

Speech-jokey

An application for editing and converting text into synthetic speech containing intonation

 **ASSIST HEIDI Designing and implementing
Assistive Tools for people with disabilities**

Rieder Nikolaus
Chiara Calvo
Martina Tumsich

What is the project for? Which problem is solved?

Assistive Technologies

Assistive technologies are tools, devices, or software designed to help individuals with disabilities perform functions that might otherwise be difficult or impossible.

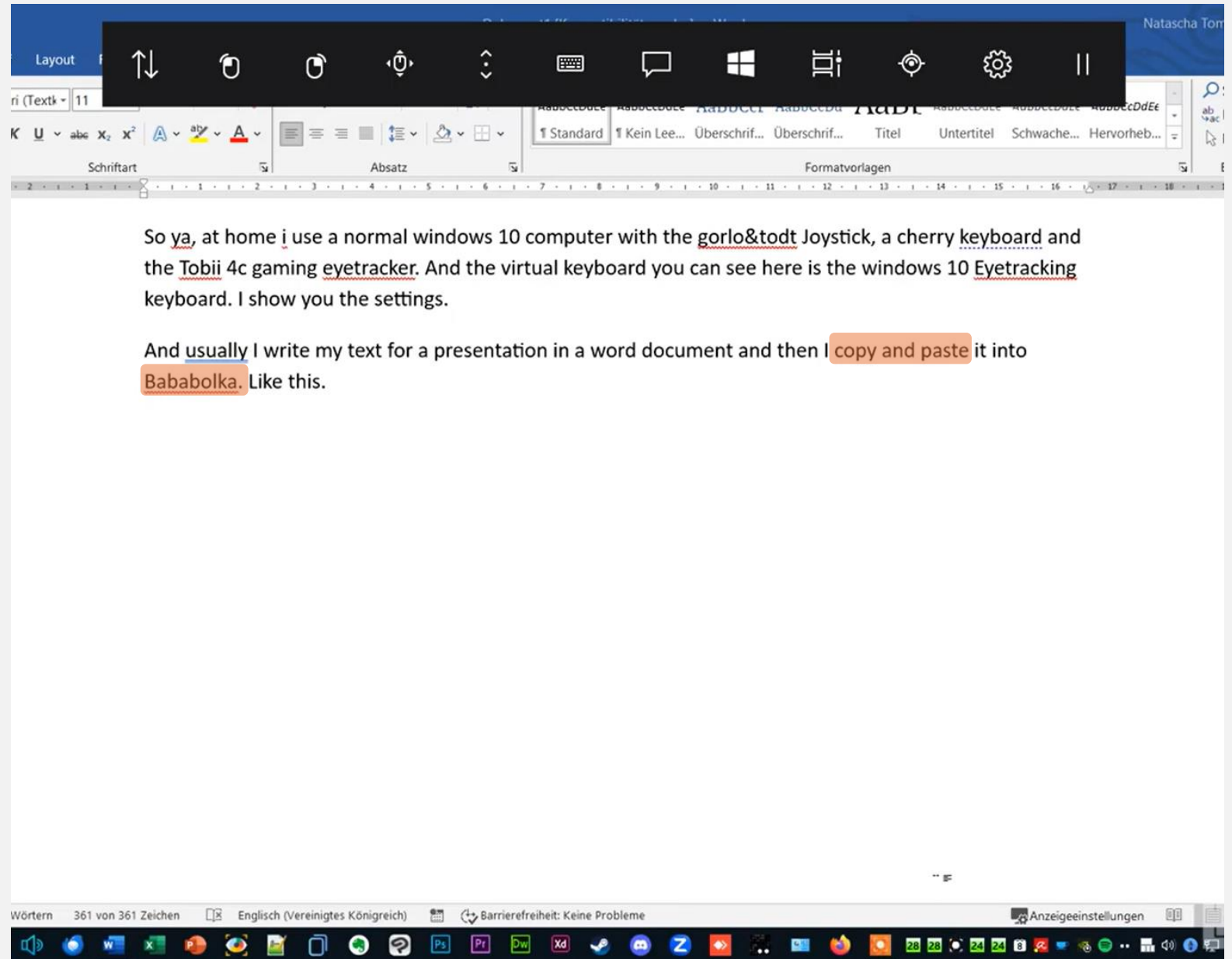


Balabolka



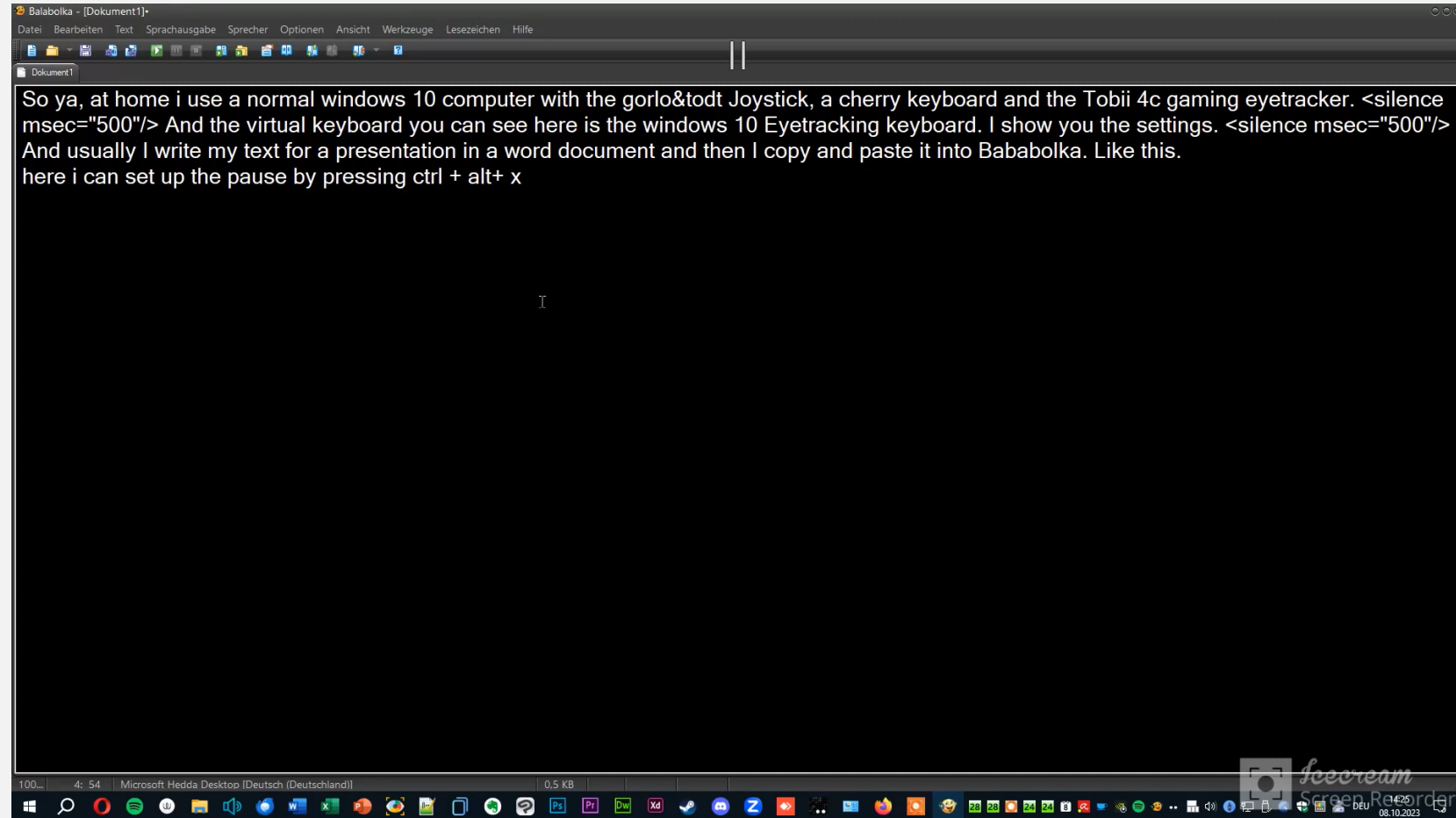
Balabolka

- Balabolka is a completely free TTS software
- Simple to use
- Intuitive interface



Balabolka

- Language selection
- Regulation of the speed
- Add pauses
- Control of the cursor position



Text to Speech

TTS is a technology that converts written texts into audible texts.

It enables individuals with speech impairment to communicate effectively thanks to computer-generated speeches.



Leveraging existing technologies

- **Revoicer**: AI text-to-speech online application: <https://revoicer.com/>
- Multiple **TTS APIs** enable developers to easily incorporate speech synthesis into their applications:
 - Microsoft Azure TTS: Realistic AI Voice Generator
 - TTS Maker
 - Many others
- Evaluation of many voices from different TTS APIs

ElevenLabs

- [Text to Speech & AI Voice Generator | ElevenLabs](#)
- It offers an advance TTS API which is able to generate voices quickly and in multiple languages
- Mimic human intonation → Serena's voice selected

SSML

- Speech Synthesis Markup Language tags
- Detailed customization of the speech synthesis output

< speak >

< prosody >

< break >

< phoneme >

< emphasis >

< p >



Main goal: Encoding intonation

Even if there are existing options with natural-sounding voices, **encoding intonation and emotions remains a significant challenge.**



Speech Jokey

Main goal: Encoding intonation

Substantial enhancement in
communication **for people
with speech impairment**



Speech Jokey


... So what is Speech Jokey?

Speech Jokey is an application which syntetize speech from user-provided text, and that gives the possibility also to add pauses, emphasis, intonation or others speacking specification.

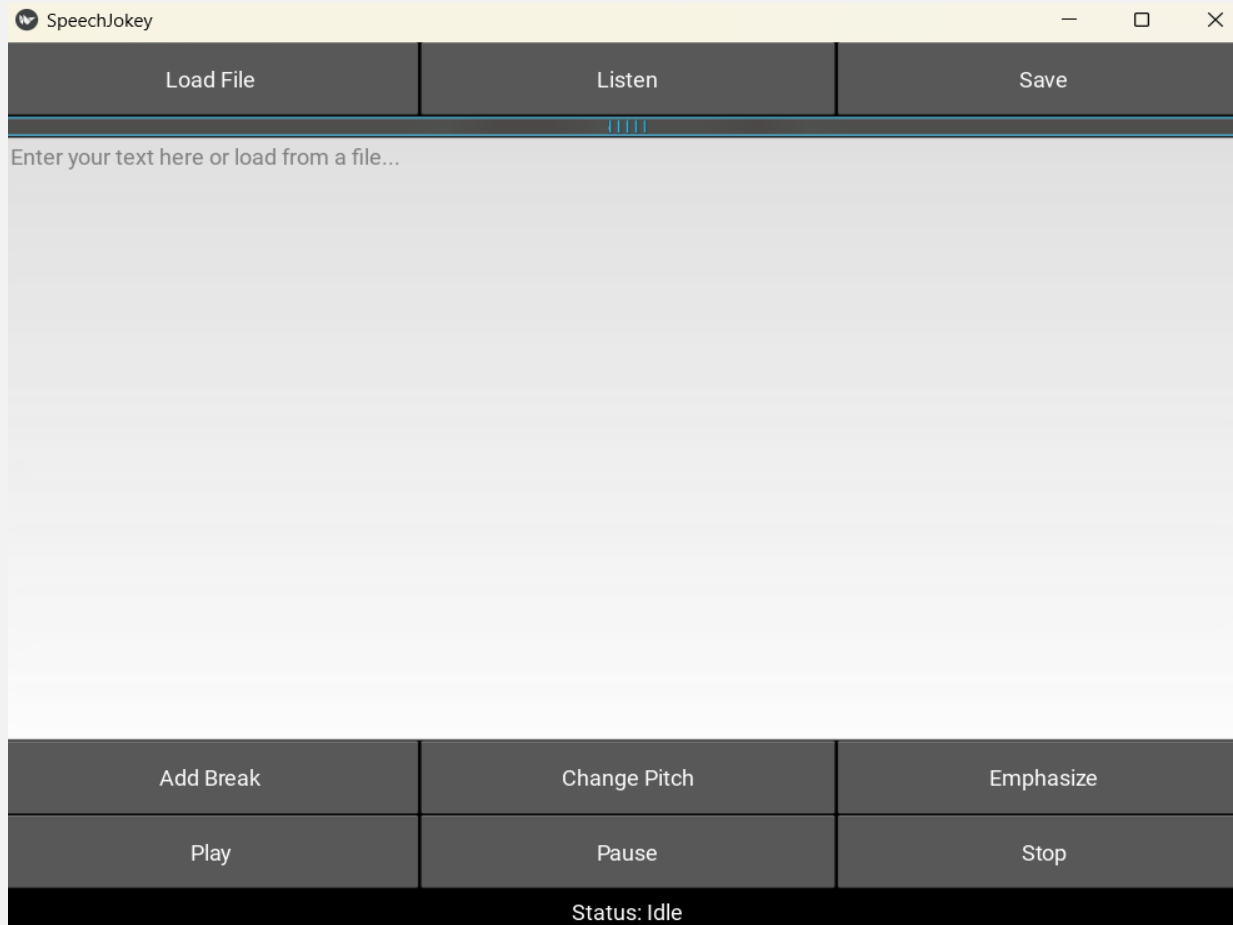
The application is written in **Python** and uses the **Kivy** framework for the user interface.



How should it work from the
user perspective

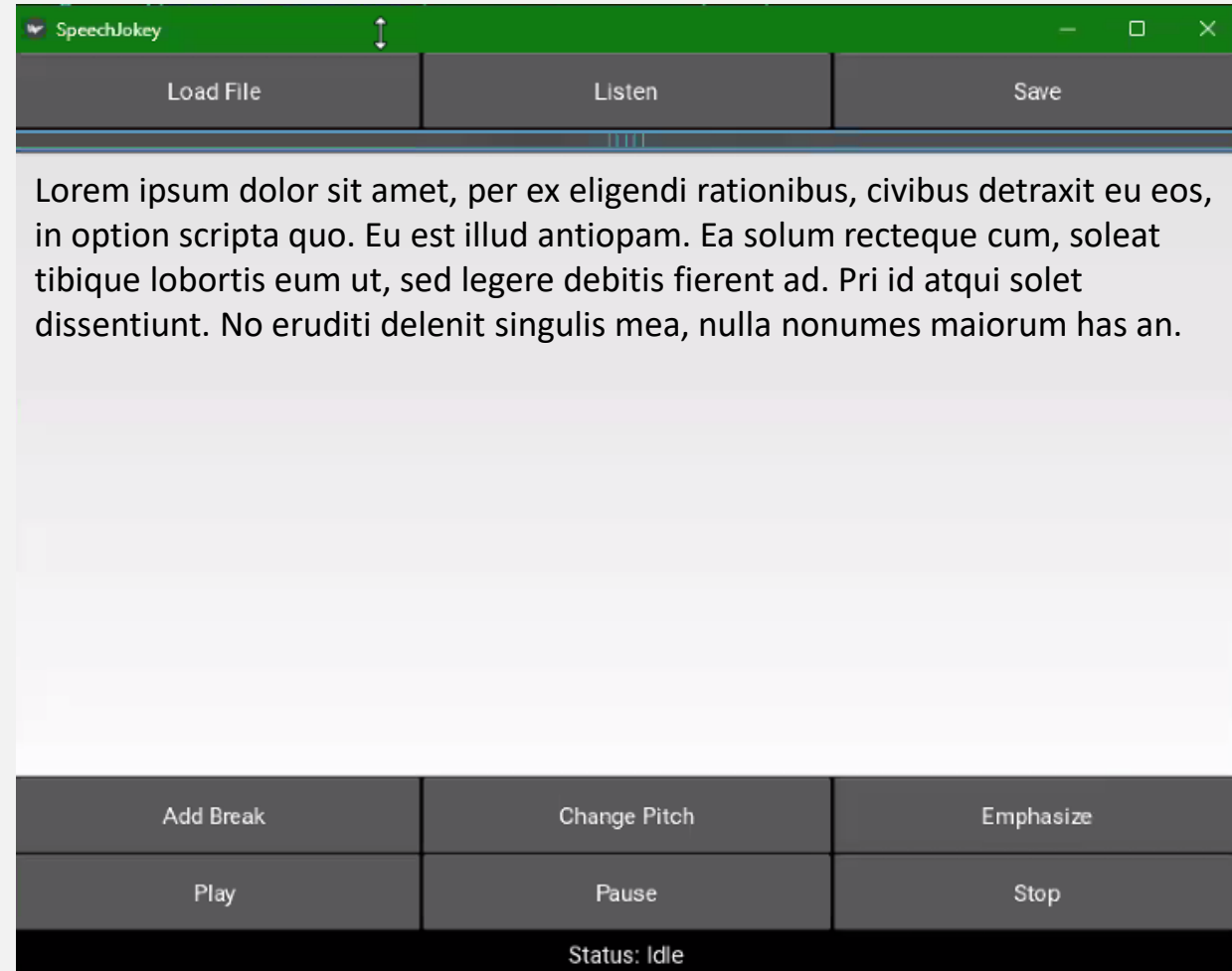
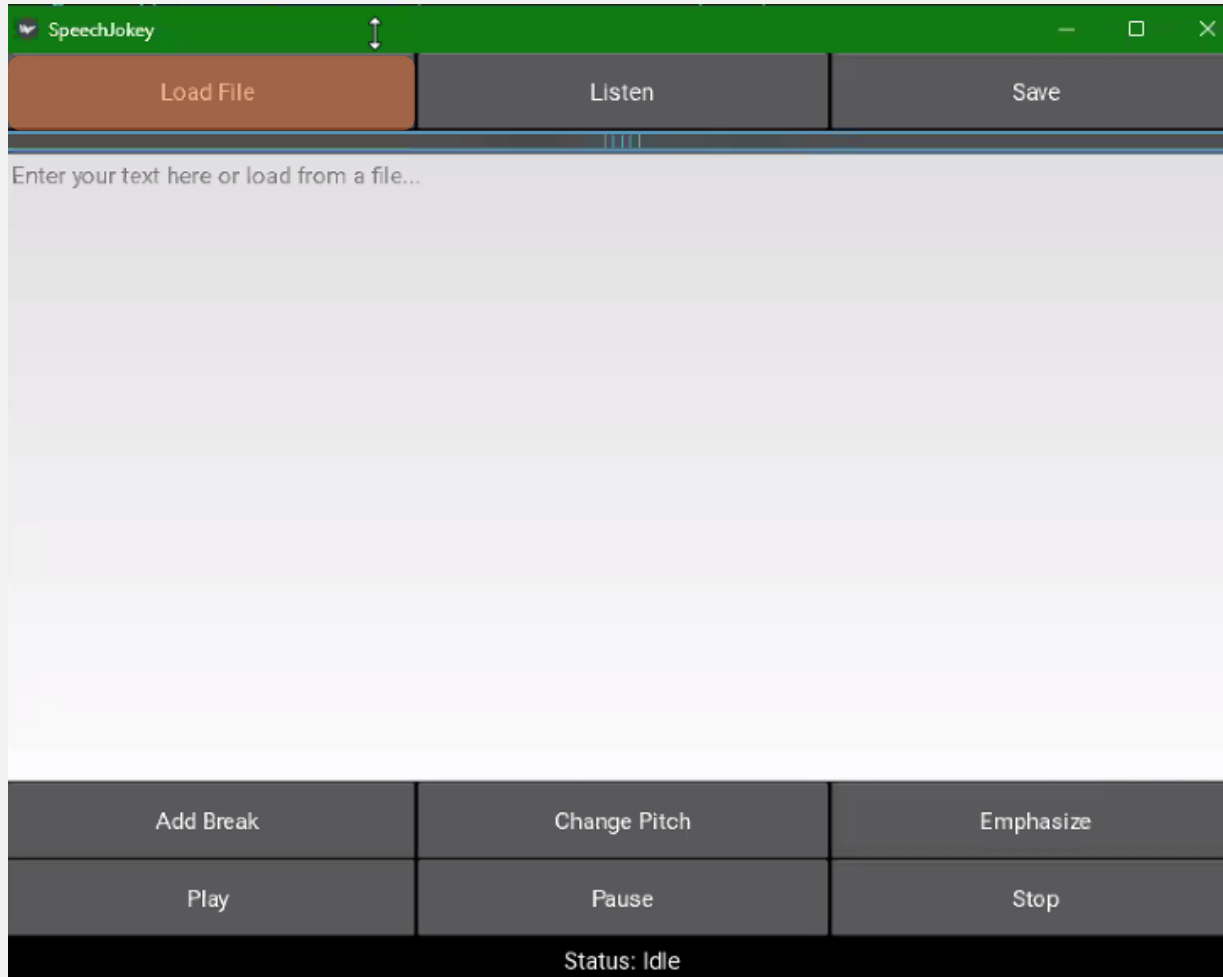


User interface overview

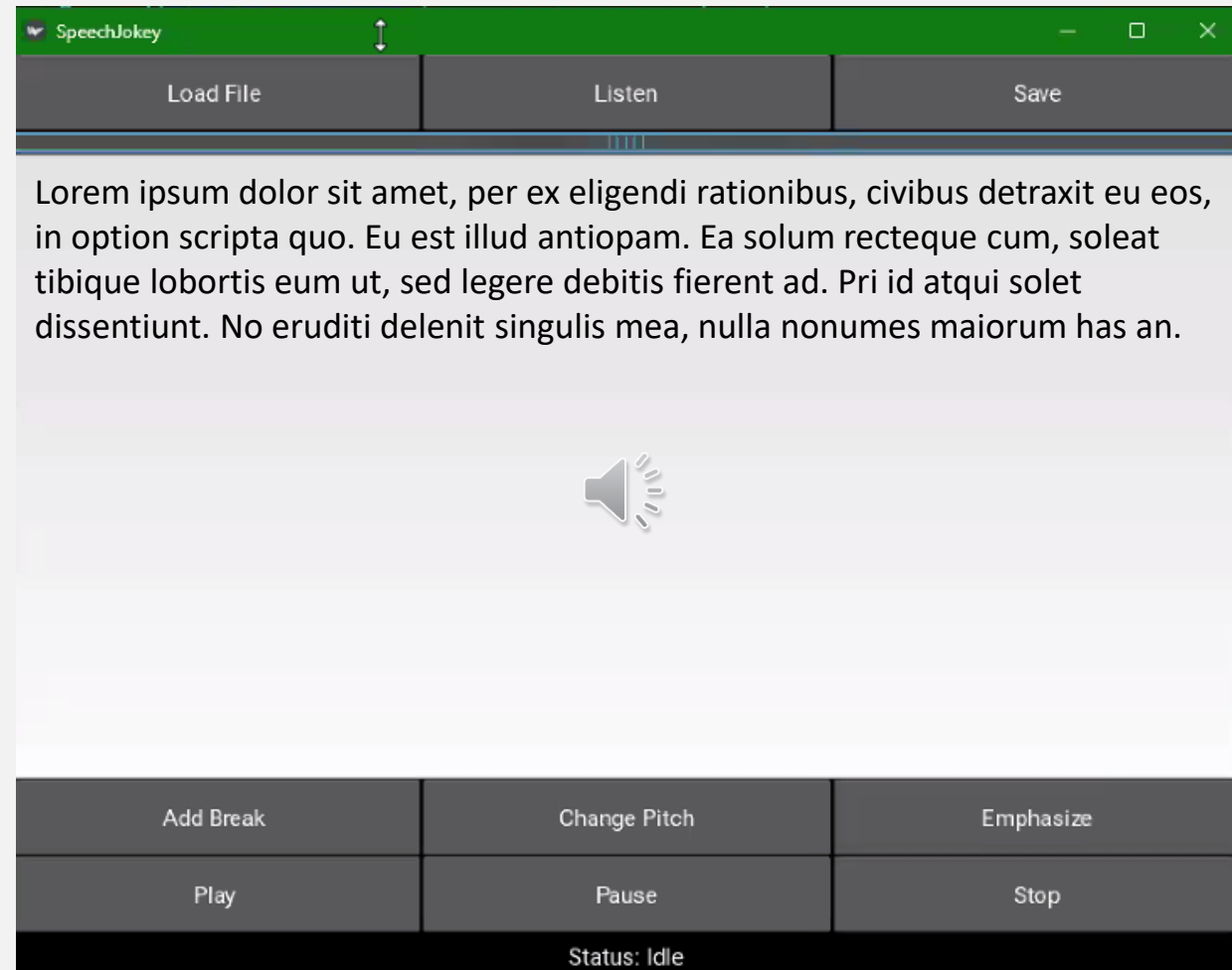
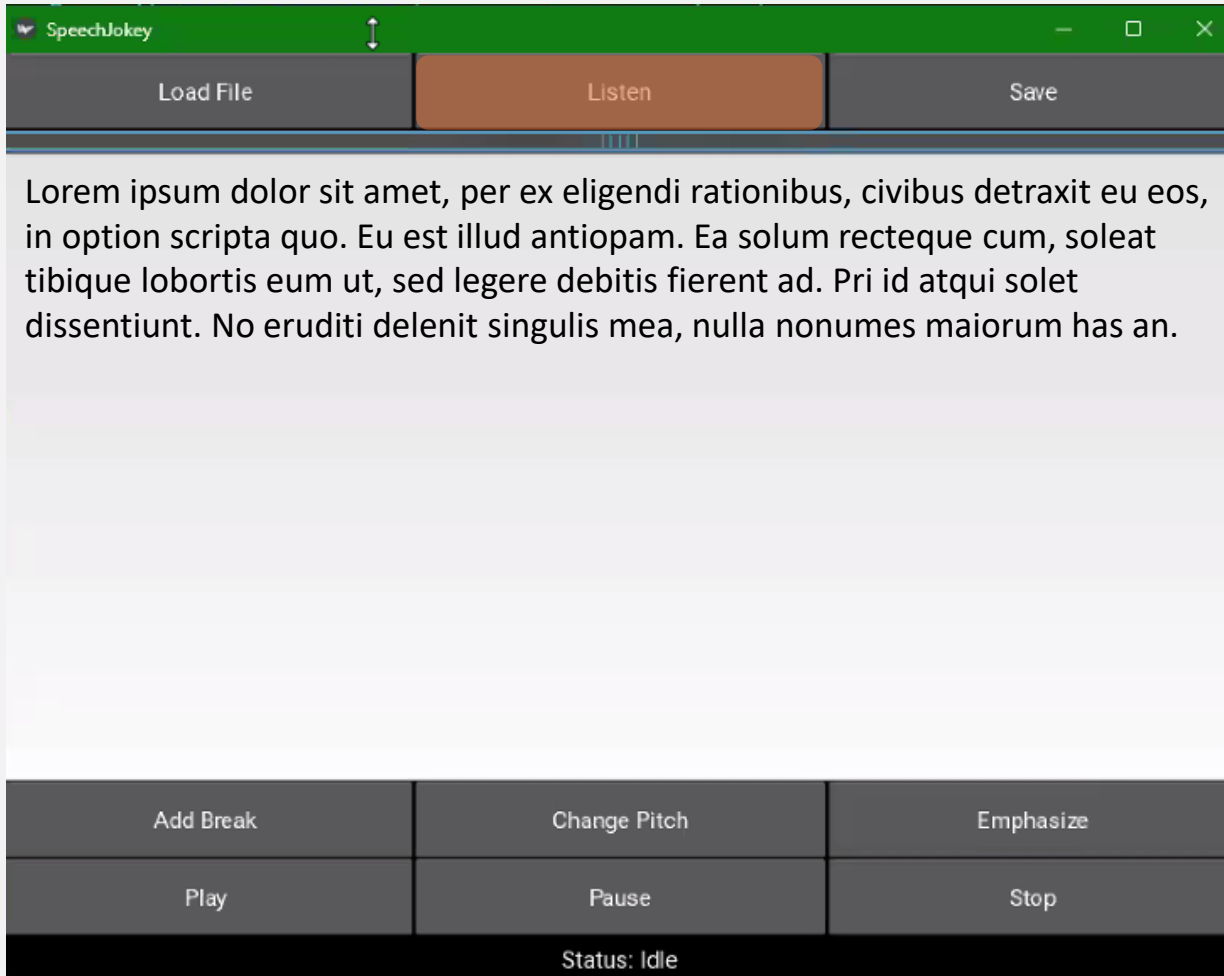


- **Main Screen:** Vertical layout featuring the central TextInput for user inputs.
- **Toolbar:** Upper BoxLayout with essential buttons: "Load File," "Listen," and "Save" for application control.
- **SSML Tags:** Bottom BoxLayout with buttons enabling the addition of SSML tags such as "Add Break," "Change Pitch," and "Emphasize."
- **Playback Control:** Last-row BoxLayout with "Play", "Pause", and "Stop" buttons for audio control.
- **Functionality:** Each button triggers specific Python functions (e.g., `load_file()`, `play_audio()`) associated with on press events.

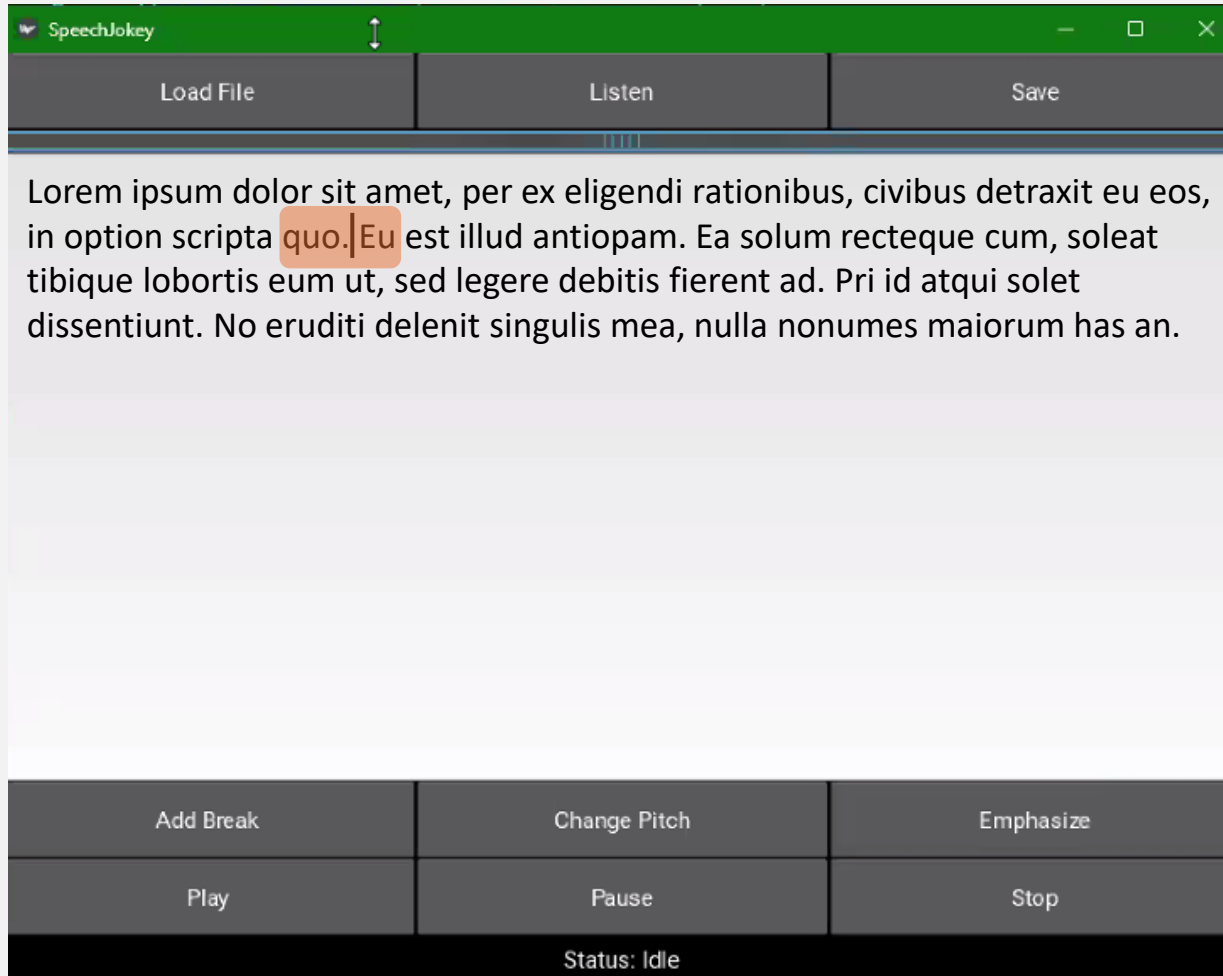
Load File



Listen button



Editing the text



The screenshot shows the SpeechJokey application window. The title bar is green with the text "SpeechJokey" and standard window controls. Below the title bar is a dark grey bar with three buttons: "Load File", "Listen", and "Save". The main text area contains the following text: "Lorem ipsum dolor sit amet, per ex eligendi rationibus, civibus detraxit eu eos, in option scripta quo. | Eu est illud antiopam. Ea solum recteque cum, soleat tibiue lobortis eum ut, sed legere debitis fierent ad. Pri id atqui solet dissentiunt. No eruditi delenit singulis mea, nulla nonumes maiorum has an." The word "quo." is highlighted in orange, and the cursor is positioned at the end of the word "Eu". Below the text area is a dark grey bar with three buttons: "Add Break", "Change Pitch", and "Emphasize". At the bottom is a black bar with three buttons: "Play", "Pause", and "Stop". The status bar at the very bottom displays "Status: Idle".

SpeechJokey

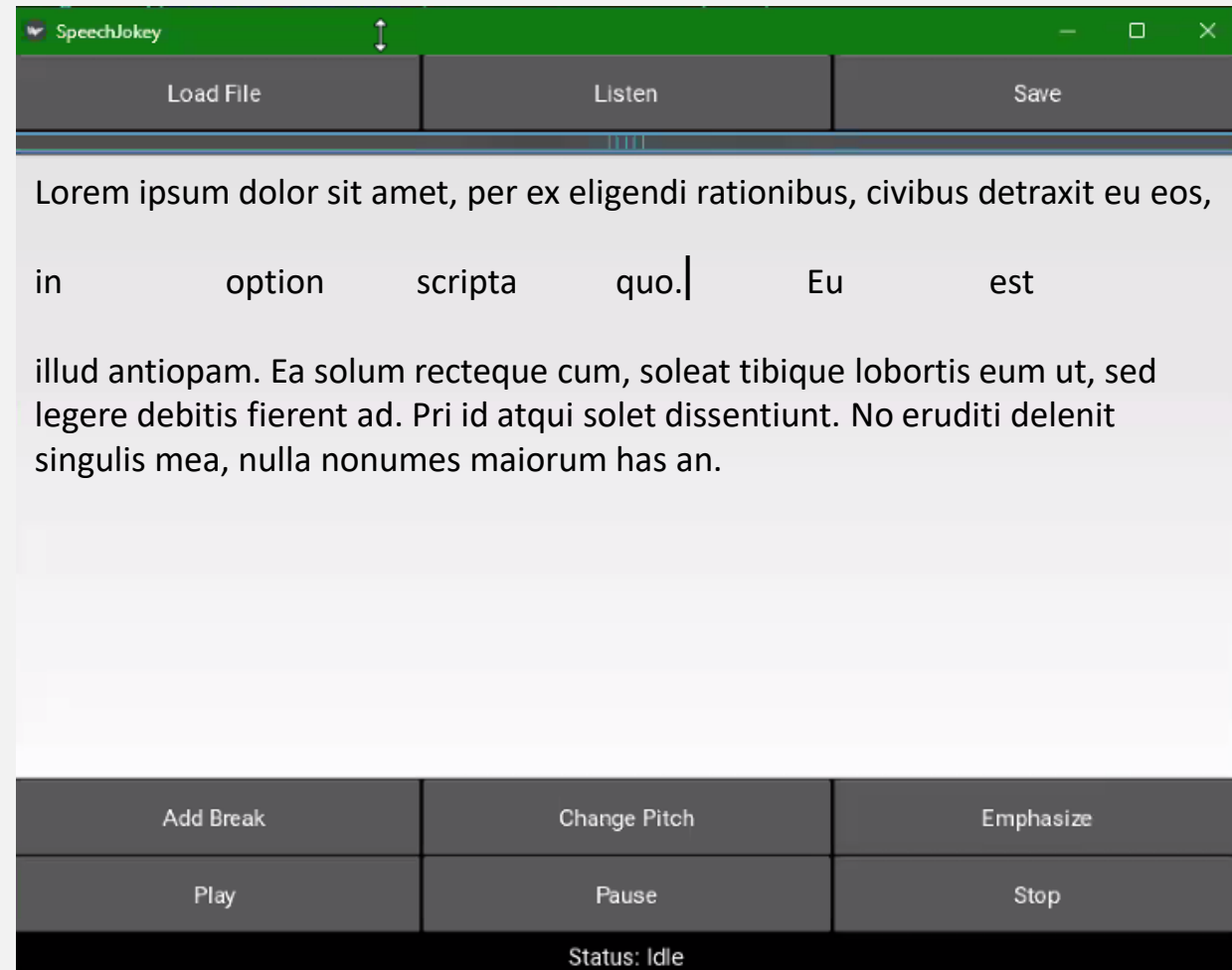
Load File Listen Save

Lorem ipsum dolor sit amet, per ex eligendi rationibus, civibus detraxit eu eos, in option scripta quo. | Eu est illud antiopam. Ea solum recteque cum, soleat tibiue lobortis eum ut, sed legere debitis fierent ad. Pri id atqui solet dissentiunt. No eruditi delenit singulis mea, nulla nonumes maiorum has an.

Add Break Change Pitch Emphasize

Play Pause Stop

Status: Idle



The screenshot shows the SpeechJokey application window. The title bar is green with the text "SpeechJokey" and standard window controls. Below the title bar is a dark grey bar with three buttons: "Load File", "Listen", and "Save". The main text area contains the following text: "Lorem ipsum dolor sit amet, per ex eligendi rationibus, civibus detraxit eu eos, in option scripta quo. | Eu est illud antiopam. Ea solum recteque cum, soleat tibiue lobortis eum ut, sed legere debitis fierent ad. Pri id atqui solet dissentiunt. No eruditi delenit singulis mea, nulla nonumes maiorum has an." The word "quo." is highlighted in orange, and the cursor is positioned at the end of the word "Eu". Below the text area is a dark grey bar with three buttons: "Add Break", "Change Pitch", and "Emphasize". At the bottom is a black bar with three buttons: "Play", "Pause", and "Stop". The status bar at the very bottom displays "Status: Idle".

SpeechJokey

Load File Listen Save

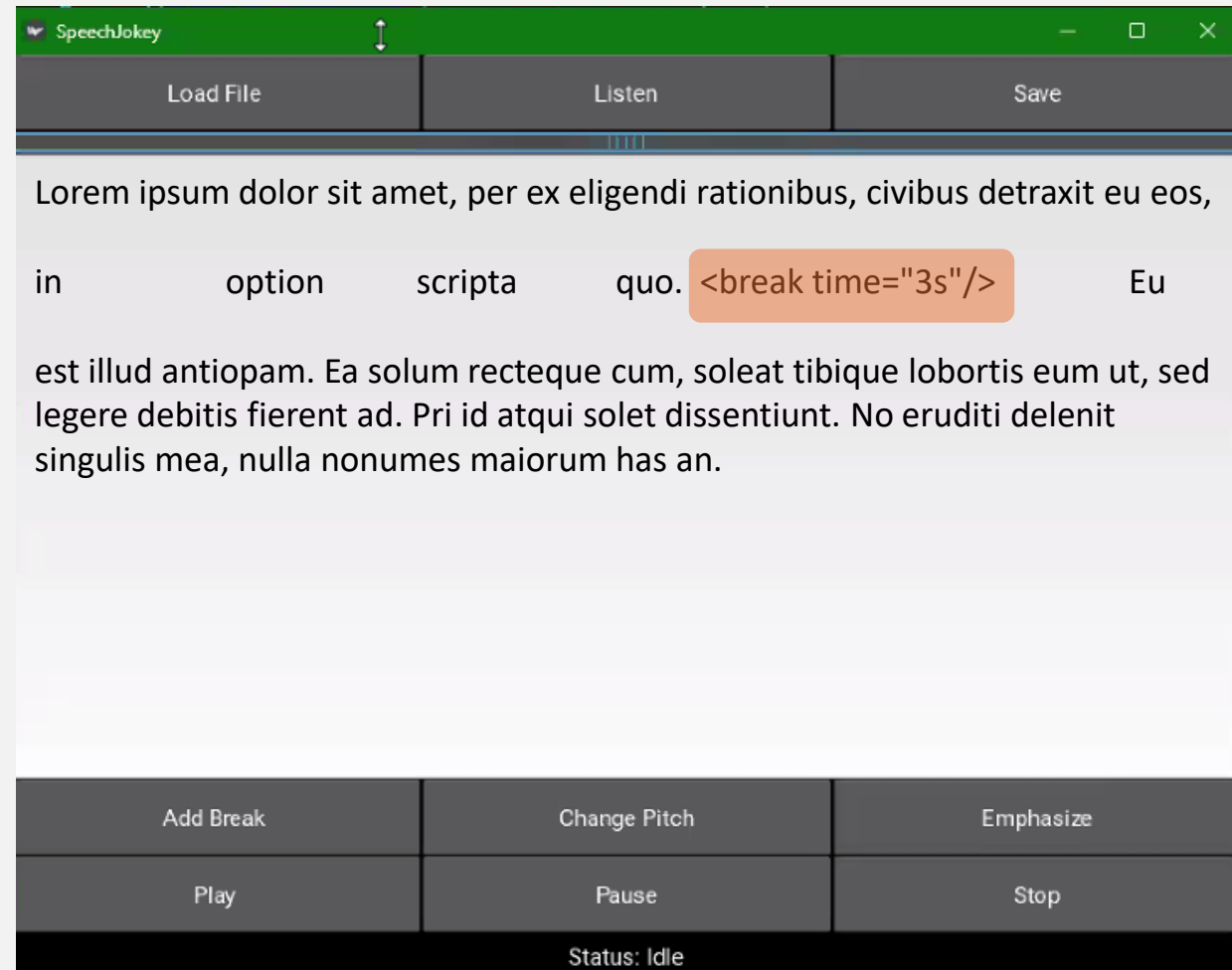
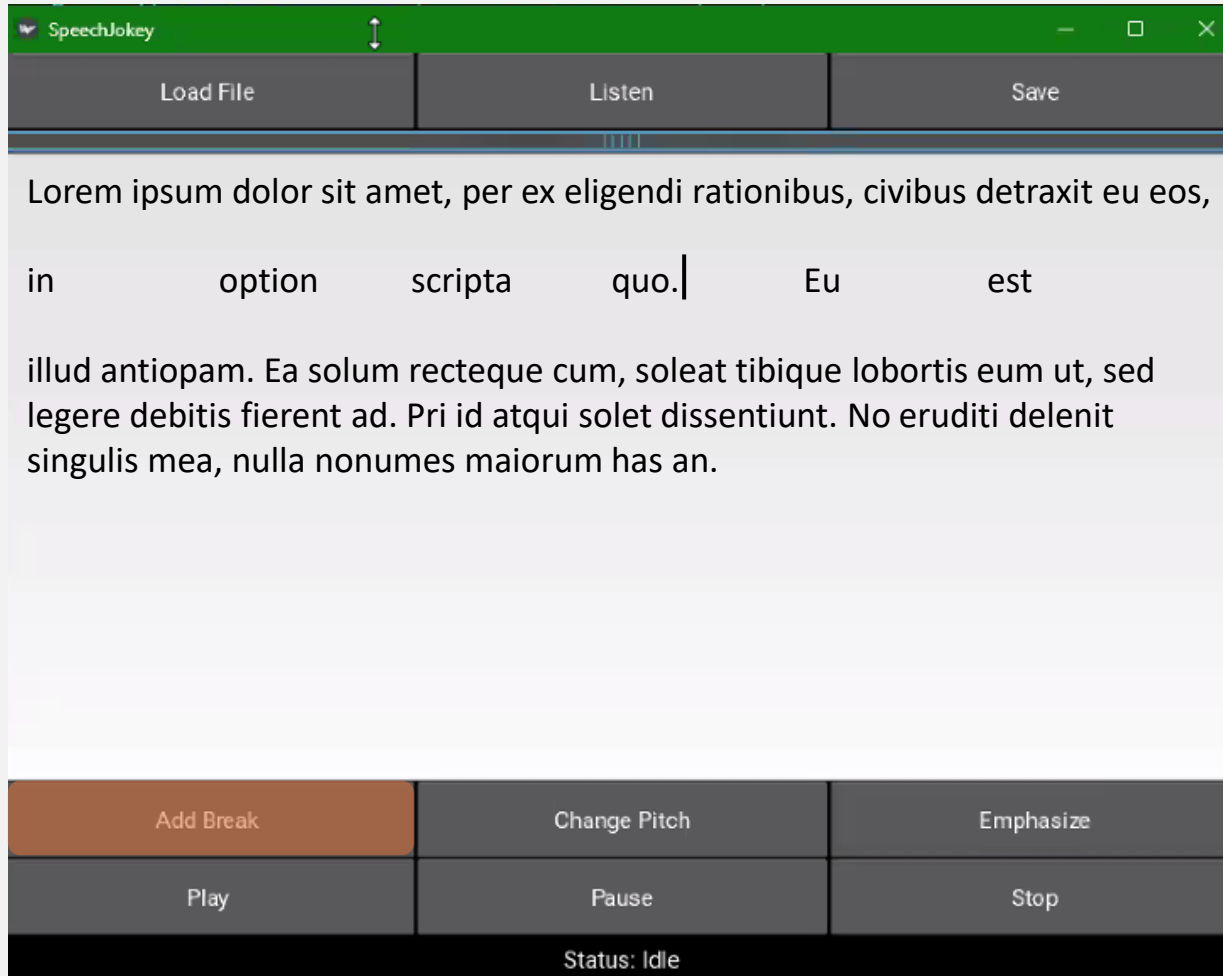
Lorem ipsum dolor sit amet, per ex eligendi rationibus, civibus detraxit eu eos, in option scripta quo. | Eu est illud antiopam. Ea solum recteque cum, soleat tibiue lobortis eum ut, sed legere debitis fierent ad. Pri id atqui solet dissentiunt. No eruditi delenit singulis mea, nulla nonumes maiorum has an.

Add Break Change Pitch Emphasize

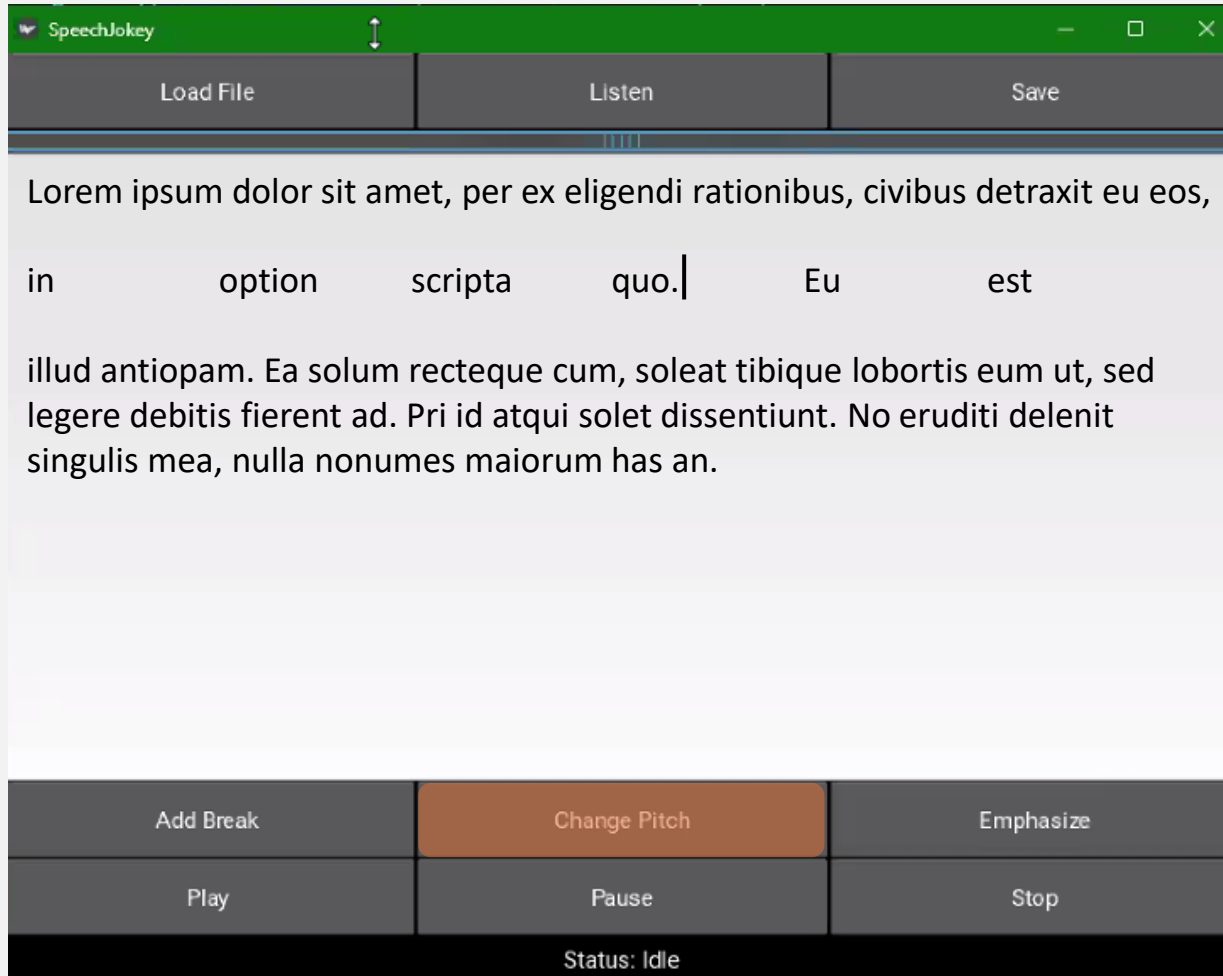
Play Pause Stop

Status: Idle

Add Break button



Change Pitch Button



SpeechJokey

Load File Listen Save

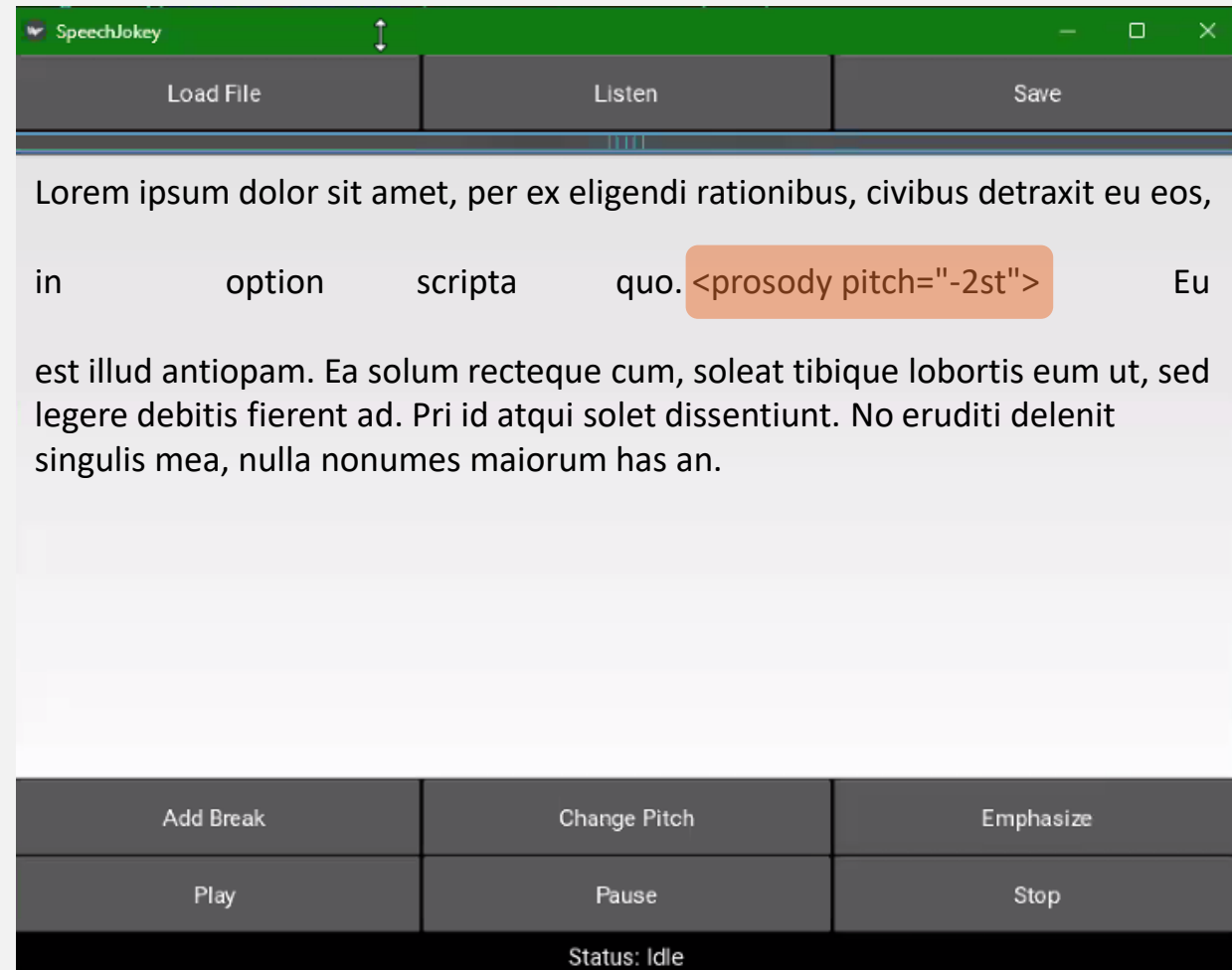
|||

Lorem ipsum dolor sit amet, per ex eligendi rationibus, civibus detraxit eu eos,
in option scripta quo. | Eu est
illud antiopam. Ea solum recteque cum, soleat tibi que lobortis eum ut, sed
legere debitis fierent ad. Pri id atqui solet dissentiunt. No eruditi delenit
singulis mea, nulla nonumes maiorum has an.

Add Break Change Pitch Emphasize

Play Pause Stop

Status: Idle



SpeechJokey

Load File Listen Save

|||

Lorem ipsum dolor sit amet, per ex eligendi rationibus, civibus detraxit eu eos,
in option scripta quo. <prosody pitch="-2st"> Eu
est illud antiopam. Ea solum recteque cum, soleat tibi que lobortis eum ut, sed
legere debitis fierent ad. Pri id atqui solet dissentiunt. No eruditi delenit
singulis mea, nulla nonumes maiorum has an.

Add Break Change Pitch Emphasize

Play Pause Stop

Status: Idle

Emphasize Button

SpeechJokey

Load File Listen Save

|||

Lorem ipsum dolor sit amet, per ex eligendi rationibus, civibus detraxit eu eos,
in option scripta quo. | Eu est
illud antiopam. Ea solum recteque cum, soleat tibi que lobortis eum ut, sed
legere debitis fierent ad. Pri id atqui solet dissentiunt. No eruditi delenit
singulis mea, nulla nonumes maiorum has an.

Add Break Change Pitch Emphasize

Play Pause Stop

Status: Idle

SpeechJokey

Load File Listen Save

|||




Lorem ipsum dolor sit amet, per ex eligendi rationibus, civibus detraxit eu eos,
in option scripta quo. **<emphasis level="strong">**
Eu est illud antiopam. Ea solum recteque cum, soleat tibi que lobortis eum ut, sed
legere debitis fierent ad. Pri id atqui solet dissentiunt. No eruditi delenit
singulis mea, nulla nonumes maiorum has an.

Add Break Change Pitch Emphasize

Play Pause Stop

Status: Idle

SSML TAGS

SSML	EMOJIE
<peak>	
<break> (time, strenght)	
<emphasis> (strong, moderate, none, reduced)	

The emojis will appear in the text input instead of the SSML tags as a visual and more intuitive reminder.

How to contribute as a developer

Git Hub



Collaborative
environnement thanks to
branches and merges



Easy distribution and
sharing with other users



Issue tracking and history
of all modifications



Assign tasks and monitor
the status of activities



Python



The project is based on Python 3.11, but it also supports lower version down to 3.9.



Python is a programming language well-suited for text-to-speech (TTS) applications.



Python is a versatile language that adapts to a wide range of development scenarios.



Extensive ecosystem of libraries, frameworks, and tools that enhance Python's capabilities.



Poetry



Poetry simplifies the process of managing dependencies



Poetry allows us to specify version constraints ensuring compatible library versions.



Poetry provides a unified tool for dependency management, packaging, and versioning



KIVY



The user interface is build in the Kivy framework:

1.4 Version is used



Offers a wide range of widgets and graphical controls to facilitate the development of intuitive user interfaces.



Kivy is an open-source Python framework for developing cross-platform applications.



Kivy apps work everywhere, either on Windows, Linux, Android, macOS



Visual Studio Code



Provides built-in support for numerous programming languages



Provides built-in support for numerous programming languages



VS Code is a code editor. It allows you to write code and to collaborate with



VS Code has an intuitive user interface and extension tools



Steps

- Installation of Python 3.11 (or lower versions until 3.9)
- Installation of Visual Studio Code (VSCode) or other source-code editor to run the code
- Clone the github repository *assist-heidi-speech-jokey* on VSCode
<https://github.com/HackXlt/assist-heidi-speech-joker>
- Install Poetry
- Install requirements using poetry
- Create virtual environment using poetry
- Edit the code
- Commit the changes

How to run/install the application

Executable

Running the application (Any OS / Development)

To run the application, execute the following command in the root of the project:

```
poetry run python src/main.py
```

Building the application executable (Windows / Local Development)

To build the application, execute the following command in the root of the project:

(You might wanna grab a coffee while running this)

```
poetry run pyinstaller src/main.py --onefile --name SpeechJokey
```

The created build application specification `SpeechJokey.spec` can now be found in the root of the project. This file needs to be modified according to the following steps:

1. Import kivy dependencies at the top of the file: `from kivy_deps import sdl2, glew`
2. Add source tree after `COLLECT(exe, : Tree('src\\'),`
3. Add source dependencies after `a.datas, : *[Tree(p) for p in (sdl2.dep_bins + glew.dep_bins)],`

After these modifications, the application can be finalized by running:

(Should be very quick after the initial build)

```
poetry run pyinstaller SpeechJokey.spec
```

Inside the `dist` output folder a folder with the name `SpeechJokey` can be found. This folder contains the final `.exe` build of the application.

For a detailed step-by-step guide on how to build a Kivy application, see [this written tutorial](#). (Keep in mind that the tutorial doesn't use `poetry`, so any command should be preceded by `poetry run`)

- All the information on how to build and run the application can be found in this github repository: <https://github.com/HackXIt/assist-heidi-speech-jokey>
- It's possible for the user to download a single `.exe` file to use the application without any further modifications.



Live preview part

A thick, hand-drawn style orange line that underlines the text "Live preview part". It starts at the left edge of the text and extends to the right, ending slightly past the end of the text.

Statements about project state

<https://github.com/HackXIt/assist-heidi-speech-joke>

User stories

The repository contains:

- README.md file for information about the project and how to use the application
- Written documentation on the intended roadmap
- Milestone of the project
<https://github.com/HackXIt/assist-heidi-speech-jokey/milestone/1>
- Further diagrams containing informations about procedures and features to implement

The end

Thank you for your attention