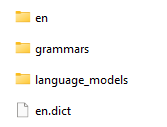
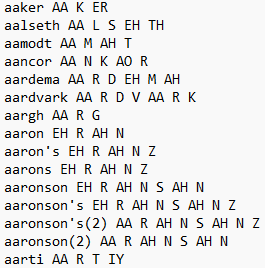
1. **Sphinx Based Plugin**

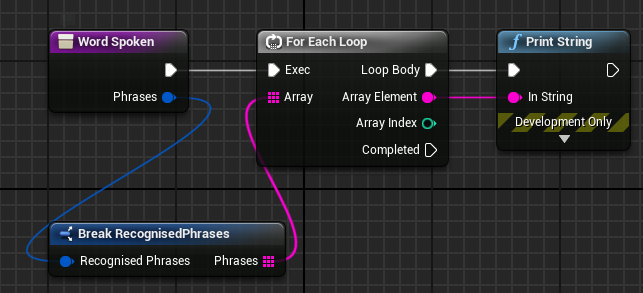
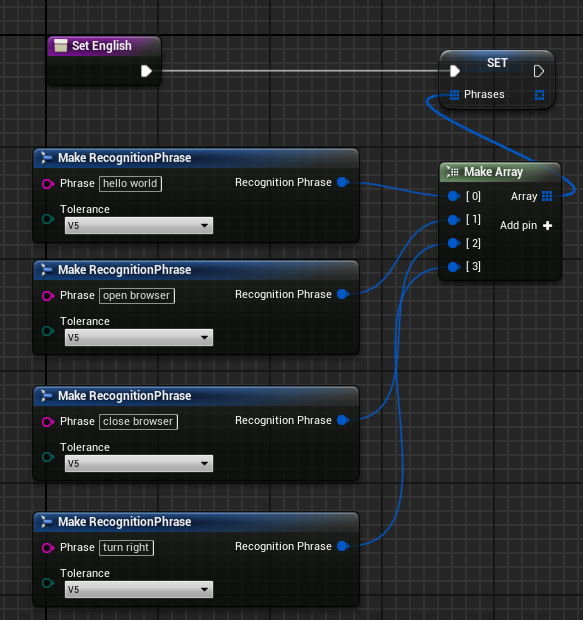
GIT : <https://github.com/shanecolb/sphinx-ue4>

Acoustic Model : Contains a statistical representation of the distinct sounds for every word in vocab and each sound corresponds to a phoneme

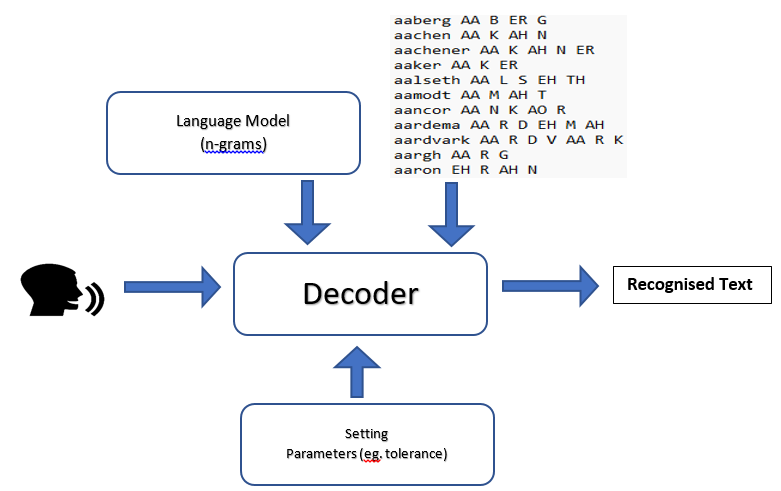
Language Model : Contains list of words and their probability of occurrence in sequence

**Fig. (a) Folder structure inside content/model directory (b) Phonemes inside the vocab**



**Fig. (a) Setting Probability tolerance for Recognised phrases (b) Reading and Displaying Recognised text**



**Fig. Overview of Sphinx plugin Speech to Text operation**

**Drawbacks**:

1. Need to add phonemes (vocabulary) for the words to get recognised

2. Performance is not good for text with 2 or more words

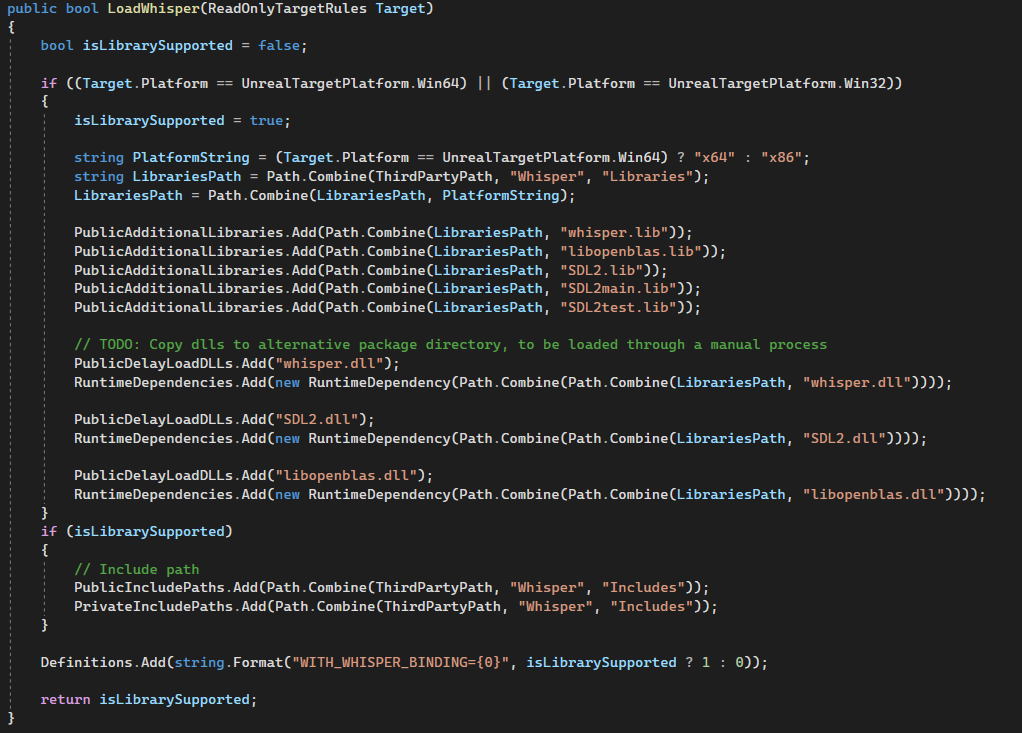
1. **Whisper Speech-to-Text Unreal Engine Plugin**

Reference : <https://github.com/ggerganov/whisper.cpp>

GIT : [.../blob/main/SpeechRecognition.zip](https://github.com/Srikanth635/IAI/blob/main/SpeechRecognition.zip)

**Libraries Used:**

1. SDL2
2. Whisper (C++)
3. Standard Library C++ 17
4. Containers : Array, Vector, Map, Set
5. Streams : fstream, iostream, sstream
6. Concurrency : thread, mutex, atomic

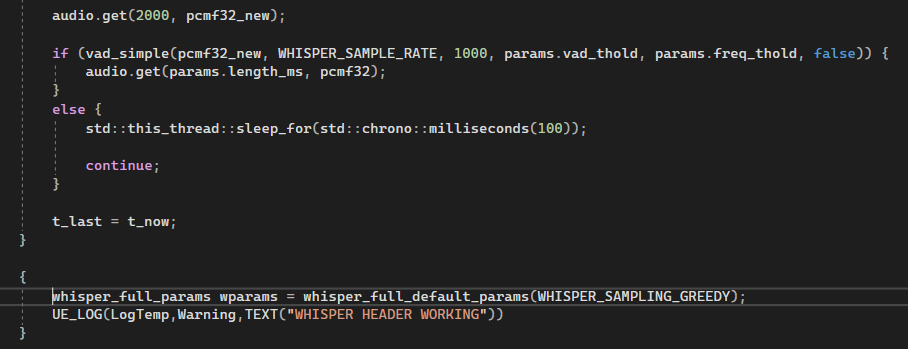


**Fig. Code snippet inside Build.cs of speech-to-text unreal engine plugin**

* Inside 'SpeechRecognition\Source\SpeechRecognition\Private\ MySpeechWorker ',

functions to record audio, scaling, filtering are found.

* Processed audio is passed on to whisper network to get transcripted text as output.



**Fig. Code snippet to retrieve audio buffer and invoke whisper for transcriptions**

**Drawbacks** :

1. Speed : The speed of transciption is greater than 8 secs which is not reliable
2. Accuracy : Obtained transcriptions doesnt match the speaker utterances always
3. **REST API (Unreal Engine) – Flask**

GIT (USemLog) : [.../USemLog/tree/SpeechRecord](https://github.com/AbhijitVyas/USemLog/tree/SpeechRecord)

GIT (Flask Python file) : [.../blob/main/SpeechREST.PY](https://github.com/Srikanth635/IAI/blob/main/SpeechREST.PY)

**FLASK** –

* Flask, a lightweight framework for building web applications in Python
* Used Python’s PyAudio to read in audio data with required format, rate etc.,
* Used routes to map URLs to functions that handle the requests
* Listened for incoming HTTP requests and responds with appropriate HTTP

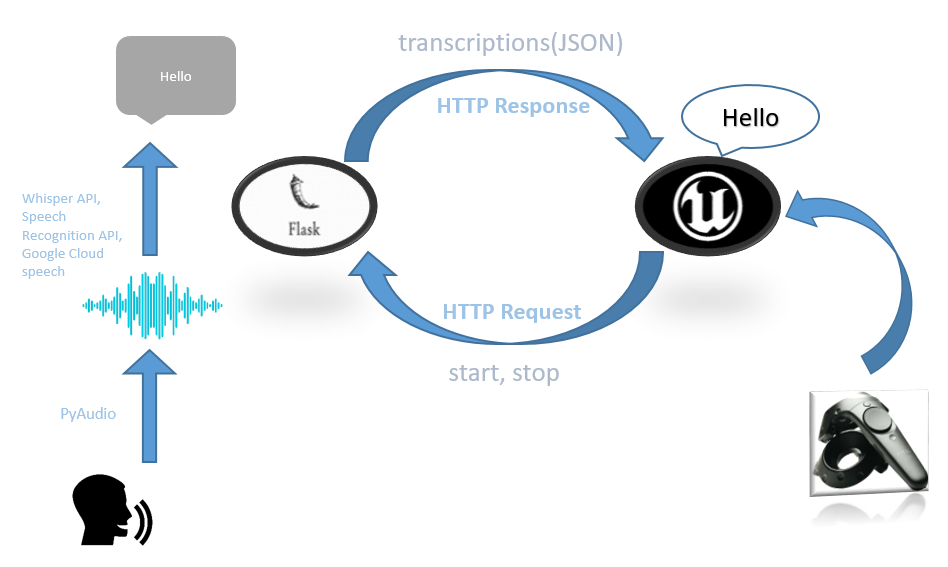
responses (transcriptions)

**Unreal REST API** –

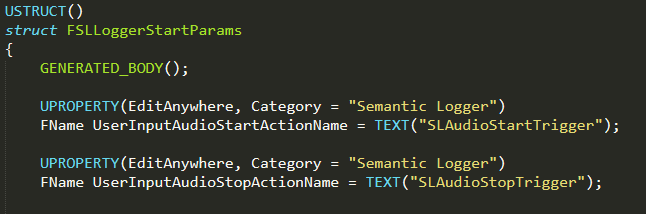
* Used Unreal engine’s C++ HTTP modules to raise API requests (start and stop recordings)
* Used parsing libraries (JSON) to parse the received response from flask

**Libraries and Tools Used :**

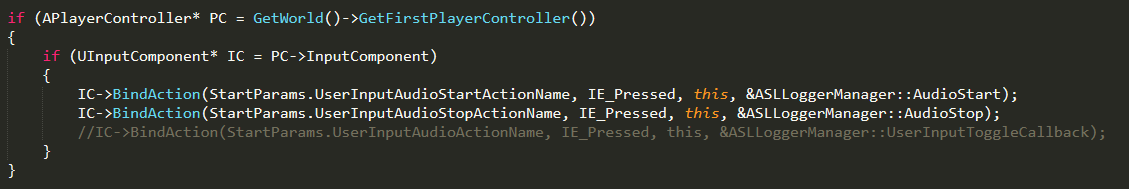
1. Unreal Engine (4.27,5.1.1)
2. USemlog Plugin
3. C++ Libraries (STL:Containers, JSON, HTTP)
4. PyCharm (Flask API)
5. Python Libraries (PyAudio, Whisper, os, wave, datetime, torch, threading)



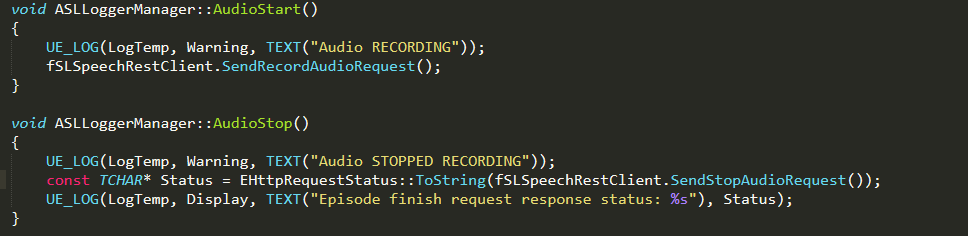
**Fig. Overview of Speech to Text Operation in ‘RobCoG’ using FLASK API**



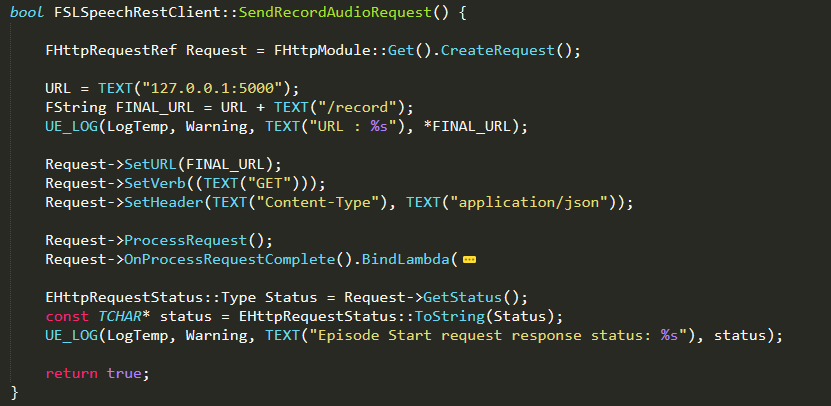
**Fig. Code snippet of variables to facilitate controller mappings**



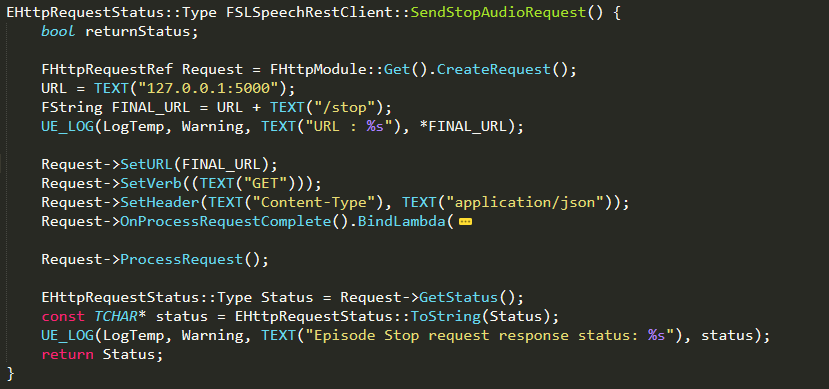
**Fig. Code snippet of controller mappings to functions**



**Fig. Code snippet of mapped functions calls definition**



**Fig. Code snippet of function to send start audio signal API request**



**Fig. Code snippet of function to send stop audio signal API request**