

Audion Voice Tools — User Guide

[Русский](#) · [About](#)

Contents

- [First Start](#)
- [Dictation](#)
- [Tray](#)
- [The Dictations Section](#)
- [The Recordings Section and the Recorder](#)
- [The Files Section](#)
- [The Journal Section](#)
- [SETUP](#)
- [SETTINGS](#)
- [What Goes to the Network](#)

First Start

1. Start **Start.exe**. The window opens, a capsule — a narrow strip — appears at the top of the screen, and the program's icon is in the tray. Windows 11 hides new tray icons behind the «^» arrow: drag the icon onto the taskbar to keep it in sight.
2. The window asks about installing the components recommended for this machine: the dictation engine, FFmpeg, GigaAM and Whisper for files; with an NVIDIA card Whisper on CUDA and Whisper large-v2 come in place of Whisper for files (why — in the Files section). Every component has a tick, the dictation engine a choice of the model: GigaAM v3 first, the best model for Russian, then the models for other languages. INSTALL puts in place what is ticked, one by one; the progress is in SETUP — the step, the percentage, the speed, the time left. "Install later" postpones the question until the next start; OPEN SETUP closes the question and opens the sheet SETUP at once.

3. While the dictation engine is being installed, the DICTATION button of the capsule is dimmed. Once the Journal section shows the line `[OK] gigaam-...: load ..., warm-up ...`, dictation is ready.

Dictation

- **Right Alt + F12** — start; again — stop and paste. **Esc** — cancel.
- Point at the capsule or click it — it opens: DICTATION and RECORDING, each with its hotkey. The right button on the capsule opens the same menu as `v`: the text modes, the engine of dictation, the place of the panel.
- The text is pasted where the cursor was, through the clipboard; the former content of the clipboard comes back.

While you speak, the capsule shows:

Left to right	What it is
×	cancel: nothing is pasted, the sound stays in the history
dot, bars, time	recording, the microphone level, the time spent
↶	start again: this take is cancelled, a new one starts at once
the last words	what was heard (with cloud streams — as you speak)
the engine's name	GigaAM, Whisper, or a cloud with the ☁ sign
red square	stop and paste
v	text modes and the dictation engine

After the stop the capsule says "Recognizing...", with a text mode — "Working over: ...", then shows the pasted text for a moment.

Dictation Engine

The Dictations section → ADVANCED, the blocks LOCAL MODELS and API: SERVICES AND MODELS; the same choice is `v` of the capsule → "Engine of dictation >" and the tray menu:

- **Local models** on transcribe.cpp — on this machine, no network, no key; a list across the window, a row a model: its name, what it is for and its languages, its size, "installed" or INSTALL. GigaAM v3 — Russian speech, the default; Parakeet V3 — 25 European languages; Parakeet EN — English;

Whisper Small, Medium, Turbo, Large v3, Large v2 — multilingual. INSTALL downloads a model from Hugging Face, checks it by SHA-256 and chooses it; SETUP has the same models. Another model chosen is loaded at once; a dictation under way finishes first. The NOW line names what runs: `gigaam-v3-e2e-rnnt-Q8_0 · Vulkan0`.

- **xAI, OpenAI Realtime, ElevenLabs** — cloud streams: the words come into the capsule while you speak.
- **OpenAI · after the stop** — one request after the stop to an endpoint compatible with OpenAI; it writes English terms best of all.

The MODEL of a cloud comes from the list of the service itself, the button with the arrow: OpenAI Realtime — `gpt-realtime-whisper` by default, OpenAI after the stop — `gpt-transcribe`; xAI and ElevenLabs choose their model themselves. The window asks the service for its list once a day, ☺ beside it — at once: the services change their models every month. SPEECH of a cloud — AUTO, RU, EN.

A cloud needs its own key (SETUP → KEYS OF THE SERVICES). If the cloud does not answer within eight seconds, GigaAM on this machine recognizes the same take, and the history marks it on the dictation.

Text Modes

√ of the capsule — the modes, as in Letterly; the same choice is Dictations → ADVANCED → TEXT AND SUMMARY. A mode chosen during a dictation already applies to it.

Mode	What the model does
Original text	nothing: the text of the engine is pasted (the default)
Magic	chooses the form itself by the text and by the window it goes into: lightly, clean, a list, parts, minutes
Lightly	punctuation, capitals, paragraphs; fillers and accidental repeats removed, the words stay
Thoroughly	a clean written text: grammar and slips corrected, the meaning and details stay
As a list	every thought or task is an item «- »
Structured	parts by meaning with short headings
Minutes of the meeting	subject, gist, decisions, tasks with owners and dates, open questions

Text modes and summaries are on with a cloud service; with the local model — when config holds an OpenAI or xAI key. SERVICE — OPENAI, XAI or OWN SERVER, MODEL — from its list. By default — OpenAI and `gpt-6-luna`: it makes the minutes of a meeting at the highest reasoning effort (`max`), the other modes at a low one (`low`), so the paste does not wait. The key of its service is needed, and goes only to it: `api.openai.com` — the OpenAI key, `api.x.ai` — the xAI key. An own server — Ollama, LM Studio — is asked without a key: an address like `http://localhost:11434/v1` and a model from its list.

The model is waited for 15 seconds, a long text longer, the minutes of a meeting up to two minutes. No answer or a wrong one — the text of the engine is pasted; the history keeps both texts.

Dictionary

The Dictionary section, the file `config\dictionary.txt`. One entry a row of the section: AS SAID → AS WRITTEN, the left field empty for a term. ROW adds an entry, NOTE a note, × removes one; the search narrows the list; the section takes the whole height of the window. The file holds the same entries, one a line:

```
Kubernetes
PyTorch
git hub actions -> GitHub Actions
kilometres per hour -> km/h
# units
degrees celsius -> °C
```

- **A term** — a word, an abbreviation, a shortening or a unit as it is written. The clouds are told it (xAI and ElevenLabs as key terms, OpenAI after the stop in its prompt), the text modes and the minutes keep it, the transcription of files takes it.
- **A replacement** `said -> written` (or `→`, `=>`) — made in the text of every dictation by whole words, whatever the case; it works for GigaAM too, which cannot be told words.
- `#` at the start of a line — a note.

The dictionary saves itself a second after typing and applies from the next dictation.

Microphone and Panel

- The microphone — Dictations or Recordings → ADVANCED → DEVICES → MICROPHONE, tray menu → "Microphone >": DEFAULT is the Windows default microphone.
- The panel — Dictations → ADVANCED → DEVICES → PANEL: TOP or BOTTOM of the screen (the tray menu or `√` of the capsule → "Place of the panel >"); SHOWN AT REST (tray menu → "Show the

floating panel"): off — the capsule appears only for a dictation or a recording.

The capsule, the tray and the window share one choice: changed in one place, it is shown in the others at once. ADVANCED is open until the first work of its page — a dictation, a recording, a run of files — then folds into one line by itself; after that it stays as it is left: the chevron on the right or a press on its head opens and folds it.

Tray

The left button on the icon opens the window. The right one — the menu:

- **Paste the last dictation** — into the window that was in front;
- **History** — the Dictations section;
- **Start a recording (Right Alt + F11)** — with a red dot while a recording goes;
- **Engine of dictation** ›, **Microphone** ›, **Place of the panel** ›, **Show the floating panel**;
- **Open Audion Voice Tools, Settings, Quit.**

Closing the window with its cross leaves the program in the tray: dictation works. Quit is only in the tray menu.

The Dictations Section

Dictations by days, the newest on top; the search finds by the beginnings of words, whatever the case. Point at a dictation — its buttons appear: copy, recognize again (GigaAM of this machine, the text is replaced), listen, download the sound to Downloads, delete (the record from the library, the sound to the Recycle Bin). The tip of the text tells where it was pasted, the engine, the mode, the length, how long the paste was waited for.

The right button: **Paste again, Copy the text of the engine — before working over** (dictations with a mode), **Pin on top** — a pinned one keeps its sound past its term.

The sound of dictations is kept 7 days (SETTINGS → SOUND OF DICTATIONS: 7 DAYS, 30 DAYS, ALWAYS, DO NOT KEEP); the text stays always. Dictations of 2.3 come into the history by themselves, without sound.

The Recordings Section and the Recorder

- **Right Alt + F11**, the RECORD button of the section or RECORDING in the capsule — start. The capsule turns red: time, level; **||** — pause (the microphone stays open, nothing is written), the red square — save, × — cancel (the file goes to the Recycle Bin).
- A recording is 48 kHz mono in `library\recordings\YEAR\MONTH\`. After a crash of the program an unfinished recording is recovered at the next start. Dictation works during a recording.
- Once saved, a recording goes to transcription by itself with the choices of the Files section (Recordings → ADVANCED → AFTER RECORDING: TRANSCRIBE AT ONCE or BY HAND; CHOOSE opens Files).

The buttons of a recording card: listen, open the transcript, **minutes of the meeting**, show in the folder, download to Downloads, transcribe (again).

The minutes of the meeting are made by the model of Dictations → ADVANCED → TEXT AND SUMMARY from the transcript: the file `NAME.protocol.md` next to the recording, opened by itself. Made, the button opens them; the right button makes them again, the old ones go to the Recycle Bin.

The Files Section

1. ADD FILES or ADD A FOLDER — recordings and videos: m4a, mp3, wav, flac, ogg, opus, aac, wma, mp4, mov, mkv, webm, avi, mxf.
2. The choices of the run — ADVANCED, the blocks LOCAL MODELS, API: SERVICES AND MODELS, RECOGNITION, RESULT:
 - **LOCAL MODELS** — GIGAAM · ONNX (Russian speech), WHISPER TURBO · WHISPER.CPP (English and mixed), WHISPER LARGE-V2 · WHISPER.CPP (more accurate on hard recordings, many times heavier). One not installed is shown dimmed. With an NVIDIA card the choice is GIGAAM and WHISPER LARGE-V2, no turbo: on CUDA turbo repeats phrases (over five minutes of English speech one phrase came 25 times in a row, 42 % of the words wrong), while large-v2 is as fast there and gets 7 % of the words wrong;
 - **API: SERVICES AND MODELS** — OpenAI, xAI, Gemini, AssemblyAI, GigaChat, each with its own key; a press on a service sends the files to the cloud; the OpenAI model comes from the list of the service, `gpt-transcribe` by default; xAI and AssemblyAI choose their model themselves;
 - **LANGUAGE** — AUTO, RU, EN;
 - **SPEAKERS** — DO NOT SPLIT, AUTO (Community-1 where it is installed, Sortformer otherwise), COMMUNITY-1 (more accurate, needs the Hugging Face key), SORTFORMER (up to four speakers, no key); longer: a pass of its own after recognition. **HOW MANY** — the number of

speakers when it is known: Community-1 takes it. With a cloud — DO NOT SPLIT or SPLIT: the service splits itself;

- **PAUSES (VAD)** with GigaAM — where a long recording is cut into pieces of 12–25 seconds: SILERO VAD — in the pauses of speech, BY TIME — at the quietest moment; no sound is thrown away either way;
- **FILES** — MD, TXT, SRT, VTT, JSON;
- **DONE ONES** — SKIP a file whose transcript is already there, or REDO.

3. RUN. The progress of every file is in the list and in the journal; STOP at the bottom stops, the finished pieces are not lost: the next run continues the file from them.

The transcript lies next to the source: `NAME.md`, `.txt`, `.srt`, `.vtt`, `.transcript.json`. The window remembers the choices of the run; recordings are transcribed with them too.

The Journal Section

The logs of the program by days — errors stay visible after a restart. The journal panel on the right: LOGS opens the `logs` folder, CONFIG — the `config` folder; copy the whole text, clear (the files stay), expand to the whole window.

SETUP

- **Components** — what is installed, what is recommended for this machine, whether there is a newer one, how much to download; INSTALL ALL puts in place everything recommended, in order. The heavy part for NVIDIA is offered only with an NVIDIA card; Whisper for files there is Whisper on CUDA with the large-v2 model, and Whisper for files on the CPU with the turbo model is "not needed on this machine".
- **KEYS OF THE SERVICES** — Hugging Face, OpenAI, xAI, ElevenLabs, AssemblyAI, Google Gemini, GigaChat: typed past the journal, CHECK asks the service itself. A key is kept in `config\api_key_*.txt` and is never shown — only "there" and its last characters.
- **MICROPHONE** — a check: the microphone is listened to for a few seconds, the level is shown as it comes; nothing is recorded.

SETTINGS

Here is what is set once:

- **HOTKEYS** — a table: dictation Right Alt + F12, its cancel Esc, recording Right Alt + F11.
- **SOUND OF DICTATIONS** — how long the sound is kept.
- **COLOURS** — four themes of the window; **FONT OF THE WINDOW, FONT OF THE JOURNAL** — the family and the size, the chevrons of the spinner step it by half a point; **BACK TO THE DEFAULTS** brings back the theme, the fonts and the sizes.

The engine, the models, the text mode, the microphone and the panel are under **ADVANCED** of the Dictations page, the dictionary in its own section.

The language of the window — RU or EN in the header.

What Goes to the Network

- The sound of a dictation — only with a cloud engine, and only to its service.
- The text of a dictation — only with a text mode other than "Original text", and only to the address of the model.
- The transcript of a recording — to the model, when you make the minutes of the meeting.
- Recordings and files are recognized on this machine unless **CLOUD** is chosen on the Files page.
- The list of models of a service — the window asks it only for the list, with its own key.