

doi:10.13582/j.cnki.1674-5884.2016.09.037

计算机辅助普通话水平测试系统的优化

周晓兰

(湖南科技大学 计算机科学与工程学院,湖南 湘潭 411201)

摘要:计算机辅助普通话水平测试是一种新型的普通话水平测试手段,目前湖南省的普通话水平测试采用科大讯飞的计算机辅助普通话水平测试系统,该系统基本上能适应普通话水平测试,但还存在某些不足之处。为更好地推动此项技术的发展,应优化语音评判系统,加强建设和搜集高分段录音样本,优化评分系统的评价模型,改进抽题策略,设计更加人性化的测试系统的交互信息,同时,设置一套有效防范作弊的措施,改进命题说话题的人工打分方式,设计智能的评判模式。

关键词:计算机辅助;普通话水平测试;系统优化

中图分类号:H102 **文献标志码:**A **文章编号:**1674-5884(2016)09-0116-04

计算机辅助普通话水平测试系统(Computer Assisted Mandarin Evaluation System)是针对全国普通话水平测试(PSC),以计算机作为辅助手段,进行无纸化、电子化的高效率测评的系统,由考务信息管理、仿真考试和录音以及机器辅助评分构成,具有考生信息数据库管理、考生语音数据采集、考试流程控制、计算机自动评测打分等功能,实现了考试过程、数据收集的自动化等功能。计算机辅助普通话水平测试作为一种利用现代信息技术新型普通话水平测试手段,在全国多个省市推广,给普通话教学与测试带来了巨大的变化。测前的准备工作相对减少,大大提高了灵活程度和工作效率,并降低了组织的难度和测试成本。但计算机辅助普通话水平测试中语音评判的准确度问题还存在一定的局限性。雷峻认为应进一步扩充丰富全国各地的录音样本,尤其是对高分段(一甲样本)的搜集,从而对高分段的测试更加准确。韦素玲以广西师范大学为例探讨了高校“机辅测试”普通话规范化管理的相关问题,提出应做好考前、考中、考后的规范化管理,重点提出实现“机辅测试”走向制度化、规范化和科学化。王苓、傅江宏提出一种新的测试模式,帮助测试管理机构有效监控测试员测评的整个过程,以便得到更加准确的测试成绩。何勇等利用遗传算子和遗传算法,依据遗传算法的编码、选择、交叉和变异等步骤,提出自动组卷算法。曾玲基于三明学院普通话测试改革,分析了计算机辅助普通话测试在高校实际运行过程中信息管理系统不够完善、硬件的兼容性较差等问题,提出优化软件平台,提高系统管理人性化自动化水平,充实提高机测题库容量与质量,实现计算机辅助普通话测试的规范化、科学化管理。为使系统能更加准确地评测前三题,给出更加客观的分数,对该测试系统进行优化,使之能对所有被测人员做出准确的评价,是研究普通话水平测试系统的重要任务。

1 目前计算机辅助普通话水平测试存在的不足之处

计算机辅助普通话测试的语音评判具有可行性、信度及经济效率,能有效解决人工测试的弊端,并提高测试的公信度、公正性和精确度。

科大讯飞股份有限公司开发出来的计算机辅助普通话水平测试系统是国家语委鉴定为当前唯一

的,并在普通话水平正式测试中应用的测试系统,基本能够满足当前高校计算机辅助普通话水平测试的要求。科大讯飞测试系统基于教育部和国家语言文字工作委员会颁布的《普通话水平测试大纲》,以先进的语音合成技术和中文语音评测技术为核心技术,通过对语料库补充完善、合成器的研发和语音参数的优化,以及对区域性语言的分析 and 研究,可准确地对朗读单音节词、朗读多音节词和朗读短文等三种测试题型实现自动评测。

科大讯飞系统为了满足不同的用户需求,开发了单机版和网络版两种版本。可以更加灵活地应用于计算机辅助普通话水平测试系统^[1]。网络版架构设计是监考机对考生测试过程进行管理监控。目前在大批量的测试中,普遍采用设置每批次20台测试机,辅助学生完成考试。每一批次进行考试时都由监考老师从监考机上发出考试指令(监考机能够随时监控考试机),考生根据监考老师在监考机发出的指令(指令有时是语音提示,有时是显示在考试机的屏幕上的文字信息)进行不同的操作。考生分步依次的操作是:准考证号输入,身份信息核对,试音以及单音节词、多音节词、朗读短文、说话等题型的考试。由测试系统自动完成考生的朗读单音节词、朗读多音节词和朗读短文的评测,最后一题说话题由测试站将测试数据上传至省中心服务器上并分配测试员后,由测试员根据录音进行网上评分。

科大讯飞的计算机辅助普通话测试系统在成绩评定上,对二级甲等到三级甲等成绩的判定基本上与人工测试的结果比对一致,大部分考生的成绩集中在二级甲等到三级甲等这个范围内,从而系统能满足高校普通话水平测试的要求。系统对高分段(比如一级甲等)和低分段(比如三级乙等及以下等级)的成绩判定与人工测试的误差大,特别是测试系统对一级甲等的成绩认定一般不能做出很好的评价。据国家语委测试数据显示,在实际测试以及测试环境完全一致的情况下,让性别不同、年龄不同以及获得证书时间不同的已获得国家语委普通话培训测试中心认定的一级甲等成绩的人员进行试验测试,最终成绩都达不到一级甲等。目前各地区在送审一甲样本时,仍旧采用人工测试。系统中最重要也是最关键的环节——语音评判、语音分析处理技术还是一个比较难解决的问题,对于全部自动化阅卷和对所有分数段的考生做出准确的评判还有待进一步改进。到目前为止还没有系统提出改进措施,特别是考生语音数据采集和计算机自动评测打分功能优化问题没有系统研究,另外尚未建立对考生在没有文字凭借情况下的命题说话题进行语音评判的模式。

测试系统的可靠性有待进一步加强^[2]。软件可靠性(software reliability)是软件产品能否在规定的条件和条件下满足规定的功能的能力,是国际标准化组织颁布的ISO9001制定的软件评价技术中的质量标准之一,是用户对软件产品的正常质量诉求。在使用过程中,科大讯飞测试系统随机偶发一些故障。比如:有时考生明明点击了“下一题”按钮,但就是无法结束本题考试;有时考试机试音成功后,第一题测试时麦克风无法拾音,提示框显示的提示信息为“失败原因:无输入信号”;有时考生参加了测试但从考试机导出的考试数据trd压缩包中没有以该考生准考证号命名的oef文件,从而信息管理系统中没有该考生的数据^[3]。

测试系统出现的有些交互信息不易被理解。测试系统在运行过程中经常出现用来提示发生错误原因的“试音失败、评测失败、操作失败”等信息,特别对于出现较多的“失败原因:截幅”,截幅是什么、导致截幅原因是什么以及如何克服截幅,系统没有提供任何解决方法。软件说明书中也查找不到截幅相关的说明。非专业的测试站工作人员无法理解这些信息,更谈不上去指导考生了。

测试系统对于考生本身作弊行为或者内外勾结重复考试的作弊行为基本上不能防范。普通话水平测试是国家级的口语测试,各测试站采取了各种各样的措施加强考场监督,但不能彻底杜绝测试中的作弊行为,特别是对于交叉替考作弊方式没有任何的防范措施。

2 对存在问题的对策及今后规范管理的思考

基于计算机辅助普通话水平测试现状,进一步研究计算机辅助普通话水平测试系统的优化,实现优化测试手段、规范测试程序、统一测试标准、降低测试成本、提高测试效率等,真正达到计算机智能普通话水平测试。结合计算机辅助普通话水平测试实际应用现状提出计算机辅助普通话水平测试中对现有

普通话中语音评判进行优化的方案,并建立对于考生在没有文字凭借情况下的命题说话题进行语音初步评判的模式。以系统思考的方法和开拓创新的精神,积极探索改进计算机辅助普通话水平测试系统的不足之处。

第一,优化计算机辅助普通话测试的语音评判系统。针对系统对高分段(比如一级甲等)和低分段(比如三级乙及以下等级)的成绩判定与人工测试误差大的情况,可以对计算机辅助普通话水平测试的评价模式进行修改。系统可以对被测样本进行分析和成绩评判,如果被测样本出属于普通段,则按照原有模式进行评判;如果被测样本属于高分段(或低分段)马上转入高分段(或低分段)处理模块进行处理(专门为高分段或低分段的样本设计一个处理模块)。另外,优化系统应考虑硬件的兼容性和大批量考生的管理需求,提高系统管理人性化自动化水平,确保省中心和测试站两级管理模式的有序协调。

第二,加强高分段录音样本的建设和搜集。目前测试系统利用了一种智能语音处理技术,即计算机辅助语言学习技术(Computer Assisted Language Learning,通过机器自动对发音进行评分、检错并给出矫正指导的技术)进行评分。测评系统利用大量的录音样本进行分析处理,并产生标准模型,再把被测者的测试样本与标准模型进行比对评判^[4]。尽管系统从全国各地搜集整理了许许多多的样本,但目前对高分段(一级甲等)的录音样本采集还很不足。为了丰富高分段的样本,可以将国家级的广播电视媒体和社会上影响力大的播音员、主持人的录音样本采集到系统语音库中,也可以将影视话剧等单位公认的普通话水平很标准的演员的录音样本大量地采集到系统语音库中。样本比较丰富了,评价系统在对比考生的语音面貌时,将更加客观。

第三,探索优化评分系统的评价模型的新算法。蚁群算法可以准确地估计评价模型向量并且不会产生局部相位的波形恶化,使得各处向量的功率谱有明显的增益^[5]。改进后的评价模型能对语音信号进行良好的识别,具有良好的通用性和全局性。使用蚁群算法对评分系统里评价模型进行优化的方法。利用计算机技术蚁群算法对评分系统中的评价模型进行优化,并建立对于考生在没有文字凭借情况下的命题说话题进行语音评判的模式。

第四,改进当前的抽题策略,并进一步建设测试试卷的题库。目前湖南省测试中心使用50套试题,每个考场从50套中抽取20套,每台考试机上固定一套试题。按照这种抽题策略,虽然是每个考生一套,但是坐在同一个位置上的考生总是做同一套试题,这样就可能出现考生作弊的行为。为了防止同一位置做相同的试题,可以每个座位上抽取不同的试题,另外每一套试题的组织应该由题库中不同类型的题目随机搭配组成。在被测室选取的试卷号,这个试卷号可以由监考老师输入测试系统。另外这50套测试卷的所有内容已成为公开的秘密,特别是命题说话题,已经被别人做出不同的标准答案,上传到网络之上或者辑印成册。测试员在评分中经常反复听到同一个话题的同一内容的标准答案,就是因为考生背熟了千篇一律的小文章^[1]。为了减少此类现象的发生,普通话水平测试题库必须进行修改和更新。题库的题目不应局限于已经存在的几十套单音节字词、双音节字词、短文及说话题目,应增加与考生生活、学习等方面相关的更加丰富的内容。在每次组织考试前,应由省测试中心或测试站负责组卷,这样考生就不会在考试前准备好标准答案。

第五,广泛协调,多方组织合作,大力开展技术改进,进行科研攻关。计算机辅助测试系统技术难度高,一般计算机辅助测试系统的研发人员是一些懂计算机技术的科研人员,这些人员一般不是普通话水平测试的一线测试员,更加不是语言教学、普通话理论研究专家。计算机技术人员不能准确把握普通话水平测试的用户的各种需求,所开发的软件自然就会出现这样那样的不合理之处。基于此,可以由教育部和国家语言文字工作委员会牵头,在全国范围内不定时地组织一些学术研讨会或者成立一些专门的研究机构,把语音处理技术上顶尖的专家、既懂计算机技术又十分熟悉普通话理论以及测试的复合型人才以及长期从事普通话教学和测试的骨干组织起来,大家联合攻关^[6]。使普通话测试工作从管理、测试、培训到研究真正成为与国际上著名的语言考试比肩的国家品牌。

第六,设计更加人性化的测试系统的交互信息。为了测试站工作人员、监考人员和考生都能准确理解提示信息,测试系统的提示信息一般应该做到考生和监考老师都能理解出现信息的含义,不能出现监

考老师和考生都不理解的信息。

第七,测试系统中可以设置一套有效防范作弊的措施。对于考生本身作弊行为中的请人代考的防止措施,可以在系统中增加现在市面上已经应用广泛的指纹识别系统或者面部识别系统。在考生报名时采集其指纹信息或者面部图像,考试时指纹、面部识别系统将正在测试的考生的生物特征与报名采集的信息比对,并将结果反馈到监考机屏幕上,监考老师根据结果确定考生是否舞弊^[3]。笔者认为用声纹识别技术更方便,考试系统不要配置额外的硬件,只要在考生报名时采集一段考生的声音,考试时候直接把报名时采集的声纹与考生当场考试的声纹进行比对即可分清是否替考。对于内外勾结重复考试的作弊行为,应由测试工作站的工作人员监视整个考场,防止发生这种情况。

第八,改进命题说话题的人工打分方式,设计智能的评判模式。测试系统对于考生在没有文字凭借情况下的命题说话题的处理还处在人工方式阶段。系统只能把考生的测试数据保存下来,而后上传至省中心服务器上,由测试系统把测试数据分发给普通话测试员,测试员根据录音进行人工打分、网上评分。基于现在语音评测技术的限制,不能完全由测试系统对命题说话题进行精确的判分,但可对当前的命题说话题进行简单的预评判,把一些没有说话的、乱说话的、完全不符合题意的剔除掉,减轻测试员的工作。

3 结语

在计算机辅助普通话水平测试应用过程中,虽然各个测试中心遇到了这样或者那样的问题,但是在教育部和国家语言文字工作委员会的指导下测试系统会不断完善。将来计算机辅助普通话水平测试系统肯定会向着智能化的方向发展,会有更加广阔的应用前景。

参考文献:

- [1] 黄谦,张小俊,赵秀宏.计算机辅助普通话水平测试存在的问题及对策[J].廊坊师范学院学报(自然科学版),2011(8):122-124.
- [2] 杨康.国家普通话水平智能测试系统的改进分析[J].考试研究,2015(4):98-101.
- [3] 杨康.ISO9001标准下的国家普通话水平智能测试系统(2.8版)改进建议[J].镇江高专学报,2015,10(4):39-40.
- [4] 雷峻.计算机辅助普通话测试的问题思考及技术对策[J].武汉理工大学学报,2010(7):160-162.
- [5] 李超,刘涛.基于蚁群算法的普通话测试系统评分机制改进[J].玉溪师范学院学报,2011(8):59-62.
- [6] 杨月君,王东波.计算机辅助普通话水平测试运行实践研究[J].教育教学论坛,2014(6):267-268.

(责任编辑 莫秀珍)