小米智能音箱“小爱同学”涉及的语音识别技术

姓名：郑雨婷 学号：2021150122

# 应用概述

物联网控制系统是通过对物品的信息采集和传输共享，对信息进行分析后发送控制指令对设备实行远程监控管理从而达到控制目标的预期。物联网控制系统具有多样化、开放化、广泛化和智能化的特点。小米公司于2017年推出的“小爱同学”智能音响是一个经典的物联网控制系统实例。

## 本文概述

本文对小米智能音箱涉及的多项技术进行了深入研究和综合分析，主要包括计算机语言、计算机声学和知识图谱三个方面。其中，重点调研总结了目前国内外语音识别技术的发展现状。接着，针对句子的情感分析问题，我们进行了系统建模，并利用IMDB数据集进行了训练，使用支持向量机算法实现了情感分析的功能。

## 小米智能音箱功能概述

小米智能音箱具备丰富的语音交互功能，包括在线音乐、网络电台、有声读物、广播电台等内容。此外，它还提供新闻、天气、闹钟、备忘录、时间、汇率、百科问答、闲聊、笑话、翻译等各类实用功能。

同时，小米智能音箱可以与部分小米智能家居产品进行捆绑，通过授权后可以通过音箱的语音交互来控制这些设备。例如，电视、机顶盒、扫地机器人、电饭煲、空气净化器、电风扇、智能灯等小米及生态链产品都可以通过小米智能音箱进行智能控制。

# 涉及的科学技术问题

“小爱同学”作为一个深受大众喜爱的物联网系统，在各个层面都运用到了前沿的技术，在计算机语音、计算机声学、知识图谱等方面均卓有成效。

## 计算机语音

计算机语音，指将人类语言信息转换为计算机可理解和处理的形式的技术。它涉及语音识别（将语音转换为文本）、语音合成（将文本转换为语音）以及语音相关的其他技术，如语音指令识别、语音情感识别等。计算机语音技术的发展使得计算机能够理解和产生自然语言，实现与人类交流的能力，广泛应用于语音助手、语音识别系统、电话自动接听等领域。

### 语音唤醒

在唤醒方面，小米语音唤醒技术为了兼顾低功耗与高性能，采用了双级唤醒策略。低功耗待机唤醒词检测模型，利用子采样与共享隐含层等技术，减少模型资源消耗的同时保证召回率在一个较高的水平。高性能误唤醒检测模型，采用粗粒度建模单元，结合局部信息与长时上下文信息，高效抑制误唤醒。通过从海量数据中自动挖掘高区分度训练样本，再经过数据扩充技术，提高唤醒模型在低信噪比与小音量场景下的鲁棒性。

### 语音识别

小米语音识别技术采用了深度学习和神经网络等先进的算法和模型，通过大规模的训练数据和端到端的训练方法，不断优化识别准确率和性能表现。它可以广泛应用于语音助手、智能音箱、语音输入、语音命令控制等方面，为用户提供便捷的语音交互体验。

小米语音识别在技术创新取得了明显的进步，主要包括3个方面：

第一，多通道端到端语音识别算法的研究取得重要进展。

第二，在细分人群和细分场景等方面取得了明显的改进。

第三，动态识别、并行推理等新技术应用改善了线上服务能力。

## 计算机声学

计算机声学技术，是以信号处理、深度学习为理论基础的，涉及阵列增强、通话降噪、智能感知、音频声场、声学测量等技术领域。

### 协同唤醒

协同唤醒技术是指通过多个设备之间的协作，实现根据具体场景和条件来选择最合适的设备进行语音唤醒操作的技术。它能够突破传统的就近唤醒原则，更加智能地选择响应设备，提供更好的用户体验。

在协同唤醒技术中，各个设备之间通常会通过网络或蓝牙等通信方式进行信息传递和协调。当用户发出语音指令时，系统会综合考虑设备距离、设备活跃状态、设备形态等条件因素，并结合场景识别、环境感知等算法进行分析，选择最优的设备来进行唤醒和响应。这样可以实现更精准、高效的语音交互，同时避免不必要的响应和干扰。

### 列阵增强

列阵增强（Beamforming）是一种信号处理技术，用于在接收端或发送端的多个微型阵列天线中，通过对接收或发送信号进行精确的空间处理，实现对特定方向上信号的增强或抑制。在接收端的列阵增强中，多个接收天线会采集到来自不同方向的信号，并通过相位权重和时间延迟的调节，将信号合成以加强感兴趣方向上的信号能量，同时减弱其他方向上的干扰信号。这种空间选择性增强可以提高接收系统的灵敏度和抗干扰能力，提供更好的语音识别、通讯质量等应用体验。

小米推出了自主研发的阵列唤醒算法，有效提升噪声场景平均唤醒率及回声场景平均唤醒率。有效解决周围说话的人距离智能设备的麦克风较远时存在的噪声、多径反射和混响，会导致麦克风收取信号的质量下降，严重影响语音识别率等问题。

## 知识图谱

知识图谱，简单理解就是知识库，知识图谱通过将各种事实和知识以节点和边的形式连接起来，形成一个全面而有机的网络结构。每个节点代表一个具体的实体或抽象的概念，而边表示节点之间的语义关系。例如，"苹果"和"水果"可以是两个节点，它们之间有一条"属于"的边，表示苹果属于水果类别。知识图谱对小爱应用场景支持更广泛，除了知识问答场景外，支持音乐、视频、古诗、菜谱、复杂推理、闲聊等场景。

小爱同学背后的知识图谱技术体现在以下几个方面：

①知识融合：多源异构知识融合技术支持了文本知识融合和多模态知识的融合；

②知识构建：知识自动构建技术支持用户定制及敏捷扩展。

③知识关联：支持复杂的关系推理和知识推荐场景；

总而言之，小米智能音箱运用到计算机语音、计算机声学、知识图谱等方面的技术。计算机语音中包括语音唤醒、语音识别；计算机声学包括协同唤醒、列阵增强等。它的目标是理解和处理人类语言，使得机器能够准确把握用户的意图，并提供相应的智能服务。

# 参考文献:

1. Help People who are Blind or Partially Sighted. https://www.orcam.com/en/
2. Hewett R, Douglas G, Keil S. Young people, visual impairment and preparing to live independently[J]. Visual Impairment Centre for Teaching and Research, University of Birmingham, 2015.
3. Bigham J P, Jayant C, Ji H, et al. Vizwiz: nearly real-time answers to visual questions[C]//Proceedings of the 23nd annual ACM symposium on User interface software and technology. 2010: 333-342.
4. https://blog.csdn.net/Rmwcf/article/details/125663393
5. [小米AI音箱\_百度百科 (baidu.com)](https://baike.baidu.com/item/%E5%B0%8F%E7%B1%B3AI%E9%9F%B3%E7%AE%B1/22046214#:~:text=%E5%B0%8F%E7%B1%B3AI%E9%9F%B3%E7%AE%B1%E5%8F%AF%E4%BB%A5%E5%92%8C%E7%B1%B3%E5%AE%B6APP%E5%BD%93%E4%B8%AD%E7%BB%91%E5%AE%9A%E7%9A%84%E9%83%A8%E5%88%86%E5%B0%8F%E7%B1%B3%E6%99%BA%E8%83%BD%E5%AE%B6%E5%B1%85%E4%BA%A7%E5%93%81%E8%BF%9B%E8%A1%8C%E6%8D%86%E7%BB%91%EF%BC%8C%E9%80%9A%E8%BF%87%E6%8E%88%E6%9D%83%E4%B9%8B%E5%90%8E%E5%8F%AF%E4%BB%A5%E9%80%9A%E8%BF%87%E9%9F%B3%E7%AE%B1%E7%9A%84%E8%AF%AD%E9%9F%B3%E4%BA%A4%E4%BA%92%E8%BF%9B%E8%A1%8C%E6%8E%A7%E5%88%B6%E3%80%82,%E5%8F%AF%E6%8E%A7%E5%88%B6%E7%94%B5%E8%A7%86%E3%80%81%E7%9B%92%E5%AD%90%E3%80%81%E6%89%AB%E5%9C%B0%E6%9C%BA%E5%99%A8%E4%BA%BA%E3%80%81%E7%94%B5%E9%A5%AD%E7%85%B2%E3%80%81%E7%A9%BA%E6%B0%94%E5%87%80%E5%8C%96%E5%99%A8%E3%80%81%E7%94%B5%E9%A3%8E%E6%89%87%E3%80%81%E6%99%BA%E8%83%BD%E7%81%AF%E7%AD%89%E5%B0%8F%E7%B1%B3%E5%8F%8A%E7%94%9F%E6%80%81%E9%93%BE%E8%AE%BE%E5%A4%87%E3%80%82)
6. [小米官方揭秘小米AI技术 一文看懂小爱同学进化的奥秘\_ZNDS资讯](https://news.znds.com/article/52204.html)
7. Y. Kong et al., "Multi-Channel Automatic Speech Recognition Using Deep Complex Unet," 2021 IEEE Spoken Language Technology Workshop (SLT), Shenzhen, China, 2021, pp. 104-110, doi: 10.1109/SLT48900.2021.9383492.