M5 · 职业版 AI 视频 · 第 5 讲

第 5 讲：数字人与配音

3 分钟口播，不露脸也能交付

阅读约 25 分钟

练习约 30 分钟

产出：30 秒口播片段

本节目录

客户会怎么提

本讲要解决的问题

本讲目标

逐步操作

- 第一步：先问清四个约束，再选技术路线
- 第二步：按七步流程推进，顺序不能反
- 第三步：打磨自然度，抓五个细节
- 第四步：交付前按清单验收并锁定报价范围

常见失败与修正

一个完整示例

五、30 秒样片的做法

六、职业版验收清单

七、报价提示

八、延伸阅读

客户会怎么提

> “我们要做 3 分钟产品口播，讲师不方便出镜。能不能用数字人？要看起来自然，周五前给第一版。”

这类需求背后，客户真正担心的通常不是“像不像人”这么抽象，而是三件具体的事：**能不能按时交付、口型和声音对不对、发出去会不会显得假**。职业交付要把这三个担心逐一拆解，用流程和验收标准来兜底。

本讲要解决的问题

数字人口播最容易翻车的地方，不在“生成”这一步，而在前后衔接：脚本不口语化导致念起来别扭、先做画面后配音导致口型对不上、一次生成整段导致返工成本爆炸、验收标准模糊导致客户反复说“不自然”。

本讲要解决四个问题：

1. **路线选择**：面对不同时长、周期、预算和信任要求，选平台数字人、开源自部署还是真人+AI 配音；
2. **流程顺序**：用可重复的步骤把脚本、配音、形象、驱动和后期串起来；
3. **质量控制**：识别口型、手势、眼神等不自然问题，并知道怎么修；
4. **交付边界**：把验收项、报价范围和授权责任说清楚，避免后期扯皮。

本讲目标

- 在三条技术路线里按时长、周期、预算和信任度做选择；
- 用七步流程做出 30 秒可用口播样片，并能扩展到 3 分钟；
- 处理口型、手势、眼神、表情与背景五类常见不自然；
- 按职业标准验收并给出可执行的报价拆项。

逐步操作

第一步：先问清四个约束，再选技术路线

不要一上来就推荐某个平台。先把约束问清楚，再匹配路线。

约束	要问清什么	对选型的影响
时长	总时长多少、是否分多条	1–3 分钟适合平台；5 分钟以上分段或考虑自部署
周期	初稿和终稿什么时候要	周内交付优先平台；自部署有环境与调试成本
出镜与信任	品牌是否要求真人、是否涉客户信任	强信任场景优先真人拍摄+AI 配音
数据与授权	素材是否涉密、形象和声音是否已授权	涉密或需私有化时考虑自部署，授权必须单独确认

三条常见路线对比：

路线	做法	适合	代价与风险
平台数字人	选模板形象或上传授权形象，输入文本生成	1–3 分钟口播、周更内容、快速试错	按分钟/次数计费，形象同质化，细节可控性有限
开源自部署	本地或云端部署驱动与合成方案	长视频、批量生产、私有数据	环境重、需要显卡或云资源，调试与维护成本高
真人拍摄 + AI 配音	真人出镜录画面，后期替换或优化声音	品牌背书、培训、对信任要求高的场景	需要拍摄场地与人员，但自然度最高

决策口诀：三分钟以内且要快，优先平台；超过五分钟、涉密或要批量，评估自部署；客户要“像真人在跟我说话”的信任感，真人拍摄仍是首选。

第二步：按七步流程推进，顺序不能反

1. **脚本口语化**：把书面语改成短句，一句不超过 20 字；把“首先/其次/综上所述”换成“第一/再看/所以”；数字和专有名词标注读法。
2. **配音**：先生成或录制声音，再定语速与停顿；口播类一般每秒 4–5 字，培训类可更慢。声音没定，不要进入驱动。
3. **形象**：选平台形象或准备真人素材，固定服装、发型、背景与光线；同一条片子不要中途换形象。
4. **分段驱动**：按段落或按 20–30 秒切段生成，不要一次生成三分钟；单段出问题只重做该段。
5. **口型校对**：逐段看嘴型与重音是否对齐，重点检查数字、品牌名、英文缩写。
6. **手势与眼神**：减少重复手势，视线尽量看镜头；手势多容易穿帮时改用半身或近景构图。

7. **后期**：加字幕、轻音乐、片头片尾与统一调色；导出比例与投放平台一致（如 9:16 竖屏）。

关键原则：**先定声音，再驱动**。声音变了，口型就要重来；一次性生成长片，返工成本会成倍增加。

第三步：打磨自然度，抓五个细节

- **断句**：逗号停半拍，句号停一拍；长句拆成两句，避免一口气读完；
- **重音**：每句只强调一个关键词，价格、数字、结论适当加重；
- **语速**：每秒 4–5 字为基准，重要信息前后略放慢；
- **停顿**：段落之间留 0.5 秒左右，给观众消化，也给画面切换留空间；
- **口头语**：保留少量“我们来看”“接下来”帮助衔接，但删掉“嗯、那个、然后然后”。

字幕与声音要匹配：每行不超过 16 字，断行跟语义走，不要把一个词拆到两行。

第四步：交付前按清单验收并锁定报价范围

验收不是“看着还行”，而是逐项过一遍可见标准。报价也要把脚本、配音、驱动、字幕、修改轮次和授权写清楚，避免客户默认“全包”。

常见失败与修正

问题	表现	常见原因	修正做法
口型不同步	字已说完嘴还在动，或嘴型与音节错位	单段过长、语速过快、先画面后配音	单段控制在 20–30 秒，先定配音再驱动，问题句单独重做
手势重复	每句话同一个挥手或点头动作	模板动作单一、驱动幅度过大	减少手势段落，改半身/近景，必要时分段选不同动作模板
眼神飘	不看镜头、频繁眨眼、视线游移	素材角度不正、头部运动过大	固定正面机位，减少头部摆动，后期裁切让视线更集中
表情僵硬	全程一个表情，与内容情绪不匹配	整段统一情绪参数、脚本无情绪提示	按语义分段生成，开场/讲解/收束分别设定情绪强度
背景穿帮	人物边缘发虚、光影与背景不一致	抠像质量差、背景光源方向冲突	优先纯色或轻虚化背景，统一光源方向，避免复杂背景

返工优先级：先修口型与读音（影响理解），再修眼神与停顿（影响可信度），最后调手势与背景细节（影响观感）。

一个完整示例

客户需求： 3 分钟产品口播，讲师不便出镜，周五前交第一版，投放微信视频号竖屏，用于新品介绍。

路线选择： 周期紧、时长 3 分钟、信任要求中等（产品介绍，非高风险背书），选择**平台数字人**路线；与客户确认形象与声音授权来源。

脚本与分段（节选 30 秒样片，约 130 字）：

- 第 1 段（0–8s，约 35 字）：“很多人出差带咖啡，总担心太麻烦。这台便携咖啡机，放进背包就能走。”
- 第 2 段（8–16s，约 35 字）：“加粉、加水、按一下，三分钟就能萃取一杯热咖啡，不用找插座。”
- 第 3 段（16–23s，约 30 字）：“杯口出液稳定，香气足，办公室、露营、车里都能用。”
- 第 4 段（23–30s，约 30 字）：“如果你也想随时喝到现磨，点击下方了解参数和价格。”

执行记录：

1. 将客户提供的书面介绍改成短句口语稿，标注“萃取”读音与“3 分钟”重读；
2. 生成女声配音，语速约 4.5 字/秒，段间停顿 0.5 秒；
3. 选定一套商务休闲形象与浅色办公背景，全片不更换；
4. 按 4 段分别驱动生成，每段生成后立即检查口型；
5. 第 2 段“萃取一杯”口型略慢，缩短该句停顿后单段重生成；
6. 加字幕（每行≤16 字）、轻音乐（音量压低，不盖人声）、结尾 logo 定格；
7. 导出 9:16、1080×1920，时长 30 秒，作为第一版样片给客户确认风格。

客户确认点： 先确认样片的“声音风格、形象气质、字幕样式、节奏快慢”，确认后再按同一标准扩展到完整 3 分钟，避免整片返工。

五、30 秒样片的做法

1. 写 80–100 字脚本，切成 4 段，标注重音与停顿（5 分钟）；
2. 生成配音，调语速与停顿，确认数字和品牌名读法（5 分钟）；
3. 选形象并固定服装、背景、机位（5 分钟）；
4. 分段驱动，逐段校口型，问题段单独重做（10 分钟）；
5. 加字幕与轻音乐，导出目标比例并自检一遍（5 分钟）。

六、职业版验收清单

- [ ] 口型与重音对齐，无明显延迟或抢字；
- [ ] 每 10 秒至少一次自然停顿，长句已拆短；
- [ ] 手势不重复，眼神基本看镜头；
- [ ] 字幕每行不超过 16 字，位于安全区且无错别字；
- [ ] 人声清晰，背景音乐不压过人声；
- [ ] 服装、背景、机位与光线前后一致；
- [ ] 导出比例、分辨率与投放平台要求一致；
- [ ] 形象与声音的授权来源已确认并记录。

七、报价提示

- 按分钟计价时，把“脚本口语化、配音、驱动、字幕、后期、2 轮修改”明确写进范围；
- 超过 3 分钟按段拆价或分里程碑验收，避免一口价被无限拖长；
- 形象授权、声音授权、平台会员或算力费用单独列出，不默认包含；
- 加急费、超轮次修改费、脚本重写费提前说明触发条件；
- 若客户后续要求更换形象或更换声音，按新增制作量计，不算“修改”。

八、延伸阅读

- 资料地图 `1.7 AI 数字人+语音合成`
- 资料地图 `1.5 AI 视频`
- 资料地图 `1.6 AI 音乐`

> 下一讲：长视频、定格与科普动画。你会用同一套流程，处理三种不同形态的片子。