



(12) 发明专利

(10) 授权公告号 CN 108831212 B

(45) 授权公告日 2020.10.23

(21) 申请号 201810689188.3

G09B 19/06 (2006.01)

(22) 申请日 2018.06.28

G10L 15/26 (2006.01)

G10L 25/51 (2013.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 108831212 A

审查员 李燕

(43) 申请公布日 2018.11.16

(73) 专利权人 深圳语易教育科技有限公司

地址 518057 广东省深圳市南山区粤海街
道粤兴六道中科纳能研发中心A座405
房

(72) 发明人 何光耀

(74) 专利代理机构 北京文苑专利代理有限公司

11516

代理人 王炜

(51) Int. Cl.

G09B 5/06 (2006.01)

权利要求书2页 说明书7页

(54) 发明名称

一种口语教学辅助装置及方法

(57) 摘要

本发明涉及一种口语教学辅助装置及方法，该装置包括处理单元以及分别与所述处理单元电连接的储存单元、语音输入单元和输出单元，所述储存单元存储多笔口语学习数据，每一笔口语学习数据包含一训练语句、多个标准音节组及一加权信息，所述训练语句由依序排列的多个标准文字所构成，该等标准音节组对应于该等标准文字，且是通过预先解析该等标准文字所获得的。本发明的语音处理能力强，能够准确辨别用户的发音并对其语音准确性进行评价，能够通过语音交互并呈现给用户虚拟现实影像，交互性强，用户体验好，身临其境的感觉强烈，从而提高了用户进行口语学习的兴趣。

1. 一种口语教学辅助装置,其特征在于,包括处理单元以及分别与所述处理单元电连接的储存单元、语音输入单元和输出单元;

所述储存单元存储多笔口语学习数据,每一笔口语学习数据包含一训练语句、多个标准音节组及一加权信息,所述训练语句由依序排列的多个标准文字所构成,所述标准音节组对应于所述标准文字,且是通过预先解析所述标准文字所获得的;所述加权信息具有一个加权文字以及一对应于加权文字的加权值,加权文字对应于所述训练语句所包含的所述标准文字中的其中一个;

所述处理单元控制输出单元输出训练语句;处理单元在接收到来自语音输入单元的一个语音时,解析语音以获得由多个待判定文字组成的一个待判定语句及多个分别对应于所述待判定文字的待判定音节组;处理单元在判定出所述标准音节组与所述待判定音节组不相符的至少一音节组时,将不相符的至少一音节组所对应的至少一标准文字作为至少一未相符文字;处理单元利用语言辨识模型,根据训练语句及至少一未相符文字,产生一相关于至少一未相符文字的原始分数;处理单元在判定出至少一未相符文字对应于加权文字时,将加权文字所对应的加权值作为目标加权值;处理单元根据原始分数及目标加权值产生一加权后分数,并根据加权后分数控制输出单元输出一个相关于语音的评价。

2. 根据权利要求 1 所述的口语教学辅助装置,其特征在于,所述输出单元包括分别与所述处理单元电连接的一显示屏幕和一扬声器。

3. 根据权利要求 1 所述的口语教学辅助装置,其特征在于,所述语音输入单元包括一麦克风。

4. 一种利用权利要求 1-3 任一项所述的口语教学辅助装置实现的口语教学辅助方法,其特征在于,包括以下步骤:

(A) 处理单元控制输出单元输出训练语句;

(B) 处理单元在接收到来自语音输入单元的语音时,解析语音以获得由多个待判定文字组成的一待判定语句及多个分别对应于所述待判定文字的待判定音节组;

(C) 处理单元在判定出所述标准音节组与所述待判定音节组不相符的至少一音节组时,将不相符的至少一音节组所对应的至少一标准文字作为至少一未相符文字;

(D) 处理单元利用一语言辨识模型,根据训练语句及至少一未相符文字,产生一相关于至少一未相符文字的原始分数;

(E) 处理单元在判定出至少一未相符文字对应于加权文字时,将加权文字所对应的加权值作为一目标加权值;及

(F) 处理单元根据原始分数及目标加权值产生一加权后分数,并根据加权后分数控制输出单元输出一个相关于语音的评价。

5. 根据权利要求 4 所述的口语教学辅助方法,其特征在于,所述步骤(C)包括:

(c1) 处理单元在判定出所述标准音节组与所述待判定音节组未对应的至少一音节组时,判定所述标准音节组的未对应的至少一音节组与所述待判定音节组不相符;

(c2) 处理单元在判定出所述标准音节组分别对应于所述待判定音节组时,判定所述标准音节组是否分别与所述待判定音节组相同,并在判定出所述标准音节组与所述待判定音节组不相同的至少一音节组时,判定所述标准音节组的不相同的至少一音节组与所述待判定音节组不相符。

6. 根据权利要求 4 所述的口语教学辅助方法,其特征在於,在所述步骤(E) 中,处理单元控制显示屏幕和/或扬声器输出评价,并控制显示屏幕显示至少一未相符文字。

7. 根据权利要求 4 所述的口语教学辅助方法,其特征在於,在步骤(A) 之前,处理单元根据口语学习数据所包含的多媒体数据,控制显示屏幕显示多媒体数据的虚拟现实影像。

8. 根据权利要求 4 所述的口语教学辅助方法,其特征在於,所述语言辨识模型为 N-Gram 模型。

一种口语教学辅助装置及方法

技术领域

[0001] 本发明属于口语教学技术领域,具体涉及一种口语教学辅助装置及方法。

背景技术

[0002] 口语学习是学习外语的必要途径之一,针对口语学习的口语教学是重要手段。现有技术的口语教学途径与方法主要有影音教学、软件教学、真人(现场或远程)教学等。现有技术的口语教学途径与方法的缺点在于:影音教学的缺点:该方式不具备与用户的交互性;软件教学的缺点:用户体验不容易实现口语使用情境之临场感,用户体验不好;真人教学的缺点:价格昂贵、有的用户遇到真人教师会难为情、开不了口。

发明内容

[0003] 针对上述现有技术中存在的问题,本发明的目的在于提供一种可避免出现上述技术缺陷的口语教学辅助装置及方法。

[0004] 为了实现上述发明目的,本发明提供的技术方案如下:

[0005] 一种口语教学辅助装置,包括处理单元以及分别与所述处理单元电连接的储存单元、语音输入单元和输出单元。

[0006] 进一步地,所述输出单元包括分别与所述处理单元电连接的一显示屏幕和一扬声器。

[0007] 进一步地,所述语音输入单元包括一麦克风。

[0008] 进一步地,所述储存单元存储多笔口语学习数据,每一笔口语学习数据包含一训练语句、多个标准音节组及一加权信息,所述训练语句由依序排列的多个标准文字所构成,该等标准音节组对应于该等标准文字,且是通过预先解析该等标准文字所获得的;所述加权信息具有一个加权文字以及一对应于该加权文字的加权值,该加权文字对应于所述训练语句所包含的该等标准文字中的其中一个。

[0009] 进一步地,所述处理单元控制输出单元输出训练语句;处理单元在接收到来自语音输入单元的一个语音时,解析该语音以获得由多个待判定文字组成的一个待判定语句及多个分别对应于该等待判定文字的待判定音节组;处理单元在判定出该等标准音节组与该等待判定音节组不相符的至少一音节组时,将不相符的该至少一音节组所对应的该至少一标准文字作为至少一未相符文字;处理单元利用语言辨识模型,根据该训练语句及该至少一未相符文字,产生一相关于该至少一未相符文字的原始分数;处理单元在判定出该至少一未相符文字对应于该加权文字时,将该加权文字所对应的加权值作为目标加权值;该处理单元根据该原始分数及该目标加权值产生一加权后分数,并根据该加权后分数控制该输出单元输出一个相关于该语音的评价。

[0010] 一种利用权利要求1所述的口语教学辅助装置实现的口语教学辅助方法,包括以下步骤:

[0011] (A) 处理单元控制输出单元输出训练语句;

[0012] (B) 处理单元在接收到来自语音输入单元的语音时,解析该语音以获得由多个待判定文字组成的一待判定语句及多个分别对应于该等待判定文字的待判定音节组;

[0013] (C) 处理单元在判定出该等标准音节组与该等待判定音节组不相符的至少一音节组时,将不相符的该至少一音节组所对应的该至少一标准文字作为至少一未相符文字;

[0014] (D) 处理单元利用一语言辨识模型,根据该训练语句及该至少一未相符文字,产生一相关于该至少一未相符文字的原始分数;

[0015] (E) 处理单元在判定出该至少一未相符文字对应于该加权文字时,将该加权文字所对应的加权重值作为一目标加权重值;及

[0016] (F) 处理单元根据该原始分数及该目标加权重值产生一加权后分数,并根据该加权后分数控制输出单元输出一个相关于该语音的评价。

[0017] 进一步地,所述步骤(C)包括:

[0018] (c1) 处理单元在判定出该等标准音节组与该等待判定音节组未对应的至少一音节组时,判定该等标准音节组的未对应的该至少一音节组与该等待判定音节组不相符;

[0019] (c2) 处理单元在判定出该等标准音节组分别对应于该等待判定音节组时,判定该等标准音节组是否分别与该等待判定音节组相同,并在判定出该等标准音节组与该等待判定音节组不相同的至少一音节组时,判定该等标准音节组的不相同的该至少一音节组与该等待判定音节组不相符。

[0020] 进一步地,在所述步骤(E)中,处理单元控制显示屏幕和/或扬声器输出该评价,并控制显示屏幕显示该至少一未相符文字。

[0021] 进一步地,在步骤(A)之前,处理单元根据口语学习数据所包含的多媒体数据,控制显示屏幕显示该多媒体数据的虚拟现实影像。

[0022] 进一步地,所述语言辨识模型为N-Gram模型。

[0023] 本发明提供的口语教学辅助装置及方法,语音处理能力强,能够准确辨别用户的发音并对其语音准确性进行评价,能够通过语音交互并呈现给用户虚拟现实影像,交互性强,用户体验好,身临其境的感觉强烈,从而提高了用户进行口语学习的兴趣,可以很好地满足实际应用的需要。

具体实施方式

[0024] 为了使本发明的目的、技术方案及优点更加清楚明白,下面结合具体实施例对本发明做进一步说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0025] 一种口语教学辅助装置,包括储存单元、语音输入单元、输出单元和处理单元,该储存单元、该语音输入单元和该输出单元分别与该处理单元电连接。

[0026] 储存单元存储多笔口语学习数据,每一笔口语学习数据包含一训练语句、多个标准音节组及一加权信息,所述训练语句由依序排列的多个标准文字所构成,该等标准音节组对应于该等标准文字,且是通过预先解析该等标准文字所获得的;所述加权信息具有一个加权文字以及一对应于该加权文字的加权重值,该加权文字对应于所述训练语句所包含的该等标准文字中的其中一个。

[0027] 所述处理单元控制该输出单元输出训练语句;该处理单元在接收到来自该语音输入单元的一个语音时,解析该语音以获得由多个待判定文字组成的一个待判定语句及多个分别对应于该等待判定文字的待判定音节组;该处理单元在判定出该等标准音节组与该等待判定音节组不相符的至少一音节组时,将不相符的该至少一音节组所对应的该至少一标准文字作为至少一未相符文字;该处理单元利用一语言辨识模型,根据该训练语句及该至少一未相符文字,产生一相关于该至少一未相符文字的原始分数;该语言辨识模型为N元(N-Gram)语言辨识模型;该处理单元在判定出该至少一未相符文字对应于该加权文字时,将该加权文字所对应的加权值作为目标加权值;该处理单元根据该原始分数及该目标加权值产生一加权后分数,并根据该加权后分数控制该输出单元输出一个相关于该语音的评价。

[0028] 该输出单元包括一显示屏幕及一扬声器,显示屏幕和扬声器分别与处理单元电连接,且该处理单元控制该显示屏幕和/或该扬声器输出该评价,并控制该显示屏幕显示该至少一未相符文字。

[0029] 该处理单元在判定该等标准音节组是否与该等待判定音节组相符时,该处理单元是在判定出该等标准音节组与该等待判定音节组未对应的至少一音节组时,判定该等标准音节组的未对应的该至少一音节组与该等待判定音节组不相符,及

[0030] 该处理单元在判定出该等标准音节组分别对应于该等待判定音节组时,还判定该等标准音节组是否分别相同于该等待判定音节组,并在判定出该等标准音节组与该等待判定音节组不相同的至少一音节组时,判定该等标准音节组的不相同的该至少一音节组与该等待判定音节组不相符。

[0031] 每笔学习数据还包含一多媒体数据,该多媒体数据具有一个相关于该等训练语句的虚拟现实影像,处理单元在控制该输出单元输出该笔学习数据的训练语句之前,还根据该笔学习数据所包含的该多媒体数据,控制该显示屏幕显示该多媒体数据的虚拟现实影像。

[0032] 该语音输入单元包括一麦克风。该口语教学辅助装置还包括一个头盔形状的外壳,储存单元、语音输入单元、输出单元和处理单元均装配在该外壳上,本口语教学辅助装置是一可穿戴在头上的头戴式显示设备。

[0033] 处理单元中安装有语音识别引擎,语音识别引擎内含三个主要模块:外文字典模块:例如本实施例使用的是英文,所以需要有一个“英文字典”作为标准;单词解析模块:将外文字典模块中的每个单词(vocabulary)解析成一连串的音节;语言模型模块:根据不同语言的使用习惯,分析该种语言各个单词在前后接连出现的比例。

[0034] 一种口语教学辅助方法,通过上述口语教学辅助装置来实现,包括以下步骤:

[0035] (A) 处理单元控制输出单元输出训练语句;

[0036] (B) 处理单元在接收到来自语音输入单元的语音时,解析该语音以获得由多个待判定文字组成的一待判定语句及多个分别对应于该等待判定文字的待判定音节组;

[0037] (C) 处理单元在判定出该等标准音节组与该等待判定音节组不相符的至少一音节组时,将不相符的该至少一音节组所对应的该至少一标准文字作为至少一未相符文字;

[0038] (D) 处理单元利用一语言辨识模型,根据该训练语句及该至少一未相符文字,产生一相关于该至少一未相符文字的原始分数;

[0039] (E) 处理单元在判定出该至少一未相符文字对应于该加权文字时,将该加权文字所对应的加权值作为一目标加权值;及

[0040] (F) 处理单元根据该原始分数及该目标加权值产生一加权后分数,并根据该加权后分数控制输出单元输出一个相关于该语音的评价。

[0041] 在该步骤(E)中,处理单元控制显示屏幕和/或扬声器输出该评价,并控制显示屏幕显示该至少一未相符文字。

[0042] 步骤(C)包括:

[0043] (c1) 处理单元在判定出该等标准音节组与该等待判定音节组未对应的至少一音节组时,判定该等标准音节组的未对应的该至少一音节组与该等待判定音节组不相符;及

[0044] (c2) 该处理单元在判定出该等标准音节组分别对应于该等待判定音节组时,判定该等标准音节组是否分别与该等待判定音节组相同,并在判定出该等标准音节组与该等待判定音节组不相同的至少一音节组时,判定该等标准音节组的不相同的该至少一音节组与该等待判定音节组不相符。

[0045] 在步骤(A)之前,处理单元根据口语学习数据所包含的多媒体数据,控制显示屏幕显示该多媒体数据的虚拟现实影像。

[0046] 在步骤(D)中,利用的语言辨识模型为N元(N-Gram)语言辨识模型。N元(N-Gram)语言辨识模型可以在NLP(Natural Language Processing,自然语言处理)中预计或者评估一个句子是否合理,可以用来评估两个字符串之间的差异程度。

[0047] 假定S表示某个有意义的句子,由一串特定顺序排列的词 $w_1, w_2, w_3, \dots, w_n$ 组成, n 是句子的长度。则S在文本中(语料库)出现的可能性也就是数学上所说的概率 $P(S)$ 为:

[0048] $P(S) = P(w_1, w_2, w_3, \dots, w_n) = P(W_1) P(W_2 | W_1) P(W_3 | W_1, W_2) \dots P(W_n | W_1, W_2, \dots, W_{n-1})$;

[0049] 该计算方法存在的缺陷在于:

[0050] 参数空间过大:条件概率 $P(w_n | w_1, w_2, \dots, w_{n-1})$ 的可能性太多,无法估算,不可能有用;

[0051] 数据稀疏严重:对于非常多词对的组合,在语料库中都没有出现,依据最大似然估计得到的概率将会是0。最后的结果是,模型仅仅能算很少的几个句子,而大部分的句子算得的概率是0。

[0052] 为了解决参数空间过大的问题,引入马尔科夫假设:随意一个词出现的概率只与它前面出现的有限的一个或者几个词有关。如果一个词的出现仅依赖于它前面出现的一个词,那么就称之为bigram:

[0053] $P(S) = P(w_1, w_2, w_3, \dots, w_n) = P(W_1) P(W_2 | W_1) P(W_3 | W_1, W_2) \dots P(W_n | W_1, W_2, \dots, W_{n-1}) \approx P(W_1) P(W_2 | W_1) P(W_3 | W_2) \dots P(W_n | W_{n-1})$
 $P(S) = P(w_1, w_2, w_3, \dots, w_n) = P(W_1) P(W_2 | W_1) P(W_3 | W_1, W_2) \dots P(W_n | W_1, W_2, \dots, W_{n-1}) \approx P(W_1) P(W_2 | W_1) P(W_3 | W_2) \dots P(W_n | W_{n-1})$;

[0054] 假设一个词的出现仅依赖于它前面出现的两个词,那么就称之为trigram:

[0055] $P(S) = P(w_1, w_2, w_3, \dots, w_n) = P(W_1) P(W_2 | W_1) P(W_3 | W_1, W_2) \dots P(W_n | W_1, W_2, \dots, W_{n-1}) \approx P(W_1) P(W_2 | W_1) P(W_3 | W_2, W_1) \dots P(W_n | W_{n-1}, W_{n-2})$;

[0056] 一般来说,N元模型就是假设当前词的出现概率只与它前面的N-1个词有关。而这些概率参数都是可以通过大规模语料库来计算,比如三元概率有:

[0057] $P(W_i|W_{i-1},W_{i-2}) \approx \text{count}(W_{i-2}W_{i-1}W_i) / \text{count}(W_{i-2}W_{i-1})$ 。

[0058] 数据平滑技术是构造高鲁棒性语言模型的重要手段,且数据平滑的效果与训练语料库的规模有关。训练语料库规模越小,数据平滑的效果越显著;训练语料库规模越大,数据平滑的效果越不显著,甚至可以忽略。数据平滑的目的有两个:一个是使全部的N-Gram概率之和为1;一个是使全部的N-Gram概率都不为0。其主要策略是把在训练样本中出现过的事件的概率适当减小,然后把减小得到的概率密度分配给训练语料中没有出现过的事件。实际中平滑算法有很多种,例如:Add-one平滑;Witten-Bell平滑;Good-Turing平滑;Katz Backoff;Stupid Backoff。

[0059] 基于N-Gram模型定义的字符串距离:

[0060] 模糊匹配的关键在于如何衡量两个长得很像的单词(或字符串)之间的“差异”,这种差异通常又称为“距离”。除了可以定义两个字符串之间的编辑距离(通常利用Needleman-Wunsch算法或Smith-Waterman算法),还可以定义它们之间的Ngram距离。假设有一个字符串S,那么该字符串的Ngram就表示按长度N切分原词得到的词段,也就是S中所有长度为N的子字符串。设想如果有两个字符串,然后分别求它们的Ngram,那么就可以从它们的共有子串的数量这个角度去定义两个字符串间的Ngram距离。但是仅仅是简单地对共有子串进行计数显然也存在不足,这种方案显然忽略了两个字符串长度差异可能导致的问题。比如字符串girl和girlfriend,二者所拥有的公共子串数量显然与girl和其自身所拥有的公共子串数量相等,但是我们并不能据此认为girl和girlfriend是两个等同的匹配。为了解决该问题,有学者提出以非重复的Ngram分词为基础来定义Ngram距离,公式表示如下:

[0061] $|GN(S1)| + |GN(S2)| - 2 \times |GN(S1) \cap GN(S2)|$;

[0062] 其中, $|GN(S1)|$ 是字符串的Ngram集合,N值一般取2或者3。以N=2为例对字符串Gorbachev和Gorbechyov进行分段,可得如下结果:

[0063] 1 Go or rb ba ac ch he ev

[0064] 2 Go or rb be ec ch hy yo ov

[0065] 结合上面的公式,即可算得两个字符串之间的距离是 $8+9-2 \times 4=9$ 。显然,字符串之间的距离越小,它们就越接近。当两个字符串完全相等的时候,它们之间的距离就是0。

[0066] 利用Ngram模型评估语句是否合理:

[0067] 从统计的角度来看,自然语言中的一个句子S可以由任何词串构成,不过概率P(S)有大有小。如果给出了某个句子的一个节选,能够猜测后续的词应该是什么,例如:

[0068] the large green_.mountain or tree?

[0069] Kate swallowed the large green_.pill or broccoli?

[0070] 假设现在有一个语料库如下,其中<s1><s2>是句首标记,</s2></s1>是句尾标记:

[0071] 1<s1><s2>yes no no no no yes</s2></s1>

[0072] 2<s1><s2>no no no yes yes yes no</s2></s1>

[0073] 评估如下这个句子的概率:

[0074] 1<s1><s2>yes no no yes</s2></s1>

[0075] 利用trigram模型来计算概率的结果为:

[0076] $P(\text{yes}|\langle s1 \rangle, \langle s2 \rangle) = 1/2, P(\text{no}|\text{yes}, \text{no}) = 1/2, P(\langle /s2 \rangle | \text{no}, \text{yes}) = 1/2, P(\text{no}|\langle s2$

$\rangle, \text{yes}) = 1P(\text{yes} | \text{no}, \text{no}) = 2/5P(\langle /s1 \rangle | \text{yes}, \langle /s2 \rangle) = 1$

[0077] 所以得到的概率等于： $1/2 \times 1 \times 1/2 \times 2/5 \times 1/2 \times 1 = 0.05$ 。

[0078] 本口语教学辅助装置是一头盔式可穿戴设备，具有头盔式外壳。在使用本装置时，使用者戴上头盔式外壳后，显示屏幕正对人的眼睛，可以自由旋转头部、从头盔式外壳观景窗中看到360度的实时三维虚拟场景与人物，虚拟的人物角色就是用户口语练习的对象；用户可以从头盔式外壳上的扬声器中听到程序预先订制好的口语教学情境与课程内容；当程序要求用户以“开口说”的方式与订制的程序进行交互时，使用者可以直接开口说，程序会透过装配在头盔式外壳上的麦克风接收用户讲出的声音（语音输入）；订制的程序中有语音识别模块，可以实时判读麦克风收到的语音输入信息，并且实时进行评价，将评价结果（说得很棒、还不错、待加强等等）反馈给使用者，让使用者可以在第一时间就知道自己的口语说得好不好；使用者也可以通过“录音/回放”功能，将自己讲出来的口语跟预置于程序中由真人事先录制的标准口语进行详细的自我比对，以改善自己的口语能力。

[0079] 储存单元中存储有要求用户以口语输入的文字数据（即会话课程），使用者以口说的方式让麦克风接收音频数据。处理单元利用预先订制的文本数据通过语音识别引擎转换成供口语输入数据比对的辨识模板。用户口语输入后，处理单元通过语音识别引擎的单词解析模块将口语输入的音频解析成一连串的音节，然后将这些音节通过语言模型模块转换成口语样本。处理单元将口语样本与预先设定好的辨识模板进行数据比对。数据比对时采用事先定义的权重参数、以模糊逻辑运算出0~1之间的浮点数表示，求出口语样本与辨识模板的差异度，0表示口语样本与辨识模板的数据完全不同、1表示完全相同、0.7表示口语样本与辨识模板的数据约有70%相同。

[0080] 通过语音识别引擎可以通过两种方式进行数据比对：

[0081] 1) 语意分析比对：首先将辨识模板中的句子拆分成单字，依照特征可以标注成不同标签，并通过前后文找出句子的语意结构，接着将用户念的句子和辨识模板中的句子结构做比对，分析两个句子的语意结构是否相似，最后算法给定一个0~1之间的分数，分数越高代表相似度越高。例如：

[0082] 例1

[0083] □辨识模板中的句子：Let's go to the three o'clock show.

[0084] □使用者1：Let's go to the show

[0085] □使用者2：Let's go to the three clock show

[0086] □分析结果：使用者1漏掉了重要的时间“three o'clock”，只能得到0.5分。使用者2把o'clock念成了clock，但是因为前后文都是正确念出，因此可以得到比较高的0.8分。

[0087] 例2

[0088] □辨识模板中的句子：Uh-huh, then go straight to the next block and the post office'll be on your left.

[0089] □使用者1：Uh-huh, then go straight to the next block and the police office'll be on your right.

[0090] □使用者2：Uh-huh, go to the next block, the post office be on your left.

[0091] □分析结果：使用者1的语句结构类似本文，但是post office和police office是

不同地方,方向也是错误,得到的分数会是偏低的0.4。使用者2虽然漏了”then”、“straight”、“and”、“will”等字,但是语意较相近本文,可以得到不错的0.7分。

[0092] 例3

[0093] 辨识模板中的句子:Thanks!Can I get off at any stop along the route?

[0094] 使用者1:Thank!Can I get off any stop along the road?

[0095] 使用者2::Thanks!Can I get on at any stop along the route?

[0096] 两位用户结构分数接近,但是关键的单词上get off,使用者2念成了get on,完全和本文语意相反,所以会扣掉非常多的分数,只得到0.6分。

[0097] 2) 单字权重比对:逐字比对辨识模板中的句子和使用者念出的句子的差异度有多少,并依照单字重要性权重给分。例如:

[0098] 辨识模板中的句子:I guess you've caught a cold.Have you checked your temperature?

[0099] 使用者:I guess you've catch a cold.Have you checked your temperature?

[0100] 分析结果:使用者把caught念成catch,可以得到的原始分数是0.91,经过加权计算,caught是个加权字,所以得分往下修正成0.78。

[0101] 比对评价:处理单元对数据比对的结果打分数,这个分数会反馈给用户知晓。例如:用户口语输入后、经过数据比对的结果为大于0.8,会给予“非常棒”的评价;又例如:用户口语输入后、经过数据比对的结果为小于0.4,系统会给予“有待加强”的评价,并且让使用者反复练习。

[0102] 差异程度:处理单元同时也会根据语音识别引擎数据比对的结果,将比对数据不相符之处标示出来,让使用者知道执行“口语输入”阶段中、哪些单词被系统评价为不相符(会以文本数据将不相符的单词用特殊的颜色标示出来、实时显示于显示屏幕上)。

[0103] 本发明提供的口语教学辅助装置及方法,语音处理能力强,能够准确辨别用户的发音并对其语音准确性进行评价,能够通过语音交互并呈现给用户虚拟现实影像,交互性强,用户体验好,身临其境的感觉强烈,从而提高了用户进行口语学习的兴趣,可以很好地满足实际应用的需要。

[0104] 以上所述实施例仅表达了本发明的实施方式,其描述较为具体和详细,但并不能因此而理解为对本发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本发明构思的前提下,还可以做出若干变形和改进,这些都属于本发明的保护范围。因此,本发明的保护范围应以所附权利要求为准。