Multimodal Interaction A.Y. 23-24

# AAL System for Elderly Care

BY LUCIA FORES 1836451

**Elderly Care** = services older people need for physical or mental impairment

**Home Care**: health care or supportive care provided by caregivers in the **individual home** where the patient or client is living

**Help for health and medical needs**: include helping the assisted in exercising, eating balanced meals and **taking medication in a timely manner**

**Caregiver** = Someone who takes care of a person who is young, old, ill, or disabled , either as a family member or friend, or as a job

# Elderly Care – Definitions and Project Context

# Problem Statement

- The **increase** of the average **health conditions** of the elders in recent years led to an **increase of** the percentage of **elderly people** who prefer to continue **living alone** and **in their own home** [1]
- Need to provide contextual monitoring system so that the **caregivers** can be in contact with their assisted and so that they can be updated about what **medication** they take and **how they feel**

# Proposed Solution

- An Ambient/Active Assist Living (**AAL**) **system** to be used by elders and caregivers in the context of the **Home Care**
- The system will be used to **remind** and **check** that the users take the **correct medications** at the correct time, to let the caregivers know when and which medication they took and to build a **constant communication channel** between them to fast communicate the need of **immediate help**

# Main Technologies – 1 –

The project was made thanks to the use of **Vosk**, **gTTS**, **PaddleOCR** and **OpenCV**

## Speech Interaction

### Vosk

**Vosk** is an open source offline speech recognition toolkit developed by **Alpha Cephei**. In the project has been used to implement the **speech recognition** module by exploiting their vosk-model-*small-it-0.22*

### gTTS

**gTTS** (Google Text-to-Speech) is a Python library that converts text to speech using the **Google Translate's API**. In the project has been used to implement the **speech synthsis** module

# Main Technologies - 2 -

## Picture Recognition

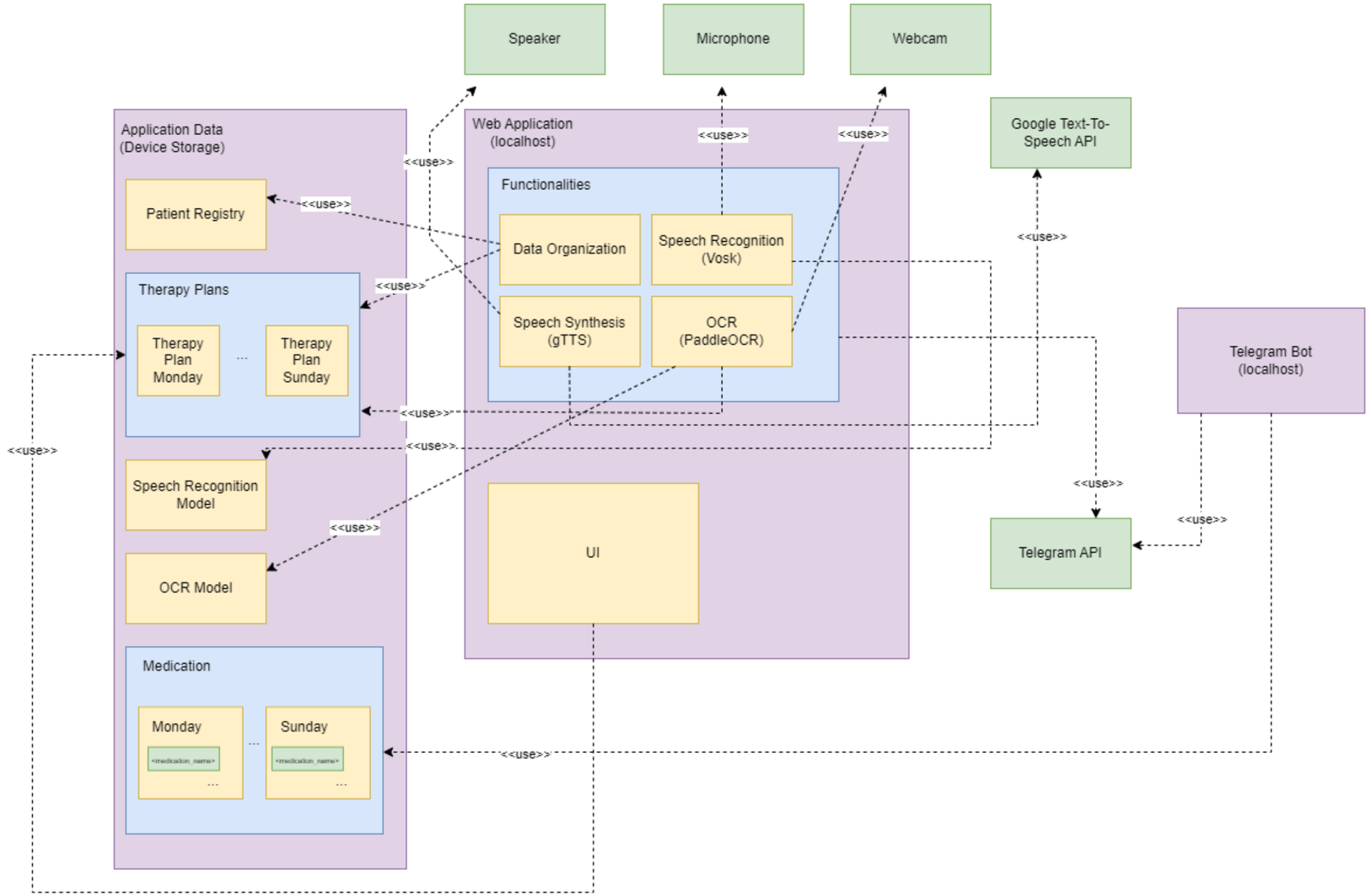
### OpenCV

**OpenCV** is the biggest open-source library for computer vision.
Thanks to its usage, a **picture** of the box of the medication took by the user (needed for the medication recognition) has been made

### PaddleOCR

**PaddleOCR** is a recent an accurate [2] OCR system based on the PaddlePaddle deep learning platform. In the project the OCR system has been used to recognize the text on the medication boxes

# System Architecture



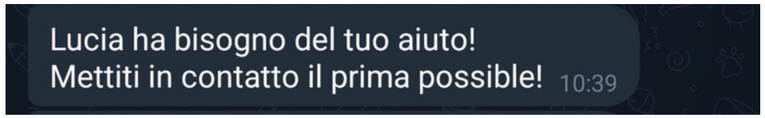
# Main Interaction

Aside from the request for help that the user can trigger at any time just by saying the word **aiuto** (which will lead to the automatic send of an help request to the caregiver through the Telegram bot), the main interaction of the system can be summed up as the following

1. The system recognize from the therapy plan that is time to take some medications
2. The system asks to the user how they feel and if they feel bad asks them if they want to send an help request to the caregiver
3. The system lists to the user the medication that they need to take and for each one performs the following
4. The system asks to the user to take the correspondent medication box, waits for the word **foto** and take a pic of the box
5. If the box is correct it says to the user the amoun of medication to be taken and then waits for the word **avanti** to proceed
6. When the user took all the medications the system says goodbye to them and then send a recap message to the caregiver

# Prototype Screens

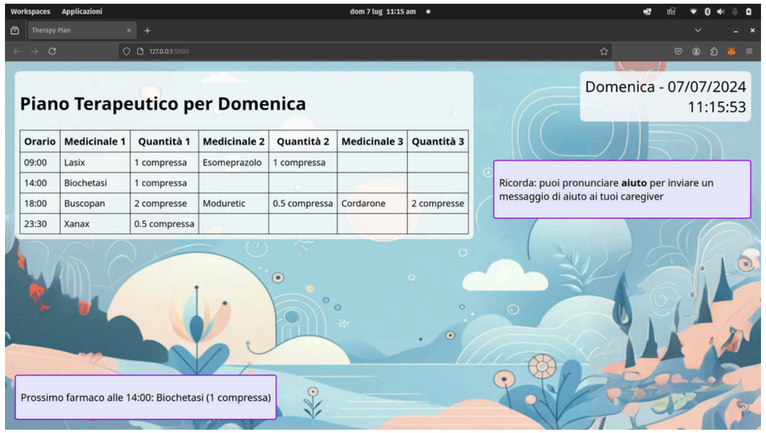
## Telegram bot messages



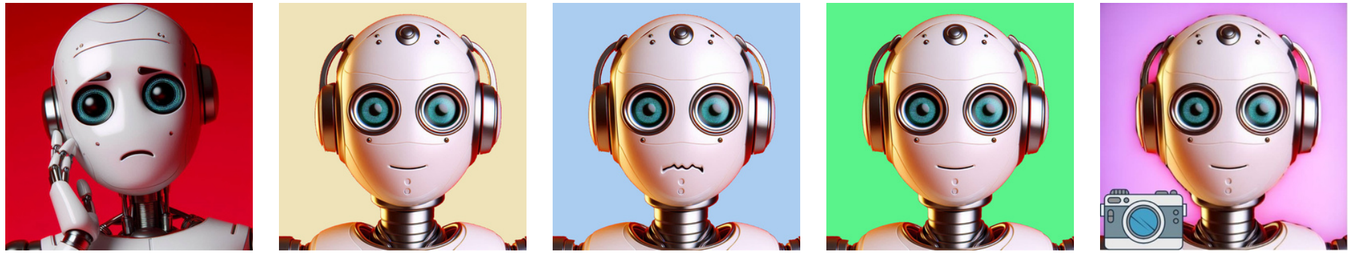

**Help message** sample


**Recap message** sample

## Web Application screens




**Idle** screen



**Interaction** screens (from left to right the **alert**, the **medication**, the **sad**, the **happy** and the **photo** one)

# Interface Evaluation – 1 –

**Visibility of System Status**

The status is **always visible** both in the idle screen or with the interaction screens. In each important moment is also given to the user a vocal feedback

**Match Between the System and the Real World**

The system communicates with the user with the **natural language**, always **explains** to the user what is happening and uses only **concepts** that the user knows

**User Control and Freedom**

The user **cannot start extended processes by error**, the **help routine cannot be stopped** because it must be an immediate one (no further confirmations can be asked)

**Consistency and Standards**

There are **no dupicated commands, situations or actions** and all the **flows** are **explained** while taking place

**Error Prevention**

The errors are **catched** and **fixed** by the system routines by presenting them to the user and suggesting them how the should correctly behave

# Interface Evaluation – 2 –

| | |
|---|---|
| **Recognition Rather than Recall** | The user **do not have to remember anything** because the system always **guide them** in the next action that they need to perform |
| **Flexibility and Efficiency of Use** | No shortcuts of usage where given due to the system target |
| **Aesthetic and Minimalist Design** | All the screens are **minimalistic** and presents only the useful information; any other information is given to the user by the system with vocal feedbacks |
| **Help Users Recognize, Diagnose, and Recover from Errors** | The errors are **instantly notified** to the users and the system always provide to them how they can **fix** them |
| **Help and Documentation** | All the extra needed **documentation** is given to the user with **vocal feedbacks** and with **textual feedbacks** in the idle screen |

# Discussion and Analysis

1. The project has been all developed using the **italian language** in order to be able to correctly test the speech recognition

2. Even thought **WhatsApp** could've been the better choice for the messaging system given the diffusion among the range of users take in consideration [3], in the end it has been decided to use **Telegram** given the limitations of the WhatsApp bot

3. At first it was thought to use **pyttsx3** as the synthesizer library (in order to have all the speech interaction module offline) but the synthesizer lacks of **prosody** and so it led to an innatural interaction with the user

4. The system suffers of some limitations like the **difficulty in recognizing some words**, the **synthesization with wrong accents** in some occasions, the necessity of a **specific setting** for the speech recognition and the difficulty in recognizing **texts** that are **not written horizontally**

# Future Work and Recommendations

- Creation of a **Robotic Interface** with the use of Arduino and of a microprocessor
- Creation of a **simpler interface** for the caregiver to compile and update the patient data and the therapy plan data

Let’s see it in action!

# References

[1] Istat. https://www.istat.it/it/archivio/259588. Accessed: 2024-06-30.
[2] Yuning Du, Chenxia Li, Ruoyu Guo, Xiaoting Yin, Weiwei Liu, Jun Zhou, Yifan Bai, Zilin Yu, Yehua Yang, Qingqing Dang, and Haoshuang Wang. Pp-ocr: A practical ultra lightweight ocr system, 2020.
[3] Valentina Turrini. Digital 2024 - i dati italiani. https://wearesocial.com/it/blog/2024/02/digital-2024-i-dati-italiani/, Feb 2024. Accessed: 2024-07-04.

# Thank you!

Questions?

BY LUCIA FORES 1836451