

# 用多模感知技术打造智能网联汽车新体验

王兴宝

科大讯飞股份有限公司

智能汽车BU 算法研发总监

2021-07-08 | 河北 沧州

## 亚太地区知名智能语音与人工智能上市公司



中文智能语音及人工智能产业领导者

002230  
股票代码



- 语音及语言信息处理国家工程实验室
- 中国第一个认知智能国家重点实验室
- 登榜2019MIT TR全球50家聪明公司
- 入选首批国家新一代人工智能开放创新平台

牵头制定中文语音技术标准

2003年、2011年，两次荣获“国家科技进步奖”

2005年、2011年，两次获得中国“信息产业重大技术发明奖”

# 科大讯飞已成为国内汽车厂商首选的语音交互产品品牌

深耕**18**年

完成**1000**余个车型项目交付

覆盖**90%**以上中国自主品牌和合资品牌

**2500**万辆及以上汽车前装搭载




# 十八年的沉淀

从2003年到2021年，科大讯飞沉淀了全闭环的汽车语音交互核心技术



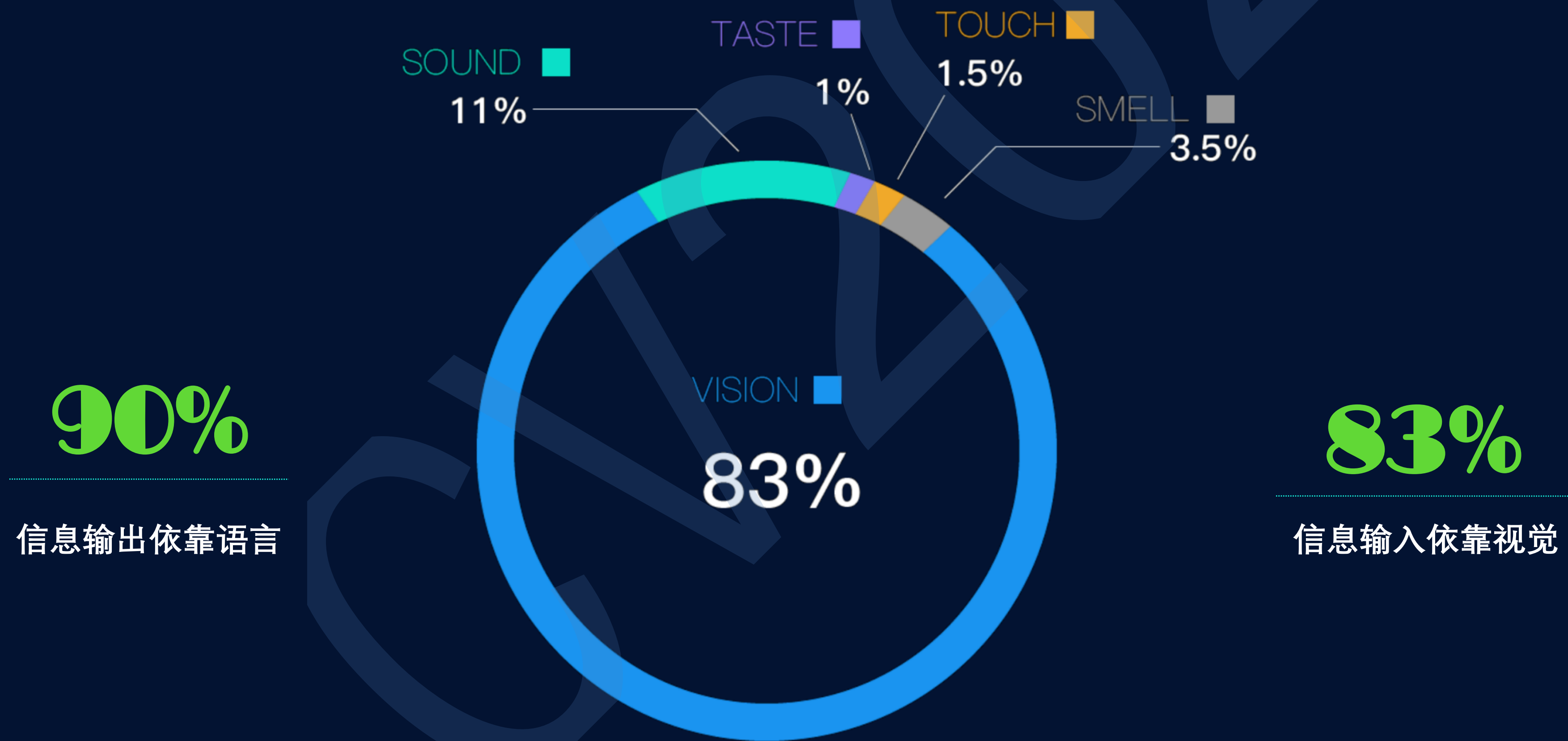
# 每台车都有一个懂你的智能助手

这位具有复杂属性的“朋友”，需要产品、技术、工程共同“培养”



# 信息输入与输出

根据美国哈佛商学院有关研究人员的分析资料表明，人的大脑每天通过五种感官接受外部信息的比例分别为：味觉1%，触觉1.5%，嗅觉3.5%，听觉**11%**，以及视觉**83%**。



**90%**  
信息输出依靠语言

**83%**  
信息输入依靠视觉

# 站在车的角度看人机交互

## 听说看显的技术、交互理解的艺术



# 听

汽车比人听得更轻松、更仔细、更准确，让服务精准送达

任意时刻，一语必达  
高噪场景识别率超90%



Anytime

任意用户，懂你所讲  
多语种/方言 识别率超87%



Anybody

任意位置，精准到“位”  
声音定位准确率超98%



Anywhere



# 听

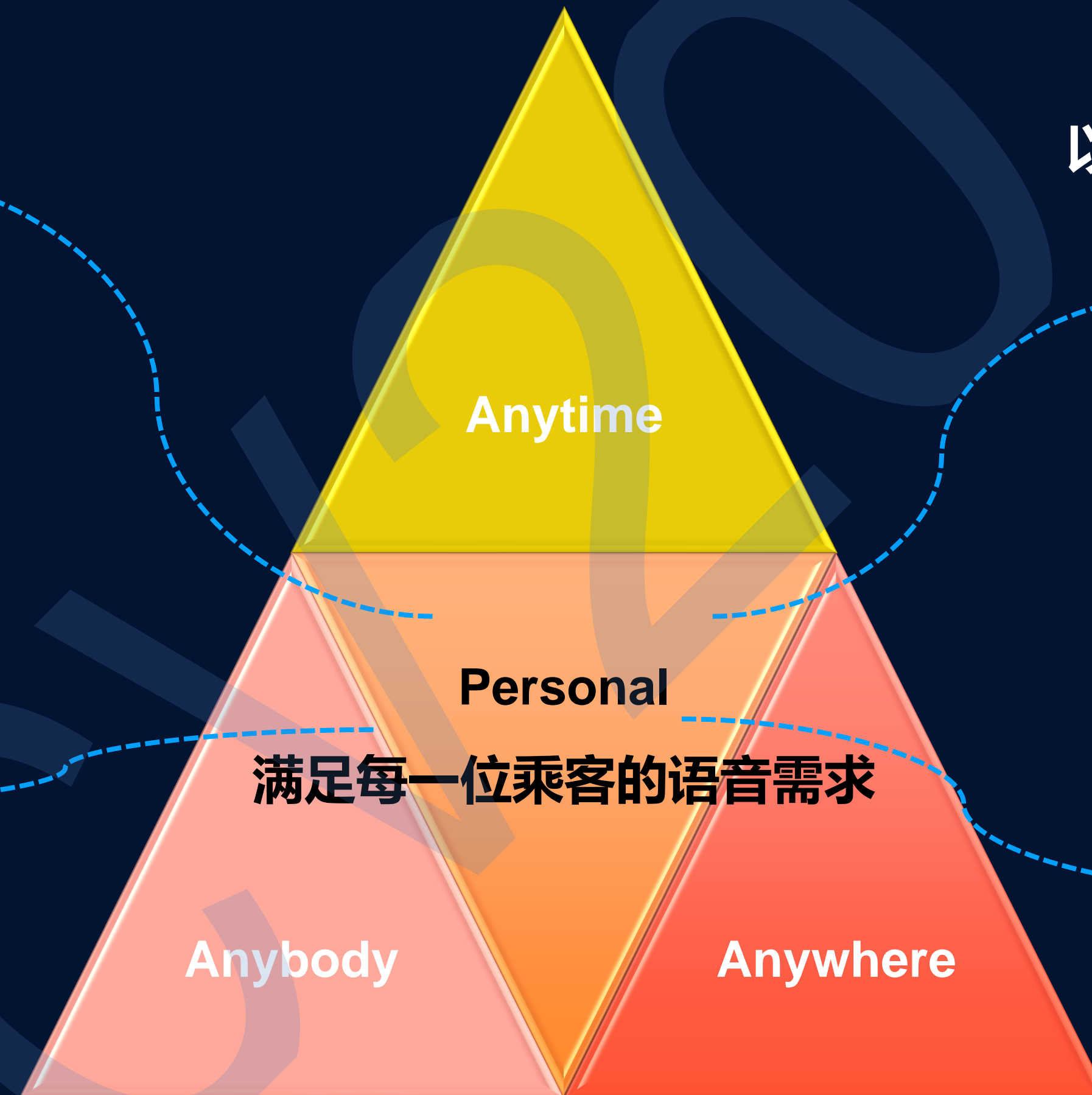
汽车比人听得更轻松、更仔细、更准确，提供贴心服务

**多模语音识别**  
让交互更加自然  
多模语音识别  
视觉VAD  
多生物鉴别VoiceID+FaceID

**个性化识别**  
以你喜欢的方式，听你所说  
唤醒/识别自适应  
语法习惯自适应  
语义自定义

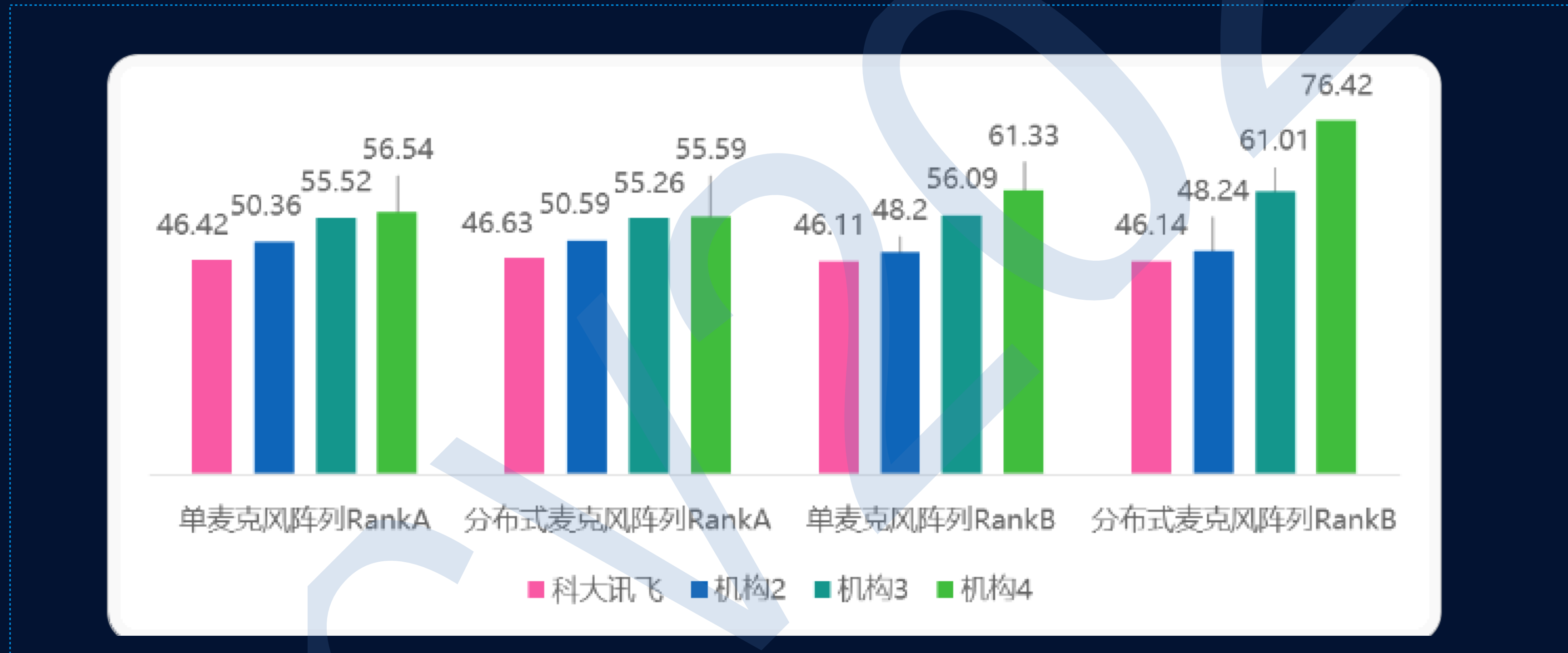
**情绪识别**  
让交互更有情感  
基于语音情绪识别  
基于文本情绪识别  
性别识别  
年龄识别

**细分领域识别**  
让识别无死点，一路畅通  
汽车知识库识别  
女性人群识别  
儿童人群识别



# 听

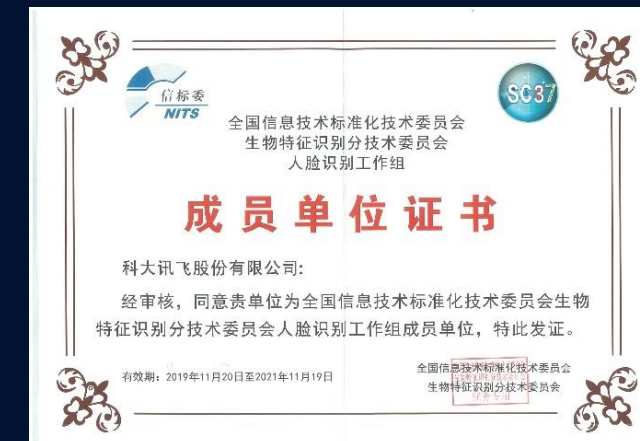
## 国际英文语音识别大赛中高噪音环境下获得突破



2019年9月包揽国际多通道语音分离和识别大赛 (CHiME) 全部四个项目的冠军，比官方系统提升绝对35个点

# 看

汽车比人看的更清楚、更仔细、更准确



人脸识别国标工作组成员

## 人机交互



### 人脸识别

无感登录+活体检测  
暗光下精准率 > 99.9%



### 注视追踪

所见即所控  
AOI精准率 > 99%  
召回率 > 95%



### 手势识别

手势识别20+  
静态手势+动态手势  
准确率 > 98%



ICCV2019

WIDER FACE AND PERSON  
CHALLENGE Cast Search by  
Portrait比赛第一名



The 20BN-jester

2020, 德国The 20BN-Jester  
Dataset手势识别评测中以  
97.26%准确率斩获冠军。

## 驾驶安全



### 疲劳检测

眼睛+嘴巴开闭度检测



### 分心检测

头姿估计+视线追踪  
路面注视精准率 > 99%  
召回率 > 98%



### 行为检测

打电话、抽烟等检测  
精准率 > 99.5%  
召回率 > 96%



### 高精度率

综合工况准确率 > 99%,  
可完成Yaw±90°, Pitch  
±45°和Roll ±45°方向内  
大角度姿态下的准确检  
测。



### 高可靠性

适应复杂光线环境, 克服  
了驾驶员佩戴眼镜、口  
罩、面部各种遮挡等复杂  
场景的影响。



### 高效低本

纯软方案, 不增加硬件成  
本。支持CPU + NPU 硬  
件加速, DMS标准版  
CPU占用低至8K  
DMIPS。

# 听+看

听和看不是孤立的，很多场景下是相互协同的，可以提高感知效果



## 多模态语音VAD

极大降低干扰  
彻底解决扬声器回声消除残留、音区串扰残留、同音区人声残留干扰影响。

## 多模态语音识别

嘈杂场景识别提升



语音识别效果提升

## 端到端意图识别

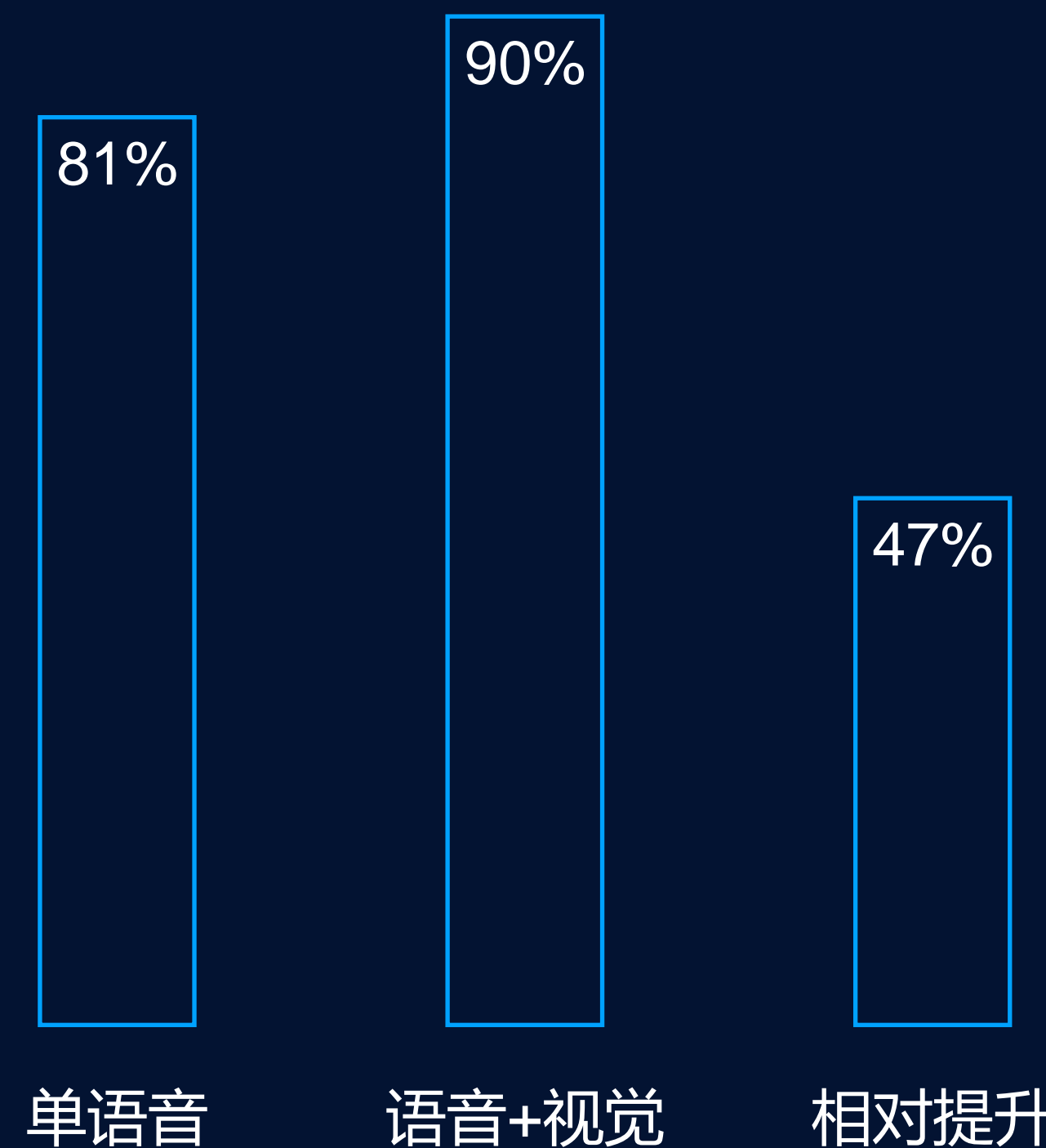
区分人人/人机



全时全场景免唤醒

首创唇音融合免唤醒技术 | 开口直说 一语直达

## 车载高噪环境下多模识别效果



# 听+看

听和看不是孤立的，很多场景下是相互协同的，可以提高感知效果

### 多模态语音技术演示

头姿  
P:7.0  
R:5.0  
Y:6.0

口部遮挡

口罩 遮挡

唇形图

媒体音

樊登读书会-《情绪急救》 播放

环境音

城市噪音-钟声车声 播放

左音区  右音区  全向

停止 设置

多模态语音 (驾驶员)

单模态语音

语音

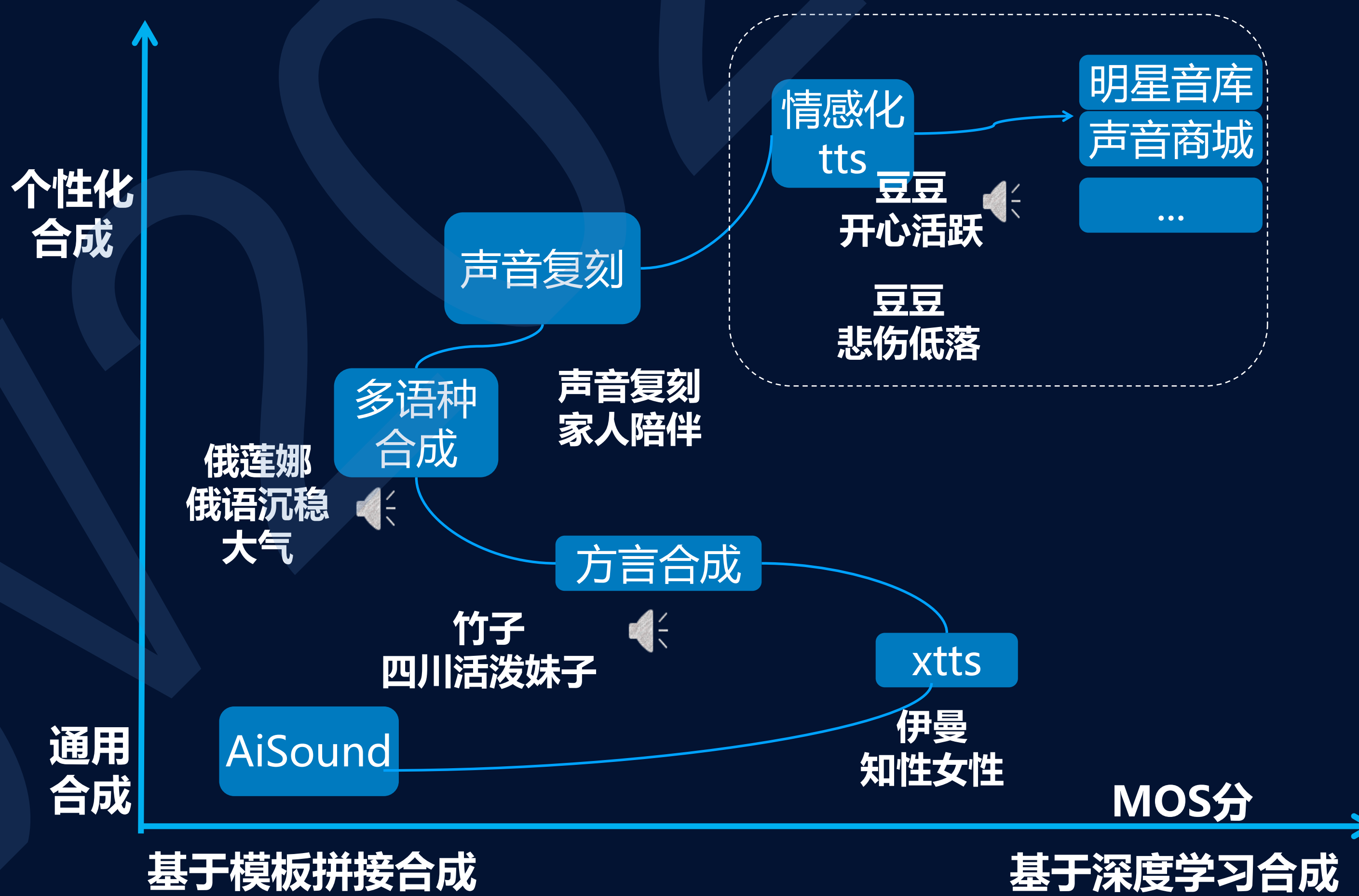
# 说

语音合成超过真人水平，让声音传递信息、温暖、情感

2019 Blizzard Challenge 语音合成连续十四年夺冠



关键指标自然度、相似度第一，自然度4.5MOS创历史新高  
科大讯飞与中科大语音及语言国家工程实验室联合参赛



# 说

## 汽车音响系统还原音乐厅效果



### 汽车音响还原目标

- 频响自然、细节丰富、无音染
- 高频细腻、中频温暖、低频结实有力，收放自如。
- 声场还原准确、舞台宽度合适
- 人声定位精准、真实
- 环境包围丰富、营造现场感
- 左右乘客位置听感接近

未专业处理过的汽车音响：没有舞台感



优秀汽车音响：音乐厅般的自然效果



说

# 讯飞具备全链路的音效综合解决方案能力

车型  
规划设计

内饰  
设计

E/E  
设计

零部件  
开发

调音

整改  
优化

宣传  
引导

量产  
跟踪

持续  
OTA

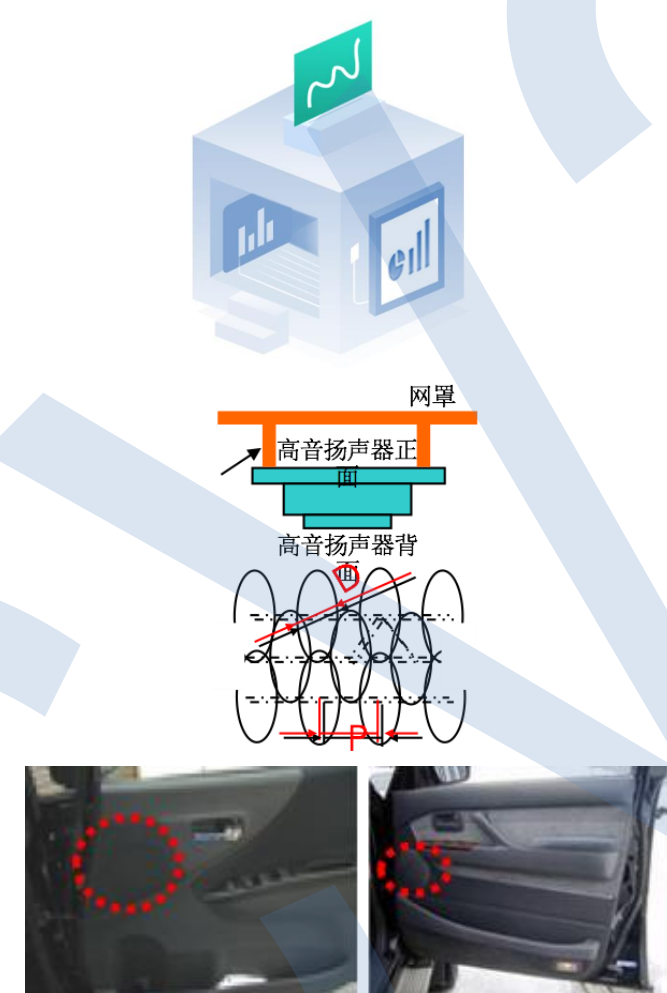
## HU音频参数测试



Output type	Variable Differential Analog
Output signal level voltage	Max 9rms+1dB variable level
Maximum output impedance	≤200Ω
Numbers of Channel	4
Frequency Response	±1 5dB @ 20Hz-20KHz
Audio Source mode balance	±1 5dB
Max Output Impedence	≤2000
S/N	≥85dB @2Vrms
THD+N	≤0.1% @ 0Vrms & 20Hz-20KHz
Channel Separation Ratio	≥75dB @2Vrms
ACC	≥15mA
Mute/Unmute	≥1mA

针对HU的音频输出做  
客观音频参数分析与主观音质分析

## 扬声器安装位置分析



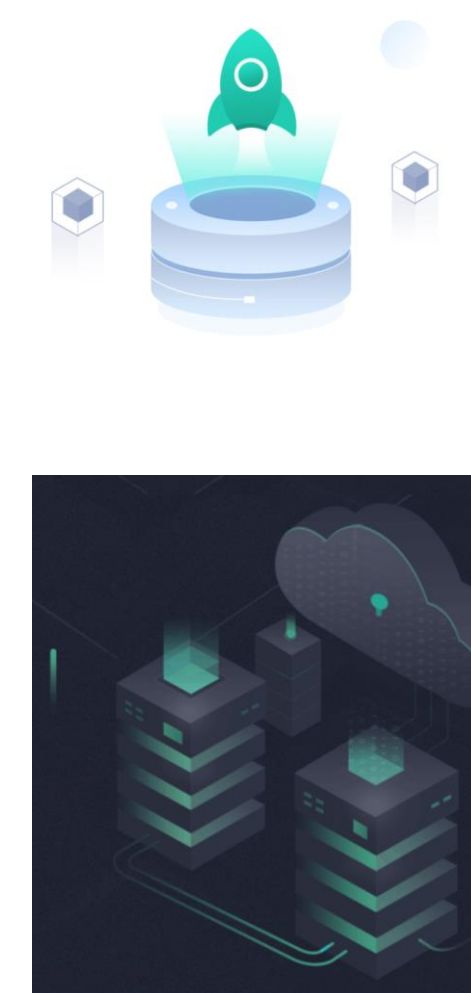
提供整套扬声器安装位置  
与音效分析服务

## 扬声器产品服务



提供整套扬声器产品  
也可以共同改善汽车原有扬声器音质

## 持续OTA服务

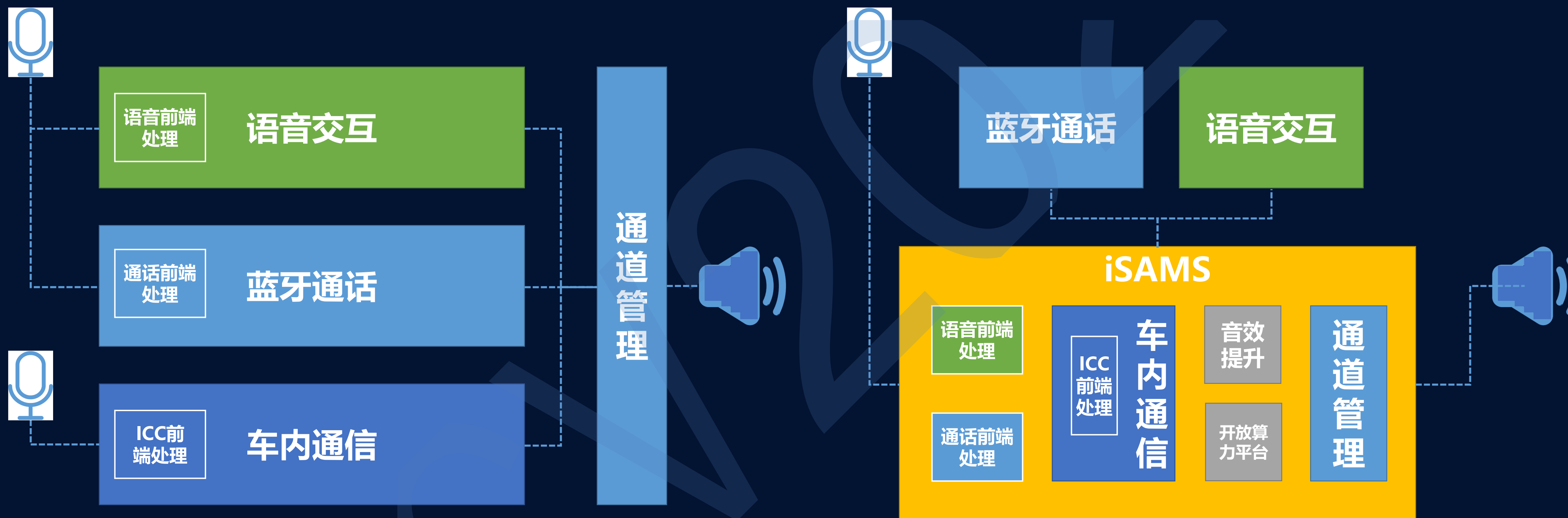


可提供持续的音效迭代优化  
甚至是功能性新增



# 听+说

听和说结合，从听到说全链路管理



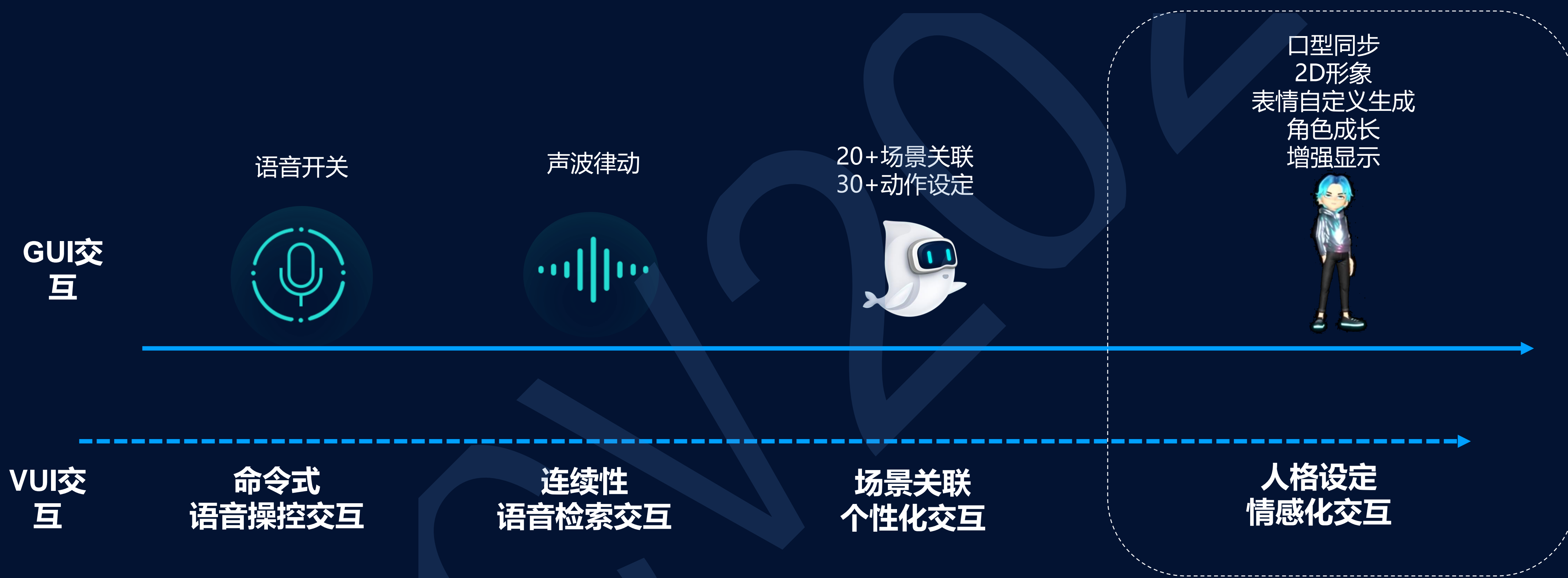
完全分离的前端处理和设备，导致软硬件  
成本高，资源开销大，效果提升难。

综合输入输出管理，以更低的硬件成本提  
供综合解决方案。

(iSAMS = iFLYTEK Speech & Audio Management System)

# 显

让汽车智能助手有颜值，有内涵，有个性多样形象交互彰显每辆车独特气质



# 显

用语音驱动口唇、表情、手势等肢体动作的变化，以达到更自然的人性化的虚拟形象展现



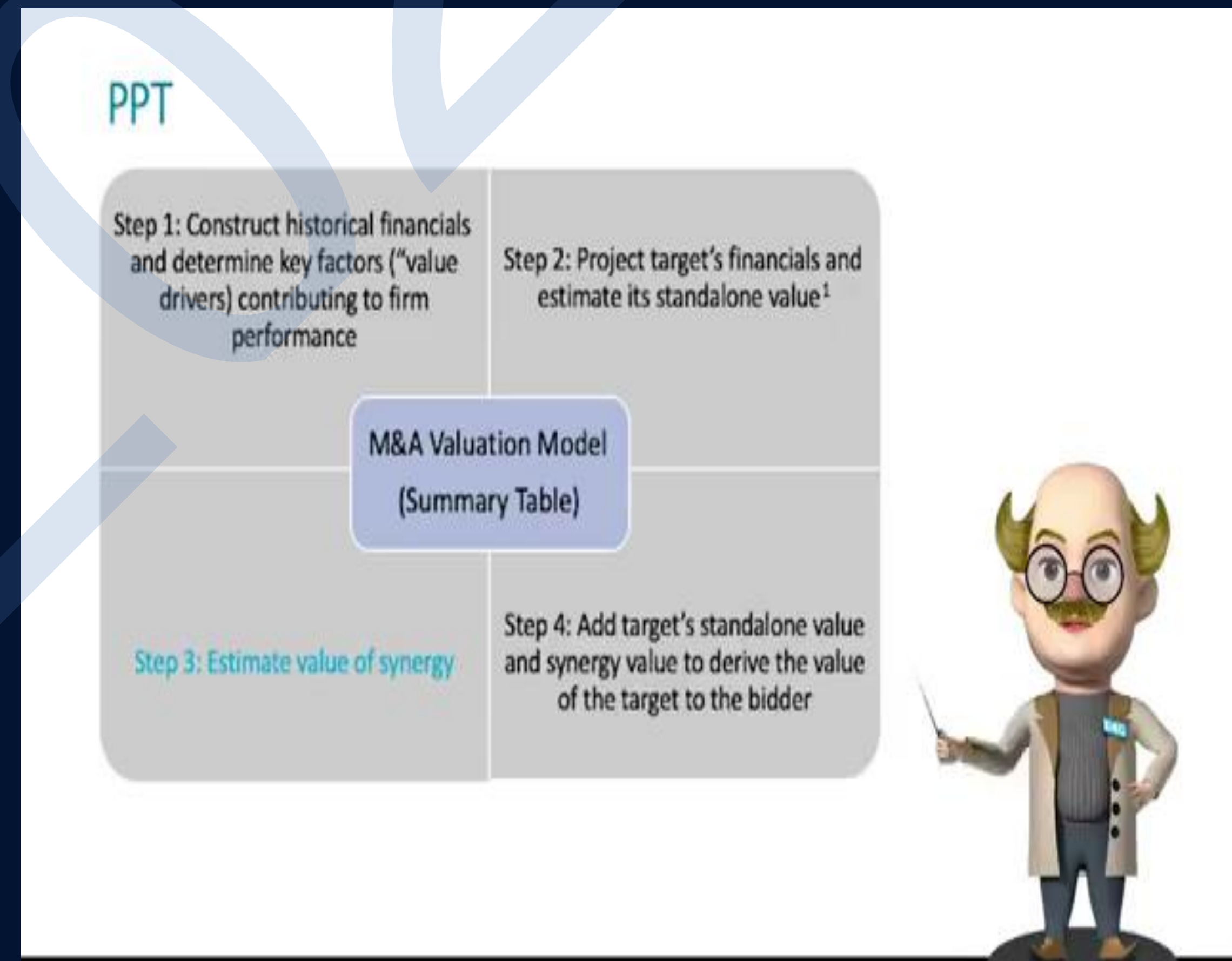
- 从单一模态走向多模态，从语音合成到视频多模态合成
- 从单一维度走向多维度，从口唇动作的合成到表情以及肢体动作的表达

显

用语音驱动口唇、表情、手势等肢体动作的变化，以达到更自然的人性化的虚拟形象展现



交互一体机



TOP虚拟教师

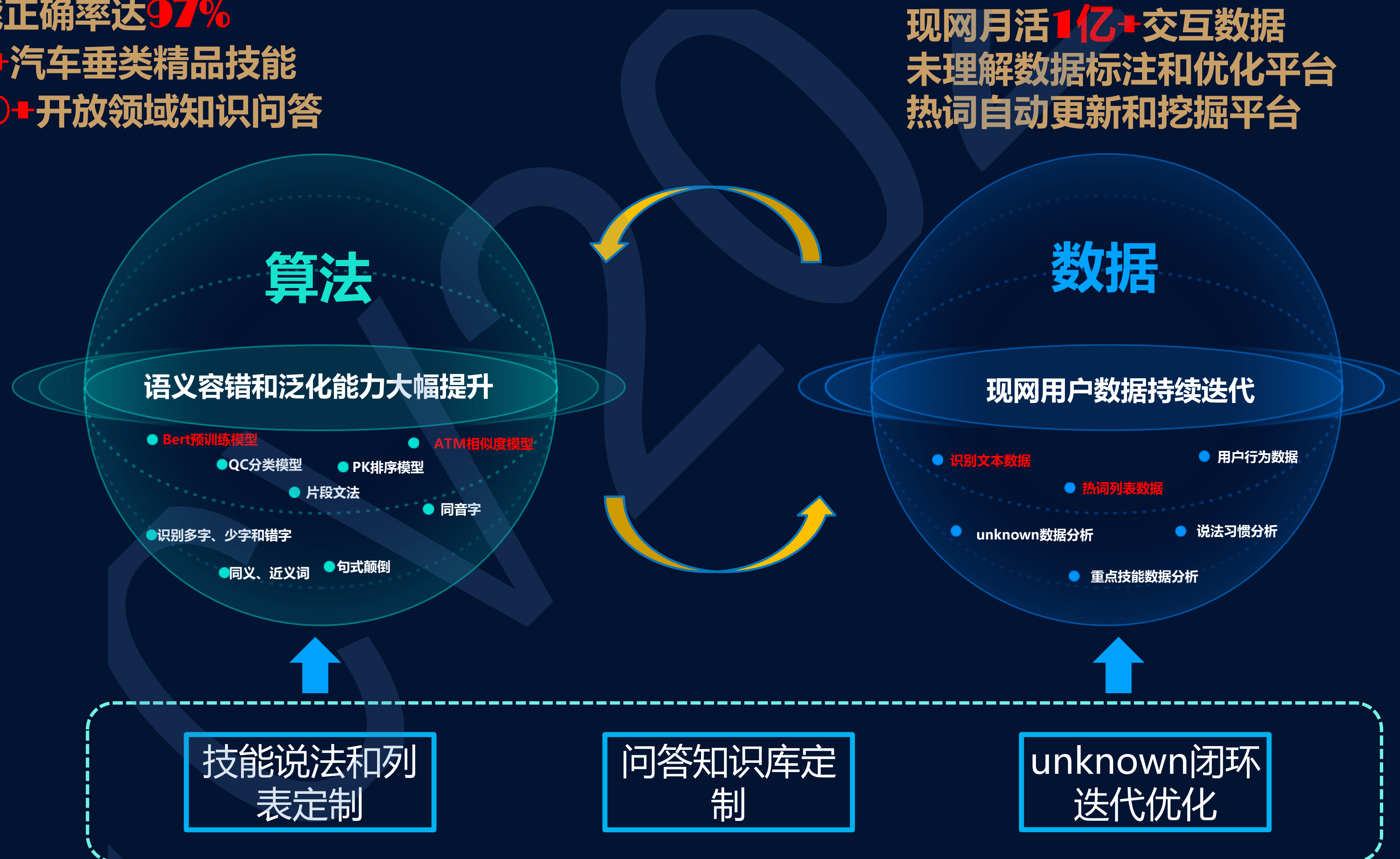
# 交互理解-语义理解

“平台+模型策略”带来语义理解真智能

现网技能正确率达**97%**  
 打造**30+**汽车垂类精品技能  
 支持**100+**开放领域知识问答

现网月活**1亿+**交互数据  
 未理解数据标注和优化平台  
 热词自动更新和挖掘平台

算法+数据双驱动



汽车AI+平台支撑

# 交互理解-知识管理

自然语言理解涉及到人工智能的本质：知识管理

我想上厕所

那边有个肯德基

校长说了，衣服上  
除了校徽，别别别的

他走了一个小时

老子天下第一

晚上“老地方见”

他走有两年了

急的像热锅上的蚂蚁

他走了一公里

## 人为什么能回答这个问题？

机器没有上厕所的需要，更不会有到肯德基上厕所的经验

问一个从没有肯德基的年代穿越到现代的人，同样答不出来

对一段文字的理解不仅仅关乎文字本身

**知识是基础：** 生活场景，背景知识，实际经验.....

# 交互管理-知识图谱

汽车知识图谱为车企打造虚拟化生产力，在用车、销售、维修、研发4大场景

倾力吸收广域知识，打造出拥有知识问答和推理的虚拟机器人



## 用车助手

- 功能使用
- 故障求助
- 维修保养
- 服务推荐



## 销售助手

- 品牌介绍
- 亮点推广
- 竞品对比
- 金融服务



## 维修助手

- 故障搜索
- 故障问诊
- 辅助诊断
- 故障统计



## 设计辅手

- 用户画像
- 设计回归
- 产品力评估
- 性能/配置定义

# 交互理解-交互设计

交互设计取决于不同的人、不同的车、不同的环境、不同的心情，这是一门艺术

打电话给张三

即将呼叫张三-科大讯飞，确认还是取消？

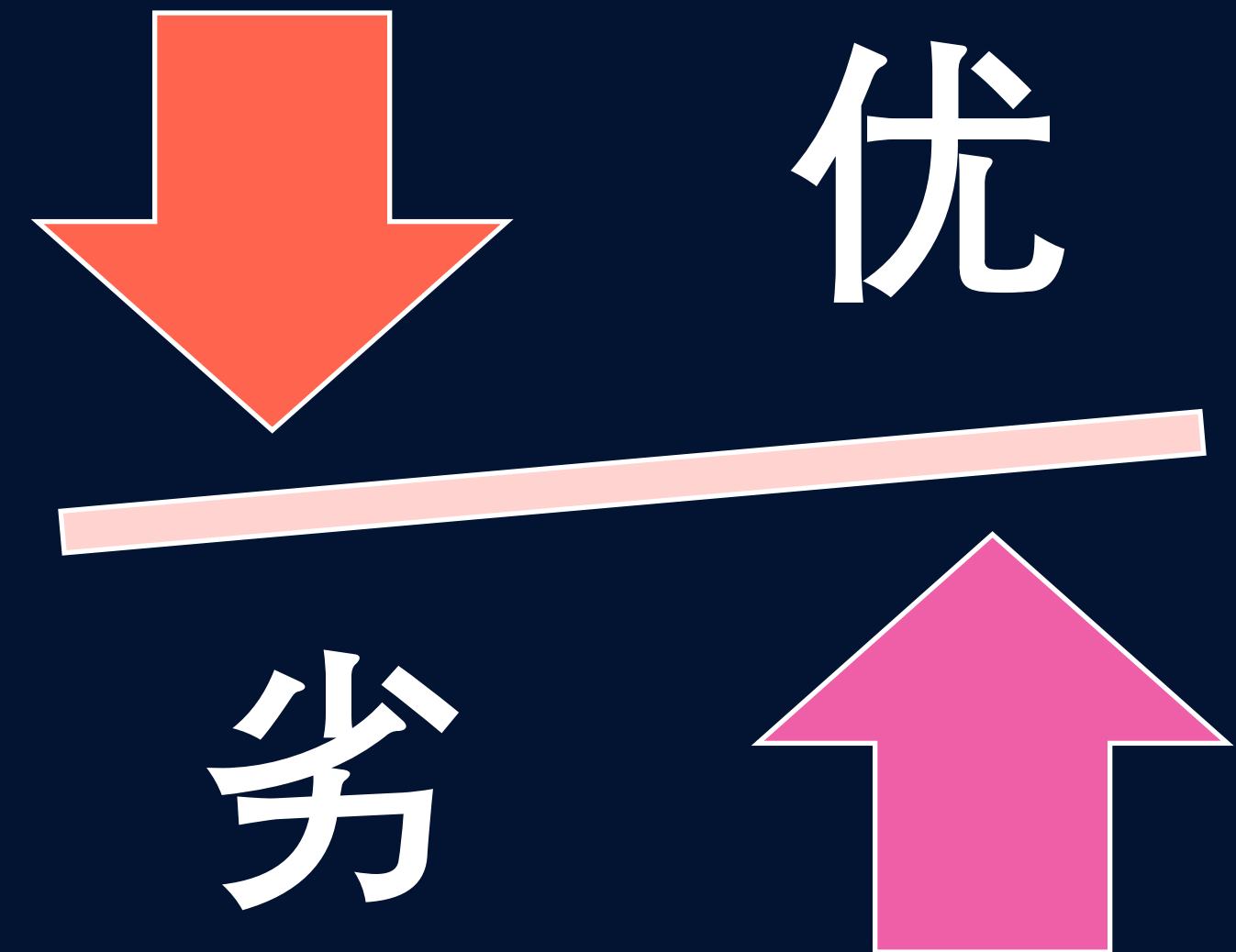
确认

正在为您呼叫张三-科大讯飞，按语音键可以取消

.....

打电话给张三

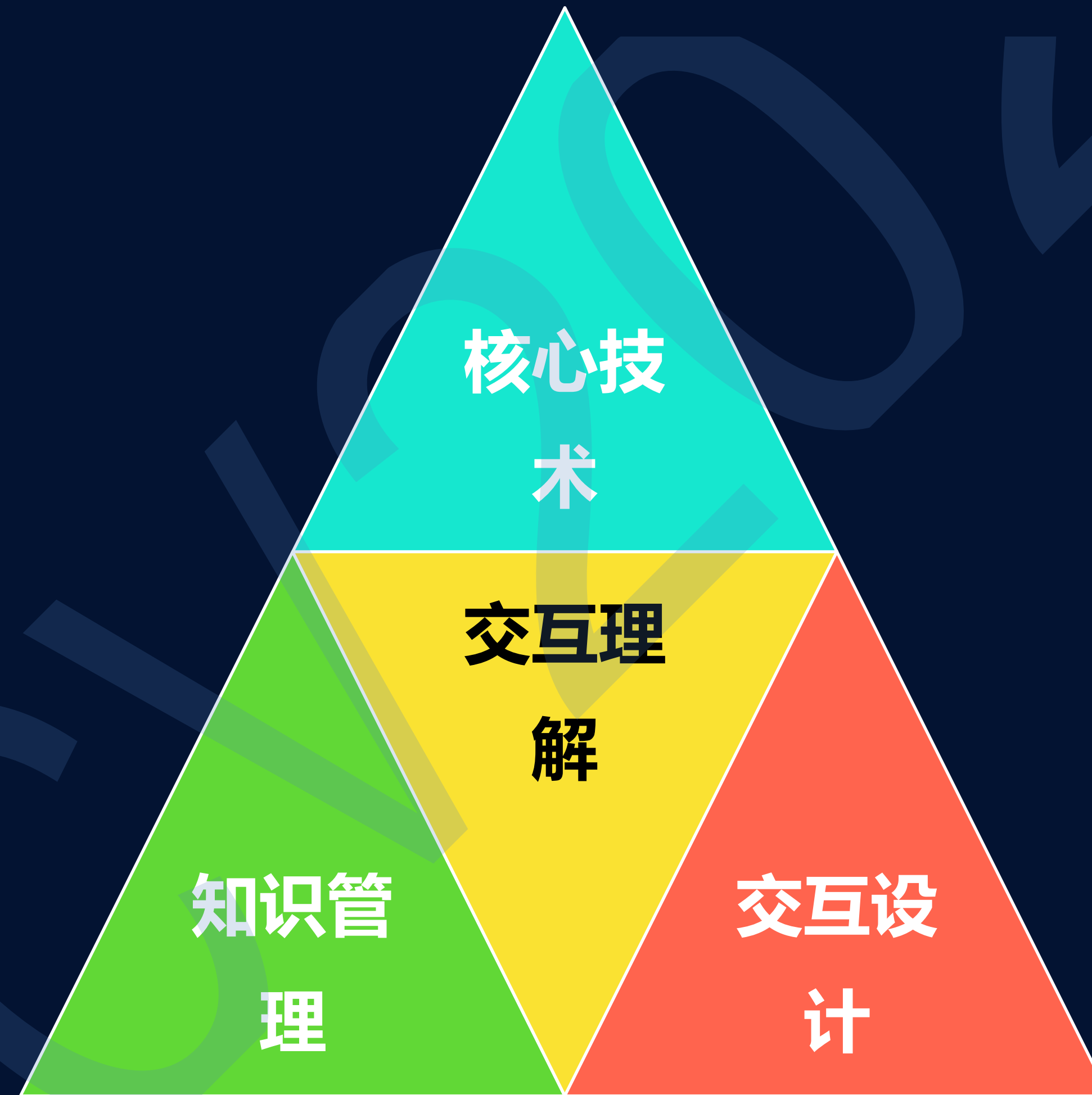
正在呼叫张三





# 交互理解

交互是技术、知识、艺术的综合应用



# 系统性创新全面升级智能网联汽车新体验

## AI赋能听说看显多模态人机交互



# 系统性创新全面升级智能网联汽车新体验

## 科大讯飞智能汽车产品全景图

<b>汽车价值链</b>	研	产	销	服	运
<b>用户价值链</b>	买车	用车	养车	卖车	
<b>解决方案集</b>	智能交互	智能音效	飞鱼系统	汽车AI销服	
<b>汽车云底座</b>	开放生态聚合平台	场景立方	汽车知识大脑	私有化AI能力平台	运营管理平台
<b>汽车AI底座</b>	语音交互	声音管理	机器视觉	知识图谱	多模态融合
<b>汽车工程底座</b>	软硬分离平台	汽车软件工程体系	用户体验设计工具	AI定制工具集	数字营销工具集

# 系统性创新全面升级智能网联汽车新体验

## 科大讯飞智能座舱全栈产品解决方案

人工智能技术持续突破

语音识别

语音合成

图像识别

语音交互

...

飞鱼系统

汽车音效

智能交互

汽车AI销服



用户画像



个性化推荐



数据结构化

建立自主的数据中台

场景和用户体验驱动的产品设计

THANK YOU!

感谢您的聆听!