

构建鲁棒的车载语音识别系统

腾讯CSIG智能平台产品部 曹松军

2020.07

腾讯云小微深耕产业互联网，在出行方向，联手腾讯车联TAI：共同打造智趣、安全、高效出行场景化解决方案。目前已合作接近30家车企，50余款车型。

关注「腾讯产业」互联网公众号
 回复「加群」加入技术交流群



>>> 腾讯车联TAI解决方案 <<<

腾讯云小微语音

腾讯地图

爱趣听

小场景



- ▶ 出行场景
通勤、出游、拥堵.....
- ▶ 社交场景
集结、路书、互动.....
- ▶ 娱乐场景
听歌、小说、游戏.....

车载微信

QQ音乐

全民K歌

.....

>>> 腾讯云小微车载语音产品及能力 <<<

车载降噪

风噪胎噪
 机器回声
 音乐噪声
 空调噪声
 车内聊天

车载语音引擎

语音唤醒引擎
 云+端语音识别
 云+端语义理解
 离线TTS引擎

车载技能生态

音乐 电台 新闻
 导航 周边美食
 电话 车控 天气
 知识查询 闲聊

前沿技术

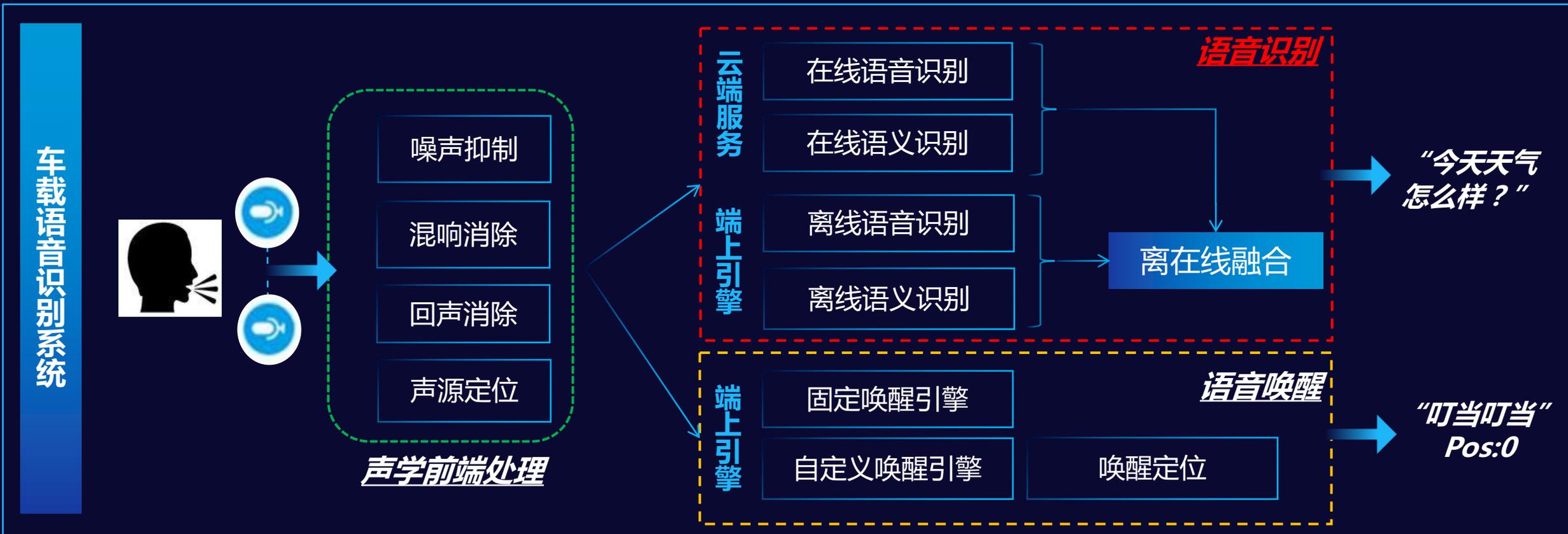
全双工 多音区 虚拟人

合作车企





云小微车载语音识别系统





车载语音识别面临的挑战



关注「腾讯产业」互联网公众号

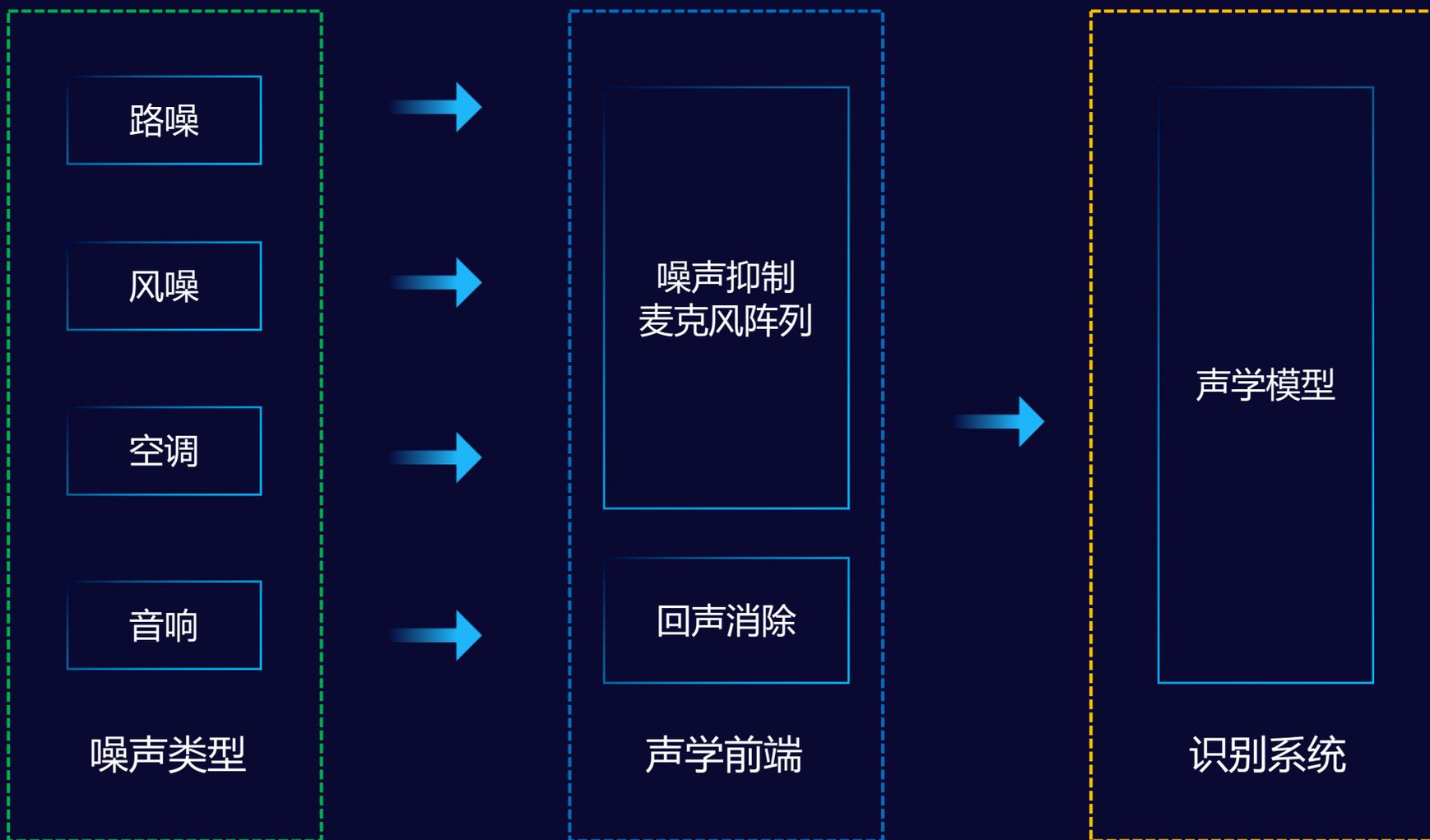
回复「加群」加入技术交流群 



挑战— 噪声

噪声问题

关注「腾讯产业」互联网公众号
回复「加群」加入技术交流群



声学前端

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群



• 打造腾讯车载声学前端方案

- 从车内语音交互场景出发，与车联网共同规划车载语音声学前端需求
- 通过内、外部合作方式共同输出车载声学前端方案，目前与**腾讯AI LAB**等进行了接入适配
- 制定并推进声学前端的标准化准入流程

• 方案的特点

- 软件、硬件方案齐备
- 软件前端方案全链条由腾讯自有技术构成
- 适用面广，覆盖前装和后装、高端和低端
- 功能完备（AEC、NS、DOA、多音区）
- 建立了第三方声学前端方案准入的标准化流程

方案	MIC数目	前端能力	目标平台	落地产品
单MIC方案	1	回声消除 噪声抑制	后装车机、改款前装车	瑞联、蘑菇车联、同行者、广汽丰田
双MIC硬件方案	2	回声消除 噪声抑制	前装车机	长安、广汽、长城等多款车型
双MIC软件方案	2	主副驾定位和拾音	前装车机	柳汽T5长城H系列
四MIC硬件方案	4	回声消除 噪声抑制	前装高端车机	-
四MIC软件方案	4	全车定位和拾音		理想One

声学模型

关注「腾讯产业」互联网公众号

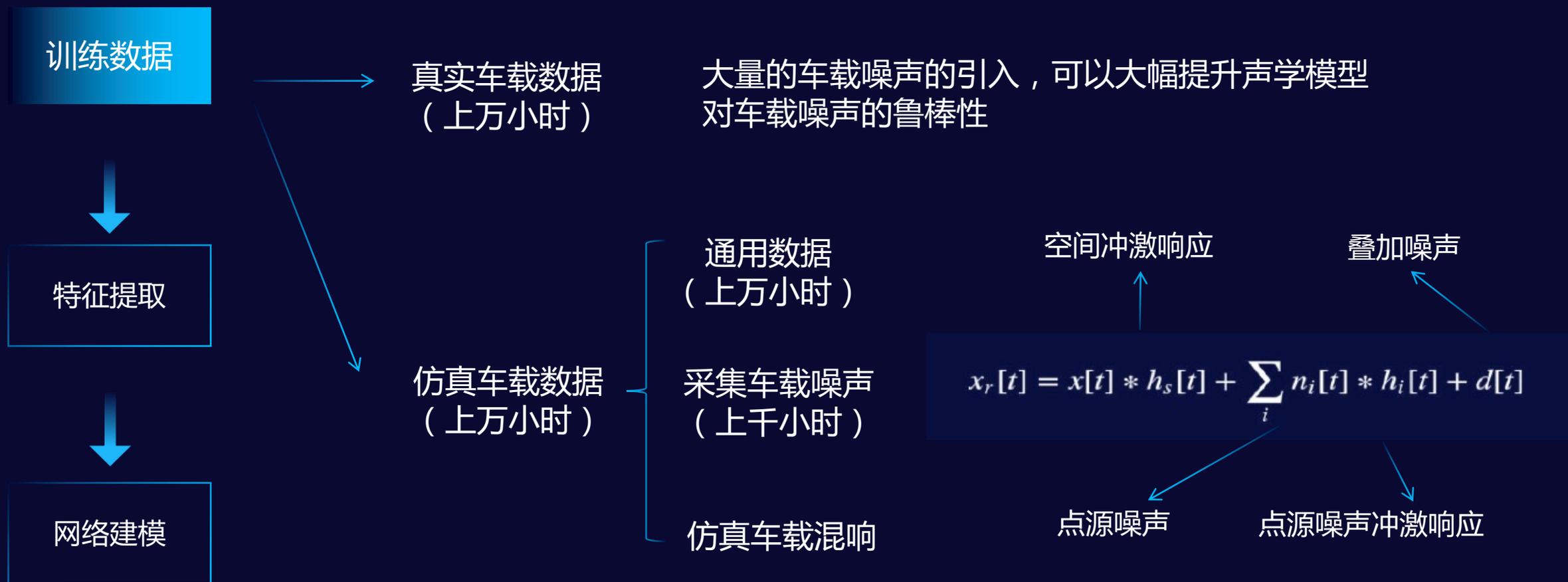
回复「加群」加入技术交流群 



声学模型

关注「腾讯产业」互联网公众号

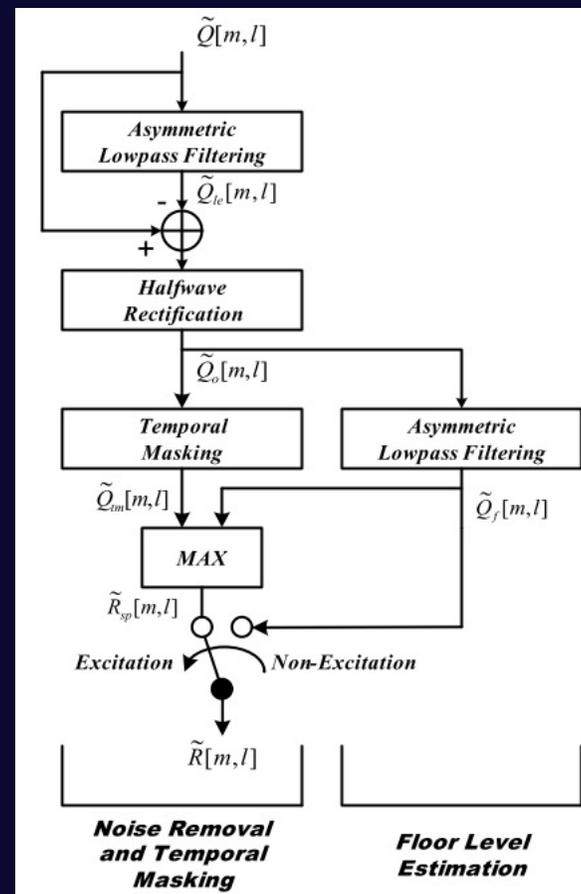
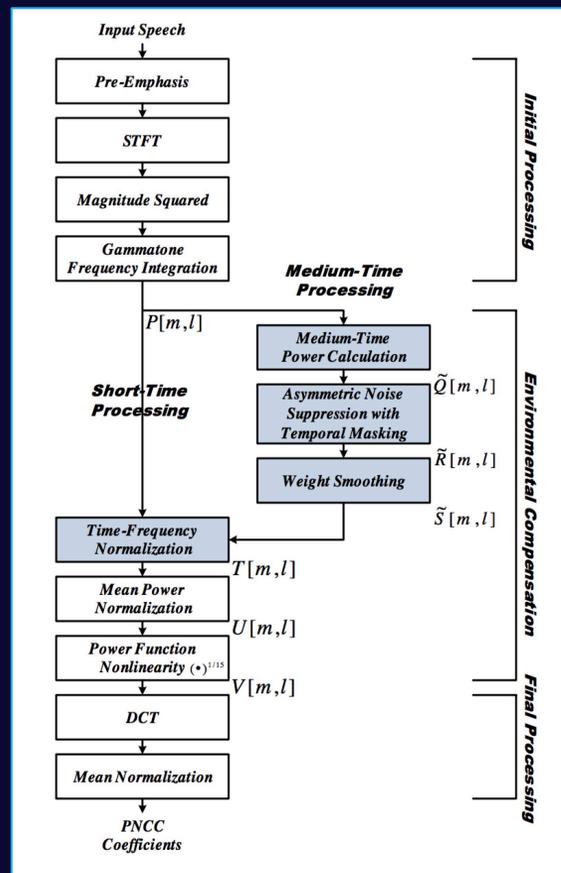
回复「加群」加入技术交流群



声学模型

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群



PNCC相比传统的MFCC，增加了时域和频域的平滑处理，具有更强的抗噪能力，噪声测试集可以获得5%的相对提升

声学模型

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群



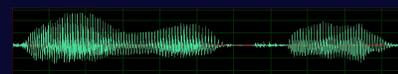
训练数据



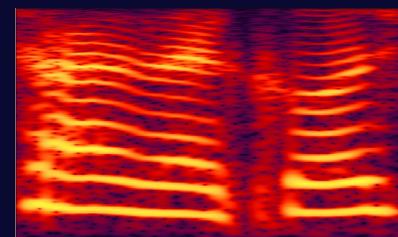
特征提取



网络建模

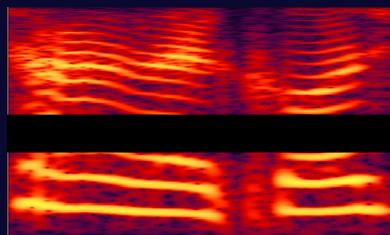


语音

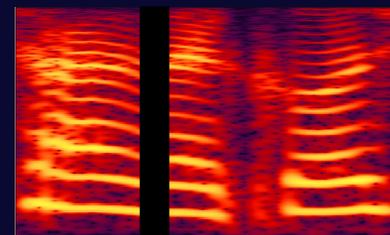


频谱图

时域mask



频域mask



SpecAugment对训练数据在时域和频域做mask，扩充数据的多样性

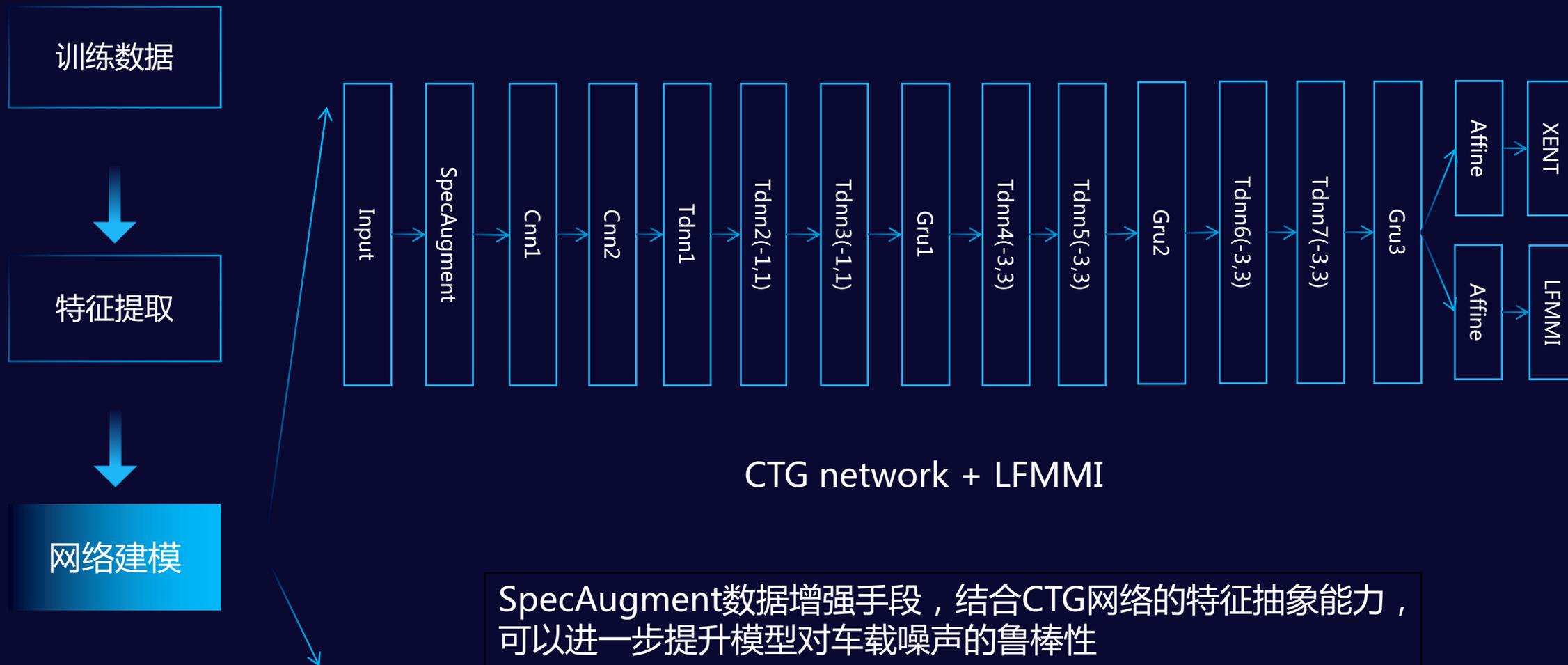
1.时域 $[t_0, t_0 + t)$ ，起点和窗长随机选取

2.频域 $[f_0, f_0 + f)$ ，起点和窗长随机选取

声学模型

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群



关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群 



挑战二 口音

口音问题

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群



中国方言区域：

MANDARIN

YUE

WU

XIANG

GAN

...

he lan欢迎你 -> 荷兰？河南？

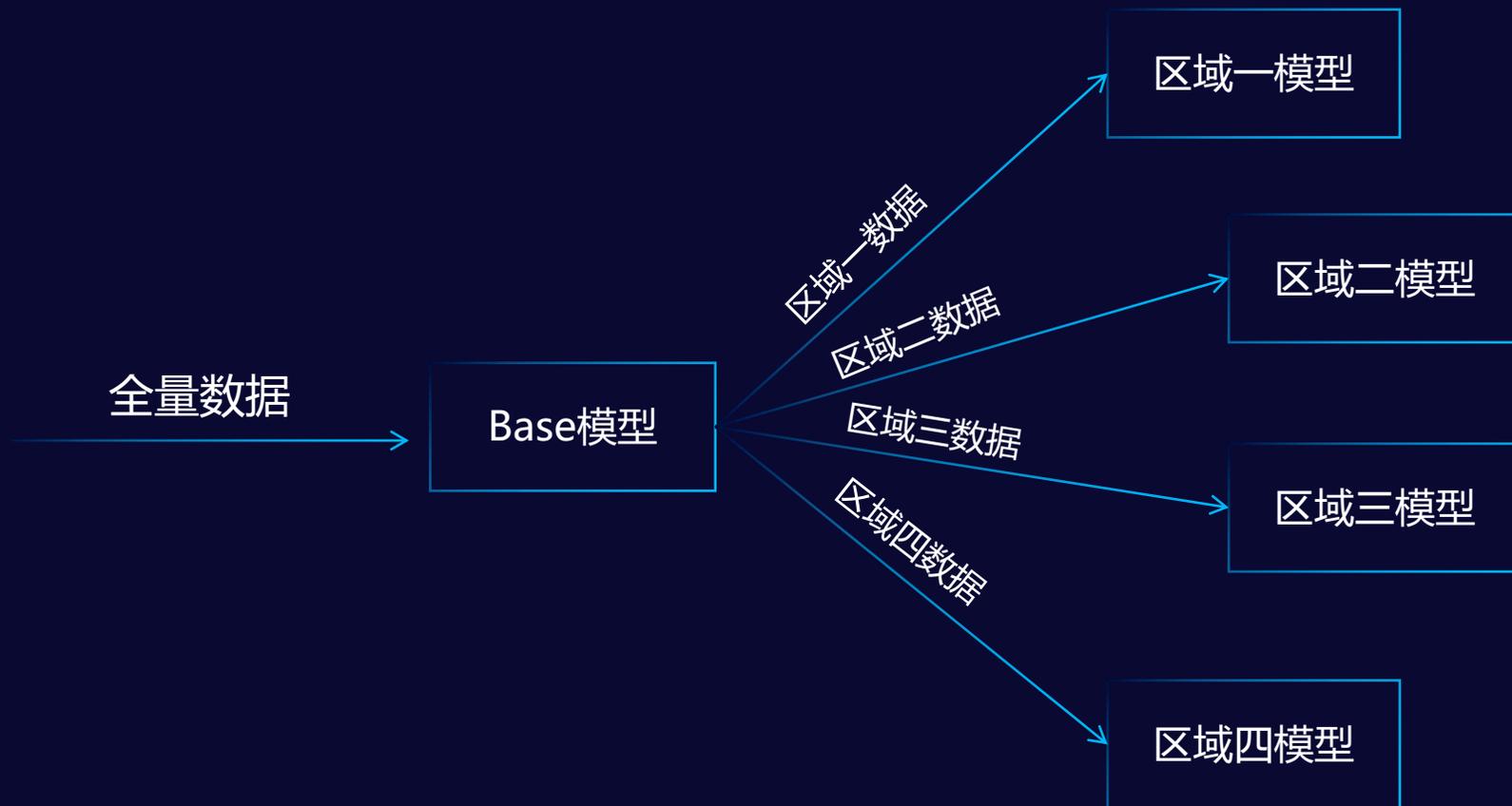
hu jian欢迎你-> 胡建？福建？



方案一 分区域声学模型

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群



将全国划分为多个不同的区域，基于base模型进行finetune得到区域模型

结果对比

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群



方案	提升幅度	特点
分地域声学模型	8%-10%	复杂，区域之间解耦
地域向量声学模型	5%-8%	简单，区域之间耦合

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群



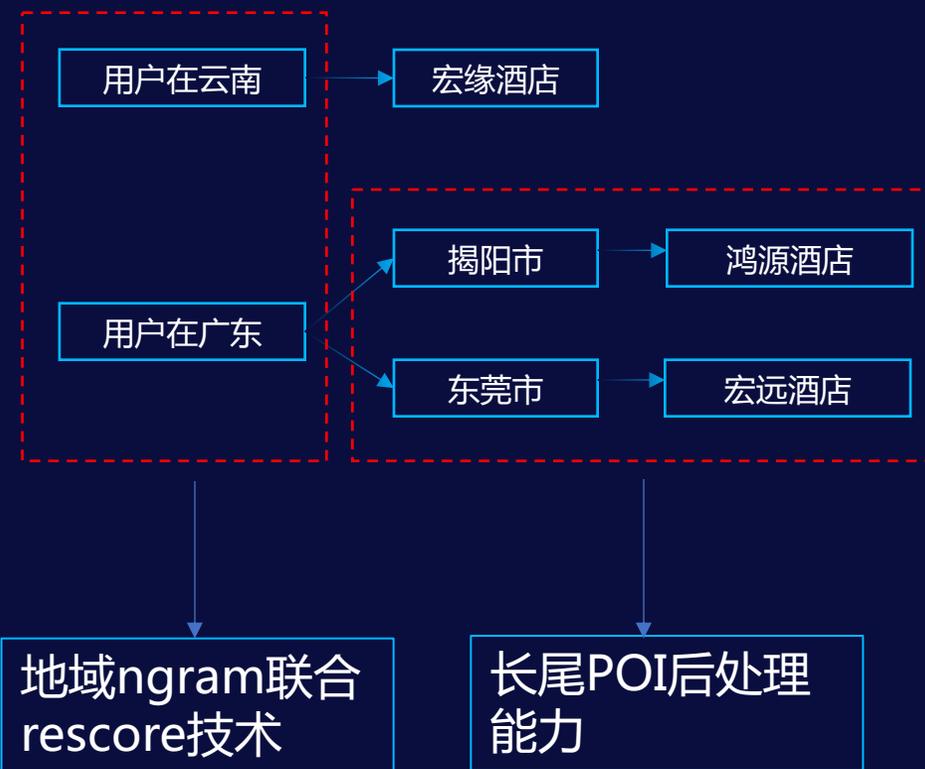
挑战三 海量POI



POI问题

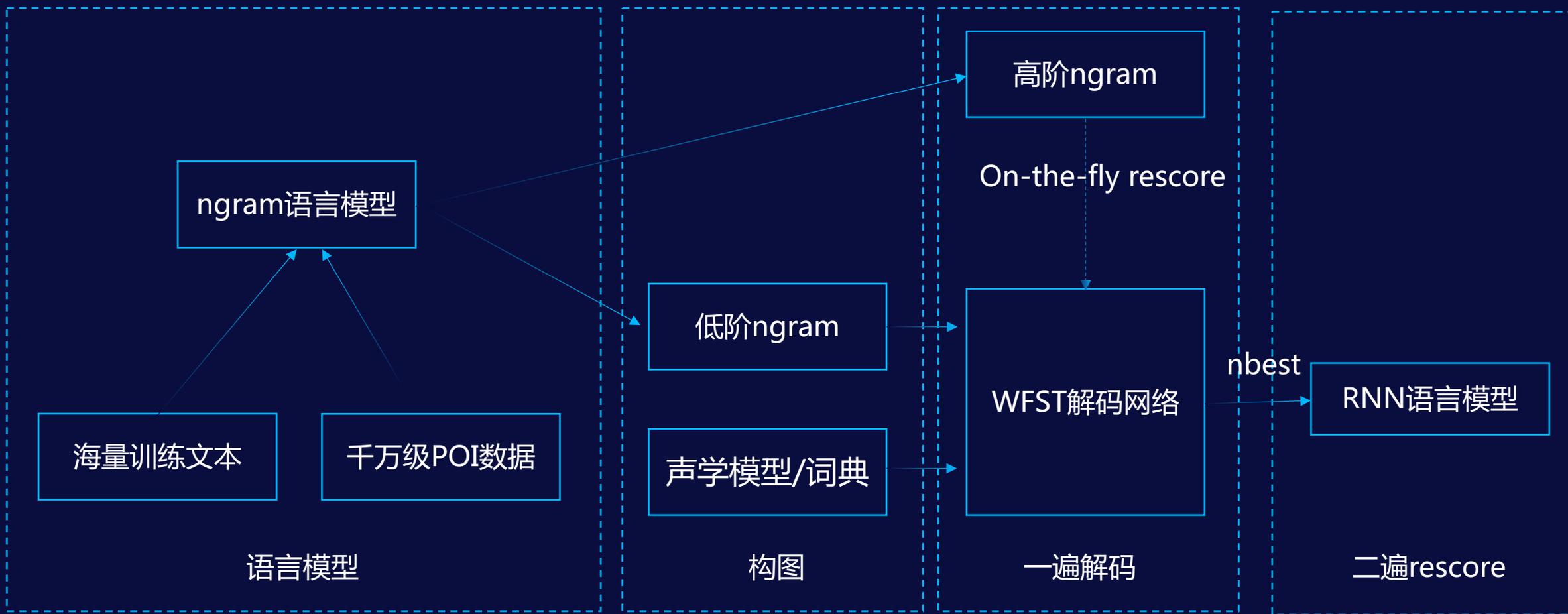
关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群





POI识别-强大的基础建模能力



关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群



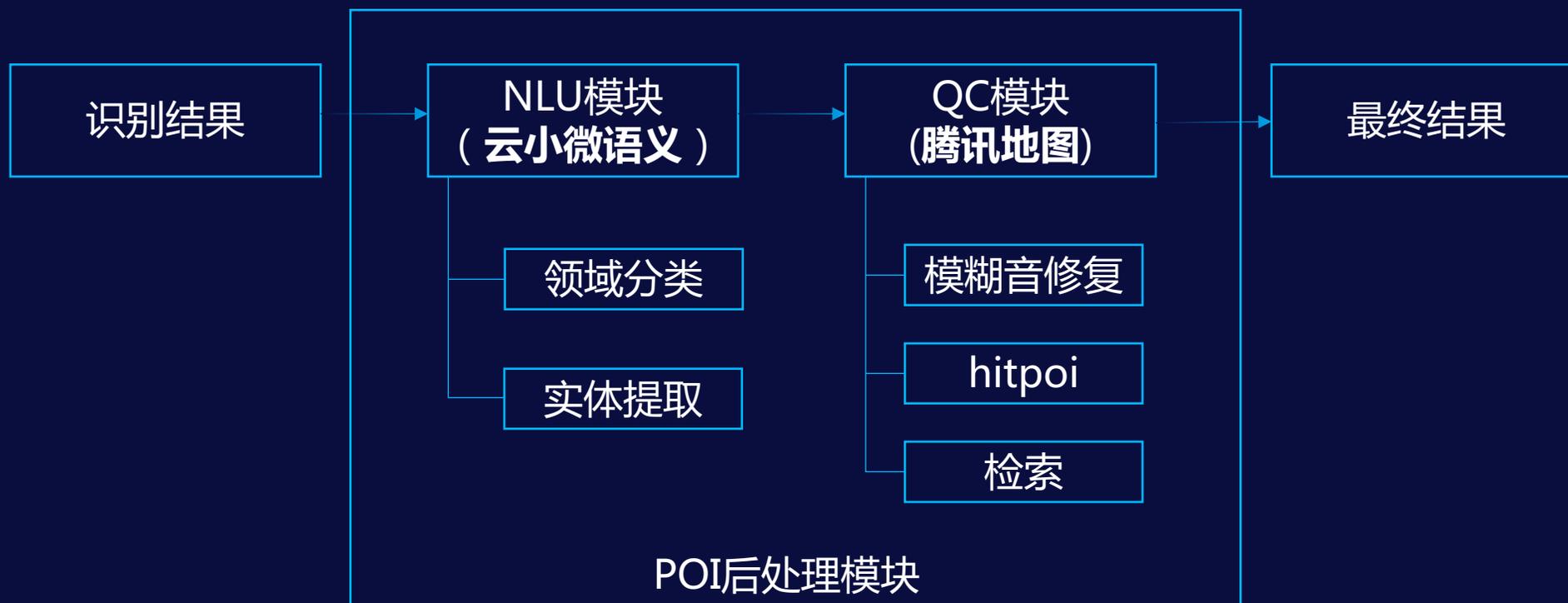
POI识别-地域ngram联合rescore



POI识别-长尾POI后处理

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群



我要去宏远酒店

我要去{宏远酒店}

{宏远酒店}+揭阳市->{鸿源酒店}

我要去鸿源酒店

POI识别率获得10%的性能提升

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群 



挑战四 场景化

场景化问题

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群 



场景一

产品的特殊词汇

车载小程序，“可见即可说”，支持识别屏幕上的词汇

场景二

产品的个性化句式

“打电话给***”等固定句式的支持

场景三

产品的个性化领域

某个车载助手接入识别服务，希望上传积累的文本数据来提升识别效果

技术方案

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群 



场景	技术方案	特点
场景化词汇	Hotfix增强技术	支持配置上千个重点词汇 支持在线更新，分钟级响应时间
场景化句式	Grammar增强技术	需要提供句式+槽位 支持在线更新，分钟级响应时间
场景化领域	文本自学习增强技术	用户提供文本 支持在线更新，小时级响应时间

效果对比

关注「腾讯产业」互联网公众号

回复「加群」加入技术交流群 



	baseline	baseline+增强技术
场景一测试集	4.85	2.40
场景二测试集	5.87	1.0
场景三测试集	2.96	1.04

三种增强技术方案可以满足不同场景的产品需求，大幅提升场景化的识别效果，可以提供更好的语音交互体验

关注「腾讯产业互联网」公众号
回复「加群」进入技术交流群



Thanks!