



Design and Development of Chinese Learning  
APP "MyTalk" Based on Speech Recognition  
Technology and Contextual Learning Theory

---

Manyun Tang

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

August 17, 2021

# 基于语音识别技术和递归分级认知理论的汉语学习 APP《MyTalk》的设计与开发

## Design and Development of Chinese Learning APP "APP" Based on Speech Recognition Technology and Contextual Learning Theory

**【摘要】** 在日益增长的汉语口语学习需求和学习类 APP 被人们广泛接受的背景下，开发设计了基于语音识别技术和递归分级认知理论的汉语学习 APP《MyTalk》，内含外国人来华生活、学习和工作等六大主题的十八门情境课程。通过人机语音对话的形式，遵照语块学习原则，给用户实时评价反馈，整个过程着重训练用户的汉语听说能力。本文介绍了软件开发设计所依据的理论背景，以及软件在交互、内容、评价机制和界面方面的设计思路。在收集用户反馈后，讨论了目前技术在模拟真实语言应用情境的不足，以及未来需要关照学习者从“虚拟”到“真实”方面的适应问题。

**【关键词】** 汉语学习；学习 APP；口语；语音识别

*Abstract: In the context of the growing demand for oral Chinese learning and the widespread acceptance of learning apps, the Chinese learning app "MyTalk" based on speech recognition technology and recursive hierarchical cognition theory was developed and designed. The software is designed for foreigners living in China. There are a total of 18 contextual courses in six themes such as study and work. Through the form of human-machine voice dialogue, following the principles of language block learning, real-time evaluation and feedback are given to users. The whole process focuses on training users' Chinese listening and speaking skills. This article introduces the theoretical background of software development and design, and the design ideas of software in terms of interaction, content, evaluation mechanism and interface. After collecting user feedback, it discussed the current technology's shortcomings in simulating real language application scenarios, and the need to take care of learners' adaptation from "virtual" to "real" in the future.*

**Keywords:** Chinese learning, learning APP, spoken language, speech recognition

### 1. 前言

语言的本质是交际而非单词或语法，人的认知是有一定次序的，语言习得要循序渐进。汉语学习者的主要目的是交际，因此汉语口语的学习至关重要。谭晓丽（2018）认为目前的汉语口语课堂教学受人数、实践、地点多方面限制，汉语学习者的口语水平如果只通过课堂学习，提高会比较缓慢，而成本低、灵活性高的汉语口语学习 APP 适应了汉语学习者快速提高汉语口语交际能力的需求。

当前人们的生活越来越离不开 APP，学习类 APP 也顺应潮流进入快速发展时期。目前，借鉴语言学习界的行业龙头 Duolingo（“多邻国”）诞生的两款热门汉语学习软件 ChineseSkill 和 HelloChinese 都侧重让知识点与闯关游戏结合，可见“交互+游戏化”的学习模式在汉语移动学习 APP 中的重要影响。市场上优秀成熟的汉语口语交互类学习 APP 集中在字词发音和阅读听说方面，在口语交际方面，现有产品（如 Mandarin Talk）提供的课程以文本跟读、学唱中文歌为主，适合低龄儿童，发展学习者成段对话能力的过程较长。

相较于英语的国际通用语言地位，汉语学习人群中，他们的交际需求更大，他们更加关注能够听懂并成段表达，对读写的要求较低。现有的 APP 能够满足在口语交际初级阶段的学习，如由外研社开发的 Talking Chinese 以游戏化的产品理念为指导，采用智能语音技术，通过动画模拟来进行教学，提供卡通动画构建的生活场景，设计有拼音、中文和英文的字幕。但这些 APP 整体话题浅显，场景比较低龄化，适合年纪较小的学习者，缺乏在具体生活场景中的成段表达训练，难以满足商务和旅游类的中高级交际需求。

因此，本研究遵循分级递归认知理论 (Recursive Hierarchical Recognition, 以下简称 RHR) 中语块学习的原则，在现有语音识别技术基础上，针对外国人来华生活、工作和学习方面的日常场景，设计开发面向具有简单汉语基础的学习者的口语学习 APP 《MyTalk》，并且结合 APP 的测试用户反馈情况，思考移动技术辅助语言学习的问题和挑战。

## 2. 研究背景

语言学家 Knowles (2008) 提出的 RHR 理论认为，发展会话技能是不同于发展书面技能的。会话的时候有时间压力。我们叫做瞬时张力。瞬时张力是一种积极的力量。它刺激大脑寻找语言模型。而书面技能中，瞬时张力要小很多了，因为文本是空间的。当处理话语时，语言被储存在短时记忆中，速度很快。这种记忆是有限的。据神经学家称，这种记忆能持续一到五秒钟，容纳四到七个语块的信息。为了用短时记忆处理更多的语言，大脑必须运用语言模型认知逻辑，把听到的语言归纳成更大的语块。

要发展归纳语块的能力，需要遵循四种技能习得路线。首先是听，然后是说，都要脱离文本。发展对于语言模型和声音的语音意识。然后读和写才能跟进。当语言模型变得熟悉和自然时，会话的流利度就提高了。会话的流利度提高的同时，还可以为书面技能提供支持，书面的技能又可以加强会话技能。书面技能也得益于处理更大语块的能力。当学生们在口头上接触和练习了语言模型后，才能通过读和写的练习再延伸。

在 RHR 理论指导下，MyTalk 的设计遵循语块学习的原则，跳过单词短语，直接进入语段的学习训练。评价机制也是从语块的准确度而不是个别单词的准确度来看。学习者在主题情境中训练发展听说技能，建立交际场景下的常用语言模型和声音的语音意识。

MyTalk 同时遵从情境学习理念的指导，按照主题情境来设计不同的学习内容模块。情境学习 (Situated learning) 是由美国加利福尼亚大学伯克利分校的让·莱夫 (Jean Lave) 教授和独立研究者爱丁纳·温格 (Etienne Wenger) 于 1990 年前后提出的一种学习方式。情境学习理论认为，学习不仅仅是一个个体性的意义建构的心理过程，而更是一个社会性的、实践性的、以差异资源为中介的参与过程。知识的意义连同学习者自身的意识与角色都是在学习者和学习情境的互动、学习者与学习者之间的互动过程生成的，因此学习情境的创设就致力于将学习者的身份和角色意识、完整的生活经验、以及认知性任务重新回归到真实的、融合的状态。情境学习理论告诉我们：学习的本质就是对话，在学习的过程中所经历的就是广泛的社会协商。而“学习的快乐就是走向对话” (崔允漦和王中男, 2012)。

在 MyTalk 中，设计的每个挑战关卡都围绕一个从社会生活中提炼出的中心话题，画面风格是简约卡通，尽量模拟真实生活场景。整体的对话情境与上下文设计也尽量向实际生活靠近，用词避免过于书面化，在追求规范准确的同时尽量贴近汉语母语者的日常使用习惯。每个关卡的评价细则也仔细考虑了真实情境中的多种回答可能，拟定与交际情境相符的多个关键词，反复测试确定关键词的匹配度。整个设计都是让学习回归真实应用情境，缩小学习环境和真实交际应用环境的差距，让学习者之后能更容易地在真实交际场景中应用所学。

### 3. MyTalk 的设计开发

MyTalk 是一款用 JAVA 语言开发的 Android 应用程序，主打“情境语音对话+智能人机交互”。程序中直接封装调用开源免费的百度语音识别 SDK，识别用户输入的语音内容。用户使用 MyTalk 的具体学习流程如图 1 所示。

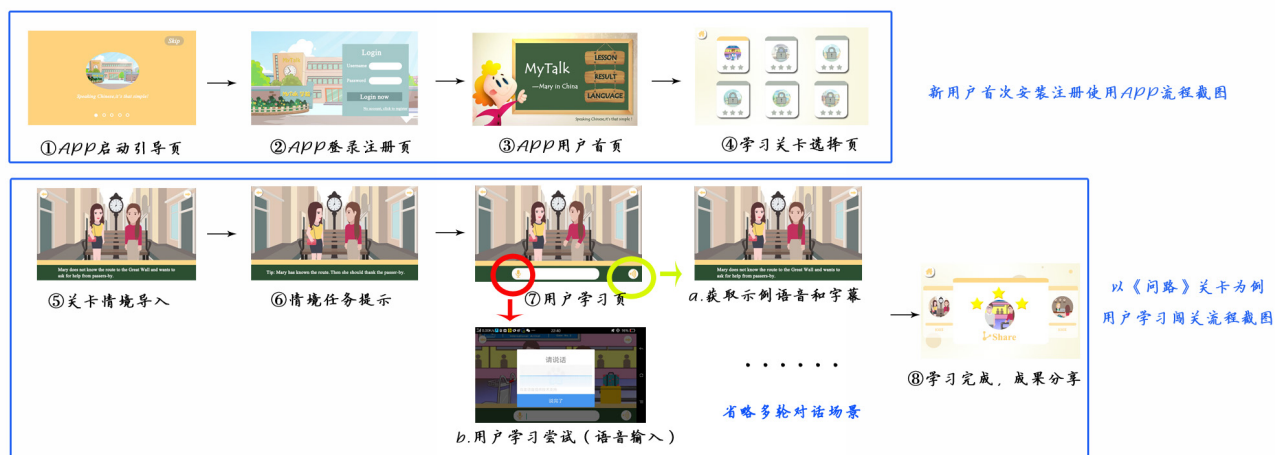


图 1 MyTalk 用户安装使用和学习流程

#### 3.1. 交互设计

每个关卡开始前会首先给学习者交际场景的语音描述，之后的多轮对话中，每次对话前还会给出更具体的任务提示，如果学习者听取提示有困难，可以点开提示字幕。整个过程中的语音提示默认中文语音，字幕提示是英语或其他用户自选语言，用户可以自主选择开启或关闭字幕提示，以此来训练学习者的中文听力能力。

当用户来到学习页面（如图 1 中的第⑦步流程截图所示），可以进行学习尝试（如图 1 中的第⑦步 b 流程截图所示），即语音输入交际话语，系统识别话语内容，根据情境关键词契合度，判断用户是否达到学习要求。如果达到学习要求，系统将给出中文语音和画面的奖励提示，增强学习者信心。如果未达到要求，将给出语音提示学习者去听取标准语音，并跟随标准语音进行反复跟读练习（如图 1 中的第⑦步 a 流程截图所示）。在多层对话后，用户该关卡总分合格，方可进入下一关卡。在软件使用过程中，用户可查看个人学习进度及系统内排名，并将学习成果分享至社交平台。

#### 3.2 内容设计

MyTalk 的故事背景中设定主人公为一名来华留学生，已具备初级汉语听说能力。从她将在中国经历的日常生活、学习、旅行等六大主题情境，设计了 18 门闯关课程（课程清单如图 2 所示）。每一个课程下将包含三道主题关卡，每个关卡的场景设计都和交际任务情景紧密贴合，以便让学习者有更强的场景代入感，丰富的背景构造将有利于语言的激活。

| 主题 | 关卡   | 内容 | 主题   | 关卡   | 内容 | 主题   | 关卡   | 内容 | 主题   | 关卡   | 内容 | 主题   | 关卡   | 内容 |       |  |
|----|------|----|------|------|----|------|------|----|------|------|----|------|------|----|-------|--|
| 初识 | 机场迎接 |    | 了解课程 | 了解课程 |    | 观看电影 | 观看电影 |    | 询问路线 | 询问路线 |    | 社交聚会 | 社交聚会 |    |       |  |
|    | 课堂互动 |    |      | 课堂讨论 |    |      | 观看视频 |    |      | 景点购票 |    |      | 公园野餐 |    | 逛街购物  |  |
|    | 饭店用餐 |    |      | 作业安装 |    |      | 健身活动 |    |      | 游玩咨询 |    |      | 话题交流 |    | 参观博物馆 |  |

图 2 MyTalk 关卡课程清单

### 3.3 评价机制设计

根据已有的研究启示，二语习得的关键是语块的掌握与运用，MyTalk 是利用第三方语音识别插件（百度语音识别 sdk）将语音转化成文本，再根据文本中的情境关键词与标准答案的契合度，评价用户汉语使用的准确性。情境关键词的准确表述和运用，将体现学习者是否掌握该交际情境下的所需语块。图 3 是以主题一中的第一关“机场迎接”为例设计的分级关键词评分细则，APP 尝试以用户的情境关键词契合度来评估用户是否胜任关卡的交际要求，决定学习者是否通过进入下一关卡。如果未达到过关要求，将提示学习者听标准答案的音频，尝试反复跟读，由此训练汉语听说能力。

1. 机场迎接

背景：Mary 下了飞机，遇到了来接她的王先生。

王：你好，我是 ABC 旅游公司的，我姓王，请问你是 Mary 吗？

提示：Mary 打招呼并且介绍自己。

Mary：你好，王先生，我就是 Mary，谢谢你来接我。

标准：

| 一级关键词     | 权重   | 二级关键词          | 权重     | 三级关键词  | 权重     | 计算公式   |
|-----------|------|----------------|--------|--------|--------|--|
| (A)       | (1)  | (B)            | (0.6)  | (C)    | (0.3)  | 总分 (S)   |
| A1: 你好/您好 | A1*1 | B1: Mary/马瑞/迈瑞 | B1*0.6 | C1: 谢谢 | C1*0.3 | $S=A1*1+A2*1+B1*0.6+C1*0.3$<br>当 $S \geq 1.3$ 即为过关 |
| A2: 我     | A2*1 |                |        |        |        |  |

王：不客气。你一路上累了吧。

提示：Mary 旅行很顺利，一点儿也不累。

Mary：不，今天很顺利，一点儿也不累。

标准：

| 一级关键词  | 权重   | 二级关键词  | 权重     | 三级关键词 (C) | 权重     | 计算公式  |
|--------|------|--------|--------|-----------|--------|---|
| (A)    | (1)  | (B)    | (0.6)  |           | (0.3)  | 总分 (S)  |
| A1: 不  | A1*1 | B1: 顺利 | B1*0.6 | C1: 今天/这次 | C1*0.3 | $S=A1*1+A2*1+B1*0.6+C1*0.3+C2*0.3$<br>当 $S \geq 1.3$ 即为过关 |
| A2: 不累 | A2*1 |        |        | C2: 飞行/航班 |        |   |

老师：那太好了，我们走吧。

图 3 内容脚本及分级评价细则示例

### 3.4. 界面设计

软件的整个界面以鲜亮的颜色为主，卡通的场景人物活泼生动但不过于低龄化，符合成人学习者的审美。软件需要用户横屏操作，字幕及对话框区域将在横屏界面的下方四分之一处填满，提示按钮布局在界面的右上角（关卡示意图如图4所示）。

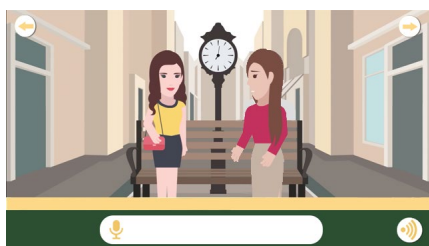


图4 闯关界面示意图



图5 美国女生人物形象

另外，用户在开始学习之前选择自己的性别和国家身份（如图5所示：美国女生 Mary 形象）。软件将支持英语、法语、西班牙语、德语和日语，根据用户的身份选择，提供对应的卡通人物形象，让学习者更有身份认同感。

## 4. MyTalk 的应用讨论

目前 MyTalk 已经小范围邀请目标用户进行体验测试，收获了一批反馈数据。对照软件开发设计初的美好畅想，实际使用情况和用户反馈不仅为 MyTalk 的后续改进提供了帮助，而且提出了许多值得思考讨论的移动技术辅助语言学习方面的问题。

### 4.1. 移动技术模拟真实情境

MyTalk 对真实情境的模拟体现在关卡界面和故事情境上，已有技术辅助语言学习的研究表示，我们需要更广阔的技术视野出发提醒我们要认清技术和语言、语境等的关系，考虑技术媒体影响上下文以及表达和交流形式的一些方式（Chun D, Kern R, & Smith B, 2016）。MyTalk 在试用测试中也收到用户反馈，实际语言对话氛围中，话轮的构建和持续远比闯关情境中复杂的多。软件中每个故事情境的话轮是两人对话，轮流发言的，而现实中会有多人对话、话轮中断或者话轮难以继续等等情况。另外，现实对话中，说话者的神态、语气和肢体动作都是在增加对话的信息。移动学习中，虚拟人物暂时没有神态、语气和复杂肢体动作（人物动态为事先绘制好的静态图像合成），所以相较真实学习环境，移动学习中对话双方的语义理解是更困难的。

所以，未来技术在模拟语言学习真实情境上还有很多需要突破的地方，或者当我们对移动技术背景下的语言学习环境和真实的语言应用环境间的差异非常清晰时，我们的学习设计能够更加有针对性，使得学习者在学习环境和应用环境间的转变更加平滑自如。

### 4.2. 从“虚拟”到“真实”的学习者心理

MyTalk 中学习者实则是和虚拟的交流者开展同步对话，所有使用非实体语言的能力都要求能够处理“真实”和“虚拟”操作的分层，语言学习者需要做的一件事就是在实体和非实体的语言使用之间转换，在“真实”和“虚拟”操作之间转换的经验（Clark, 1999）。

学习者在一开始接触工具，开展对话时往往充满好奇心和新奇感，学习积极性很高，但一段时间后，他们会摸清技术评价机制，例如有 MyTalk 的用户提出，在发现系统是根据关键词准确度来判定是否过关后，他意识到可以只说出一些关键词，而不再练习完整句子，这

是一种“讨巧”方法，虽然作为一个主动学习者，他没有实际这样去做。诚然在实际语言交流中，我们仅通过关键词也可以猜测对话另一方的表达意思，但长远来看，希望培养高水平的语言使用能力还是要掌握完整句意表达技能的。

另外，因为是面对虚拟交流者，知道有容错，可能很多学习者心理上更加放松。但转换到实际场景中，因为有真实交流对象，且对象的身份变化，对话目的不再是学习和练习，很多学习者有恐惧、紧张的心理。而且，很多学习者不知道如何将虚拟层面的练习成果转化到真实场景，开口难的问题依然存在。

所以，移动技术和语言学习结合的研究不仅是单纯技术，或者单纯语言层面，还应该重视应用的社会情境研究和学习者心理分析，帮助学习者适应“虚拟”和“真实”场景变化，构建比较清晰的转化策略。

## 参考文献

- 谭晓丽. 汉语口语学习 APP 的分析研究. (Doctoral dissertation, 上海外国语大学).
- 崔允灏, & 王中男. (2012). 学习如何发生: 情境学习理论的诠释. *教育科学研究*(7), 28-32.
- Chun, D., Kern, R., & Smith, B. (2016). Technology in language use, language teaching, and language learning. *Modern Language Journal*, 100(S1), 64-80.
- Clark, H. H. (1999). How do real people communicate with virtual partners? *Proceedings of 1999 AAAI Fall Symposium, Psychological Models of Communication in Collaborative Systems* (pp. 43-47). Falmouth, MA: AAAI.
- Knowles, Phillip. (2008). Recursive Hierarchical Recognition: A Brain-based Theory of Language Learning. *FEELTA/NATE Conference Proceedings, 2008*