

Audion Voice Tools

[English](#) · [Руководство](#)

Содержание

- [Зачем это сделано](#)
- [Что умеет](#)
- [Что нужно доустановить](#)
- [Принцип](#)
- [Где что лежит](#)
- [Дальше](#)
- [Техническая часть](#)
 - [FFmpeg и драйвер NVIDIA](#)
 - [Портативность](#)

Диктовка в любое окно и расшифровка записей — одна программа для Windows. Нажали хоткей, говорите — текст встаёт туда, где курсор. Записи совещаний и видео с диска превращаются в текст, субтитры и протокол встречи рядом с исходником.

Зачем это сделано

Распознавание речи нужно в двух разных обстоятельствах.

Быстро и сейчас. Продиктовать сообщение, письмо, задачу — и сразу продолжить работу. Здесь важны отклик и то, чтобы программа не мешала: капсула вверху экрана, хоткей, текст на месте курсора, буфер обмена возвращается прежним.

Много и надолго. Часовое совещание, запись целого дня, папка видео. Здесь важнее, чтобы ничего не терялось и не уезжало наружу без нужды: запись пишется на диск сразу, распознаётся на этой машине, текст ложится рядом с исходником.

Что умеет

- **Диктовка** — Right Alt + F12 начинает и заканчивает, Esc отменяет. Капсула показывает уровень, время и последние слова. Распознаёт локальная модель на этой машине — GigaAM v3 для русского по умолчанию, Parakeet и Whisper на выбор — без сети и без ключа — или облако по своему ключу: xAI, OpenAI, ElevenLabs.
- **Режимы текста** — кнопка  на капсуле: Исходный текст, Магия, Слегка, Значительно, Списанием, Структурированно, Протокол встречи. Текст оформляет языковая модель после остановки, до вставки: по умолчанию GPT-6 Luna, или xAI, или свой сервер; модели — из списка самого сервиса.
- **Словарь** — свой раздел, по строке на запись: термины, аббревиатуры, сокращения и единицы измерения, как их писать; замены «сказано -> написано».
- **История диктовок** — по дням, с поиском: копировать, прослушать, распознать заново, скачать звук, вставить ещё раз.
- **Рекордер** — Right Alt + F11: запись с микрофона в библиотеку, пауза, восстановление после сбоя, расшифровка сразу после остановки.
- **Протокол встречи** — по расшифровке записи: тема, суть, решения, задачи с исполнителями и сроками, открытые вопросы.
- **Файлы** — записи и видео с диска: GigaAM для русской речи, Whisper для английской и смешанной, разделение по говорящим; MD, TXT, SRT, VTT и JSON рядом с исходником.
- **SETUP** — компоненты с докачкой и ходом установки, ключи сервисов с проверкой, проверка микрофона.

Что нужно доустановить

Движки и модели в раздачу не входят: у каждого своя лицензия, они ставятся с сайтов авторов в папку программы.

При первом запуске окно само предлагает то, что советуется этой машине: движок диктовки, FFmpeg, GigaAM и Whisper для файлов — без видеокарты NVIDIA всего около 3,7 ГБ. «Позже» откладывает вопрос; всё то же — в SETUP, там видно, что стоит и сколько скачивать.

Whisper зависит от видеокарты. Без NVIDIA это модель turbo на процессоре: из моделей Whisper процессору посильна только она. С видеокартой NVIDIA Whisper для файлов — Whisper на CUDA с моделью large-v2, и первый запуск предлагает всего около 5,6 ГБ. Turbo на такой машине не скачивается: на CUDA она повторяет фразы — на пяти минутах английской речи одна фраза шла 25 раз подряд, ошибок по словам было 42 %, — а large-v2 там так же быстра (пять

минут — за 18 секунд) и ошибается в 7 % слов. Turbo, оставшаяся с процессора, при установке Whisper на CUDA уходит в Корзину.

С видеокартой NVIDIA SETUP предлагает и тяжёлое — по желанию: GigaAM на CUDA, разделение по говорящим на видеокарте (Community-1, нужен ключ Hugging Face).

Принцип

Окно никогда не ждёт. Микрофон, распознавание, вставка, облако и модель режима работают на своих потоках; окно и капсула только показывают, что происходит.

Ничего не теряется. Облако не ответило — тот же дубль распознаёт GigaAM на этой машине. Модель режима не ответила или ответила не тем — вставляется текст движка. Отменённая диктовка хранит свой звук, запись после сбоя восстанавливается. Каждый такой случай отмечен в истории.

Где что лежит

Папка	Что там
library\	библиотека: диктовки и записи (<code>library.sqlite3</code>), записи — в <code>recordings\ГОД\МЕСЯЦ\</code> с расшифровками
config\	настройки: диктовки (<code>dictation.json</code>), словарь (<code>dictionary.txt</code>), окна (<code>window_settings.json</code>), файлов (<code>app_settings.yaml</code>); ключи сервисов — <code>api_key_*.txt</code>
logs\	журналы по дням
Tools\ , models\	движки и модели, поставленные из SETUP

Дальше

- [Руководство](#) — диктовка, режимы, словарь, записи, файлы, SETUP по шагам.

Техническая часть

FFmpeg и драйвер NVIDIA

Каждая сборка FFmpeg собрана под конкретную версию заголовков аппаратного кодирования, и каждая требует своего минимума драйвера. Самая свежая сборка на старом драйвере не ускоряет, а ломает аппаратный путь — поэтому сборка подбирается по драйверу, а не по номеру версии.

Драйвер NVIDIA	Какую FFmpeg ставит программа
видеокарты NVIDIA нет	самую новую стабильную сборку Guap, не ниже 9.0.2
610.88 (LTS) и новее	самую новую стабильную сборку Guap, не ниже 9.0.2
570.0 – 610.87	8.0.1 — доступны не все операции
551.76 – 569.xx	7.1 — доступны не все операции
ниже 551.76	никакую: установщик останавливается и говорит почему

610.88 — LTS-драйвер, под который собраны программы. Выйдет 9.1 — будет 9.1; если новая сборка не прошла пробу NVENC на драйвере, установщик возвращается к 9.0.2. Ниже 610.88 ставится прежняя ветка, при которой работает NVENC, и установщик говорит, что доступны не все операции. Сборка подбирается при установке: сменился драйвер — FFmpeg, поставленная снова, подберётся под него.

Портативность

Программа живёт в своей папке: в Windows ничего не пишется, установки не нужно. Окно — `App\AudionVoiceTools.exe` со своей средой .NET, запуск — `Start.exe`. Расшифровку файлов делает консольная команда `voice_cli.cmd` на Python из `runtime\` той же папки.