



(12) 发明专利申请

(10) 申请公布号 CN 113611287 A

(43) 申请公布日 2021. 11. 05

(21) 申请号 202110730385.7

G01H 13/00 (2006.01)

(22) 申请日 2021.06.29

(71) 申请人 深圳大学

地址 518060 广东省深圳市南山区南海大道3688号

(72) 发明人 伍楷舜 王璐 王泰华 涂栋亮

(74) 专利代理机构 北京市诚辉律师事务所  
11430

代理人 耿慧敏 朱伟军

(51) Int. Cl.

G10L 15/02 (2006.01)

G10L 15/06 (2013.01)

G10L 15/16 (2006.01)

G10L 15/25 (2013.01)

G10L 21/0208 (2013.01)

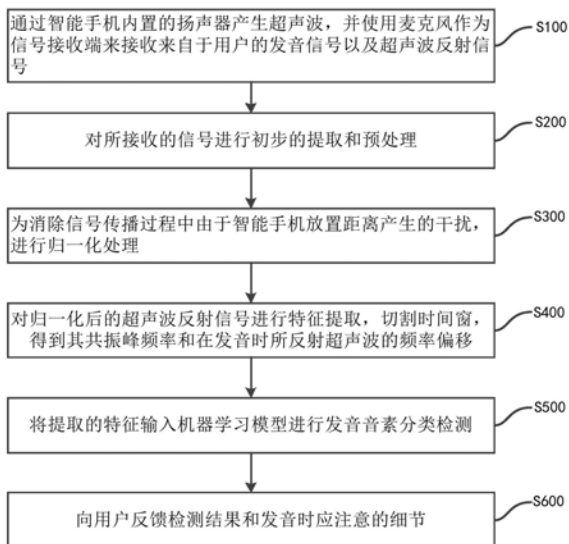
权利要求书2页 说明书9页 附图4页

(54) 发明名称

一种基于机器学习的发音纠错方法和系统

(57) 摘要

本发明公开一种基于机器学习的发音纠错方法和系统。该方法包括：通过电子设备内置的扬声器产生超声波，并使用麦克风作为信号接收端接收来自于用户的发音信号以及超声波反射信号；对用户发音信号以及超声波反射信号进行去噪，提取出有效发音信号和对应的有效超声波反射信号；对有效发音信号和有效超声波反射信号统一进行峰值归一化处理；对归一化后的超声波反射信号进行时间窗分割，计算发音信号的共振峰频率，并基于多普勒效应计算超声波反射信号的频率偏移值；将一个时序中共振峰频率值以及对应的超声波反射信号的频率偏移值输入经训练的机器学习模型，识别出语音音素分类。本发明可准确识别多种语言的发音并辅助发音训练，具有普适性。



1. 一种基于机器学习的发音纠错方法,包括以下步骤:

步骤S1,通过电子设备内置的扬声器产生超声波,并使用麦克风作为信号接收端接收来自于用户的发音信号以及超声波反射信号,该超声波反射信号是超声波所反射的用户唇部的声音信号;

步骤S2,对接收到的用户发音信号以及超声波反射信号分别进行去噪处理,以提取出有效发音信号和对应的有效超声波反射信号;

步骤S3,对所述有效发音信号和所述有效超声波反射信号统一进行峰值归一化处理;

步骤S4,对归一化后的超声波反射信号进行时间窗分割,计算发声信号的共振峰频率,并基于多普勒效应计算超声波反射信号的频率偏移值;

步骤S5,将一个时序中共振峰频率值以及对应的超声波反射信号的频率偏移值输入经训练的机器学习模型,识别出语音音素分类。

2. 根据权利要求1所述的方法,其中,步骤S2包括:

利用陷波滤波器,分别去除背景声音信号以及超声波反射信号中的高频和低频信号;

利用最小均方滤波器,去除声音信号中噪声;

通过设定阈值检测有效发音信号的起点和终点,起点设置在超过该阈值的第一个点之前,取M1个采样点,终点设置为在信号连续M1个采样点后低于该阈值的第一个点之后的数据点。

3. 根据权利要求1所述的方法,其中,在步骤S4中,利用线性预测编码对发音信号计算共振峰频率,并利用傅里叶变换分析频域上由于唇部运动对所反射超声波带来的多普勒效应产生的频率偏移值。

4. 根据权利要求1所述的方法,其中,根据以下步骤训练所述机器学习模型:

利用训练集训练所构建的机器学习模型,该训练集反映时序共振峰频率值、超声波反射信号的频率偏移值与发音音素类型之间的对应关系;

利用测试集评估经训练的机器学习模型,输出发音音素的分类标签,该测试集包含测试样本的时序共振峰频率值、对应的超声波反射信号的频率偏移值,并且将超过训练集样本的共振峰值数据按信号值为1处理。

5. 根据权利要求1所述的方法,其中,该方法还包括:

向用户显示所识别的语音音素分类;

根据时间窗中共振峰频率信息和多普勒频率偏移的变化,确定用户嘴唇和舌头的位置,并将反馈发音时嘴唇和舌头位置信息变化进行显示,以供用户辅助发音训练。

6. 根据权利要求1所述的方法,其中,所述机器学习模型是支持向量机。

7. 根据权利要求3所述的方法,其中,根据以下步骤计算所述共振峰频率:

利用递归算法计算每个发音信号的自相关函数,得到预测器的系数;

分解线性预测滤波器的多项式系数,使用特征值分解方法查找多项式根 $\theta_i$ ,并根据该多项式根计算共振峰频率 $f_i$ ,表示为:

$$f_i = \frac{\theta_i}{2\pi T}$$

其中,T是采样周期值。

8. 一种基于机器学习的发音纠错系统,包括:

信号接收单元:用于通过电子设备内置的扬声器产生超声波,并使用麦克风作为信号接收端接收来自于用户的发音信号以及超声波反射信号,该超声波反射信号是超声波所反射的用户嘴唇的声音信号;

预处理单元:用于对接收到的用户发音信号以及超声波反射信号分别进行去噪处理,以提取出有效发音信号和对应的有效超声波反射信号;

归一化处理单元:用于对所述有效发音信号和所述有效超声波反射信号统一进行峰值归一化处理;

特征提取单元:用于对归一化后的超声波反射信号进行时间窗分割,计算发声信号的共振峰频率,并基于多普勒效应计算超声波反射信号的频率偏移值;

分类识别单元:用于将一个时序中共振峰频率值以及对应的超声波反射信号的频率偏移值输入经训练的机器学习模型,识别出语音音素分类。

9. 一种计算机可读存储介质,其上存储有计算机程序,其中,该程序被处理器执行时实现根据权利要求1至8中任一项所述方法的步骤。

10. 一种计算机设备,包括存储器和处理器,在所述存储器上存储有能够在处理器上运行的计算机程序,其特征在于,所述处理器执行所述程序时实现权利要求1至8中任一项所述的方法的步骤。

## 一种基于机器学习的发音纠错方法和系统

### 技术领域

[0001] 本发明涉及机器学习技术领域,更具体地,涉及一种基于机器学习的发音纠错方法和系统。

### 背景技术

[0002] 近年来,各专业领域对多种语言人才的需求不断增长,全球外语学习市场出现大幅增长。然而,在外语教育中,发音学习相比其它语言技能的学习更具挑战性。在实际教学过程中,由于缺乏发音教学资源 and 课时有限,许多教师很难教会学生正确的发音,因此,通常需要学生自己练习。目前,自动语音识别软件已应用于电脑辅助发音训练以促进发音学习,但自动语音识别软件仍需要不断提高其准确率。在语言评估方面,采用自动语音识别软件是成功的。例如,以前的托福或英语考试已经由人听改为计算机测试。一些教育服务提供商使用被称为“语音评分器”的自动语音识别系统对考生的发音和口音进行评分。

[0003] 然而,自动语音识别系统在语言教学中也存在问题和挑战。自动语音识别系统的发展更趋向于侧重评估情景发音的语调和发音的熟悉程度,使用自动语音识别系统通过自我纠正来调整发音的过程并没有达到理想的效果。例如,对于很多非母语人士,不了解语调发音的具体状况,而是根据所听到的声音盲目模仿语调,从而给发音学习带来了错误的指导。近年来,自动语音识别系统吸引了不同学科的研究人员,包括计算机科学、语言学、语音识别、信号处理、心理语言学、教育学、听觉和发音研究等。然而,在实践中,大多数自动语音识别系统集中在发音错误检测和评估,并且只关注发音的语调和流利程度,而不是发音本身。当学习者通过查看分数意识到自己的发音不正确时,他们并不知道如何通过根据分数或语音信号来调整正确的发音。

[0004] 总的来说,现有的自动语音识别系统主要着重于注重评估分数,而不是鼓励对语言学习者进行自我学习。因此,大多数第二语言学习者无意中被发音训练应用误导。例如,学习者试图通过调整自己的音高或语调(韵律水平)接近示例,而不是根据发音规则(语音水平)来调整自己的发音来达到最高分数。尽管自动化语音评分程序存在上述许多弊端,但使用发音系统的正确性并没有得到权威的评估,市场上现有的发音训练应用主要集中在两个方面:如何发音音素、单词或句子;通过语音识别技术评估学习者的发音质量,指出不正确的音素。也就是说,基于发音错误检测评分的方法并不能对发音错误进行诊断,诊断具体的错误形式。

[0005] 此外,第二语言(非母语)学习者通常使用与母语(第一语言)最接近的一个外语音素。这种情况会导致语音替换、插入和删除,语言学中的迁移模型解释了这一点。这种情况的发生使语言学习者的学习过程产生了非常大的弊端,让使用者无法有效的学习一门外语。例如,英语学习者在学习普通话时存在元音问题,因为英语中不存在元音/y/。所以,以英语为母语的人通常会发/u/而不是/y/,这会让使用者经常性的替换音节而不是主动地去纠正这种错误。由于/u/和/y/的共振峰频率在同一个频段,因此也无法利用对声音信号进行识别的语言模型区分出来使用者的错误发音。

## 发明内容

[0006] 本发明的目的是克服上述现有技术的缺陷,提供一种基于机器学习的发音纠错方法和系统。

[0007] 根据本发明的第一方面,提供一种基于机器学习的发音纠错方法。该方法包括以下步骤:

[0008] 步骤S1,通过电子设备内置的扬声器产生超声波,并使用麦克风作为信号接收端接收来自于用户的发音信号以及超声波反射信号,该超声波反射信号是超声波所反射的用户嘴唇的声音信号;

[0009] 步骤S2,对接收到的用户发音信号以及超声波反射信号分别进行去噪处理,以提取出有效发音信号和对应的有效超声波反射信号;

[0010] 步骤S3,对所述有效发音信号和所述有效超声波反射信号统一进行峰值归一化处理;

[0011] 步骤S4,对归一化后的超声波反射信号进行时间窗分割,计算发声信号的共振峰频率,并基于多普勒效应计算超声波反射信号的频率偏移值;

[0012] 步骤S5,将一个时序中共振峰频率值以及对应的超声波反射信号的频率偏移值输入经训练的机器学习模型,识别出语音音素分类。

[0013] 根据本发明的第二方面,提供一种基于机器学习的发音纠错系统。该系统包括:

[0014] 信号接收单元,其用于通过电子设备内置的扬声器产生超声波,并使用麦克风作为信号接收端接收来自于用户的发音信号以及超声波反射信号,该超声波反射信号是超声波所反射的用户嘴唇的声音信号;

[0015] 预处理单元,其用于对接收到的用户发音信号以及超声波反射信号分别进行去噪处理,以提取出有效发音信号和对应的有效超声波反射信号;

[0016] 归一化处理单元,其用于对所述有效发音信号和所述有效超声波反射信号统一进行峰值归一化处理;

[0017] 特征提取单元,其用于对归一化后的超声波反射信号进行时间窗分割,计算发声信号的共振峰频率,并基于多普勒效应计算超声波反射信号的频率偏移值;

[0018] 分类识别单元,其用于将一个时序中共振峰频率值以及对应的超声波反射信号的频率偏移值输入经训练的机器学习模型,识别出语音音素分类。

[0019] 与现有技术相比,本发明的优点在于,利用使用者在发音时语音共振峰的改变以及超声波信号反射产生的多普勒效应,分别判断发音时舌头的位置和嘴唇的运动,提出更准确的发音鉴别方案。并且,本发明可搭载在商用的电子设备上(如智能手机),无需额外的设备,降低了应用成本。

[0020] 通过以下参照附图对本发明的示例性实施例的详细描述,本发明的其它特征及其优点将会变得清楚。

## 附图说明

[0021] 被结合在说明书中并构成说明书的一部分的附图示出了本发明的实施例,并且连同其说明一起用于解释本发明的原理。

[0022] 图1是根据本发明一个实施例的基于机器学习的发音纠错方法的流程图;

- [0023] 图2是根据本发明一个实施例的基于机器学习的发音纠错方法的过程示意图；
- [0024] 图3是根据本发明一个实施例的基于机器学习的发音纠错系统的框架结构示意图；
- [0025] 图4是根据本发明一个实施例的实验布置图。

### 具体实施方式

[0026] 现在将参照附图来详细描述本发明的各种示例性实施例。应注意到：除非另外具体说明，否则在这些实施例中阐述的部件和步骤的相对布置、数字表达式和数值不限制本发明的范围。

[0027] 以下对至少一个示例性实施例的描述实际上仅仅是说明性的，决不作为对本发明及其应用或使用的任何限制。

[0028] 对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论，但在适当情况下，所述技术、方法和设备应当被视为说明书的一部分。

[0029] 在这里示出和讨论的所有例子中，任何具体值应被解释为仅仅是示例性的，而不是作为限制。因此，示例性实施例的其它例子可以具有不同的值。

[0030] 应注意到：相似的标号和字母在下面的附图中表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一步讨论。

[0031] 本发明旨在提出使用常用的电子设备进行语音评估和智能发音，可以根据舌位和唇形的语音规则识别发音，特别是单音元音和语音发音等。本发明可搭载在商用的电子设备上进行发音识别和分析，其中电子设备包括但不限于智能手机、平板电子设备、台式电脑、个人数字助理(PDA)、车载设备、智能穿戴设备等。为清楚起见，以下以智能手机为例，并以用户(或称使用者)进行元音发音为例进行说明。

[0032] 具体地，结合图1和图2所示，所提供的基于机器学习的发音纠错方法包括以下步骤。

[0033] 步骤S100，通过智能手机内置的扬声器产生超声波，并使用麦克风作为信号接收端来接收来自于用户的发音信号以及超声波反射信号。

[0034] 发音信号是指用户读音时直接产生的声音信号，超声波反射信号是指利用超声波所反射的用户嘴唇运动或面部运动的声音信号，利用麦克风可接收这两种声音的混合信号。在实际应用中，可将智能手机放置在用户面前，并通过内置的扬声器产生频率为20KHz的超声波。

[0035] 在一个实施例中，步骤S100包括以下子步骤：

[0036] 步骤S101，采集智能手机麦克风的的声音信号(包括用户发音信号和发射超声波信号)；

[0037] 步骤S102，对采集到的声音信号进行数据处理，去除高频环境噪声，电磁噪声的干扰。

[0038] 例如，使用巴特沃斯低通滤波器进行去噪，其截止频率设置为5500Hz，麦克风的数据采样率设置为44KHz。

[0039] 步骤S200，对所接收的信号进行初步的提取和预处理。

[0040] 在一个实施例中，可利用智能手机执行以下子步骤：

[0041] 步骤S201,利用陷波滤波器,去除背景人声声音信号和反射超声波中的高频和低频信号;

[0042] 步骤S202,利用最小均方滤波器,去除声音信号中的噪声;

[0043] 步骤S203,利用阈值法检测声音信号中发音的起点和终点。

[0044] 例如,在进行信号检测时,对超过设定阈值的信号进行时间上的判断,当信号持续超过阈值并且维持了一段时间 $t$ 时( $t$ 大于 $1s$ ),则认为该信号不是噪声信号,而是需要提取的有效信号,进一步地对这个信号进行切割。将切割起点和切割终点的信号提取出来。具体地,判断切割信号的方法为,首先判断信号的起点,起点选择例如是在超过阈值的第一个点之前,取 $M1$ 个采样点( $M1$ 的大小约为采样率与一次发音时间的乘积);然后,判断该次发音信号的终点,例如,在信号连续 $M1$ 个采样点后低于阈值的第一个点之后的数据点。

[0045] 在该步骤S200中,处理人声语音信号时所使用的陷波滤波器例如是 $100\sim 5500Hz$ ,处理反射超声波的语音信号时所使用的陷波滤波器例如是 $19.90\sim 21.10KHz$ ,最小均方滤波器的采样时间长短例如设置为 $10ms$ 。

[0046] 步骤S300,为消除信号传播过程中由于智能手机放置距离产生的干扰,进行归一化处理。

[0047] 在一个实施例中,步骤S300包括以下子步骤:

[0048] 步骤S301,寻找所有信号中训练集声音信号的峰值,并将接收到的所有声音信号进行峰值归一化处理;

[0049] 步骤S302,在测试样本中,将超过训练集样本的峰值数据按信号值为1进行处理。

[0050] 对接收的混合声音信号(包括发音信号和超声波反射信号)进行归一化处理,能够消除由于智能手机与用户嘴唇之间的距离差异而引起的干扰,从而提供后续发音音素识别的精确度。例如,归一化处理公式表示为:

$$[0051] \quad y = \frac{(x - x_{\min})}{x_{\max} - x_{\min}} \quad (1)$$

[0052] 其中 $x$ 和 $y$ 分别对应归一化前后的信号, $x_{\min}$ 是归一化前的最小峰值, $x_{\max}$ 是归一化前的最大峰值。

[0053] 步骤S400,对归一化后的超声波反射信号进行特征提取,切割时间窗,得到其共振峰频率和在发音时所反射超声波的频率偏移。

[0054] 在一个实施例中,步骤S400包括以下子步骤:

[0055] 步骤S401,对信号使用傅里叶变换进行处理,并只提取所得声音信号频率谱中 $100-5500Hz$ 的频率幅值信息,利用线性预测编码技术得到其共振峰频率;

[0056] 步骤S402,将所分割的数据进行时间窗分割,利用多普勒效应通过傅里叶变换检测在 $20KHz$ 频段左右用户反射智能手机所发射超声波的频率偏移。

[0057] 具体地,可利用线性预测编码通过确定一组预测器的系数来预测实际的语音采样值,通过对预测器的系数的计算可以得到此时的共振峰频率,从而映射出此时舌头的位置以及发音的音素。线性预测的基本思想是,语音采样值之间存在着相关性,因此可以通过组合若干个过去的语音样本值来逼近当前的语音样本值。通过使实际的语音采样值和预测值之间的差值的平方和最小,即均方误差最小,来确定唯一的一组预测系数。

[0058] 例如,使用莱文逊-杜宾(Levinson-Durbin)的自相关法进行求解,利用递归算法

计算每个发音信号的自相关函数,得到预测器的系数。

[0059] 获得预测系数后,分解线性预测滤波器的多项式系数,如使用特征值分解方法查找复杂的多项式根 $\theta_i$ ,根据复杂根及采样周期值T计算共振峰频率 $f_i$ ,表示为:

$$[0060] \quad f_i = \frac{\theta_i}{2\pi T} \quad (2)$$

[0061] 通过对此共振峰频率 $f_i$ 对嘴唇位置的映射可以大致确定此时发音的音素的种类,但由于类似相似的元音音素,例如/u/和/y/的共振峰频率相近,仍不能通过此方法完全分离。根本原因在于发该元音时舌头位置基本相同。因此,为了解决这个问题,本发明进一步利用多普勒频移测量嘴唇的方法来检测嘴唇的运动。

[0062] 多普勒效应是发射波源和接收目标存在相对运动时,观察者接收波的频率与波源发出的频率出现偏移的现象,在本发明中,由于所检测的使用者在发不同元音的时候嘴唇的运动不同,会产生多普勒效应,从而引入了不同的频移。

[0063] 在一个实施例中,使用短时傅里叶变换的方法采集由手机扬声器发射并反射回麦克风的20KHz附近频率的信号,当用户发声时,嘴唇向后运动会产生负频率偏移,反之,嘴唇向前运动会产生正频率偏移。例如,当发/a/、/i/、/e/这三个元音时,嘴唇呈扁平状,整体是向后缩的,与接收端之间的距离变大,多普勒频率偏移表现为反射波的中心频率向低频段运动。而当发/o/、/u/、/ü/这三个元音时,嘴唇呈圆形,具有靠近接收端的运动趋势,多普勒频率偏移表现为反射波的中心频率向高频段运动。

[0064] 在该步骤S400中,利用线性预测编码技术来计算信号共振峰频率,并利用傅里叶变换方法计算频域上由于嘴唇运动的多普勒效应所反射超声波的频移值,该频移值既能够反映频移的方向,也可以反映频率的变化程度,从而根据发音规则捕捉嘴唇和舌头的位置。

[0065] 步骤S500,将提取的特征输入机器学习模型进行发音音素分类检测。

[0066] 在该步骤中,将所提取的特征,包括共振峰频率值和发音时所反射超声波的频率偏移值等输入机器学习模型,即可识别发音音素类型。例如,将一个时序中所筛选的特征输入支持向量机进行分类,也可采用其他的机器学习模型,如卷积神经网络等。

[0067] 在一个实施例中,步骤S500包括以下子步骤:

[0068] 步骤S501,基于支持向量机的分类方法,将处理好的时序共振峰频率信息输入支持向量机;

[0069] 步骤S502,支持向量机具有训练状态与测试状态,在训练状态,根据数据所给的标签对支持向量机所建立的超平面参数进行训练,如果是测试状态,则返回根据该支持向量机分类所显示的标签结果。其中在进行支持向量机模型训练时,输入的特征为通过S100-S400处理后的声音信号在分割出的各个时间窗内的共振峰频率,输出的分类标签为分类识别的元音音素,即元音音素类型。

[0070] 在上述实施例中,优选使用支持向量机模型分类元音发音信号。支持向量机模型是作为单声道分类器的合适模型,支持向量机内核将低维输入空间转换为更高维空间。换句话说,支持向量机模型通过向它添加更多维度,将不可分离的问题转换为可分离的问题。该模型采用了高斯核函数,因为它通过添加更多维度将不可分离的问题转换为可分离的问题。在这种情况下,通过共振峰的选取以及多普勒频移的确定可以对具有相同舌头位置或具有相同唇形的元音进行分类。



[0071] 综上,在该步骤S500中,通过使用支持向量机模型检测用户发音正确或错误,并提供发音反馈,以使用户了解发音时嘴巴和舌头的位置。

[0072] 应理解的是,利用训练数据集对支持向量机进行训练,并利用测试数据集评估模型的训练效果,即可获得优化的支持向量机模型,进而用于实际发音时音素种类识别和分析。训练数据集反映时序上共振峰频率值、发音时所反射超声波的频率偏移值与发音音素种类之间的对应关系。

[0073] 步骤S600,向用户反馈检测结果和发音时应注意的细节。

[0074] 在获得检测结果后,可通过智能手机告知用户发音时应注意的细节。例如,步骤S600包括以下子步骤:

[0075] 步骤S601,根据步骤S500所返回的结果,对分类的发音结果进行显示;

[0076] 步骤S602,根据步骤S500所分析的时间窗中共振峰频率信息的变化,判断用户嘴唇和舌头的位置,并将这一变化显示在智能手机上;

[0077] 步骤S603,分类结果和共振信息的变化反映用户在发音时需要注意的细节,例如发音时应该将舌头上移或下移,亦或是嘴唇应该向前或向后,以提供辅助训练。

[0078] 例如,实际采集不同国籍的母语者在说不同语言的元音发音,通过数据分析以及分类训练,能够对多种语言的元音发音实现矫正并对使用者反馈分析结果。

[0079] 相应地,本发明还提供一种基于机器学习的发音纠错系统,用于实现上述方法的一个方面或多个方面,其中涉及的模块或单元可采用专用处理器或FPGA等实现。例如,参见图3所示,该系统包括感知和处理模块、计算模块和反馈模块。

[0080] 感知和处理模块用于执行以下过程:在用户发声时,通过使用智能手机的扬声器发生超声波并在智能手机的麦克风上接收返回的声波信号。归一化声波信号,并通过使用巴特沃斯带通滤波器,陷波滤波器,最小均方滤波器处理原始的声波信号,并利用线性预测编码技术来计算共振峰频率以及利用傅里叶变换处理超声波反射的信号计算多普勒频移。在用户一个时间段的采集中,将声波信号切割成很多个时间单元,对每个单元的共振峰频率进行计算,利用多普勒效应可以判断用户在发音时嘴唇和舌头的运动。

[0081] 计算模块用于执行以下过程:计算得到在一次采样中多个时间窗口的共振峰频率,通过支持向量机的分类模型对上述多个窗口共振峰频率进行训练,并反馈给该机器学习模型。在得到多个训练样本的数据后,该模型能够对所得到的多个时间窗口的共振峰频率数据进行分类预测;

[0082] 反馈模块用于执行以下过程:反馈给用户所发元音的音素,同时,能够将用户发音过程中通过对时间窗上的共振峰频率的变化判断出舌头的位置和嘴唇的形状,将这一变化进行可视化显示。并且,通过参考在时间窗口中共振峰频率的标准库中对此次元音音素的发音提出建议,告诉用户应该张大或是抿起嘴唇发音。

[0083] 在一个实施例中,所述感知和处理模块包括以下单元:

[0084] 语音发射单元,使用扬声器发送频率规律呈正弦波震荡的超声波;

[0085] 语音采集单元,使用麦克风用于反射用户面部(发声器官)的声音信号进行采集,并通过对阈值的判断采集时间片信号;

[0086] 归一化单元,用于消除用户发声大小带来的影响;

[0087] 滤波单元,利用滤波器对声音信号数据进行过滤;

- [0088] 模数单元,用于将模拟的声音信号转化成数字信号;
- [0089] 提取单元,利用线性预测编码技术来计算共振峰频率以及利用傅里叶变换处理超声波反射的信号计算多普勒频移。
- [0090] 在一个实施例中,所述计算模块包括以下单元:
- [0091] 时间片切割单元,用于对信号时间片信号的切割;
- [0092] 计算单元,利用机器学习模型对所分割单元中的特征进行分类预测。
- [0093] 在一个实施例中,所述反馈模块包括以下单元:
- [0094] 纠错单元,用于显示使用机器学习模型对输入的特征进行分类预测所得到元音音素的发音种类;
- [0095] 可视化单元,用于通过对时间窗分割时序特征的变化生成嘴唇和舌头位置的反馈图,并通过对比标准图的方法使用户可视化的明确其发声的错误。
- [0096] 本发明关注用户学习目标语言的语音清晰度,通过对元音音素的判断和纠正实现辅助发音训练。经验证,在实际应用中,本发明可以通过嘴唇和舌头的位置识别四种语言(法语、日语、韩语和普通话)中的30个元音。如图4所示,本发明可内置于智能手机中,用户进行语音纠正时,朗读所选择需要训练的元音音素,在朗读结束后,告知使用者朗读是否正确,其发声部位应该如何调整才可以更好的发出所学习语言的元音音素。
- [0097] 综上所述,本发明提供基于机器学习的发音纠错的实现方法,是一种创新的发音训练方式;利用使用者的声波中的共振并使用超声波基于多普勒效应对使用者的舌头和嘴唇位置分别检测,提高了检测精确度;通过使用巴特沃斯带通滤波器,陷波滤波器,最小均方滤波器,时间切片,归一化,傅里叶变换等处理方法对信号处理,显著提高了信号的准确度;所提出的使用支持向量机对元音音素分类识别率高;本发明还实现了检测结果反馈功能,进一步提高了可用性和功能性,不仅可以识别唇形和舌头的位置,还可以进行元音发音训练。
- [0098] 此外,传统方法不能通过分析/i/和/y/等语音波形对圆润和不圆润元音进行分类,因为它们依赖于唇形。针对该问题,本发明通过分析超声信号的频移方向,建立了一种圆形和非圆形的元音分类器模型,称为频移模型,能够提取更多标准元音音素的发音特征。
- [0099] 本发明可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括计算机可读存储介质,其上载有用于使处理器实现本发明的各个方面的计算机可读程序指令。
- [0100] 计算机可读存储介质可以是保持和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质例如可以是但不限于电存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或者上述的任意合适的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括:便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPR0M或闪存)、静态随机存取存储器(SRAM)、便携式压缩盘只读存储器(CD-ROM)、数字多功能盘(DVD)、记忆棒、软盘、机械编码设备、例如其上存储有指令的打孔卡或凹槽内凸起结构、以及上述的任意合适的组合。这里所使用的计算机可读存储介质不被解释为瞬时信号本身,诸如无线电波或者其他自由传播的电磁波、通过波导或其他传输媒介传播的电磁波(例如,通过光纤电缆的光脉冲)、或者通过电线传输的电信号。
- [0101] 这里所描述的计算机可读程序指令可以从计算机可读存储介质下载到各个计算/

处理设备,或者通过网络、例如因特网、局域网、广域网和/或无线网下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光纤传输、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或者网络接口从网络接收计算机可读程序指令,并转发该计算机可读程序指令,以供存储在各个计算/处理设备中的计算机可读存储介质中。

[0102] 用于执行本发明操作的计算机程序指令可以是汇编指令、指令集架构 (ISA) 指令、机器指令、机器相关指令、微代码、固件指令、状态设置数据、或者以一种或多种编程语言的任意组合编写的源代码或目标代码,所述编程语言包括面向对象的编程语言—诸如 Smalltalk、C++、Python 等,以及常规的过程式编程语言—诸如“C”语言或类似的编程语言。计算机可读程序指令可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络—包括局域网 (LAN) 或广域网 (WAN)—连接到用户计算机,或者,可以连接到外部计算机 (例如利用因特网服务提供商来通过因特网连接)。在一些实施例中,通过利用计算机可读程序指令的状态信息来个性化定制电子电路,例如可编程逻辑电路、现场可编程门阵列 (FPGA) 或可编程逻辑阵列 (PLA),该电子电路可以执行计算机可读程序指令,从而实现本发明的各个方面。

[0103] 这里参照根据本发明实施例的方法、装置 (系统) 和计算机程序产品的流程图和/或框图描述了本发明的各个方面。应当理解,流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合,都可以由计算机可读程序指令实现。

[0104] 这些计算机可读程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理器,从而生产出一种机器,使得这些指令在通过计算机或其它可编程数据处理装置的处理器执行时,产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中,这些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作,从而,存储有指令的计算机可读介质则包括一个制品,其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

[0105] 也可以把计算机可读程序指令加载到计算机、其它可编程数据处理装置、或其它设备上,使得在计算机、其它可编程数据处理装置或其它设备上执行一系列操作步骤,以产生计算机实现的过程,从而使得在计算机、其它可编程数据处理装置、或其它设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

[0106] 附图中的流程图和框图显示了根据本发明的多个实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分,所述模块、程序段或指令的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。对

于本领域技术人员来说公知的是,通过硬件方式实现、通过软件方式实现以及通过软件和硬件结合的方式实现都是等价的。

[0107] 以上已经描述了本发明的各实施例,上述说明是示例性的,并非穷尽性的,并且也不限于所披露的各实施例。在不偏离所说明的各实施例的范围和精神的情况下,对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择,旨在最好地解释各实施例的原理、实际应用或对市场中的技术改进,或者使本技术领域的其它普通技术人员能理解本文披露的各实施例。本发明的范围由所附权利要求来限定。

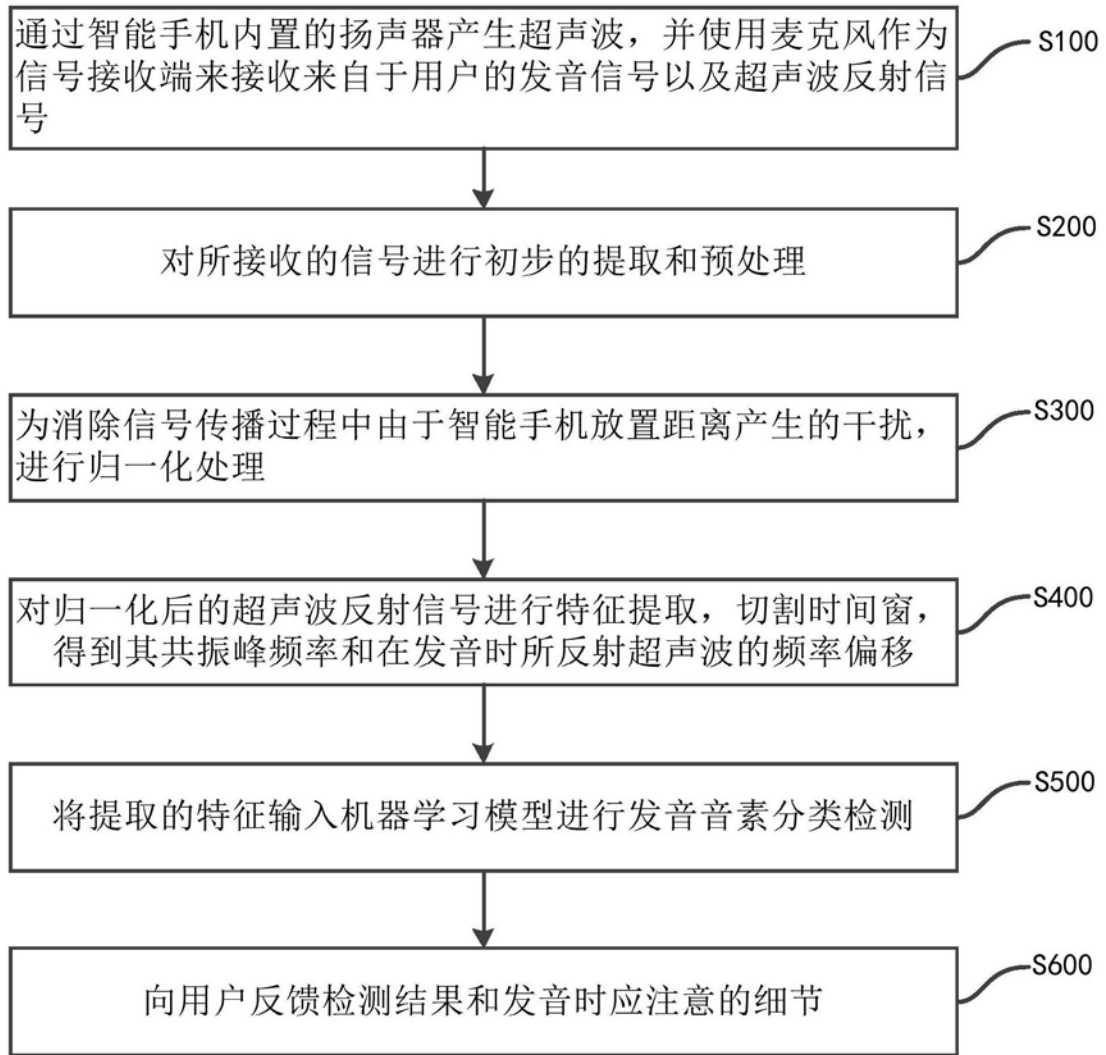


图1

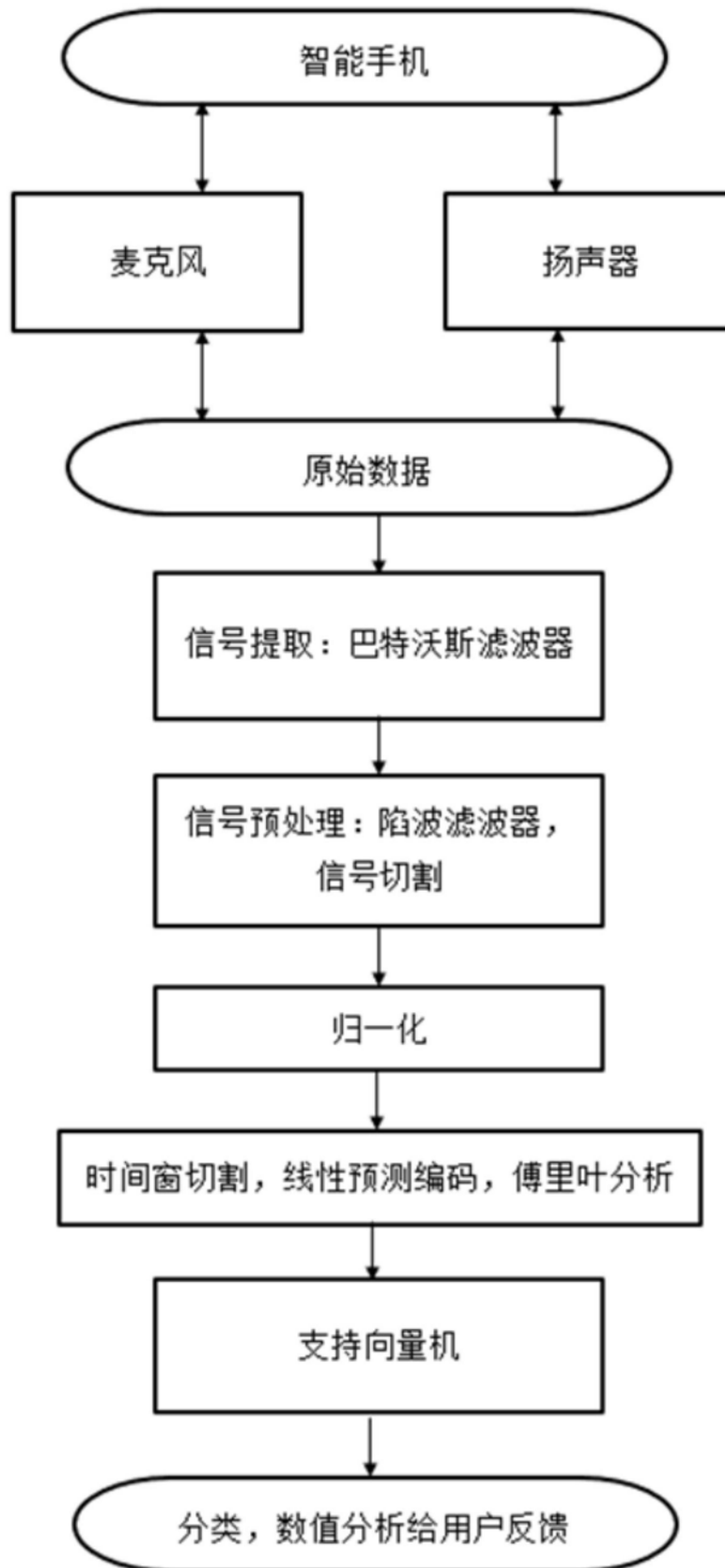


图2

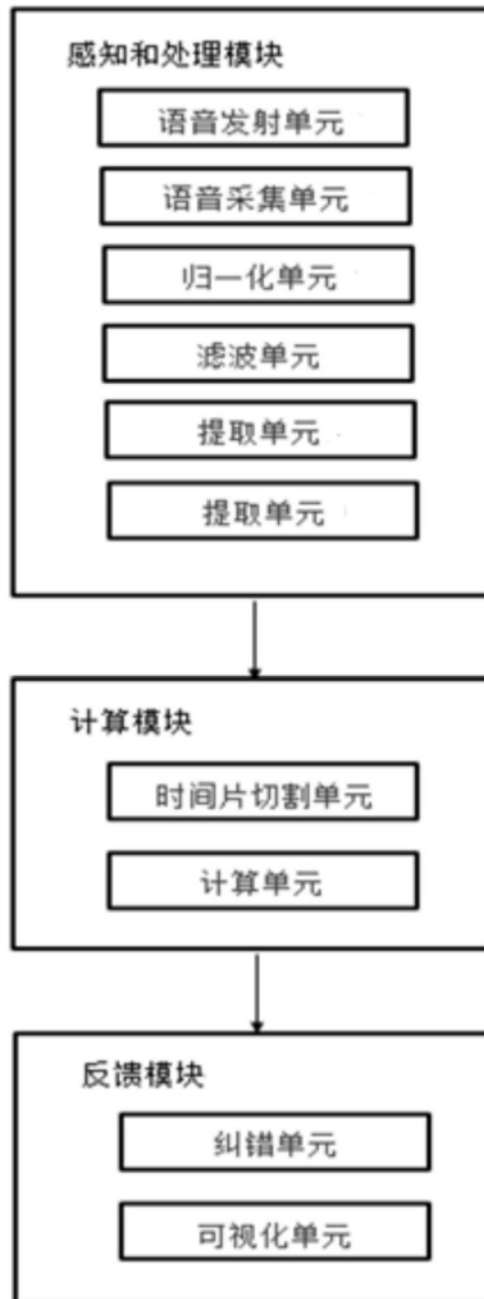


图3



图4