Overview of VCIVING

VCIVING is a project that intends to provide voice control for IVI. User speech is converted into internal commands that run on the system.

Operation of VCIVING is divided into 4 subsequent steps.

1. Reading user's speech through microphone.
2. Using a Speech-to-Text (STT) model / Speech Recognition Engine to obtain the text version of the speech.
3. Interpretation of the recognized text using a trained model.
4. Executing the task/process requested by the user/meant by the recognized text.

1. Reading user's speech through microphone
   - This step involves grabbing all inputs through the microphone.
   - Some technologies will be used later on to determine whether user's speech is really meant for VCIVING to process or not.

2. Using a STT model/Speech Recognizer to obtain the text version of the speech
   - Speech from the microphone is obtained as an audio stream.
   - This needs to be converted to text to be processed further.
   - After eliminating noise, pre-trained models are used to convert the audio to text.
   - If there is any valid speech in the audio received, its text version is returned through the model/recognizer.

3. Interpretation of the recognized text of speech using a trained model
   - After text version of the speech is obtained, our next attempt is to interpret it.
   - What we expect from the interpretation is the grabbing the underlying meaning of the text obtained.
   - Text is passed through another pre-trained model which previously trained to classify phrases into different tasks we expect the system to perform for us.
   - The pre-trained model would have a set of tasks that were defined before the training process, so the model would find most suitable task for the phrase given as the input.
   - Then the retrieved task is executed.

4. Executing the task/process requested by the user/meant by the recognized text
   - When the task is retrieved and executed, VCIVING interacts with the core IVI systems (and core systems of the smart vehicle) to run the desired task.
   - The commands meant by the recognized text is passed to these tasks and processed inside them to grab necessary data to be passed into the core IVI system.
   - This merely consists of interfaces which connects VCIVING to the core of IVI.