

Speech to Text User Guide

Table of Contents

User Guide	1
Evaluation Licences	1
Speech to Text Models	2
Dictation App	2
Game Chat App	3
Transcribe File App	4
Technical Information	4
GPU Preferences	4
Hotkey Setup	5
Transcribe File GUI	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-softver.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 15 of them for completeness, but I only recommend two.

- Voxtral Mini BCML — for graphics cards with 4 GB VRAM.
- Voxtral Mini BCML text — for graphics cards with at least 6 GB VRAM.

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 models 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 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.

The downloader tool stores partially downloaded content on disk and automatically resumes downloads on launch.

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 allowed text destinations with checkboxes, 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 newly transcribed text as it becomes available.

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

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, the newly transcribed text is displayed in green colour, and the transcribed text is held in a buffer, as shown in the screenshot to the right. When this occurs, stop the dictation session using the same hotkey — the status window will become interactive again, allowing you to copy the buffered text directly from it. Alternatively, pressing the transcription hotkey once more will flush the entire buffered text into whichever text editor currently holds the keyboard focus.



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.

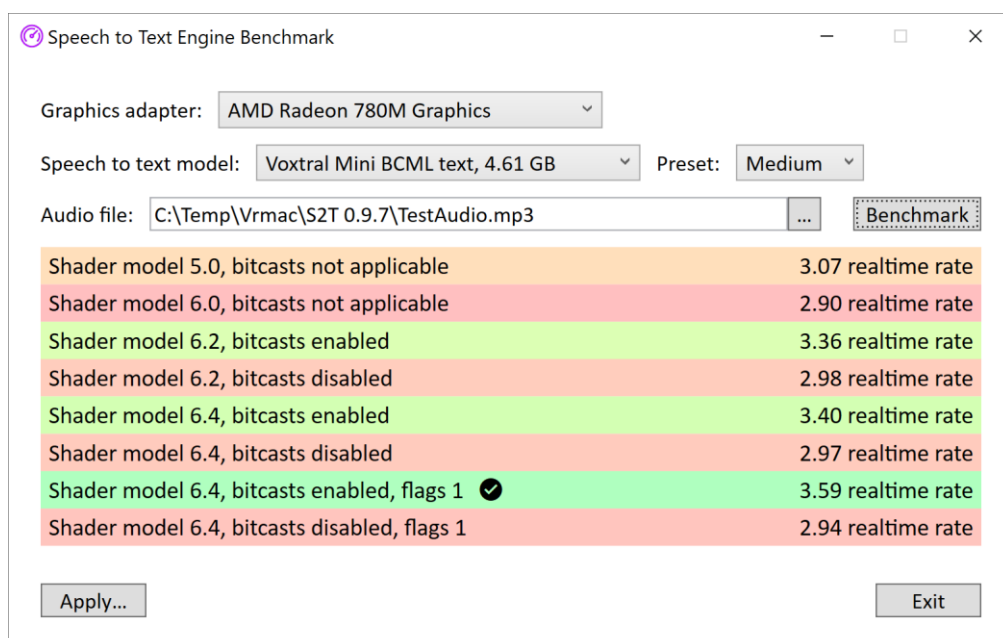
The workflow is quite similar: load a model on the initial screen, test on the next screen, transcribe files on the final one.

Technical Information

GPU Preferences

All 3 applications have “GPU Settings...” button on the initial screen where you select a model to use. 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, switch the “Graphics adapter” preference.

I do not recommend using window for advanced preferences, though. Instead, run Benchmark.exe located in the installation folder. That tool transcribes a test audio file with the supported combinations of advanced parameters, and has an “Apply...” button to save the found optimal settings for use by speech-to-text applications.



² On my desktop PC, the integrated Radeon 780M GPU delivers 3.6 realtime rate with Voxtral BCML model.

Hotkey Setup

Both Dictation and Game chat applications trigger speech to text 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.

Transcribe File GUI

When using Whisper models, the “Advanced options...” button on the test screen opens a window where you can adjust the speech-to-text engine parameters. The default settings should be suitable for most use cases. At the time of writing, the software remains in pre-release, with ongoing R&D aimed at improving output quality and performance. I expect significant changes to that window; therefore, the parameters are not yet documented in this user guide.

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

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.

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 speech-to-text engine can make mistakes, including substituting words that are phonetically similar but have entirely different meanings. Always review transcribed text carefully before sending or publishing, especially for important communications.

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