

语音评测技术 助力英语口语教学与评价

□文 / 魏思、吴奎、竺博、王士进



魏 思

中国科学技术大学博士，科大讯飞 AI 研究院副院长，科大讯飞首席科学家，“讯飞超脑”计划的主要负责人。带领团队在语音识别、语音合成、语义理解、知识图谱、机器学习等领域取得多项成果，多次摘取国际权威比赛桂冠。入选 2017 年《麻省理工科技评论》全球 35 位 35 岁以下科技创新青年中国区榜单。

中国科学技术大学硕士，科大讯飞研究院研究主管，主导开发科大讯飞第一代音乐检索系统，目前负责中英文语音评测技术研发工作，在朗读评分、发音诊断和口头表达评分上有深入的理解和实践，拥有数项专利。

吴 奎



竺 博

中国科学技术大学硕士，科大讯飞 AI 研究院副院长，广东讯飞启明科技发展有限公司副总经理。长期从事文档图像识别、语音识别、智能评测等领域人工智能方向相关研究，负责人工智能技术在教育领域的应用落地等工作，先后参与和主导科大讯飞输入法手写识别、科大讯飞智能考试评卷辅助系统的研发，在相关国际会议和期刊上有多篇学术论文，拥有数项专利。

中国科学院自动化研究所博士，科大讯飞 AI 研究院副院长，个性化教育研发部总经理，长期从事语音语言、人工智能、教育信息化技术方向研究；2009 年作为技术负责人，带领团队在国内首次将自动语音评测技术应用于正式英语口语考试；2014 年获得 IWSLT 评测中英和英中翻译方向的第一名；2015 年获得 NIST OpentMT 评测中英翻译方向的可用性人工评分第一名。在相关国际会议和期刊上有多篇学术论文，拥有十余项专利和软件著作权。

王士进



在学习一门语言时，不仅要做到能看会写，还要做到能听会说。然而，对于学生而言，后者通常比前者更具难度。如何让学生更好地将语言学以致用，成为了很多老师深入思考的问题。随着语音评测技术的发展，这一问题有望得到很好的解决。本文将以前科大讯飞研发的英语口语智能考试及评分系统、英语听说教考平台、小学口语作业平台为例，解读技术对语言学习带来的改变。

对于大多数中国学校和老师而言,英语口语的教学评价难度远高于书面英语。同样,对于大多数中国学生而言,提升英语口语能力也远比提升书面写作能力困难。随着人工智能技术的发展,特别是语音评测技术的发展,大规模口语考试评分已成为现实,英语口语的教学效率得到大幅提升。本文将科大讯飞研发的英语口语智能考试及评分系统、英语听说教考平台和小学口语作业平台为例,围绕中国英语口语教学评价的现状、语音评测的基本原理和关键技术等问题展开分析,阐述语音评测技术在英语口语教学评价中的应用,探索未来的更多可能。

一、中国英语口语教学与评价现状

外语能力是国民综合素质的一部分,外语水平的提升有助于一个民族在世界上获得更多话语权。目前,英语是世界通用语言,掌握了英语就有了与世界沟通、交流的工具。学习语言最重要的目的在于交流,听说能力是最重要的语言技能。早期,受应试教育模式的影响,我国英语教学过于重视读写能力的培养,忽略了对口语表达能力的锻炼。在英语课堂上,多以老师讲述语法、词汇为主,口语表达的机会寥寥无几,可以说基本上没有把口语作为教学的重点,造成中国学生普遍存在“哑巴英语”、“中式英语”的现象,严重影响学生的未来发展。

早期,我国英语口语的教学难点可总结为三点。首先,大班教学,学生多,课堂上很难让每个学生都有开口说话的机会;其次,在课后没有能够让学生进行口语练习且提供即时评测反馈的平台;最后,日常口语考试的组织难度大,需要有专门的语音机房、配套软硬件机考系统、人工评分系统以及考试题型资源,老师需要花费数倍于批改书面试题卷的时间,才能完整地听完学生的作答音频并给出评分。

这种状况正在慢慢发生变化。近年来,为适应社会需求,推动素质教育改革,全国各地的教育主管部门在中高考的英语口语考试方面展开了积极探索,部分教育发达地区将英语口语考试纳入了中高考。中高考英语口语考试题型众多,考生人数众多。传统的英语口语考试,采用人工面对面考试或计算机录音、人工评分的方式进行,这种考试方式组织难度大、实施成本高,评分结果容易受到评卷人的主观性影响,不利于大规模开展。因此,在2011年以前,各省份的高考只有英语口语加试,不计入高考成绩。2011年,广东省率先不再设置英语口语加试,直接将英语听说成绩计入高考总分,参加英语科目考试的考生必须参加该项考试,考试成绩通过人工双评分方式获得。2014年,科大讯飞研发的计算机智能语音评测系统正式应用于广东高考的英语听说考试,在两天内完成了所有的评分工作,大幅度降低了考试的组织难度,提高了评分效率,标志着大规模英语口语考试真正成为现实,进入计算机智能评测时代,具备全国推广可行性。目前,广

东、上海高考已包含对英语口语的考核，并将成绩计入高考总分，北京、江苏、广州、温州等 32 个省市将英语听说考试纳入中考，其中近 15 个省市将成绩计入中考总分，科大讯飞的计算机智能评分系统在辅助评分方面得到了广泛应用。

在“以考促教”思想的指导下，学校重视口语教学，学生口语能力大幅度提升。一些地区已经建立起较为完善的口语教学考试系统平台，可在专用的语音机房中进行口语教学、练习和模拟考试，借助语音评测技术，教师可即时获得学生的评价得分反馈，初高中英语口语的教学和评价效率得到大幅度提升。在教育改革的大环境下，英语口语考试已是中高考必然趋势。目前，能够实施英语口语教学和考试的地区只占少数，大部分地区受限于师资资源，无法开展有效的口语教学和考试活动，但借助语音评测技术和其他地区的成功经验，这种状况可以很快得到改善。

相对于初中生和高中生，小学生没有升学压力，且处于语言学习的最佳阶段，英语口语受到家长和学校的重视。然而，由于缺乏语言环境，40 分钟的课堂教学难以让每个学生都有开口表达的机会。为了让学生能够在课后有机会开口说话，早期，老师会借助微信群布置课文朗读作业，学生以语音方式提交。由于学生数量较多，检查语音作业会耗费老师较多精力，导致老师对学生的口语作业评价不力，学生也随之表现出很大的随意性。从 2012 年开始，口语作业平台相关应用开始兴起，智能语音评测技术替代老师和家长，对学生口语作业自动进行评分和发音错误进行反馈，解放了老师，提升了学生的学习积极性。

大学英语教学也在发生改变，开始将口语教学列入教学重点。越来越多的大学引入类似 FiF（外研讯飞）的口语训练和测试系统，学生课后可以通过手机 APP 随时随地完成口语作业和自主学习，极大地改善了语言环境。

综上所述，中国英语口语教学评价发生了重大改变，学生的口语水平大幅提升。其中，语音评测技术起到了不可替代的作用。

二、语音评测简介与关键技术

在现代语言学习中，计算机辅助语言学习系统（Computer-Assisted Language Learning, CALL）扮演着越来越重要的角色，为语言学习者提供了更真实和更丰富的语言学习环境。CALL 系统的核心模块是语音评测技术，指利用计算机自动对学习者的语音进行评价，具体包括发音错误检测、发音水平以及口语表达能力评价，该技术从上世纪 90 年代开始得到广泛研究，发展迅速。语音评测包含朗读评测和口头表达评测两项关键技术。前者主要包括如字、词、句、篇的朗读等题型，考察重点是学习者的发音错误情况和发音质量；后者主要包括如口头翻译、口头复述、看图说话、话题表述等题型，

主要考察学习者的逻辑思维能力和语言组织能力。

朗读评测技术研究较早，目前已达到成熟水平。典型的朗读评测原理框图如图 1 所示。

朗读评测最基础的一步是评测文本与评测音频的时间对齐，即音频切分，朗读评分和检错都基于音频切分结果进行特征提取。音频切分的一般步骤为：基于朗读文本，首先生成一个音素切分网络，然后结合标准发音模型，获

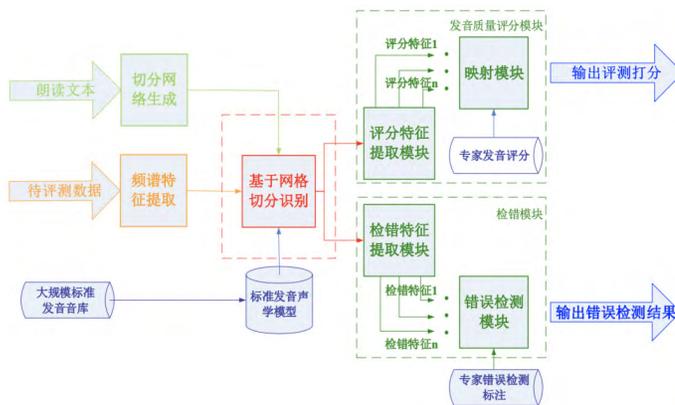


图 1: 朗读评测原理框图

取朗读文本中每个音素在评测音频中的时间边界。一般来说，切分网络的结构设计较为灵活，面对漏读、增读、回读等各种复杂情况，都能够准确地进行边界切分。发音质量一般包括准确度、流畅度、完整度三个维度分，以及三个维度的综合分。在声学模型层面，2010 年以后，神经网络模型取代 GMM-HMM 模型，成为语音识别声学模型的主流结构，发音错误自动检测和发音质量自动评价也随之步入神经网络时代，效果得到进一步提升。

口头表达评测的技术难度远大于朗读评测，直接通过音频评价学习者的口头表达能力是极其困难的。目前的做法通常是先识别学习者的口头表达内容，表示成文本，然后基于识别文本提取内容相关特征，进一步将识别文本作为评测文本，提取发音水平相关特征，与内容相关特征组合在一起，构成完整的评分特征，如图 2 所示。

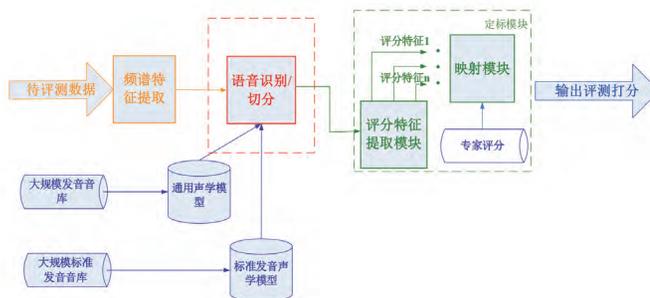


图 2: 口头表达评测原理框图

如上所述，语音识别准确性和内容特征矢量化表征是文本无关评测的核心。得益于语音识别和自然语言处理技术的发展，特别是深度学习在这两个领域的突破，无文本语音评测才逐渐达到使用水平。在语音识别方面，主流的语音识别系统以 LSTM 神经网络为声学模型，识别率已经达到 90% 以上，满足评分需求。在内容矢量表示方面，基

于 Word2Vect、LSTM、CNN、Decoder-Encoder 的内容表示亦成为目前的主流方法。

对于看图说话、话题表述题型，虽然可以通过现在主流的深度神经网络进行回答内容的矢量化表征，但由于参考答案空间无限大，可以有不同的角度、不同的词汇和不同的语法结构，因此计算机无法直接进行评分。目前，通常采用专家定标，机器进行学习的技术方案，称为“定标方案”，如图3所示。具体过程是针对每个题目选取一批覆盖高、低分的数据集，

称之为“定标集”，让机器学习人工专家的评分结果，形成与人工专家定标标准相一致的评分标准，在学习过程中，机器会掌握不同数据所具有的特性。

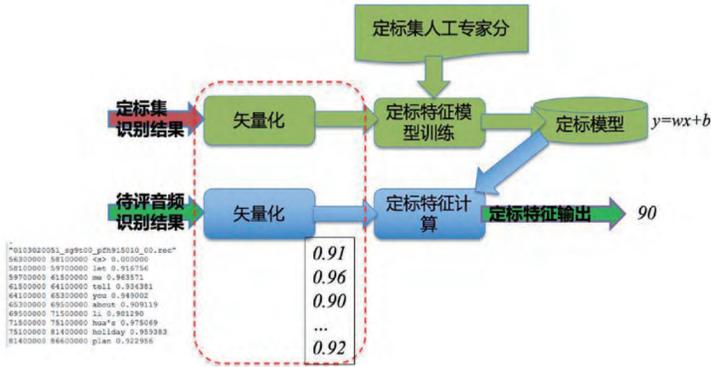


图3：定标方案

朗读评测技术和口头表达评测技术相结合，可以全面地评价学习者的语言水平。目前，上述技术已成功应用于众多大规模口语考试的自动评分，国外以ETS为代表，国内以科大讯飞为代表。机器自动评分极大地减少了人工评分的工作量，并保证了评分的公正性。

三、科大讯飞英语口语智能考试及评分系统

目前，科大讯飞英语口语智能考试及评分系统已在包括北京、上海、广东在内的6个省市的高考，北京、广州、江苏在内的32个省市的中考以及CET-SET在内的2个社会化考试中成功应用，截至目前，已经累计测试

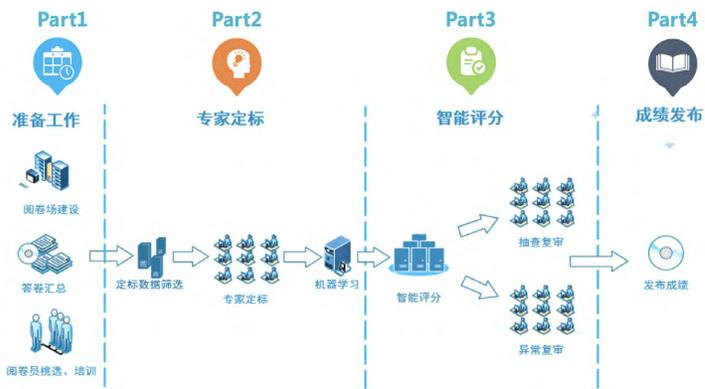


图4：智能化听说考试流程

2000万考生。英语口语智能考试及评分系统包括考务和智能评测两大核心系统,前者包括人机对话机考系统、人工评分系统两个子系统,后者包括专家定标和智能评分两个步骤。智能化听说考试流程如图4所示。

下面以广东高考英语口语智能评测为例,详细阐述智能评测的过程。广东高考英语听说考试包括模仿朗读、角色扮演、故事复述三种题型,前者属于朗读评测技术范畴,后两者属于口头表达评测技术范畴。由于口语考试每年都采用多套新试卷,试卷难度难以统一,为消除试卷间的差异性,每套试卷的所有题目都采用定标评分方案,使智能评分尺度与专家评分尺度保持一致。为保证评分的准确性和公平性,科大讯飞在评测技术和评分流程上进行了优化。在评测技术方面,考虑到考场噪声大,考生答题录音受到噪声污染,科大讯飞专门设计了鲁棒性强的声学模型和打分特征噪声补偿技术,最终使评分系统具备较好的抗噪性能。此外,基于当次考试的所有考生数据进行语音分析,在正式评测前完成声学模型、语言模型的自适应学习,显著改善了评测系统在正式评测运算中对考生语音的识别性能。在评分流程方面,计算机完成对全体考生的自动评分后,通过专门设计的多系统校验算法筛选出机器自动评分置信度低的考生,组织少量人工专家对其进行人工打分,即异常复审环节。同时,为验证自动评分效果达到可发布水平,增加抽查复审环节,在全体考生中随机抽取少量考生进行专家评分,将专家评分与自动评分进行对比。2012年11月30日,在广东省科学技术厅的组织下,来自于教育部考试中心、中国科学院自动化研究所以及部分高校的专家,对科大讯飞的英语听说智能评测系统进行了科学鉴定,认定为达到“国际先进”水平。

智能语音评测技术在广东高考英语听说考试中的成功应用,标志着大规模英语口语考试真正成为现实,大幅度提升英语口语考试的评分效率,可以说直接推动全国的英语口语教学向前迈出了一大步,具有巨大的社会价值。

四、科大讯飞英语听说教考平台

科大讯飞英语听说教考平台是在中高考改革背景下,在“以测促学、以学定教”思路指导下,依托科大讯飞智能语音及人工智能技术开发的,集英语听说教、学、考、评于一体的,区级英语听说教学、考试综合解决方案,目标在于满足各地市教育主管部门、学校教师和学生对于英语听说模拟考试、教学和练习的迫切需求。该平台提供丰富的教学资源 and 考试资源,支持组织区级联考、校级模考和班级日常测试;也可帮助教师开展英语听说课堂教学活动,支持学生进行个性化自主学习;还能自动汇总、统计学生的考试记录和学习记录,形成考情分析报告和学情分析报告,指导教师制定并修改教学计划,辅助教学主管部门进行教学研究和决策。整个系统采用“网站平台+客户端”的



图 5：科大讯飞英语听说教考平台架构图

设计方式，满足更为多样的实际教学需求，具体架构如图 5 所示。

科大讯飞英语听说教考平台主要分为区域管理平台、学校终端应用两个层级。区域管理平台在服务层上主要提供统一资源管理、统一用户管理及统一数据管理服务，在应用层上主要结合区域用户的业务管理需要，提供面向常态化听说考试、听说教学工作的考务管理、教学管理、评价分析和资源管理应用。校端应用分为英语听说模拟考试系统和英语听说教学系统，部署在学校机房、多媒体教室，提供具体的听说考试、听说教学场景化应用服务。

教考评价分析模块基于科大讯飞的口语智能识别、评测技术，对学生在教学和考试中的互动、答题数据进行评分，生成客观全面的成绩和分析报告。学科教师组织完成班级听说模拟测试后，教考评价分析模块自动分析班级考试情况，并生成考试讲评报告，教师可以查看班级的学生成绩报告并进行试卷讲解。学科教师完成听说教学活动后，可通过教学管理模块，查看课堂教学的互动记录及分析报告，掌握学生听说学习的整体情况及薄弱项。学生通过教考评价分析模块，可以查看自己的历次听说考试分析报告、考试成绩单、听说课后练习完成情况，以及基于课堂互动、课后练习综合评估的学情分析报告。

科大讯飞英语听说教考平台支持字、词、句、篇朗读，情景问答，话题表述等各地中高考题型；智能评分过程自动、客观、准确，无需人工参与，大大解放了老师进行听说教学和考试的负担，同时也为学生提供了良好的学习环境。

目前，英语听说教考平台已在北京、上海、辽宁、江苏、重庆、广东、山东、沈阳、湖南、浙江等全国将近 16 个省市的中高考中使用，为上百万师生提供服务。其中，北京市房山区、门头沟区、通州区、密云区、东城区等多个地区已在 2017 年先后建设了英语听说教考平台，四川省绵阳市三台县，浙江省舟山市黄岩区，北京市西城区、昌平区等地在 2018 年陆续建设了英语听说教考平台，并形成了常态化应用，有效提高了学

生的听说水平。上海市已有近百所学校使用科大讯飞英语听说教考平台,探索听说教学改革,其中包括上海外国语大学附属外国语学校、曹杨中学、敬业中学等一大批知名学校。

五、科大讯飞小学口语作业平台

现阶段,小学家长和学校注重口语教学,课上、课后的口语练习环节不可或缺。由于缺乏技术手段,老师对学生的英语口语作业评价不力,学生积极性不高,家长不会教。为解决这些痛点问题,科大讯飞依托语音评测技术,打造了英语及语文学科的在线口语作业平台,平台分为老师端和学生端,框架图如图6所示。

老师端一键发布口语作业,一键批改,自动获取学生的作答情况,高效、省时、省力,全面释放老师压力,极大地减轻了老师布置作业及批改作业负担,将老师从重复单调的工作中解放出来,将更多的精力用于教学研究与辅导。学生端完成口语作业练习后,平台即时反馈评测结果,指出发音错误,并给出针对性修正意见,学生可以立即进行改正,极大地提升学生的学习积极性。对于家长来说,可以解决不会教和不敢教的难题,改善亲子关系。

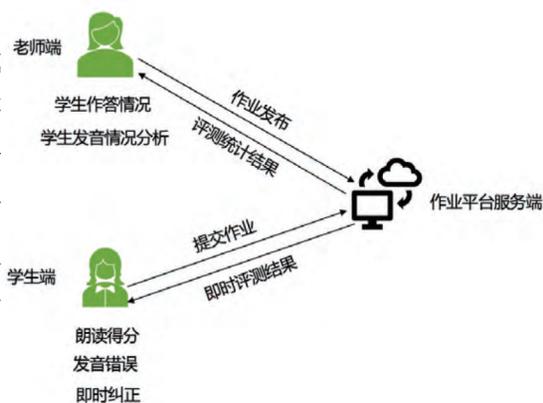


图6: 科大讯飞小学口语作业平台框图

六、总结与展望

英语口语教学与评价已成为中国英语教育必不可少的一个环节,传统的教学和评价方式与口语教学与评价需求之间的矛盾日益突出。人工智能技术的发展,特别是智能语音评测技术的发展,已经解决或者正在解决这样的矛盾。基于计算机自动评测技术,大规模的英语口语考试评价已成为现实,全国各地已逐渐将英语口语考试纳入到升学考试。在需求更为复杂的教学场景,人工智能技术开始发挥越来越重要的作用,辅助老师进行课堂英语口语教学,辅助学生进行课堂课后自我练习。智能评分和反馈只是人工智能在英语口语教学评价中一个非常小的突破,完整的教学和学习的课程体系、基于大数据技术的个性化推荐、拟人化的反馈指导、智能对话、虚拟现实的学习场景等是人工智能技术在英语口语教学与评价领域,需要继续研究和探索的方向。



查看内容精选