

讯飞听见智能会议系统是基于科大讯飞业界领先的语音识别技术所研发，可满足公司发布会、电视节目直播、大型会议、培训等一系列需要进行实时文字转写的场景。

本文档主要介绍了讯飞听见智能会议系统的研发背景、设计理念、产品框架及功能等，以帮助读者对讯飞听见智能会议系统有快速的了解。

本文档适用于需要对讯飞听见智能会议系统进一步了解的用户，如还有其他疑问可联系我们的技术支持人员，我们会进一步为您解答。

产品研发背景

3.1 市场需求

进入 21 世纪以来，人类逐渐进入多媒体信息时代，当今大众传媒主要有互联网、电视、手机等，多媒体信息逐渐成为生活中不可缺少的重要部分。作为传媒载体的三要素，声音、文字、图像三者的结合即现场直播是人们最直接传递信息和理解的方式，而在发布会、大型会议、电视直播、教育培训等场景中体现尤为明显。

发布会的目的是为了传递团队或者企业的产品、文化、品牌，最好的传递方式就是用视频直播，每年发布会超过 400 万次，总时长在 1600 万小时以上。

电视直播主要为体育比赛、电子竞技等直播，为广大体育、电子竞技等爱好者提供及时、快捷的信息获取渠道。每年体育比赛、电子竞技举办次数超过 100 万场。

截止 2014 年中国的教育培训机构总量约为 14 万家，视频教学已经成为常见的教学方式，为学习人员提供基础教育、语言类、职业技术等辅导，2014 年总培训辅导时长超过 28000 小时。

另外，任何需要以声音、图像、文字作为载体传播的讯息，如采访、会议、法律纠纷、医生问诊等，都需要有一套可以提供实时上屏的产品系统。

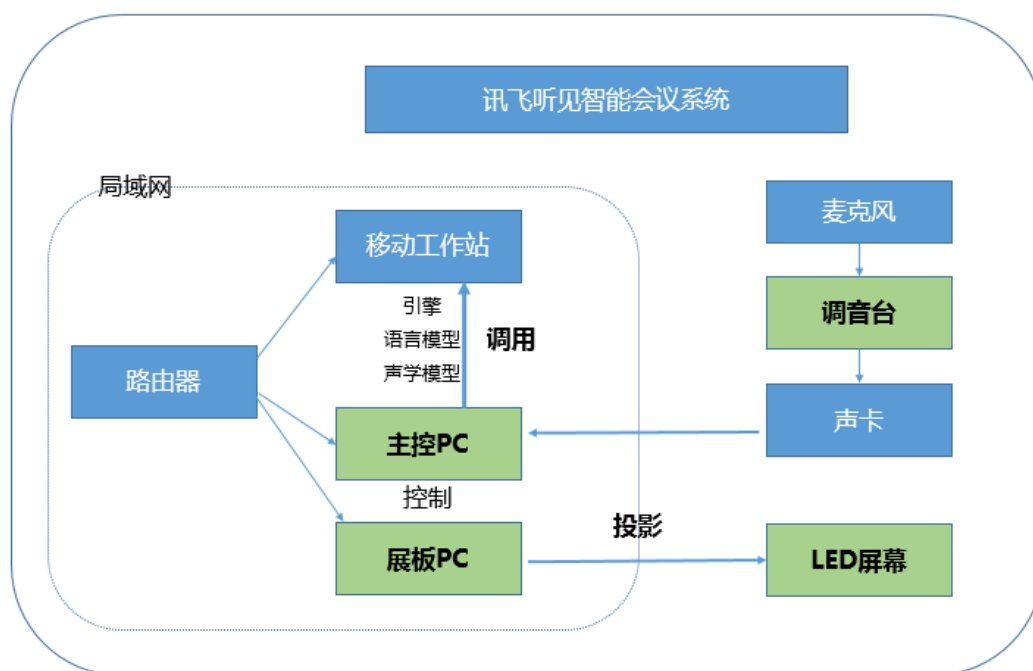
3.2 市场现状

为了在上述环境中实现现场直播，传统的解决方法是：在现场录制过程中，配有专业的速记团队对音频进行文字转写和校对，转写完成后再和视频或者图文匹配，完成后进行发布，实现现场直播，这样解决的方法具有以下点局限性：

- 消息滞后性，因为视频是通过人工后期转写后发布，与现场存在一定时间差。
- 信息获取缺乏效率，相比较而言，人类获取文字信息的速度要快于获取语音信息，且文字信息可以跳过一些无需关注的内容。
- 后续整理耗费资源，进行现场视频直播时，需耗费人力对转写文字和视频进行时间戳校对形成字幕。

产品概述

4.1 产品架构



4.2 产品功能详述

1.VAD 检测

进行准确的 VAD 端点检测，可避免背景噪音和人声噪音的误识别，使识别结果更加准确。

2.语音识别

利用科大讯飞的先进语音转写技术，集成说话人自适应、文本后处理、文本顺滑、异常检测等技术，将语音流逐字实时转换成文字，并实时推送到客户端。

3.语音后处理

进行标点、数字、英文等后处理，使识别结果更加准确。

4.实时上屏

将转写结果实时投影到现场的大屏幕上，进行实时字幕直播。

4.3 产品包含组件

序号	名称	功能
1	移动工作站	识别引擎及模型资源部署
2	主控客户端	控制录音和展板端
3	展板客户端	现场实时上屏展示
4	路由器	局域网
5	声卡	音频处理
6	鹅颈麦克风	收音设备
7	加密狗	保障数据和服务安全

4.4 简单部署流程

1. 引擎及语言模型声学模型资源部署、配置；
2. 拉通测试；
3. 打开主控端和展板端 PC，进行录音及投影操作；

4.5 产品实现页面展示

1. 主控端

主要功能：控制录音开始于暂停，控制展板页面切换和清屏，可选择模型，并可将转写结果

保存和导出。



2. 展板端

主要功能：展板端接收主控台的音频转写结果，将通过投影仪将转写结果展示在大屏幕上，实现现场实时上屏，并可检测音量大小。

