



科大讯飞，中国语音智能生活的实践者

语音，被各大科技公司认为智能家居竞赛中最具潜力的人机交互方式。

总部设在合肥的科大讯飞不久前发布了AIUI开放平台3.0版本，

这所在中国智能语音技术上占据领先地位的公司，

将以AIUI为技术连接，打造一个智能家居的开放生态，尝试用声音连接的中国各行各业。



“智能语音肯定是要走入未来生活的各个方面。”

当我们抱着智能语音技术和AIUI开放平台究竟能为日常生活带来哪些改变的疑问，走访科大讯飞位于合肥的总部时，现担任科大讯飞AIUI开放平台负责人的赵艳军赵艳军首先指出：从2010到2016年底，整个AIUI平台拥有20多万开发者，但到2017年，这个数量增加到50多万，现在则已经突破80万了，“大家主要都奔着万物互联和智能硬件这波浪潮。但从真正落地的情况来看，发展快速的还是在强刚需的应用场景，比如车载、电视、音箱这几个领域。现在市场上比较多的都是一些能真正解决用户实际问题的产品，还远远没有达到能够让每一个家庭都装上智能家居的阶段，我们现在更多在尝试一些场景化的方案。”

而这些运用了讯飞智能语音技术的场景化解决方案，都陈列于科大讯飞位于大楼边角空间的201室，这里是科大讯飞自主研发成果的展示厅，每天从早到晚都接受着无数团队和个人的预约。

进入展厅，悬挂在大门一侧的两块显示屏上实时显示着当前AIUI开放平台参与的开发者总人数、分布地区和产品数量等信息，它旁边的另一块屏幕上则正播放着一段特朗普的语音合成视频，在这个视频中，特朗普先用英文，再用中文说道，“科大讯飞真是太棒了”，但事实上，特朗普并不会讲中文，这是该公司利用其技术玩的一个宣传小花招；往前走，被放置在桌上的一个小小黑匣子，是刘庆峰曾在两会期间演示过的讯飞晓译翻译机，其准确率较高的实时双向翻译功能赢得了不少国外出行者的喜爱，两个造型十分可爱，颇为吸引眼球的阿尔法大蛋、小蛋是科大讯飞推出的儿童智能陪护型机器人，它们正闪烁着笑脸等待用户的唤醒；桌上最为人熟悉的另一款产品则是2015年科大讯飞联合京东推出的叮咚智能音箱，这也是中国第一款自主研发的智能语音音箱；另外两个非常引人注目的是站立于展厅中的两个与人等高的机器人，其中一台名为“小途”的星途机器人能够为人们政务服务，而另一个机器人“晓医”则在不久前刚刚通过了国家医疗执照考试，取得了超出标准分96分的成绩，成为历史上第一个取得这种成就的机器人，可以作为“智医助理”帮助医生捕捉和分析患者信息，提供人工智能辅助诊疗；此外，一辆受到众多男性参观者喜爱的汽车驾驶舱则展示了科大讯飞语音技术在车载系统中的运用，智能车载是如今智能语音技术被广泛应用的一个领域，奥迪、宝马、奔驰、通用、上汽和奇瑞等国内外汽车制造厂商已经和科大讯飞进行合作，将为中国用户提供便捷的智能服务。

1
洛杉矶、纽约都有店铺，并且我们打算在接下来的两个月一直在说我一直是一个梦想家和信徒。

1
洛杉矶、纽约都有店铺，并且我们打算在接下来的两个月一直在说我一直是一个梦想家和信徒。



从中文语音智能技术特有的难点出发

从一个大学生创业项目，到现在占据中国智能语音市场70%份额的上市公司，随着公司业务的逐渐扩大，科大讯飞位于合肥的总部已经聚集了九千多名员工，为了满足进一步的需求，几座新大楼的建设工程正在进行，而周边，也隐隐以此为中心聚集着新的商圈，科大讯飞正逐步实现着印刻在各个大楼醒目的那句口号“让世界聆听我们的声音”，试图将智能语音从合肥这座甚少被提及的中部城市传向世界，打造自己的声音硅谷。

“我们在定义AIUI时希望实现一个理想的交互环境，也就是人和机器沟通的无障碍，可以通过语音、图象、手势等方式模拟人与人之间真实的沟通方式，进行持续、自然和双向的交流。”赵艳军笑道。

中文语音智能技术，有着自己独特的难点，首先不同于英语标准的语法规定，中文同样意思的表达有着非常多不同的句式，比如“你什么时候下班”和“你工作到今天几点”，语法规则完全不同，但却表达着相同的意思。其次，对于很多用户而言，家里的父母和老人，他们带着各地口音的普通话和方言更是成为家庭智能设备中语音识别的难点。赵艳军介绍道，到今年，科大讯飞普通话的识别率已经可以达到98%，而这个数据在2010年刚发布讯飞开放平台时仅为60%至70%，经过不断的迭代训练以及优化，去年达到97%。“越到后面要提高识别率就越困难”，赵艳军解释道，将识别率从70%提高至98%，其相对错误率需要下降30%，但如果要从98%提高到99%，它的相对错误率则要下降50%以上。

针对方言问题，AIUI开放平台在3.0新版本中为用户开放了23种方言的识别技术，其中超过一半的识别率可以达到90%以上。训练机器进行语音识别时，某种语言覆盖的用户人群所产生的数据样本越大，优化迭代速度就越快，这就是为什么每天超过2亿活跃用户的普通话，其识别率能够快速提高。方言中，像粤语和四川话这些人群覆盖度比较大，识别率也会相对较高。

走入家庭：小的体型，快的反应速度、海量内容数据

同时，赵艳军认为在设计和考虑家庭用智能设备时，“重型的或者大型的机器人走入家庭里非常难。比如像日本的Pepper这种中大型的机器人现在基本上都是针对于行业服务，放在家庭场景中，都是像音箱或者桌面型的小型机器人才有可能。”

不难看出，其中颇受家庭欢迎的阿尔法蛋便是其典型案例。这款儿童教育陪伴智能机器人按体型分为“大蛋”和“小蛋”两款，大蛋也不大，才32厘米高，小蛋更小，只有17厘米高，两款都有着圆滚滚的可爱造型，并搭载了讯飞AIUI智能系统，拥有“类人脑”，有着很强的理解能力和表达能力。

“现在阿尔法蛋要完成一个接受命令并作出反应的过程，大概需要800毫秒的时间。”赵艳军表示这个速度处于业内领先水平，快速的识别能力令阿尔法蛋能够对对话进行快速反应，更加模拟真人对话场景。“在语音识别翻译的场景中，只涉及一个语音识别加翻译的过程，也就是识别出文字后就可以即时翻译，显示到屏幕上。但是对于阿尔法蛋，它需要理解你的意思，判断你什么时候说话说完，需要去做断句。断句要么是你说完话后停顿，要么是由机器检测这段话是不是完整。不能我说话一半时候就给我回复，这时候我还没完全表达我的意思呢。所以断句是一个必需的过程，可能需要500毫秒，而整个语音识别加上语义理解也就是约300毫秒，整个交互完成则合计是800毫秒，在1秒以内。”

并且，赵艳军还分享道，基于AIUI开放平台背后的海量数据可以看出，“用户最常用第一是内容点播”。比如在阿尔法蛋中有音乐、视频、舞蹈、评书、诗词国学、课程教材、百科知识等内容，所以AIUI开放平台也与百度音乐、喜马拉雅FM和蜻蜓FM等内容平台都达成了合作，令用户可以通过AIUI获取这些平台的海量内容点播歌曲，听相声等。“智能家居在市场上还不够普及的主要原因也是因为内容及其应用场景还不够丰富，如果不能持续的更新或加以扩充，重复的交互会导致用户对它的新鲜感越来越低。”赵艳军指出。

越发自然和情感化的家庭语音服务

在电影《Her》中，主人公西奥多在经历了婚姻的失败后，接触了最新款人工智能系统OS1的化身萨曼莎，由斯嘉丽·约翰逊配音的萨曼莎拥有着迷人的声线，她时而风趣幽默、时而温柔体贴，在两人不断的交流过程中，西奥多竟爱上了这位虚拟女友。在现实生活中，我们看到，林志玲“性感语音”发布后令高德地图的下载量超过以往6倍，而郭德纲“段子语音”一经推出便帮助高德地图的新增用户速率提升230%。阿尔法蛋则占据着国内儿童类智能玩具销售前三，它反应灵敏的童声受到了孩子和家长的喜爱。

“我们尝试令语音合成的播报更加自然和有情感，针对不同的语音智能产品，我们会设定一个虚拟IP，它有自己一个人设，规定好它的性别和年龄等特征，再根据这个人设去设置相应的语音以及符合人设的回复话术，让用户产生亲近感。”赵艳军介绍道，AIUI3.0开放平台平台中的声纹技术，可以分析对话者的声音特征，比如小孩有小孩的声音特征，女主人和男主人又有另外的声音特征。这样，在家庭成员使用家庭智能产品时，它就能够判断出使用者是哪位家庭成员，从而作出符合身份的个性化互动，例如对孩子回复萌萌的童声，对成年男性回复甜美的女声，对成年女性则回复磁性的男声。“我们曾给深圳一家做陪伴型机器人的公司提供过技术支持，它的合成音库里面有中性、高兴、悲伤等不同情感的声音，根据对用户不同话语的理解，提供不同情感的声音回复。”

想象在未来的家庭生活中，你无须带钥匙可以通“芝麻开门”打开大门；智能电视一边为你提供节目检索推荐服务，一边用妈妈的声音提醒你不要长时间观看；智能音箱中传来女朋友温柔的声音喊你起床，提醒你今天下雨记得要打伞；儿童智能玩具中用到爸爸的声音给孩子们讲故事、背诵古诗文；智能家庭医师助理使用着奶奶的声音，像一个家长似的告诉你发烧了在家休息吃点什么药；车载导航系统中妻子、孩子的声音提醒你小心驾驶，注意安全……

这些未来场景从未离我们如此之近。