

第一次打开软件-【批量API管理器操作文档】

前言

如果你是第一次使用，请不要因为看到字多而觉得害怕，其实并没有那么复杂，一旦你成功启动工具软件，你将打开新世界的大门。

硬件建议：如果没有20系列以上的Nvidia显卡，CPU就要求至少4核8线程，内存至少12G。

1、请先下载GPT-SoVITS v2

如果没有，请先下载GPT-SoVITS v2，批量API管理器是一款用于启动GPT-SoVITS高效使用的桌面应用工具，你必须电脑已经下载有GPT-SoVITS v2。

GPT-SoVITS是B站UP主花儿不哭大佬开源的一款声音工具，这个工具是免费的，但由于文件体积较大，下载工具或网盘可能会收费，建议先尝试迅雷。

【GPT-SoVITS作者原下载地址分享，尝试迅雷下载】

<https://huggingface.co/lj1995/GPT-SoVITS-windows-package/blob/main/GPT-SoVITS-v2-240821.7z>

【迅雷云盘搬运】

分享文件：GPT-SoVITS-v2-240821.7z

链接：<https://pan.xunlei.com/s/V09IAGPTkmUjCHUuRB3CnW2XA1#>

提取码：jshq

【夸克网盘搬运】

链接：<https://pan.quark.cn/s/88bad928becf>

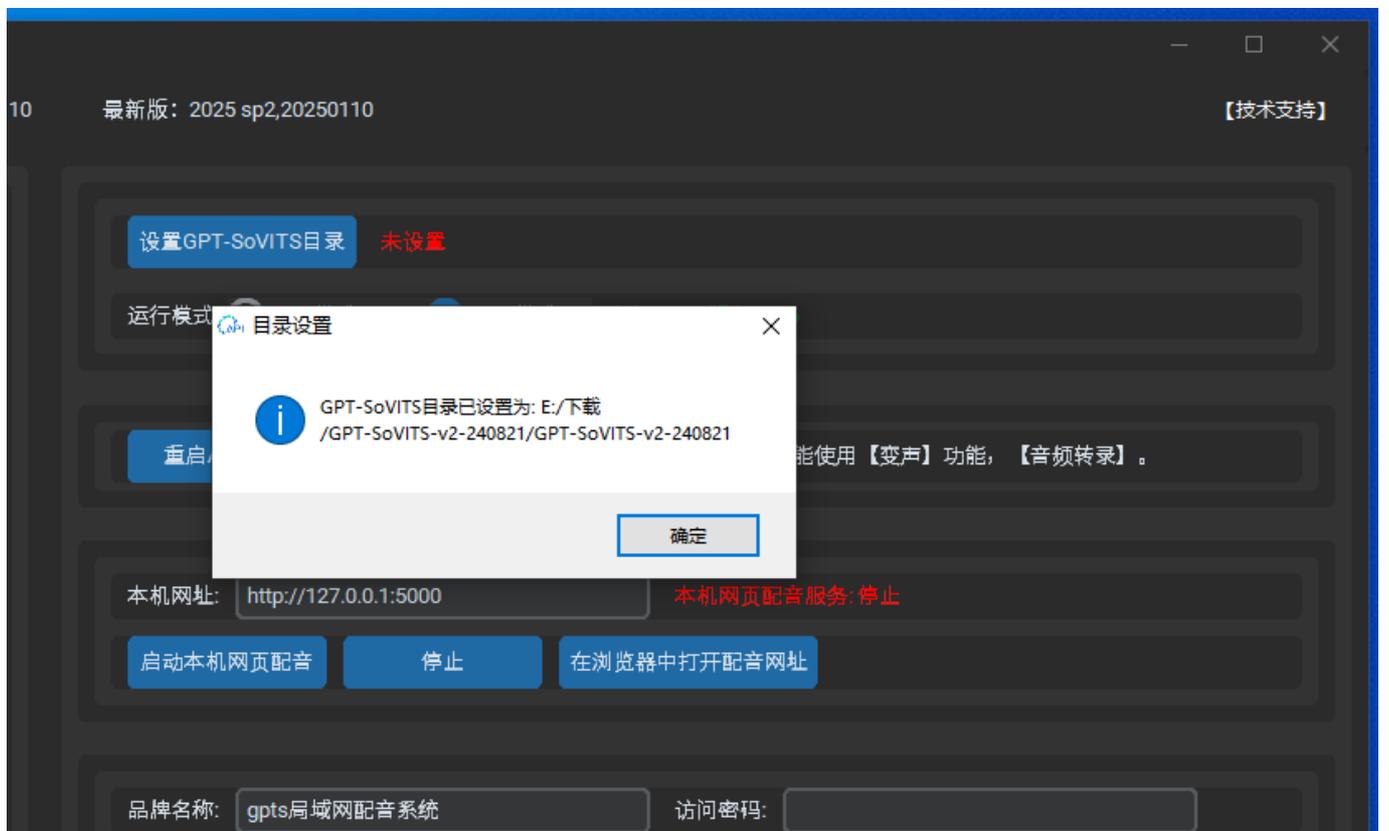
提取码：6sU5

下载后，解压缩放到任意一个分区下就行。

2、下载并配置批量API管理器

当用户下载软件包之后，找到名称为startapi_v2025sp2.7z，

进行解压缩后，并打开解压缩得到的文件夹，找到并运行startapi.exe，打开后设置界面。



要想玩转各中个性化的配音，你还需要不同的声音模型。当用户了解GPT-SoVITS v2之后，一般都会去自己制作不同的声音模型，或到网上搜索别人分享的模型，B站上就有不少UP主分享的。

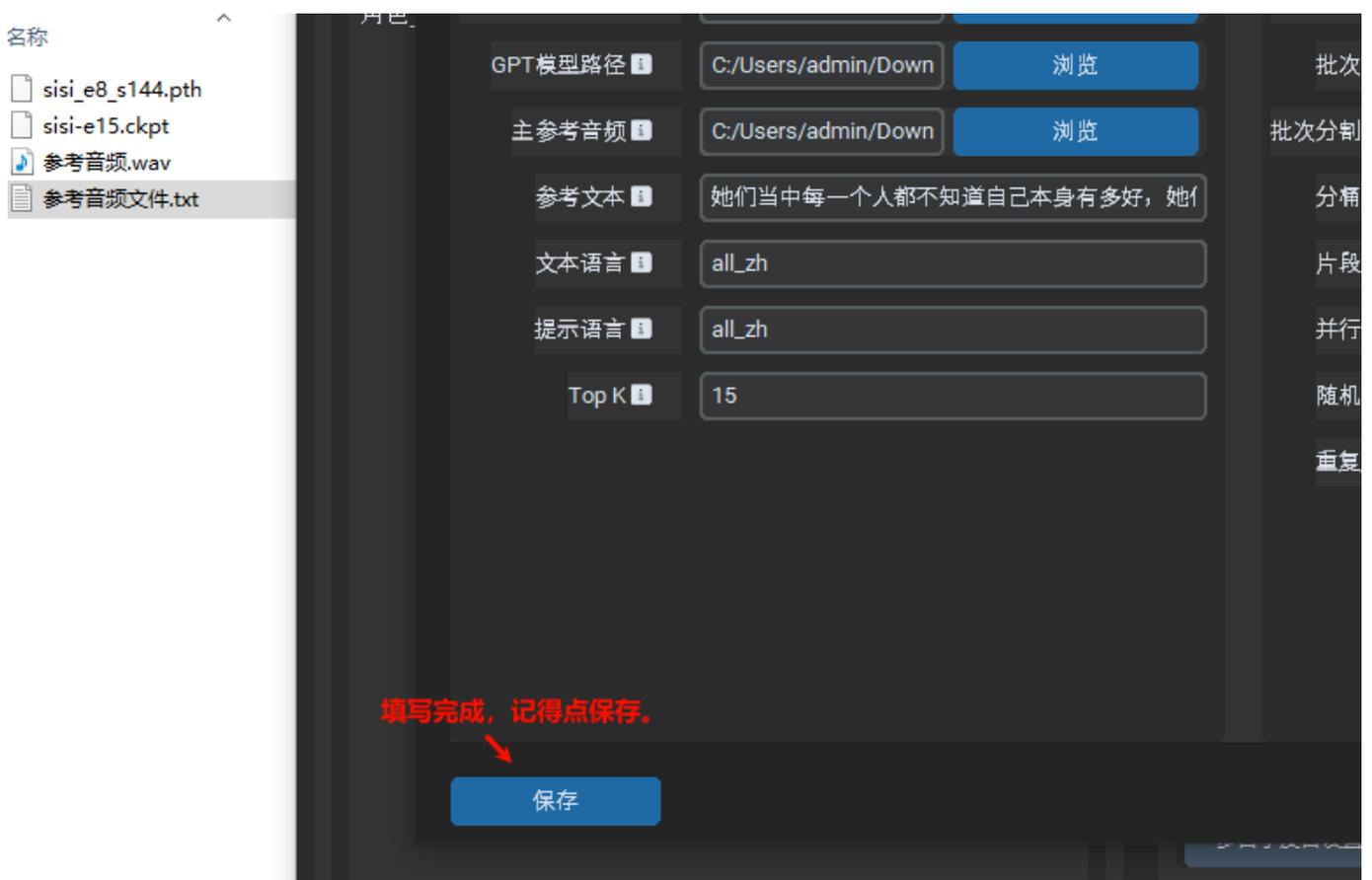
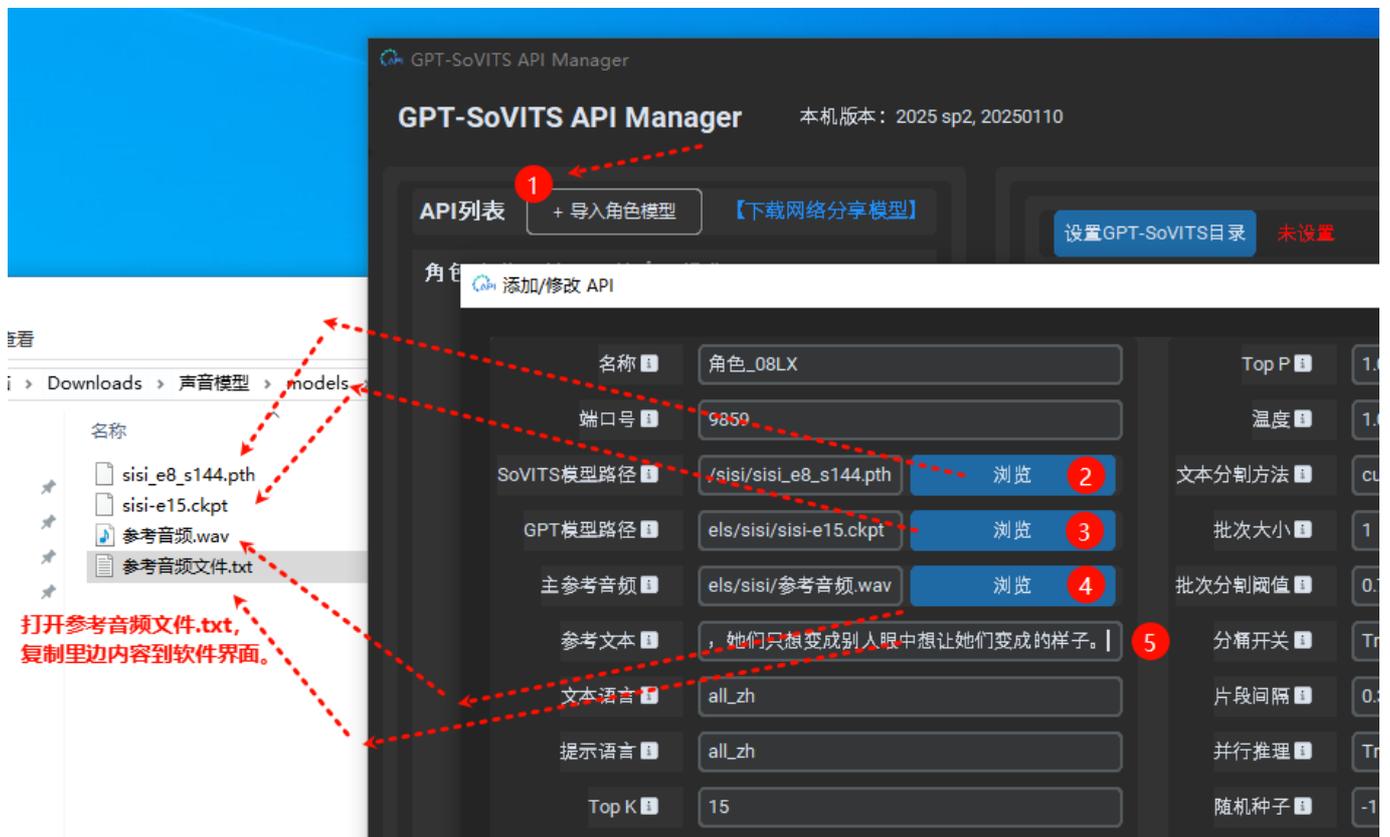
Downloads > 声音模型 > models > sisi

名称	修改日期	类型	大小
sisi_e8_s144.pth	2024/10/11 2:18	PTH 文件	83,016 KB
sisi-e15.ckpt	2024/10/11 2:56	CKPT 文件	151,673 KB
参考音频.wav	2024/10/11 10:07	波形声音	452 KB
参考音频文件.txt	2025/1/9 22:38	文本文档	1 KB

**一般模型，至少有两个模型文件+一个参考音频+参考音频的文本。
有些模型，会有多个参考音频。**

下图，点击【导入角色模型】，弹出的窗口虽然很多东西，但只要填写4项就行了。其它的不理解可先忽略。





填写完成，记得点保存。

保存

点击【保存】之后，就在API列表中出现一个刚添加的角色了。



当运行状态处于运行中, 点击【在浏览器中打开配音网址】, 得到下边的窗口, 可以开始文本合成语音了。

GPT-SoVITS Web TTS

127.0.0.1:5000

GPT-SoVITS Web TTS

单文本配音 批量TXT配音 变声(转录)

默认角色 插入1s静音 纠音设置 多角色选择

大家好，今天是2025年的1月10日的凌晨。
夜间真的很安静，放着喜欢的音乐，写着软件使用教程。

1 在这里输入要合成语音的文本

音量 100

语速 1.0

Seed -1 随机 音频片段间隔 0.3 (秒)

2 点击开始合成语音按钮，等一会就有结果了。

开始合成语音

单文本配音 批量TXT配音 变声(转录)

默认角色 插入1s静音 纠音设置 多角色选择

大家好，今天是2025年的1月10日的凌晨。
夜间真的很安静，放着喜欢的音乐，写着软件使用教程。

音量 100
语速 1.0

Seed 2021027 随机 音频片段间隔 0.3 (秒)

开始合成语音

播放合成的语音 0:00 / 0:09

这个按钮可以下载合成的语音音频文件mp3

试听配音效果：

file:///F:/班底录屏/bandicam%202025-01-10%2001-53-56-885.mp4

批量txt合成音频【GPT-SoVITS 批量API管理器】

成功启动角色模型和本机网页配置之后，点击打开配音网址按钮。

The screenshot displays the GPT-SoVITS API management interface. On the left, there is an 'API列表' (API List) table with columns for '角色名称' (Character Name), '端口' (Port), '状态' (Status), and '操作' (Actions). The table lists three instances: '奇听' (Qiteng) on port 9846 (status: 已停止), '提米' (Temi) on port 9879 (status: 运行中), and '思思' (Sisi) on port 9839 (status: 运行中). A red circle labeled '1' highlights the '运行中' status of the '提米' instance. On the right, there are configuration sections. The top section shows '设置GPT-SoVITS目录' (Set GPT-SoVITS directory) as '已设置' (Set), and '运行模式' (Running mode) with 'GPU模式' (GPU mode) selected. Below this, there is a '重启ASR服务' (Restart ASR service) button and 'ASR状态: 未运行' (ASR status: Not running). The middle section shows '本机网址' (Local URL) as 'http://127.0.0.1:5000' and '运行状态: 运行中' (Running status: Running), with a red circle labeled '2' around the status. Below this are buttons for '启动本机网页配音' (Start local web voice), '停止' (Stop), and '在浏览器中打开配音网址' (Open voice URL in browser), with a red circle labeled '3' around the last button. The bottom section shows '品牌名称' (Brand name) as 'gpts局域网配音系统' (GPTS LAN voice system) and '访问密码' (Access password) as empty. Below this are '内网URL' (Intranet URL) as 'http://192.168.247.1:5001, http://192.168.247.1:5000' and '内网网页配音服务: 停止' (Intranet web voice service: Stop). Below this are buttons for '启动内网网页配音' (Start intranet web voice), '停止' (Stop), and '在浏览器中打开配音网址' (Open voice URL in browser). At the bottom, there are buttons for '多音字发音设置' (Polyphonic pronunciation settings), '修复拼音设置不生效' (Fix pinyin settings not taking effect), and '生成tts json' (Generate tts json).

角色名称	端口	状态	操作
奇听	9846	已停止	启动 停止 测试 修改 删除
提米	9879	运行中	启动 停止 测试 修改 删除
思思	9839	运行中	启动 停止 测试 修改 删除

点击网页菜单【批量TXT配音】 - 【选择TXT文件】 -开始批量合成语音

GPT-SoVITS Web TTS

1

单文本配音

批量TXT配音

变声(转录)

默认配音角色

提米

限制每个文件最大读取10000行(推荐)

选择TXT文件

2

开始批量合成语音

3

GPT-SoVITS Web TTS

单文本配音

批量TXT配音

变声(转录)

默认配音角色

提米

限制每个文件最大读取10000行(推荐)

选择TXT文件

1.txt

等待处理

删除

2.txt

等待处理

删除

3.txt

等待处理

删除

4.txt

等待处理

删除

5.txt

等待处理

删除

总体处理进度



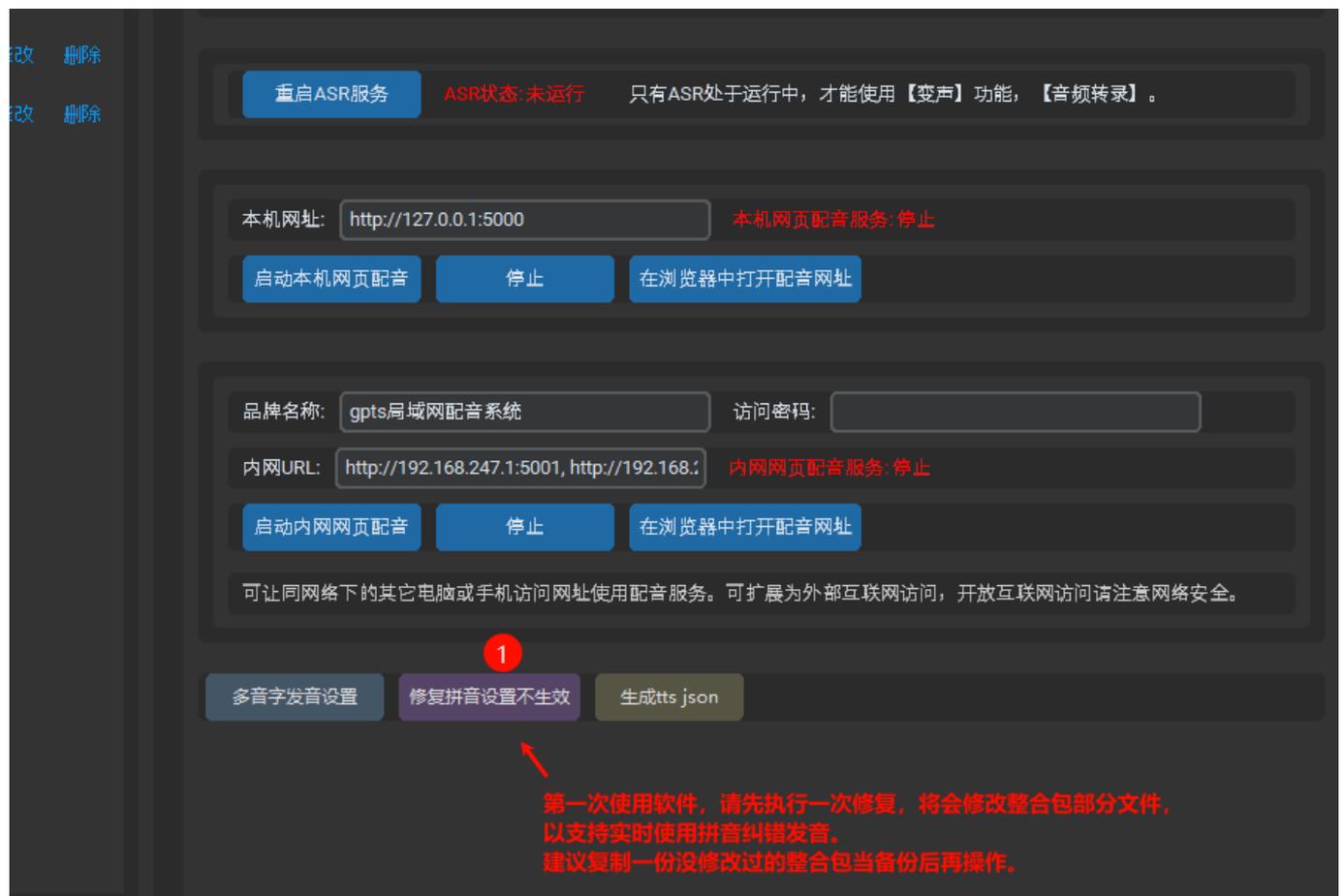
开始批量合成语音

实时拼音纠正TTS发音【GPT-SoVITS 批量API管理器】

前言

感谢花儿不哭大佬，GPT-SoVITS v2原生支持多音字。但由于是原生是配置文件类型的方式，修改之后要重启，在实际使用中不是很方便，我们对整个包少量文件进行了修改，以支持实时修改进行拼音纠正发音。

第一次使用软件，要使用实时拼音纠正发音，必须点击一次【修复拼音设置不生效】。建议你在修复前备份一次整个包。



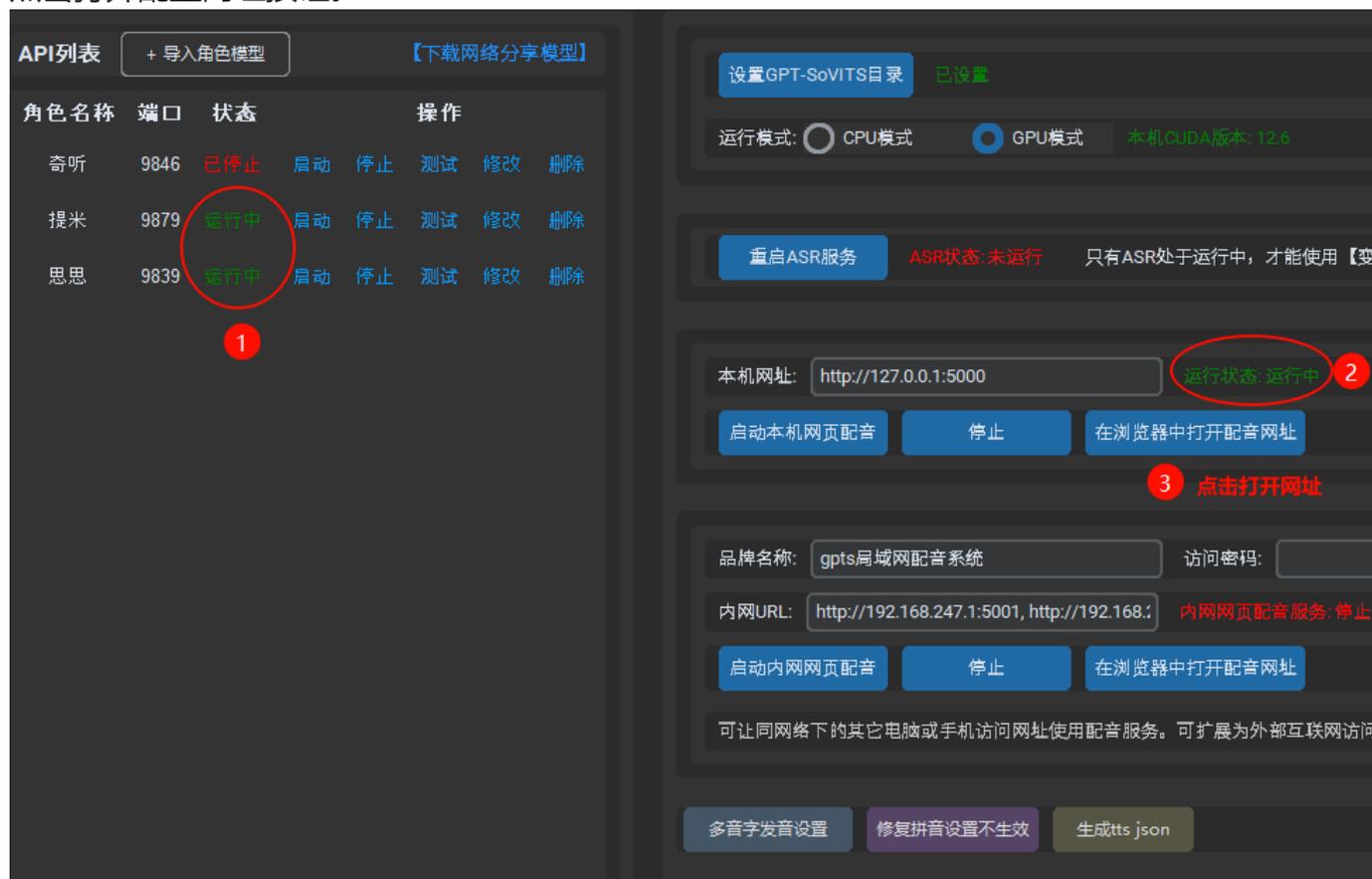
纠错TTS发音

设置拼音有两个地方：一个是软件界面的【多音字发音设置】，另一个是网页配置界面的【发音设置】。

软件界面的设置能对所有使用者生效，而网页配音界面的发音设置只对使用网页配音的用户生效，这个在多用户使用网页配音时很有用。

我们主要展示网页配音上的发音设置。

首先：启动需要的模型角色，处于运行中之后，再启动本机网页面配置，使其也处于运行中。点击打开配置网址按钮。



在打开的网页配置界面上，点击【纠音设置】

GPT-SoVITS Web TTS

单文本配音

批量TXT配音

变声(转录)

默认角色

插入1s静音

纠音设置

多角色选择

大家好，我是2024级新生，我喜欢打篮球。

音量 100

语速 1.0

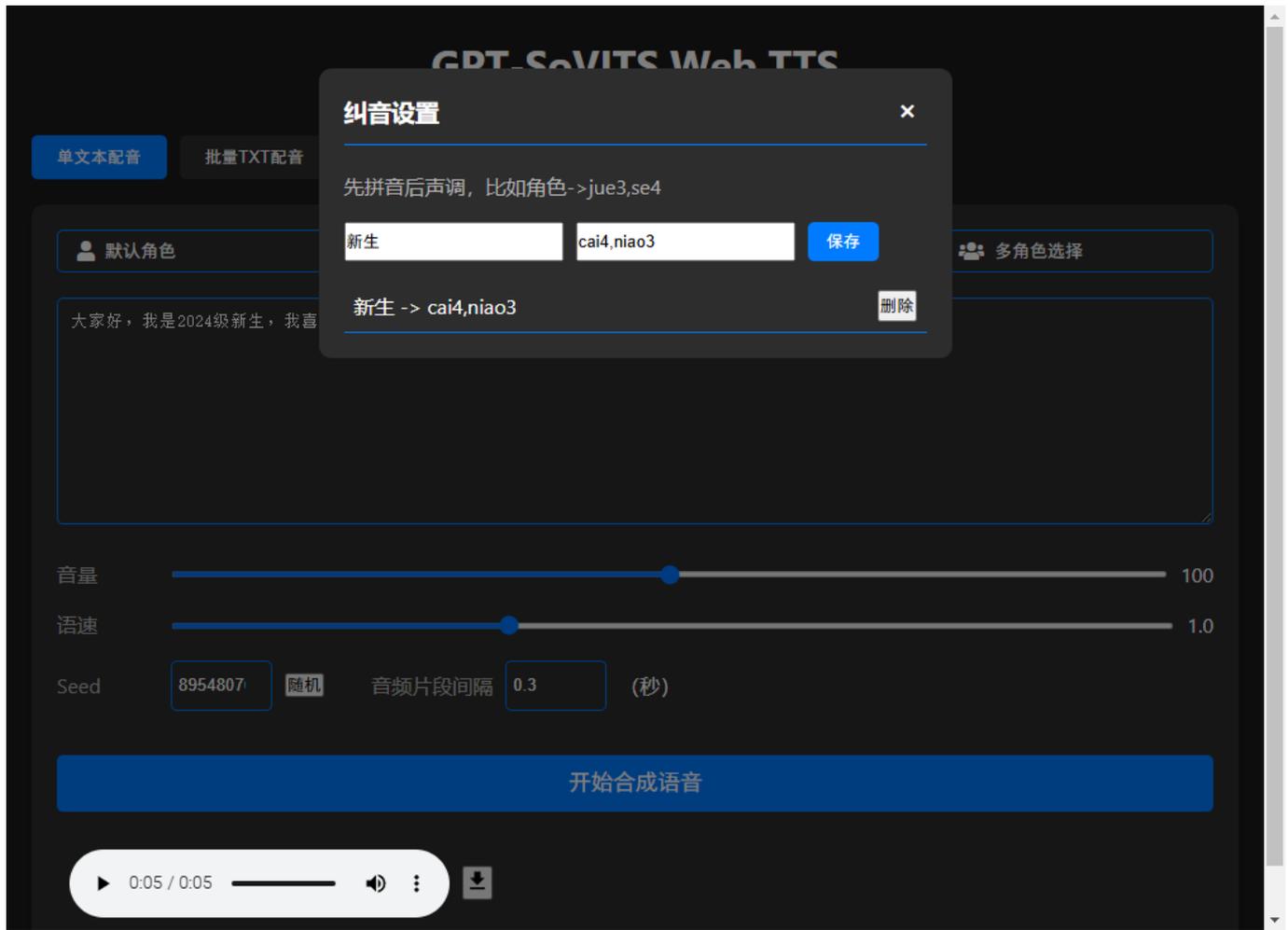
Seed 随机 音频片段间隔 (秒)

开始合成语音

▶ 0:05 / 0:05



这里的纠音是中文拼音的方式，但声调在拼音后边。



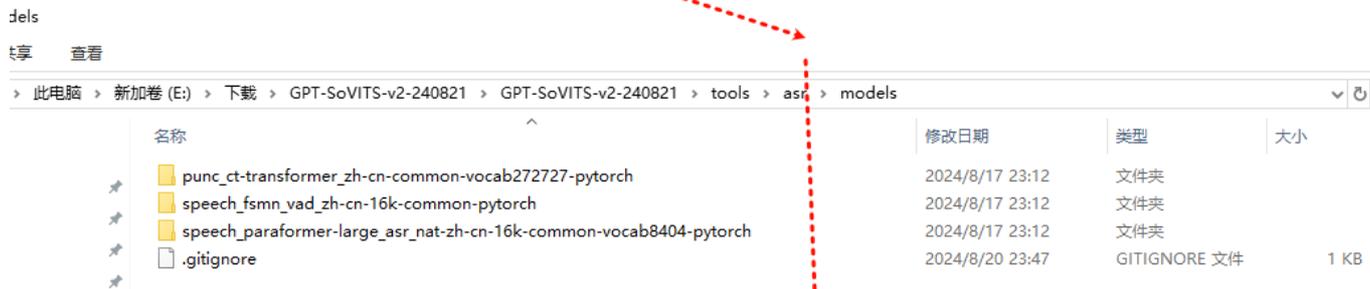
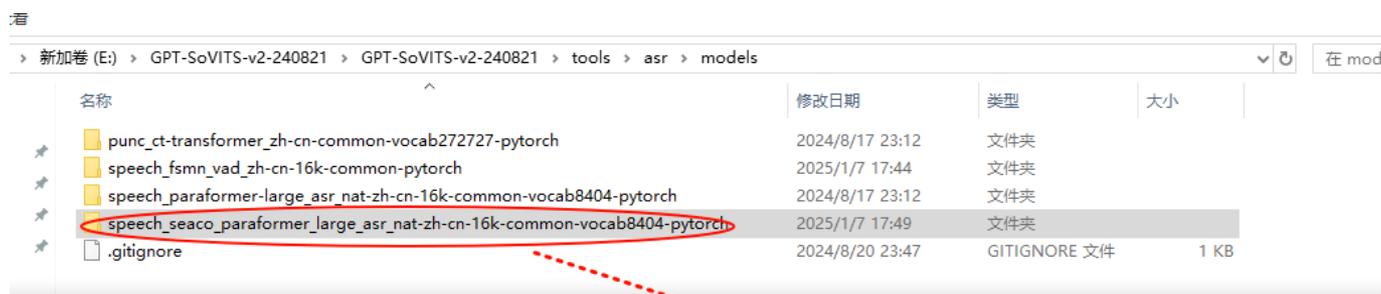
这里的纠音不能乱用，如果你硬要把没有歧义的词语修改发音，也不一定成功，但修复歧义多音字还是不成问题。

变声（音频转录）【GPT-SoVITS 批量API管理器】

模型准备

speech_seaco_paraformer_large_asr_nat-zh-cn-16k-common-vocab8404-pytorch

模型是实现使用GPT-SoVITS实现语音转录的关键，根据名称网上自行搜索下载。



要实现语音转录，整合包的模型列表中，必须增加一个模型。
复制名称，从网上下载解压缩后复制到整合包中。

软件界面启动ASR

和以往TTS配音不同。使用变声（音频转录）功能，必须在软件界面上启动ASR服务，使ASR状态也处于运行中。

接着就可以进行音频转录了，如果你使用手机版，而能实时录音畅玩变声。

GPT-SoVITS Web TTS

单文本配音

批量TXT配音

变声(转录)

变声角色

提米

选择音频文件

曾小贤.wav

已完成



0:13 / 0:13

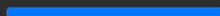


下载音频

删除

开始变声转录

总体处理进度



使用GPT-SoVITS制作一个角色声音模型

双击GPT-SoVITS整合包中的go-webui.bat打开软件

1、背景声音分离。

(1) 如果您的音频本身就很干净只有纯人声，这一步可免除

如果是自己录声音，尽量1分钟以上，加一点停顿 语气 情绪 之类。可朗读一个文章试下。

如果是网络视频音频则直接下载，声音存在背景声音或音乐的情况下需要使用这个步骤单独提取纯人声

打开软件后，一般是网址 <http://localhost:9874/>



(2) 指定要进行人声提取的原始音频，以及指定输出主人声文件夹



2、分割音频文件

不论是一个多音频还是多个短音频，都进行这一步

(1) 音频自动切分输入路径，可文件可文件夹

比如 F:\音频素材\XX说话.mp3，要换斜杠F:/音频素材/XX说话.mp3

如果有多个文件，则只需要F:/音频素材/

(2) 输出根目录一定要填一个实在的，没有自己可以创建一个，

比如F:/音频素材/分割输出

0b-语音切分工具

音频自动切分输入路径，可文件可文件夹

切分后的子音频的输出根目录

threshold:音量小于这个值视作静音的备选切割点

min_length:每段最小多长，如果第一段太短一直和后面段连起来直到超过这个值

max归一化后最大值多少

alpha_mix:混多少比例归一化后音频进来

1 声音文件路径

2 设置一个分割后音频的保存路径

3 点击开启语音切割

开启语音切割

min_interval:最短切割间隔

hop_size:怎么算音量曲线，越小精度越大计算量越高（不是精度越大效果越好）

max_sil_kept:切完后静音最多留多长

切割使用的进程数

语音切割进程输出信息

0bb-语音降噪工具

比如实际填写如下

音频自动切分输入路径，可文件可文件夹

切分后的子音频的输出根目录

threshold:音量小于这个值视作静音的备选切割点

min_length:每段最小多长，如果第一段太短一直和后面段连起来直到超过这个值

max归一化后最大值多少

alpha_mix:混多少比例归一化后音频进来

1

2

3

开启语音切割

min_interval:最短切割间隔

hop_size:怎么算音量曲线，越小精度越大计算量越高（不是精度越大效果越好）

max_sil_kept:切完后静音最多留多长

切割使用的进程数

语音切割进程输出信息

3、批量ASR，意思就是AI提取音频文字

把上一步分割后音频的保存路径，填写到第1标记处，再设置一个输出文件夹路径，之后 点击 按钮。

输入文件夹路径

输出文件夹路径

ASR 模型

ASR 模型尺寸

ASR 语言设置

1

2

3

开启离线批量ASR

processing | 0.6/20.2

4、进行数据标注，主要是纠错

把上一步批量ASR的输出路径，找到最终生成的文件，连路径与文件名一起填写到

list标注文件的路径填写框架中，点击勾选 打开WEBUI

是否开启打标WebUI

list标注文件的路径

5、数据标注

对于不需要的，勾选yes后点击delete audio按钮。当前页下的校验完成后，点SAVE FILE ,SUBMIT TEXT两个按钮，如果还有音频没在当前页，可能在下一页的，点击NEXT INDEX。同样操作。

检验的方式：主要就是如果自动识别的文本与声音表达的不同，人工修改文内容。

Change Index Submit Text Merge Audio Delete Audio Previous Index

Next Index

Index 0

Audio Split Point(s) 0

Split Audio Save File Invert Selection

Text 0 Yes

Text 1 Yes

Text 2 Yes

Text 3 Yes

6、进行模型预训练数据格式化

0-前置数据集获取工具 1-GPT-SoVITS-TTS 2-GPT-SoVITS-变声

*实验/模型名: zawei 1

显卡信息: 0 NVIDIA GeForce RTX 3080

预训练的SoVITS-G模型路径: GPT_SoVITS/pretrained_models/s2G488k...

预训练的SoVITS-D模型路径: GPT_SoVITS/pretrained_models/s2D488k...

预训练的GPT模型路径: GPT_SoVITS/pretrained_models/s1bert25...

1A-训练集格式化工具 1B-微调训练 1C-推理

输出logs/实验名目录下应有23456开头的文件和文件夹

*文本标注文件: F:/FFOutput/zawei/cut_asr/cut.list 2

*训练集音频文件目录: F:/FFOutput/zawei/cut 3

1Aa-文本内容

GPU卡号以分割, 每个卡号一个进程: 0-0

预训练的中文BERT模型路径: GPT_SoVITS/pretrained_model.../chinese_sbert_base

4 开启文本获取

文本进程输出信息: 文本进程成功

1Ab-SSL自监督特征提取

GPU卡号以分割, 每个卡号一个进程: 0-0

预训练的SSL模型路径: GPT_SoVITS/pretrained_model.../ssl

5 开启SSL提取

SSL进程输出信息: SSL提取进程结束

1Ac-语义token提取

GPU卡号以分割, 每个卡号一个进程: 0-0

6 开启语义token提取

语义token提取进程输出信息: 语义token提取进程结束

1Aabc-训练集格式化—键三连

7 开启—键三连

一键三连进程输出信息: 一键三连进程结束

7、进行微调训练，这里batch_size根据显存大小来调。个人12G的3080就默认6差不多。

0-前置数据集获取工具 1-GPT-SoVITS-TTS 2-GPT-SoVITS-变声

*实验/模型名: zawei

显卡信息: 0 NVIDIA GeForce RTX 3080

预训练的SoVITS-G模型路径: GPT_SoVITS/pretrained_models/s2G488k...

预训练的SoVITS-D模型路径: GPT_SoVITS/pretrained_models/s2D488k...

预训练的GPT模型路径: GPT_SoVITS/pretrained_models/s1bert25...

1A-训练集格式化工具 1B-微调训练 1C-推理

1Ba-SoVITS训练。用于分享的模型文件输出在SoVITS_weights下。

每张显卡的batch_size: 6

总训练轮数total_epoch, 不建议太高: 8

文本模块学习率权重: 0.4

保存频率save_every_epoch: 4

是否仅保存最新的ckpt文件以节省硬盘空间

是否在每次保存时间点将最终小模型保存至weights文件夹

GPU卡号以分割, 每个卡号一个进程: 0

1 开启SoVITS训练

SoVITS训练进程输出信息: SoVITS训练完成

1Bb-GPT训练。用于分享的模型文件输出在GPT_weights下。

每张显卡的batch_size: 6

总训练轮数total_epoch: 15

是否开启dpo训练选项(实验性)

是否仅保存最新的ckpt文件以节省硬盘空间

是否在每次保存时间点将最终小模型保存至weights文件夹

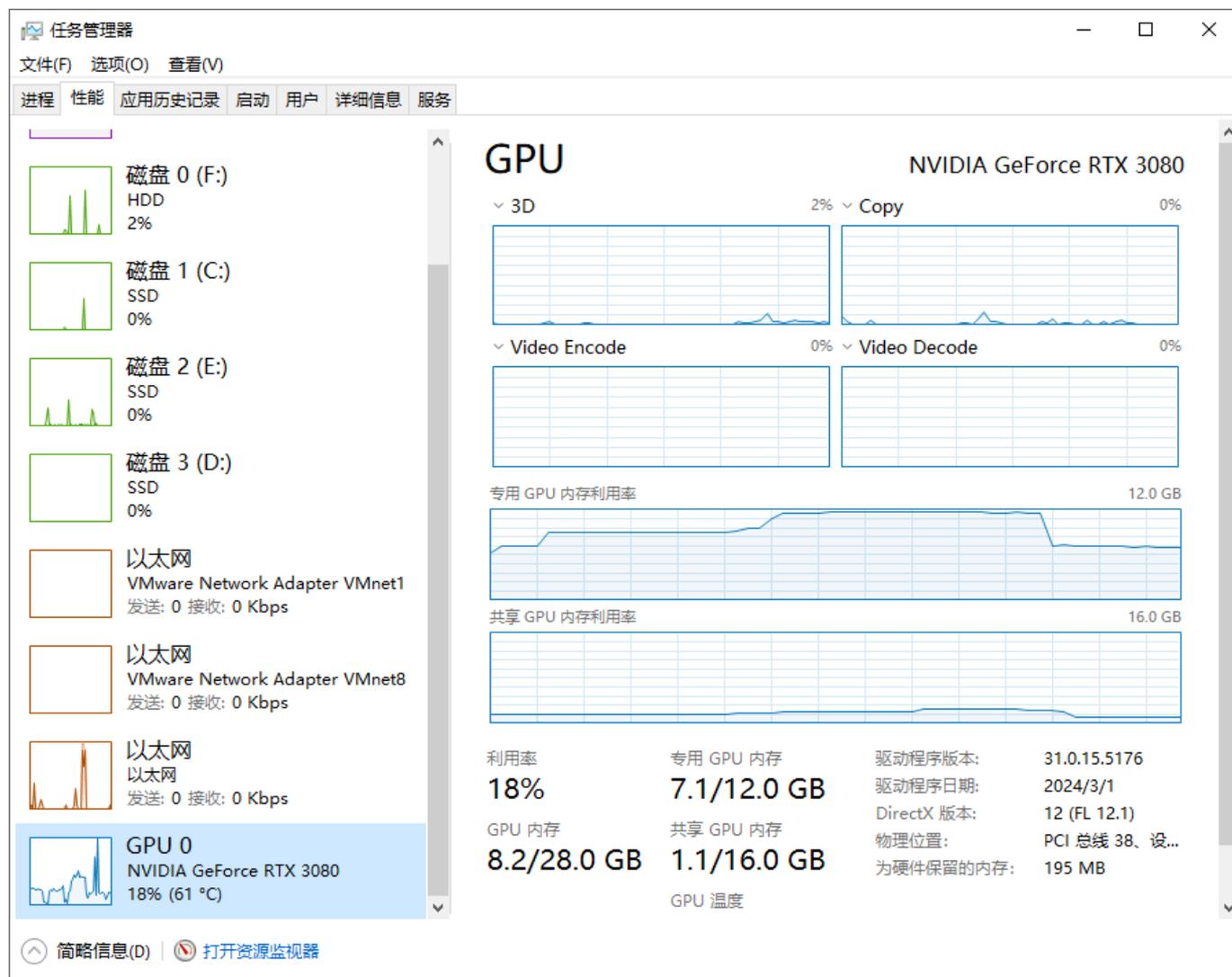
保存频率save_every_epoch: 5

GPU卡号以分割, 每个卡号一个进程: 0

2 开启GPT训练

GPT训练进程输出信息: GPT训练完成

下图可看到，两个训练同时进行，显存几乎跑满12G



8、完成，现在可生成音频了。

- (1、2) 选刚才的模型
- (3) 上传之前用过的随便一个声音
- (4) 填写那个声音的文本内容
- (5) 填写要生成语音的文本
- (6) 点击按钮
- (7) 试听

模型切换

GPT模型列表 **1**
GPT_weights/ziwei-e15.ckpt

SoVITS模型列表 **2**
SoVITS_weights/ziwei_e8_s136.pth

刷新模型路径

*请上传并填写参考信息

请上传3-10秒内参考音频，超过会报错! **3**

0:03 / 0:03

参考音频主要是情感参考

开启无参考文本模式。不填参考文本亦相当于开启。

参考音频的语种
中文

使用无参考文本模式时建议使用微调的GPT，听不清参考音频说的啥(不晓得写啥)可以开，开启后无视填写的参考文本。

参考音频的文本 **4**
你认为...

*请填写需要合成的目标文本和语种模式

需要合成的文本
因为科技的尽头就是金融，所以小罗不直接跳跃? **5**

需要合成的语种
中文

gpt采样参数(无参考文本时不要太低):

top_k 5

top_p 1

temperature 1

怎么切
 不切 凑四句一切
 凑50字一切 按中文句号。切
 按英文句号。切 按标点符号切

合成语音 **6**

输出的音 **7**

0:04

文本切分工具。太长的文本合成出来效果不一定好，所以太长建议先切。合成会根据文本的换行分开合成再拼起来。

9、提醒

请遵守开源软件GPT-SoVITS的使用要求，同时注意：未获得别人同意情况下不要使用别人的声音，更不要传播以别人声音制作的模型。

可以录自己的声音或找一些无版权的声音来练习！

下一章将介绍 API管理器与创建tts配音文件