Deliverable 1.1
User study, analysis of requirements and definition of the application task

Authors: Fiorenza Arisio, Roberto Manione (New Amuser), Luca Cristoforetti, Marco Matassoni, Maurizio Omologo (FBK)

Affiliations: New Amuser, FBK

Date: May 31st, 2012

Document Type: R

Status/Version: Final 1.0

Dissemination Level: PU

Project Reference	FP7-ICT-2011-7 - 288121
Project Acronym	DIRHA
Project Full Title	Distant-speech Interaction for Robust Home Applications
Dissemination Level	PU
Contractual Date of Delivery	April 30, 2012
Actual Date of Delivery	May 31, 2012
Document Number	DIRHA_D1.1_20120531
Type	Deliverable
Status & Version	Final 1.0
Number of Pages	3+28
WP Contributing to the Deliverable	WP1
WP Task responsible	Fiorenza Arisio (New Amuser)
Authors (Affiliation)	Fiorenza Arisio, Roberto Manione (New Amuser), Luca Cristoforetti, Marco Matassoni, Maurizio Omologo (FBK)
Other Contributors	Franco Giovanazzi, Samuel Ramella (Domoticarea), Gernot Kubin, Martin Hagmüller, Hannes Pessentheiner (TUGraz), Gerasimos Potamianos (ATHENA RC –IAMU)
Reviewer	
EC Project Officer	Alexandra Wesolowska, Pierre Paul Sondag

Keywords: User study, Analysis of requirements, Application tasks, Scenarios, Interviews, Prototype functionalities, System architecture.

Abstract: This document reports on a user study which was conducted during the first months of the project, in order to analyse user requirements and to define application scenarios. The result of this study will characterize most of the activities of the next phase of the project and, in particular, the early design of components to eventually integrate into the intermediate prototype.

Summary

Summary iii
1. Introduction 1
2. Aim of requirements gathering 2
3. Research procedure and analysis method 2
3.1 Research techniques 3
3.2 Participants to the qualitative research 4
4. Content analysis 5
4.1 User expectations about voice-enabled interfaces 5
4.2 User expectations about system placement and possible features 6
4.2.1 Dialogue features and error recovery strategies 7
4.3 Integration of special devices through the Internet of Things approach 9
4.4 Check list results 9
5. Possible application scenarios and related tasks 11
5.1 Management of house facilities 12
5.1.1 Control of doors, windows and shutters 12
5.1.2 Control of light and electric sockets 13
5.1.3 Control of temperature 13
5.2 Management of communications 14
5.2.1 Control of interphone 14
5.2.2 Answer the entry-phone 14
5.2.3 Control of telephone 15
5.3 Management of leisure facilities 15
5.3.1 Management of HiFi audio system 15
5.3.2 Management of TV system 16
6. Identified functionalities vs system architecture 16
6.1.1 Control of doors, windows and shutters 19
6.1.2 Control of interphone 19
7. Discussion and conclusions 20
Annex A - Interview guidelines 23
Annex B - Voice automated services check list 24
Annex C - Preliminary analysis of requirements for the end-user available in Graz 26

1. Introduction

This deliverable is the first contribution to work package 1, "User requirements and system validation", and describes the activities so far conducted under task T1.1 "User study and analysis of requirements" and, partially, under task T1.2 "Application scenario and user interface definition".

According to Annex I – Description of Work (DoW), its purpose is to provide an overall overview of motor-impaired user needs in relation to the use of voice-enabled technologies with home automation. This is a very first step towards the definition of some guidelines for the design of the targeted DIRHA prototype, and of its components, from the microphone network for acoustic scene analysis, to speech recognition, understanding and spoken dialogue management. This activity, in fact, includes not only the results of this user study but also an analysis of the emerging scenarios and of the functionalities which should be addressed in the next phase of the project.

Note that a parallel activity has been conducted during the same period (i.e., first four months of the project) in order to define experimental tasks that can support the various research activities on the addressed technologies. From a possible convergence between scenarios, identified under the former user-driven activity, and experimental tasks, being defined under the latter technology-oriented one, the DIRHA consortium will take several important decisions, from those regarding acoustic-speech and text corpora collection, multi-microphone set-ups, etc., to those concerning the characteristics of speech understanding and spoken dialogue management components. So, the given work plays a relevant role in the definition of the main features of the intermediate prototype.

In a User Centered Design (UCD) approach, user requirement collection is useful to determine what is technically and economically feasible, keeping in mind what needs are most important for users. For this purpose, user needs, available functionalities, logistics and costs are considered in order to jointly maximize the chance of success and user satisfaction at the end of the project.
Different methodologies can be adopted to collect user requirements. Due to the fact that under DIRHA real end-users will mainly be persons with physical disabilities, we decided to use a qualitative approach to know their needs. For this reason, we chose to make the user requirements collection through in-depth interviews and filling a check list about possible home-automation services.

This approach was followed involving seven Italian motor-impaired individuals, most of them clients (or future clients) of DomoticArea, and potential end-users who will test the targeted DIRHA prototype once installed in their homes. Moreover, the document reports on an interview that was realized with another potential end-user in Graz (Austria). In the latter case, it is worth noting that due to logistics reasons and to the fact that the subject can not speak either Italian or English, we could not adopt the same methodology as done for the other seven end-users. As a consequence, the information regarding this interview is available in the Annex C of the document, and is not discussed in the main part of it.

The deliverable is structured as follows: Sections 2 and 3 introduce the methodology that was adopted, and provide information about the end-users who contributed to this user study; Section 4 reports on the results of this analysis, identifying the main requirements which emerged at end-user level; Section 5 aims to outline a possible set of scenarios and application tasks, while Section 6 outlines a possible hardware-software set-up based on the identified system functionalities. Finally, Section 7 draws some conclusions, discussing on the possible way this study can impact the next phases of the project.

2. Aim of requirements gathering

At the beginning of a project life-cycle, the phase known as requirements gathering represents the basis for the implementation of services.

To choose what kind of services/functionalities have to be implemented in an application of voice-enabled home automation, it is necessary to understand the end-user desires/needs, their priorities, and how related functionalities can be realized using state-of-the-art or targeted technologies.

A specificity and a real challenge of the DIRHA project is the introduction of distant-speech interaction in an application context, the home automation, in which voice-enabled systems are still not commonly used. Here, distant-speech interaction means that in principle the system can work, in any room, in an "always-listening mode" without the requirement of any push-to-talk button for its activation.

Another specificity of the project is that motor-impaired end-users can belong to different case studies, because they can have impairments of different nature. However, in the context addressed by DIRHA we feel that the most important aspect to consider is to involve in the study motor-impaired end-users who have some familiarity or experience with home automation. Depending on their impairment, they can have different experiences with automated services, but their expertise is of fundamental importance in order to understand possible scenarios which may be developed in the future, and eventually discuss on how to meet their needs/desires and the limitations of the technologies under study, which is one of the main challenges of the project.

Since innovative systems, as that conceived under DIRHA, can change the environment and the relationships between people, in this case the requirements gathering represents a critical phase at psychological level. For example, the preference that a subject can express about a particular kind of service/feature, can be affected by the fact that he/she is looking for a way to reach a certain level of autonomy within a particular home context (e.g. in the bathroom). However, it can be difficult, or it can be considered inappropriate, to express this kind of need or desire, both in relation with the researcher and with someone of the family, because of the private nature of the need. For this reason, we decided to use a qualitative approach, instead of a quantitative one, and to use the interview as the most appropriate technique to inquire this field of research. It is also important to highlight that requirements gathering is not a market analysis, and so that it is possible to use a limited number of subjects. Even increasing substantially the number of subjects, the resulting analysis would not have a statistical significance, due to the above-mentioned variety in the specific needs of the given end-users.

3. Research procedure and analysis method

The basic paradigm for a qualitative research is to gather an in-depth understanding of the user behaviour keeping in mind the particular context of research field. Although the sample is generally small it is always focused on a particular topic, and methods such as focus groups are often used to this purpose. In our case, since the subjects had evident problems and could not move to a common place, the focus group method could not be adopted, i.e., we went to each home to interview them separately.

As a matter of fact, this choice was advantageous since in this way we could observe the automated home and services already available at each site, understand how they are used by the end-users, and discuss with them in an easier way about possible application of voice-enabled interfaces in this context.

3.1 Research techniques

Two different techniques were applied to investigate the needs and the problems that mainly affect the subjects that already have an automated service in their home. These techniques are: an interview and a check list to prioritise services.

The **interview** is a "structured interview", which means that it is conducted by following a list of predefined topics (see Annex A), while end-user responses were recorded. The aim of the interview was to identify usability issues that should be the most important for people with motor disabilities and to understand their expectations about the possibility that a vocal interface could improve their everyday life in an automated home. This kind of interview was preferred to that based on a questionnaire, both because we decided to favour a qualitative approach instead of a quantitative survey, and because optional avenues of questioning can be explored depending on answers to earlier questions.

Advantages of the interview

This technique can be used at any stage of system development, depending on the questions that are asked in the survey. Another advantage is that interactive interviews typically bring more information than a questionnaire and can provide a deeper level of detail. Moreover, a structured approach can also gather more reliable, quantifiable data than an open-ended interview, and can be designed rigorously to avoid biases in the line of questioning.

Disadvantages of the interview

Interviews are good for getting subjective reactions, opinions, and insights into how people reason about issues, but they imply the employment of time or a lot of interviewers. Another constraint is that in order to analyse the results it is necessary to use the "content analysis" method, which presents some problems due to the alignment between different languages and cultures.

The aim of the **check list**, is to prioritise the home-automation services functionalities that meet user needs. All the automated services we thought can be based on a voice-enabled interface, were listed as grouped in the following four areas: Communication, Healthcare/security, House management, Leisure/entertainment (see Annex B). Each area included from three to five different services (e.g., within the Healthcare/security area were listed: Management of electronic case-history, Alerting service for emergency situations, and Voice video-camera management). At the end of each interview, the check list was proposed to the subjects, asking them to use two different evaluation criteria:

- give a score to the service listed within each area, using a scale from 1 to 10 (where 1 = useless and 10 = very useful);
- order by priority all services that received a good score in the first judgment (independently of the area they belong to).

In this way we tried to get two different kinds of data: on the one hand, there was an assessment about how much useful can a service be, when compared with other services of the same area; on the other hand, there was an assessment about an absolute priority, depending on the home automation services considered by the subjects as ideal.

Advantages of the check list

The list was composed by all the possible home automation services that can be used by a voice-enabled interface, so it was useful also to suggest new solutions to the people interviewed.

Moreover, the possibility to "vote", giving a judgement within each area, and then among all the areas, consented the users to assess both their personal needs and an ideal prioritisation of home automated services.

Disadvantages of the check list

Since the check list was comprehensive of all the possible services that we thought could be improved by a voice-enabled interface, some of them were not appropriate in relation to the main disability of some subjects, to their age or their habits (e.g. for an adult person without any possibility to do rehabilitation exercises, the item "Video-games or strength-fitness training management/setup" was inappropriate).

3.2 Participants to the qualitative research

Since the motor disabled population is not a homogenous group in terms of physical disabilities, we decided to interview persons with different histories related to their impairments, so that to have a different point of view referred to different disabilities.

To investigate this heterogeneity, especially from the functional point of view, 7 interviews have been conducted with Italian persons (5 women and 2 men, with age ranging from 30 to 60) that have each one important, but different, physical impairments (excluding physical disabilities as visual impairments, hearing impairments, pulmonary disease, respiratory disorders and epilepsy). All these persons are not able to use lower limbs, and a lot of them also have muscular atrophy in the upper limbs.

We selected this small sample among people that were physically disabled for a long time, to avoid involving in an interview persons who have just been traumatized, since they would not have the sufficient serenity to talk about their daily life and discuss on the adaptation of their house to their new needs. Within this sample, we did not distinguish among persons with congenital physical disabilities and those who had an injury after birth or at a later time; moreover, during the interviews we did not ask them what kind of pathology they have. Those who wanted to specify the genesis of their impairment, of course were free to explain it, but we did not consider important to have a medical classification of possible disabilities, because the main focus of the interview was to understand the desires related to the improvement of their daily life in their house and how they imagine that a voice-enabled interface can facilitate their approach to home automated services.

Among these seven persons (most of them are clients of DomoticArea), only one still does not have any automated service in her home, but her purpose is to equip soon her house with these facilities. So, since she does not have any experience about domotics, we discussed with her about her availability to arrange her house in relation to her expectations on house automation[1].

About the familiarity with voice-enabled services, there were different ranks of knowledge: one person knows very well these technologies, since two years ago a voice-enabled command-and-control system has been installed (by FBK and DomoticArea) in her bathroom. Among the other six participants, some of them had experience with telephone voice services (some customer care or timetable train services), but it was in any case an occasional use. So, they can be classified as "naive users".

The persons without any knowledge of voice recognition services were informed about how these services work, with a description of their advantages and disadvantages according to different possible dialogue solutions which can be realized.

[1] We decided to interview this person not only because she is interested to improve her house with home automated services, but also because she is the president of an association that can diffuse information about these technologies to other disabled people. So, her experience during the whole project could also facilitate other people that will ask her for advice about the possble installation of these technologies in their home.

4. Content analysis

Content analysis is a method of analyzing written, verbal or visual communication and can be used in qualitative research. Since we used a "structured interview", the contents were already defined, but of course each person answered in a different manner, explaining her/his motivations. The structured interview was prepared thinking to investigate, among others, the following topics:

- Expectations based on their knowledge of state-of-the-art voice-enabled interfaces;
- How applying voice-enabled interfaces can facilitate using home automated services;
- Possible use of speaker identification, user profiling, different set ups for different rooms, etc.;
- Availability to change location of some furniture and to install microphones on walls and room ceiling;
- Investigation about the availability to have an installation including internet of things;
- Preferred feedback from the system according to the different functionalities and to the room where the service is available; preference on pre-recorded utterances vs text-to-speech when voice feedback is foreseen;
- Set up of default system state;
- Error tolerance in different possible contexts, and possible strategy for error recovery.

4.1 User expectations about voice-enabled interfaces

Background knowledge on speech technologies

It is likely that user expectations will change depending on the experience she/he will progressively acquire using the voice-enabled interface. However, a fundamental aspect of this study is the background on speech technology that the end-user has.

As highlighted in Section 3.2, only one person of our sample did not have any experience about home automated services; her experience with voice-enabled interfaces was referred only to telephones services (in the past, before using internet, she used the train timetable service calling the related number of Italian railways).

The other subjects involved in the interviews, have experience of home automated services, but only two of them had used in the past voice enabled interfaces not related to telephone services (i.e. one subject used a dictation software some years ago, while another person is using a voice interface in the bathroom as already reported in previous sections).

Emerging needs at end-user level

Starting from their daily experience, all the subjects were initially interested in understanding how a voice interface could be integrated with the services they already know and use. For this reason, usually we started our interviews asking them what kind of home automated services they would like, first of all, to control by voice. After that, since they were solicited to imagine that they can use a voice interface to operate also home automated services they currently do not have, we tried to investigate their priorities and expectations asking them to imagine what kind of services and functionalities they would like to have and which of them could be addressed by voice. From this analysis it turns out that:

- the most mentioned feature regards a voice-enabled interface to control open/close doors (mainly the front door), windows and rolling shutters (that usually have already been automated).

- immediately after this service, there is the possibility (important mainly for persons with muscular atrophy in upper limbs) to switch on and off the lights and to manage the entry-phone.
- the possibility to control by voice the room temperature set up, an interphone (to communicate with people that live in the same home, but also in a very near house), an alarm service, in particular for emergency situations, is also relevant to some of them;
- the possibility, just with basic commands, to control leisure devices as HiFi, TV, docking station or multimedia/photograph archives; the latter one was usually mentioned when the subjects were further solicited to imagine what would they like to have.

Other important features, depending on the characteristics of the house where the subjects live or on the experience they already have (living alone or with other persons), are: the possibility to control the lift and the possibility to adjust the tents and the water temperature in the bathroom.

Two persons also mentioned, as very specific feature, the possibility to move down and up the wall cupboards (mainly in the kitchen) to take objects also from the wheelchair, while a woman who uses a particular "hook" to raise fallen objects, asked if it could be possible to have a kind of robot that helps to raise things from the floor.

Another subject, being already accustomed to use a voice interface, suggested also the possibility to create "special commands" to activate specific "scenarios", which would require several interactions, one for each related functionality. Some arrangements related to the regulation of the tents, the adjustment of lights for chromotherapy, the play of some particular kind of music to relax, could be automatically set just using a single command that includes the ensemble of all these dispositions (e.g., when one gets up in the morning, the command "DIRHA, please set the scenario morning" can set up the automated home with a specific configuration jointly involving several devices and actions).

4.2 User expectations about system placement and possible features

Location of system resources

There are no particular constraints about the possibility to place microphones in different rooms. All the subjects are available to move a little the furniture or to install the microphones on the ceiling. The only person who highlighted a problem is the one that already uses a voice interface; she is available to put microphones everywhere, but she would prefer to avoid a wireless installation, due to some doubts about the possible influence of wireless technology on health problems[2]. The preferred rooms where to start with the installation of microphones and the activation of related services are the bedroom and the bathroom, followed by the kitchen and the living room/lounge.

Speaker identification/verification

The possibility to have a feature as speaker identification/verification, for restricted access to the system, is not considered so important. Its relevance depends on the fact that the subjects live alone or with other persons.

User profiling

User profiling is considered more interesting than speaker verification by the subjects who do not live alone. Although they are available to learn carefully ways and specific commands to control each home automated service, they think that voice control could also be useful for their relatives, and that probably their relatives would not be so available to learn specific commands as well as rationale and mechanisms on which the dialogue management is based (see Section 4.2.1). They also think that the

[2] Actually, under DIRHA the use of wireless technology is not foreseen, at least in the real installation at end-user homes.

possible management of multiple users in different rooms addressed by the DIRHA project represents a very important functionality in some contexts. For instance, in the case of an incoming call, the possibility to easily transfer it to a person in another room would solve a typical problem tackled frequently by some of them.

Facilities status check and trigger by remote

Generally speaking, most of the subjects think that is more interesting to have a functionality (like a kind of interphone) that allows one to give commands from a room to another one, without the necessity to move themselves to another place (especially to control by remote the windows, the temperature of different rooms, or the HiFi). Of course, to decide what kind of command is needed to change a facility status it is necessary to know the state of the item, making some queries to the system in order to know how, and how much, to modify it.

4.2.1 Dialogue features and error recovery strategies

Dialogue strategy

Related to user profiling, another feature that persons who do not live alone consider useful is the possibility to run different types of dialogue with the system according to expertise, attitude, etc., of the current user. During the interview, we explained to people who did not have direct experience with voice-enabled interfaces, that there is the possibility to design the dialogue with the system at least in two different ways:

- a dialogue more based on "system initiative", that is more reliable, but also more structured (which also means that the user has to learn how to interact with the system in pre-defined ways);
- a dialogue that is more "user friendly" and flexible (e.g., based on more advanced capabilities of understanding spontaneous speech), allowing the user to speak in a "natural" way, but probably introducing a higher recognition error rate.

Most of the persons interviewed are available to use a structured dialogue and to learn specific keywords to interact with a more reliable system, but those who live with other persons (not impaired but simply elders) highlighted that their relatives can have some difficulties to learn precise keywords. In this case, the possibility to set different profiles was considered very useful because they imagined that, depending on the user profile, the system can react with a different kind of dialogue (i.e. for persons who do not have access to a voice-enabled service for necessity, it is easier to interact with a system that can make some errors because, in case of error, they can reach their goal also manually).

System feedback

As for the feedback provided by the system to the user during the interaction (including actions for error recovery), most of the subjects would prefer a voice feedback instead of a visual one. The visual feedback is considered interesting only when it is necessary to set up the system, or particular services which imply the need to choose among a large number of possible options (e.g., programming the thermostat, or having access to a selection of songs available in the multimedia archive).

Talking about the interaction with services that are used in a "contingent" way, the voice feedback is considered more useful because it is possible to know if the system reacts correctly even when the command is given to modify something in a room different from the one where the person stays. Moreover, it is not possible to install a lighted led (that could be used to give a feedback in rooms where no screen is available) in all the rooms, or to see it from all the slants in the same room. A subject (the most expert one on voice-enabled interfaces) suggested also foreseeing in the dialogue

the possibility to make queries to the system in order to know its state (e.g., having the possibility to make a check on the "wind alarm" state without the need to go in another room to see if the alarm is on).

Since they want to have a "relaxing" voice that talks with them, most of the subjects said they would prefer a pre-recorded voice instead of a TTS; there is no agreement about the kind of voice that should be used, because women would prefer to have a male voice, and men would prefer a female voice.

Error recovery strategies

For the same reason (the dialogue with the system has to be amiable or comfortable), most of the subjects think that an error recovery process should foresee at maximum 2 or 3 attempts to try to recover from an error. After those 2-3 attempts, it would be better to have available a possible command that starts the dialogue from the initial prompt, which is particularly useful if the system is making a "critical" error.

For example, if the system "understands" that the temperature has to be adjusted to 39 Celsius degrees instead of 29, the error is acceptable at user level, since the operation will not be executed immediately (it takes some time to increase a room temperature). and, for this reason, a confirmation will always be required before to perform the related action. If the user does not confirm, after having received the system feedback, it will be possible to correct easily the recognition error, spending a little more time to recover the right interaction within the "thermostat service".

As a counterexample, if the system makes a "critical" error. e.g. the user wants to switch on the light and instead of this the system understands that it has to switch on the HiFi, an immediate recovering of the unwanted operation may be difficult because of the possible sound emitted by the HiFi. In this case, restarting the dialogue from scratch, simply using a command to reset the interaction, is surely easier than trying to interact with the system in order to recover from the given error.

System activation mode

During the interviews we explained that it could be possible to start the interaction with the system in two different ways:

- the system is in "always listening" mode, and it can execute any command that is captured in the environment and understood by the system as a possible request coming from the user; this strategy has the problem that there may be a high probability of false alarms, which can cause misunderstanding at system level (e.g., sometimes the system may interpret as a command a word that the user says while speaking with other people, or a noise event);
- before a short dialogue for the execution of a command, the system has to be activated by uttering a particular keyword (which, in case, will be sufficiently long or strange to avoid possible misunderstanding with other words people normally pronounce while speaking); once detected this keyword, after few seconds the system will be ready to interact with the user.

Most of the subjects prefer the latter, definitely more "safe", strategy. They said it is preferable to have a particular keyword to activate the interaction with the system and most of them said also that they prefer to have some kind of remote control (as already available for their home automated services) to have such a backup solution when the interaction fails for any possible problem.

4.3 Integration of special devices through the Internet of Things approach

Internet of things is a technology that allows one to access virtually any kind of device from within the "internet" cloud (e.g. electric appliances, toys, and others). Within the DIRHA scenario, introducing this approach would bring in the ability to implement solutions that take advantage of devices which are not usually accessible from a computer.

The difference of the Internet of Things technology from other interconnection technologies defined in the past years (e.g. Konnex, Fieldbus, Echelon, X11 and many others) is that devices using it are natively able to interact with any PC or server through the TCP interface; this will mean lower system complexity and lower costs. As a result, devices like special remote controls or simple light feedback systems able to natively interconnect with the rest of the DIRHA system could be implemented easily.

We asked the subjects if they thought that could be useful for them to use this technology, taking into account that the possibility to interact with a remote computer can bring fears related to the fact that some users are afraid to introduce new technologies in their houses. In the end, three people said that they are not in favor of this technology, exhibiting some kind of concern about their own privacy (one subject) and about the risk that telecommunication lines could break from time to time (two subjects that live together). It must be noticed, however, that the above risks will not be there in the scenario that we are envisioning, as the Internet of Things paradigm will be instantiated only for devices within the house, excluding their exposure on the Internet network.

4.4 Check list results

In addition to interviews, we used also a check list to ask the subjects what are the most important services they would mainly use to interact with a voice-enabled interface.

As explained in Section 3.1 we used two kinds of scale to evaluate services with the check list: at the beginning, we asked to assign the scores to assess how useful a service can be within a certain area (e.g. Communication, Healthcare, etc.); afterwards, considering only the services that came out as the most important within each area, we asked the subject to assign them another score to prioritize the home automation services that the subjects imagine as ideal (for this reason, with this second assessment, services evaluated near 1 are considered more useful than the ones with a higher score).

In Table 1, these two different evaluations are represented in two sets of columns; services that collected the best scores (both for importance and for priority) are colored in a gradual scale where the orange color represents the services that received higher scores, while no color represents the lower scores.

Regarding the section related to Absolute priority, since each subject prioritized only those services that were judged with high scores in the "Internal area" assessment, both the average score and the number of votes are reported (or, in case, where there was no occurrences, no average score is shown).

Considering the best scores (both for importance and for priority), as illustrated in the Maslow hierarchy of needs[3], it came out that users' autonomy passes mainly through "basic" facilities as the possibility to easily move in their homes (controlling doors), easily manage the internal light

[3] "A Theory of Human Motivation", A.H. Maslow. Motivation and personality. Harper and Row New York, New York 1954

(controlling rolling shutters and electric lights and sockets), and adjust the temperature (controlling the thermostat).

Once these facilities for the "internal comfort" are managed, naturally users' interest gradually expands itself to a context that includes the possibility to communicate with other people, for example managing the entry-phone and the interphone (to communicate with closer people) or a telephone assistant (to communicate with persons far away and to manage possible emergency situations).

Finally, when physiological, safety and belonging needs are satisfied, cognitive and aesthetic needs are considered, consequently also facilities related to leisure and entertainment become interesting to be controlled by voice (HiFi or TV management).

Internal area scores	AVER	***Absolute priority***	AVER	OCCUR
Communication area		**Communication area**		
Telephone assistant	7,7	Telephone assistant	4,4	5
Voice browsing of favorites websites	5,9	Voice browsing of favorites websites	5	1
Interphone	8,4	Interphone	3,5	6
Electronic notice board	7,0	Electronic notice board	5,5	2
Entry-phone management	8,6	Entry-phone management	2,6	5
Healthcare/security area		**Healthcare/security area**		
Management of electronic case-history	4,4	Management of electronic case-history		
Alerting service	8,3	Alerting service	3	2
Voice video-camera management	6,7	Voice video-camera management	5	1
House management area		**House management area**		
Lights and sockets voice control	9,7	Lights and sockets voice control	2,4	7
Windows and doors voice control	9,7	Windows and doors voice control	1,6	7
Rolling shutter voice control	9,7	Rolling shutter voice control	1,6	7
Rooms temperature set up control	8,7	Rooms temperature set up control	4,4	7
Internet connection voice control	5,1	Internet connection voice control		
Management and set up of different household appliances	6,9	Management and set up of different household appliances	5,0	4
Leisure/entertainment area		**Leisure/entertainment area**		
Hi-fi, docking station management, multimedia/photograph archives	8,3	Hi-fi, docking station management, multimedia/photograph archives	7,3	4
TV (with all advanced services), webTv and VOD voice control	8,6	TV (with all advanced services), webTv and VOD voice control	6,0	6
Newspapers, e-book reading	5,6	Newspapers, e-book reading		
Video-games or strength-fitness training management/setup	4,7	Video-games or strength-fitness training management/setup		

Table 1: The table reports on check list results. Two kinds of scale were adopted to evaluate the given services.

5. Possible application scenarios and related tasks

Based on the analysis of the above interviews and of the answers to the check list, a number of application scenarios can be envisioned. Such scenarios and the related initial proposal for the dialogues will serve as starting point for the design of the real services to be developed within the DIRHA project. In particular, the storyboards that are reported in the following sections can represent a starting point for next activities, which include the definition of experimental tasks, of dialogue strategies, ontologies and concepts to investigate at speech understanding level, hand-crafted grammars, language models and vocabularies to work with at recognition level, etc. The intermediate and final prototypes, as well as the targeted showcases, will surely be designed taking into account the guidelines which come out from this user study.

In order to effectively address the specific scenarios, a set of common assumptions and guidelines is reported in the following (and not reprised again case by case in the following examples of services and related dialogues):

- ***Location of devices and system resources*:** the system shall be usable from all the rooms of the house; as a consequence, microphones and loudspeakers shall be installed in all (or most of) the rooms and interconnected to a centralized platform where the proper processing will take place, both at acoustic level and at dialog level; this processing will be implemented in such a way that multiple concurrent dialogues could take place inside different rooms. On the other hand, visual resources (e.g., screens or TV) will not necessarily be available inside all the rooms; for this reason, the dialogue management component shall be aware of user location (i.e., room) for which it is running, in order to choose whether to use visual prompts or not.
- ***System activation mode***: the system, in always-listening mode, is invoked by a keyword that activates the interaction (e.g., "Dirha Open Sesame").
- ***Dialogue strategy*** will be Mixed Initiative, that is, in case the system needs more data than what gathered so far from the user, it will ask the user to specify it, and in case it will implicitly answer to a previous question while giving a new feedback (see for instance the next dialogue regarding the control of temperature).
- ***Error recovery strategy***: user input will mostly be vocal, with the system making 2 or 3 attempts to recover errors, possibly providing increasingly verbose prompts to linguistically induce the user towards using some known words; should this procedure fail, the command will be ignored, allowing the user to start the interaction from the beginning.
- ***Command cancellation***: should the user need to abort the command just accepted by the system (e.g., in case the system understood a different command than the intended one) he/she will use specific keywords.
- ***Alternate command mode*** a remote control will always be available as a backup, supporting the same basic command set as defined for the vocal dialogues.
- ***Feedback***: the system feedback will be vocal in most of the cases. When possible, a pre-recorded set of prompts will be used instead of TTS, for better voice quality; when the task is simple and not harmful, no confirmation will be requested and the command will be immediately executed; for some more complex commands, visual output will be provided through a display, a TV, or a lighted led.
- ***Facilities status check and trigger by remote***: where meaningful, the system will implicitly assume that dialogs involve objects located within the current room, or anyway close to the

user (e.g. turn on the light, raise the temperature, what is the room temperature?). However, dialogs shall allow, where appropriate, users to talk about objects located in other rooms explicitly naming them at the end of the utterance (e.g., turn on the light of the kitchen; which is the temperature of the living room?). Moreover, if the user wants to change the state of an object that is in an unknown status for the user, he/she could ask to know the status or the system can warn the user that the requested change is not necessary.

- ***End of interaction***: after a command has been executed, or the goal of a given user query has been achieved, the dialogue is stopped (e.g., based on a conclusive sentence as "Thanks Dirha, Bye" or "Dirha, close Sesame", or based on the fact that the user does not speak anymore for a period of a given length) and the system goes back in its idle state, waiting for another command. During the transition a chime is played, letting the user know the command is over; different type of chimes will be handled, according to the result of the interaction (e.g. command executed, command not executed, command not understood, command inconsistent, command cancelled, etc.) .
- ***Speaker identification***: verification of the user identity may reinforce the confidence of the system in order to avoid a confirmation or deny access to specific actions.

5.1 Management of house facilities

This section provides some examples of interaction with the DIRHA system for what concerns the management of house facilities (i.e., control of doors, windows, lights, sockets, temperature). Some related guidelines are reported in the following:

- each item (door, window, light, etc.) will be given a specific name, possibly with synonyms (e.g. "kitchen light"): this name will be used in case the user needs to explicitly refer to an object;
- the system compares the understood command with the current status of the item at hand (thanks to the link with the home automation system) and may either execute it, or inform the user of the inconsistency of the command (e.g. try to open a door that is already open), or finally propose an alternative (e.g. "did you mean to close the door?"), trying to recover the possible error in the input request, as the error could be brought in by the speech recognition and understanding components.

5.1.1 Control of doors, windows and shutters

This service offers direct control of doors, windows and shutters. The characteristics of the related scenario are the following:

- the user requests the opening/closing/locking of a door (or window or shutter).

This interaction example shows the feature that allows checking the facility status:

User Eva: "Dirha Open Sesame,… Eva is speaking"
DIRHA system: [after few seconds] "Tell me, Eva"
User Eva: "Please lock the front door"
DIRHA system: "The front door is already locked"
User Eva: "Ok, good night"
DIRHA system: [sound prompt to let the user understand that the system switched to the idle state]

5.1.2 Control of light and electric sockets

This service offers direct control of lights and electric sockets (possible powering electric appliances). The characteristics of the related scenario are the following:

- the user requests the on/off switching of lights or electric sockets and, when available, the adjustment of particular lights.

This interaction example shows the feature that allows to set the chromotherapy light:

User David: "Dirha Open Sesame,… David is speaking"
DIRHA system: [after few seconds] "Tell me, David"
User David: "Please turn on the chromotherapy lights"
DIRHA system: "Do you want to start with default light or do you prefer to change?"
User David: "The default set up is ok, thanks"
DIRHA system: [executes the command and then produces a corresponding sound prompt]

5.1.3 Control of temperature

This service offers direct control of temperature in different rooms. The characteristics of the related scenario are the following:

- "contextual" change: it is possible to control the thermostats located in different rooms specifying where a change of temperature (in the whole house or only in one room) is needed for a limited time span; in this case, only a voice feedback is provided and the system requests a confirmation before executing the command;
- temperature set up: it is possible to set up the thermostats for a long time, also scheduling the temperature level for different periods; in this case, it is necessary to provide a visual feedback and the command of confirmation can be given by voice, but supported also by a graphic interface (i.e. computer or TV remote control).

This interaction example shows that the system initially understood a wrong command, which is cancelled by the user; furthermore, we have two users accessing the system from different rooms. Due to the fact that two users are involved in the interaction from different home places, and that in such a case several technologies are involved (e.g., acoustic echo cancellation, concurrent dialogue management, possible feedback through screens, etc.), this interaction represents one of the most complicated examples of dialogue which will be considered under DIRHA.

User Eva: "Dirha Open Sesame,… Eva is speaking" [in the bedroom]
DIRHA system: [after few seconds] "Tell me, Eva"
User Eva: "Please increase the bathroom temperature"
DIRHA system: "Now bathroom temperature is set on 20 degrees, how much I have to increase it?"
User Eva: "Please increase it by three degrees"
DIRHA system: "I'm starting increase by thirty degrees in the bathroom"
User Eva: "Dirha, cancel; increase the temperature by three degrees"
DIRHA system: "OK, I will increase the temperature by three degrees in the bathroom. Please confirm."
User Eva: "Dirha, now it is correct. Thanks."
User David: "Dirha Open Sesame,… Please set tomorrow's temperature" [in the kitchen]

DIRHA system: "Hi David. Note that also Eva is changing the temperature in the house. Do you want to set tomorrow temperature for the kitchen or for the whole home?"
User David: "The whole home"
DIRHA system: "I'll give you a feedback through the TV. Please go in the living room to watch the default state for each room"
User David: [in the living room] "Dirha, from 9am to 5pm please reduce the temperature to 18 degrees in all the rooms"
DIRHA system: [shows changes by the thermostat graphic] "18 degrees from 9am to 5pm. Is that all?"
User David: "Yes, thanks. Bye bye"
DIRHA system: [sound prompt of system switching to idle state]

5.2 Management of communications

This service offers the possibility to manage the incoming and outcoming calls (both calls addressed to other telephones and calls addressed to other parts of the house or linked apartments); the characteristics of the related scenario are the following:

- outcoming "calls": to talk with any addressee (no matter if in the apartment near the user or in another part of the country) the users have to specify what kind of service they want to activate (the interphone or the telephone) and specify a name only if the telephone service is activated;
- incoming "calls": the system activates the speakers in the room where someone accepts to answer the "call" or to the door bell; if the person who took the call is not the addressee of the call, it will be possible to transfer the call to another room, specifying where.

5.2.1 Control of interphone

This service offers the possibility to "call" persons in another part of the same house, or in a house near the one where the user is. The characteristics of the related scenario are the following:

- the interphone has to be activated always specifying where are the addressees of the "call" (another room in the same house, or another apartment).

This interaction example shows a "call" activated to talk with people in another apartment:

User David: "Dirha Open Sesame,… David is speaking"
DIRHA system: [after few seconds] "Tell me, David"
User David: "Activate interphone for <parents> apartment"
DIRHA system: "Interphone activated in <parents> apartment. You can speak"
User David: "Hi mom, are you there?...."
` [after a while] "Dirha, please close the connection. Close Sesame"
DIRHA system: [sound prompt of switching to idle state]

5.2.2 Answer the entry-phone

This service offers the possibility to answer the entry-phone. The characteristics of the related scenario are the following:

- the entry-phone is managed as an incoming call: the system activates the speakers in the room where someone accepts to answer;
- if some camera is connected to it, the image will be displayed by the available screens.

This example shows the interaction with a "simple" entry-phone (without any camera):

DIRHA system: [the entry-phone is ringing]
User Eva: "Answer the entry-phone"
DIRHA system: "Entry-phone on. You can speak"
…

5.2.3 Control of telephone

This service offers the possibility to answer the telephone, for emergency calls, incoming calls, normal calls. The characteristics of the related scenario are the following:

- outcoming calls: the user has to activate the telephone service and then pronounces the number or the name of the person to call.
- incoming calls: the system activates the speakers in the room where someone accepts to answer the call; it is possible to transfer the call to another room.
- if the telephone line is directly linked to a mobile phone, the incoming sms and e-mails can be read by TTS.

This interaction example shows an incoming call that is transferred to another addressee in another room:

DIRHA system: [the telephone is ringing]
User David: "Dirha answer the call… Hello" [in the bedroom]
Incoming call Speaker: "Hello, it's Jack. May I speak with Eva?"
User David: "Just one moment please". "Open Sesame, transfer the call to the kitchen"
DIRHA system: [In the bedroom: sound prompt of switching the call]
[in the kitchen] "There is an incoming call"
User Eva: "Thanks Dirha, I take it!" "Hi…."

5.3 Management of leisure facilities

This scenario allows to control the HiFi audio and TV systems. The characteristics are the following:

- authentication of the user is required only if sensitive data is involved.
- the system knows the position of the user, in case the user wants to activate a device located in another room the system may request a confirmation to be sure.
- it is possible to adjust the speakers volume and to change music piece or channel.

5.3.1 Management of HiFi audio system

This service offers the possibility to manage the HiFi system. The characteristics are the following:

- if the user wants to listen to the music in a different room where he/she is, the room has to be specified;
- although the system may know every single song of every album that is in the archive, recognition performance may be unsatisfactory if all the titles are put together in a single grammar. A refinement based on single steps, addressing first artist name and then album name, can allow to create small grammars with few titles; this approach would improve recognition rate.

This interaction example shows the choice of a piece and its playing in a different room:

User Luca: "Dirha, Open Sesame… it's Luca speaking" [from the kitchen]
DIRHA System: "Hi Luca"
User Luca: "Activate the HiFi and let me listen Megadeth in the living room"
DIRHA System: "Do you want the music in the living room?"
User Luca: "Yes"
DIRHA System: "Which album of Megadeth do you want?"
User Luca: "The one containing Holy Wars / The Rust in Peace CD"

Music playing

DIRHA System: [sound to return to idle state]

5.3.2 Management of TV system

This service offers the possibility to manage the TV system. At this moment, only basic commands are foreseen. The characteristics are the following:

- it is possible to adjust the volume and to change channel.
- authentication of the user is required only if sensitive data is involved.

This interaction example shows the activation of the TV and the choice of a channel with authentication.

User Luca [in the living room]: Dirha, Open Sesame … turn on the TV.
DIRHA System: "Which channel do you want?"
User Luca: "CNBC please."

TV on

DIRHA System: [sound to return to idle state]

alternate version:
…
User Luca: "Playboy channel"
DIRHA System: "Sorry, but I did not recognize your voice / give me the password"
User Luca: "It's Luca"

TV on

DIRHA System: [sound to return to idle state]

6. Identified functionalities vs system architecture

This section aims to analyze the relationship between the above-outlined functionalities and the DIRHA prototype architecture currently under analysis, and to examine the consequent possible impact on the whole system design. At this moment, the prototype architecture still represents an

open issue, which will be influenced by several aspects as for instance performance, robustness, and complexity of the technologies that will be tested in the next phase of the project. Hence, the following discussion just represents a starting point to progress towards the system design, which will be finalized during the first half of the project and then will determine the real features of the intermediate prototype.

A preliminary block diagram of the targeted prototype was reported in the Annex I – DoW, for what regards the processing steps that range from the microphones to the spoken dialogue management component. Figure 1 shows a similar block diagram, which in this case differs from the original one for what regards home automation, here highlighted by a single block also including audio output generation. It is worth noting that the latter process might be a topic of further investigation, since the acoustic echo cancellation component highly depends on quality of reference signals as well as on synchronization with other system components.

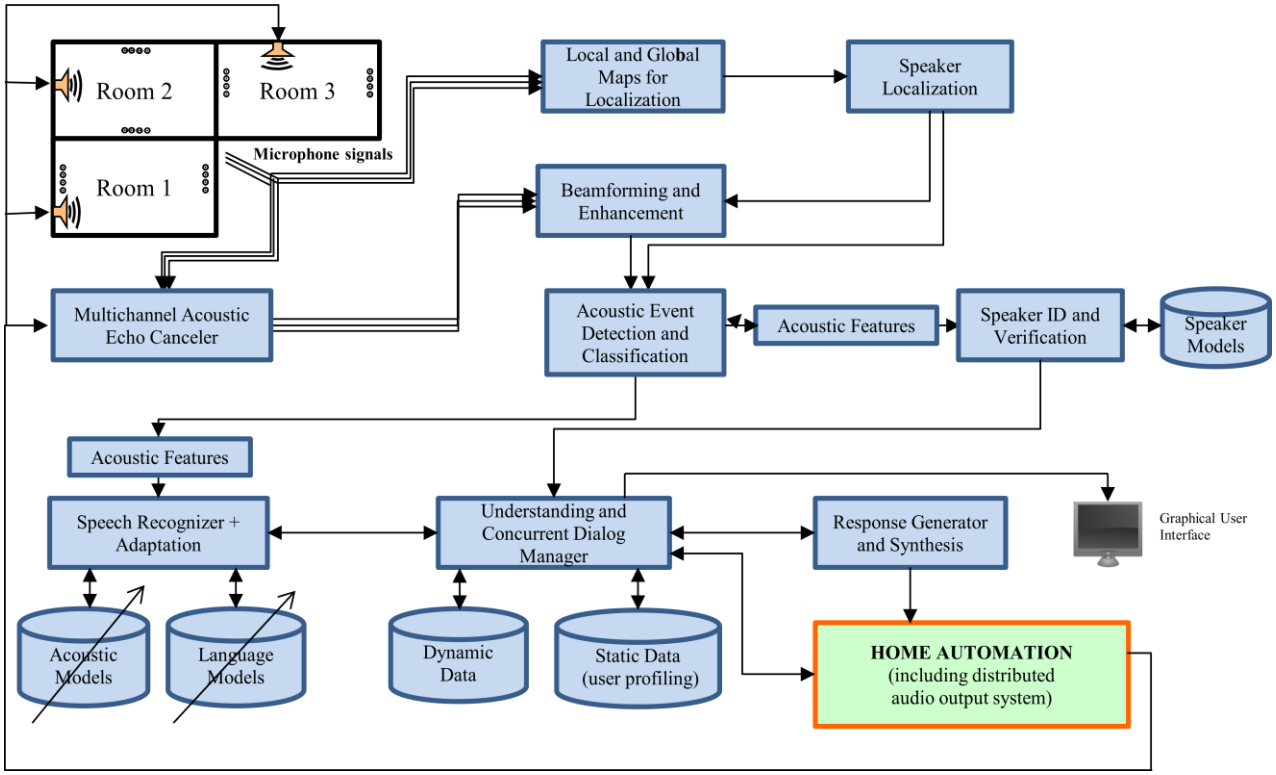

Figure 1: Possible block diagram of the targeted DIRHA prototype. The diagram highlights the components of the distributed microphone network based interaction system, while the home automation part is represented by the block in the right lower part of the scheme.

As the block diagram of Figure 1 does not provide any insights on the home automation part of the targeted system, the complementary Figure 2 is here introduced to describe a possible network comprising a central unit, home devices, and communication among them. Based on this framework, at the moment we assume that the software regarding the microphone network based interaction system runs on a PC, while the home automation is managed by another dedicated PC and related hardware[4].

[4] This is a simplifying assumption, due to the purpose of this document. Actually, during the first half of the project, specific activities of WP6 will be devoted to finalize the prototype architecture, and to establish the number of PCs which

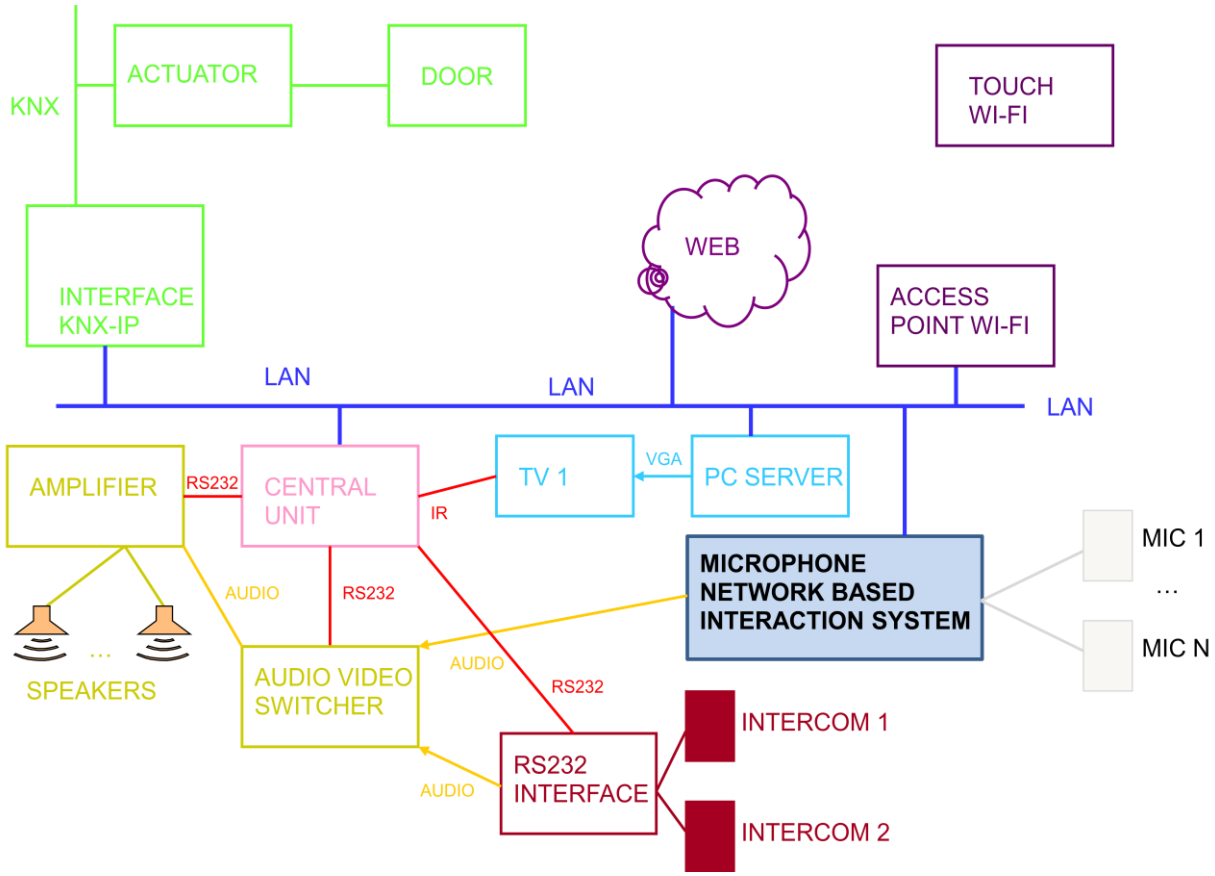

Figure 2: Possible architecture of the home automation system of the targeted DIRHA prototype. For the sake of simplicity, the diagram does not show all the devices (e.g., windows, shutters, lights, etc.). The processing steps ranging from microphone network processing to concurrent dialogue management, and response generation, are represented by the block "Microphone network based interaction system". It is worth noting that in this provisional scheme the two sub-systems share the same LAN, and that audio output is delivered by the home automation part of the whole system.

Note also that the given block diagram has been inspired by the facilities that will be available inside the ITEA apartment where the first prototype will be realized.

The next paragraphs address two scenarios and related dialogues introduced in Section 5, with the aim of highlighting some mechanisms which will be involved in order to perform each step of the interaction with the user. The purpose is just to sketch a possible relationship between the above-defined functionalities and the system architecture; at this moment, we can not have any ambition to make any exhaustive and fully realistic analysis, as we are in a very preliminary phase of the project, with several aspects still to analyze at different levels.

will be necessary, according to the characteristics of the DIRHA system and the real-time response requirements. If feasible, the target is to have a single PC on which the whole DIRHA system can be implemented.

6.1.1 Control of doors, windows and shutters

The dialogue presented in Section 5.1.1 represents a quite simple example of possible interaction with the DIRHA system. The goal is just to open a door, in the same way as one could interact to make a similar action on a window or a shutter. For this type of action, the spoken dialogue management component needs to share with the home automation system the status of each device. The most likely way to implement it will be to have an updated status of each door, window, shutter (i.e., open, closed, partially closed, etc.) always available at dialogue management level. This will happen thanks to the fact that when a device changes its status it promptly informs the central unit about this change; the same information is consequently provided instantaneously to the spoken dialogue management component, where the "intelligence" and "decision making' mechanisms are implemented.

Once understood by the system that someone is speaking[5], a possible process to run is speaker verification in order to establish the likelihood that the active speaker is Eva. Once verified it, and detected that Eva is asking to lock the front door[6], the system has to react according to the status of that door:

- If it is locked (as reported in the given example), the system has just to inform the user of this fact.
- On the other hand, if the door were unlocked, it should be locked by the home automation system. In the latter case, the door status will eventually be updated, and the system will inform the user of the fact that the desired action has been accomplished. As for the mechanisms implemented at home automation level, in this case the central unit interacts via KNX-IP interface with the actuator that performs the corresponding action.

It is also worth noting that at the beginning of the interaction, once the system has detected that Eva needs to interact with it, the central unit activates, via RS232, the loudspeaker available in the room where Eva is [7], which means that the audio output providing answers to Eva will not be diffused in all the rooms.

Once completed the interaction, based on the final message "Ok, good night", the spoken dialogue management may conclude the session with a corresponding acoustic prompt; consequently, the formerly activated loudspeaker will return to a stand-by mode, and the system will switch again to a mode which corresponds to "waiting for a new interaction".

6.1.2 Control of interphone

As second example, we address the scenario described in Section 5.2.1, which regards the activation of the interphone to communicate with the apartment of David's parents.

As in the previous case, there is a preliminary phase during which the system detects the keyword that starts the interaction, and in case verifies the identity of the speaker. The following request formulated by David, i.e., to activate the interphone, triggers a certain set of operations at central unit level (which are not addressed here in detail because they are out of the scope of this document): the final target is that the home automation system establishes a connection between the room where

[5] The system, in fact, is initially in a stand-by always-listening mode, and it changes mode based on a successful spotting of the triggering keyword "DIRHA Open Sesame…".

[6] This input utterance involves all the processing chain from the multi-microphone front-end to the understanding component, since the system must comprehend the nature of the request and the targeted device (i.e. the front door).

[7] With this regard, we expect that the speaker localization component will provide the spoken dialogue management component with this detail. In general, the same evidence may also be confirmed by the status of other devices, as presence sensors, when available.

David is and the room where his parents are; then, it switches on the respective loudspeakers, and finally it informs David that the audio connection is now on (i.e., "Interphone activated … You can speak"). During the following conversation, the spoken dialogue management system will be in a "waiting" mode, from which it will be waked up once a command as "DIRHA, please close the connection…" will be detected. After it, the home automation system will perform the necessary operations (at central unit level, with communication via RS232 and ethernet to various devices and system components) in order to switch off the devices which were involved in the communication, and then it will quit the session.

7. Discussion and conclusions

The purpose of this document was to report on the results of the user study conducted during the first four months of the project, to analyse the resulting user requirements, to outline functionalities and related application scenarios, and to provide a preliminary analysis on their possible impact on the architecture of the DIRHA prototype.

According to Annex I – Description of Work (DoW), the aim of requirements gathering was to obtain guidelines to design the dialogue and prototype in the most proper way, and to characterize a possible set-up at each user site. To this purpose, two techniques were chosen to conduct a qualitative research (interview and check list), aiming to achieve the following goals:

1) **which** real end-users are interested to participate to the project, and what is their flexibility for possible changes in their homes due to the installation of a multi-microphone based distant-speech interface;
2) **what** domotic services/functionalities are useful to these end-users, and what priority they have to facilitate their daily life;
3) **how** the end-users imagine a dialogue with the system could be organized, in order to facilitate home automated service management.

1) Regarding the first point, we observed that the motivation of all the subjects is very high. They are available not only to contribute to the project, testing the DIRHA technology to understand how available home automation services would take benefit of it, but also to learn the best way to use distant-speech interfaces. Moreover, they do not mind about microphone installation, accepting possible movement in the furniture, if this allows improved system performance.

2) Regarding the second goal, since in a UCD approach the design is focused on the users' needs, we had to understand what kind of facilities can meet their goals and increase their satisfaction. This is particularly important for end users as the ones involved in DIRHA project, because the possibility to manage the home automated services by a voice-enabled interface can give them more autonomy, restoring the possibility to cope with a lot of daily problems and improving their self esteem.

Even if the sample was very small, so that it can not have any statistical significance, we could interview persons with different kinds of motor impairment and family structures, and in different contexts (e.g. house or apartment, country or town). This gave us the opportunity to understand different requirements based on a good range of user typologies (e.g. persons that are able to use upper limbs vs persons that do not, persons that mainly live alone vs persons that live with many other relatives, young people vs elder people, etc...).

Despite all these differences, and independently from the diverse motor impairments that the people interviewed show, it came out that users' priority can be represented following the Maslow's "pyramid of needs", where the services/facilities considered most useful are those satisfying an "internal comfort" consisting in some "basic" home functions. Once these facilities are managed,

users' interest gradually expands itself to a context that includes the possibility to communicate with the outside and, at the end, to more complex services/facilities related to leisure and entertainment.

This prioritized list of interests can also determine a scheduling in the application of DIRHA technologies to the targeted home automated services. In other words, since control of doors, windows, lights, and rolling shutters, was considered as the most useful service, the design and development of the dialogue management will address it first.

3) Regarding the third aim of requirements gathering, during the interviews some subjects provided specific indication on dialogue features related to: system robustness, different system's feedbacks, error recovery strategies and different facilities related to the default system state, speaker localization, identification and related user profiling. This evidence depends on their daily experience with home automation services, although most of them are not experts in voice-enabled interfaces. For instance, a voice feedback is preferred to a visual one, since it can be difficult to pay attention to the latter one if this requires to change position/posture.

Another important factor to consider designing the dialogue is the users' motivation: they are available to try to recover from errors multiple times, but most of them stated they are also available to learn specific keywords to interact with a more robust system. For the same reason, since system robustness is deemed important, the dialogue should follow a prudent strategy, for example activating the recognition system with a specific long statement (to avoid false triggering), using explicit/implicit confirmations depending on the complexity of the service, implementing voice command to abort an operation, and supporting an alternate command modality (e.g., touch screen or remote control).

In any case the envisaged functionalities do not represent major problems in terms of device handling: the technical challenges are mainly related to the proper design and implementation of the multi-channel acoustic front-end and the concurrent dialogue management.

Based on the study so far conducted, some critical open issues remain, as summarized in the following:

- limited number of users - voice-operated features are strictly dependent on the characteristics of the house where the subjects live or the presence in the house of relatives/carers;
- limited scope - as expected, desired features are related to "basic" functionalities that cannot be easily migrated to a generic user (with no motor impairments);
- strong motivation of the involved users – this may reduce the chance to obtain a fair subjective evaluation of the DIRHA system.

Nevertheless, the availability of motor-impaired people as real end-users for DIRHA represents an extraordinary opportunity to develop a concrete system able to solve real needs and facilitate independent life in the home environment. Due to the specificity of the emerging requests, however, an additional effort will be required to investigate further extension of system capabilities to general-purpose functionalities for a wider spectrum of users.

As for the next steps of the project, this document for sure can represent an important reference for the definition of tasks to manage at spoken dialogue management level, and for concepts, ontologies, grammars, etc. to adopt at recognition and understanding level. It will then be relevant in the design of the mechanisms with which the dialogue management component will operate connected to the home automation system and to the other system components. The document also represents a starting point for other activities, which comprise the creation of experimental tasks, the possible collection of Wizard of Oz sessions, etc. It can finally provide a relevant contribution to the forthcoming definition of evaluation criteria and metrics to apply in the analysis of prototype behaviour at user level.

It is foreseen that some contents of this document will be revised and updated during the first half of the second year, based on the progress of the project towards the definition of the real functionalities which will be implemented in the intermediate prototype. During the first months of the second year, a new set of interviews with the end-users is foreseen. In particular, this analysis will be discussed in Deliverable D1.3 which will also deal with the design of the user interface.

Annex A - Interview guidelines

- General explanation of the aim of the project and of the interview.
- How applying voice-enabled interfaces can facilitate using home automated services.
- Expectations based on their knowledge of state-of-the-art voice-enabled interfaces and request of examples of most useful/comfortable services (in relation to their disability) they would like being automated first.
- Possible use of speaker identification, user profiling (depending on the age/disability of the user), different set ups for different rooms, etc.
- Availability to change location of some furniture and to install microphones on walls and room ceiling.
- Investigation about the availability to have an installation including internet of things.
- Preferred feedback from the system according to the different functionalities and to the room where the service is available; preference on pre-recorded utterances vs text-to-speech when voice feedback is foreseen.
- Set up of default system state (always listening vs activation by a particular sound or pressing a button) and contextual trigger or long term set up of different services.
- Error tolerance in different possible contexts, and possible strategy for error recovery and for silent alarm service activation.

Annex B - Voice automated services check list

In the following tables you will find a list of different services that can be automated and controlled by vocal commands. They have been grouped by different areas and they need to be ordered depending on the importance that they have in your opinion. In order for us to understand the services that must be controlled by vocal commands, in the right column (Internal vote), we ask you to give a score, using a scale from 1 to 10 (where 1 = useless and 10 = very useful). In this case you would judge each service within the area where it has been listed, you do not need to give an order but just a score (different services can also have the same score). After having assigned these scores, in the left column, we ask you to write, only for those services that have a high score, a number from 1 to 5 to prioritize those services that you already judged as the most important (Absolute priority: where 1=higher priority and 5=lower priority).

Absolute priority	**Communication area**	Internal vote
	Telephone assistant to make a call/send messages (sms or voice e-mails), or to redirect an incoming call to different addressees in different rooms (transferring calls from speakerphone to the receiver)	
	Voice browsing of favorites website through TTS reading	
	Interphone through different rooms	
	Electronic notice board to annotate and remind appointments and deadlines	
	Entry-phone management	

Absolute priority	**Healthcare/security area**	Internal vote
	Management of electronic case-history to monitor drugs consumption and send health parameters to the family doctor (blood pressure, calorie-counting, etc.)	
	Alerting service for emergency situations	
	Voice video-camera management (switching-on or moving the camera toward a specific direction)	

Absolute priority	**House management area**	Internal vote
	Lights and sockets voice control (also for chromotherapy)	
	Windows and doors voice control	
	Rolling shutter voice control (for windows and French-doors)	
	Rooms temperature set up control by voice	

	Internet connection voice control	
	Management and set up of different household appliances like: oven, washing machine, lawn sprinkler, mower, etc.	

Absolute priority	**Leisure/entertainment area**	Internal vote
	Management and set up of different devices as: HiFi, docking station, multimedia/photograph archives	
	TV (with all advanced services), webTV, and VOD voice control	
	Newspapers, e-book reading	
	Video-games or strength-fitness training management/setup	

If you have other suggestions, please write them here and give them a vote:

.....

.....

.....

.....

Annex C - Preliminary analysis of requirements for the end-user available in Graz

The patient/end-user:
On the 4th of April 2012 we visited the apartment of a possible end-user suffering from the myopathy syndrome: a muscular disorder in advanced stage (level of care in Austria: 7). The mid-twenties young man, who passed the Austrian Higher School Certificate (ger.: Abitur) and who works 4-8 hours per week in a business outside his apartment, is confined to a wheelchair and is not able to move hands and fingers sufficiently in order to, e.g., hold a glass of water, push a light switch or any switches or buttons of the same or bigger size. Furthermore, he is not able to enter the bed, to wash himself, or to go to the toilette all alone. By day, a social worker or attendant supports him in activities of daily living. In the evening, his mother takes care of him. One target of the DIRHA-project is to increase his sphere of personal privacy and independence of other people.

The Apartment:
Bedroom: The patient spends most of the time in this room (see Figures 3-4).
* automated nursing/care bed controlled by a remote control to move the bedhead
  + bed is always at the same place
  + solution: control bed with voice command, e.g., move bedhead up, down, stop, etc.
* television
  + television is always at the same place
  + solution: turn on/off the television and switch programs with voice command
* telephone
  + now: phone calls by using a mobile phone; problem: low life of battery;
  + phone calls via Skype by using his laptop placed on the work desk
    - using Skype while lying in bed (!), e.g., in case of an emergency if no one is at home in the morning or evening
  + solution: voice-operated dialing

Living Room:
* television and sound system
  + four different remote controls (too many)
  + solution: turn on/off the television and switch programs with voice command

Bathroom:
* portable mobile lifting aid for shower, bath, and toilette (see Figure 5)
  + operated by the attendant using a remote control
  + attendant needs both hands for patient, thus, hands-free operation required (!)
  + also used during travels
  + solution: voice-operated lifting aid

Front and Vestibule Door:
* surveillance via camera and laptop or television
* opening/closing both doors automatically on command

General problems with the handling of ...
– bed: control of the bedhead-movement (!)
– front and draughtproof door opening/closing/locking/surveillance via camera (!)
– television and sound system
– light switch
– heating system: turn on/off
– lifting aid, e.g., for toilette, used by the social worker or mother (3rd hand required)
– emergency button, better: emergency call via DSR system

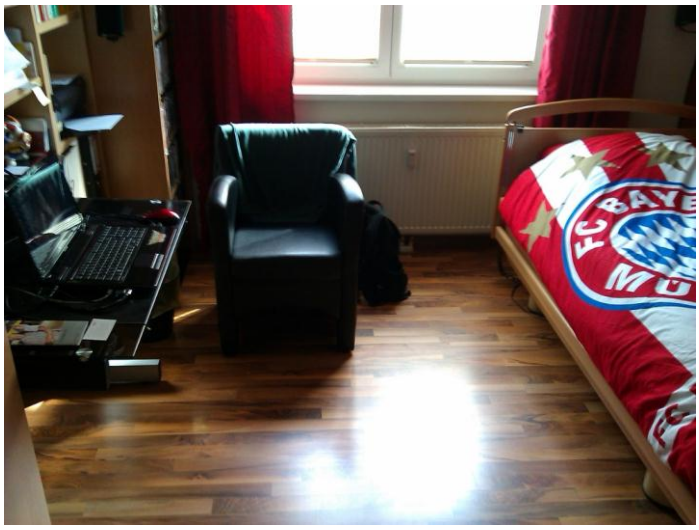

Figure 3: Bedroom of the end-user. The laptop is placed on the work desk located on the right side of the bed as seen from the perspective of the lying patient. The distance between the bedhead and the laptop is approximately two meters.

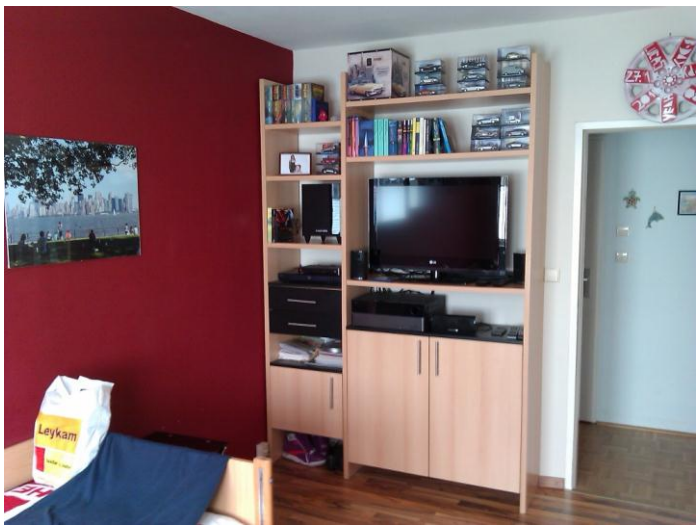

Figure 4: Bedroom of the end-user. The television is located in front of the bed.

Figure 5: Bathroom of the end-user. This picture shows the lifting aid operated by the attendant by using the green remote control. The patient is moved on a swing made out of robust bandages fixed on the front of the lifter.