

从通信到交互 - AI时代语音技术的挑战

小米大脑 相非

2017.09





远场语音前端面临的挑战

- **产品场景变化**
 - 通信 VS 识别
 - 近场 VS 远场
- 声学环境复杂
- 工程化挑战
- 测试标准缺失
- 技术成熟度低

场景变化

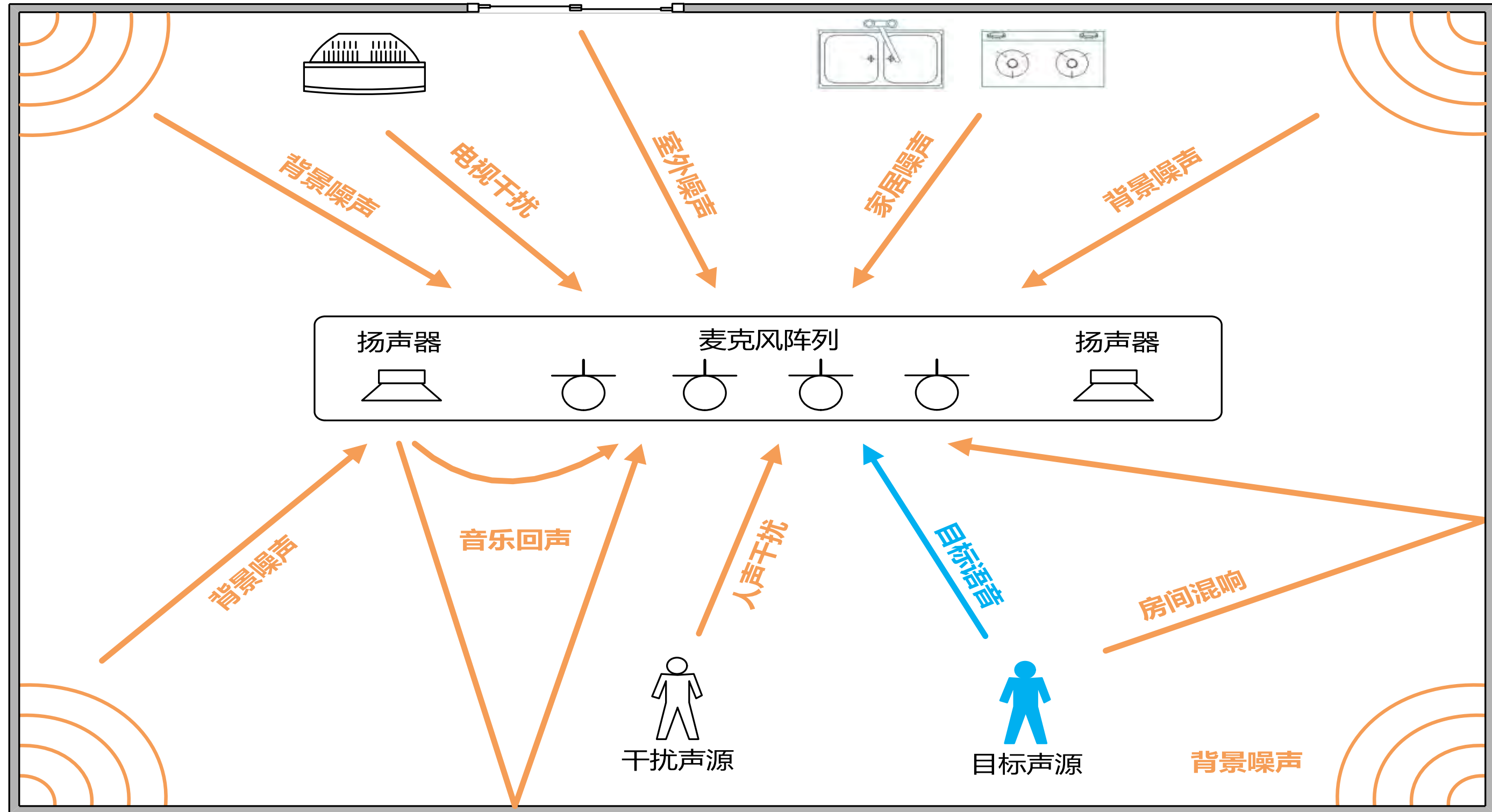
- **通信**

- 场景：近场为主
- 目标：语音清晰度、可懂度
- 技术指标：SNR、MOS分
- 测试标准：3GPP、ITU

- **识别**

- 场景：近场 + 远场
- 目标：语音特征分布？
- 技术指标：SNR？频域失真？
- 测试标准：？

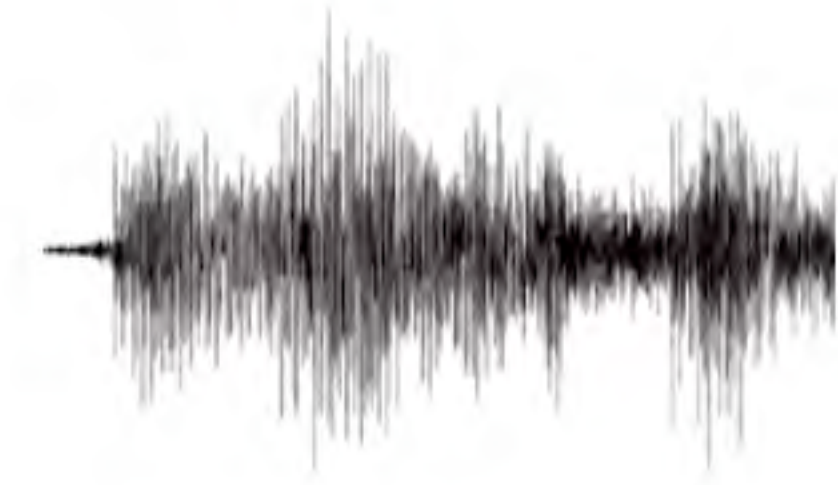
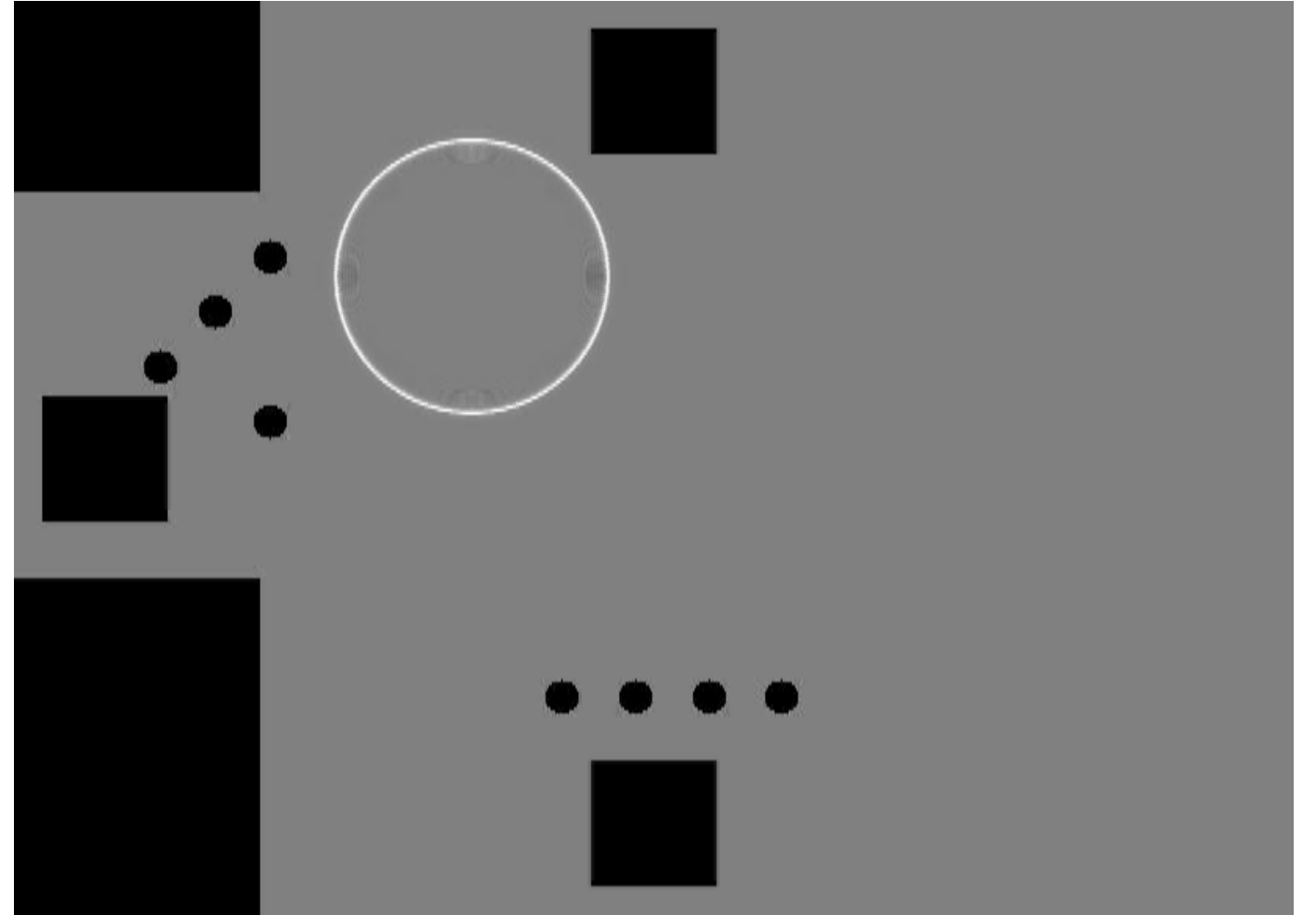
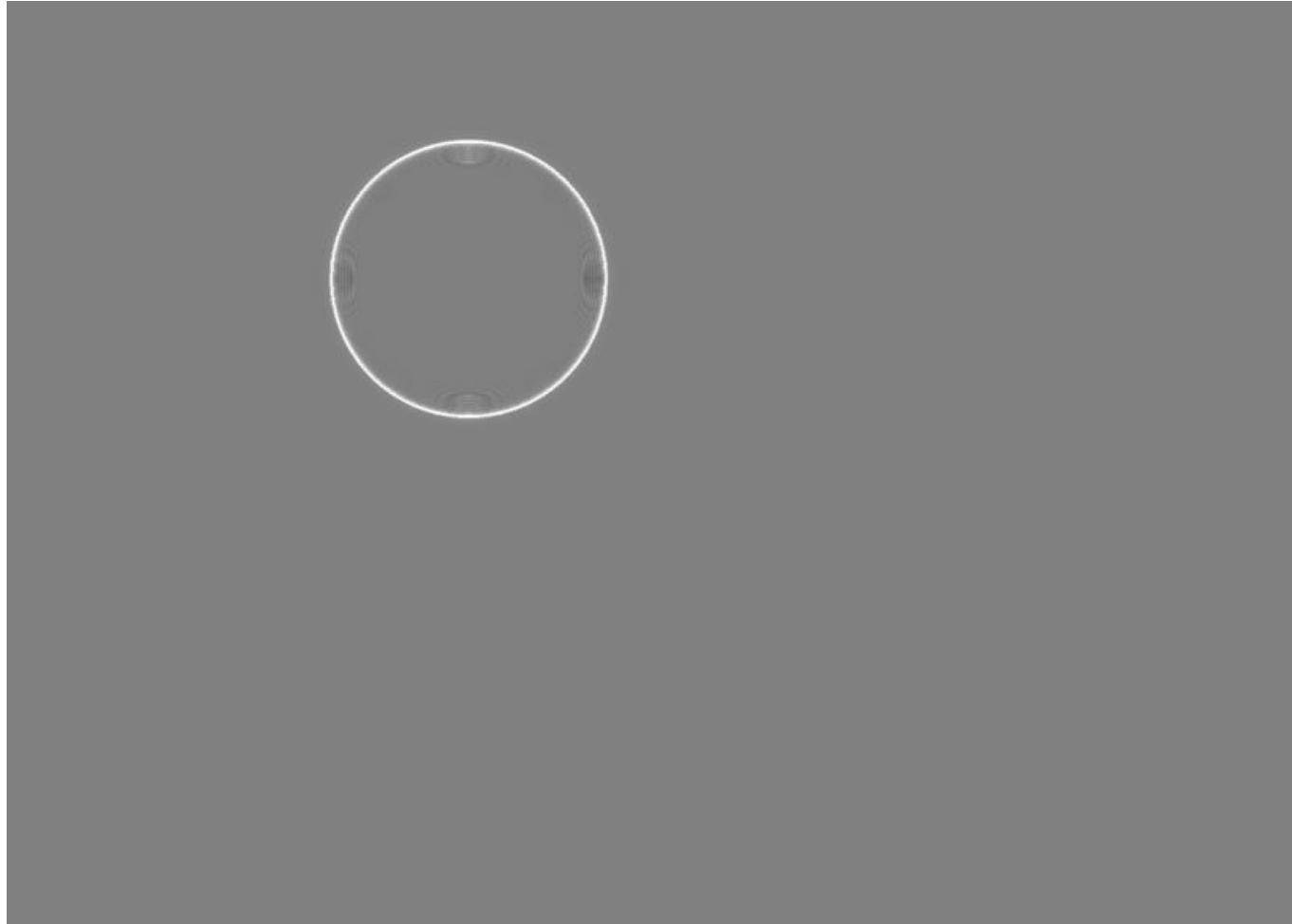
声学环境



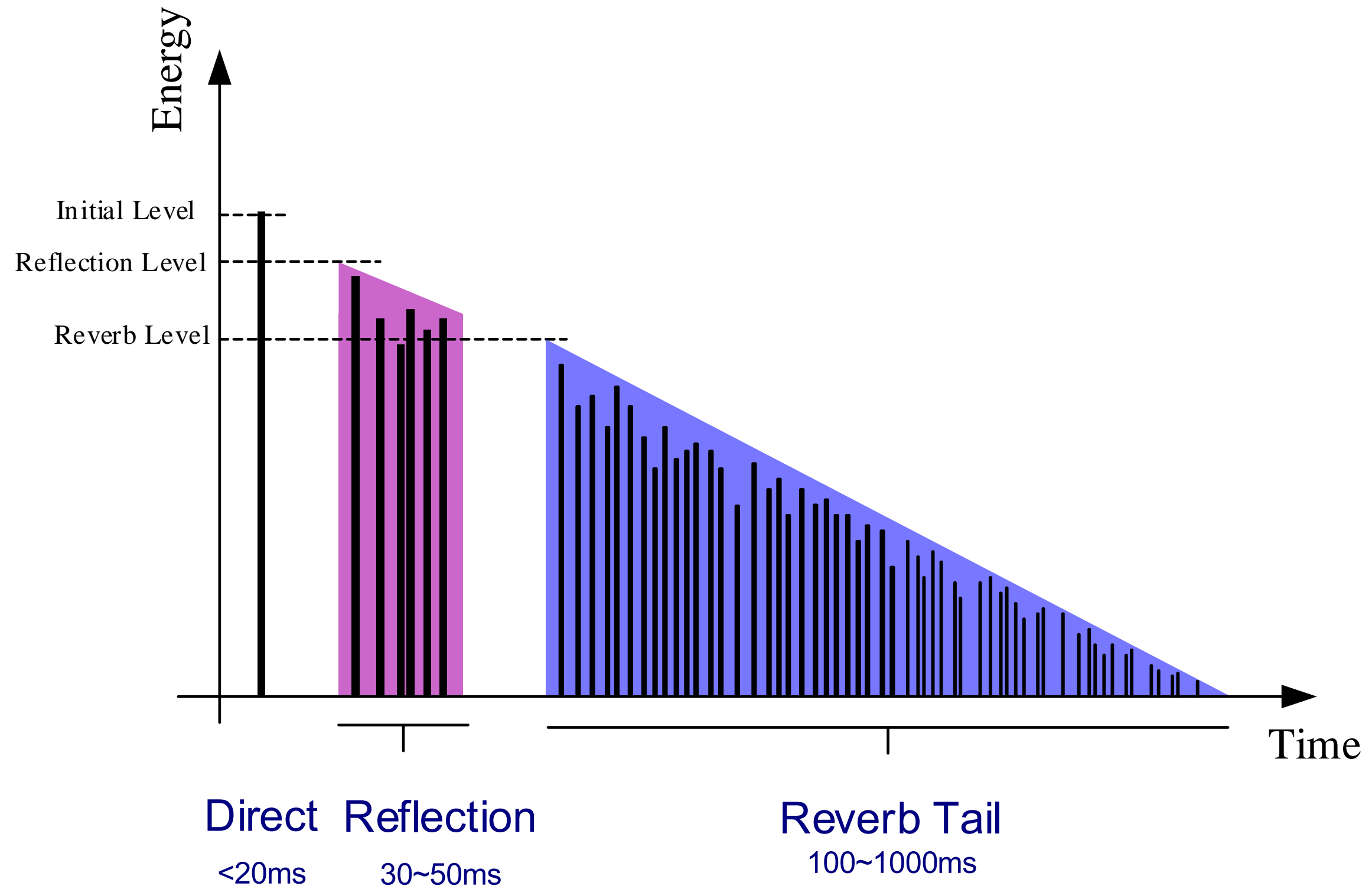
声学环境的挑战

- 目标移动
- 声学回波
- 背景噪声
- 房间混响
- **干扰声源**
 - 电视/音箱干扰
 - 多干扰源并存（鸡尾酒会）
 - 干扰声源与目标声源的切换

房间混响



房间混响模型



工程化挑战

- **硬件一致性要求**
 - 算法性能严重依赖麦克风阵列的一致性
 - 扬声器频响特性对回波抑制影响
- **嵌入式平台的运算性能瓶颈**
 - 多通道处理带来运算量的成倍增加
 - 家居产品的成本限制
- **声学调试**
 - 不同产品场景对声学指标要求的差异
 - 算法对硬件声学设计的依赖性
 - 模块间依赖和耦合关系

解决方案

技术架构

针对回波、噪声、混响的高鲁棒性算法设计
单声源向多声源的算法迭代
单点拾音向分布式拾音的技术演进

工程优化

硬件一致性自动检测及补偿功能
基于CPU和GPU的性能深度优化
灵活可配置的工程架构和算法模块设计

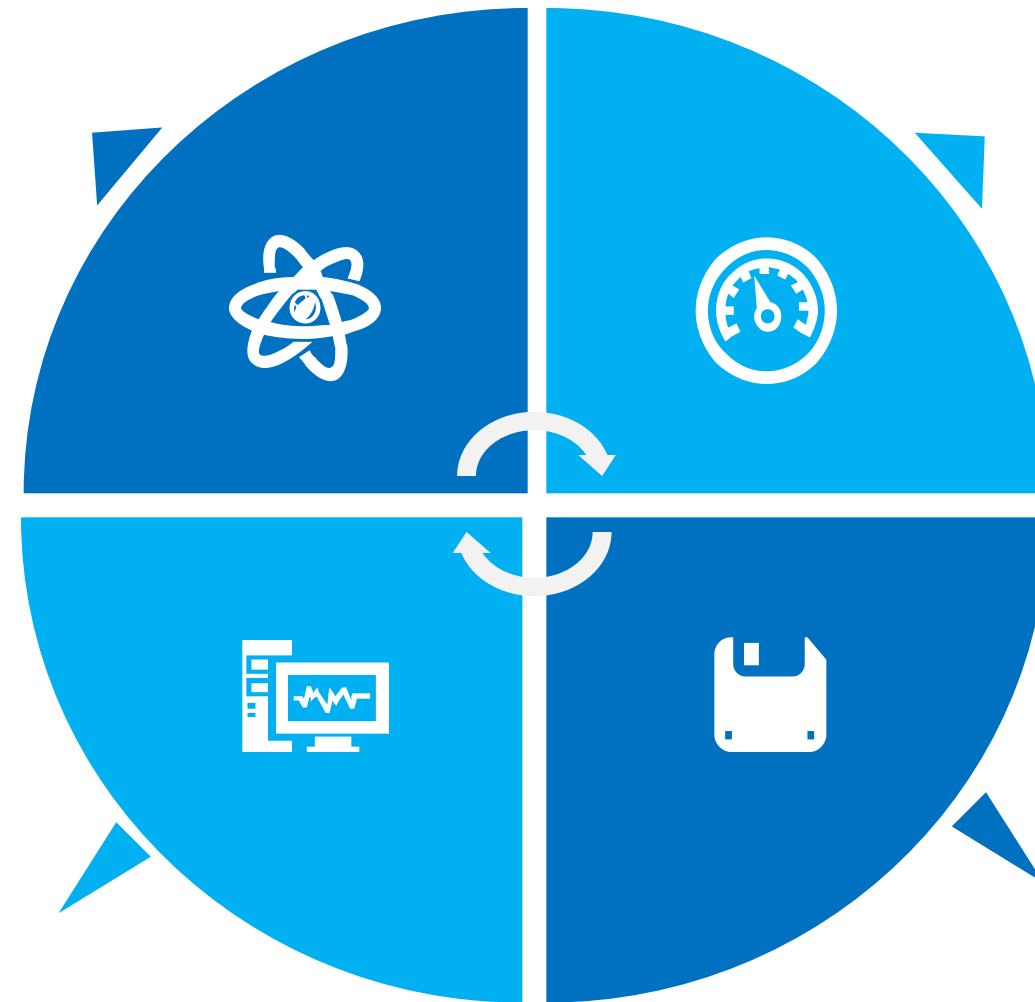
测试评估

确定识别率与前端指标的大致对应关系，实现指标量化
结合智能家居场景设立测试标准，进行横向和纵向对比
建立技术运营体系，不断发现和解决BadCase

智能语音能力建设

核心技术体系

物理声学、麦克风阵列设计
多通道/单通道语音增强
语音识别、命令识别、语音唤醒
语音合成



测试评估体系

声学测试、麦克风测试
增强算法测试
唤醒率、识别率测试
合成主客观测试

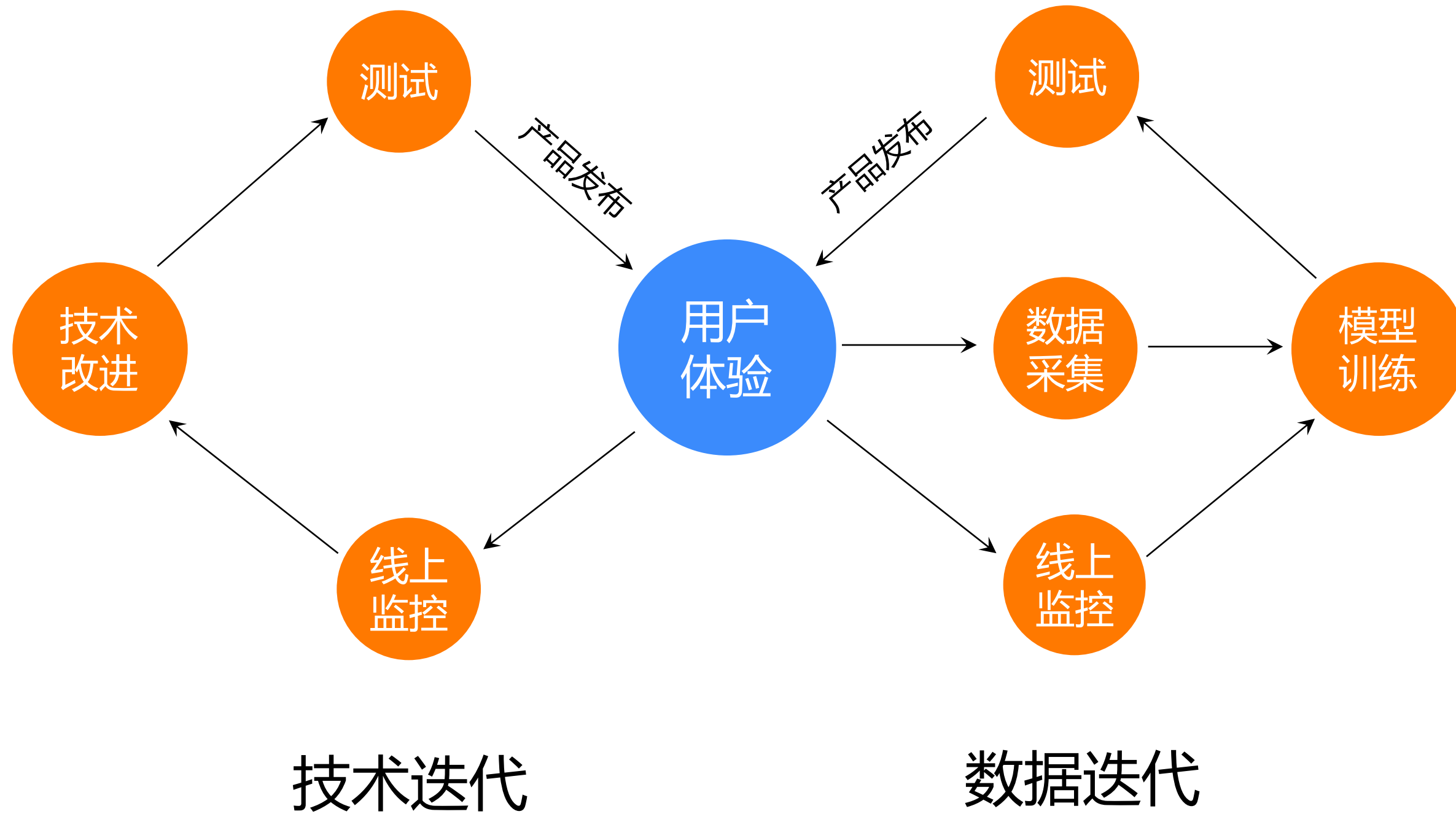
技术运营体系

业务指标、技术指标监控
线上服务质量评估
云端控制系统

数据迭代体系

数据采集与清洗
数据标注
模型迭代

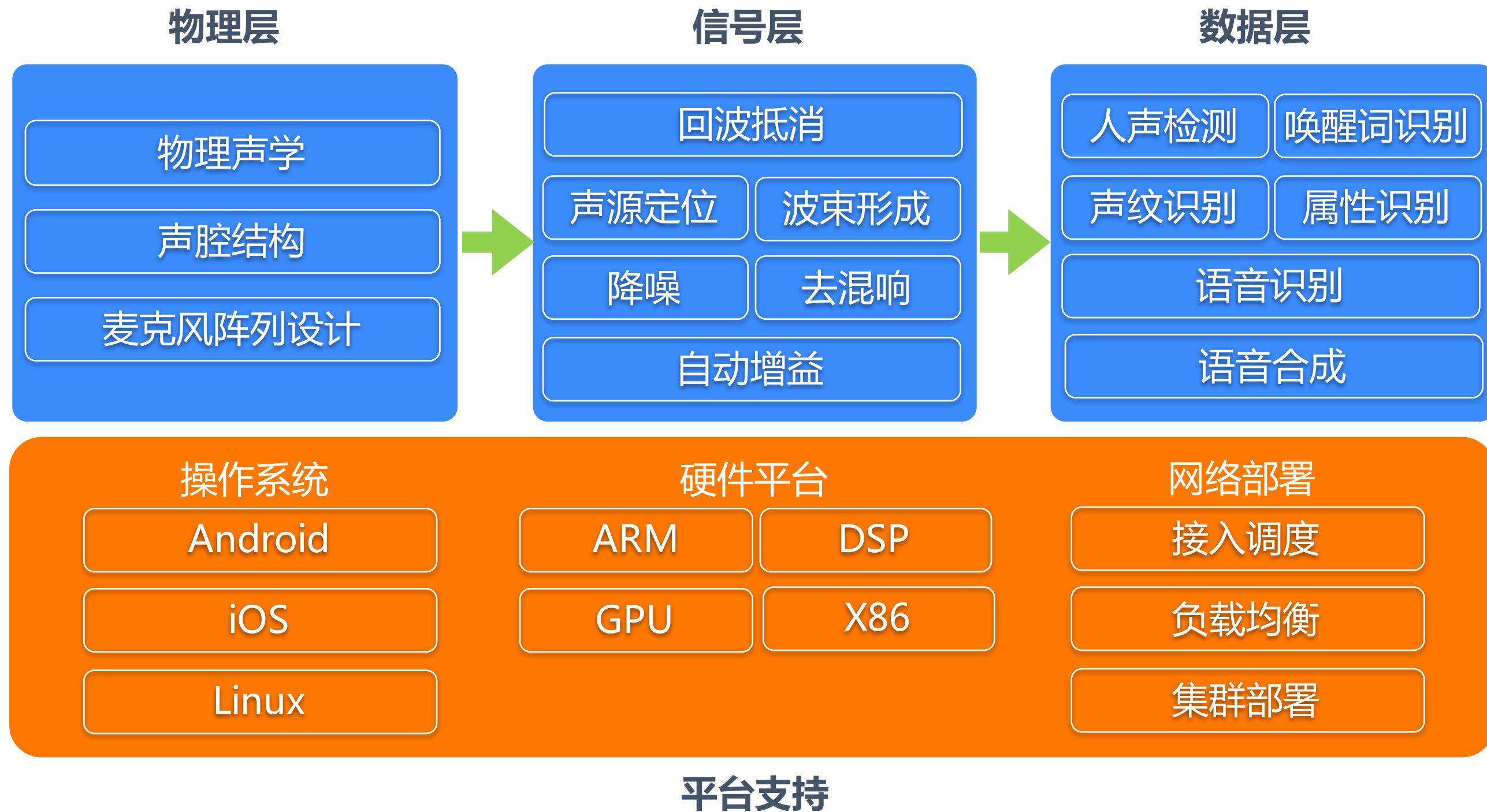
用户体验改进模型



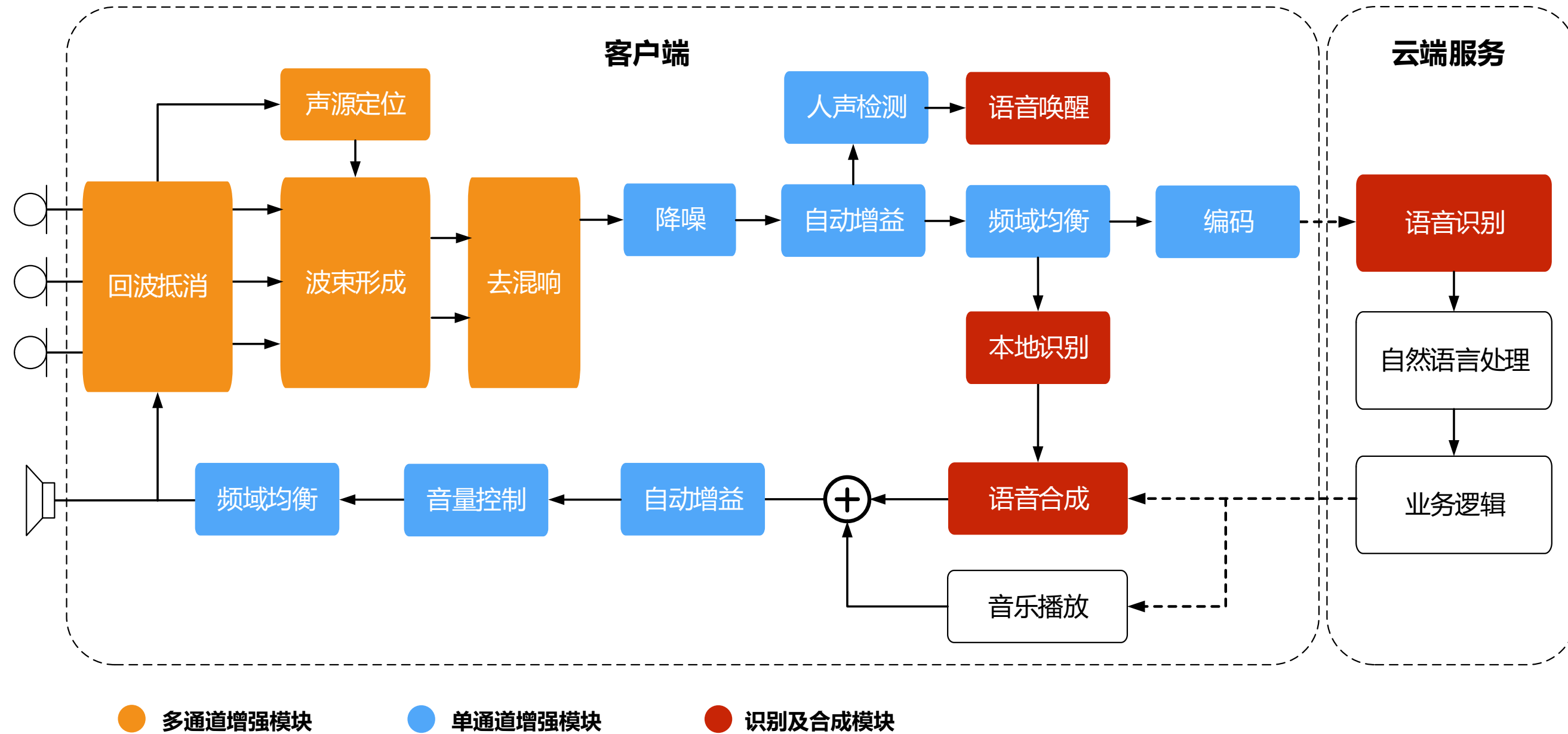
核心技术体系-语音通信



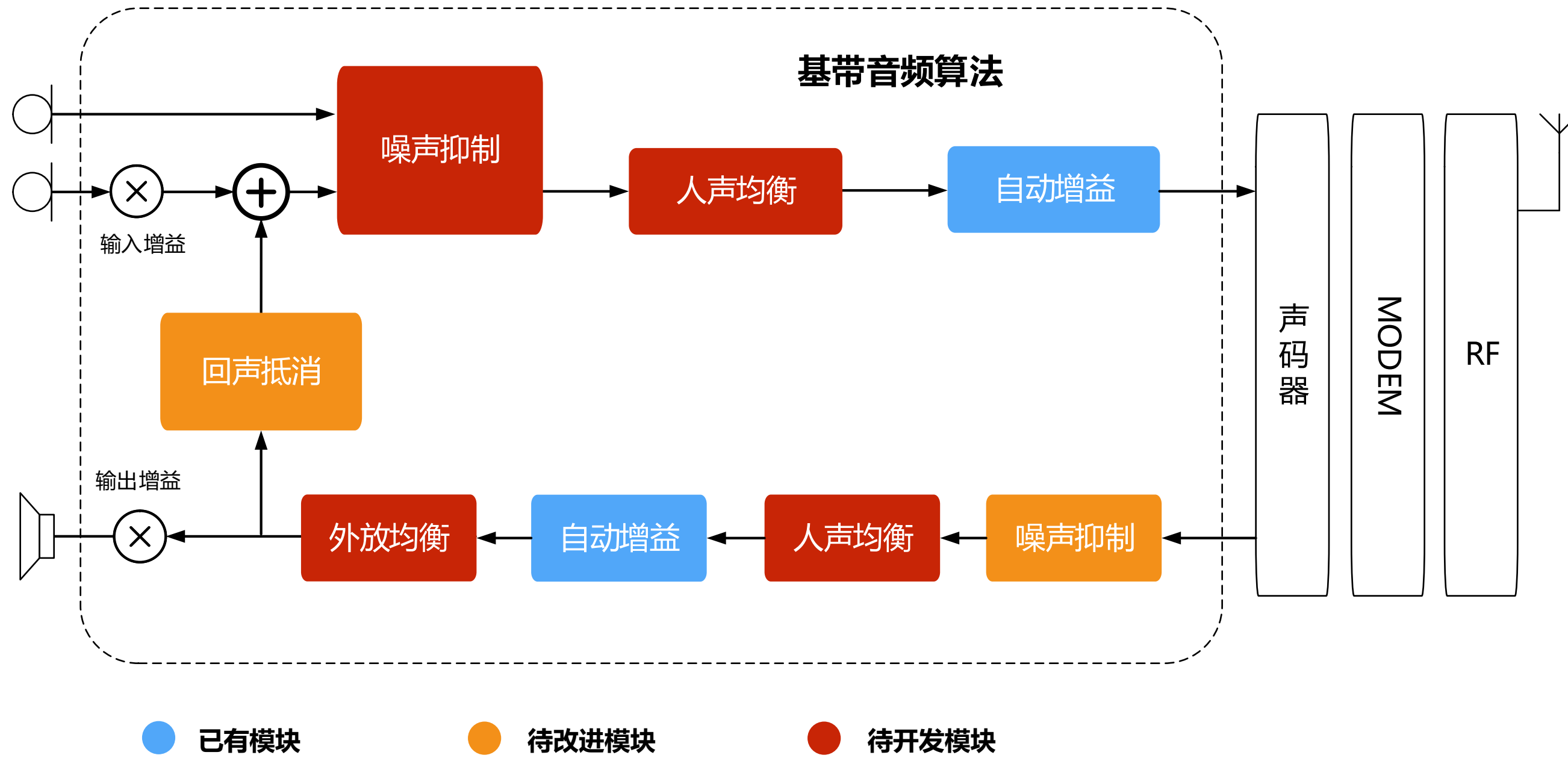
核心技术体系-语音交互



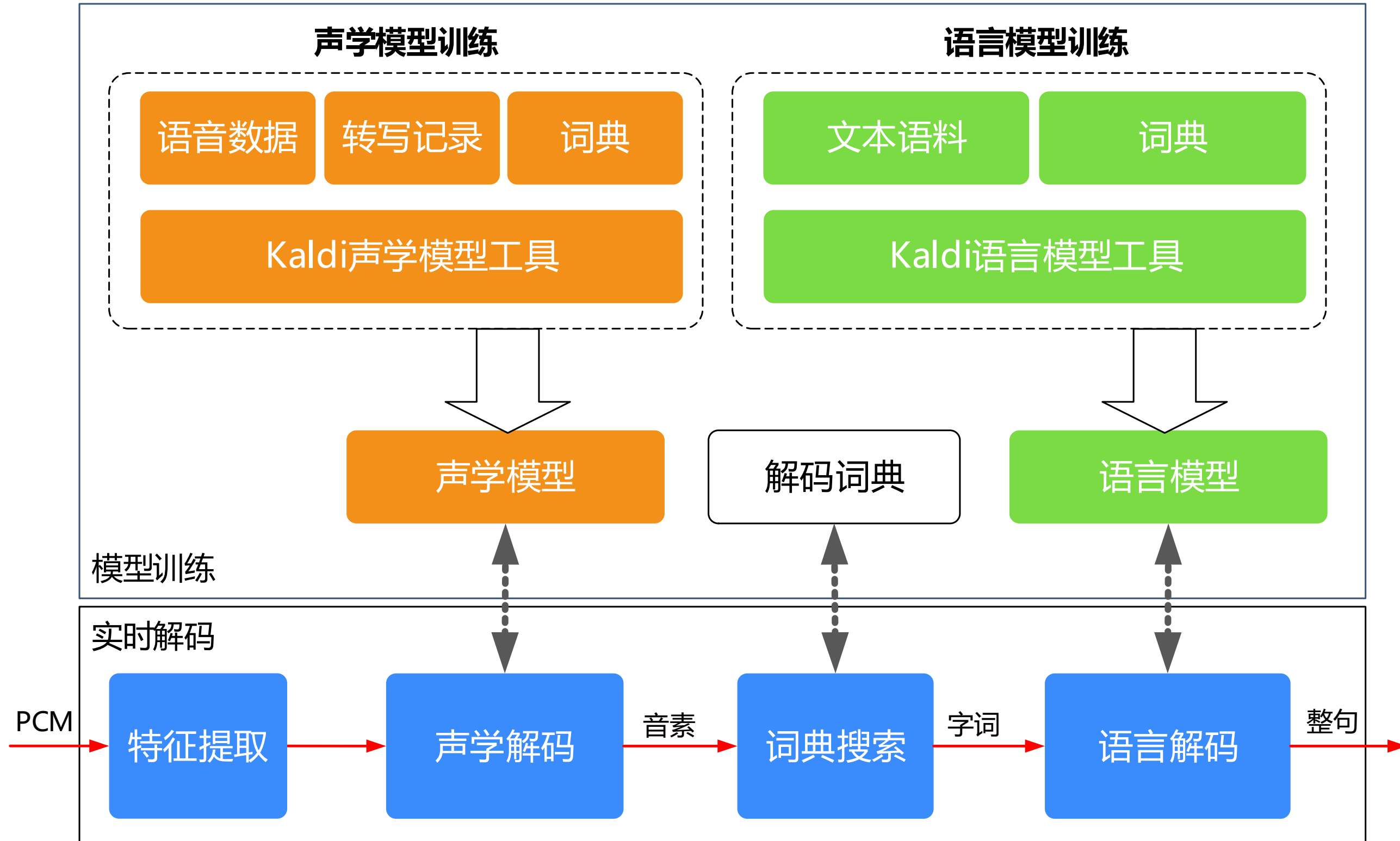
技术架构-语音交互



技术架构-语音交互



语音识别技术架构



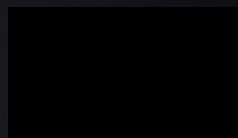


小米人工智能电视





小米AI音箱





小米AI音箱目前已经支持的技能



听音乐



天气



闹钟



听新闻



广播电台



相声小品



精品小说



脱口秀



路况



算数



时间



查找手机



评书



留言



儿童故事



备忘



单位换算



股票



汇率



公开课



亲戚称呼



听声音



问答百科



闲聊笑话

谢谢大家，欢迎交流！