

Speech to Text User Guide

Table of Contents

User Guide	1
Evaluation Licences	1
Speech to Text Models	2
Dictation App	3
Game Chat App	3
Transcribe File App	4
Technical Information	4
GPU Preferences	4
Hotkey Setup	5
Known Issues	5
Uninstalling	6
Disclaimers	6

User Guide

Since you are reading this document, you have probably already installed the software.

Launch one of the Start menu shortcuts: “Live Dictation”, “Game Chat”, or “Transcribe File”.

Evaluation Licences

If your computer has an internet connection, click **Yes** when prompted with “Allow online activation?”. On the next screen, click “Evaluate”.

If the computer is offline, click **No** when asked “Allow online activation?”. Copy the computer ID by clicking the Copy button and paste it into a text file. Transfer this file to another computer with internet access using a USB flash drive or another suitable method. On the internet-connected computer, paste the Computer ID into the form on the following web page <http://vrmac-software.com/speech2text/evaluation> and click “Submit”. The server will generate a ZIP file containing the licence for your computer. Save this file to the USB flash drive and transfer it back to the offline computer. On the offline software activation screen, click “Import...” and open the ZIP file from the flash drive.

Speech to Text Models

The next setup step is to download a speech to text model. I distribute 12 of them for completeness, but I only recommend two.

- Whisper Turbo BCML — use this when you do not need translation.
- Whisper Large BCML — use this when you want to translate non-English speech into English text.

On all the computers in my household, the medium models perform worse than the Turbo model. Although BCML is a lossy codec, I have not observed any loss of transcription accuracy. By contrast, uncompressed models have a substantial memory overhead, using approximately three times more memory. However, if you have a discrete GPU with sufficient VRAM, the uncompressed models contain weights that are bit-for-bit identical to the upstream Whisper models from OpenAI, with no loss introduced during packaging.

The model downloader takes some time to start downloading, please give it a couple of minutes after clicking the download button.

If you have purchased the software but unable to download models due to some networking shenanigans¹, please contact support. I have alternative methods to distribute model files but these are not automated, require some manual intervention.

If you want to copy-paste model across computers to save time and network bandwidth, they are stored in `%LOCALAPPDATA%\Speech to Text\Models`, feel free to copy-paste these *.cgm1 files. Paste the model files into that location on the destination computer, then click “Download models...”. The downloader will verify the file hashes to ensure integrity, and the newly added models will become available to all speech to text apps in this software.

The downloader tool does not verify software licences. If you have purchased a licence and wish to download models for use on another computer, install the software and run ModelDownloader.exe from the installation folder. Download the required models, then copy them to the computer where you want to use speech-to-text.

When testing the downloader tool over Wi-Fi, I have noticed that downloads sometimes stop partway through and the progress indicator freezes. This issue will be debugged and fixed in due course. In the meantime, close the downloader tool and relaunch it. The tool stores partially downloaded content on disk and automatically resumes downloads on launch.

¹ Model downloader utility is based on [BitTorrent](#) protocol.

Dictation App

The application continuously transcribes live speech captured from the microphone and sends the final output to third-party text editors.

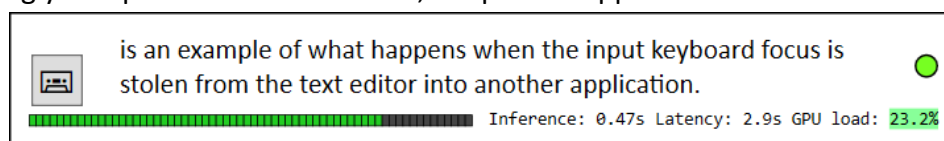
After loading a model, use the next screen to select the audio input device, specify the language, and test speech to text by clicking “Capture”.

On the following screen, assign a global hot key, optionally restrict the allowed text destinations, and click Next. The application will then switch to a small, always-on-top status window. When idle, this window can be moved by clicking and dragging it with the mouse.

Alt+Tab to the text editor where you want the text, press the configured hotkey to start the dictation session, and begin speaking. Press the same hotkey again to stop the session.

That status window shows pending text. That text can and does change as you continue speaking. The final text is sent to the text editor that had keyboard focus when you pressed the hotkey.

If you alt+tab away from the text editor to another window, or if the input focus is stolen by a rogue pop-up, the application will stop sending text to the editor but will continue transcribing your speech. In this situation, a tape icon appears on the left of the window:



When this happens, stop the dictation session using the hotkey, then click the tape icon. A panel will open showing the text that could not be delivered to your text editor. Buttons on the right allow you to copy the text to the clipboard or clear the buffer.

As with keyboard input, dictated text replaces the current selection. If you select all content in the editor with Ctrl+A, press the dictation hotkey, and begin speaking, the existing content will be replaced.

While a dictation session is in progress, the status window does not respond to mouse input. To move the window or exit the application, stop the session first.

Game Chat App

The application transcribes individual spoken sentences and places the resulting text on the Windows clipboard.

The configuration steps are similar to those for dictation: select and load a model, test speech to text, and set up a global hotkey. Clicking Next transforms the tool into a small circular window with a microphone icon, as shown in the screenshot to the right.



By default, you can move that window around the screen using the mouse. If you right click the circle and select “Disable mouse events” from the context menu, the window will no

longer handle mouse input and clicks will pass through to whatever is behind it. To re-enable mouse input or exit the application, right click the system tray icon that appears after selecting “Disable mouse events”.

Once you are satisfied with the position of the status indicator window, press the global hotkey and start speaking. Press the same hotkey again to stop recording and begin transcription. Alternatively, you can simply stop speaking. At the time of writing, the default setting for maximum silence duration is 1.5 seconds, this means that as soon as the app detects a continuous 1.5 seconds of audio without speech, transcription starts automatically.

After a short delay, a pop-up displaying the transcribed text will appear. By the time this pop-up is shown, the text has already been copied to the Windows clipboard and can be pasted into other applications.

Unlike dictation, the game chat app is designed to run alongside other GPU demanding software. Once setup is complete, the user interface is minimal, and the speech-to-text engine uses significantly fewer GPU resources than dictation.

Transcribe File App

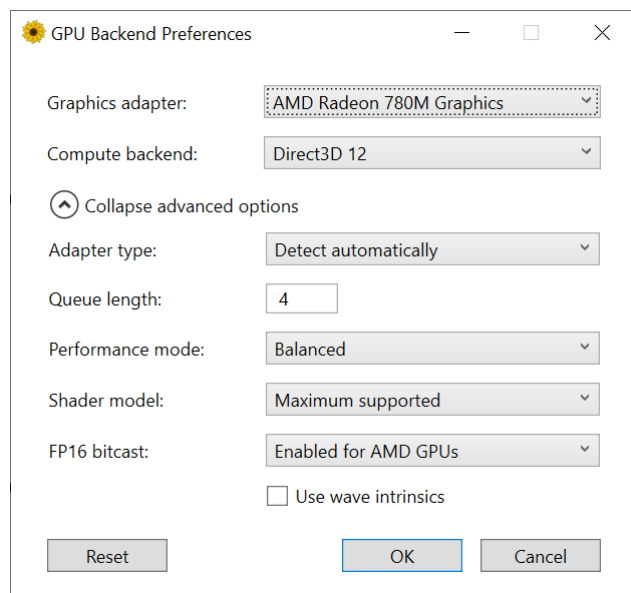
The application transcribes audio files from disk with the same speech to text engine.

The workflow is quite similar: load a model on the initial screen, test on the next screen, transcribe files on the final one.

Technical Information

GPU Preferences

All 3 applications have “GPU Settings...” button on the initial screen where you select a model to use. Note the advanced options are collapsed by default, click on the “Expand advanced options” to access these parameters. I hope the default values should be good enough for most systems, but I recommend testing custom parameters in the following cases.



- If you have a computer with Intel GPU, integrated or discrete, try switching the FP16 bitcast option to “Enabled for all GPUs”. On nVidia and AMD GPUs the performance consequence of that setting is huge, disabling for AMD or enabling for nVidia decreases performance approximately by half.

- If you have a computer with two GPUs integrated and discrete, the integrated one is performant², and you want discrete GPU for some other demanding tasks running in parallel with speech to text.
- If you have a computer with AMD Strix Halo integrated graphics, try to switch the adapter type from the default “Detect automatically” to “Discrete graphics” and test the performance.

Queue length and performance mode preferences have non-trivial performance impact with the legacy Direct3D 11 backend, minimal impact with the Direct3D 12 backend.

Hotkey Setup

Both Dictation and Game chat applications trigger speech to test with a hotkey. And there are two mutually exclusive ways to specify that hotkey: keyboard hotkey, and raw input.

- Keyboard hotkeys rely on the corresponding [OS API function](#). The registered hotkeys are not specific to a particular keyboard device, and the hotkey device is never lost even if you disconnect all keyboards from the computer. If the hotkey you want is already registered by some other application running on the same PC, you will see an error message and the application will require you to specify another hotkey.
- Raw input supports mouse in addition to keyboard, and commands from other [human interface devices](#) like game controllers. The registered command is specific to the input device. Reconnecting the input device into another USB port changes device ID, you will have to set up the hotkey again. If the hotkey is already handled by some other running program, you will see no warning messages, the OS will deliver input events to both programs. And if you will use a keyboard-based raw hotkey, some security software might falsely flag dictation and game chat programs as keyloggers³.

Known Issues

If you are trying to transcribe OGG Vorbis audio file but it fails complaining about missing codec, please install [Web Media Extensions](#) from windows store.

If you are monitoring the GPU tab in Windows Task Manager while transcribing audio, you may notice that the GPU appears idle, even though the computer’s fans are spinning and the system is clearly under load. This is a display issue in Task Manager. To correct it, select a different graph from the drop-down list. The appropriate graph depends on the GPU model and driver version. When using the default Direct3D 12 backend, GPU activity is shown on one of the graphs labelled “Compute <something>”, such as “Compute 0” or “Compute_1”. If

² On my desktop PC, the integrated Radeon 780M GPU delivers 8x realtime speed with Turbo model.

³ If you are wondering why I support raw keyboard hotkeys despite the drawbacks, it is because some high-end mice present themselves to the operating system as composite USB devices, with one endpoint being a fake keyboard. When pressed, certain additional buttons beyond the five standard ones (left, right, middle, X1, X2) generate keyboard events.

you switch the backend to the legacy Direct3D 11, the transcription workload will instead appear on the default “3D” graph.

When using the dictation app to input text into Apache OpenOffice 4.1.1, it may take two or three attempts to start a dictation session. I think this is due to an accessibility-related bug in that version of OpenOffice. At least on my computer, OpenOffice 4.1.16 works correctly.

Uninstalling

The installer supports two distinct installation types: system wide, and current user. You can even install both ways if you want, the setup program does not stop you.

To uninstall the system wide installation, open either control panel “Uninstall a program”, or settings “Apps & features”. Type or copy-paste “Speech to text” into the search box, click the uninstall button.

To remove the program installed for current user only, copy-paste the following into the address bar of Windows explorer: `%LOCALAPPDATA%/Speech to Text` Press Enter, and delete `Binaries` folder there. Then copy-paste the following into the address bar: `%APPDATA%\Microsoft\Windows\Start Menu\Programs` Press Enter, and delete `Speech to Text` folder.

Disclaimers

The quality of the speech-to-text engine is generally good but not perfect. It can, and does, make mistakes, especially when the input speech is not in English. If you intend to use the software to dictate important e-mails, please ensure you carefully review the output text before clicking the “Send” button.

See the "Pre-existing IP.txt" text file in the installation folder for copyright notices.