

Audion Voice Tools — руководство

[English](#) · [О программе](#)

Содержание

- [Первый запуск](#)
- [Диктовка](#)
- [Трей](#)
- [Раздел «Диктовки»](#)
- [Раздел «Записи» и рекордер](#)
- [Раздел «Файлы»](#)
- [Раздел «Журнал»](#)
- [SETUP](#)
- [SETTINGS](#)
- [Что уходит в сеть](#)

Первый запуск

1. Запустите `Start.exe`. Открывается окно, вверху экрана появляется капсула — узкая полоска, в трее — значок программы. Windows 11 прячет новые значки троя за стрелкой «^»: перетащите значок на панель задач, чтобы он был на виду.
2. Окно спрашивает об установке компонентов, которые советуются этой машине: движок диктовки, FFmpeg, GigaAM и Whisper для файлов; с видеокартой NVIDIA вместо Whisper для файлов — Whisper на CUDA и Whisper large-v2 (почему — в разделе «Файлы»). У каждого компонента галочка, у движка диктовки — выбор модели: первой стоит GigaAM v3, лучшая модель для русского языка, ниже — модели для других языков. УСТАНОВИТЬ ставит отмеченное по очереди; ход виден в SETUP — шаг, проценты, скорость, оставшееся время. «Установить позже» откладывает вопрос до следующего запуска; ОТКРЫТЬ SETUP закрывает вопрос и сразу открывает лист SETUP.

3. Пока ставится движок диктовки, кнопка ДИКТОВКА в капсуле притушена. Когда в разделе «Журнал» появилась строка `[OK] gigaam-...: load ..., warm-up ...`, диктовка готова.

Диктовка

- **Right Alt + F12** — начать; ещё раз — закончить и вставить. **Esc** — отменить.
- Наведите указатель на капсулу или нажмите на неё — она раскроется: ДИКТОВКА и ЗАПИСЬ, каждая со своим хоткеем. Правая кнопка по капсуле — то же меню, что `v`: режимы текста, движок диктовки, положение панели.
- Текст вставляется туда, где был курсор, через буфер обмена; прежнее содержимое буфера возвращается.

Пока вы говорите, капсула показывает:

Слева направо	Что это
×	отменить: текст не вставляется, звук остаётся в истории
точка, полоски, время	идёт запись, уровень микрофона, сколько прошло
↶	начать заново: этот дубль отменяется, новый начинается сразу
последние слова	что услышано (у облачных потоков — по ходу речи)
название движка	GigaAM, Whisper или облако со значком ☁
красный квадрат	закончить и вставить
v	режимы текста и движок диктовки

После остановки капсула пишет «Распознаю...», при режиме текста — «Обработка: ...», затем на мгновение показывает вставленный текст.

Движок диктовки

Раздел «Диктовки» → ДОПОЛНИТЕЛЬНО, блоки ЛОКАЛЬНЫЕ МОДЕЛИ и API: СЕРВИСЫ И МОДЕЛИ; тот же выбор — `v` капсулы → «Движок диктовки » и меню трея:

- **Локальные модели** на `transcribe.cpp` — на этой машине, без сети и без ключа; список во всю ширину окна, строка на модель: название, для чего она и её языки, размер, «установлена» или УСТАНОВИТЬ. GigaAM v3 — русская речь, по умолчанию; Parakeet V3 — 25 европейских языков; Parakeet EN — английский; Whisper Small, Medium, Turbo, Large v3, Large v2 —

многоязычные. УСТАНОВИТЬ скачивает модель с Hugging Face, сверяет по SHA-256 и выбирает её; те же модели — в SETUP. Выбор другой модели загружает её сразу, диктовка под ней не прерывается. Строка СЕЙЧАС называет то, что работает: `gigaam-v3-e2e-rnnt-Q8_0 · Vulkan0`.

- **xAI, OpenAI Realtime, ElevenLabs** — потоки облака: слова идут в капсулу, пока вы говорите.
- **OpenAI · после остановки** — один запрос после остановки к входу, совместимому с OpenAI; лучше всех пишет английские термины.

МОДЕЛЬ облака — из списка самого сервиса, кнопка со стрелкой: OpenAI Realtime — по умолчанию `gpt-realtime-whisper`, OpenAI после остановки — `gpt-transcribe`; xAI и ElevenLabs выбирают модель сами. Список окно спрашивает у сервиса раз в сутки, ☹ рядом — сразу: модели у сервисов меняются каждый месяц. ЯЗЫК РЕЧИ облака — АВТО, RU, EN.

Облаку нужен свой ключ (SETUP → КЛЮЧИ СЕРВИСОВ). Облако не ответило за восемь секунд — тот же дубль распознаёт GigaAM на этой машине, а история отмечает это у диктовки.

Режимы текста

√ капсулы — режимы, как у Letterly; тот же выбор — «Диктовки» → ДОПОЛНИТЕЛЬНО → ОБРАБОТКА И САММАРИЗАЦИЯ. Режим, выбранный во время диктовки, действует уже на неё.

Режим	Что делает модель
Исходный текст	ничего: вставляется текст движка (по умолчанию)
Магия	сама выбирает оформление по тексту и по окну, куда он встанет: слегка, грамотно, списком, по частям, протоколом
Слегка	пунктуация, заглавные буквы, абзацы; убраны слова-паразиты и случайные повторы, слова те же
Значительно	грамотный письменный текст: исправлены грамматика и оговорки, смысл и подробности те же
Списком	каждая мысль или задача — пункт «- »
Структурированно	смысловые части с короткими заголовками
Протокол встречи	тема, суть, решения, задачи с исполнителями и сроками, открытые вопросы

Обработка и саммаризация включаются с сервисом облака; с локальной моделью — когда в config есть ключ OpenAI или xAI. СЕРВИС — OPENAI, XAI или СВОЙ СЕРВЕР, МОДЕЛЬ — из его

списка. По умолчанию — OpenAI и `gpt-6-luna`: протокол встречи она составляет с наибольшим уровнем рассуждения (`max`), остальные режимы — с низким (`low`), чтобы вставка не ждала. Нужен ключ своего сервиса, и уходит он только ему: `api.openai.com` — ключ OpenAI, `api.x.ai` — ключ xAI. Свой сервер — Ollama, LM Studio — спрашивается без ключа: адрес вроде `http://localhost:11434/v1` и модель из его списка.

Модель ждёт 15 секунд, длинный текст — дольше, протокол встречи — до двух минут. Не ответила или ответила не тем — вставляется текст движка; история хранит оба текста.

Словарь

Раздел «Словарь», файл `config\dictionary.txt`. Строка раздела — одна запись: КАК СКАЗАНО → КАК ПИСАТЬ, у термина левое поле пустое. СТРОКА добавляет запись, ЗАМЕТКА — заметку, × убирает; поиск сужает список; раздел растягивается на всё окно. В файле те же записи, по одной на строку:

```
ФГИС ТП
Kubernetes
гитхаб экшнс -> GitHub Actions
километров в час -> км/ч
# единицы
градусов цельсия -> °C
```

- **Термин** — слово, аббревиатура, сокращение или единица, как её писать. Его подсказывают облакам (xAI и ElevenLabs — как ключевые слова, OpenAI после остановки — в подсказке), его соблюдают режимы текста и протокол встречи, его берёт расшифровка файлов.
- **Замена** `сказано -> написано` (или `→`, `=>`) — делается в тексте каждой диктовки по целым словам, без учёта регистра; работает и для GigaAM, которому слова не подсказать.
- `#` в начале строки — заметка.

Словарь сохраняется сам, через секунду после ввода, и действует со следующей диктовки.

Микрофон и панель

- Микрофон — «Диктовки» или «Записи» → ДОПОЛНИТЕЛЬНО → УСТРОЙСТВА → МИКРОФОН, меню трея → «Микрофон »»: ПО УМОЛЧАНИЮ — микрофон Windows по умолчанию.
- Панель — «Диктовки» → ДОПОЛНИТЕЛЬНО → УСТРОЙСТВА → ПАНЕЛЬ: ВВЕРХУ или ВНИЗУ экрана (меню трея или √ капсулы → «Положение панели »); ВИДНА В ПОКОЕ (меню трея → «Показывать плавающую панель»): выключено — капсула появляется только на время диктовки или записи.

Выбор у капсулы, трея и окна один: изменённый в одном месте сразу виден в других. ДОПОЛНИТЕЛЬНО открыто, пока на странице не было первой работы — диктовки, записи, прогона файлов, — затем само сворачивается в строку; дальше оно такое, каким его оставили: шеврон справа или нажатие на заголовок открывает и сворачивает его.

Трей

Левая кнопка по значку открывает окно. Правая — меню:

- **Вставить последнюю диктовку** — в окно, которое было впереди;
- **История** — раздел «Диктовки»;
- **Начать запись (Right Alt + F11)** — с красной точкой, пока запись идёт;
- **Движок диктовки >, Микрофон >, Положение панели >, Показывать плавающую панель;**
- **Открыть Audion Voice Tools, Настройки, Выход.**

Заккрытие окна крестиком оставляет программу в трее: диктовка работает. Выход — только из меню трея.

Раздел «Диктовки»

Диктовки по дням, новые сверху; поиск — по началу слов, без учёта регистра. Наведите указатель на диктовку — появятся кнопки: копировать, распознать заново (GigaAM этой машины, текст заменится), прослушать, скачать звук в «Загрузки», удалить (запись — из библиотеки, звук — в Корзину). Подсказка текста — куда вставлено, движок, режим, длительность, сколько ждали вставку.

Правая кнопка: **Вставить ещё раз, Копировать текст движка — до обработки** (у диктовок с режимом), **Закрепить наверху** — закреплённая хранит звук и после срока.

Звук диктовок хранится 7 дней (SETTINGS → ЗВУК ДИКТОВОК: 7 дней, 30 дней, всегда, не хранить); текст остаётся всегда. Диктовки 2.3 приходят в историю сами, без звука.

Раздел «Записи» и рекордер

- **Right Alt + F11**, кнопка ЗАПИСАТЬ в разделе или ЗАПИСЬ в капсуле — начать. Капсула красная: время, уровень; **■** — пауза (микрофон открыт, в файл ничего не пишется), красный квадрат — сохранить, **×** — отменить (файл — в Корзину).

- Запись — 48 кГц моно в `library\recordings\ГОД\МЕСЯЦ\`. После сбоя программы недописанная запись восстанавливается при следующем старте. Диктовка во время записи работает.
- После сохранения запись сама идёт в расшифровку с выбором раздела «Файлы» («Записи» → ДОПОЛНИТЕЛЬНО → ПОСЛЕ ЗАПИСИ: РАСШИФРОВЫВАТЬ СРАЗУ или ВРУЧНУЮ; ВЫБРАТЬ открывает «Файлы»).

Кнопки карточки записи: прослушать, открыть расшифровку, **протокол встречи**, показать в папке, скачать в «Загрузки», расшифровать (заново).

Протокол встречи составляет модель «Диктовки» → ДОПОЛНИТЕЛЬНО → ОБРАБОТКА И САММАРИЗАЦИЯ по расшифровке: файл `ИМЯ.protocol.md` рядом с записью, открывается сам. Готовый протокол кнопка открывает; правая кнопка — составить заново, прежний уходит в Корзину.

Раздел «Файлы»

1. ДОБАВИТЬ ФАЙЛЫ или ДОБАВИТЬ ПАПКУ — записи и видео: m4a, mp3, wav, flac, ogg, opus, aac, wma, mp4, mov, mkv, webm, avi, mx4.
2. Выбор прогона — ДОПОЛНИТЕЛЬНО, блоки ЛОКАЛЬНЫЕ МОДЕЛИ, API: СЕРВИСЫ И МОДЕЛИ, РАСПОЗНАВАНИЕ, РЕЗУЛЬТАТ:
 - **ЛОКАЛЬНЫЕ МОДЕЛИ** — GIGAAM · ONNX (русская речь), WHISPER TURBO · WHISPER.CPP (английская и смешанная), WHISPER LARGE-V2 · WHISPER.CPP (точнее на трудных записях, в разы тяжелее). Неустановленный показан притушенным. С видеокартой NVIDIA выбор — GIGAAM и WHISPER LARGE-V2, turbo нет: на CUDA turbo повторяет фразы (на пяти минутах английской речи одна фраза шла 25 раз подряд, ошибок по словам — 42 %), а large-v2 там так же быстра и ошибается в 7 % слов;
 - **API: СЕРВИСЫ И МОДЕЛИ** — OpenAI, xAI, Gemini, AssemblyAI, GigaChat, у каждого свой ключ; нажатие на сервис отправляет файлы в облако; модель OpenAI — из списка сервиса, по умолчанию `gpt-transcribe`; xAI и AssemblyAI выбирают модель сами;
 - **ЯЗЫК** — АВТО, RU, EN;
 - **ГОВОРЯЩИЕ** — НЕ РАЗДЕЛЯТЬ, АВТО (Community-1, где он установлен, иначе Sortformer), COMMUNITY-1 (точнее, нужен ключ Hugging Face), SORTFORMER (до четырёх говорящих, без ключа); дольше: отдельный проход после распознавания. **СКОЛЬКО** — число говорящих, если оно известно: его берёт Community-1. У облака — НЕ РАЗДЕЛЯТЬ или РАЗДЕЛИТЬ: разделяет сам сервис;

- **ПАУЗЫ (VAD)** у GigaAM — где резать длинную запись на куски по 12–25 секунд: SILERO VAD — в паузах речи, **РОВНО ПО ВРЕМЕНИ** — в самом тихом месте; звук не выбрасывается ни в одном случае;
- **ФАЙЛЫ** — MD, TXT, SRT, VTT, JSON;
- **ГОТОВЫЕ** — ПРОПУСКАТЬ файл, у которого расшифровка уже есть, или ПЕРЕДЕЛАТЬ.

3. ЗАПУСТИТЬ. Ход каждого файла — в списке и в журнале; СТОП внизу останавливает, готовые куски не пропадают: следующий прогон продолжит файл с них.

Расшифровка ложится рядом с исходником: `ИМЯ.md`, `.txt`, `.srt`, `.vtt`, `.transcript.json`.
Выбор прогона окно запоминает; с ним же расшифровываются записи.

Раздел «Журнал»

Журналы программы по дням — ошибки видны и после перезапуска. Панель журнала справа: LOGS открывает папку `logs`, CONFIG — папку `config`; копировать весь текст, очистить (файлы остаются), развернуть на всё окно.

SETUP

- **Компоненты** — что стоит, что советуется этой машине, есть ли новее, сколько скачивать; УСТАНОВИТЬ ВСЁ ставит по порядку всё рекомендованное. Тяжёлое для NVIDIA предлагается только при видеокарте NVIDIA; Whisper для файлов там — Whisper на CUDA с моделью large-v2, а у Whisper для файлов на процессоре с моделью turbo — «этой машине не нужен».
- **КЛЮЧИ СЕРВИСОВ** — Hugging Face, OpenAI, xAI, ElevenLabs, AssemblyAI, Google Gemini, GigaChat: ввод мимо журнала, ПРОВЕРИТЬ спрашивает сам сервис. Ключ хранится в `config\api_key_*.txt` и никогда не показывается — только «есть» и последние символы.
- **МИКРОФОН** — проверка: микрофон слушается несколько секунд, уровень виден по ходу; ничего не записывается.

SETTINGS

Здесь то, что задают один раз:

- **ХОТКЕИ** — таблица: диктовка Right Alt + F12, её отмена Esc, запись Right Alt + F11.
- **ЗВУК ДИКТОВОК** — сколько хранить звук.

- **ЦВЕТА** — четыре темы окна; **ШРИФТ ОКНА, ШРИФТ ЖУРНАЛА** — семейство и размер, шевроны счётчика меняют его на полпункта; К УМОЛЧАНИЯМ возвращает тему, шрифты и размеры.

Движок, модели, обработка текста, микрофон и панель — в ДОПОЛНИТЕЛЬНО страницы «Диктовки», словарь — в своём разделе.

Язык окна — RU или EN в шапке.

Что уходит в сеть

- Звук диктовки — только при облачном движке, и только его сервису.
- Текст диктовки — только при режиме текста, кроме «Исходного», и только по адресу модели.
- Расшифровка записи — модели, когда вы составляете протокол встречи.
- Записи и файлы распознаются на этой машине, пока в разделе «Файлы» не выбрано ОБЛАКО.
- Список моделей сервиса — окно спрашивает у него только список, его же ключом.