VOICE2TEXT AI

用户手册

Voice2Text

从声音到文字，支持多种语言。

使用人工智能转写音频和视频文件，修改文本，并保存到任何位置。

离线 Whisper · 12 种格式 · 云存储 · 9 种语言



版本 1.1 · macOS · Windows · Linux · 中文（简体）

目录

01	欢迎使用 Voice2Text	3
02	安装	4
03	主窗口	5
04	选择引擎和语言	6
05	在线服务：音频将离开你的电脑	8
06	转写过程中	9
07	修改转写	10
08	保存和导出	12
09	保存到云端	14
10	设置和语言	16
11	键盘快捷键	18
12	常见问题	19

01 欢迎使用 Voice2Text

> Voice2Text 可将音频和视频录音转换为文字。采访、会议、课程、播客、语音消息：把文件拖进窗口，选择人工智能引擎，几分钟后就能得到可编辑、可保存的文本。

主要功能

- **支持任意格式。**MP3、M4A、WAV、FLAC、OGG、AAC、AIFF，以及 MP4、MOV、MKV、WebM 视频等；安装 ffmpeg 后几乎支持所有格式。
- **无需联网也能用。**Whisper 直接在电脑上运行，音频不会离开你的桌面。
- **识别说话人。**使用 AssemblyAI 和 Deepgram 时会区分不同的声音，并可为每个人命名。
- **轻松修改。**文本按句子划分并与音频关联：点一下即可重听对应位置。
- **多种导出格式。**Word、PDF、ODT、RTF、TXT、Markdown、HTML、SRT 和 VTT 字幕、CSV、JSON，以及可重新打开的项目文件。
- **直接保存到云端。**通过电脑上的同步文件夹保存到 Google Drive、OneDrive、Dropbox、Box、Proton Drive、iCloud 等。
- **九种界面语言。**意大利语、英语、法语、西班牙语、德语、中文、日语、俄语和阿拉伯语；翻译是可修改的 XML 文件。

系统要求

macOS 11 或更高版本（Apple 芯片或 Intel 的 Mac）、64 位 Windows 10/11，或 64 位 Linux（Ubuntu、Debian、Fedora 及其衍生版）。使用离线 Whisper 建议至少 8 GB 内存；在线服务只需联网。

02 安装

> 下载适合你系统的文件，像安装其他软件一样安装即可。

macOS

打开 `Voice2Text_1.1.1_universal.dmg`，将 Voice2Text 图标拖到**应用程序**文件夹。首次打开时 macOS 可能提示无法验证开发者：打开**系统设置 → 隐私与安全性**并选择**仍要打开**，或右键点击应用并选择**打开**。只需操作一次。

Windows

运行 `Voice2Text_1.1.1_x64-setup.exe`。如果出现蓝色的 Windows SmartScreen 窗口，点击**更多信息**，再点击**仍要运行**。安装无需管理员权限。

Linux

在 Ubuntu 和 Debian 上安装 `.deb` 包（双击或运行 `sudo apt install ./Voice2Text_1.1.1_amd64.deb`）；在 Fedora 上安装 `.rpm` 包。也可以将 `.AppImage` 设为可执行（`chmod +x`）后直接运行。

> **可选：ffmpeg。**Voice2Text 能自行读取常见格式。对于少见格式（OPUS、WMA、AMR、AVI…），请安装 ffmpeg：Mac 上运行 `brew install ffmpeg`，Windows 上运行 `winget install ffmpeg`，Linux 上运行 `sudo apt install ffmpeg`。程序会自动找到它。

03 主窗口

> 所有操作都在一个窗口中完成：左侧是文件列表，中间是当前选中的文件。



图 1. 开始界面：把文件拖到框内，或点击“选择文件”。

- **顶部栏**：**添加文件**、**打开转写**（已保存的 `.v2t` 项目）以及**设置**齿轮。
- **左侧文件列表**：每个文件显示状态（就绪、排队中、正在转写、已转写、错误）和进度。文件名后的圆点表示有未保存的修改。
- **中间区域**：转写设置、进度或待修改的文本。

添加文件

将一个或多个音频或视频文件拖到窗口，或点击**添加文件**（`⌘O` / `Ctrl+O`）。数量不限，会依次转写。拖入 `.v2t` 文件即可重新打开已保存的转写。

04 选择引擎和语言

转写前，先选择由谁来完成：电脑上的 Whisper，还是在线服务。

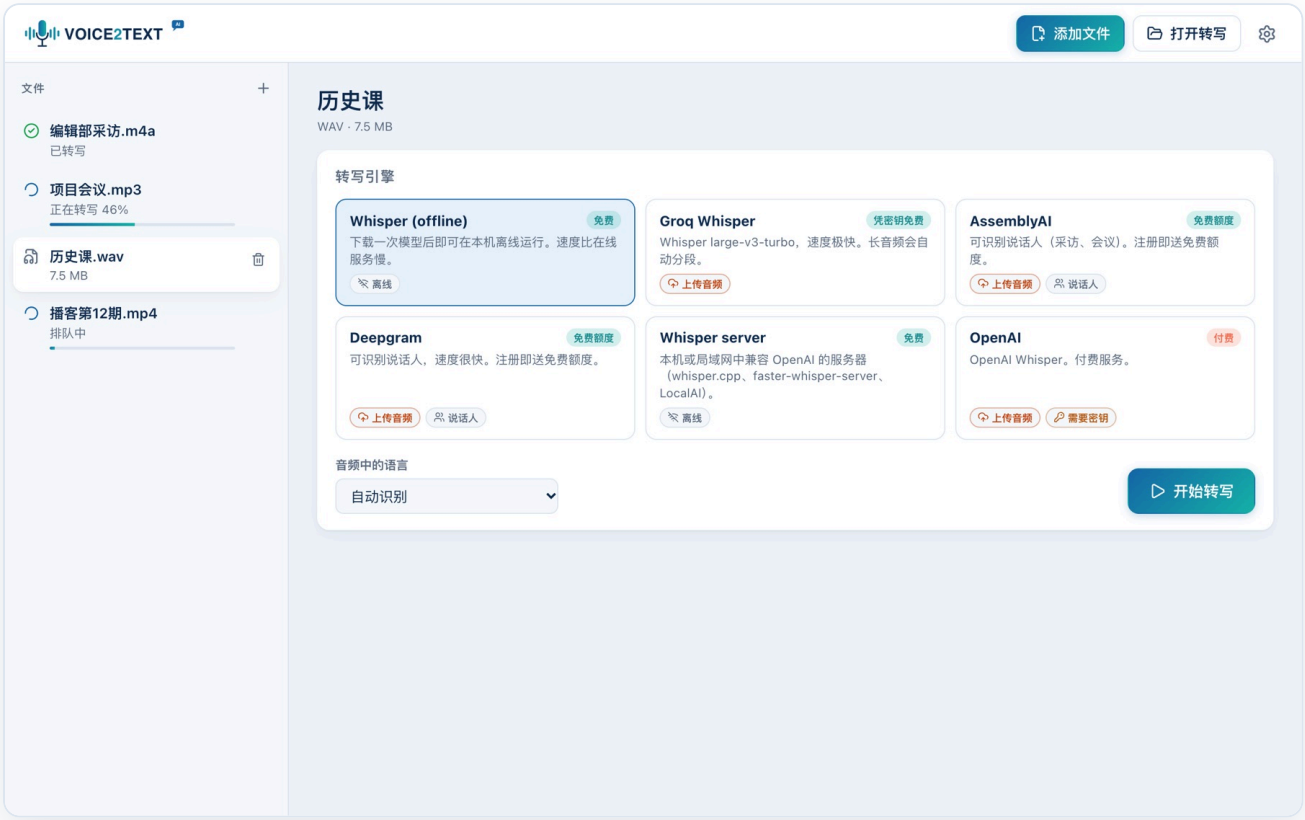


图 2. 为选中的文件选择引擎和语言。

引擎	费用	特点
Whisper（离线）	免费	无需联网；模型只需下载一次（45 至 600 MB）。速度比在线服务慢。
Groq Whisper	凭密钥免费	速度极快；长音频会自动分段。
AssemblyAI	免费额度	可识别说话人，适合采访和会议。
Deepgram	免费额度	可识别说话人，速度很快。
Whisper server	免费	本机或局域网中的 Whisper 服务器（whisper.cpp、faster-whisper-server、LocalAI）。
OpenAI	付费	OpenAI Whisper。

音频语言

选择**自动识别**时，Voice2Text 会自行判断所说的语言。如果你知道语言，最好直接选择：结果会更准确，尤其是开头几秒。

在线服务的密钥

Groq、AssemblyAI、Deepgram 和 OpenAI 需要个人密钥：在**设置 → 引擎**中，每个服务都有对应的输入框和“获取密钥”链接。注册 Groq 免费；AssemblyAI 和 Deepgram 注册即送免费额度。



图 3. 设置 → 引擎：Whisper 模型和各服务的密钥。

点击**开始转写**即可开始。有多个文件等待时，侧边栏还会出现**全部转写**按钮。

05 在线服务：音频将离开你的电脑

使用在线人工智能服务时，音频文件会通过互联网上传。Voice2Text 会明确告知并请你确认。



图 4. 首次上传到在线服务前必须阅读的提示。

首次使用 Groq、AssemblyAI、Deepgram 或 OpenAI 时会出现此提示。要继续，必须勾选**“我已阅读并理解”**；也可以直接改用**离线 Whisper**。

音频文件将上传到外部服务的服务器，该服务会依据其自身条款和隐私政策处理文件：可能会保存一段时间，也可能在其他国家/地区处理。

请勿上传包含机密信息的录音，也不要上传隐私或保密内容，以及个人、健康或第三方数据，除非你拥有相应权利并已获得被录音者的同意。

Voice2Text 及其作者不承担任何责任，包括上传文件的内容、保密性以及外部服务对文件的使用。责任完全由上传文件的人承担。

同意仅对单个服务有效，并会记录日期。只要选择了在线服务，就会显示橙色提示框。在**设置 → 关于**中可以重新阅读此提示，并让它再次出现。

音频涉及机密？请使用 **Whisper（离线）**：转写完全在你的电脑上进行。

06 转写过程中

> Voice2Text 会随时显示工作进度。

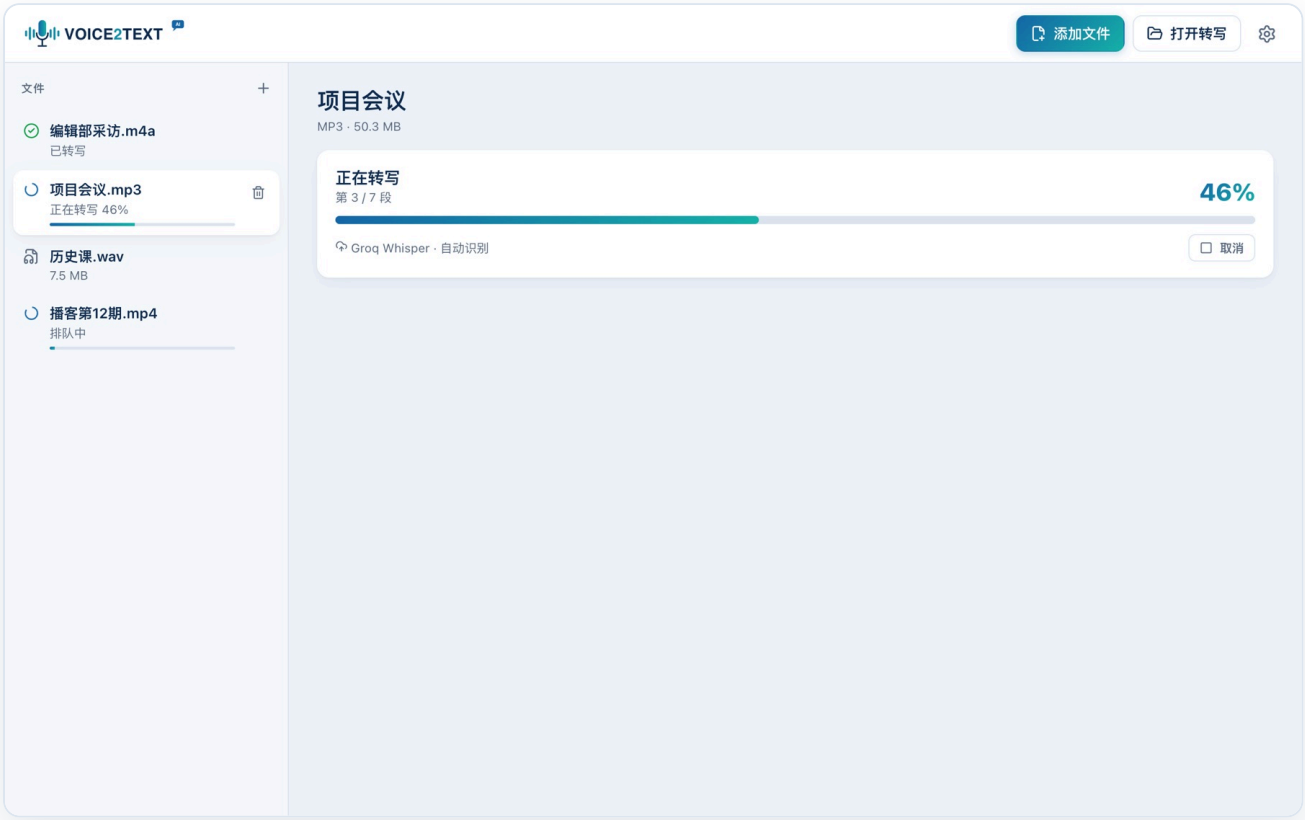


图 5. 使用 Groq 进行转写：进度和正在处理的分段。

- **正在读取音频**：在电脑上解码文件。
- **正在准备模型**（仅离线 Whisper，首次使用时）：下载并保存所选模型。
- **正在上传音频**和**正在转写**：长文件会显示当前处理的分段（“第 3 / 7 段”）。
- **等待服务**：使用免费方案时若超过请求限制，Voice2Text 会自动等待并重试。

你可以随时点击**取消**，也可以同时处理其他文件：文本完成后会收到通知。

07 修改转写

> 文本按句子划分，每句都有时间，若能识别还会标出说话人。



图 6. 编辑器：上方是音频播放器，下方是可编辑的句子；正在播放的句子会高亮显示。

收听

上方的播放器播放原始音频：空格键开始或暂停（未在输入时），按钮可前后跳转 5 秒，速度可在 0.75× 到 2× 之间选择。点击句子的时间即可收听；开启**跟随音频**后，正在播放的句子始终可见。

编辑

- 点击句子即可输入，修改立即生效。
- Enter 在光标处拆分句子；Shift+Enter 在同一句内换行。
- 在句首按 Backspace 会与上一句合并；**合并**按钮会与下一句合并。
- 垃圾桶图标可删除句子（例如被误识别为文字的噪音）。



图 7. 查找替换和说话人名称。

查找和替换

查找（⌘F / Ctrl+F）会高亮包含该词的句子；**全部替换**可一次性更正整篇转写中写错的名字。

说话人

使用 AssemblyAI 和 Deepgram 时，每句都有标签（说话人 A、说话人 B…）。在**说话人**面板中输入真实姓名，它会出现在所有地方，包括导出的文件。可在时间旁的菜单中更改某句的说话人、添加新的说话人，即使引擎无法识别也可以手动指定。

08 保存和导出

> 保存项目以便稍后继续修改，并按需要的格式导出文本。

.v2t 项目

保存（`⌘S` / `Ctrl+S`）会生成 `.v2t` 文件，包含转写、修改内容和说话人名称。通过**打开转写**（`⇧⌘O` / `Ctrl+Shift+O`）或拖到窗口中即可重新打开。`⇧⌘S` 可另存为其他名称。

导出

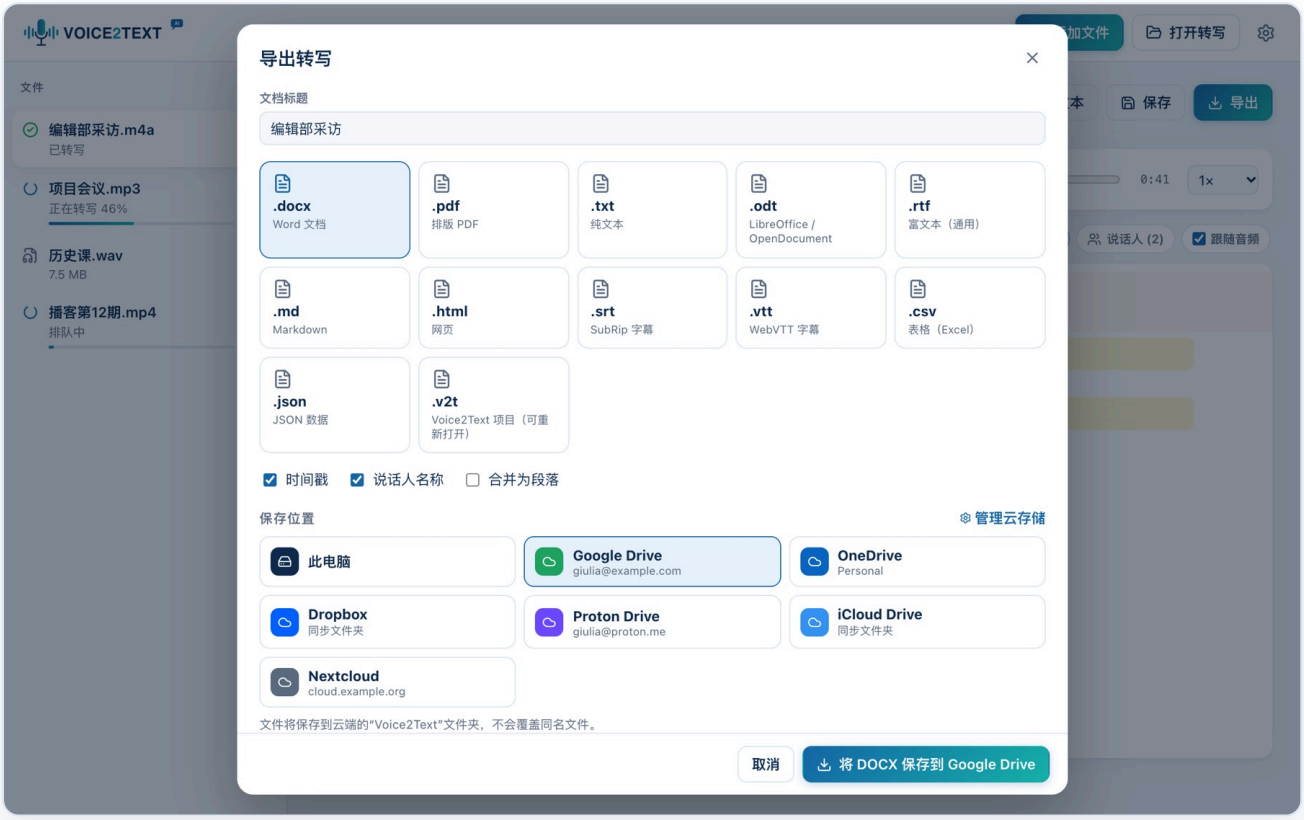


图 8. 导出窗口：格式、选项和保存位置。

点击**导出**（`⌘E` / `Ctrl+E`），输入文档标题并选择格式：

格式	适用于
.docx Word · **.odt** LibreOffice · **.rtf**	可编辑的文档，含标题和录音信息
.pdf	用于打印或发送；通过系统打印窗口生成（“存储为 PDF”）
.txt · **.md** · **.html**	纯文本、Markdown、网页
.srt · **.vtt**	视频字幕（YouTube、Premiere、DaVinci、VLC…）
.csv · **.json**	表格（Excel、Numbers）及供其他程序使用的数据
.v2t	可重新打开的项目，也可保存到云端

选项可决定是否包含**时间戳**和**说话人名称**，以及是否**合并为段落**。阿拉伯语等从右向左书写的语言会正确排版。

09 保存到云端

> Voice2Text 通过云服务应用在电脑上创建的同步文件夹，直接把文件保存到你的云端。无需连接账户，也无需输入密码。

工作原理

安装 Google Drive、OneDrive、Dropbox、Box、**Proton Drive**、iCloud Drive、pCloud、MEGA、Nextcloud 或 Sync.com 的应用后，电脑上会出现一个同步文件夹。Voice2Text 会自动找到它：在**导出**窗口的**保存位置**中选择云存储并点击按钮。文件会写入 `Voice2Text` 子文件夹，云服务应用随即上传。

- 文件绝不会被覆盖：如果已有同名文件，新文件会命名为“名称 (2)”。
- 任何格式都可以保存到云端，包括 `.v2t` 项目。
- PDF 通过打印窗口生成：在其中选择“存储为 PDF”和云存储文件夹。



图 9. 设置 → 云存储：电脑上找到的文件夹和 WebDAV 连接。

设置 → 云存储

- **文件夹名称**：在云端创建的子文件夹（默认 `Voice2Text`）。
- **同步文件夹**：找到的云存储列表；点击文件夹图标可打开。若缺少某个服务，下方链接会指向需要安装的应用。
- **添加其他文件夹**：用于其他云服务或网络存储（NAS）。
- **WebDAV**：适用于未安装应用的 Nextcloud、ownCloud、Koofr、pCloud、Yandex Disk 以及群晖/威联通 NAS。选择服务，填写地址、用户名和应用专用密码，然后点击**测试连接**。

Proton Drive 不允许其他应用直接访问账户，因此 Voice2Text 使用它的同步文件夹，并保持相同的端到端加密。

10 设置和语言

右上角的齿轮（`⌘,` / `Ctrl+,`）可打开设置。



图 10. 设置 → 常规。

- **常规**：界面语言、浅色/深色/自动外观、默认引擎和音频语言、自动开始转写。
- **引擎**：离线 Whisper 模型（极小、基础、小型、大型 turbo）、模型占用空间、服务密钥、本地 Whisper 服务器、是否检测到 ffmpeg。
- **云存储**：同步文件夹和 WebDAV（见上一章）。
- **语言**：界面翻译文件。
- **关于**：版本、许可、在线服务提示。

语言是 XML 文件

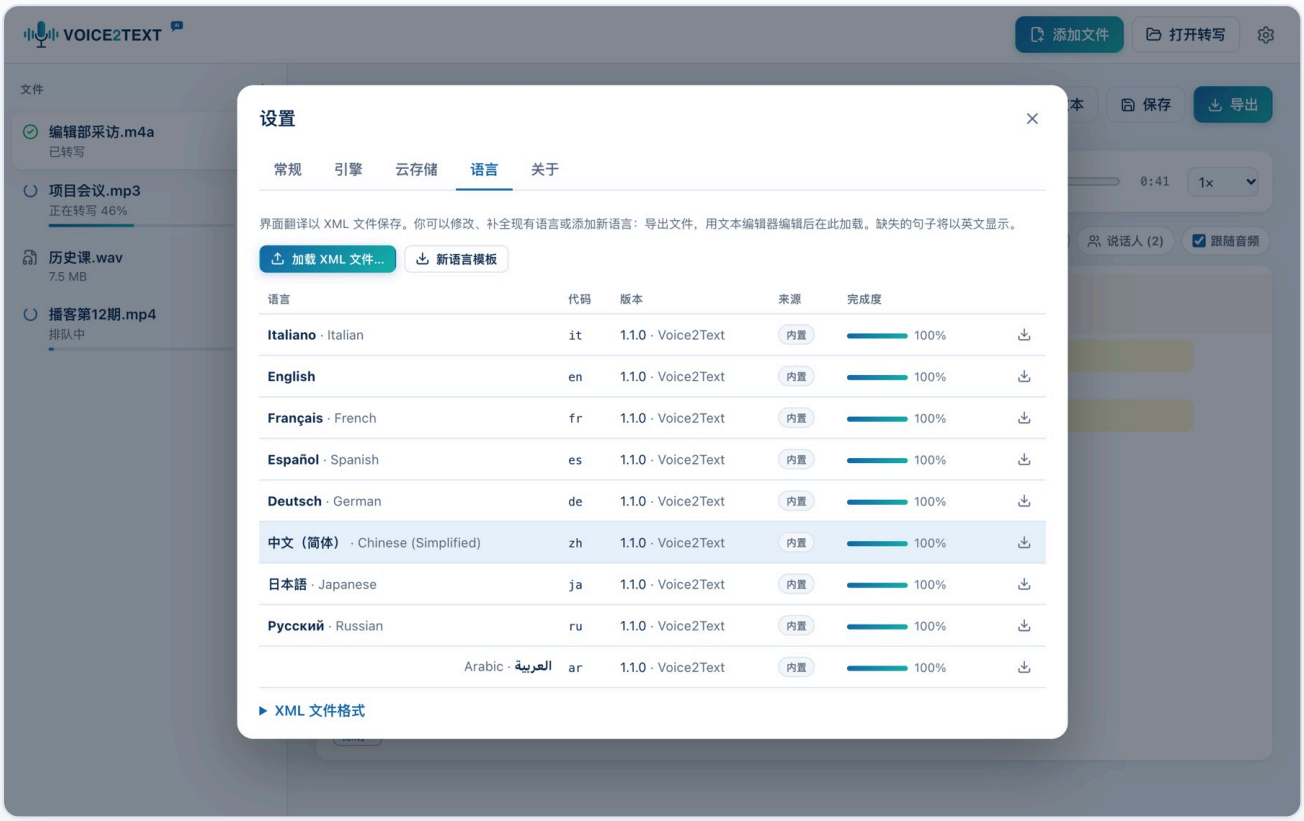


图 11. 设置 → 语言：可用语言、完成度和 XML 文件工具。

每种界面语言都是一个 XML 文件。要修改翻译或添加新语言：

1. 导出该语言文件（行内箭头），或下载**新语言模板**。
2. 用文本编辑器打开，只翻译标签之间的文字，保留 `key` 属性和花括号占位符（如 `{n}`）。
3. 点击**加载 XML 文件...**：该语言会立即出现在列表和语言菜单中。

```
<language code="pt" name="Português" english="Portuguese" dir="ltr" version="1.0">
  <string key="app.addFiles">Adicionar ficheiros</string>
  <string key="progress.piece">parte {i} de {n}</string>
</language>
```

与内置语言代码相同的文件会替换内置语言（用于修正），随时可以恢复原版。缺失的句子以英文显示；**完成度**一栏显示已翻译的比例。从右向左书写的语言请使用 `dir="rtl"`。

11 键盘快捷键

> 在 Windows 和 Linux 上用 Ctrl 代替 ⌘。

按键	操作
⌘O	添加音频或视频文件
⇧⌘O	打开已保存的转写（.v2t）
⌘S · ⇧⌘S	保存 · 另存为
⌘E	导出
⌘F	查找和替换
⌘,	设置
空格键	播放 / 暂停（不在文本中时）
Enter · Shift+Enter	拆分句子 · 句内换行
句首按 Backspace	与上一句合并
Esc	退出文本 · 关闭当前窗口

12 常见问题

> 最常见问题的解答。

无法读取我的文件格式。
安装 ffmpeg（见“安装”一章），或使用 AssemblyAI 或 Deepgram，它们接收原始文件，因此支持任意格式。

离线 Whisper 很慢。
在设置 → 引擎中选择**基础**或**极小**模型。“大型 turbo”最准确，但没有显卡时较慢。长文件用 Groq 会快得多。

文本是另一种语言或被翻译了。
不要使用自动识别，直接指定音频语言，然后点击**重新转写**。

“服务密钥无效”。
在设置 → 引擎中检查密钥：从服务网站重新复制，不要带空格。

我的云存储没有出现。
确认云服务应用已安装并正在同步，然后在设置 → 云存储中点击**重新查找**。也可以手动添加文件夹。

播放器显示“音频不可用”。
保存项目后，原始音频文件被移动或重命名了。文本仍然可用；如需重听，请把文件放回原位置。

我的密钥和设置保存在哪里?
只保存在你的电脑上，在应用的设置文件中，仅当前用户可读，绝不会发送给 Voice2Text。



VOICE2TEXT

从声音到文字，支持多种语言。

Voice2Text 由 overminder 开发。采用 PolyForm Strict 1.0.0 许可：允许非商业使用；不允许再分发、修改或创作衍生作品。

© 2026 overminder · 版本 1.1