Multimodality and Adaptation for an Enhanced Mobile Medication Assistant for the Elderly

Figure 1: Medication Assistant's main screen.

António Teixeira
Dep Electronics Telec. &
Informatics/IEETA
University of Aveiro
3810-193 Aveiro, Portugal
ajst@ua.pt

José Casimiro Pereira
Polytecnic Institute of Tomar
2300-531 Tomar, Portugal
casimiro@ipt.pt

Flávio Ferreira
IEETA
University of Aveiro
3810-193 Aveiro, Portugal
flaviommferreira@gmail.com

Samuel Silva
IEETA
University of Aveiro
3810-193 Aveiro, Portugal
sss@ua.pt

Nuno Almeida
IEETA
University of Aveiro
3810-193 Aveiro, Portugal
nunoalmeida@ua.pt

Alexandra Queirós
Health School (ESSUA)
University of Aveiro
3810-193 Aveiro, Portugal
alexandra@ua.pt

Ana Filipa Rosa
IEETA
University of Aveiro
3810-193 Aveiro, Portugal
filiparosa@ua.pt

André Oliveira
IEETA
University of Aveiro
3810-193 Aveiro, Portugal
andremota@ua.pt

CHI 2013 Mobile Accessibility Workshop, April 28, 2013, Paris, France.

Abstract
In order to improve medication adherence a mobile multimodal application is proposed, targeting the elderly population, and exploring speech as both input and output modality.

Author Keywords
Multimodality; speech; natural language generation; adaptation; medication; elderly.

ACM Classification Keywords
H.5.2 [Information interfaces and presentation]: User Interfaces—Natural Language, Voice I/O.

Introduction
The number of mobile applications targeting the elderly population, to improve quality of life and health conditions, has been increasing. The elderly present high levels of non-adherence to medication and, in that sense, it is important to develop strategies and applications that contribute to improve medication adherence [6]. Providing instructions and ways of solving problems, such as alerts, cues, reinforcements and self-monitoring tools, increases their motivation and reduces the number of errors in taking medication [5]. Therefore, applications that allow medication management are clearly useful.

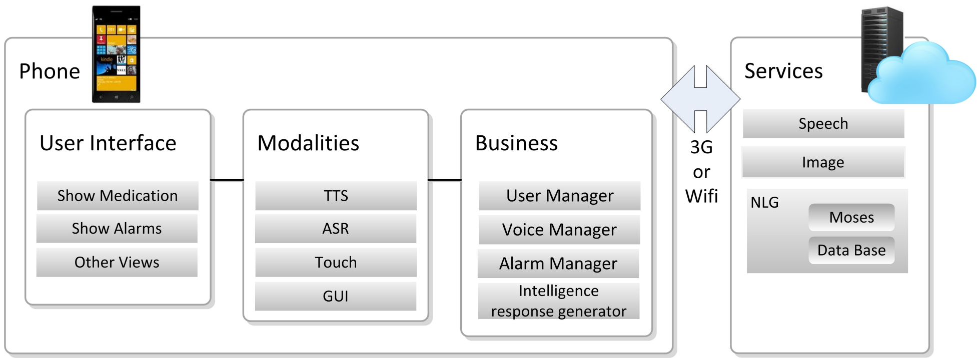

Figure 2: System architecture depicting the different modules.

Most of existing applications consist of reminders to take medicines and registration of the intake (e.g., True-Kare [1] and Pillboxie [2]).

Ageing is accompanied by vision, hearing and memory losses, reduced motor skills and decreased touch sensibility. Regarding interaction, existing applications focus on touch, which sometimes may not be feasible for the elderly population. The use of modalities such as speech has the potential to make the application more accessible to the elderly population, according to their needs and specificities [4].

We propose a mobile application (fig. 1) targeting the elderly population to improve adherence to medication and minimize errors, that can also serve as a testbed for research in the best uses of different modalities with a particular focus on speech.

[1] https://www.true-kare.com
[2] http://pillboxie.tumblr.com/

Application Overview

Development follows the methodology described in [2]. After gathering requirements, a prototype is proposed, evaluated and used in further refining the requirements. This iterative methodology continues with additional prototypes and evaluations towards an increasingly refined application.

From Analysis to Requirements: In order to get the context scenarios and the system requirements we followed the five stages method proposed by [1] and have identified three notable expectations of the Persona: 1) to be able to use the application when suffering from arthrosis; 2) to have an application to help prevent gaps in medication, control intakes, report side effects and provide alerts; and 3) to be able to use the application despite their little aptitude for working with electronic devices.

The following functional requirements were identified: 1) provide medication alerts to remember users about medication schedules; 2) provide help for elders in common medication related questions, such as "*What if I forget to take my medication?*"; and 3) provide medication insertion and management by third parties so that seniors do not need to perform this task.

Careful analysis also lead to the following user requirements: 1) inform users in simple everyday language avoiding technical terms; 2) provide touch and speech interaction for all features in order to tackle existing physical/cognitive limitations; 3) reliable and credible application performance to instil trust on its advices and alerts; 4) avoid content overload on screen.

System Architecture and Features: The application is divided in several modules (with self explanatory names) and views as depicted in figure 2. Beyond those modules,

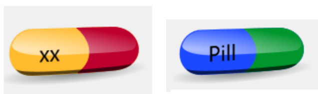

Figure 3: Pill model and derived pill image after adapting the model to a specific medication characteristics.

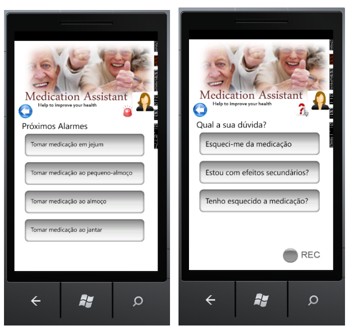

Figure 4: The Alert View (left) enables the senior to view the next alerts and plan the remaining hours of the day accordingly. The Advisor View (right) can be interacted with through speech or touch and allows the senior to obtain answers to questions related to the medication.

there are three services used by the application: the images service that, based on models of each type of pill (in SVG format) generates images for the medications that do not contain the pictures that are needed for the mobile application (fig. 3);the user service, that manages the user and gets the user alarms information; and the voice service, which was one of the main focuses of the work carried out, and provides speech capabilities to the application including speech recognition and speech synthesis.

The first prototype was developed for Windows Phone 7 providing two main working functionalities: the generation and visualization of alerts and the information/advice service providing information regarding the medication (e.g., packaging and side effects) and action to take in case of forgetting to take medication (fig. 4).

Multimodal Input

The system can get input by speech or by touch. The input by touch is accepted for all the operations, thus making speech an auxiliary input. In most cases, beyond the advantages associated with the use of voice over the use of touch [3], voice inputs can lead to a quite fast and intuitive interaction.

For the case of visualization of a medication and the details associated, seniors can use their voice to get information about side effects, the purpose of the medication and medication intake schedules.

In the advices menu, speech usage is quite intuitive and can lead to faster responses from the system. For example, to obtain information regarding what to do if a medication intake was forgotten, if touch is used, the user must start by choosing the option "I forgot my medication," followed by "I forgot a medication" and finally choosing the medicine that was forgotten. Through speech, the elder just has to say a sentence (e.g. "I forgot to take medication X" or "I forgot X") in order for the application to be able to provide an answer.

Speech recognition (part of the Speech Service) is performed using the MSP, converting the audio stream from the user to text using the proper grammar.

Multimodal Output

The output is provided through speech, text and graphics and aiming for content redundancy over all modalities, to ensure system flexibility towards user profile and capabilities.

The list of medications to take consists of a panoramic view (fig. 5) that allows the user to see three different perspectives: 1) list of medication names to take (and dosage); 2) image with all medications to be taken; 3) pictures of packages of the medications to take. By clicking the images or the list it is possible to access medication details. These views are complementary with each other and have the advantage of providing an image to be used to compare with the real set of tablets that the elder will have to take.

Information can also be provided through speech. The speech service receives the information to be read to the user and, using the voice installed (by default in Portuguese), synthesizes the message to speech using the Microsoft Speech Platform (MSP). Other parameters can be provided such as volume and speech rate. For the latter, the default value has been set to 1 based on evidence provided by recent experiments in our group regarding synthetic speech perception by the elderly. Additionally, the Natural Language Generator (NLG) is a service used to generate sentences, to be read to the user.

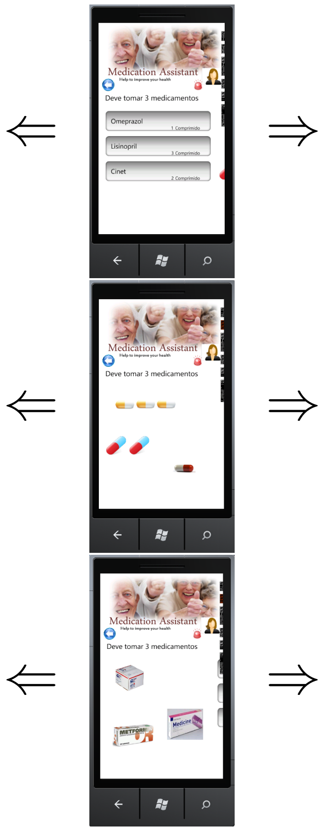

Figure 5: Panorama used to show medication details: 1) medication name and dosage; 2) pills colour/shape and quantity; 3) medication boxes.

It uses Moses [3], a statistical machine translation system featuring a trained language model for the purpose of the project. Moses receives a set of parameters (user, medication name, quantity, application method and intake interval) and returns the generated sentence (e.g., "You should take one capsule of X").

Adaptation

An important feature of the application is its ability to adapt the output to context. It can be configured by the user and/or by measuring some environment variables.

Currently, two variables to evaluate context are considered: 1) distance between the user and the smartphone, by using the smartphone camera turned to the user, detecting the face and computing the ratio between face area and total image area; and 2) ambient noise level, by measuring the mean noise level from a sound sample recorded when the user is not speaking. These parameters are stored in the context model and, if there are changes, the size of items on screen is recalculated or the output volume adjusted.

Conclusions

A mobile multimodal system to help improve adherence to medication was presented. At this stage it provides features to support medication intake (what pills to take and when) and advise the user regarding medication characteristics or how to act when an intake is missed. Speech has been successfully added as both input (along with touch) and output (along with text and graphics). Preliminary work regarding adaptability to context allows improving interaction by dealing with possible vision difficulties or usage in noisy environments.

[3] http://www.statmt.org/moses/

Further Work: Formal evaluation of the developed system is being carried out based on the methodology proposed in [2] and we aim to further explore and improve the uses of speech for input/output and adaptation features to users and contexts.

Acknowledgements

This work is part of the Smart Phones for Smart Seniors (S4S) project, a QREN project (QREN 21541), co-funded by COMPETE and FEDER.

References

[1] Cooper, A., Reimann, R., and Cronin, D. *About Face3: The Essentials of Interaction Design*. Willey Publishing, 2007.

[2] Martins, A. I., Queirós, A., Cerqueira, M., Rocha, N., and Teixeira, A. The international classification of functioning, disability and health as a conceptual model for the evaluation of environmental factors. *Procedia Computer Science* (2012).

[3] Nass, C., and Brave, S. *Wired for Speech: How Voice Activates and Advances the Human-computer Relationship*. MIT Press, 2007.

[4] Pieraccini, R. *The Voice in the Machine: Building Computers that understand speech*. MIT Press, 2012.

[5] Rosen, M. I., Rigsby, M. O., Salahi, J. T., Ryan, C. E., and Cramer, J. A. Electronic monitoring and counseling to improve medication adherence. *Behaviour Res. and Therapy 42*, 4 (2004), 409–422.

[6] Zygmunt, A., Olfson, M., Boyer, C., and Mechanic, D. Interventions to improve medication adherence in schizophrenia. *Am. J. Psychiatry 159*, 10 (2002), 1653–1664.