

基于云与语音识别的手机智能 通话系统——SkyTalk

施天麟 / 上海市上海中学
指导教师 毛黎莉 王颂赞

◎ 简介

本项目提出一种手机智能通话系统 (SkyTalk)，能够方便人们在双方通话中简单、快捷、实时地通过语音获取信息、完成用户指定任务，从而避免大量复杂、耗时的操作。在拨出 SkyTalk 智能电话时，手机中的基站编号、GPS 坐标、移动速度、通讯录等数据会上传至服务器，并由服务器建立智能通话。在通话的过程中，服务器作为第三方通过语音识别和语音合成响应用户的语音请求，可以发送通讯录中的号码、说出用户的具体位置、计算到达对方的时间、播送天气预报、进行实时语音翻译等。SkyTalk 通过云计算提供反向地址解析、获取互联网信息，不仅简化了服务器构架，还能为用户提供更多精彩应用。

移动电话拉近了人与人之间的距离。电话的发展，从贝尔到摩托罗拉，再到今天的 4G，技术不断革新，但人们的通话方式却始终没有改变。SkyTalk 手机智能通话系统通过云资源、语音识别与合成，可以使用户在通话过程中，不必挂断手机、亲自编辑短信、查阅信息或打开 GPS 定位，而只需说出问题，SkyTalk 便能够用语音加

以回答。总之，SkyTalk 是信息化、智能化社会下人们通话方式的革新。

◎ 项目原理

➤ 基本构架

SkyTalk 通话系统在电话语音通道之外建立一条互联网“智能”通道。手机上的 GPS 位置、通讯录、个人偏好等信息通过这条通道传输到中间服务器，在“云”的辅助下对之进行处理与交换。而中间服务器通过语音识别 (SR) 和文本到语音转换 (TTS) 在电话通道上与用户双方进行交互。这样，一方面可以利用互联网通道交换基本数据，另一方面又利用现有电话通道进行语音传输，从而避免产生大量的语音数据包。更为详细的 SkyTalk 的构架 (如图 1) 分为 3 个部分：手机客户端 (SkyTalk Client)、服务器 (SkyTalk Server) 和云 (Cloud)。

客户端 (SkyTalk Client) 一方面取代原有电话的拨号界面，另一方面通过手机平台的 API 获取手机的各种信息 (GPS、电子罗盘、通讯录等)，并通过 UI 交互获取用户的定制偏好及信息。当用户拨打一个电话时，客户端 (SkyTalk Client) 向服务器

(SkyTalk Server) 发送一个请求，并附上原先获取的信息。

服务器 (SkyTalk Server) 收到请求后，建立对等手机和服务器的三方通话 (Phone-Server-Phone)。服务器作为第三方根据语音识别引擎分析用户的语音命令，并结合用户的手机信息进行相应处理。

云端 (Cloud) 集成了用户可能需要的信息。这片“云”是现有的互联网服务集合。

➤ 基本流程

SkyTalk 基本工作流程见图 2。

本项目的客户端基于 Qt (cute) 开发。由于涉及访问手机系统，用 Java 平台实现是不合适的。Java 虽然提供了此方面的 JSR 开发包，但由于不同厂商支持的 JSR 不同，实现也不同，因此兼容性很差。于是，本项目采用效率更高、操作系统支持更强的 Qt (C++) 实现。

◎ 功能实现

➤ 基于 Skype 的通话系统

一方面为了保障正常通话质量，另一方面为了使服务器的语音介入成为可能。本项目 (SkyTalk) 采用 Skype 构

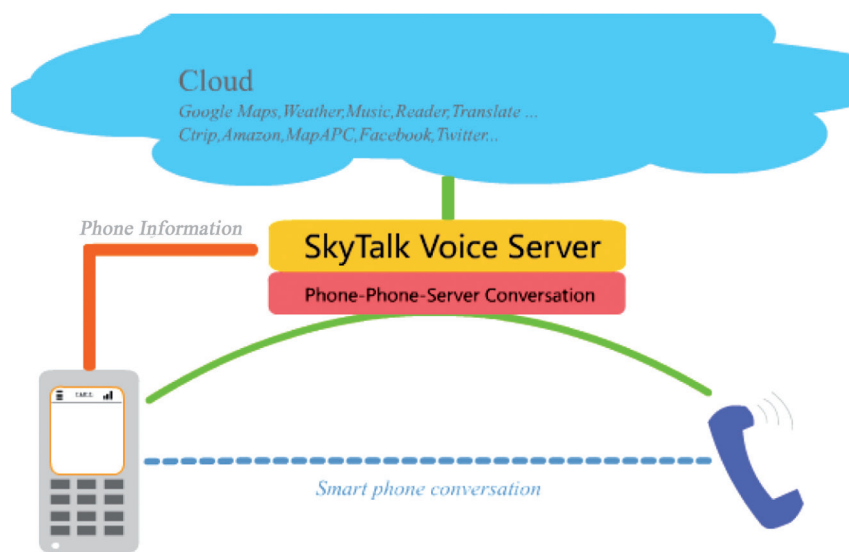


图 1 基本构架

建通话系统。Skype 在 Windows 平台上为开发者提供了 COM 接口。本项目利用 COM 接口与 Skype 进行通信。

➤ 音频数据定向传输——语音管道

本项目通过虚拟硬件的驱动实现语音定向传输。

➤ 语音识别与合成

为了更好地实现语音功能，本项目基于微软 SAPI 5.4 语音引擎构建。这一引擎在 Windows 7 和 Windows Server 2008 中内置。

➤ 已实现的几项具体功能

- ① 在通话中查询当地天气。
- ② 在通话中查询当地时间。
- ③ 在通话中提供地理位置服务（包括通话者的位置，与另一方的距离和到达对方的时间）。
- ④ 在通话中说出通讯录中某人的号码。
- ⑤ 在通话中发送通讯录中某人的号码。
- ⑥ 在通话中播放音乐。

⑦ 在通话中分享 Google Reader。

⑧ 通话中的实时翻译。

◎ 创新点

◆ 本项目 (SkyTalk) 使用 Skype 进行语音传输功能的实现，从而能够利用现有移动通讯网传输语音数据。这不仅大大减少了语音传输的网络消耗，也使得客户端可以对任何移动手机、固定终端拨打智能电话。

◆ SkyTalk 能够根据用户的需求在电话语音通话中提供实时的语音信息服务。这使得用户在通话的过程中能够获得互联网资讯、交换手机信息、解析地理位置、分享各类数据，而不需要事先准备或挂断电话后进行操作。因此用户的交谈、讨论得到大大的简化。

◆ 通话的过程中进行其他操作（如翻阅通讯录、用计算机上网查找信息）并不实际。SkyTalk 利用语音技术解决这一问题。语音识别技术识别用户的命令请求，而语音合成返回服务器的

响应。

◆ 通过互联网上的云资源，SkyTalk Server 能够获取各类计算功能和各类信息，如反向地址解析与天气资讯。云资源大大拓展了 SkyTalk Server 的能力，使其能够提供各类通话中的语音服务。

◆ SkyTalk 能够通过利用用户的个人标志进行个性化智能服务。这样做有 3 个优点：① 可以加载相应的语音识别配置文件，达到更高的识别准确率。② 可以访问用户的互联网或私有云的数据，从而能在通话中利用这些信息。③ 解析用户的个人偏好语言。

◆ SkyTalk 对用户通讯录、用户地理位置信息的存储和传输进行了加密，以保障用户隐私及安全。

◎ 展望

SkyTalk 在功能上和性能上都有很多可以改进的地方。通过改进与扩展，本项目将可以实现更多的应用，达到更

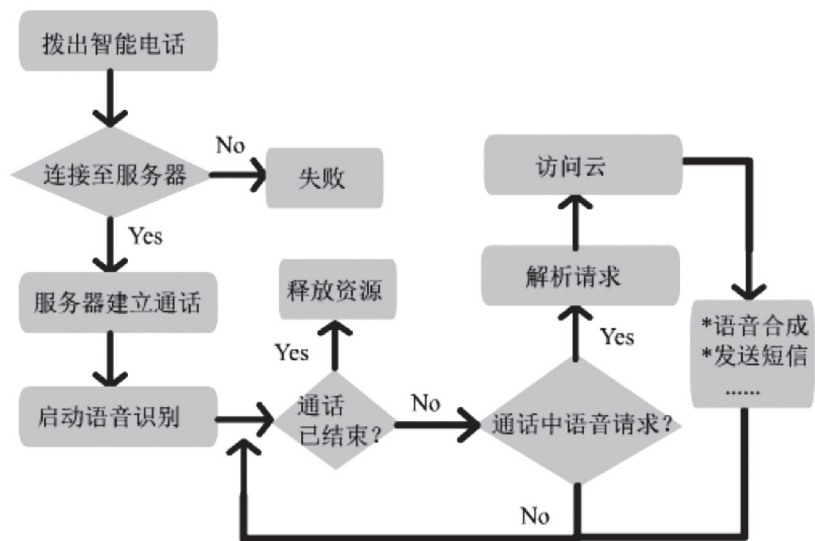


图2 基本流程

好的性能和用户体验。这些改进与扩展包括：

◆ 更强的兼容性。目前手机客户端（SkyTalk Client）基于诺基亚平台实现。将来，可以为不同的手机平台实现其客户端，增加 SkyTalk 系统的应用范围。

◆ 更高的质量与性能。在实验中，本项目有多处可以改进之处。通过整合服务器资源，SkyTalk 可以减少拨出电话时间、减少延迟。通过优化服务器框架、选择更好的云资源，SkyTalk 可以提高响应速度。通过采用更好的语音识别算法，SkyTalk 可以增强识别率。通过这些改进，SkyTalk 可以达到更高的性能。

◆ 更安全的个人标志。尽管 SkyTalk 已经通过较难伪造的手机号来作为用户的个人标志，但不法分子仍然可以通过某些技术手段达到伪造的目的。为了保护个人隐私安全，SkyTalk 可以采用声纹识别进行更强的个人信息认证。声纹和指纹一样，是每个人的生

物特征。声纹从个人语音声波中提取，不易伪造。如果结合原先的手机号验证，那么将大大增强 SkyTalk 的安全性。

◆ 开放式语音平台。这一改进将允许第三方开发平台的插件，将大大增强 SkyTalk 的应用性。因此，本项目完全可以发展成一个开放的语音应用平台，使各种 web 网站或服务都可以通过语音应用的方式出现在这个平台上，为交谈中的人们带来更多的便利。

◆ 面向个人的应用——更智能的 IVR。IVR（Interactive Voice Response）在电话系统中的应用相当广泛。利用 SkyTalk，则可以进一步改进 IVR 服务系统。传统的 IVR 基于企业数据库，旨在代替电话接线员，向企业的用户提供顾客咨询服务，或是提供电话点餐、点歌。改进后的 SkyTalk IVR 则基于云技术和用户个人标志，可以将电话和用户的互联网生活紧密相连。例如，用户可以像听收音机一样，收听自己 Google Reader 的最新资讯；可以收听个人电子邮箱中的最新邮件

还可以利用语音远程操纵自己的计算机——只需说出文档的名字，就可以将文档提取、传输到手机。对于普通用户而言，SkyTalk IVR 提供了一种全新的方式去访问互联网。而对于残疾人士而言，则可以通过它来享受互联网的乐趣。

该项目获得第 26 届全国青少年科技创新大赛创新成果竞赛项目中学组计算机科学一等奖。

专家评语

项目提出和实现了一种手机智能通话系统 SkyTalk，内容涉及云资源信息服务、语音识别与合成等技术，方便了人们获取实时语音服务信息。■