

Undertone - Offline Whisper AI Voice Recognition

Getting Started

Welcome to the Undertone documentation! In this section, we'll walk you through the initial steps to start using the tools. We will explain the various features of Undertone, how to set it up, and provide guidance on using the different models for voice recognition.

Models

Undertone offers both English-only and multilingual models. The plugin comes with a default English-only model, `tiny.en`. Available model types include `tiny`, `base`, `small`, `medium`, and `large`. Smaller models are more suitable for devices with limited resources, like phones, while larger models can be used on computers with more processing power.

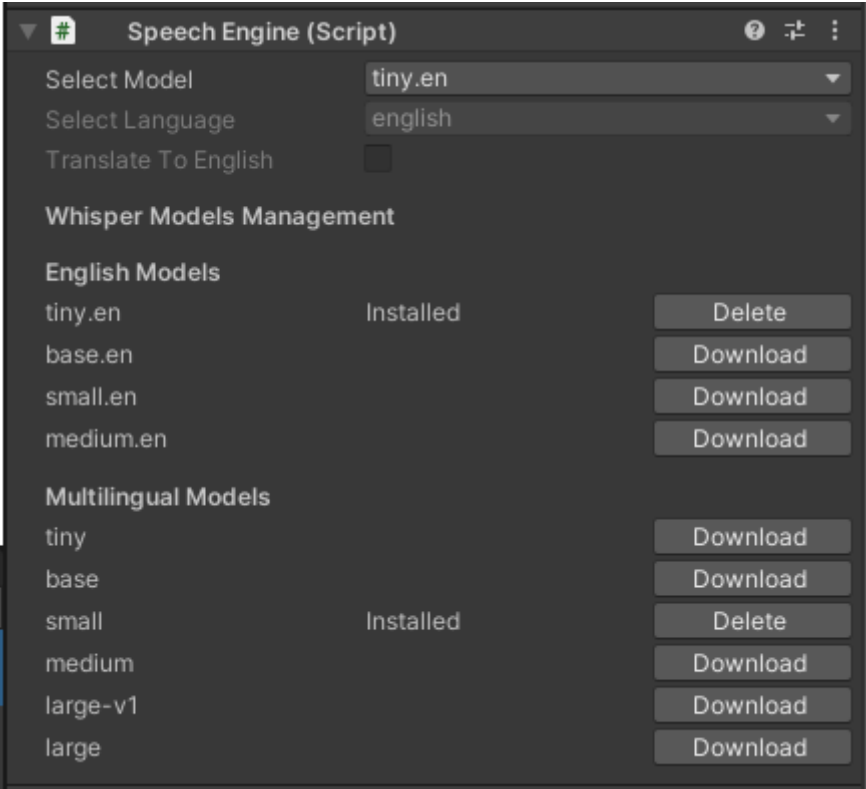
Model comparison

The following table provides a comparison of the various models in terms of disk space and memory usage:

Model	Disk	Mem
tiny	75 MB	~125 MB
base	142 MB	~210 MB
small	466 MB	~600 MB
medium	1.5 GB	~1.7 GB
large	2.9 GB	~3.3 GB

How to download models

The SpeechEngine component provides a convenient interface for downloading the models



with just a click.

Demos

The plugin contains two demos to demonstrate transcription functionality: realtime transcription and push-to-record transcription.

Realtime transcriber

This demo captures a window of data every 1-2 seconds and transcribes it in the background. It uses the RealtimeTranscriber.cs script.

Push to talk

This demo captures up to 100 seconds of audio

Scripts

SpeechEngine

This class loads the model into memory. Since the model is not multi-threaded, one SpeechEngine should be loaded per concurrent thread that will be transcribing. The SpeechEngine allows you to select the language and translation.

Script Reference for SpeechEngine.cs

Property	Type	Description	Default Value
Selected Model	string	A string representing one of the downloaded models. The tiny.en model is included in the asset package.	tiny.en
Selected Language	string	A string representing the language the audio will be in. Use "auto" for auto-detecting the lang.	en
Translate to english	bool	Translate the result into english. For example if the audio says "Hola" in spanish the resulting text will be "Hello"	false
Suppress Blank	bool	Suppress blanks in transcription	true
Speed Up	bool	Experimental speed up, at the loss of quality. Useful on long amounts of audio	false
Verbose	bool	Extended logging	false

RealtimeTranscriber

This script captures audio from the microphone continuously and transcribes it in windows, displaying the transcription on the screen. Smaller windows result in faster transcribing at the cost of overloading the system. The transcription window adapts depending on how the load is being handled.

Script Reference for RealtimeTranscriber.cs

Property	Type	Description	Default Value
Engine	SpeechEngine	The speech engine used for processing	null
OnTextTranscribed	event	Event triggered when text is transcribed	Empty
InitialStepSizeInSeconds	float	Initial step size for processing audio in seconds	1.5

Property	Type	Description	Default Value
AutoAdjustStep	bool	Automatically adjust the step size based on the input	true
MaxWindowLengthInSecs	float	Maximum window length for processing audio in seconds	12
WriteTimestamps	bool	Include timestamps in the transcription output	true

PushToRecord

`PushToRecord.cs` is a script that captures audio input and processes it using a specified speech engine. The script is designed to transcribe the audio, providing an output with optional timestamps. Users can set a maximum recording time to limit the duration of the audio capture. The script requires an instance of a `SpeechEngine` to perform the transcription and can be customized based on the user's requirements.

Script Reference for `PushToRecord.cs`

Property	Type	Description	Default Value
Engine	SpeechEngine	The speech engine used for processing	null
WriteTimestamps	bool	Include timestamps in the transcription output	true
MaxRecordingTime	int	Maximum recording time in seconds	100

Troubleshooting

Common issues

Transcription quality is poor

There could be several factors contributing to this issue:

- **Background noise:** The model might struggle with accurate transcription when there is substantial background noise or music. Try reducing the noise for better results.
- **Small model:** While small models offer portability and speed, their transcription quality may not be as high. Consider using base or larger models for improved accuracy.
- **Multilingual for English:** If your application is primarily focused on supporting English, it is advisable to use an English-specific model. These models typically perform better on English tasks compared to their multilingual counterparts.

Other

For any questions, issues or feature requests don't hesitate to email us at help@leastquares.io or join the [discord](#). Very are happy to help and have very fast response times :)

About us

[LeastSquares](#) is an independent game studio focused on building tools for game developers.

For any inquiries, feel free to contact us at hello@leastquares.io or contact us on the [discord](#)

Check out our other assets [LeastSquares Unity publisher store](#)

Appendix

Supported languages

Undertone multilingual models support the following languages:

- auto
- english
- chinese
- german
- spanish
- russian
- korean
- french

- japanese
- portuguese
- turkish
- polish
- catalan
- dutch
- arabic
- swedish
- italian
- indonesian
- hindi
- finnish
- vietnamese
- hebrew
- ukrainian
- greek
- malay
- czech
- romanian
- danish
- hungarian
- tamil
- norwegian
- thai
- urdu
- croatian
- bulgarian
- lithuanian
- latin
- maori
- malayalam
- welsh
- slovak
- telugu
- persian
- latvian
- bengali

- serbian
- azerbaijani
- slovenian
- kannada
- estonian
- macedonian
- breton
- basque
- icelandic
- armenian
- nepali
- mongolian
- bosnian
- kazakh
- albanian
- swahili
- galician
- marathi
- punjabi
- sinhala
- khmer
- shona
- yoruba
- somali
- afrikaans
- occitan
- georgian
- belarusian
- tajik
- sindhi
- gujarati
- amharic
- yiddish
- lao
- uzbek
- faroese
- haitian creole

- pashto
- turkmen
- nynorsk
- maltese
- sanskrit
- luxembourgish
- myanmar
- tibetan
- tagalog
- malagasy
- assamese
- tatar
- hawaiian
- lingala
- hausa
- bashkir
- javanese
- sundanese