

语音识别技术在中文教学中的应用：  
一堂汉语拼音练习课的启示  
(The Application of Speech Recognition Technology in Chinese  
Language Learning: What Can Be Learned From a Pinyin  
Lab Session)

笄骏

(Da, Jun)

中田纳西州立大学

(Middle Tennessee State University)

jun.da@mtsu.edu

**摘要：** 本文通过对一次汉语拼音课堂练习活动的观察初步探讨了借助语音识别技术进行拼音练习的三个基础问题。我们观察到目前的中文语音识别技术，比如像 Google 公司的语音识别引擎，可以给学生提供有意义的练习机会。这表现在 Google 语音引擎识别常用且有意义的言语时准确率更高，不能正确识别时给出的（出乎意料的）反馈让学生们觉得有趣，降低了他们的学习焦虑，学习变得更主动。虽然本研究有限的观察数据并不能充分证明语音识别技术对二语习得的益处，但我们至少观察到了学生学习行为的一些积极改变，这些改变是有助于二语习得的。我们的观察还表明在借助语音识别技术帮助学生进行拼音练习时最好使用（接近）真实的言语提示，避免使用无语义的音节组合或在真实生活中使用频率低的言语表达。

**Abstract:** This paper reports the findings from a language lab session where speech recognition technology was used to give students practice opportunities in Chinese Pinyin pronunciation. Our limited observations suggest that current Chinese ASR technologies, as exemplified by Google's Web Speech API, can provide opportunities for students to engage in meaningful pronunciation practices. This is demonstrated by students' more willingness to practice, and relaxed reactions to both the expected and unexpected feedbacks from the Google speech engine. Though we do not have sufficient data to demonstrate the effectiveness of current speech technologies in improving second language acquisition, our observations do indicate that it helps to lead to some good learner behaviors in language learning, which are acknowledged to be good indicators of better language acquisition. Our observations also suggest that when designing Pinyin activities it is better for instructors to provide students with meaningful and frequently used prompts rather than isolated syllables groups or low-frequency expressions.

**关键词：**语音识别，中文教学，拼音练习，二语习得，教学法

**Keywords:** Automatic speech recognition, Chinese language teaching, Pinyin practice, Second language acquisition, Pedagogy

## 1. 引言

自然语言处理中的自动语音识别(Automatic Speech Recognition, ASR)是指将语音转换成文本(Speech-to-text)的技术,其起点可以追溯至上世纪三十年代 Bell Labs 的研究<sup>1</sup>。在互联网和移动终端普及之前,为用户提供语音识别服务主要是通过安装在单机上的应用程序和有限能力的中央控制系统来实现,比如 Windows 操作系统的声控功能, Rosetta Stone 外语学习软件<sup>2</sup>、和在美国常见的电话语音客服系统等。近几年来随着云计算、移动(互联网)和大数据技术的广泛使用,尤其是其本身的不断进步,语音识别技术在更广的范围内得到了应用。比如各大互联网公司都为其操作系统和服务软件提供了语音识别支持(如 Apple 公司的 Siri<sup>3</sup>、Google 和 Microsoft 公司的语音检索服务等)。不少移动终端制造商也给其智能手机和平板电脑配备了语音输入和应用,用户可以通过智能移动终端的语音输入和 Google 公司的 Chrome 等浏览器使用语音识别服务。

在外语教学中使用语音识别技术是一件很自然的事情。在语音识别技术发展(Huang, Baker 和 Reddy, 2014)的各个阶段,外语学习者和教师、外语教学研究者以及应用开发厂商都尝试过将语音识别技术用于各种语言的教学,探讨过其应用于语言学习各个方面的可能和问题(Holland 和 Fisher, 2007)。被关注的话题包括语音识别处理各个语言的可能和有效性(如最近的 Liakin, Cardoso 和 Liakina 等, 2013; Mushangwe, 2015),在外语学习中使用的有效性和学习者对其的感知(Cordier, 2009),以及使用语音识别技术的教学方法(如 Morton, Gunson 和 Mervyn, 2012)等。早期较为他人熟知的带有语音识别功能的外语学习软件包括 Rosetta Stone<sup>4</sup>和 TELLMEMORE<sup>5</sup>等。

较之于早期使用语音识别技术,目前的语音识别技术在其使用场景和便利性、尤其是可识别的语言内容开放性、反馈信息、以及第三方应用开发等方面有了很大的改观,主要表现在以下三个方面:

### (1) 可使用语音识别技术场景和便利性的改变

<sup>1</sup>参见: [https://en.wikipedia.org/wiki/Speech\\_recognition](https://en.wikipedia.org/wiki/Speech_recognition)

<sup>2</sup>参见: <http://www.rosettastone.com/>

<sup>3</sup>参见: <https://www.apple.com/ios/siri/>

<sup>4</sup>参加: <http://www.rosettastone.com/>

<sup>5</sup>参见: <http://tellymore.com/home.aspx#&panel1-1>

早期的语音识别大都通过安装在单板机上的软件进行。用户使用语音识别受制于机器的可移动性，可使用语音识别软件的场所有限。但是随着智能移动终端和移动互联网的普及以及使用成本的降低，用户可以更为方便地使用语音识别技术应用。与早期只能在单板机上使用语音识别软件的不便相比，他们现在可以更方便地在比如在手机上或通过网页浏览器在更多的场所提交需要识别的语音，并获得反馈信息。

### (2) 可识别的语言内容和识别反馈信息的改变

早期的语音识别应用软件都是使用了 HMM 等统计算法进行语音模式匹配<sup>6</sup>，安装在单板机上，与外部没有数据交换，用户的语音输入和获得的识别反馈只是局限于事先编程好的有限内容。与此相比，目前用户大都可以通过移动端或网页浏览器更随意地提交语言输入给处于云端的语音识别引擎，并从那里获得识别反馈信息。这些基于云端的语音识别引擎不仅使用了传统的统计算法，而且还借助了语音和文本语料库等大数据提高识别的精度和智能反馈，极大地丰富了可识别的语音内容和与用户的交互。比如使用 Apple 公司的 Siri 引擎，用户不仅可以声控自己的 iPhone 或 iPad 进行功能性操作，而且还可以通过互联网使用 Apple 公司提供的语音搜索服务查询诸如天气和场所等真实世界信息。前者与早期的语音识别应用服务并无本质上的差异，但后者则是语音识别技术应用的一个本质性改变。正是由于这种基于云端模式的语音识别和大数据的支持，语音识别已经从仅提供单纯的语音模式识别发展到能够智能地给用户提提供真实世界的信息。

### (3) 开发语音识别应用软件门槛的降低

早期的语音识别应用软件开发门槛很高，需要有专业编程人员开发。所以一般外语教师都无缘于自行开发使用语音识别技术的教学软件或课件。近年来不少掌握语音识别技术的大（互联网）公司都提供了开放性语音识别应用接口(Speech API)。比如，可用于汉语普通话语音识别的开放平台有百度公司的语音开放平台<sup>7</sup>、科大讯飞的讯飞开放平台<sup>8</sup>以及 Google 公司通过其 Chrome 浏览器提供的语音识别接口(Web Speech API)。这些接口使得第三方人员开发语音识别应用的门槛降低了很多。例如，使用 Google 语音识别接口开发最简单的应用只需掌握一些 JavaScript 的编程知识。这使得外语教师自己或借助其他程序员的帮助很容易开发一些带有语音识别功能的外语学习应用。同时，它也改变了以往学习者只能借助于键盘的书面语言输入方式与学习软件进行交互。例如，学生在外语学习平台上完成一个填空练习，以往只能通过键盘输入答案，现在也可以通过语音输入来完成。

这些语音识别技术的新发展给开发外语学习软件，设计各种教学活动带来了许多新的可能性。在使用这些新技术，尤其是将语音识别智能信息反馈融合到语言教学软件中之前，一些基本教学问题仍然需要得到验证，其中包括：1) 基于目前开

<sup>6</sup>参见：<http://cmusphinx.sourceforge.net/wiki/tutorialconcepts>

<sup>7</sup>参见：<http://yuyin.baidu.com/fc.html>

<sup>8</sup>参见：<http://www.xfyun.cn/>

放平台的语音识别准确率，尤其是对外语初学者多样化语音的识别准确率如何？例如，就汉语学习者而言，怎样的言语输入能被成功识别：是单纯的单字音节，还是一段有语义的言语(像汉语中的成语和常见习语表达)？2) 基于云端和大数据的语音识别引擎可以智能地提供关于真实世界的信息反馈，这些模拟了真实世界的人际交互除了给学生提供更多的口语练习机会外，是否还能提高学生们的学习兴趣？3) 将这些新的语音识别技术应用到外语教学中，有哪些需要我们关注的教学法问题？

本文借助于一次使用了语音识别技术的汉语拼音课堂练习活动试图对以上三个问题做一个初步回答。本文以下将分为三部分：第二节将描述这次拼音练习活动的基本情况；第三节将提供笔者观察到的本次拼音练习活动的一些数据，并借助于这些数据讨论使用语音识别新技术给外语教学带来的益处和教学上的注意事项；第四节将是本文的小结。这里需要事先提及的是由于参与本次课堂练习的学生人数较少，本文的讨论将只是抛砖引玉，其结论不具备统计上的定论意义。

## 2. 一堂汉语拼音练习课

### 2.1 参与拼音练习的学生

参与本次拼音练习的是笔者所在学校的初级汉语班零起点学生，共有 10 位，他们都主修其他专业。在本次拼音练习课之前，这些学生已经从零开始学习了六周的汉语，每周三课时，使用《Chinese Link (2nd ed.)》教材(吴素美等，2010)。除了学期开始两周学习拼音和其他学习内容外，他们在课外并无额外的练习汉语口语的机会，在参与本次练习活动时仍属于 ACTFL 初低水平。

### 2.2 本次练习活动

本次学习任务是借助之前已经学过的语言表达来复习学习过的拼音。较之于以往教师带领的汉语语音学习和通过小组活动等进行的个人练习，本次是让学生通过 Chrome 浏览器使用语音识别来口头完成一个在线拼音练习。语音识别练习直接使用了 Google 公司提供的语音识别测试页面<sup>9</sup> (图一)。练习提示(Prompts)由教师提供，分五类，包括仅有声调差异的单字音节组、数字、类似音节但有声调差异的有语义的表达、学生已经学过的汉语常用表达、和中文经典短语(见表一)。之前任课教师在教拼音时使用过数字作为教学材料。因为直接使用了 Google 公司的语音识别页面，所以这五类练习提示都打印发给学生，同时提供拼音和汉字。教师要求学生每组拼音至少说两遍。

<sup>9</sup>参见：<https://www.google.com/intl/en/chrome/demos/speech.html>

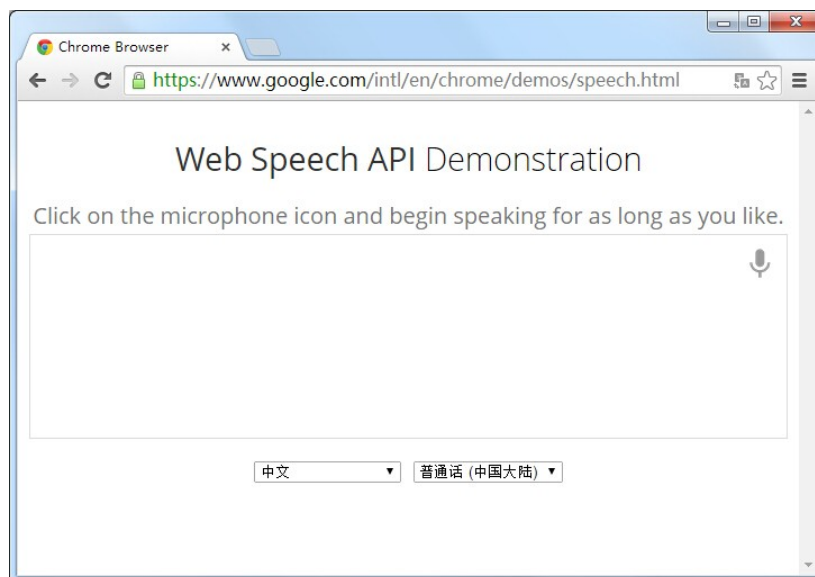


图 1: Google 语音识别测试页面

表 1: 拼音练习的提示

拼音练习种类	练习提示
1) 仅有声调差异的单字音节, 无语义	mā      má      mǎ      mà 妈      麻      马      骂 māo     máo     mǎo     mào 猫      毛      柳      帽
2) 数字	yī      èr      sān     sì      wǔ 一      二      三      四      五
3) 类似音节但音调不同, 有语义, 但真实世界里使用的频率很低	māmamǎimǎ 妈妈买马。 māmamàimǎ 妈妈卖马。
4) 学生已经掌握的语言表达, 有语义, 现实中常用	nǐhǎo ,wǒjiàowángdàzhōng. shìzhōngguó rén. 你好! 我叫王大中, 是中国人。
5) 中文经典表达, 汉语初学者一般没学过, 也不会使用	yuǎnqīnbùrú jìn lín 远亲不如近邻

### 3. 语音识别结果和讨论

#### 3.1 识别结果

为了比较 Google 语音识别结果, 笔者在课后又找了一位 ACTFL 中级水平的汉语学生和一位汉语母语者进行了语音识别对比。表二至表六穷举了各种识别结果, 但没有统计频次。

表 2: 仅有声调差异的单字音节但无语义的识别结果

语言提示	妈 麻 马 骂	猫 毛 柳 帽
本班初学者	么么么么 么么马来吗 马马马马 马 888	猫猫猫道, 猫猫猫猫 毛毛妈露曼 猫猫猫
中级水平学生	妈妈妈妈	猫猫猫猫
汉语母语者	妈妈妈妈	猫猫猫猫

表 3: 数字的识别结果

语言提示	一 二 三 四 五
本班初学者	12345
中级水平学生	12345
汉语母语者	12345

表 4: 类似音节但音调不同, 有语义, 但真实世界里使用的频率很低

语言提示	妈妈买马 vs. 妈妈卖马
本班初学者	那么差吗 那么慢吗 吗吗卖吗 妈妈卖吗 妈妈卖吗 慢慢买吗
中级水平学生	妈妈买吗 妈妈卖吗
汉语母语者	妈妈卖吗

表 5: 学生已经掌握的有语义且现实中常用的表达

语言提示	你好! 我叫王大中, 是中国人。
本班初学者	你好我叫王的中学中国人 你好我叫王大将是中国人 你好我叫王大众是中国人 你好佛教网双十中国人 你好我交往的装修窗口 你好我叫王的升学中文中国人 你好我叫王大中学中国 你好我叫王大洲是中国人 你好我叫黄的钟是中国人
中级水平学生	你好我叫王大众是中国人
汉语母语者	你好我叫王大中我是中国人

表 6: 中文经典表达, 汉语初学者一般没学过, 也不会使用

语言提示	远亲不如近邻
本班初学者	人亲不如近邻 yuan 亲不如近邻 远亲不如近邻 娘亲不如近邻 养亲不如近邻 情不如精灵 燃情不如精灵 yvonne 情不容情
中级水平学生	远亲不如近邻
汉语母语者	远亲不如近邻

从上述各表列出的识别结果我们可以得出以下的观察:

(1) 如果只是使用无语境语义的单字音节来练习发音, Google 语音识别引擎总体上可以识别音节, 但无法区分声调。这对初级和中级水平的学生, 以及汉语母语者都一样(表二)。

(2) Google 语音识别引擎可以准确地识别数字, 无论是针对什么水平的说话者(表三)。

(3) 当需要识别类似音节但音调有差异的言语时, Google 语音识别引擎仍然只能识别出音节, 但无法区分声调, 虽然这个言语表达有语义, 汉语教学中常用, 但在真实世界里使用的频率很低(如妈妈卖马, 见表四)。这种结果类似于使用单字音节的情况。

(4) 无论是哪种水平的说话者, 如果使用真实世界里常用的有语义的语音输入, Google 语音引擎识别率明显提高(表五)。

(5) 如果需要识别的是中文的经典短语或表达, 则识别率也明显提高, 类似于第四种情况(表六)。

综上所述, 我们可以推论 Google 的语音识别引擎除了使用了模式匹配算法之外, 还结合了语料库的使用帮助其提高识别的精度。这就是为什么需要识别有语义的汉语常用表达时无论说话者是谁它的识别准确率都会更准确。

### 3.2 学生在完成语音识别练习时的行为表现

在完成这次学习任务过程中, 笔者还观察到学生们的以下表现:

(1) 因为是第一次使用语音识别帮助学生练习发音, 他们对这个新颖的练习形式很好奇。较之以往的课堂练习, 学生们在本次练习过程中更加全神贯注;



(2) 学生对某些识别结果的出其不意感到很有趣, 尤其是在教师用英文解释了识别反馈信息之后。较之以传统课程上的小组训练或教师领读练习, 学生做这样的练习时更放松。

(3) 由于识别的结果时常是出奇不意, 学生在觉得很好玩的同时还更愿意主动反复尝试, 并自我纠错, 增加了练习的频率。

### 3.3 本次练习尝试对使用当前语音识别技术的教学启示

基于以上识别结果和对学生完成学习任务过程的观察, 我们可以得出这样的一些初步结论和教学建议:

(1) 使用语音识别技术进行汉语拼音练习可以提高学生的学习兴趣, 增加他们的学习机会。这种练习机会对很少或无机会接触汉语母语者的学生来说更有价值, 因为提高学习兴趣是帮助学习者进行二语习得的关键因素 (Dörnyei, 1998)。

(2) 学生使用语音识别技术制作的教学软件进行拼音练习, 因为识别结果时常很有趣, 所以他们在练习时也很放松, 降低了学习过程中的焦虑, 更加注重练习结果, 并能进行自我纠错 (Krashe, 1981)。如果在学生自我纠错的过程中能给他们提供参考发音, 将更有助于他们的语言习得。

(3) 如果使用语音识别技术来进行语音练习, 教师最好选取有语义且常用的短语表达, 避免使用无语义或很少使用的言语提示。使用 (接近) 真实语料可以有助于学生在与智能语音识别引擎交互的过程中模拟真实的交际任务。这个做法符合交际法教学原则, 有助于学习者的二语习得 (Richards, 2006)。

## 4. 小结

本文借助于对一次汉语拼音课堂练习活动的观察初步探讨了使用语音识别新技术进行汉语拼音学习的三个基础问题。我们观察到目前基于云端的语音处理技术在识别真实世界常用的言语时其准确性更高, 学生使用时会感觉有趣, 并因此会主动地反复练习和自我纠错。这里需要指出的是本次研究的数据有限, 今后需要更多的研究和数据以确认语音识别新技术对于二语习得的益处和其在外语教学中的最佳使用方法。希望本次初步的研究能起到一个抛砖引玉的作用。

## 参考文献

Cordier, D. (2009). Speech recognition software for language learning: Toward an evaluation of validity and student perceptions (Unpublished doctoral dissertation).



- University of South Florida, Tampa, FL . Retrieved from <http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=2908&context=etd>.
- Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching*, (31)3, 117 - 135. DOI: 10.1017/S026144480001315X, Published online: 12 June 2009
- Holland, M., & Fisher, F. P. (2007). The path of speech technologies in computer assisted language learning: from research toward practice. New York, NY: Routledge.
- Huang, X., & Baker, J., & Reddy, R. (2014). A historical perspective of speech recognition. *Communications of the ACM*, (57)1, 94-103.
- Krashen, S. (1981). Second language acquisition and second language learning. New York, NY: Pergamon Press.
- Liakin, D., Cardoso, W., & Liakina, N. (2013). Mobile-assisted learning in the second language classroom. *International Journal of Information Technology & Computer Science*, (8)2, 58-65.
- Mushangwe, H. (2015). Using voice recognition software in learning of Chinese as a foreign language pronunciation. *The Journal of Language Teaching and Learning*, 1, 52-67.
- Morton, H., Gunson, N., & Mervyn, J. (2012). Interactive language learning through speech-enabled virtual scenarios. *Advances in Human-Computer Interaction*, 1-14. Retrieved from <http://www.hindawi.com/journals/ahci/2012/389523/>.
- Richards, J. (2006). *Communicative language teaching today*. Cambridge, UK: Cambridge University Press.
- Wu, S., Yu, Y., Zhang, Y., & Tian, W. (2010). *Chinese link (2nd Edition)*. New York, NY: Pearson.