Scene description

1. Experimental purpose

This chapter learns how to combine the online large language model to realize the process of scene description with the robot dog. This case only requires audio equipment and camera equipment.
Notes:

1. **Before running this case, you need to close the startup program**, please refer to the Raspberry Pi system configuration section**9.Open and close the APP control program**This tutorial ends the startup program.
2. You need to fill in the API_KEY of the large model, please refer to the operation method of **AI large model section "1. Prerequisites for using the large model"**.

2. Experimental steps

(This tutorial takes the Chinese version of the effect diagram as an example)

1. Terminal input

```
cd /home/pi/DOGZILLA/Samples/4_Big_Modle
python3 pic_comprehension/sp_AI_Image_en.py
```

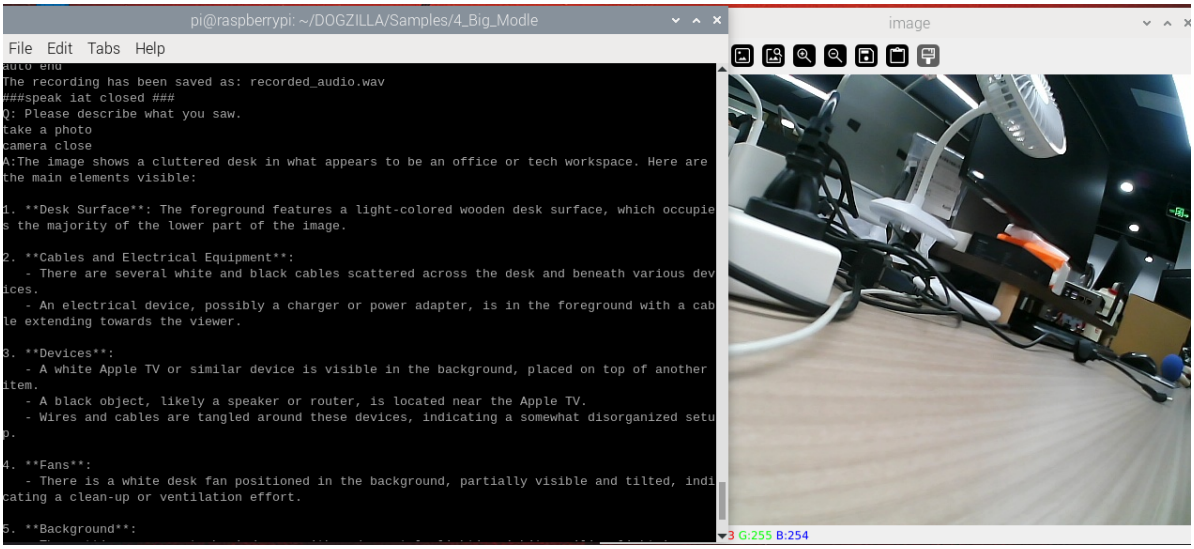
2. Wake-up operation
   Wake-up word: **hello yahboom**

```
pi@raspberrypi:~/DOGZILLA/Samples/4_Big_Modle $  python3 pic_comprehension/sp_AI_Image.py
serial /dev/myspeech open
Waiting for keyword...
```

After waking up, you will hear a "ding" sound, and then you can express your thoughts to the robot dog.

```
pi@raspberrypi:~/DOGZILLA/Samples/4_Big_Modle $ python3 pic_comprehension/sp_AI_Image.py
serial /dev/myspeech open
Waiting for keyword...
Keyword detected: 05.Jun 2025 09:52:56
Playing WAVE './ding.wav' : Signed 16 bit Little Endian, Rate 16000 Hz, Mono
```

3. The robot dog will be processed by the big model to understand the customer's ideas, and then feedback the corresponding text results and audio playback.



4. At this point, the interactive process ends. If you need to express your ideas again, just wake up again.
   At the same time, in this case, you can interrupt the previous interactive dialogue by waking up during the **audio broadcast stage**.

3. Analysis of the main source code of the experiment

In the path "/home/pi/DOGZILLA/Samples/4_Big_Modle/pic_comprehension/sp_AI_Image_en.py" is a main function entry

```
# Main function flow
    while True:
        if detect_keyword():
            cv2.destroyAllWindows()
            os.system("pkill mplayer")
            time.sleep(.2)

            start_recording()
            content = rec_wav_music_en()

            if content != "":
                print("Q:"+content)

                take_photo()
                time.sleep(1)

                mymytext = dogGPT_Image_en(content+'reply English')#image
description

                time.sleep(1)

                print("A:"+mymytext)

                try:
                    response = mymytext
                    tts_thread = threading.Thread(target=Speak_Vioce)
                    tts_thread.daemon = True
                    tts_thread.start()

                except:
                    pass
```

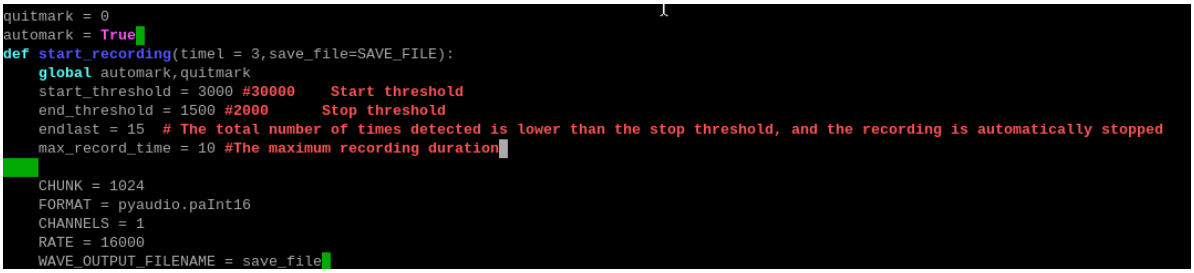
```
            if content == 0:
                break

        time.sleep(0.1)
```

1. Program flow: detect wake-up word->listen to expression semantics and take a picture->understand the big model->feedback answer
2. xinghou_Image: This is an interface combined with the big model. The big model used in this function is **OpenRouter**
   If you want to change to another big model, you can refer to **xinghou_UltraAPI.py**, or write a py file according to the source code provided by the big model you want to use. For example, the file name is: **mychatgpt.py**, then import this file at the beginning of the sp_AI_Image_en.py file, writing method: form mychatgpt.py import *,
   Then replace the xinghou_Image interface provided by the routine with the interface function in mychatgpt.py that can call the large model. This involves a lot of DIY operations, and it is not recommended for novices to replace the model.
3. If you want to change the threshold for recording start and the duration of recording, you can change this file. Enter in the terminal

```
nano /home/pi/DOGZILLA/Samples/4_Big_Modle/audio.py
```

Change the recording part of this file as shown in the figure below

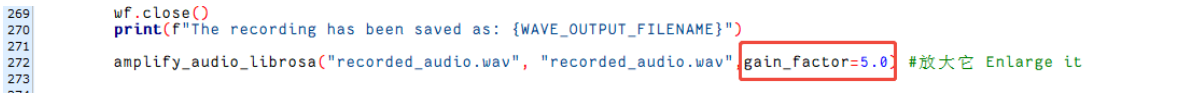


Parameter meaning:

- start_threshold = 3000 #Start recording when a sound louder than this value is detected. This value changes according to the environment
- end_threshold = 1500 #Sounds lower than this value are detected. This value changes according to the environment
- endlast = 15 #Stop recording when the number of sounds lower than end_threshold is detected. Here it is 15 times
- max_record_time = 5 #The duration of the recorded audio. Here it is 5
  Note: start_threshold must be greater than end_threshold (start_threshold>end_threshold)
  In general, the ideal value of end_threshold is half of start_threshold, which can be adjusted according to your own environment.

4. If you feel that the recorded audio cannot be recognized by the online large model because the sound is too small, you can adjust the value here to amplify the recorded audio.
   Terminal input

```
nano /home/pi/DOGZILLA/Samples/4_Big_Modle/audio.py
```



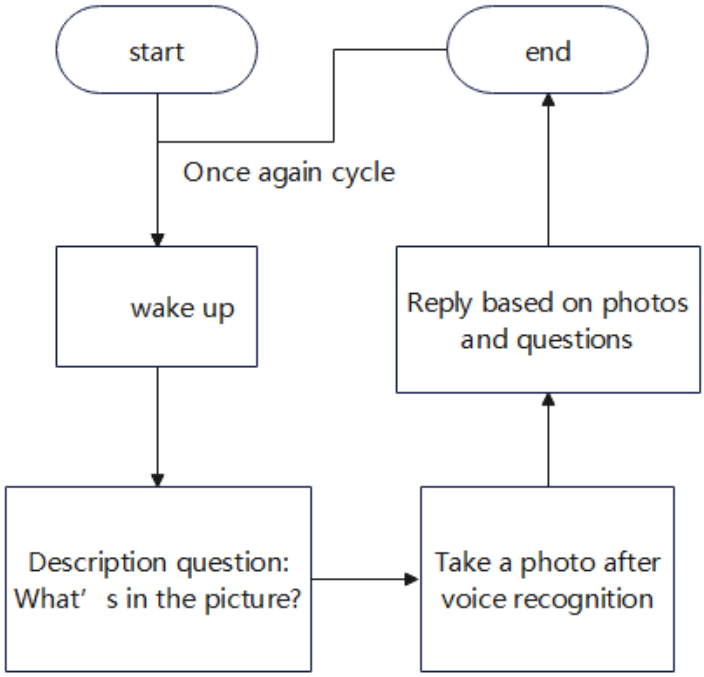
Here it is amplified 5 times, here you can make an adjustment according to the distance of the sound source.
Note: If the distance is too far to record audio at all, adjusting the parameters here will be meaningless.
It is recommended that the distance from the sound source of the recorded audio should not be greater than 1.5m.

5. In the path of "/home/pi/DOGZILLA/Samples/4_Big_Modle/pic_comprehension", the directory structure description is as follows:

```
├── rec.jpg #Photographed image
├── sp_AI_Image_en.py #English moderator program interface
├── sp_AI_Image.py #Chinese moderator program interface
├── xinghou_ImageAPI.py #iFlytek Spark large model interface
├── xinghou_speak_iat.py #iFlytek Spark platform speech recognition interface
└── xinghou_tts.py #iFlytek Spark platform synthesis audio interface
```

4. Experimental summary

Based on the above description, the flowchart of this case is as follows:



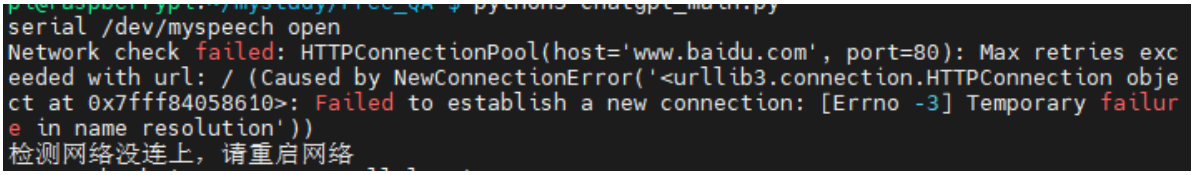
If you don't know what to express in this case, here are some reference examples
Example:

1. Request + content type For example: Please describe what is in the picture
2. You can also ask some scene-related questions For example: How many colors are there in the picture?

And so on, you can use your imagination, and I won't elaborate here. The example questions of free dialogue are also applicable here.

Notes

1. If this error occurs when the program starts, you can press "ctrl+C" to end the program and then restart the program.



2. If you want to terminate this case, press "ctrl+C" to end the program.