

Whisper Transcriber – Super Simple Guide

Hi! This little book shows you how to use Whisper Transcriber step by step. No big words. Just click, wait, and read.

What this app does

- It listens to your audio files (like `.mp3`, `.wav`, `.m4a`).
- It writes down what it hears as text (a transcript).
- It also saves “segments” — small pieces with time stamps, like `[0.00 - 3.12] Hello!`
- Optional: It can glue nearby segments together to make bigger chunks that are easier to read.

1) Pick the model

- Find the box that says “Whisper model”.
- Choices: tiny, base, small, medium, large.
- Big model = smarter but slower.
- Tip: “large” is most accurate; if your computer is slow, try “small”.

2) Add your audio files

- Drag & drop files into the big box, or click “Browse...” to pick files.
- You can select more than one file.
- You’ll see the file names listed.

3) Transcribe

- Click “Transcribe Selected Files”.
- You’ll see a progress bar and an ETA.
- When done:
- “Full transcript” is the full text for each file.
- “Segments” is the time-stamped version.
- If you checked “Combine full transcripts into one file”, it will ask where to save a single big `.txt` file.

4) Open the folder

- Click “Open Output/Audio Directory” to open the place where your results are saved.

Optional: Segment Parser (the glue tool)

Think of segments like little Lego blocks of text. Sometimes there's a tiny gap between two blocks that really belong together (like a comma). The Segment Parser can glue them.

The “threshold” number

- You'll see a box that says “Merge if gap < (sec)”. This is the threshold.
- Example: If threshold is 1.0 seconds, and two segments have a gap of 0.5 seconds between them, they get glued into one longer segment.
- If the gap is bigger than the threshold, they stay separate.

Simple examples

- Imagine these two segments:
 - [0.00 - 2.00] I like apples
 - [2.30 - 4.00] and bananas
- The gap is $2.30 - 2.00 = 0.30$ seconds.
- If your threshold is 1.0 seconds, $0.30 < 1.0$ so they glue together:
 - [0.00 - 4.00] I like apples and bananas
- Another pair:
 - [5.00 - 6.00] This is sentence one.
 - [8.50 - 10.00] This is sentence two.
- The gap is 2.50 seconds.
- If your threshold is 1.0, then $2.50 > 1.0$, so they do not glue. They stay as two segments.

How to use it

- 1) Click “Browse Segments File” and select a file that looks like `my_audio_segments.txt`.
- 2) Set the threshold (like 0.8 or 1.5). Start with 1.0 if you're not sure.
- 3) Click “Parse Selected File”.
- 4) You'll get a new file with a name like `my_audio_parsed_t1.0.txt`.

When to change the threshold

- If the text looks too choppy (too many tiny lines), raise the number (e.g., 1.5 or 2.0) to glue more.

- If the text looks too smushed (too long lines that shouldn't be together), lower the number (e.g., 0.5) to glue less.

Tips & FAQ

- My transcript is wrong in places! Try a bigger model (like “large”).
- It's slow. Try smaller model sizes, or fewer files at once.
- Where are files saved? Next to your audio, unless you pick a different place for the combined file.
- Can I use dark mode? Use the Theme dropdown (if available) to pick a different theme.

Quick Glossary

- Transcript: All the words in your audio, as text.
- Segment: A small chunk of the transcript with start and end times.
- Threshold: The “glue power”. If the gap between two segments is smaller than this number (in seconds), they will be merged.