as reliable, other categories of users (e.g., elderly people) may be addressed in future projects.

Another impact of the project regards the portability of the foreseen solutions to other possible domains. In fact, the DIRHA approach and the resulting technologies could eventually be applied to several application contexts characterized by noisy environment and by the need of talking far from the microphone as, for instance, robotics, surveillance, telepresence, gaming, industry sector and manufactory.

1.5 Achievements during the first year

1.5.1 General achievements

- A **User study** was conducted during the first months of the project in order to analyze requirements of a set of motor-impaired end-users who are available to experiment the DIRHA prototype. Following this activity, a set of scenarios and related functionalities were outlined, which also represented a guideline for the definition of the user interface to adopt. In particular, it emerged that functionalities as doors, windows and rolling shutter management and lights, temperature and entry-phone/interphone control are the most important features to be supported by voice for motor-impaired people. Based on this, during the last months of Year 1 Wizard-Of-Oz experiments were conducted to test the usability of a voice interface aimed at managing the above mentioned devices and other ones chosen by the users. In the meantime, the WoZ experiment allowed us to collect spontaneous speech of 11 subjects while giving voice commands to the simulated system.
- A first set of **Experimental tasks** was defined to support all the experimental activities which are being conducted under scientific and technological work packages. Both simulated and real acoustic and speech corpora were created. Real data includes collection of distant-speech material as well as acoustic measurements in home environments. Simulated data were produced thanks to a technique that reconstructs, in a very realistic manner, multi-microphone front-end observations of typical scenes occurring in a domestic

environment. To this purpose, an ITEA (Istituto Trentino per l'Edilizia Abitativa S.p.A.) apartment is available in Trento; this apartment will also represent the site where DIRHA prototypes will initially be developed and tested.

- Preliminary research activities on **Multi-microphone front-end processing** were conducted. Some algorithms have been selected and tested to assess their performance in the real application scenario and to determine their most suitable combination for a prototype implementation. The target is to derive, starting from the signals of the available multiple microphones, an input of sufficient quality for an effective speech recognition process. The possibility of using novel solutions based on MEMS digital microphone arrays is also explored.
- As far as **Distant-speech and speaker recognition** is concerned, preliminary activities were conducted on the development and evaluation of baseline recognition components in the four languages of interest in DIRHA and of the baseline distant speaker identification and verification. In addition to baselines description and experimental results, some novel research activities have been started.
- As for **Speech Understanding**, two approaches have been investigated, the more traditional one based on the use of hand-crafted grammars and the other based on a statistical framework. In particular, the latter topic is being investigated towards the development of a robust system for speech understanding in order to overcome the limitations offered by the former approach. A first set of related tools has been developed to this purpose.
- Activities on **Spoken Dialogue Management** regarded the definition of the project-wide strategy for the handling of Concurrent dialogues, each one taking place in separated spaces of the house, the design and development of a reusable Concurrent Dialogue Manager based on the StateCharts paradigm, and the Modelling of the User+House State and Profile.
- A significant effort was devoted to define the hardware setup to be used and the **System architecture** for the next development of the intermediate DIRHA prototype which will run in a real environment fully connected to an automated home (i.e., ITEA apartment).
- Nine **Showcases** were realized, which represent proofs of concept for what regards the state-of-the-art technologies on multi-microphone front-end processing, ASR and dialogue management, available at DIRHA partner sites. Some of the showcases refer to targeted functionalities for the intermediate prototype. In particular, a showcase was implemented in the ITEA apartment and does already support real-time execution of speech input requests for command-and-control of some devices.

1.5.2 Dissemination

- An official and publicly available project web site has been established (see http://dirha.fbk.eu), which represents the main means for presentation and dissemination of project achievements outside the consortium, including the most important information regarding the project and allowing the download of publicly available documents. At that site, most of the information about DIRHA can be found and downloaded (the brochure, state-of-the-art documents, etc.). The web site was regularly updated during the first year.
- General information about the project, promotional material (e.g. the DIRHA brochure) have already been made available to a wide range of specialists with different disciplines both inside scientific-technical communities and in the home automation related industrial field. In particular, the project was presented at LREC 2012 - EU Village in Istanbul, and at META-FORUM 2012 in Brussels.

- The project was presented in invited lectures and other events. Some scientific papers have already been presented at international conferences and workshops, while other ones have been submitted for review at forthcoming events.
- A public deliverable (D7.1) was produced, which will be made available soon in the project web site, in order to provide further information about the project as well as dissemination plan.