

Control Thymio Robot via Voice Commands

Using vosk Speech to Text Library

By Ahmad Agbaryah

Abstract

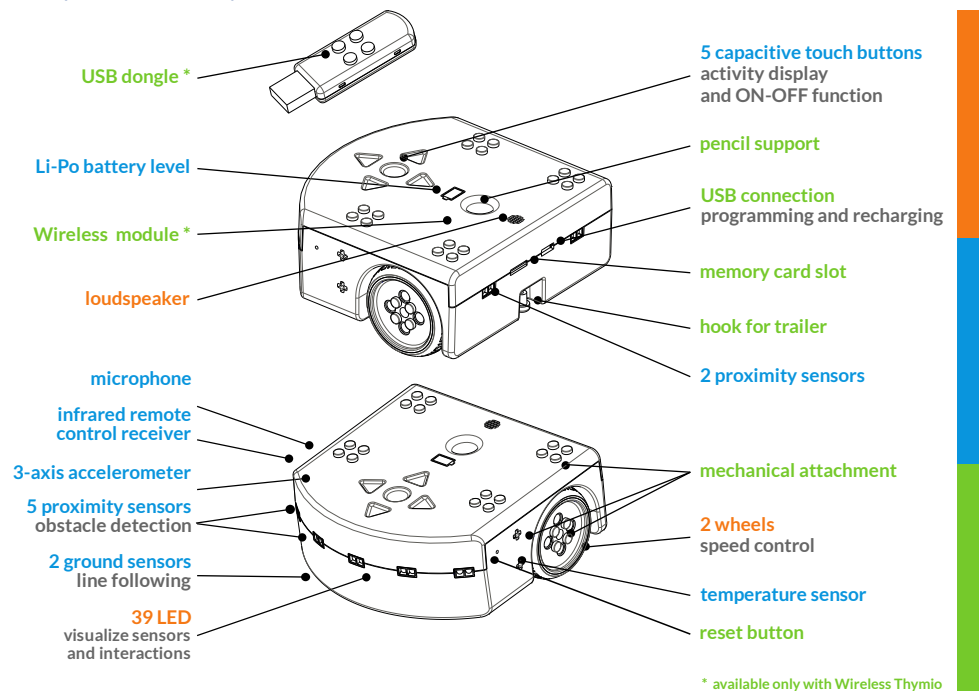
This proof of concept combining a wireless Thymio II with a python software on computer to control the movements of the robot. Speech to text is performed on the computer, using vosk library with an offline model. Once the python application recognizes new voice, it turns it to text and based on the result, the python app decides what command to send to the Thymio robot.

First, let's have some base knowledge on the pieces of the proof of the concept.

What Thymio is?

Thymio is an open-source educational robot designed by researchers from the EPFL, in collaboration with ECAL, and produced by Mobsya, a nonprofit association whose mission is to offer comprehensive, engaging STEAM journeys to learners of all ages. For more information, visit the official web site at: <https://www.thymio.org/>

What is Thymio composed of?¹



What is VOSK?

Vosk is a speech recognition toolkit. The best things in Vosk are:

- 1) Supports 20+ languages and dialects - English, Indian English, German, French, Arabic and many more (for full list of supported languages, visit the official site: <https://alphacephei.com/vosk/>).
- 2) Works offline, even on lightweight devices - Raspberry Pi, Android, iOS
- 3) Installs with simple pip3 install vosk
- 4) Provides streaming API for the best user experience.

¹ Image from: <http://wiki.thymio.org/en:thymiospecifications>

What is Needed for the Demo?

To duplicate the demo on your end, you need the following:

- 1- Wireless Thymio II (with wireless dongle)
- 2- Laptop (With microphone)

Software Setup

Follow the following steps to have the needed software pieces:

- 1) Install python v3 on your computer. You can find more details at the Python official site:

<https://www.python.org/downloads/>

- 2) Install the Thymio Suite on your computer. More details at the Thymio official site at:

<https://www.thymio.org/download-thymio-suite/>

- 3) Prepare Python environment:

- On your command line, type the following command (specify your own path):

```
python3 -m venv c:\work\thymio
```

- Activate the environment by typing the following command on your command shell:

```
C:\work\thymio\Scripts\Activate.bat
```

- Install required libraries by typing the following command:

```
(thymio) c:\work\thymio>pip install vosk sounddevice tdmclient
```

The first library, vosk, is the responsible for speech to text recognition (offline)

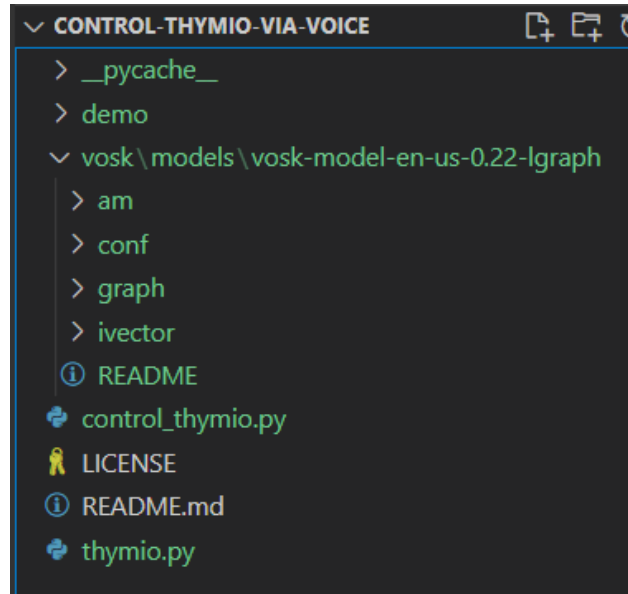
The second one, sounddevice, wraps the audio devices such as microphone and convert audio streamed from the microphone to raw data. Later, this raw data is fed into vosk model to convert it to text.

The last one is the Python Thymio wrapper to control the robot.

- Download proof of concept code from github to your local disk, e.g.,
c:\work\thymio\src\control-thymio-via-voice. Github code:
<https://github.com/ahmad081177/control-thymio-via-voice>
- Download the VSOK model for offline recognition. Visit the page:
<https://alphacephei.com/vosk/models> and pick any model you prefer. In my proof of concept, I've used both models: vosk-model-en-us-0.22 and vosk-model-en-us-0.22-lgraph.

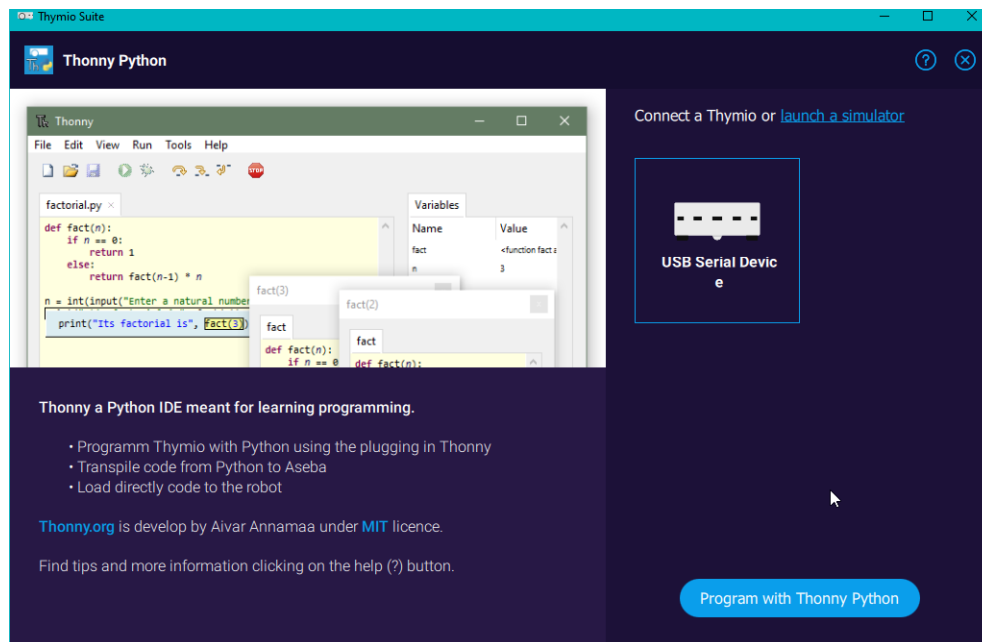
Place the model in your local disk, e.g., in c:\work\thymio\src\control-thymio-via-voice\vosk\models (If you place it otherwise, you may reflect the new path in control_thymio.py)

- Now you may have the following file structure:



Run the Software

Make sure you have Thymio II turned on and the wireless dongle is plugged in your computer. If you run Thymio Suite software, you may see something like:



In your command shell, type:

```
(thymio) c:\work\thymio\src\control-thymio-via-voice>python
control_thymio.py
```

The system will initialize the connection to the microphone, then initialize the connection to the Thymio II. VOSK library will transcript your voice. The python script will convert the text command to thymio's command and activate it.

Here is sample output for one of the demo's run:

```
(thymio) c:\work\thymio\src\control-thymio-via-voice>python control_thymio.py
To create new Thymio interface
To start Thymio interface
Thymio interface has been started successfully
==> Initial Default Device Number:1 Description: {'name': 'Microphone Array (Realtek(R) Au', 'hostapi': 0,
'max_input_channels': 2, 'max_output_channels': 0, 'default_low_input_latency': 0.09,
'default_low_output_latency': 0.09, 'default_high_input_latency': 0.18, 'default_high_output_latency': 0.18,
'default_samplerate': 44100.0}
==> Build the model and recognizer objects. This will take a few minutes.
LOG (VoskAPI:ReadDataFiles():model.cc:213) Decoding params beam=13 max-active=7000 lattice-beam=6
LOG (VoskAPI:ReadDataFiles():model.cc:216) Silence phones 1:2:3:4:5:11:12:13:14:15
LOG (VoskAPI:RemoveOrphanNodes():nnet-nnet.cc:948) Removed 0 orphan nodes.
LOG (VoskAPI:RemoveOrphanComponents():nnet-nnet.cc:847) Removing 0 orphan components.
LOG (VoskAPI:ReadDataFiles():model.cc:248) Loading i-vector extractor from
D:\workdir\units\weizmann\thymio\vosk\models\vosk-model-en-us-0.22/ivector/final.ie
LOG (VoskAPI:ComputeDerivedVars():ivector-extractor.cc:183) Computing derived variables for iVector extractor
LOG (VoskAPI:ComputeDerivedVars():ivector-extractor.cc:204) Done.
LOG (VoskAPI:ReadDataFiles():model.cc:279) Loading HCLG from C:\work\thymio\src\control-thymio-via-
voice\vosk\models\vosk-model-en-us-0.22/graph/HCLG.fst
LOG (VoskAPI:ReadDataFiles():model.cc:294) Loading words from C:\work\thymio\src\control-thymio-via-
voice\vosk\models\vosk-model-en-us-0.22/graph/words.txt
LOG (VoskAPI:ReadDataFiles():model.cc:303) Loading winfo C:\work\thymio\src\control-thymio-via-voice\vosk\models\vosk-
model-en-us-0.22/graph/phones/word_boundary.int
LOG (VoskAPI:ReadDataFiles():model.cc:310) Loading subtract G.fst model from C:\work\thymio\src\control-thymio-via-
voice\vosk\models\vosk-model-en-us-0.22/rescore/G.fst
LOG (VoskAPI:ReadDataFiles():model.cc:312) Loading CARPA model from C:\work\thymio\src\control-thymio-via-
voice\vosk\models\vosk-model-en-us-0.22/rescore/G.carpaLOG (VoskAPI:ReadDataFiles():model.cc:318) Loading RNNLM
model from C:\work\thymio\src\control-thymio-via-voice\vosk\models\vosk-model-en-us-0.22/rnnlm/final.raw==> Begin
recording. Press Ctrl+C to stop the recording
{
  "text" : "move"
}
{
  "text" : "speed up"
}
{
  "text" : "speed"
}
{
  "text" : "left"
}
no input sound
{
```

```

    "text" : "stop"
  }
  {
    "text" : "start"
  }
  no input sound
  {
    "text" : "left"
  }
  {
    "text" : "stop"
  }
  {
    "text" : "start"
  }
  {
    "text" : "right"
  }
  no input sound
  {
    "text" : "stop"
  }
  {
    "text" : "back"
  }
  {
    "text" : "speed"
  }
  ==> Finished Recording

```

Demo

You may find demo video in the github page <https://github.com/ahmad081177/control-thymio-via-voice> or in the following YouTube link: <https://youtu.be/3wKyVvp2RC0>

Conclusion

It is not so hard to build the proof of concept in a school environment. You may try different VOSK models and different languages to control the Thymio robot (note, you need to modify the thymio.py file to react to the new commands). It is also recommended to give it a try to the pupils and let them build their own voice control.

The demo can be enhanced easily with more commands, such as, control the colors and play sounds on the robot.

Summary

We have seen how easy it is to setup the environment. It is highly recommended to let the students try it. students may extend the commands they control the Thymio robot.

We recommend integrating such capability to the Thymio framework such as, Scratch.