Report for Voice Recognition Application

1752095 范思琪

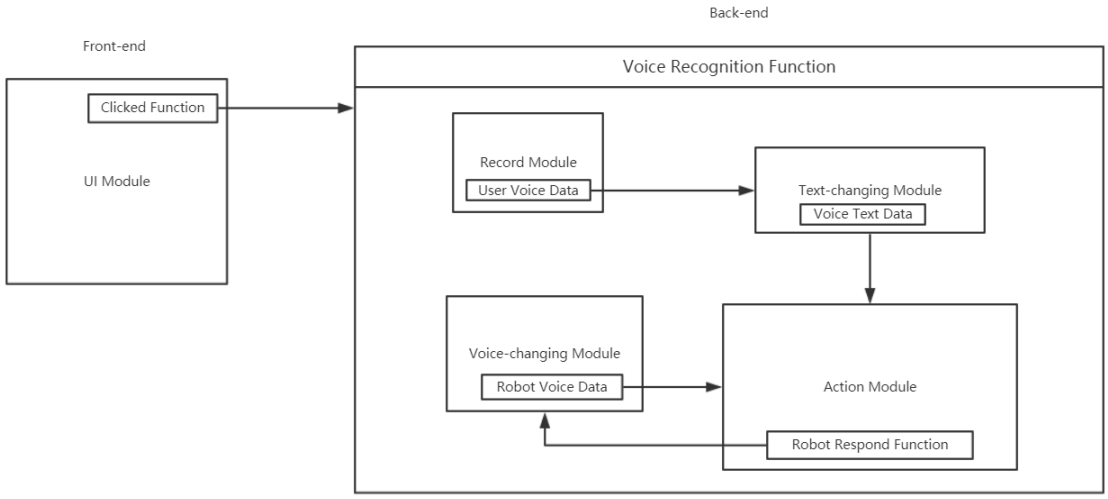
1752662 龙　宇

目录

1. Description

1.1. Structures

The voice recognition application for Windows can help users carry out actions quickly. In our voice recognition application, there are five modules constituting the front-end and back-end. The following image is the structural diagram of the five modules.



Seen from the image above, we can easily understand the structure among modules. UI module will show window of the application as soon as the program is started. If users' click-action is caught, the clicked function will be called. The function will call the voice recognition function, which enables voice interaction with the user.

Firstly, the voice recognition function will run into the record module, where a user voice data will be recorded and sent to text-changing module.

Secondly, the text-changing module will change the voice data into text data. The action module will make action according to the text data sent by text-changing module. If robot response function in action module is called, the voice-changing module will provide a piece of robot voice data back to action module according to the given robot response.

In conclusion, the application will make certain action and give spoken message in response of each piece of voice command the user say.

1.2. Modules

UI Module

This module shows the main window of the application and provide a clicked function for users' clicking action. The clicked function ties to voice recognition function.

Record Module

This module is part of the voice recognition function. The users' voice commands are recorded here and saved as a wave file, which will be deal afterwards.

Text-changing Module

This module is part of the voice recognition function. It will receive the wave file sent by record module and change it into words. The words are saved in a string.

Action Module

This module is part of the voice recognition function. It will receive the string sent by text-changing module. The string is classified to different types by key words. Different type of commands has different kind of reaction. When one of the actions is called, the application will create a message.

Voice-changing Module

This module is part of the voice recognition function. It will receive the message sent by action module and change it into a piece of voice saved in a wave file. Then, it will play the file out aloud.

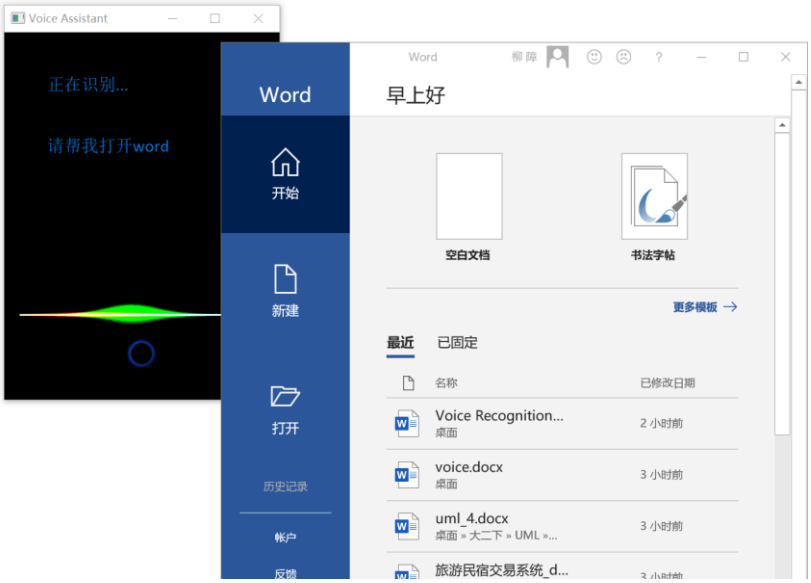
2. Implemented Requirements

2.1. Open Notepad

If "Notepad" in Chinese could be found in your command, the application will open notepad for you while giving a spoken message.

2.2. Open Word

If “Word” could be found in your command, the application will open word for you while giving a spoken message.

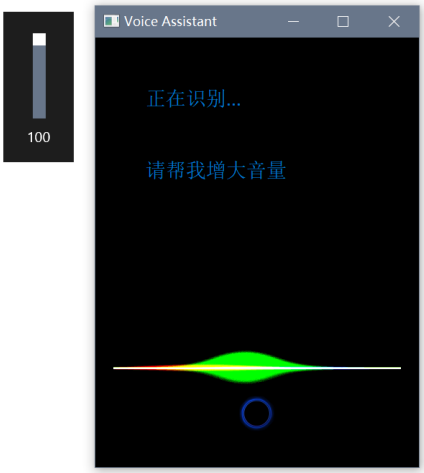


2.3. Play Music

If “Music” in Chinese could be found in your command, the application will open music player for you while giving a spoken message.

2.4. Increase the Volume

If “increase the volume” in Chinese could be found in your command, the application will increase the volume to highest for you while giving a spoken message.

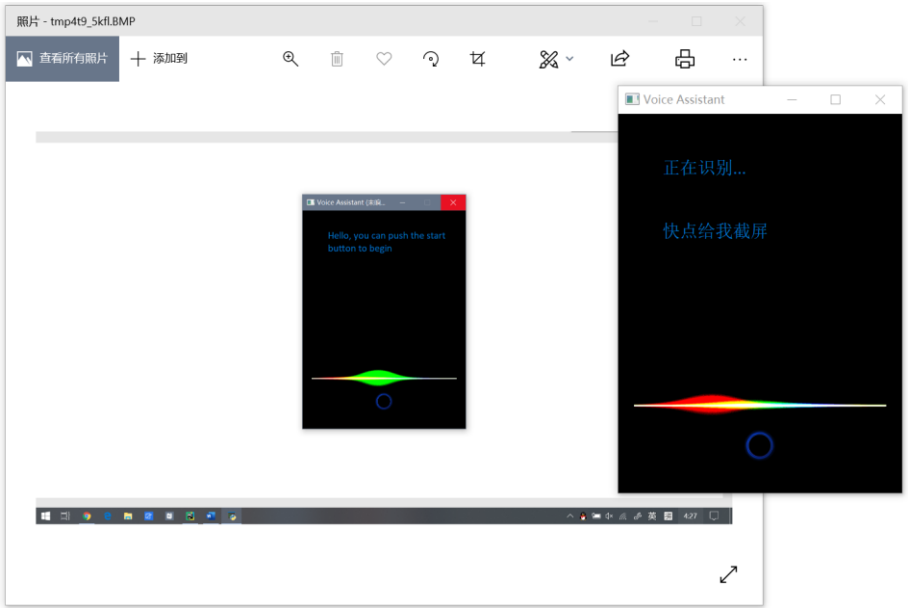


2.5. Decrease the Volume

If “decrease the volume” in Chinese could be found in your command, the application will decrease the volume to lower for you while giving a spoken message.

2.6. Screenshot

If “screenshot” in Chinese could be found in your command, the application will screenshot as well as show the picture for you while giving a spoken message.

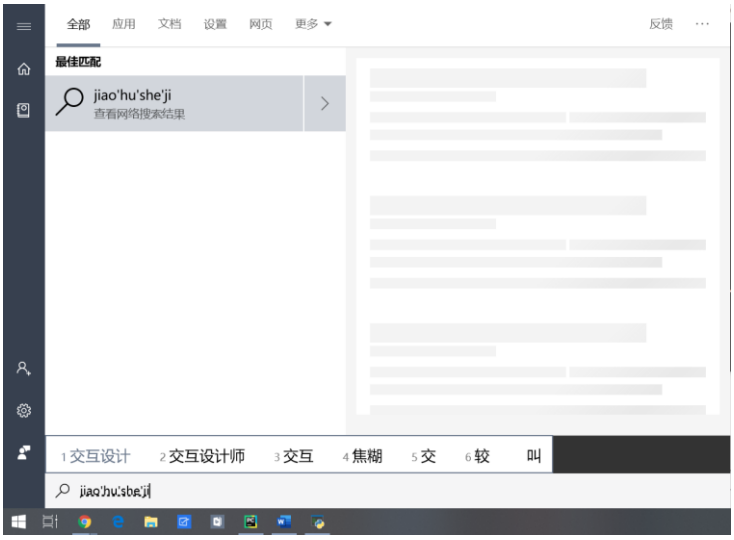


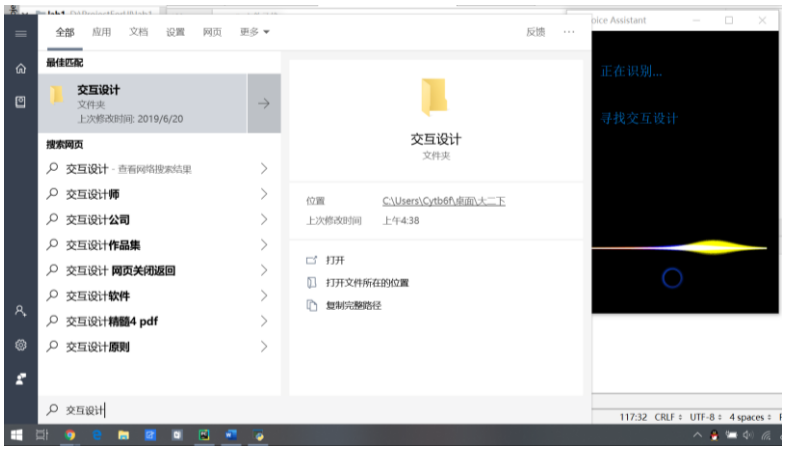
2.7. Web Search

If “search” in Chinese could be found in your command, the application will search the words after “search” on the internet for you while giving a spoken message.

2.8. Lookup in Computer

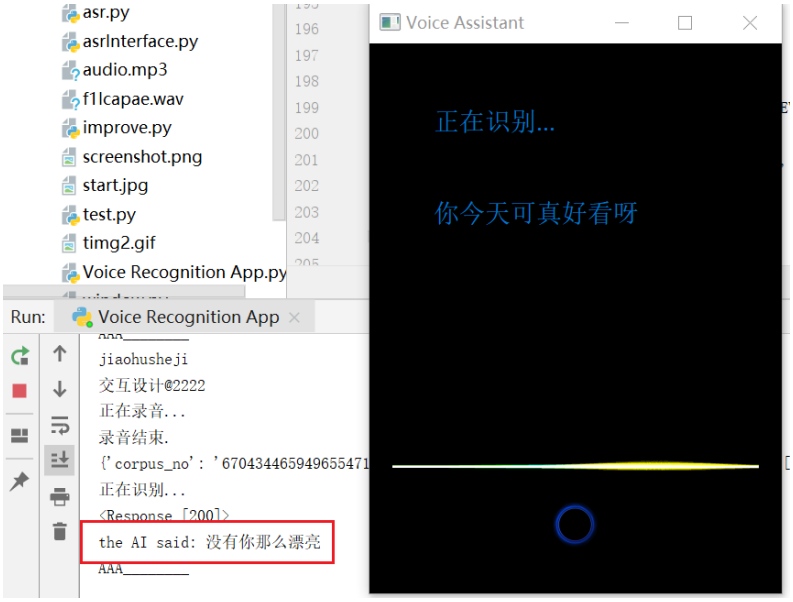
If “lookup” in Chinese could be found in your command, the application will lookup the words after “lookup” for you on computer while giving a spoken message.





2.9. Chat with Robot

If none of the key words is in your command, the robot will give a response and talk with you.



3. Advantages and Disadvantages

3.1. Advantages

- The application will give the user an answer whatever the user says.
- The interaction between users and robot will bring a better experience for users.

- The function of searching on the internet and looking up on the computer will better cater to the needs of users.
- The interaction is only available after the button pressed. This can prevent false operations.

3.2. Disadvantages

- The connection to Baidu Aip will drop occasionally.
- Some function only available for commands in Chinese.
- The avoidance of false operations is realized by button. A better option is by automatic detection of ambient sounds.
- Fewer error commands can be accommodated.
- The interface is a little poor.
- Feedback output on the interface is not real-time.
-

4. Improvement

Improvement of front-end

- A better interface can be design.
- The show of users' command could be able to edit like Cortana. In this way, users can correct the errors in speech recognition.

Improvement of back-end

- A larger optional instruction set should be enlarge, to cater to users' commands in English and commands of doing more things.
- More error commands should be found and deal.
- Find a way to strong the connection to Baidu Aip.