(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 103151042 B
(45)授权公告日 2016.02.24

(21)申请号 201310034371.7

(22)申请日 2013.01.29

(66)本国优先权数据
201310027676.5 2013.01.23 CN

(73)专利权人 中国科学院深圳先进技术研究院
地址 518055 广东省深圳市南山区西丽大学城学苑大道1068号

(72)发明人 王岚 宋阳 陈蒙 金晓虎 李阳

(74)专利代理机构 深圳市科进知识产权代理事务所(普通合伙) 44316
代理人 宋鹰武

(51)Int.Cl.
G10L 15/30(2013.01)
G10L 15/02(2006.01)
G10L 15/19(2013.01)
G10L 25/51(2013.01)

(56)对比文件
CN 102509483 A,2012.06.20,
CN 102509483 A,2012.06.20,
CN 101740024 A,2010.06.16,
US 2010/0185435 A1,2010.07.22,
CN 101393103 A,2009.03.11,
CN 102034475 A,2011.04.27,
CN 101551947 A,2009.10.07,
CN 101246685 A,2008.08.20,

审查员 游晓梅

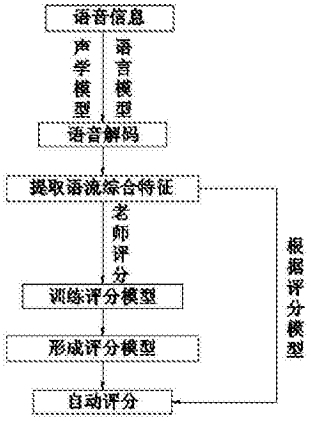
权利要求书2页 说明书9页 附图2页

(54)发明名称

全自动口语评测管理与评分系统及其评分方法

(57)摘要

本发明涉及计算机辅助教育领域，提供一种全自动口语评测管理与评分系统，其包括依次连接的客户端、落地服务器以及网络服务器，所述考生作答结果包括朗读题的口语测评结果以及自发性口语表述题的口语测评结果中的一种，所述落地服务器还包括评分装置，该评分装置包括识别模块和评分模块，所述识别模块包括声学子模块、语言子模块以及识别子模块；所述评分模块包括特征提取子模块和评分子模块。本发明还提供一种全自动口语评测管理与评分方法。本发明的全自动口语评测管理与评分系统及其评分方法可以完全由计算机自动控制考生的准备答题时间和作答时间，无需人工干预，并且比人工计时更为准确，进一步确保了评测的公平性。



CN 103151042 B

1.一种全自动口语评测管理与评分系统，其包括依次连接的客户端、落地服务器以及网络服务器，其中，网络服务器主要负责评测结果的整理、收集以及试卷的分发，落地服务器主要负责机器自动评分，客户端主要负责评测，试卷从网络服务器通过落地服务器分发到客户端，考生作答结果从客户端通过落地服务器上传到网络服务器，所述考生作答结果包括朗读题的口语测评结果以及自发性口语表述题的口语测评结果中的一种，其特征在于：所述落地服务器还包括评分装置，该评分装置包括识别模块和评分模块，所述识别模块包括声学子模块、语言子模块以及识别子模块，所述声学子模块提取考生作答音频的声学特征得到声学模型，所述语言子模块根据题目信息及训练文本得到语言模型，所述识别子模块通过声学模型和语言模型对考生作答音频进行解码得到识别结果；所述评分模块包括特征提取子模块和评分子模块，所述特征提取子模块用于提取所述识别结果中的语流综合特征，所述语流综合特征包括口语测评中发音准确度方向的特征、流利度方向的特征以及文本语义相似度方向的特征，所述文本语义相似度方向的特征包括语义相关度特征以及语法结构相似度特征；

所述特征提取子模块用于提取所述语义相关度特征，包括：计算识别结果中的每一个单词和标准答案中每一个单词的语义相似度得分；计算识别结果中的每一个单词和标准答案中每一个句子的语义相似度得分；计算识别结果中每一个单词和标准答案中每一个句子中的语义相似度得分最大值或者平均值作为单词与句子之间的相似度得分；计算考生答案和标准答案之间的相似度得分；

所述特征提取子模块用于提取所述语法结构相似度特征，包括：分别为识别结果的每一个句子建立语法顺序矢量；分别求出识别结果中的每一个句子和标准答案中的每一个句子的语法结构相似度得分，取识别结果中每一个句子语法结构相似度得分最大值作为这个句子的语法结构相似度得分；通过对识别结果中每一个句子语法结构相似度得分加权平均计算考生答案和标准答案之间的语法结构相似度特征；

所述识别模块采用基于大规模连续语音识别的解码系统，所述声学模型采用基于隐马尔科夫模型，所述语言模型采用基于元文法的语言模型，在进行解码时采用基于多遍解码技术，所述多遍解码包括直接解码、基于最大线性似然回归的非监督性自适应及二次解码；所述评分子模块对语流综合特征进行评分训练，得到评分模型，并依据评分模型对识别结果进行评分。

2.如权利要求1所述的全自动口语评测管理与评分系统，其特征在于：所述特征提取子模块还用于提取所述发音准确度方向的特征，包括：

将识别结果与正确文本进行强制对齐，计算每个音素的强制对齐得分；

构建单音素解码模型并解码每个音素，计算每个音素的最大似然得分；

利用强制对齐得分与最大似然得分的差值得到发音准确度方向的特征。

3.如权利要求1所述的全自动口语评测管理与评分系统，其特征在于：网络服务器包括调度模块，用于在落地服务器与网络服务器之间调度评测信息；采用联网评测模式，使得分布在不同地点的考场之间落地服务器的通信由所述网络服务器调度模块统一调度管理。

4.如权利要求1所述的全自动口语评测管理与评分系统，其特征在于：所述系统包含三种不同权限的角色：考生、教师和管理员，考生主要负责评测与作答；教师主要负责制卷、发布评测、管理评测、检查评测结果以及评分工作，评分方面采用系统评分和老师评分

相结合的方法；管理员主要负责评测的管理以及试卷分发的时间掌控。

5. 一种全自动口语评测管理与评分方法，其特征在于：其包括如下几个步骤：

A0、选取若干考生进行如步骤 A1 ～ A5 所述的过程，然后将语流综合特征与老师评分结合进行自动评分模型训练，形成评分模型；

A1、收集考生作答音频；

A2、提取考生作答音频的声学特征得到声学模型，并根据题目信息及训练文本得到语言模型；

A3、根据建立的声学模型和语言模型对考生作答音频进行解码得到识别结果；

A4、提取识别结果中的语流综合特征，所述语流综合特征包括口语测评中发音准确度方向的特征、流利度方向的特征以及文本语义相似度方向的特征；所述文本语义相似度方向的特征包括语义相关度特征以及语法结构相似度特征；

获取所述语义相关度特征包括以下步骤：针对考生作答音频进行解码识别，得到识别结果；计算识别结果中的每一个单词和标准答案中每一个单词的语义相似度得分；计算识别结果中的每一个单词和标准答案中每一个句子的语义相似度得分；计算识别结果中每一个单词和标准答案中每一个句子中的语义相似度得分最大值或者平均值作为单词与句子之间的相似度得分；计算考生答案和标准答案之间的相似度得分；

获取所述语法结构相似度特征包括以下步骤：针对考生作答音频进行解码识别，得到识别结果；分别为识别结果的每一个句子建立语法顺序矢量；分别求出识别结果中的每一个句子和标准答案中的每一个句子的语法结构相似度得分，取识别结果中每一个句子语法结构相似度得分最大值作为这个句子的语法结构相似度得分；通过对识别结果中每一个句子语法结构相似度得分加权平均计算考生答案和标准答案之间的语法结构相似度特征；

采用基于大规模连续语音识别的解码系统进行解码识别，所述声学模型采用基于隐马尔科夫模型，所述语言模型采用基于元文法的语言模型，在进行解码时采用基于多遍解码技术，所述多遍解码包括直接解码、基于最大线性似然回归的非监督性自适应及二次解码；

A5、根据语流综合特征形成的评分模型进行自动评分。

全自动口语评测管理与评分系统及其评分方法

技术领域

[0001] 本发明涉及计算机辅助教育领域，尤其涉及一种全自动口语评测管理与评分系统及其评分方法。

背景技术

[0002] 随着我国英语教学的普及，哑巴英语的问题也越来越严重，因此英语口语的教学与评测也日益受到重视。英语的口语评测已经在我国多个省市开展起来，其中广东省已经把英语口语评测纳入到了普通高考的必考范围。由于口语评测方式与内容的特殊性，目前英语口语的评分主要还是通过人工判分的方式来解决。人工评分有着很多很难克服的缺点：

[0003] 老师与学生面对面口语评测缺乏公平性与准确性。不同老师针对同一考生的评分很难保证完全一致，同一老师针对不同考生的评分标准也很难保证完全一致，这一切都将极大的降低评测的公平性。本系统制定一套大家都认可的公正统一的机器自动评分标准，从而可避免因人工音素而造成的评分差异，确保了评测的公平性。

[0004] 人工评分口语评测评分效率十分低下。本发明设计的一套全自动全自动口语评测管理与评分系统，以机器评分代替人工评分，从评测试卷的分发，到评测的开始和结束，到考生答案音频的保存，再到自动评分全部是由机器自动完成的，评分效率大大提高。

[0005] 传统的口语评测组织管理十分复杂。传统的人对人口语评测由于采用面试的形式，需要大量经过专业培训的口语考官；与此同时，每场评测只能对几名学生进行考评，如果要对全年级甚至全校学习英语课程的学生开展一次口语评测，不但费时耗力还存在漏题泄题的风险。本发明设计的一套全自动全自动口语评测管理与评分系统可以同时为全年级甚至全校学生同时开展一次公平、安全、高效的英语口语水平评测，并可极大地节约人力物力，还可避免因为分批次评测而带来的泄题风险。

[0006] 传统口语评测需要人工控制评测开始和结束的时间，每次更换题目，甚至每道小题都需要重新计时，非常麻烦。

发明内容

[0007] 本发明为解决上述技术问题，提供一种无需人工干预、计时更为准确且一步确保评测公平性的全自动口语评测管理与评分系统，其包括依次连接的客户端、落地服务器以及网络服务器，其中，网络服务器主要负责评测结果的整理、收集以及试卷的分发，落地服务器主要负责机器自动评分，客户端主要负责评测，试卷从网络服务器通过落地服务器分发到客户端，考生作答结果从客户端通过落地服务器上传到网络服务器，所述考生作答结果包括朗读题的口语测评结果以及自发性口语表述题的口语测评结果中的一种，所述落地服务器还包括评分装置，该评分装置包括识别模块和评分模块，所述识别模块包括声学子模块、语言子模块以及识别子模块，所述声学子模块提取考生作答音频的声学特征得到声学模型，所述语言子模块根据题目信息及训练文本得到语言模型，所述识别子模块通过声

学模型和语言模型对考生作答音频进行解码得到识别结果；所述评分模块包括特征提取子模块和评分子模块，所述特征提取子模块用于提取所述识别结果中的语流综合特征，所述语流综合特征包括口语测评中发音准确度方向的特征、流利度方向的特征以及文本语义相似度方向的特征，所述评分子模块对语流综合特征进行评分训练，得到评分模型，并依据评分模型对识别结果进行评分。

[0008] 优选地，所述发音准确度方向的特征通过以下步骤获得：

[0009] 针对考生作答音频进行解码识别，得到识别结果；

[0010] 将识别结果与正确文本进行强制对齐，计算每个音素的强制对齐得分；

[0011] 构建单音素解码模型并解码每个音素，计算每个音素的最大似然得分；

[0012] 利用强制对齐得分与最大似然得分的差值得到发音准确度方向的特征。

[0013] 优选地，所述流利度方向的特征包括语速特征以及短时停顿时长特征，所述语速特征通过以下步骤获得：

[0014] 针对考生作答音频进行解码识别，得到识别结果；

[0015] 根据识别结果统计出语音识别结果中每个音素对应的帧数；

[0016] 利用音素总个数与所有音素的持续时间的比值得到语速特征；

[0017] 所述短时停顿时长特征通过以下步骤获得：

[0018] 针对考生作答音频进行解码识别，得到识别结果；

[0019] 根据识别结果统计出语音识别结果中每个音素对应的帧数和音频总的帧数；

[0020] 利用所有音素短时停顿的时长的总和与总的发音时长的比值得到短时停顿时长特征。

[0021] 优选地，所述文本语义相似度方向的特征包括语义相关度特征以及语法结构相似度特征。

[0022] 优选地，所述语义相关度特征包括以下步骤：

[0023] 针对考生作答音频进行解码识别，得到识别结果；

[0024] 计算识别结果中的每一个单词和标准答案中每一个单词的语义相似度得分；

[0025] 计算识别结果中的每一个单词和标准答案中每一个句子的语义相似度得分；

[0026] 计算识别结果中每一个单词和标准答案中每一个句子中的语义相似度得分最大值或者平均值作为单词与句子之间的相似度得分；

[0027] 计算考生答案和标准答案之间的相似度得分。

[0028] 优选地，所述语法结构相似度特征包括以下步骤：

[0029] 针对考生作答音频进行解码识别，得到识别结果；

[0030] 分别为识别结果的每一个句子建立语法顺序矢量；

[0031] 分别求出识别结果中的每一个句子和标准答案中的每一个句子的语法结构相似度得分，取识别结果中每一个句子语法结构相似度得分最大值作为这个句子的语法结构相似度得分；

[0032] 通过对识别结果中每一个句子语法结构相似度得分加权平均计算考生答案和标准答案之间的语法结构相似度特征。

[0033] 优选地，网络服务器包括调度模块，用于在落地服务器与网络服务器之间调度评测信息；采用联网评测模式，使得分布在不同地点的考场之间落地服务器的通信由所述网

络服务器调度模块统一调度管理。

[0034] 优选地，所述系统包含三种不同权限的角色：考生、教师和管理员，考生主要负责评测与作答；教师主要负责制卷、发布评测、管理评测、检查评测结果以及评分工作，评分方面采用系统评分和老师评分相结合的方法；管理员主要负责评测的管理以及试卷分发的时间掌控。

[0035] 优选地，对于角色扮演题可以利用每道小题已经给出的参考答案作为文本语义相似度分析的参考文本，进而扩展出其它参考答案；对于复述题可以利用音频内容的转写文本作为文本语义相似度分析的参考文本，进而扩展出其它参考答案。

[0036] 本发明另外提供一种全自动口语评测管理与评分方法，其包括如下几个步骤：

[0037] A0、选取若干考生进行如步骤A1～A5所述的过程，然后将所述语流综合特征与老师评分结合进行自动评分模型训练，形成评分模型；

[0038] A1、收集考生作答音频；

[0039] A2、提取考生作答音频的声学特征得到声学模型，并根据题目信息及训练文本得到语言模型；

[0040] A3、根据建立的声学模型和语言模型对考生作答音频进行解码得到识别结果；

[0041] A4、提取识别结果中的语流综合特征，所述语流综合特征包括口语测评中发音准确度方向的特征、流利度方向的特征以及文本语义相似度方向的特征；

[0042] A5、根据语流综合特征形成的评分模型进行自动评分。

[0043] 本发明的全自动口语评测管理与评分系统及其评分方法：在传统的朗读题口语测评基础上增加了文本语义相似度方向的特征，从而可以针对自发性口语表述进行测评；设计了独特的云端评测系统架构，确保评测高效进行，同时充分利用了整个系统的资源，大大提升了组织口语评测效率，节约了大量的人力物力；可以在大范围同时组织口语评测，满足了省市范围内甚至全国范围内同时组织口语评测的需求；口语评测的形式也将更加多元化，对考生口语水平的考察更为全面公正。

附图说明

[0044] 图1为本发明一实施例全自动口语评测管理与评分方法流程示意图；

[0045] 图2为本发明一实施例全自动口语评测管理与评分系统架构示意图；

[0046] 图3为本发明一实施例语义相关度特征提取流程示意图；

[0047] 图4为本发明另一实施例语义相关度特征提取流程示意图；

[0048] 图5为本发明一实施例语法结构相似度特征提取流程图。

具体实施方式

[0049] 下面将结合附图以及具体实施例来对本发明作进一步详细说明。

[0050] 实施例：

[0051] 本发明提供一种全自动口语评测管理与评分系统，其包括依次连接的客户端、落地服务器以及网络服务器，其中，网络服务器主要负责评测结果的整理、收集以及试卷的分发，落地服务器主要负责机器自动评分，客户端主要负责评测，试卷从网络服务器通过落地服务器分发到客户端，考生作答结果从客户端通过落地服务器上传到网络服务器，所述考

生作答结果包括朗读题的口语测评结果以及自发性口语表述题的口语测评结果中的一种，所述落地服务器还包括评分装置，该评分装置包括识别模块和评分模块，所述识别模块包括声学子模块、语言子模块以及识别子模块，所述声学子模块提取考生作答音频的声学特征得到声学模型，所述语言子模块根据题目信息及训练文本得到语言模型，所述识别子模块通过声学模型和语言模型对考生作答音频进行解码得到识别结果；所述评分模块包括特征提取子模块和评分子模块，所述特征提取子模块用于提取所述识别结果中的语流综合特征，所述语流综合特征包括口语测评中发音准确度方向的特征、流利度方向的特征以及文本语义相似度方向的特征，所述评分子模块对语流综合特征进行评分训练，得到评分模型，并依据评分模型对识别结果进行评分。

[0052] 如图 1 所示，为本发明一实施例全自动口语评测管理与评分方法流程示意图，其包括如下几个步骤：

[0053] A0、选取若干考生进行如步骤 A1 ～ A5 所述的过程，然后将所述语流综合特征与老师评分结合进行自动评分模型训练，形成评分模型；

[0054] A1、收集考生作答音频；

[0055] A2、提取考生作答音频的声学特征得到声学模型，并根据题目信息及训练文本得到语言模型；

[0056] A3、根据建立的声学模型和语言模型对考生作答音频进行解码得到识别结果；

[0057] A4、提取识别结果中的语流综合特征；

[0058] A5、根据语流综合特征形成的评分模型进行自动评分。

[0059] 在上述实施例中，所述语言模型包括话题相关文本、书面相关文本以及口语风格文本的一种或多种。所谓的话题相关文本是指确定一个话题，标准答案可以给考生看到或者不给考生看到，在评测时由考生根据话题以及标准答案形成的文本；所谓的书面相关文本是指考生按照标准答案形成的语音文本；所谓口语风格文本是指没有参考任何标准答案，完全由考生自由发挥形成的语音文本，所述老师评分是指聘请专业的老师对考生作答音频进行评分，实行双评，即以两位评卷老师给出成绩的平均分为该考生的最终成绩，如果两位评卷老师给出的成绩相差 3 分以上，采取三评方式。

[0060] 如图 2 所示，为本发明一实施例全自动口语评测管理与评分系统架构示意图，其主要分为以下三大部分：

[0061] 客户端：电脑、平板电脑、高端智能手机均可以作为本发明的客户端，客户端必须具备独立的音频视频播放功能和独立的音频输入设备。所述客户端包括评测模块以及发布模块，所述评测模块用于考生进行口语评测，包括试题发布、评测、收卷等，并处理和传输考生作答音频至落地服务器，对于考生作答音频，评测模块还设有初处理模块用于对考生作答音频的格式转换与特征提取。评测结束后，考生的评测成绩也将通过发布模块发布在客户端上。

[0062] 落地服务器：落地服务器指的是安装在考场的高性能计算机，落地服务器包括通讯模块、采集模块以及评分模块，所述通讯模块用于将接收于网络服务器的评测信息传输给客户端，其在特定时间提供试卷给客户端并控制评测时间，所述采集模块用于从客户端收集考生作答音频，所述评分模块用于对考生答卷进行识别、解码、评分，评分完成后将评测结果通过通讯模块及时反馈回客户端。根据考场规模和计算任务量，落地服务器可以选

择多台高性能计算机组建计算机集群的形式，以便加快评分与解码的速度。

[0063] 网络服务器：网络服务器主要负责集中控制分散在各个考场的落地服务器，其包括调度模块、分析模块以及查询模块，网络服务器包括调度模块，用于在落地服务器与网络服务器之间调度评测信息；采用联网评测模式，使得分布在不同地点的考场之间落地服务器的通信由所述网络服务器调度模块统一调度管理，统一将试卷分发给各个落地服务器并控制其评测开始和结束的时间，这样做的目的是便于在大范围同时组织口语评测，这样每组织一次评测，就无需提前为每个考生准备一份考卷，网络服务器会将考卷分发给各个落地服务器，落地服务器会将考卷再分发给每个客户端，极大地降低了泄题的风险；所述分析模块用于评测结束后对考生作答信息以及其得分情况做集中的分析与处理，将考生总分、单项得分与排名等信息统计出来，所述查询模块用于老师和学生们随时查询考生总分、单项得分与排名等信息。

[0064] 所述系统包含三种不同权限的角色：考生、教师和管理员，考生主要负责评测与作答；教师主要负责制卷、发布评测、管理评测、检查评测结果以及评分工作，评分方面采用系统评分和老师评分相结合的方法；管理员主要负责评测的管理、试卷分发的时间掌控以及整体评测系统的维护。

[0065] 本实施例中评分装置所述识别模块采用基于大规模连续语音识别的解码系统，声学模型采用基于隐马尔科夫模型的声学模型，语言模型采用的是基于 N 元文法的语言模型。多遍解码时，由于未知考生作答内容所以采用直接解码、基于最大线性似然回归的非监督性自适应、二次解码等处理。另外解码时会采用多编解码技术以提高识别率。解码完成后，通过评分模块从识别结果中提取需要的语流综合特征，训练评分模型，最后根据评分模型对新来的音频进行机器自动评分。

[0066] 本实施例全自动口语评测管理与评分系统的核心算法采用基于模式识别的机器学习方法，所述基于模式识别的机器学习采用线性回归方法、基于事例推理的学习、关联规则学习、神经网络或者支持向量机中的一种。

[0067] 上述实施例中所述发音准确度方向的特征通过以下步骤获得：

[0068] 针对考生作答音频进行解码识别，得到识别结果；

[0069] 将识别结果与正确文本进行强制对齐，计算每个音素的强制对齐得分；

[0070] 构建单音素解码模型并解码每个音素，计算每个音素的最大似然得分；

[0071] 利用强制对齐得分与最大似然得分的差值得到发音准确度方向的特征。

[0072] 所述发音准确度方向的特征具体实施例：

[0073] 用 O 表示语音学的声学特征，p 表示语音学所对应的发音音素，Q 表示所有的音素集合。则给定声学特征 O 的前提下，音素 p 的后验概率为：

[0074] $PP = P(p|o) = P(o|p)P(p) / \sum_{q \in Q} P(Q|q)P(q)$

[0075] 为了方便计算，做出如下假设：所有的音素都是等概率出现的，既有 $P(p) = P(q)$，并且分母中的求和可以用最大估计值，这样就得到了发音准确度的定义：

[0076] $GOP(p) = \log(P(o|p)/MAX_{q \in Q^P}(o|q))$。

[0077] 特征提取：发音准确度可以通过以下公式得到

[0078] $GOP = S_{FA} - S_{PL}$。

[0079] 上述实施例中所述流利度方向的特征包括语速特征以及短时停顿时长特征，所述

语速特征通过以下步骤获得：

[0080] 针对考生作答音频进行解码识别，得到识别结果；

[0081] 根据识别结果统计出语音识别结果中每个音素对应的帧数；

[0082] 利用音素总个数与所有音素的持续时间的比值得到语速特征；

[0083] 所述短时停顿时长特征通过以下步骤获得：

[0084] 针对考生作答音频进行解码识别，得到识别结果；

[0085] 根据识别结果统计出语音识别结果中每个音素对应的帧数和音频总的帧数；

[0086] 利用所有音素短时停顿的时长的总和与总的发音时长的比值得到短时停顿时长特征。

[0087] 关于所述语速特征具体实施例：

[0088] 在第二语言学习中，语速特征（ROS）能很好的表征说话的流畅度和口语的熟练水平，对于初学者尤其如此。用 T_S 表示所有的音素的持续时间，N_P 表示音素总个数，则有：

[0089] $ROS = N_p/T_S$。

[0090] 关于所述短时停顿特征具体实施例：

[0091] 语音中词与词之间的短时停顿也反映了发音流利度。一般来说，停顿时间越长，对应的流利度也越差。用 D_{SP} 代表停顿时占总长的比重，T 代表总的发音时常，T_i 代表第 i 个停顿的时常，N_{SP} 代表短时停顿的总个数，则有：

[0092]

$$D_{SP} = \sum_{i=1}^{N_{SP}} Ti / T \quad 。$$

[0093] 上述实施例中所述文本语义相似度方向的特征包括语义相关度特征以及语法结构相似度特征。

[0094] 所述语义相关度特征包括以下步骤：

[0095] 针对考生作答音频进行解码识别，得到识别结果；

[0096] 计算识别结果中的每一个单词和标准答案中每一个单词的语义相似度得分；

[0097] 计算识别结果中的每一个单词和标准答案中每一个句子的语义相似度得分；

[0098] 计算识别结果中每一个单词和标准答案中每一个句子中的语义相似度得分最大值作为单词与句子之间的相似度得分；

[0099] 计算考生答案和标准答案之间的相似度得分。

[0100] 所述语义相似度特征还可以通过以下步骤获得：

[0101] 针对考生作答音频进行解码识别，得到识别结果；

[0102] 计算识别结果中的每一个单词和标准答案中每一个单词的语义相似度得分；

[0103] 计算识别结果中的每一个单词和标准答案中每一个句子的语义相似度得分；

[0104] 计算识别结果中每一个单词和标准答案中每一个句子语义相似度得分平均值作为单词与句子之间的相似度得分；

[0105] 计算考生答案和标准答案之间的相似度得分。

[0106] 所述语法结构相似度特征包括以下步骤：

[0107] 针对考生作答音频进行解码识别，得到识别结果；

[0108] 分别为识别结果的每一个句子建立语法顺序矢量；

[0109] 分别求出识别结果中的每一个句子和标准答案中的每一个句子的语法结构相似

度得分，取识别结果中每一个句子语法结构相似度得分最大值作为这个句子的语法结构相似度得分；

[0110] 通过对识别结果中每一个句子语法结构相似度得分加权平均计算考生答案和标准答案之间的语法结构相似度特征。

[0111] 所述语义相关度特征具体实施例：

[0112] 如图 3 所示，为本发明一实施例语义相关度特征提取流程图，语义相关度能够极大的反应考生答案和标准答案之间的语义相关程度，一般来讲两句话的语义相关度越高，考生的答案也就越接近标准答案。词与词之间的语义相似度可以通过词网络 (WordNet) 来计算，WordNet 是一个覆盖范围宽广的英语词汇语义网。名词、动词、形容词和副词各自被组织成一个同义词的网络，每个同义词集合都代表一个基本的语义概念，并且这些集合之间也由各种关系连接(一个多义词将出现在它的每个意思的同义词集合中)。以扩展语义覆盖的方法来计算两个单词之间的语义相似度为例，例如我们要计算英语中 drawing 和 decal 的语义相似度，首先我们要通过 WordNet 来获得两个单词的语义：

[0113] drawing:paper that is specially prepared for use in drafting。

[0114] decal:the art of transferring designs from specially prepared paper to a wood or glass or metal surface。

[0115] 两个单词的语义中共有三个单词的语义覆盖，其中 paper 为一元词我们积 1 分，而 specially prepared 为一个二元词组我们积 4 分，drawing 和 decal 之间的语义相似度为 5 分。这样我们就形成了一个打分机制：

[0116] 1) 两个单词之间的语义中出现一个一元词我们积一分，出现一个 n 元词组我们积 n 的平方分，因为出现一个 n 连词的概率要远远比同时出现 n 个一元词的概率低。通常在 WordNet 中每个单词都会包含多种语义。

[0117] 用 ω，ω’ 分别表示两个单词 S(ω) 表示单词 ω 在词网络中所有语义，c_i表示单词 ω 在词网络中的第 i 个语义，在计算两个单词的相似度时取两个单词间相似度得分最大值的两个单词间的语义。

[0118] $SCORE_{lesk}(\omega, \omega') = MAX_{ci} \in s(\omega')c_j \in s(\omega')rel(c_i c_j)$。

[0119] 有了词与词之间的语义相似度后，就可以计算词与句子之间的相似度得分了，取单词与句子里单词之间的相似的得分平均值作为单词与句子之间的相似度得分，用 |U| 表示句子 U 里的单词的个数。

[0120] $$SCORE_W(\omega, U) = \frac{1}{|U|}\sum_{\omega \in u} SCORE_W(\omega, \omega)$$

[0121] 最后就可以计算考生答案 U 和标准答案 P 之间的相似度得分了，用 uniq(U) 表示考生答案的去重单词数，则 U 与 P 之间的相似度得分为：

[0122] $$SCORE_U(U, P) = \frac{1}{|U|}\sum_{\omega \in u} SCORE_W(\omega, P)uniq(U)$$。

[0123] 如图 4 所示，为本发明另一实施例语义相关度特征提取流程图，语义相关度能够极大的反应考生答案和标准答案之间的语义相关程度，一般来讲两句话的语义相关度越高，考生的答案也就越接近标准答案。用 ω，ω’ 分别表示两个单词 S(ω) 表示单词 ω 在词网络中所有语义，c_i表示单词 ω 在词网络中的第 i 个语义，在计算两个单词的相似度时

取两个单词间相似度得分最大值的两个单词间的语义。

[0124] 有了词与词之间的语义相似度后，就可以计算词与句子之间的相似度得分了。取单词与句子里单词之间的相似的得分最大值作为单词与句子之间的相似度得分，用U表示句子。

[0125] $SCORE_W(\omega,U) = \max_{\omega i} \in {}_u SCORE(\omega, \omega i)$。

[0126] 最后就可以计算考生答案U和标准答案P之间的相似度得分了。

[0127] 所述语法结构相似度特征具体实施例：

[0128] 如图5所示，为本发明一实施例语法结构相似度特征提取流程图。

[0129] 考虑以下两个句子：

[0130] T_1: $\omega_1\omega_2\omega_3\omega_4\omega_5\omega_6\omega_7\omega_8\omega_9$。

[0131] T_2: $\omega_1\omega_2\omega_3\omega_9\omega_5\omega_6\omega_7\omega_8\omega_4$。

[0132] 句子T_1和T_2所包含的单词是一样的，如果单纯的通过两句话中所包含的单词进行语义相似度分析，那么得出的结果将是T_1与T_2两句话所表达的语意是完全一样的。两个句子的词序是不一样的，这里将提出一种通过词序进行句法结构相似度分析的方案，我们对T_1和T_2中每一个单词都进行重新编号，比如T_1中dog的编号是4，over的编号是6。定义句子T_1与T_2语法顺序矢量为r_1和r_2则有：

[0133] r_1=（123456789）

[0134] r_2=（123956784）

[0135] 那么句子T_1与T_2的语法结构相似度可以通过以下公式进行计算：

[0136] $$S_r = 1 - \frac{\|r_1 - r_2\|}{\|r_1 + r_2\|}$$。

[0137] 实际情况中T_1与T_2的所包含的单词并不一定是完全一样的，这里需要定义一个句子之间的单词对应规则。

[0138] 如果在T_2中的一个词ω_i也包含在T_1当中，我们将该单词在句子T_1中第一次出现的编号作为它在T_2中的编号。否则，我们就要在T_1中寻找与ω_i语义最接近的词ϖ_i。

[0139] 如果ω_i和ϖ_i的相似度超过预先设定的阈值，那么ω_i在T_1中的编号就设为ϖ_i在T_2中的编号。

[0140] 如果前两种搜索都失败了，那么ω_i的编号就设为0。

[0141] 根据以上规则，我们可以提取得到语法结构相似度特征。

[0142] 利用公式$S_r = 1 - \frac{\|r_1 - r_2\|}{\|r_1 + r_2\|}$分别求出识别结果中的每句话和标准答案中的每句话的语法结构相似度，取识别结果中每句话语法结构相似度最大值作为其语法结构相似度。

[0143] 最终考生答案识别结果和标准答案之间的语法结构相似度S_{REC}可以表示为

[0144] $$S_{REC} = \frac{1}{M}\sum_{T1 \in REC} MAX_{T2 \in REF} S_r(T_1, T_2)$$。

[0145] 其中M为每个考生识别结果的句子数，REC代表识别结果，REF代表参考答案。

[0146] 得到了以上语流综合特征之后我们需要调整特征权重使评分模型效果最好，可以基于期望最大化算法进行特征权重优化。

[0147] 在以上任一实施例中，针对客观性试题没有固定参考答案的问题，我们提出了一

种基于文本语义相似度分析的答案扩展方案：对于角色扮演题可以利用每道小题已经给出的参考答案作为文本语义相似度分析的参考文本，进而扩展出其它参考答案；对于复述题可以利用音频内容的转写文本作为文本语义相似度分析的参考文本，进而扩展出其它参考答案，考生不一定完全按参考答案进行作答，所以不能进行强制对齐。解码过程包括行初次解码，说话人自适应及二次解码识别；二次解码识别完成以后从识别结果中提取语流综合特征。然后将语流综合特征与老师评分进行评分模型训练，得到评分模型，进行自动评分。

[0148] 本发明的全自动口语评测管理与评分系统无需预知考生作答文本，考生可以根据话题内容进行自发性口语表述，评分前只需知道考生作答话题相关文本即可评分；相对于传统朗读题口语测评，本发明不但能够进行朗读题的口语测评，还可以进行自发性口语表述题的口语测评；评分更加全面公正，可以在考生自发性表述的情况下考察考生的发音准确度与流利度，更能反映考生实际的口语水平；考生作答文本不再受到限制，自动评分评测题型也将不再仅限于朗读题，这样就可以考察考生在自发性口语表述的情况下对语言的理解、运用与表达能力；语流综合特征中增加了语义相关度特征以及语法结构相似度特征，这样既能考察考生口语运用的语义相关性，又能考察考生口语的语法水平，设计了独特的云端评测系统架构，确保评测高效进行，同时充分利用了整个系统的资源，大大提升了组织口语评测效率，节约了大量的人力物力；可以在大范围同时组织口语评测，满足了省市范围内甚至全国范围内同时组织口语评测的需求；口语评测的形式也将更加多元化，对考生口语水平的考察更为全面公正。

[0149] 可以理解的是，对于本领域的普通技术人员来说，可以根据本发明的技术构思做出其他各种相应的改变与变形，而所有这些改变与变形都应属于本发明权利要求的保护范围。

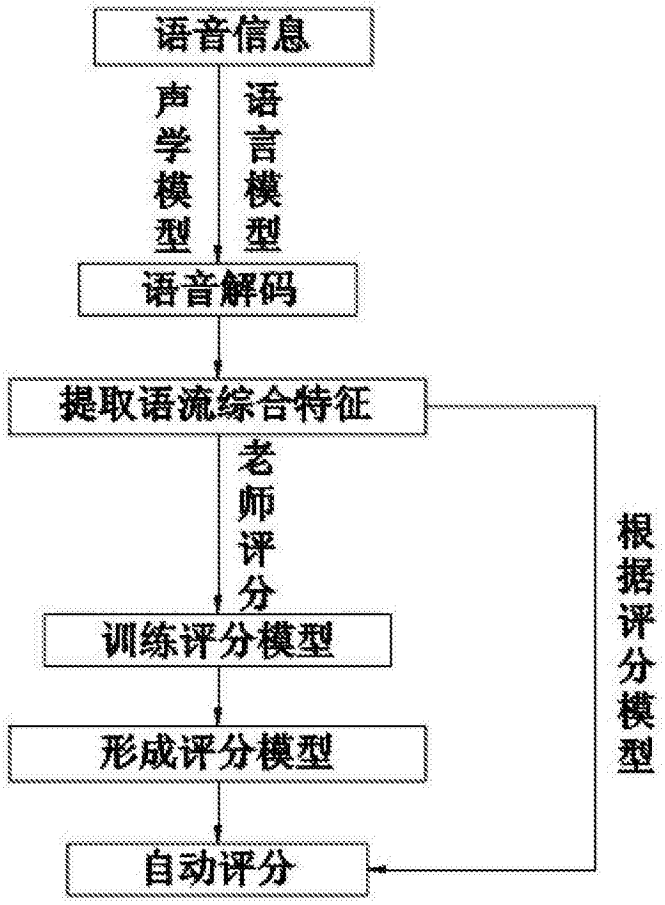

语音信息
声学模型
语言模型
语音解码
提取语流综合特征
老师评分
根据评分模型
训练评分模型
形成评分模型
自动评分

图 1

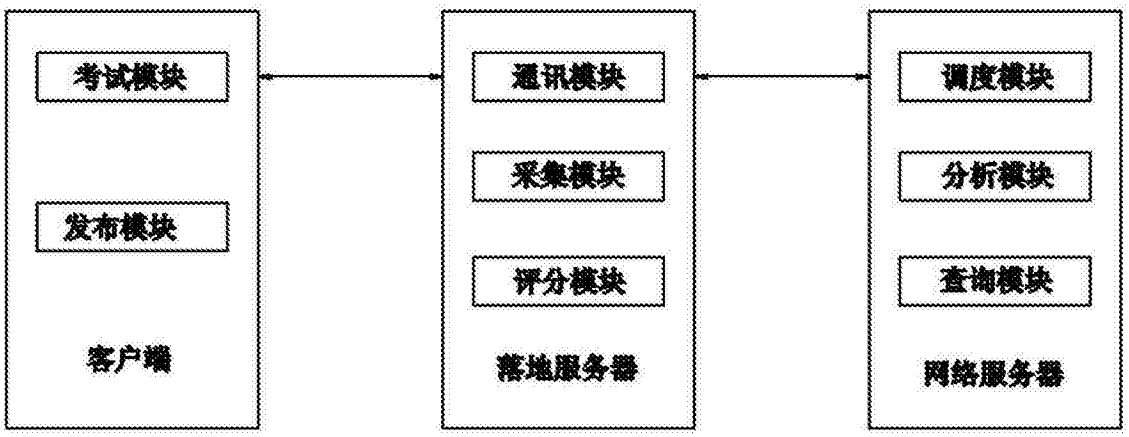

考试模块
发布模块
客户端
通讯模块
采集模块
评分模块
落地服务器
调度模块
分析模块
查询模块
网络服务器

图 2

识别结果
标准答案
词与词之间的相似度得分
词与句子之间的相似度得分
平均值
识别结果与标准答案之间的相似度得分

图 3

识别结果
标准答案
词与词之间的相似度得分
词与句子之间的相似度得分
最大值
识别结果与标准答案之间的相似度得分

图 4

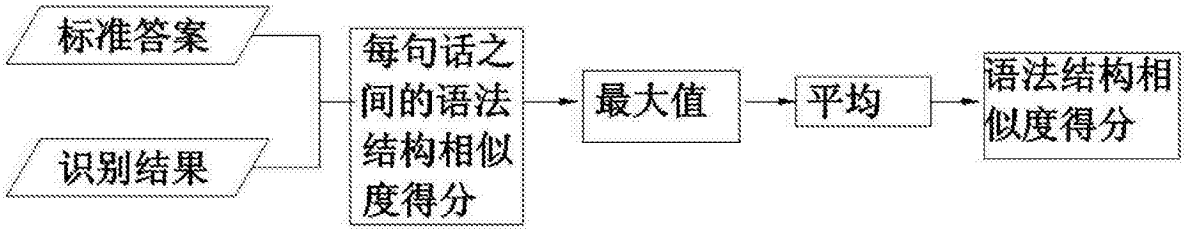

标准答案
识别结果
每句话之间的语法结构相似度得分
最大值
平均
语法结构相似度得分

图 5