Operating System Voice Control

- Bruno Bastos 93302
- Leandro Silva 93446

Introduction

This project enables the user to control some functionalities of the Windows Operating system using the voice. It is a prototype that explores different ways of interacting with the operating system. In a further implementation, it would require the ability to control the Windows OS with more freedom, as some of the functionalities are limited by the APIs provided by Microsoft. All the functionalities were tested using Windows 11, so there is a possibility that not all of them work on other versions.

Implementation

This section lists and explains the available operations:

- Volume - the user can increase or decrease the volume. Some interactions also allow the user to mute or put the volume to the max.
  Example: “Aumentar Volume”, “Aumentar o som”, “Não consigo ouvir nada”, “Silêncio”, “Diminuir volume”
- Brightness - works in the same way as the volume by allowing the user to increase or decrease the brightness of the screen. It was only tested on the main screen in a laptop, and might not work on other screens.
  Example: “Aumentar brilho”, “Diminuir luz”
- Open Applications - there are a set of applications that can be opened by voice. All of them share a common way of opening, but there are others which have some specific phrases associated.
  Example: “Abrir VSCode”, “Inicializar bloco de notas”, “Quero fazer contas”, “Quero escrever um texto”
- Close Applications - apart from the command that closes every application in the current workspace, only the applications that are available can be closed.
  Example: “Fechar calculadora”, "Terminar bloco de notas”, “Fechar todas as janelas”
- Minimize/Maximize - it is possible to minimize or maximize the windows. In the case of minimization, it is possible to minimize all of them.
  Example: “Minimizar calculadora”, “Maximizar bloco de texto”, “Esconder tudo”
- Move Windows - a window can be moved to a certain position on the screen, one of the corners or one of the directions.

Example: “Mover para cima à direita a calculadora”, “Colocar à esquerda o bloco de notas”.

- Show Windows - if a window is minimized, it can reappear in front.
  Example: “Mostrar calculadora”
- Camera - the camera app can be used as a normal app, however it has an extra functionality that allows the user to take a photo with a delay of three seconds.
  Example: “Tirar uma foto”
- File Explorer - the file explorer has a set of defined known directories that can be opened. It was supposed to behave like a normal app, but we weren’t able to perform other operations on it.
  Example: “Abrir fotografias”, “Ver explorador de ficheiros no ambiente de trabalho”
- Shortcuts - the graphical interface of the program, allows the user to provide a set of paths that will be used as shortcuts. To open the file explorer in one of those shortcuts, they must be called by their position.
  Example: “Abrir atalho um”
- Virtual Desktops - virtual desktops can be created by the user. It is also possible to navigate to the next/previous desktop and delete a desktop.
  Example: “Criar ambiente de trabalho”, “Ir para o próximo ambiente de trabalho”, “Mover para o ambiente de trabalho anterior”, “Apagar àrea de trabalho”
- Browser - some websites are quickly accessible by voice. Some of them have specific phrases but can also be opened like a normal app. There is no implementation to close them nor the browser.
  Example: “Abrir facebook”, “Quero ver um vídeo"(youtube), “Será que já saiu as notas”(Elearning)

Permissions and Feedback:

The program will tell the user whenever there is an action it cannot perform using the voice synthesizer. Some of the actions can be classified as dangerous, so the program will ask for feedback. This way is much less likely for the program to misunderstand some action. Actions are only valid if the program’s confidence is above a certain threshold, which can be higher depending if the action is dangerous or not.

Conclusion:

During the development of this project, we learned the concepts and the objectives of the interaction using voice. Not only that, we managed to figure out some of the possible scenarios where there could be a failure or a decrease in the user experience and tried to find ways of circumventing those problems. There were things that could be refined, but this prototype served its purpose of understanding whether or not an application that controls the operating system using voice could be viable and useful for the users.