Output Matters! Adaptable Multimodal Output for New Telerehabilitation Services for the Elderly

António Teixeira[1], Carlos Pereira[1], Miguel Oliveira e Silva[1]
Joaquim Alvarelhão[2], António Neves[1] and Osvaldo Pacheco[1]

[1]Dep. Electronics Telecommunications and Informatics, IEETA
University of Aveiro, 3810-193 Aveiro, Portugal
[2]Health School, University of Aveiro, 3810-193 Aveiro, Portugal

Abstract. Technologies capable of being used by the older adults, with their specificities, to provide new services in the areas of telehealth and Ambient Assisted Living are needed. In this paper, we describe a new service, and its first prototype implementation, in the area of elderly health support at home. Special care was taken to improve the usability, adaptability to user and context and inclusiveness capabilities of the output. The basis for intelligent adaptation of the multimodal output – called AdaptO – is proposed and first versions of the needed services and agents were created.

1 Introduction

The continued introduction of technology in the domestic environment is a reality. Increasingly, the majority of the population sees technology as something natural and very useful.

Within the older adults group - getting bigger and bigger by each year and representing a significant percentage of the population at least in developed countries - this technology can have a positive effect on their living conditions, particularly for those staying at home, if these technologies for domestic environments are made accessible and usable. One of the most common problems with elderly people is mobility. The need to go to a Health facility (ex: clinic) to make a physiotherapy session or to be monitored for an extended period are factors that have repercussions in the daily life of these people and of their families.

With suitable natural interfaces and the possibilities offered by next generation networks, the introduction of technological solutions can facilitate the daily life of the elderly, fighting isolation and exclusion, increasing their pro-activity, work capacity and autonomy. Examples of services are multimedia information access and exchange of personal data; **telehealth** and automatic medication delivery; support of daily activities and community, social and civic life; and automatic management of the environment to improve both the quality of life and security.

Technologies usable by the older adults, with their specificities, to provide new services in the areas of telehealth and Ambient Assisted Living (AAL) are needed. In this paper, we present work in progress in such a new service, for telerehabilitation or

Teixeira A., Pereira C., Oliveira e Silva M., Alvarelhão J., Neves A. and Pacheco O..
Output Matters! Adaptable Multimodal Output for New Telerehabilitation Services for the Elderly .
DOI: 10.5220/0003308000230035
In *Proceedings of the 1st International Living Usability Lab Workshop on AAL Latest Solutions, Trends and Applications* (AAL-2011), pages 23-35
ISBN: 978-989-8425-39-3

even telefitness.

This work is part of the Living Usability Lab for Next Generation Networks project (www.livinglab.pt), LUL for short, a Portuguese industry-academia collaborative R&D project, active in the field of live usability testing, focusing on the development of technologies and services to support healthy, productive and active citizens. The project adopts the principles of universal design and natural user interfaces (speech, gesture) making use of the benefits of next generation networks and distributed computing [1].

1.1 Telerehabilitation

Rehabilitation, Training and Assistive Technologies, have been one of the major concepts in the elderly care sector [2]. From the many ways to define rehabilitation we adopt the interesting definition used by AAL Forum 2010 Track R5 (Rehabilitation, Training, Assistive Technologies)[2]:

> *"You must again be the director of your own life! You must train and rehearse to again be able to help yourself and gain independent living*".

As an example, a typical exercise program for frail elders in seated position [3] includes four components: Warm-up, Aerobic training, resistance training and cool-down. Warm-up includes exercises like postural awareness, breathing, joint range of motion, stretching; Resistance training comprises Body-weight and resistance exercises; Cool-down includes stretching and relaxation.

Telerehabilitation has some advantages [4]. Particularly relevant for the older adults are: availability of therapists, rehabilitation at home, reduced therapist cost, reduced isolation. The disadvantages, also according to [4], are: equipment cost, network bandwidth, technical expertise, safety at home, sterilization for redeployment, efficacy studies, psychological factors. The next generation networks will certainly decrease the problems of network bandwidth; the technical expertise can be addressed by easy-to-use interaction; costs will certainly drop as the number of users increase. It is reasonable to expect that some older adults may exercise less without direct therapist intervention, since they feel they get less attention than they deserve. It is also reasonable to expect that others will accept perfectly less human contact [4].

According to [5] "the one-to-one paradigm of therapy will also change, with one therapist performing "multiplexed" telerehabilitation. This is expected to reduce treatment costs while also increasing access to therapeutic care worldwide". New technologies can help provide therapy anywhere, anytime, addressing current limitation due to location, lack of transportation or limited therapist availability [5]. [5] predicts that cloud computing will be extended to cloud rehabilitation, where the library of disability-specific software simulations or games will reside on a third-party "cloud"; the clinicians will log on to set up exercises programs, follow up progress, insure compliance and monitor safety.

1.2 Vision and Hearing in Older Adults

Age-related eye problems are a major cause of vision loss or distortion in people over

40. The risk for developing serious visions problems increases, as you get older.

Presbycusis [6, 7] is age-related hearing loss. It becomes more common in people as they get older. People with this kind of hearing loss may have a hard time hearing what others say or may be unable to stand loud sounds. The decline is slow; it can develop at different rates. The degree of hearing loss varies from person to person.

Another common complaint by older adults regarding hearing is their **difficulty in understanding rapid speech**. Studies have demonstrated older adults' special difficulty in encoding rapid speech due to age-related changes in peripheral and/or central hearing abilities [8].

1.3 Multimodal Adaptive Interaction with the Elderly

According to [9] the current migration from WIMP interfaces towards multimodal interaction has benefits for seniors. As an example, speech can accompany a text signal in a warning message box. In this way different capabilities, expertise, preferences or expectations of the seniors can be accommodated.

The diversity of environments, systems and user profiles leads to a contextualisation of the interaction. Initially the interaction had to be adapted to a given application and for a specific interaction context. Nowadays, the interaction has to be adapted to different situations and to a context in constant evolution [10].

Environment and context of use can have high impact in an application or service use. For example, when the user is not looking towards a monitor on whom textual information is presented this can result in a critical failure. Examples of environment relevant alterations are the increase of ambient noise level and the changes of illumination. It is essential to determine these variations in context and adapt the system interaction inputs and outputs accordingly.

Another factor often neglected is the user himself/herself. For example, in a scenario in which the hearing ability of a user is reduced it makes sense to provide output for other senses, such as vision, thus increasing the likelihood that system messages are received and perceived by the user. Without the redundant or alternative use of vision, the interaction between the application and the user has a strong possibility of failure.

Efficient multimodal interfaces should be able to take into account the user requirements and needs. This is particularly relevant for a group such as the older adults, with some potential additional difficulties.

2 Objective: A New Generation of Remote Rehabilitation Services for the Elderly

Combining the needs of the elderly to have professionally monitored exercises without leaving their homes with other factors, such as the availability and costs of qualified health professionals, a Telerehabilitation Service was considered as one of the priority new services to develop and test in the scenario of our Living Usability Lab. The service is based, but not equal, to the rehabilitation service proposed by project Persona [11]. In very general terms, the service should allow supervised remote exercise sessions at home or community centres, as a mean to maintaining health and prevent illness. Table 1 presents essential information on the Service.

Table 1. Service description, based on the Persona Remote Rehabilitation Service [11].

Name	LUL Telerehabilitation Service with Multimodal Interaction
Description (What)	Remote and supervised exercise sessions at home or community centres, for maintaining health and prevent illness. Sessions carried out concurrently at several sites via networked multimodal applications. It should be possible for the user to wear biosensors while doing the exercises/rehabilitation. Eventual use of vibration or similar to give feedback (ex: to indicate wrong/right execution of the exercise). A health professional supervises everything from the training centre/hospital, including the biosensors signals captured remotely and the (multiple) cameras images. The system also includes mechanisms to request and process information regarding effort level from the user.
End user	People taking rehabilitation sessions at home and people wanting to do exercise in case it is just training sessions.
Stackeholders and Roles (Who)	Health/Sport professionals (physiotherapists, gerontologists, etc) who configures the sessions and directs them. The training centre or other health services provider, which should install and maintain the platform.
Technological description (How)	The main user interface is a large size computer monitor (acting as a large size TV) combined with speakers and video cameras. In addition, it should be possible to use a set of biosensors. Sensors gather the vital signals from the patient and send them to the health professional.

Table 2. Analysis of our system according to AMITUDE properties.

Application Type	Home rehabilitation system for elderly people. Physician supervision. Agent based system with context-enabled functionalities.
Modalities	*Input, health professional:* 2D analogue haptics pointing, 2D static Portuguese typed text, Portuguese spoken conversation. *Output, health professional:* Portuguese spoken conversation, Video imaging, 2D static Portuguese typed text.
	Input, elderly user: 2D static Portuguese typed text, Portuguese spoken conversation, Video Streaming *Output, elderly user:* 2D static Portuguese typed text, Portuguese spoken conversation, 3D dynamic graphics describing exercises and involved routines.
Interaction	Two-way communication (deliberate). Computer-mediated human-human communication.
User task, other activity, domain	*Task goal:* execute a telerehabilitation session *Generic task:* complete a session of rehabilitation exercises according to a defined program, with direct supervision from a professional (ex: physician) *Domain:* e-Health *Issues:* Context-aware environment, customization
User (Personas)	Elderly user. See the description of the Personas in next section.
Devices	*Input, elderly user:* microphone and speech recognition, video camera, biosensors, context-aware sensors. *Output, elderly user:* loudspeaker and text-to-speech, large graphics display.
	Input, health professional: microphone and speech recognition, keyboard, mouse and touch-enabled graphics monitor. *Output, health professional:* loudspeakers, graphical display.

3 Requirements

Because the user-centred development and usability are central to LUL project, the method adopted involves user inputs from the requirements stage. The usability will not be evaluated only at the end of the development process; a first prototype was developed to integrate even in the requirements stage. To enable the creation of this prototype as earlier as possible, a first draft version of requirements was obtained based on the so-called imagination methods [12], particularly by using Personas and the knowledge of specialists (namely the Gerontologists and Physicians that integrate the project team). This requirements, summarized below, were obtained based on the characteristics of the Personas and their goals and descriptions of usage scenarios (not included due to space limitations), method inspired by the PSG (Persona + Scenario + Goal) method of [13].

3.1 Personas

Three main Personas were considered essential for service definition and created with the help of specialists. For the system user Persona, WHO ICF [14] classifications are used (inside parenthesis).

Mrs Zulmira at age 70, retired secretary from a small company in a semi-rural area has osteoarthritis (b7102) and hypertension (b4200). Despite her presbyopia (b2150) and hearing problem (b2304), her family physician (e5800) makes the referral for a community exercise program (d9201). Their relatives, who live near by, work full time. The restriction in using public transportation (d4702) prevents her of attending the exercise class (d9201).

Flávia Conceição, 32 years, is a Gerontologist specialized in adapted physical activity. She started her career working with frail elders, developing activities and exercise programs at private charities. She is familiar with the use of the web, such as Microsoft Messenger and social networks.

Dra Filipa - Physician, specialized in rehabilitation, at a hospital, aged 55, married, with some responsibilities in taking care of her parents (living in nearby). Professional use of computer applications (such as email and word processors), but with not much interest for computer use. She does not have a laptop or smartphone.

3.2 Goals

Personal Goals of the Elderly: to be able to perform fitness exercises or rehabilitation exercises with health professionals monitoring – speciality dependent of the exact objective - without the need to abandon their residence; to have faster access to such services and at the most convenient times according to daily activities; use this new health service in a attractive, simple and secure way.

Health Professionals Goals: to be able to extend there services outside the limits of the traditional place they develop their activity; to be able to provide the services to a larger number of persons, to persons that can not leave their homes easily and even to rural areas / more isolated regions; to be able to cross borders to provide transnational planning and monitoring of rehabilitation; to have the possibility to easily combine their competences with other, possibly remote, to create an multidisciplinary monitoring.

Health Institutions and the National Health System: to be able to sent people home earlier from hospitals by assuring control over rehabilitation plan at home; to improve the health of relevant groups of the population, reducing health problems and increasing productivity.

3.3 Functional Requirements for the Service

The diagram in Fig. **1** represents the first, simplified version, of the functional requirements for the service, elaborated based on the scenario and goals already presented.

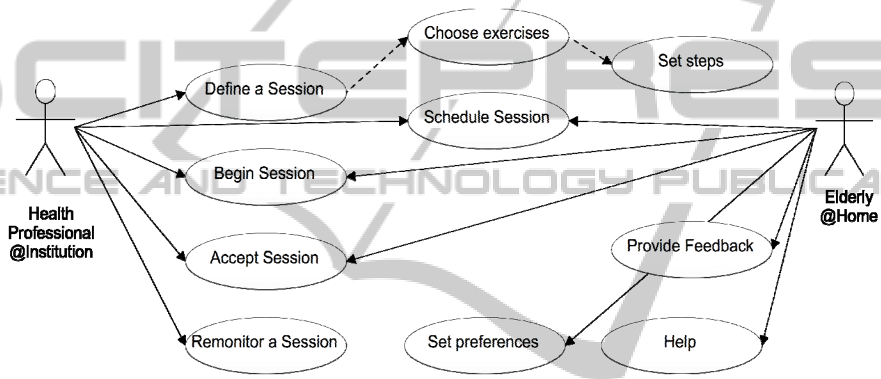

Fig. 1. Main functionalities for the Service.

3.4 Non-functional Requirements

There are many non-functional requirements that can be applied to this service. We will only present the ones considered most relevant:

1. Reliable communications with enough Bandwidth - In order for the prototype to work and connect all participants, the system must be supported with enough bandwidth;
2. System must be reliable - Given the type of operations being performed, it is crucial that the system remains operational and stable; it is crucial for the system to possess fault-tolerant capabilities in order to continue operating even in the event of a failure.
3. Distributed and heterogeneous computing environment – to cover the geographically distributed nature of the service and to avoid the need for specific computing environments;
4. The system must be scalable and extensible - It must be ready to grow in the future and to be capable of integrating new functionalities. The future inclusion of new input and output modalities should be easy;

5. The system must log all system and users activities - This is crucial for the Living Lab paradigm;

The system must be highly usable. This is without doubt the most important of all non-functional requirements to us. All of the goals proposed should be easy for the users (elderly and health professionals) to achieve; the user must be comfortable interacting with the system; the system should be simple and natural to him. If it is not, then probably the purpose of the system will eventually fail. This requirement is further detailed below.

Human Computer Interaction (HCI) focused non-functional requirements given the high interaction needed for our service:

1. The text interface should be readable at a distance, have large and clear text capable , and adapting itself according to the user's distance;

2. Interaction with users must use simple words and sentences;

3. An highly adaptable and intelligent multimodal output system, able to adapt itself to changing environment conditions (light, noise, distance, etc.) and to its users needs, limitations and personal choices:

– Redundancy of modalities must be used in order to increase the chances of message delivery;

– Output modalities with capabilities, based on the environment and user, to decide to activate/deactivate themselves. Ex: there is no reason to keep active an Text-to-Speech (TTS) output when the user is deaf;

– Use of several output modalities to make the system usable by speech and hearing impaired persons;

– Output characteristics, such as the volume of the synthesized speech, must adjust themselves according to the user and the environment (ex: distance to speakers, noise level and users hearing acuity);

– Speech rate must be adapted to listener and listening conditions;

– Based on their preferences, allow users to be informed through their preferred modality(ies).

4. Multi-touch input should be available on the health professional side in order to allow him to easily access some information or quickly select a course of action.

4 A First Prototype

In order to introduce as early as possible the elderly in the development loop and usability evaluation of the new service we opted to start by creating a minimally functional prototype, capable of enabling the interaction between a health care professional (in the hospital for example) and an elderly at home or institution. Use of this prototype by seniors will provided the needed information for a first refinement of the requirements.

Creation of a prototype for the service depends on: (1) the development of two application with suitable Human Computer Interaction, one for the elderly at home, other for the health professional planning, monitoring and evaluating the session; (2) the necessary network connections and services (ex: to transmit video); (3) the existence of the health professionals and elderly to provide and use the service. Here we will only address the first.

The two applications use multimodal input and output, with particular emphasis in the use of speech and text. The use of speech derives from its potential to be usable by visually disabled people and to enable interaction hands free and at some distance from the devices. This capability of receiving information and giving commands to a computer at a couple of meters of the TV/computer display is essential when the aim is making all body movement exercises.

4.1 Architecture and Services

The main characteristics of the first prototype's architecture is its SOA (Service Oriented Architecture) approach, implemented with Jade [15], the decentralization and autonomy of all its service agents, and its adaptability to changing environment conditions (context) and user preferences and limitations.

The choice of using an agent based platform seemed one of the best available options, because it is a mature and quite stable technology, able to provide us a solution for a distributed heterogeneous system, supported on a standard communication protocol (FIPA), which may simplify future integration of third party services supporting new input and output devices.

Agents also provided us a versatile solution for the need of a decentralized, scalable, adaptable, and intelligent system. **Fig. 2** presents a global view of the prototype.

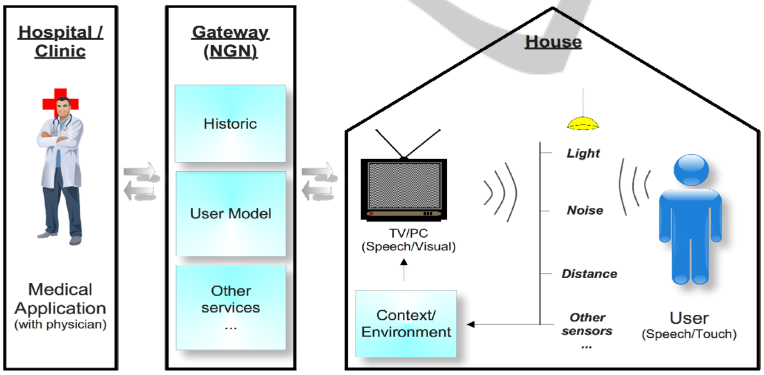

Fig. 2. General view of the service architecture.

To provide a solution to the adaptability of our applications to changing environment conditions we have defined a context service. This service, implemented as an agent, is responsible to register, in real-time, all the relevant environment conditions such as noise, light, distance of the user to the output devices, etc.. The information for this service comes from a set of specialized producer agents that, in general, are connected, directly or indirectly, to sensors, such as microphones and cameras.

A user model service is also provided, in order to register and fetch specific user related **capabilities** and **preferences.** Examples of capabilities are vision and hearing acuity and mobility capabilities. As preferences we can have, for example, the personal preference for receiving information visually or for a specific color for text.

When considering possible approaches to incorporate the different input and output devices we were faced with two basic choices. The first one was to make them simple dummy devices responsible only for sending input information to the system and to receive output messages already adapted to the context and user from it. This approach has several serious problems, such as the need for a complex fission coordination service very knowledgeable of all the available output devices, and able to coordinate all the relevant context and user conditions in order to generate proper outputs. Also, that approach would make it more difficult to scale and extend out multimodal applications to new input and output devices.

To avoid these problems, it was decided to provide intelligent agent output devices, able to adapt themselves to changing context and user conditions. To remove from the context and user services the knowledge of all the available output and input devices, those services are implemented as simple reactive information repositories in which the device agents register themselves to be listeners of specific variables. For instance, a speech output device may register itself in the user model service to become aware of possible hearing or comprehension user problems, and register in the context service to receive changes in the user distance. With such knowledge, this device may change the volume of the sound, and the rate of the speech to maximize a successful transmission of the messages required.

There are also two more important services (also implemented as agents): a logger service able to register all the relevant history information for latter uses, such as for the creation of user specific information for the user model. A coordination fission service able to ensure the reliability of the application as a whole. This service would be able to ensure fault tolerance, notifying the application when an output was not transmitted (for example, due to the unavailability of proper output devices).

4.2 Implementation

Our prototype implementation (Fig. 3) includes three main blocks, running in 3 different networked computers, resulting from the subdivision of the application at the elderly side in two: one, in a more capable device (a server), including the more performance demanding operations; the other (in the Home TV/ Personal PC) responsible by interaction and the devices for environment monitoring.

Besides the main computing devices, the system includes input and output devices such as microphones, speakers and video cameras, required for the interaction between the users and the platform, and sensors to detect environment factors. In order to facilitate interoperability, all communication processes are done using TCP and RTP protocols over IP networks.

Agents (JADE) communicate with each other via FIPA messages. A FIPA message normally includes a destination agent, a title, an attachment and a flag that indicates the type of message. This last one became important to us because it allowed us to filter the messages according to its type and process the response more easily.

In order to facilitate the search for agents, JADE provides a service-registration index upon which the agents may inscribe themselves. This was very useful to us in respect to output agents. If an agent perceives that it can resolve some types of output messages (such as text), then he inscribes himself under the registry for that given action. This allows other agents that may need to output some message to instantly search and find an agent that may answer their request. Vice-versa, if the output agents that perceive that they can't no longer perform a given action to unsubscribe themselves thus reducing possible miscommunication or system errors.

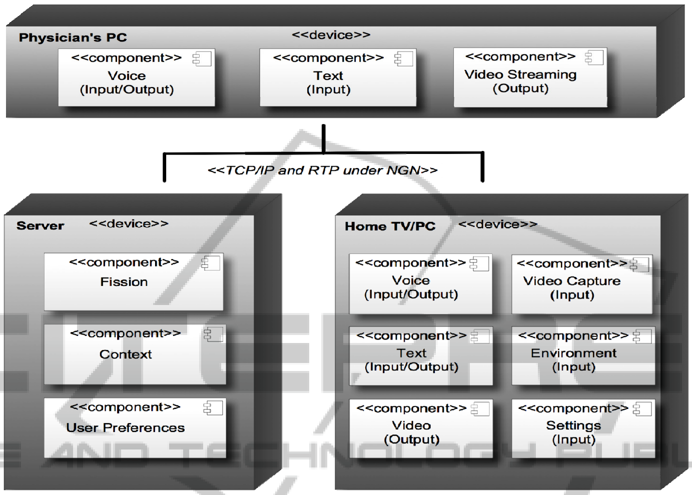

Fig. 3. Deployment diagram.

Text-to-Speech and Speech-to-Text European Portuguese capabilities of the system, made available by a speech recognizer and a text synthesizer separated agents, are implemented using Microsoft's Speech Platform [16]. Video streaming between the participants was also enabled using JMF (Java Media Framework) over RTP (Real Time Protocol).

At the time of writing, 3 environment monitoring agents are available: an environment background “noise” level, implemented by using and event of the speech recognizer, AudioLevelUpdated; one on the light conditions; a third one, providing an estimate of the distance of the user to the screen/display. The light conditions are evaluated based on statistical measures - Mean Sample Value (MSV) - of the intensities histogram calculated on the acquired image. The distance is obtained using algorithms based on background subtraction to estimate the position of the person in the image. Using the properties of the vision system (position, camera and lens properties, etc.), it is possible to estimate the position of the person related to the camera.

5 AdaptO – Adaptive Multimodal Output

Focus on the communication between the system and the user - multimodal output - is the main novelty of our architecture and prototype. In general, systems in this area of application incorporate various modalities to communicate with the user (such as voice, text or images) but they are completely devoid of any autonomy, i.e. simply output the messages they receive using default definitions. In the proposed model, it is intended that these modalities have some independence and self-adaptability to user and context of use (ex: environment).

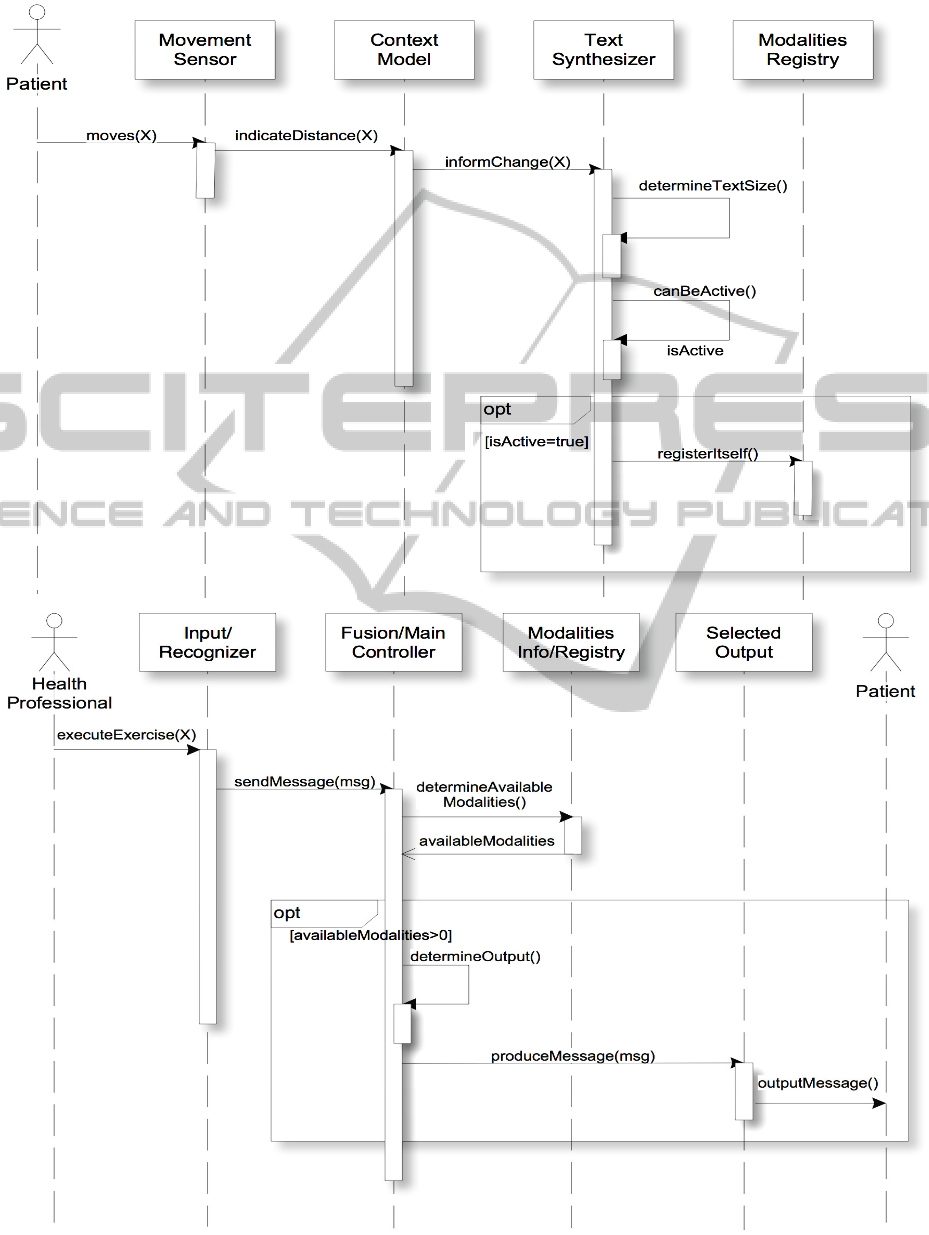

Fig. 4. Sequence diagrams of actions in response to: user movement relative to screen (at top); instructions regarding a new step of the exercises program sent by the health professional (bottom).

The current agents capable of transmitting information from the system to the user – synthesizers in multimodal interaction literature - include two important mechanisms: The first is capable of deciding if, in the current environment conditions and taking in consideration user capabilities, it is in a position to be active and fulfill the request. For example, if the user is hearing impaired or the noise level is too high the synthesizer deactivates itself.

The second changes some of the properties of the message to be transmitted also based on contextual and user information. Presently, and as proof of concept, the text synthesizer varies, using simple heuristics, font size as function of user vision capacity, environment lighting conditions and distance of the user to the screen. The speech synthesizer is capable of varying both volume and speech rate. This increases the chances that the message is received and understood. The context and user models, implemented as services, are crucial to make possible these two mechanisms.

Illustration of the mechanisms in action is presented in fig. 4 using an UML sequence diagram. The diagram presents the sequence of most relevant actions in response to a change in user position and when the health professional sends information on how to execute a part of an exercise.

6 Conclusions

In this paper, we present the description of a new service in the area of elderly health support at home and of the first prototype implementation. As was considered crucial to make the service accessible and usable by the target group – older adults – despite their visual and hearing capabilities, one of the more worked aspects of the development so far relates to the adaptability and inclusiveness capabilities of output. The basis for an intelligent adaptation of output – that we called AdaptO, a Portuguese word meaning adapt – was proposed and first versions of the needed services and agents were created. In addition, particular attention was given to the definition of the architecture and on the use of services and agents to create an implementation. In the near future real user test on the service and applications will be performed. The interaction aspects of the applications for the health professionals and elderly at home and the evaluation plan are being finished.

Acknowledgements

This work is part of the COMPETE - Programa Operacional Factores de Competitividade and the European Union (FEDER) under QREN Living Usability Lab for Next Generation Networks (http://www.livinglab.pt/).

References

1. LUL Consortium. Living Usability Lab for Next Generation Networks. 2010; Available from: www.livinglab.pt.

2. AAL Forum. Track R - Robotics enabled Assistive Living. 2010; Available from: http://www.aalforum.eu/page/track-r-robotics-enabled.
3. Best-Martini, E. and K. BOTENHAGEN-DIGENOVA, Exercises for frail Elders. 2008: Human Kinetics.
4. Burdea, G. C., Virtual rehabilitation--benefits and challenges. Methods Inf Med, 2003. 42(5): p. 519-23.
5. Burdea, G. C., Rubber Ball to Cloud Rehabilitation - Musing on the Future of Therapy in Virtual Rehabilitation International Conference. 2009: Haifa, Israel p. 50-50.
6. NIDCD. Presbycusis. Available from: http://www.nidcd.nih.gov/health/hearing/presbycusis.html.
7. Huang, Q. and J. Tang, Age-related hearing loss or presbycusis. European Archives of Oto-Rhino-Laryngology, 2010. 267(8): p. 1179-1191.
8. McCoy, S. L. et al., Aging in a Fast-Paced World: Rapid Speech and Its Effect on Understanding. The ASHA Leader., 2005.
9. O'Connell, T. A., The Why and How of Senior-Focused Design, in Universal Usability, J. Lazar, Editor. 2009, Wiley.
10. Rousseau, C., Y. Bellik, and F. Vernier, Multimodal output specification / simulation platform, in Proceedings of the 7th international conference on Multimodal interfaces. 2005, ACM: Torento, Italy. p. 84-91.
11. Persona, User Requirements and AAL Service Specification. 2008, Person Consortium.
12. Bernsen, N. O. and L. Dybkjaer, Multimodal Usability. 2009: Springer.
13. Aoyama, M. Persona-Scenario-Goal Methodology for User-Centered Requirements Engineering. in 15th IEEE Int. Requirements Eng. Conference. 2007: IEEE.
14. World Health Organization (WHO). International Classification of Functioning, Disability and Health (ICF). 18 October 2010]; Available from: http://www.who.int/classifications/icf/en/.
15. Bellifemine, F. et al., JADE (Java Agent DEvelopment Framework) - A White Paper, in exp. 2003.
16. Microsoft. Developing Speech Applications. 2010; Available from: http://www.microsoft.com/speech/developers.aspx.