

Nuance——世界领先的语音识别技术

一、 概述

近年来语音识别技术越来越广泛地用于商业。每天，千千万万的人通过拨打运行语音识别、语言理解和声纹鉴别软件的电话，进行出游预订、股票交易、与其它通讯媒体、企业和互联网系统进行交往等活动。

基于电话的语音识别软件容纳了丰富的词汇，不需要对用户进行培训，可以支持任何类型的电话。现行的语音系统支持口语中的常用短语，如“我想查明天北京的天气怎样？”

语音识别由某一网络服务器而不是电话本身进行处理。这种配置使系统充满生机，同时还可以升级，以支持数百万的用户，而且还能识别用户对同一信息所提要求可能使用的所有途径。用户还可以在已经获得想要的信息时进行打断(中断系统提示)。

对于任何一种新的服务来说，如果想成为一种主流，使用灵活、方便是必备的两个条件。许多公司都在建设语音驱动服务，因为对于任何年龄、任何生活方式的人们，语音识别都不失为一种直观的方法，使他们可以从任何一部电话上获取想要的信息，并与感兴趣的人们交流。

二、 博雅思语音识别解决方案

博雅思引进并在其自助语音系统 Edify 中产品集成了 Nuance 的语音识别技术。Nuance 是产品语音商务联盟的发起成员，并领导创建了语音应用程序开发的开放标准世界领先的 Nuance 公司语音识别技术。美国航空、Bell Atlantic、Charles Schwab、家庭购物网络、Lloyds TSB、Sears、UPS 等大公司使用 Nuance 的软件为客户提供更好的服务，同时也大幅度降低了成本。

1. 方案特色:

- **海量词汇、独立于讲话者的健壮识别功能**
系统能可靠地对多种语言进行大词汇量的识别，并可提供识别结果的置信度。该系统对商业上使用的大量词汇提供最准确的语音识别技术，在市场上具有最高的准确率。生产中的应用程序经测试，准确性超过96%。
- **内建的自然语言理解**
通过Nuance系统可以开发自然语言理解系统，它以句子为输入，返回句子意义的解释性表达。应用程序可以根据用户的请求采取相应的动作。系统也提供基于档次的置信评分，它能更加接近地判别可能准确（或不准确）识别的短语各部分。然后可更加自然和有效地修改应用程序，处理错误检查或重新提示。
- **基于主机的客户/服务机结构**

系统基于开放式客户/服务机结构，特别为大型应用程序所需的健壮性和可伸缩性而设计。呼叫者的讲话由客户端收集，而识别和鉴别处理的负载被平均分配到网络上的多个分开的服务器上。

➤ **集成的讲话人校验**

声纹鉴别使用户可通过其声音的生物学统计特征进行校验。允许同时进行讲话识别和用户甄别。系统可对多种讲话进行校验，包括自定义口令、数字和应用程序命令。系统检查要求的话是否已讲，同时根据以前保存的音纹对用户进行辨别。其所提供的安全性可与指纹媲美，而无需特殊的设备。在任何时候，用户所需要一部电话，依靠独特的语音便可被安全地鉴别。

➤ **动态语法**

动态语法是由运行的应用程序创建和修改的语法。这在应用程序词汇必须在运行时才能完全确定的情况下是必不可少的。

➤ **无线和免提准确性**

系统提供强壮的无线和免提电话处理功能，在嘈杂环境下有特别高的准确率，语音识别高质可靠。

➤ **单个词汇校正**

也叫按档置信评分，如果一个长句中的一个词语未被识别，应用程序可提示用户重复该片段，而不是整个句子。

➤ **热词识别**

热词识别使系统能对讲话者进行偷听，等待特定的词汇或短语，将控制交还该应用程序。可在应用程序使用该功能，使识别器能够静静倾听，直到用户说出特定的短语进行请求时才与用户交互。

➤ **智能断句**

断句是对进来的样本流确定语句的起始和终止的处理过程。当找到语句的起始和终止点后，语句区向前后分别延伸预先确定的长度。一旦检测到语句的起始点，样本开始流向RecServer，直到发现语句的终止点。通过这种方式，RecServer在用户仍在讲话的同时，实际上已经开始处理讲话的内容，而又不处理讲话的起止处多余的空白，从而节约CPU时间和网络带宽。

➤ **打断功能**

打断功能使用户可以打断提示、作出响应，无需等到提示结束播放。打断功能使用户和系统间的交流更加快捷、自然，特别是系统的经常用户。

➤ **N-Best处理**

对于有些应用程序，可能需要识别引擎产生可能的识别结果集，而不是一个最好的结果。系统的N-best识别处理方法便有这个功能，它提供了可能的识别结果列表，并按可能性从高到低排列。

➤ **语法概率**

系统允许对呼叫者所讲的特定词语或短语的在语法中的概率进行指定。当被讲的词语或短语的概率可根据实际使用进行估计时，非常有用。对语法增加概率可提高识别的准确率和速度。

➤ **降低噪音**

当进来的呼叫包含稳定的背景噪音时，系统通过一种机制，使识别服务器更准确地进行识别。识别服务器将进来的话语进行增强，以有效地将语气、嗡嗡声、哼叫声、嘘嘘声等噪声过滤。如果相当数量的电话均含有稳定的背景噪声，比如在汽车上免提打电话时，这个机制效果较理想。

➤ **提示回放**

系统允许播放预先录好的提示和由文本到语音转换系统产生的提示。如果应用程序使用多个文本到语音转换服务器，资源管理器将对这些服务器的转换负载进行平衡，以提高硬件效率。

➤ **SNMP支持**

系统为远程监控提供了简单网络管理协议(SNMP)支持，独具可视化工具，便于进行配置、管理和操作。

➤ **用户资料**

用户资料将特定呼叫者的所有有关信息结合在一条数据库记录里，从而提供了一个方便的机制。可存储的信息包括声纹鉴别的声纹或您想增加的任何特定应用的数据。

2. 系统结构的优点

分布式结构具有伸缩性强、成本效率高和配置灵活等优点，具有支持数百万呼叫的能力；客户/服务机结构使语音识别应用程序威力强大，具有极其高效的容错性和载入平衡能力，保证了应用程序的可靠性。

➤ **分布式结构。**资源管理器在RecServer间进行载入均衡，从而保证硬件的利用效率。对CPU强度大的识别可以由非运行应用程序和音频接口的远程机器来执行。这种结构优化了内存和CPU资源的使用，因为每个RecServer可以支持多个客户端，资源管理器可以将呼叫平均分布到几个RecServer上。另外，对于小型系统或者原型，可在同一台机上运行RecClient和RecServer。

➤ **高密度接口。**将客户端的少量处理从CPU密集型的服务器处理隔离，允许客户端有高密度的接口又能提高服务器端CPU的使用效率。

➤ **容错和可靠性。**即使个别服务器失效，也不会使系统崩溃，甚至不会错过一个电话。当一个RecServer失效时，资源管理器自动停止向其发送请求，当服务

器恢复时，自动开始向它发送请求。呼叫者不会被断线，最多是系统要求其重复最后一句话，并转向另外一个可用的服务器。

- **维护方便。**可以关闭一个RecServer进行维修，而对整个系统的性能没有影响，或者影响很小。一些类型的维护甚至可以不关闭RecServer进行。
- **无缝的鉴别功能。**系统将识别过程和鉴别处理集成在一起。所以资源管理器不仅对所有RecServer的识别负载，同时还有鉴别负载，进行均衡。这个方案比分开进行校验和识别处理的方案具有更高的效率、需要更少的硬件。
- **可伸缩性。**随着呼叫量的增加，可增加RecServer、RecClient和应用的实例，无须停止任何运行着的应用程序或关闭IVR系统。