

唯爱AI音乐学苑 第三讲

中文声线与华乐音色

(再谈 RSI)

讲者:	Michael Huo
时间:	2026-10-10 周六 17:30-18:30 PDT
美东: 20:30-21:30 EDT 北京: 10月11日 周日 08:30-09:30	
形式:	Zoom 公益免费 (60分钟, 含5分钟休息)

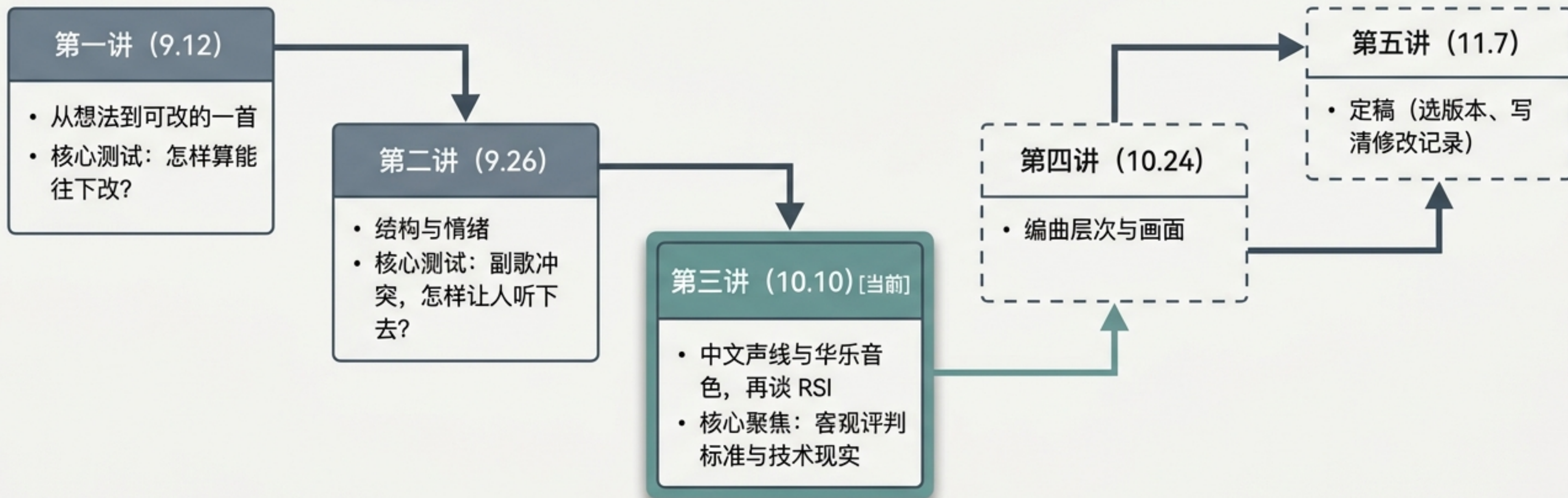
主办: 北卡浙大校友会、北美高校联盟AI创业公益俱乐部
承办: 唯爱AI基金会、唯爱AI公益基金会

接入凭证面板

Zoom会议号:
952 9981 2540

密码:
18970521

资料链接:
music.weiai.ai





上半场：音乐与声线

只评“听得见 / 听不见”
与“稳不稳”

剥离主观口味，不空谈审美。
寻找物理发声的客观“信号”。

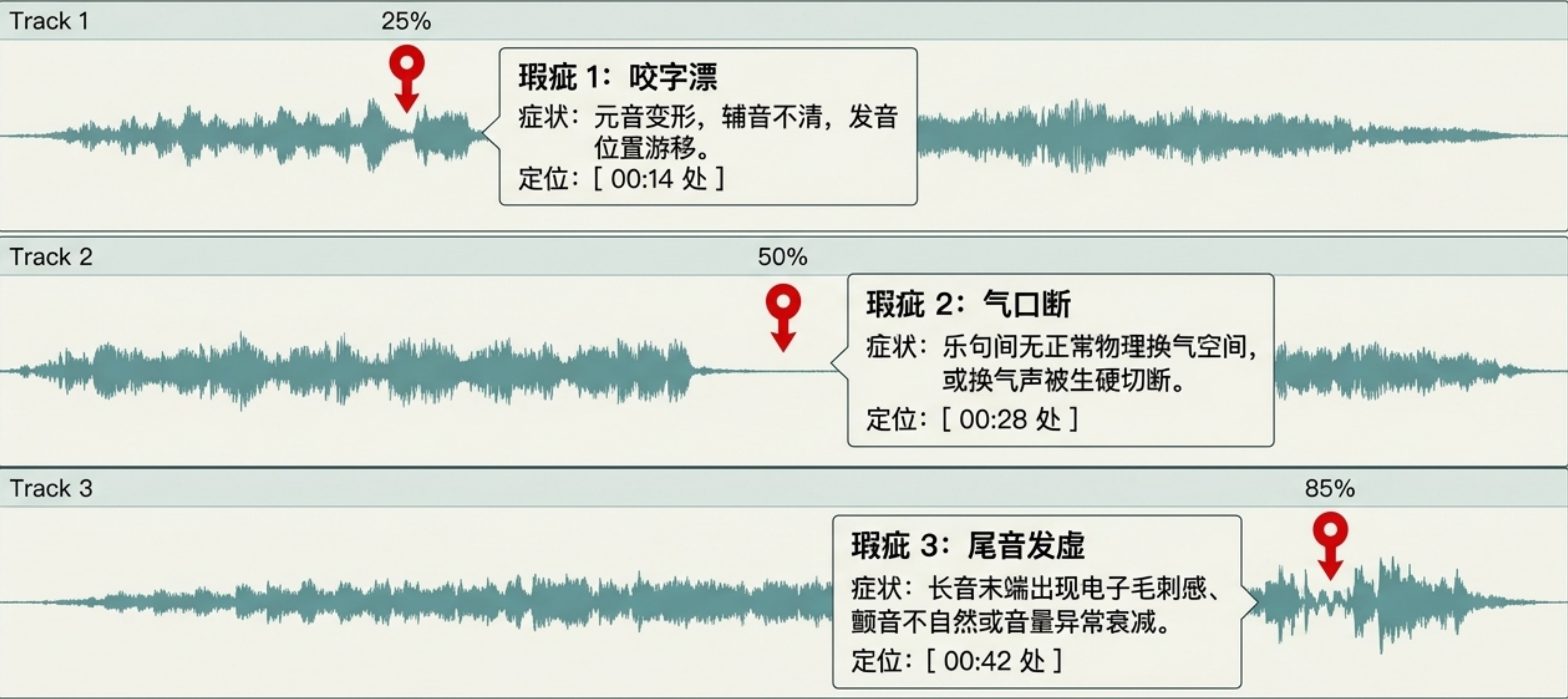


下半场：再谈 RSI

只看“谁改了模型权重”，
不盲从“厂商自述”

穿透科技圈泡沫的“噪声”，
寻找真正突破的客观“信号”。

AI声线的“塑料感”体检报告



声线对照矩阵

测试基准语句：[指定中文测试歌词文本]

检测维度	声线版本 A	声线版本 B
音色物理质感	偏沉 / 颗粒感重	偏亮 / 清透
咬字稳定性	[●] 稳固 (参考 00:15)	[✕] 游移 (参考 00:15)
气口连贯性	[✕] 断裂 (参考 00:28)	[●] 连贯 (参考 00:28)

核心结论：这还是同一首歌吗？

否。情感基调与物理表现已完全偏移。

华乐音色检视面板

验真标准：二元检出

二胡

提示词：已写入



听得见

琵琶

提示词：已写入



听得见

古筝

提示词：已写入



听不见

笛

提示词：已写入



听不见

笙

提示词：已写入



听得见

诊断总结：风格词写了「华乐」，不等于AI合成了该乐器的物理特征。
耳朵的物理接收是唯一验证标准。

编曲意图对撞表

Intent

我们写了什么？ (Input)

- 明确的中文咬字要求
- 指定的华乐配器标签
- 特定秒数的情绪爆发设定



Execution

系统交付了什么？ (Output)

- 核心人声稳定度：[**✗**] **FAIL**
- 华乐配器分离度：[**✗**] **FAIL**
- 情绪动态符合度：[**✓**] **PASS**

核心洞察：提示词只是「愿望清单」，不是「控制代码」。

同一句词，换声线就等同于换歌。

剥离对主观口味的争执，听觉客观性才是AI音乐落地的第一基石。

中场休息 / INTERMISSION

05:00

系统即将切换至数据研判模式：再谈 RSI (Recursive Self-Improvement)

RSI 的严谨定义界线

边界内改进（已实现）



修改输出 / 纠错



优化提示词



改变搜索策略

跨代际 RSI 尚未发生

跨代际跃升（未发生）

完整、无约束地自主设计与训练下一代模型架构与权重。

RSI 基准：系统用自己的输出改进下一步（含提示词、工具、代码或评测）

2026 前沿厂商自述全景

现实检验：厂商 PR 滤镜

Reality Gauge (Anthropic 2026.08)

协作及以上代码生成（人类主导）

→ >90%

AI 主导研发工作
（特定框架内）

→ 约 26%

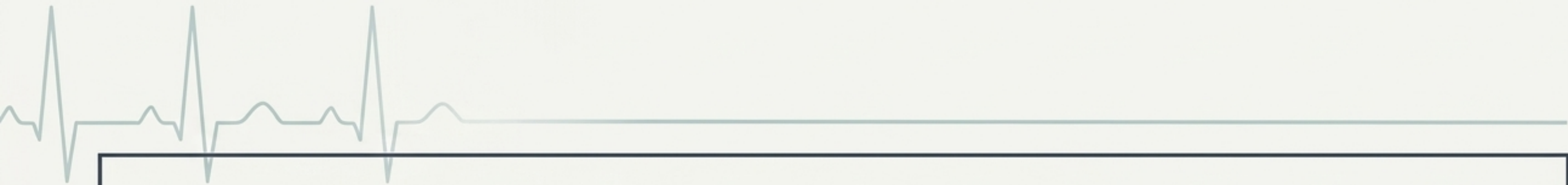
完全自主研发后继模型 → 0%

OpenAI 声明节点 (2026.09.06)

- 当前宣称：
自动化研究实习生
- 期货目标 (2028.03)：
自动化 AI 研究员

法医备注：人类仍在
『定方向』。

大佬的冷却剂



We Must Pace the Frontier.

— Dario Amodei, 2026-09-12

核心动作

主动放慢前沿能力的抬升节奏

核心目的

给『对齐 (Alignment)』和『核验 (Verification)』留出充足时间

结论：前沿玩家主动踩下刹车，而非加速走向失控的完全 RSI。

9月模型与Agent进化图谱

Autopsy Matrix				
文献	提交日	核心机制	修改提示词/策略?	修改底层模型权重?
Roadmap	9.10	纯路线图与文献综述	N/A	N/A（非演示）
Dream-RSI	9.14	用历史搜索记录当模拟器	[√] 是	[X] 否
AIDE^2	9.22	自主运行8天，7次改进	[√] 是	[X] 否
RRSI	9.22	用正则压制过拟合	[√] 是	[X] 否

具身智能的 RSI 尝试

EmbodiedSmith (10月6日)

场景：纯仿真环境

机制：场景与任务互相修改，用于扩充训练数据。

现实检验：没有真机成功率数据。

EmbodiRSI (9月)

场景：桌面任务

机制：仿真成功率升至83.5%；真机少量修正至83.1%。

现实检验：论文自述。改的仅是数据和修正轨迹。

核心研判： 机器人并没有在物理世界中修改自己的模型结构。

RSI 终极鉴定矩阵



- ☑ - 当前所谓的 RSI，改的都是数据和策略。
- ☑ - 最高维度的目标与任务，依然是人定的。
- ☑ - 最终生效的模型权重，依然是人批的 (Human-approved)。

2026年还剩两个月。
抛开论文自述与厂商泡沫，
明年此时，我们再验真伪。