ESP32 AI 音箱 · 课前准备手册

90 分钟线上实战课 —— 用 AI 编程工具从零做出一台会说话、能对话的 USB 音箱

⏱ 90 分钟 | 🔧 硬件实操 | AI 辅助编程 | 云端语音服务

0 这门课做什么

先看清楚终点，再准备行装。90 分钟内，你将从一块裸板和一个云账号出发，做出下面 3 个效果：

效果	说明	用到的能力
🎧 ① 播放音乐	ESP32 直接播放放在板子里的音频文件	MicroPython · I2S · 功放
🗣 ② 说任意文字	输入一句话，喇叭用云端合成的语音读出来	以上 + TTS（文字转语音）+ 后端代理
💬 ③ AI 对话	对网页说话 → AI 回答 → 从喇叭里播出	以上 + 语音识别 + 大模型

1 硬件清单

总共 6 样，电商搜关键词即可。

物品	推荐型号 / 规格	数量	说明
主控板	**ESP32-S3 N16R8** 开发板	1	引脚以本手册接线表为准
功放模块	**MAX98357A** I2S 功放模块	1	7 针脚的那种
喇叭	**4Ω / 3W** 小喇叭（8Ω 也可，功率略小）	1	带焊点或端子的

杜邦线	**母对母**（两头插孔，连 ESP32 ↔ 功放）+ **公对公**（两头插针，连功放 ↔ 喇叭），10cm	母母 ≥5 根、公公 ≥2 根	多备几根防坏；没公对公也可：喇叭线剥线露铜丝直接压进绿色端子（见 3.4 接线表 ⊕/⊖ 行）
一字螺丝刀	小号，刀头宽 2~3mm	1	拧松 / 拧紧绿色螺钉式接线端子 —— **有无公对公杜邦线都要用**
USB 数据线	与你的板子匹配（多为 USB-C）	1	**必须是数据线**，纯充电线不行

2 云平台账号（阿里云百炼）

我们的音箱要「听、说、想」，这些能力由云端大模型提供 —— 课堂统一使用**阿里云百炼**：语音能力一站齐备（合成 / 识别 / 对话），免费额度充足，**国内手机号即可注册，免绑卡、免国际网络。**

2.1 开通与配置（约 5 分钟）

1. **注册 / 登录阿里云**：打开 `bailian.console.aliyun.com`，用手机号注册或登录（免费额度**无需实名认证**即可领取使用；想按量付费才需要认证）。确认右上角地域是**华北2（北京）**—— 免费额度只发给北京地域的模型，课上的 API 域名也是北京区的。
2. **开通模型服务**：首次进入会弹出服务协议，同意即可开通；**没弹窗 = 之前已开通过**，正常。额度开通后**自动发放，不用手动领取**（最长 2 小时生效）。规则：每个模型**独立**赠送额度 —— LLM 通常是 100 万 tokens、有效期 90 天；**语音模型按字符/秒计费**，新模型（如 CosyVoice v3、Qwen3-ASR）送 90 天一次性额度，老模型（如 cosyvoice-v1、paraformer 系列）是每月重置的长期额度（见下表），并不是「每个模型都 100 万 tokens」。
3. **拿到业务空间 ID**：**不用创建** —— 百炼已默认建好一个**默认业务空间**，直接用它。入口：控制台 → 模型服务 → **业务空间管理**，复制「默认业务空间」的 ID（形如 `llm-xxxxxxxx`）—— 2026 年起语音接口走业务空间专属域名，**必填。**默认业务空间的**模型权限自动全部开通**（含语音识别类），无需单独设置。

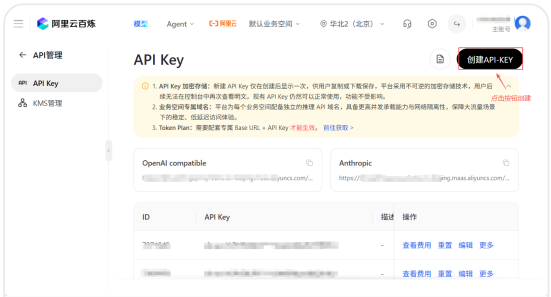


复制「默认业务空间」的 ID（llm- 开头）

4 **创建 API Key**：左下角 **API-KEY** 按钮 → **创建 API-KEY** → 弹窗默认选中「默认业务空间」，直接点「确定」（生成的 Key 形如 `sk-xxxxxxxx`）。



① 左下角「API-KEY」按钮入口



② 点击「创建 API-KEY」



③ 创建弹窗：默认选中「默认业务空间」，直接点「确定」



④ 点击复制，存进记事本

5 **查看自己的免费额度**：直接打开 `bailian.console.aliyun.com/cn-beijing/costing-balance/free-quota`（控制台左侧菜单 → **免费额度** 页也能进）—— 每个模型的剩余量、到期时间、「用完即停」开关都在这里；也可以在模型广场点开模型详情页看单个模型的额度（几分钟刷新一次）。



① 免费额度页：各模型剩余量与到期时间



② 免费额度页（续）：「用完即停」开关

6 **建议开启「用完即停」**：在「免费额度」页把用完后的行为设为停用，杜绝意外扣费。

7 **保存好两样东西**（课上要填进配置文件）：

```
DASHSCOPE_API_KEY      = sk-xxxxxxxx          （API Key）
DASHSCOPE_WORKSPACE_ID = llm-xxxxxxxx         （业务空间 ID）
```

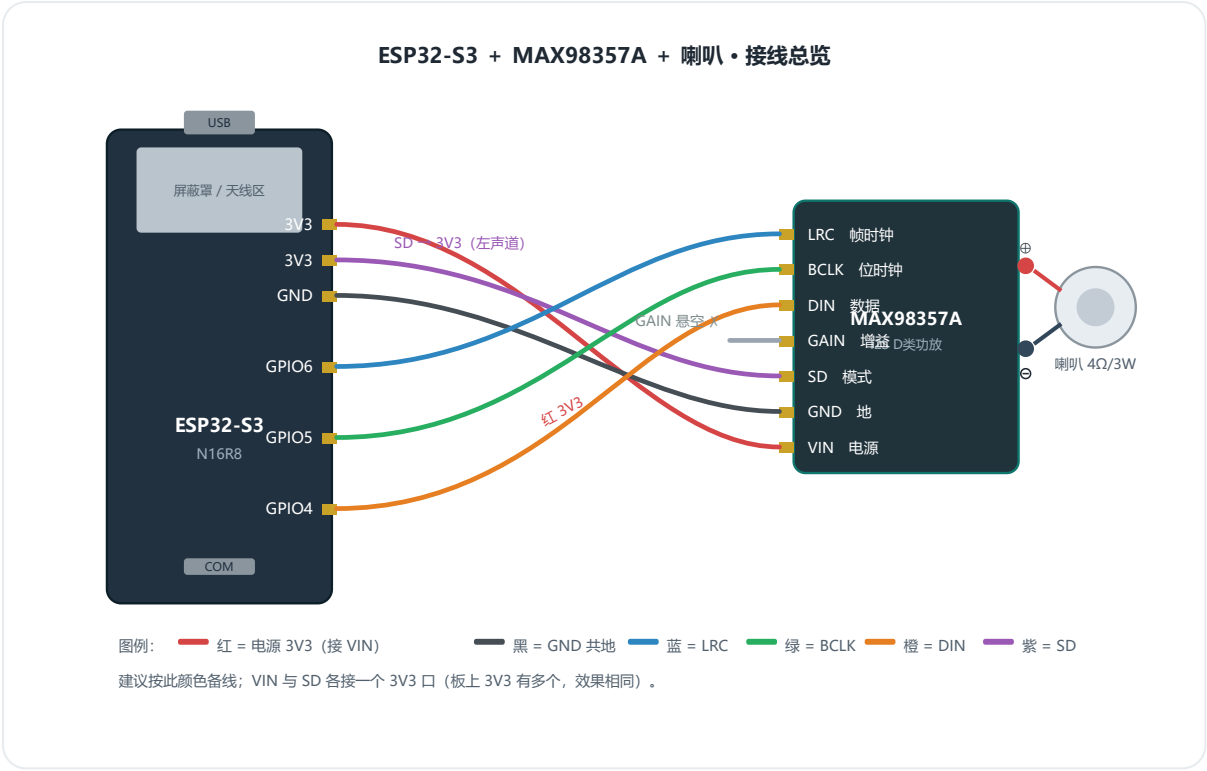
本课程用到的模型清单（对照着在「免费额度」页核对一遍）

能力	模型	免费额度
🗣 **说话** 语音合成	`cosyvoice-v3-flash`	1 万字符（新用户免费额度，90 天内有效）
👂 **听** 语音识别·文件转写	`paraformer-v2`	3.6 万/月，每月 1 日重置，长期有效
👂 **听** 语音识别·实时流式	`paraformer-realtime-v2`	3.6 万/月，每月 1 日重置，长期有效
💡 **想** AI 对话	`qwen3.8-flash`	100 万 tokens（90 天内）

3 接线（课前完成，课上复查）

ESP32-S3 通过 6 根线驱动 MAX98357A，功放直接推喇叭。**请在课前接好**，课上只做复查与通电自检。 接线时保持断电，接完检查再上电。

3.1 接线图



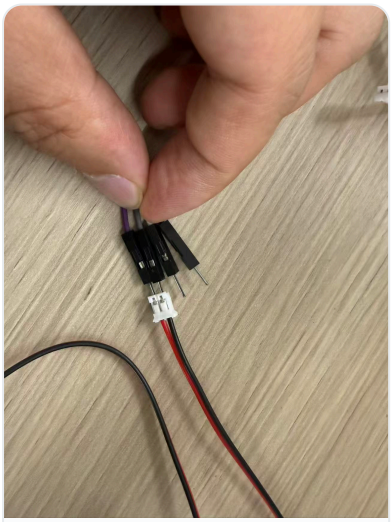
SVG 示意图（矢量可缩放）。引脚名称以你的实物模块丝印为准；实拍照片见 3.2。

3.2 实拍接线图（对照你的实物）

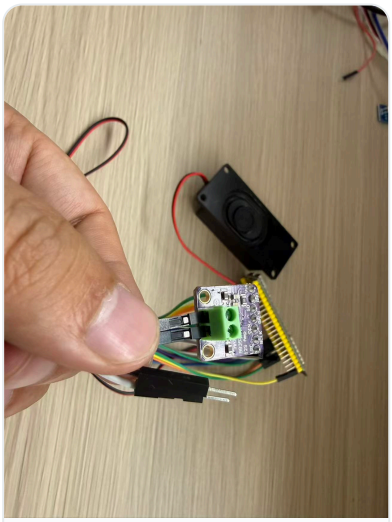
本课套件的实拍（ESP32-S3 N16R8 + MAX98357A + 4Ω/3W 喇叭）。先看接好的整机，再逐张认清两端的接口， 然后照着 3.4 接线表逐根插线。



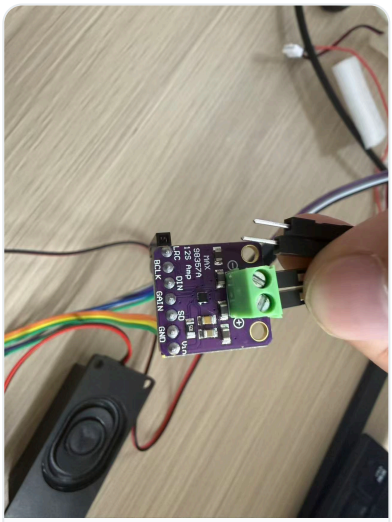
① 接好的整机：ESP32-S3（左下）→ 6 根杜邦线 → MAX98357A 功放（中）→ 红黑喇叭线 → 4Ω 喇叭（右）



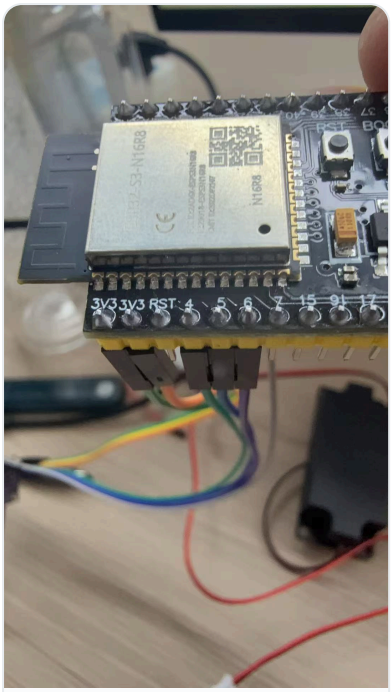
② 认识端子：黑色母头是杜邦线（插排针用）；白色端子是喇叭自带线（红黑两芯）



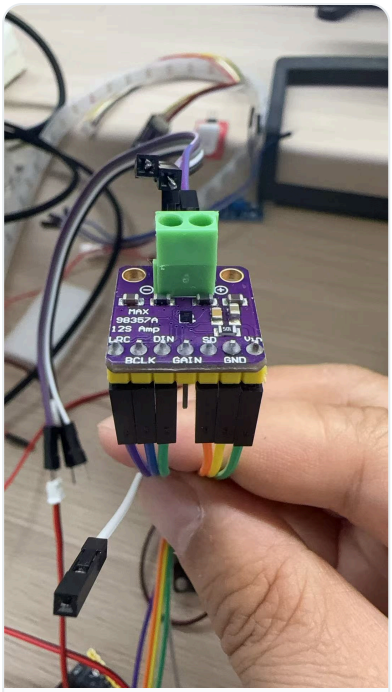
③ 喇叭线接这里：功放的绿色**螺钉式接线端子**（⊕ ⊖ 输出）——用一字螺丝刀拧松螺丝，喇叭线剥掉外皮露铜丝再插入，拧紧压住



④ 功放认针：丝印依次为 LRC / BCLK / DIN / GAIN / SD / GND / VIN，另一端 ⊕ ⊖ 是喇叭输出



⑤ ESP32 认针：从丝印左起 3V3 3V3 RST 4 5 6 7 ...；本课用两个 3V3 与 GPIO4 / 5 / 6



⑥ 功放正面全貌：一整排信号针脚 + 绿色输出端子（⊕ ⊖），杜邦线已插在针脚上

3.3 文字线稿图（打印友好）

```
ESP32-S3 N16R8                    MAX98357A 模块（针脚丝印顺序固定）
┌──────────────────┐
│              GPIO6├──────────┬──────────► LRC    帧时钟（WS）
│              GPIO5├──────────┼──────────► BCLK   位时钟
│              GPIO4├──────────┼──────────► DIN    I2S 数据
│                   │          │           GAIN   悬空（默认 +9dB）
│                3V3├──────────┼──────────► SD     接 3V3（= 左声道）
│                GND├──────────┼──────────► GND    共地
│                3V3├──────────┴──────────► VIN    电源（2.5~5.5V）
│                   │          │           ⊕ ┐
│                   │          │              ├◄ 喇叭 4Ω/3W
└──────────────────┘          │           ⊖ ┘
（BTL 输出，喇叭无正负极之分）
```

3.4 接线表（逐行核对）

MAX98357A	接 ESP32-S3	颜色建议	作用

LRC	**GPIO6**	蓝	帧时钟 / 字选择（Word Select），频率 = 采样率，指示当前是左/右声道的数据
BCLK	**GPIO5**	绿	位时钟，每一个脉冲传 1 个比特，典型 64× 采样率
DIN	**GPIO4**	橙	串行音频数据（高位在前）
GAIN	**悬空**	—	增益默认 +9dB（其他档位见课上讲解）
SD	**3V3**（另一口，与 VIN 分开）	紫	模式选择：>1.4V = 输出左声道（详见 3.5）
GND	**GND**	黑	共地（所有信号以地为参考，不接则全是噪声）
VIN	**3V3**	红	供电。模块支持 2.5~5.5V；接 5V 功率更大，本课 3V3 够用
⊕ / ⊖	喇叭两脚	任意	模块上的**绿色螺钉式接线端子**：一字螺丝刀拧松螺丝，插入喇叭线（无公对公杜邦线就剥线露铜丝直接插）再拧紧；BTL 桥接输出，**不分正负**，也**绝不接地**

3.5 SD 引脚进阶

SD 电压	模式	怎么接
< 0.16V	**关断静音**（不是右声道！）	接 GND
0.16 ~ 0.77V	(L+R)/2 **左右混合**	10kΩ(上) + 2kΩ(下) 分压 ≈ 0.55V
0.77 ~ 1.4V	只放右声道	分压（如 10kΩ+3.3kΩ ≈ 0.82V）
> 1.4V	只放左声道	**直接接 3V3（本课接法）**

3.6 接线自查表

#	检查项	通过标准
1	6 根线两端都插紧	轻拉不脱落；杜邦线本身没断（可用万用表蜂鸣档量通断）
2	电源极性	VIN→3V3、GND→GND，**没有接反**（接反是唯一会立刻损坏的错法）

3	三根信号线一一对应	LRC→6、BCLK→5、DIN→4，没有交叉错位
4	喇叭	只接 ⊕⊖ 两个端子，没碰其他任何针脚
5	GAIN	悬空（没插任何线）

4 课前自检清单

全部打勾即完成课前准备。勾选状态会自动保存在本页（同一浏览器）。 （已完成 0/8）

硬件

- [ ] ESP32-S3 开发板到手
- [ ] MAX98357A 模块 + 4Ω/3W 喇叭 + 母对母 ≥5 根杜邦线 + 一字螺丝刀到手；公对公 ≥2 根或喇叭线剥线接端子二选一
- [ ] 按接线图接好 6 根线 + 喇叭，并通过 3.6 自查表 5 项

云账号

- [ ] 百炼已开通，拿到 **API Key**（sk- 开头）
- [ ] 已拿到**默认业务空间**的 ID（llm- 开头，不用自己创建）
- [ ] 百炼默认业务空间可用（模型权限自动开通，无需单独设置）

课上环境

- [ ] 新建一个空文件夹（如 `esp32-speaker`），课上从这里从零开始

- [ ] 把百炼 Key 和业务空间 ID 放进一个记事本，课上要用