(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 108198547 B
(45)授权公告日 2020.10.23

(21)申请号 201810048223.3
(22)申请日 2018.01.18
(65)同一申请的已公布的文献号
申请公布号 CN 108198547 A
(43)申请公布日 2018.06.22
(73)专利权人 深圳市北科瑞声科技股份有限公司
地址 518051 广东省深圳市南山区高新区南区深港产学研基地大楼西座四楼W406室
(72)发明人 黄石磊 刘轶 王昕
(74)专利代理机构 广州华进联合专利商标代理有限公司 44224
代理人 谢曲曲
(51)Int.Cl.
G10L 15/02(2006.01)
G10L 15/08(2006.01)
G10L 19/038(2013.01)
G10L 21/0216(2013.01)
G10L 25/87(2013.01)

审查员 徐联微

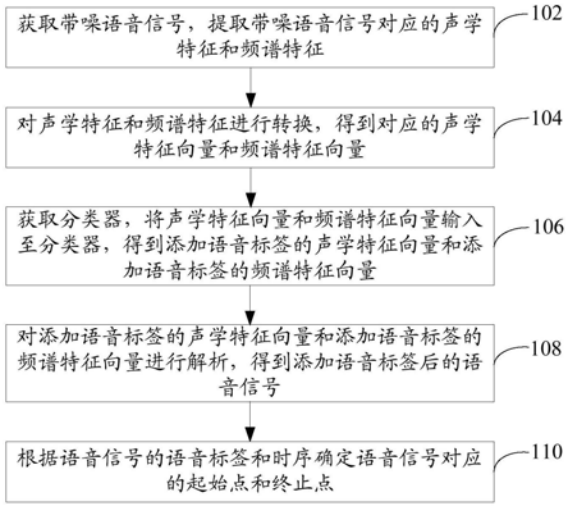
权利要求书2页 说明书12页 附图2页

(54)发明名称

语音端点检测方法、装置、计算机设备和存储介质

(57)摘要

本申请涉及一种语音端点检测方法、装置、计算机设备和存储介质。该方法包括:获取带噪语音信号,提取所述带噪语音信号对应的声学特征和频谱特征;对所述声学特征和频谱特征进行转换,得到对应的声学特征向量和频谱特征向量;获取分类器,将所述声学特征向量和频谱特征向量输入至所述分类器,得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量;对所述添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析,得到对应的语音信号;根据所述语音信号的时序确定所述语音信号对应的起始点和终止点。采用本方法能够有效提高语音端点检测的准确性。



CN 108198547 B

1.一种语音端点检测方法,包括:

获取带噪语音信号,提取所述带噪语音信号对应的声学特征;

提取所述带噪语音信号的带噪语音幅度谱、噪声幅度谱和语音幅度谱;

根据所述带噪语音幅度谱、所述噪声幅度谱和所述语音幅度谱生成所述带噪语音信号对应的频谱特征;

对所述声学特征和频谱特征进行转换,得到对应的声学特征向量和频谱特征向量;

获取分类器,将所述声学特征向量和频谱特征向量输入至所述分类器,得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量;

对所述添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析,得到对应的语音信号;

根据所述语音信号的时序确定所述语音信号对应的起始点和终止点。

2.根据权利要求1所述的方法,其特征在于,在所述提取所述带噪语音信号对应的声学特征和频谱特征之前,还包括:

将所述带噪语音信号转换为带噪语音频谱;

对所述带噪语音频谱进行时域分析和/或频域分析和/或变换域分析,得到所述带噪语音信号对应的声学特征。

3.根据权利要求1所述的方法,其特征在于,所述提取所述带噪语音信号的带噪语音幅度谱、噪声幅度谱和语音幅度谱,包括:

将所述带噪语音信号转换为带噪语音频谱,根据所述带噪语音频谱计算带噪语音幅度谱;

根据所述带噪语音幅度谱对所述带噪语音频谱进行动态噪声估计,得到噪声幅度谱;

根据所述带噪语音幅度谱和所述噪声幅度谱估计纯净语音信号的语音幅度谱。

4.根据权利要求1所述的方法,其特征在于,所述对所述声学特征和频谱特征进行转换包括:

提取所述声学特征和所述频谱特征中当前帧的前后预设数量帧;

通过利用当前帧的前后预设数量帧计算当前帧对应的均值矢量和/或方差矢量;

对计算当前帧对应的均值矢量和/或方差矢量后的声学特征和频谱特征进行对数域转换,得到转换后的声学特征向量和频谱特征向量。

5.根据权利要求1所述的方法,其特征在于,所述获取分类器的步骤之前还包括:

获取添加语音类别标签的带噪语音数据,通过对所述带噪语音数据进行训练,得到初始分类器;

获取第一验证集,所述第一验证集中包括多个第一语音数据;

将多个第一语音数据输入至所述初始分类器,得到所述多个第一语音数据对应的类别概率;

对多个第一语音数据对应的类别概率进行筛选,对选出的第一语音数据添加类别标签,得到添加类别标签的验证集;

利用所述添加类别标签的验证集和所述添加语音类别标签的带噪语音数据进行训练,得到验证分类器;

获取第二验证集,所述第二验证集中包括多个第二语音数据;

将多个第二语音数据输入至验证分类器，得到所述多个第二语音数据对应的类别概率；

当多个第二语音数据对应的类别概率达到预设概率值时，得到所需的分类器。

6.根据权利要求1至5任一项所述的方法，其特征在于，利用所述分类器对所述声学特征向量和频谱特征向量进行分类的步骤包括：

将所述声学特征向量和频谱特征向量作为分类器的输入，得到所述声学特征向量和频谱特征向量对应的决策值；

当所述决策值为第一阈值时，对所述声学特征向量或频谱特征向量添加语音标签；

当所述决策值为第二阈值时，对所述声学特征向量或频谱特征向量添加非语音标签。

7.一种语音端点检测装置，包括：

提取模块，用于获取带噪语音信号，提取所述带噪语音信号对应的声学特征；提取所述带噪语音信号的带噪语音幅度谱、噪声幅度谱和语音幅度谱；根据所述带噪语音幅度谱、所述噪声幅度谱和所述语音幅度谱生成所述带噪语音信号对应的频谱特征；

转换模块，用于对所述声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量；

分类模块，用于获取分类器，将所述声学特征向量和频谱特征向量输入至所述分类器，得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量；

解析模块，用于对所述添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到对应的语音信号；根据所述语音信号的时序确定所述语音信号对应的起始点和终止点。

8.根据权利要求7所述的装置，其特征在于，所述转换模块还用于提取所述声学特征和所述频谱特征中当前帧的前后预设数量帧；通过利用当前帧对应的前后预设数量帧计算当前帧的均值矢量和/或方差矢量；对计算当前帧对应的均值矢量和/或方差矢量后的声学特征和频谱特征进行对数域转换，得到转换后的声学特征向量和频谱特征向量。

9.一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，其特征在于，所述处理器执行所述计算机程序时实现权利要求1至6中任一项所述方法的步骤。

10.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现权利要求1至6中任一项所述方法的步骤。

语音端点检测方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及信号处理技术领域，特别是涉及一种语言端点检测方法、装置、计算机设备和存储介质。

背景技术

[0002] 随着语音技术的不断发展，语音端点检测技术在语音识别技术中占有十分重要的地位。语音端点检测是从一段连续的噪声语音中检测出语音部分的起始点和终止点，从而能够有效地识别出语音。

[0003] 传统的语音端点检测方式有两种，一种是根据语音和噪声信号的时域和频域的特征不同，提取每一段信号的特征，将每一段信号的特征与设定的阈值进行比较，从而进行语音端点检测。但这种方式仅适用于平稳噪声条件下检测，噪声鲁棒性差，很难区分纯净语音和噪声，导致语音端点检测的准确性较低。。另一种则是基于神经网络的方式，通过利用训练模型对语音信号进行端点检测。然而大多模型的输入向量只含有带噪语音的特征，使得噪声鲁棒性差，从而导致语音端点检测的准确性较低。因此，如何有效提高语音端点检测的准确性成为目前需要解决的技术问题。

发明内容

[0004] 基于此，有必要针对上述技术问题，提供一种能够有效提高语音端点检测的准确性的语音端点检测方法、装置、计算机设备和存储介质。

[0005] 一种语音端点检测方法，包括：

[0006] 获取带噪语音信号，提取所述带噪语音信号对应的声学特征和频谱特征；

[0007] 对所述声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量；

[0008] 获取分类器，将所述声学特征向量和频谱特征向量输入至所述分类器，得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量；

[0009] 对所述添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到对应的语音信号；

[0010] 根据所述语音信号的时序确定所述语音信号对应的起始点和终止点。

[0011] 在其中一个实施例中，在所述提取所述带噪语音信号对应的声学特征和频谱特征之前，还包括：

[0012] 将所述带噪语音信号转换为带噪语音频谱；

[0013] 对所述带噪语音频谱进行时域分析和/或频域分析和/或变换域分析，得到所述带噪语音信号对应的声学特征。

[0014] 在其中一个实施例中，在所述提取所述带噪语音信号对应的声学特征和频谱特征之前，还包括：

[0015] 将所述带噪语音信号转换为带噪语音频谱，根据所述带噪语音频谱计算带噪语音

幅度谱；

[0016] 根据所述带噪语音幅度谱对所述带噪语音频谱进行动态噪声估计，得到噪声幅度谱；

[0017] 根据所述带噪语音幅度谱和所述噪声幅度谱估计纯净语音信号的语音幅度谱；

[0018] 利用所述带噪语音幅度谱、所述噪声幅度谱和所述语音幅度谱生成所述带噪语音信号对应的频谱特征。

[0019] 在其中一个实施例中，所述对所述声学特征和频谱特征进行转换包括：

[0020] 提取所述声学特征和所述频谱特征中当前帧的前后预设数量帧；

[0021] 通过利用当前帧的前后预设数量帧计算当前帧对应的均值矢量和/或方差矢量；

[0022] 对计算当前帧对应的均值矢量和/或方差矢量后的声学特征和频谱特征进行对数域转换，得到转换后的声学特征向量和频谱特征向量。

[0023] 在其中一个实施例中，所述获取分类器的步骤之前还包括：

[0024] 获取添加语音类别标签的带噪语音数据，通过对所述带噪语音数据进行训练，得到初始分类器；

[0025] 获取第一验证集，所述第一验证集中包括多个第一语音数据；

[0026] 将多个第一语音数据输入至分类器，得到所述多个第一语音数据对应的类别概率；

[0027] 对多个第一语音数据对应的类别概率进行筛选，对选出的第一语音数据添加类别标签，得到添加类别标签的验证集；

[0028] 利用所述添加类别标签的验证集和所述训练集进行训练，得到验证分类器；

[0029] 获取第二验证集，所述第二验证集中包括多个第二语音数据；

[0030] 将多个第二语音数据输入至验证分类器，得到所述多个第二语音数据对应的类别概率；

[0031] 当多个第二语音数据对应的类别概率达到预设概率值时，得到所需的分类器。

[0032] 在其中一个实施例中，所述利用所述分类器对所述声学特征向量和频谱特征向量进行分类的步骤包括：

[0033] 将所述声学特征向量和频谱特征向量作为分类器的输入，得到所述声学特征向量和频谱特征向量对应的决策值；

[0034] 当所述决策值为第一阈值时，对所述声学特征向量或频谱特征向量添加语音标签；

[0035] 当所述决策值为第二阈值时，对所述声学特征向量或频谱特征向量添加非语音标签。

[0036] 一种语音端点检测装置，包括：

[0037] 提取模块，用于获取带噪语音信号，提取所述带噪语音信号对应的声学特征和频谱特征；

[0038] 转换模块，用于对所述声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量；

[0039] 分类模块，用于获取分类器，将所述声学特征向量和频谱特征向量输入至所述分类器，得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量；

[0040] 解析模块，用于对所述添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到对应的语音信号；根据所述语音信号的时序确定所述语音信号对应的起始点和终止点。

[0041] 在其中一个实施例中，所述转换模块还用于提取所述声学特征和所述频谱特征中当前帧的前后预设数量帧；通过利用当前帧的前后预设数量帧计算当前帧对应的均值矢量和/或方差矢量；对计算当前帧对应的均值矢量和/或方差矢量后的声学特征和频谱特征进行对数域转换，得到转换后的声学特征向量和频谱特征向量。

[0042] 一种计算机设备，包括存储器、处理器，所述存储器存储有计算机程序，所述处理器执行所述计算机程序时实现以下步骤：

[0043] 获取带噪语音信号，提取所述带噪语音信号对应的声学特征和频谱特征；

[0044] 对所述声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量；

[0045] 获取分类器，将所述声学特征向量和频谱特征向量输入至所述分类器，得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量；

[0046] 对所述添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到对应的语音信号；

[0047] 根据所述语音信号的时序确定所述语音信号对应的起始点和终止点。

[0048] 一种计算机可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现以下步骤：

[0049] 获取带噪语音信号，提取所述带噪语音信号对应的声学特征和频谱特征；

[0050] 对所述声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量；

[0051] 获取分类器，将所述声学特征向量和频谱特征向量输入至所述分类器，得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量；

[0052] 对所述添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到对应的语音信号；

[0053] 根据所述语音信号的时序确定所述语音信号对应的起始点和终止点。

[0054] 上述语音端点检测方法、装置、计算机设备和存储介质，获取带噪语音信号，提取带噪语音信号对应的声学特征和频谱特征；通过对声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量。获取分类器，通过将声学特征向量和频谱特征向量输入至分类器，得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量，由此能够有效地对声学特征向量和频谱特征向量进行分类，从而能够有效识别出语音和非语音。对添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到对应的语音信号；语音信号的时序确定语音信号对应的起始点和终止点，由此能够准确地识别带噪语音信号的起始点和终止点，从而能够有效提高语音端点检测的准确性。

附图说明

[0055] 图1为一个实施例中语音端点检测方法的流程图；

[0056] 图2为一个实施例中语音端点检测装置的的内部结构图；

[0057] 图3为一个实施例中计算机设备的内部结构图。

具体实施方式

[0058] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本发明进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定申请。可以理解，本发明所使用的术语“第一”、“第二”等可在本文中用于描述各种元件，但这些元件不受这些术语限制。这些术语仅用于将第一个元件与另一个元件区分。

[0059] 在一个实施例中，如图1所示，提供了一种语音端点检测方法，以该方法应用于终端为例进行说明，包括以下步骤：

[0060] 步骤102，获取带噪语音信号，提取带噪语音信号对应的声学特征和频谱特征。

[0061] 通常而言，实际采集到的语音信号通常含有一定强度的噪音，当这些噪音强度较大时，会对语音应用的效果产生明显的影响，比如语音识别效率低，端点检测准确性下降等。

[0062] 终端可以获取用户通过语音输入装置输入的语音。其中，终端可以是智能手机、平板电脑、笔记本电脑、台式电脑等终端，终端还包括语音输入装置，例如，可以是麦克风等具有录入语音功能的装置。终端获取到的用户输入的语音通常为含有噪声的带噪语音信号，带噪语音信号可以是用户输入的通话语音、录音音频、语音指令等带噪语音信号。终端获取带噪语音信号后，提取出带噪语音信号对应的声学特征和频谱特征。其中，声学特征可以包括带噪语音信号的清音、浊音，元音、辅音等特征信息。频谱特征可以包括带噪语音信号的振动频率、震动幅度以及带噪语音信号的响度、音色等特征信息。

[0063] 具体地，终端获取带噪语音信号后，对带噪语音信号进行加窗分帧。例如，可以采用汉宁窗将带噪语音信号分为多个帧长为10-30ms(毫秒)的帧，帧移可以取10ms，从而可以将带噪语音信号分为多帧带噪语音信号。终端对对带噪语音信号进行加窗分帧后，对加窗分帧后的带噪语音信号进行快速傅里叶转换，由此得到带噪语音信号的频谱。终端则可以根据带噪语音的频谱提取出带噪语音信号对应的声学特征和频谱特征。

[0064] 步骤104，对声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量。

[0065] 终端提取出带噪语音信号对应的声学特征和频谱特征后，对提取出的带噪语音信号对应的声学特征和频谱特征进行转换，将声学特征转换为对应的声学特征向量，将频谱特征转换为对应的频谱特征向量。

[0066] 步骤106，获取分类器，将声学特征向量和频谱特征向量输入至分类器，得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量。

[0067] 终端获取分类器，分类器为在进行语音端点检测之前训练好的分类器，分类器可以通过向声学特征向量和频谱特征向量添加语音标签和非语音标签，将输入的声学特征向量和频谱特征向量分为语音类的声学特征向量和频谱特征向量和非语音类的声学特征向量和频谱特征向量。终端通过将带噪语音对应的声学特征向量和频谱特征向量输入至分类器，利用分类器对输入的声学特征向量和频谱特征向量进行分类。当输入的声学特征向量或频谱特征向量为语音类别时，为声学特征向量或频谱特征向量添加语音标签；当输入的声学特征向量或频谱特征向量为非语音类别时，为声学特征向量或频谱特征向量添加非语

音标签，由此能够准确识别出语音和非语音。终端利用分类器对声学特征向量和频谱特征向量后，就可以得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量。

[0068] 进一步地，终端将声学特征向量和频谱特征向量作为分类器的输入，还可以得到声学特征向量和频谱特征向量对应的决策值。终端可以根据得到的决策值对声学特征向量和频谱特征向量添加语音标签或非语音标签。从而实现对声学特征向量和频谱特征向量进行准确分类。

[0069] 步骤108，对添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到添加语音标签后的语音信号。

[0070] 步骤110，根据语音信号的语音标签和时序确定语音信号对应的起始点和终止点。

[0071] 终端对声学特征向量和频谱特征向量进行分类后，需要对添加了语音标签的声学特征向量和添加了语音标签的频谱特征向量进行解析。具体地，终端将添加了语音标签的声学特征向量和添加了语音标签的频谱特征向量进行解析，得到添加了语音标签的声学特征和频谱特征对应的频谱。终端根据带噪语音信号的时序将添加了语音标签的声学特征和频谱特征对应的频谱转换为对应的语音信号，由此能够解析得到对应的语音信号。

[0072] 带噪语音信号具有时序，添加了语音标签后的语音信号的时序仍然与带噪语音信号的时序相对应。终端将添加了语音标签的声学特征向量和添加了语音标签的频谱特征向量解析为对应的添加了语音标签的语音信号，终端由此能够根据语音信号的语音标签和时序确定带噪语音信号对应的起始点和终止点。

[0073] 例如，终端通过分类器对输入的声学特征向量和频谱特征向量进行分类后，得到的决策值可以是一个0到1之间的值。当得到的决策值为1时，终端对声学特征向量或频谱特征向量添加语音标签。当得到的决策值为0时，终端对声学特征向量或频谱特征向量添加非语音标签。由此能够准确地对声学特征向量和频谱特征向量进行准确分类。终端将添加了语音标签的声学特征向量和添加了语音标签的频谱特征向量进行解析后，就可以得到添加了语音标签后的语音信号。根据添加了语音标签后的语音信号的时序，当第一次出现添加了语音标签的语音帧则为带噪语音信号的起始点，当最后一次出现语音标签对应的语音帧则为带噪语音信号的终止点。进一步地，还可以根据决策值0到1的跳转来确定语音信号的起始点，根据决策值1到0的跳转来确定语音信号的终止点。由此可以准确地确定带噪语音信号对应的起始点和终止点。

[0074] 本实施例中，终端获取带噪语音信号后，提取带噪语音信号对应的声学特征和频谱特征，通过对声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量。通过将声学特征向量和频谱特征向量输入至分类器，得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量，由此能够有效地对声学特征向量和频谱特征向量进行分类，从而能够有效识别出语音和非语音。终端通过对添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到对应的语音信号。终端根据语音信号的时序确定语音信号对应的起始点和终止点，由此能够准确地识别带噪语音信号的起始点和终止点，从而能够有效提高语音端点检测的准确性。

[0075] 在一个实施例中，在提取带噪语音信号对应的声学特征和频谱特征之前，还包括：将带噪语音信号转换为带噪语音频谱；对带噪语音频谱进行时域分析和/或频域分析和/或变换域分析，得到带噪语音信号对应的声学特征。

[0076] 在语音学中，语音特征可以分为元音、辅音、清音、浊音以及静音等声学特征。终端获取带噪语音信号后，对带噪语音信号进行加窗分帧。例如，可以采用汉宁窗将带噪语音信号分为多个帧长为10-30ms(毫秒)的帧，帧移可以取10ms。从而可以将带噪语音信号分为多帧带噪语音信号。终端对对带噪语音信号进行加窗分帧后，对加窗分帧后的带噪语音信号进行快速傅里叶转换，由此得到带噪语音信号的频谱。

[0077] 进一步地，终端可以对带噪语音频谱进行时域分析和/或频域分析和/或变换域分析，从而能够得到带噪语音信号对应的声学特征。

[0078] 例如，终端可以采用MFCC(Mel-Frequency Cepstrum Coefficients，梅尔频率倒谱系数)方式提取出带噪语音信号对应的声学特征。终端对带噪语音信号进行加窗分帧后，将带噪语音信号转换为带噪语音信号的频谱。终端将带噪语音信号的频谱变换为带噪语音倒谱，终端根据带噪语音倒谱进行倒谱分析，将带噪语音倒谱进行离散余弦变换，得到每一帧的声学特征，从而能够得到带噪语音有效的声学特征。

[0079] 在一个实施例中，在提取带噪语音信号对应的声学特征和频谱特征之前，还包括：将带噪语音信号转换为带噪语音频谱，根据带噪语音频谱计算带噪语音幅度谱；根据带噪语音幅度谱对带噪语音频谱进行动态噪声估计，得到噪声幅度谱；根据带噪语音幅度谱和噪声幅度谱估计纯净语音信号的语音幅度谱；利用带噪语音幅度谱、噪声幅度谱和语音幅度谱生成带噪语音信号对应的频谱特征。

[0080] 终端获取带噪语音信号后，对带噪语音信号进行加窗分帧。例如，可以采用汉宁窗将带噪语音信号分为多个帧长为10-30ms(毫秒)的帧，帧移可以取10ms。从而可以将带噪语音信号分为多帧带噪语音信号。终端对对带噪语音信号进行加窗分帧后，对加窗分帧后的带噪语音信号进行快速傅里叶转换，由此得到带噪语音信号的频谱。其中，带噪语音信号的频谱可以为经过快速傅里叶转换之后的带噪语音的能量幅度谱。

[0081] 进一步地，终端利用带噪语音频谱可以计算出带噪语音幅度谱和带噪语音相位谱。终端根据带噪语音幅度谱和带噪语音相位谱对带噪语音频谱进行动态噪声估计。具体地，终端可以利用改进最小受控递归平均算法对带噪语音频谱进行动态噪声估计，从而可以得到噪声幅度谱。终端根据带噪语音幅度谱、带噪语音相位谱和噪声幅度谱估计出语音信号的语音幅度谱。例如，终端可以利用对数幅度谱最小均方差估计法，估计出语音信号的语音幅度谱。

[0082] 终端利用估计出的带噪语音幅度谱、噪声幅度谱和纯净语音信号的语音幅度谱生成带噪语音信号对应的频谱特征，由此终端就可以有效地提取出带噪语音信号对应的频谱特征。

[0083] 在一个实施例中，对声学特征和频谱特征进行转换包括：提取声学特征和频谱特征中当前帧的前后预设数量帧；通过利用当前帧的前后预设数量帧计算当前帧对应的均值矢量和/或方差矢量；对计算当前帧对应的均值矢量和/或方差矢量后的声学特征和频谱特征进行对数域转换，得到转换后的声学特征向量和频谱特征向量。

[0084] 终端获取带噪语音信号后，对带噪语音信号进行加窗分帧。从而可以将带噪语音信号分为多帧带噪语音信号。终端对对带噪语音信号进行加窗分帧后，对加窗分帧后的带噪语音信号进行快速傅里叶转换，由此得到带噪语音信号的频谱。终端就可以根据带噪语音的频谱提取出带噪语音信号对应的声学特征和频谱特征。

[0085] 终端提取出带噪语音信号对应的声学特征和频谱特征后，将声学特征和频谱特征转换为声学特征向量和频谱特征向量。终端提取声学特征向量和频谱特征向量中当前帧的前后预设数量帧。终端通过利用当前帧的前后预设数量帧计算出当前帧对应的均值矢量或方差矢量，从而可以对声学特征和频谱特征进行平滑处理，得到平滑后的声学特征向量和频谱特征向量。

[0086] 例如，终端可以获取声学特征或频谱特征当前帧的前推和后续各五帧，总共11帧带噪语音频谱。通过计算这11帧的平均值，可以得到当前帧的均值矢量。具体地，终端可以获取滤波器组，其中，滤波器的形状为三角形，三角窗口表示滤波窗口。每个滤波器具有三角形滤波器的特性，在带噪语音频谱范围内，这些滤波器可以是等带宽的。终端可以利用滤波器组计算当前帧的均值矢量，由此可以对带噪语音频谱进行平滑处理，得到平滑后的声学特征向量和频谱特征向量。

[0087] 终端对带噪语音频谱进行平滑处理后，对平滑后的声学特征向量和平滑后的频谱特征向量计算对数域，得到转换后的声学特征向量和频谱特征向量。具体地，终端可以计算每个滤波器输出的声学特征和频谱特征的对数能量，由此可以得到声学特征向量的对数域和频谱特征向量的对数域，从而能够有效地得到转换后的声学特征向量和频谱特征向量。

[0088] 在一个实施例中，获取分类器的步骤之前还包括：获取添加语音类别标签的带噪语音数据，通过对带噪语音数据进行训练，得到初始分类器；获取第一验证集，第一验证集中包括多个第一语音数据；将多个第一语音数据输入至分类器，得到多个第一语音数据对应的类别概率；对多个第一语音数据对应的类别概率进行筛选，对选出的第一语音数据添加类别标签，得到添加类别标签的验证集；利用添加类别标签的验证集和训练集进行训练，得到验证分类器；获取第二验证集，第二验证集中包括多个第二语音数据；将多个第二语音数据输入至验证分类器，得到多个第二语音数据对应的类别概率；当多个第二语音数据对应的类别概率达到预设概率值时，得到所需的分类器。

[0089] 在获取分类器之前，需要利用大量带噪语音数据训练出分类器，这些大量带噪语音数据可以是终端从数据库中获取的带噪语音数据，也可以是终端从互联网中获取的带噪语音数据。在训练分类器时，首先通过人工对带噪语音数据进行标注，利用人工标注后的带噪语音数据进行训练得到分类器。

[0090] 具体地，终端提取带噪语音数据对应的声学特征和频谱特征后，对声学特征和频谱特征进行转换，转换为对应的声学特征向量和频谱特征向量。工作人员可以根据类别对照表对声学特征向量和频谱特征向量进行标注，对每一帧带噪语音信号添加语音标签或非语音标签。终端获取工作人员根据类别对照表对带噪语音数据进行标注后的带噪语音数据。

[0091] 终端将添加标签后的声学特征向量和频谱特征向量组合起来输入至LSTM(Bidirectional Long Short-term Memory，双向长短期记忆神经网络)的输入层，LSTM神经网络中的非线性隐藏层可以从输入的向量中学习到新的特征，通过激活函数计算出输入向量的所属类别。具体地，每个LSTM单元中有三个门，分别为遗忘门、候选门和输出门。具体的计算公式可以为：

[0092] $$f_i = \sigma(W_f^T \times h_{t-1} + U_f^T \times h_i + b_f)$$

[0093] 其中，σ表示激活函数，W_f^T 表示遗忘门权重矩阵，U_f^T 是遗忘门输入层与隐层之间的权重矩阵，b_f表示遗忘门的偏置，遗忘门是通过将前一隐层的输出h_{t-1}与当前的输入x_t进行了线性组合，然后利用激活函数将其输出值压缩到0到1之间。当输出值越靠近1，则表明记忆体保留的信息越多；反之，越靠近0，则表明记忆体保留的信息越少。

[0094] 侯选门计算当前输入的单元状态，具体公式可以为：

[0095] $$C_i = \tanh(W_c^T \times h_{t-1} + U_c^T \times h_c + b_c)$$

[0096] 其中，C_i表示当前输入的单元状态，通过tanh激活函数可以把输出值规整到-1和1之间。

[0097] 输出门可以控制用于下一层网络更新的记忆信息的数量，公式可以表示为：

[0098] $$O_t = \sigma(W_O^T \times h_{t-1} + U_O^T \times h_t + b_O)$$

[0099] 其中，O_t表示用于下一层网络更新的记忆信息的数量。

[0100] 通过LSTM单元可以计算得到最后的输出，公式可以表示为：

[0101] $h_t = O_t \times \tanh(C_t)$

[0102] 由正向和反向计算得到最后的声学特征向量或频谱特征向量，公式可以表示为：

[0103] $$h_i = \left| \overrightarrow{h_i}, \overleftarrow{h_i} \right|$$

[0104] 其中 $\overrightarrow{h_i}$ 为正向的输出向量，$\overleftarrow{h_i}$为反向的输出向量，h_i为最后的标注了类别标签的多个声学特征向量或频谱特征向量。

[0105] 进一步地，LSTM中的输出层可以根据预设的决策函数计算出输出单元C_i的值。其中，输出单元C_i的值可以为0至1之间的值，1代表语音类，0代表非语音类。

[0106] 终端利用标注了语音类别标签的多个声学特征向量和频谱特征向量计算出每个声学特征和频谱特征属于类别对照表中语音类别和非语音类别的概率，提取声学特征向量和频谱特征向量在类别对照表中的概率值最大的类别，对声学特征向量或频谱特征向量添加与概率值最大的类别对应的语音类别标签。

[0107] 终端利用添加了语音类别标签的带噪语音数据进行训练，得到初始分类器；终端获取第一验证集，第一验证集中包括多个第一语音数据。终端将多个第一语音数据输入至分类器，得到多个第一语音数据对应的类别概率后，对多个第一语音数据对应的类别概率进行筛选。工作人员利用终端对选出的第一语音数据添加语音类别标签，终端获取添加语音类别标签后的第一语音数据，利用添加语音类别标签后的第一语音数据生成添加语音类别标签的验证集。终端利用添加类别语音标签的验证集和带噪语音数据再进行训练，得到验证分类器。终端获取第二验证集，第二验证集中包括多个第二语音数据；将多个第二语音数据输入至验证分类器，得到多个第二语音数据对应的类别概率。终端筛选出类别概率在预设范围的第二语言数据，并将筛选出的第二语音数据再进行标注，将标注后的第二语音数据和添加标签后的带噪语音数据重新进行训练得到新的分类器。一直持续训练，直到所有验证集中预设数量的声学特征向量或频谱特征向量的概率值在预设概率范围值之间时，停止训练，就可以得到所需的分类器。由此可以得到准确率较高的分类器，从而实现对声学

特征向量和频谱特征向量进行准确分类，进而能够准确地识别出语音和非语音。

[0108] 在一个实施例中，利用分类器对声学特征向量和频谱特征向量进行分类的步骤包括：将声学特征向量和频谱特征向量作为分类器的输入，得到声学特征向量和频谱特征向量对应的决策值；当决策值为第一阈值时，对声学特征向量或频谱特征向量添加语音标签；当决策值为第二阈值时，对声学特征向量或频谱特征向量添加非语音标签。

[0109] 终端获取带噪语音信号后，提取带噪语音信号对应的声学特征和频谱特征。终端对声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量。终端获取分类器后，将声学特征向量和频谱特征向量输入至分类器。分类器对输入的声学特征向量和频谱特征向量进行分类后，可以得到声学特征向量和频谱特征向量对应的决策值。当得到的决策值为预设的第一阈值时，终端对声学特征向量或频谱特征向量添加语音标签。其中，第一阈值可以是一个范围值。当得到的决策值为预设的第二阈值时，终端对声学特征向量或频谱特征向量添加非语音标签。通过利用分类器对声学特征向量和频谱特征向量进行准确分类，从而能够准确地识别出带噪语音信号中语音信号和非语音信号。

[0110] 例如，得到的决策值可以是一个0到1之间的值。预设的第一阈值可以是1，预设的第二阈值可以是0。当得到的决策值为1时，终端对声学特征向量或频谱特征向量添加语音标签。当得到的决策值为0时，终端对声学特征向量或频谱特征向量添加非语音标签。由此能够准确地对声学特征向量和频谱特征向量进行准确分类。

[0111] 在一个实施例中，如图2所示，提供了一种语音端点检测装置，包括提取模块202、转换模块204、分类模块206和解析模块208，其中：

[0112] 提取模块202，用于获取带噪语音信号，提取带噪语音信号对应的声学特征和频谱特征。

[0113] 转换模块204，用于对声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量。

[0114] 分类模块206，用于获取分类器，将声学特征向量和频谱特征向量输入至所述分类器，得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量。

[0115] 解析模块208，用于对添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到对应的语音信号；根据语音信号的时序确定语音信号对应的起始点和终止点。

[0116] 在一个实施例中，提取模块202还用于将所述带噪语音信号转换为带噪语音频谱；对所述带噪语音频谱进行时域分析和/或频域分析和/或变换域分析，得到所述带噪语音信号对应的声学特征。

[0117] 在一个实施例中，提取模块202还用于将带噪语音信号转换为带噪语音频谱，根据带噪语音频谱计算带噪语音幅度谱；根据带噪语音幅度谱对带噪语音频谱进行动态噪声估计，得到噪声幅度谱；根据带噪语音幅度谱和噪声幅度谱估计纯净语音信号的语音幅度谱；利用带噪语音幅度谱、噪声幅度谱和语音幅度谱生成带噪语音信号对应的频谱特征。

[0118] 在一个实施例中，转换模块204还用于提取所述声学特征和所述频谱特征中当前帧的前后预设数量帧；通过利用当前帧的前后预设数量帧计算当前帧对应的均值矢量和/或方差矢量；对计算当前帧对应的均值矢量和/或方差矢量后的声学特征和频谱特征进行对数域转换，得到转换后的声学特征向量和频谱特征向量。

[0119] 在一个实施例中，该装置还包括训练模块，用于获取添加语音类别标签的带噪语音数据，通过对带噪语音数据进行训练，得到初始分类器；获取第一验证集，第一验证集中包括多个第一语音数据；将多个第一语音数据输入至初始分类器，得到多个第一语音数据对应的类别概率；对多个第一语音数据对应的类别概率进行筛选，对选出的第一语音数据添加类别标签，得到添加类别标签的验证集；利用添加类别标签的验证集和添加语音类别标签的带噪语音数据进行训练，得到验证分类器；获取第二验证集，第二验证集中包括多个第二语音数据；将多个第二语音数据输入至验证分类器，得到多个第二语音数据对应的类别概率；当多个第二语音数据对应的类别概率达到预设概率值时，得到所需的分类器。

[0120] 在一个实施例中，分类模块206还用于将声学特征向量和频谱特征向量作为分类器的输入，得到声学特征向量和频谱特征向量对应的决策值；当决策值为第一阈值时，对声学特征向量或频谱特征向量添加语音标签；当决策值为第二阈值时，对声学特征向量或频谱特征向量添加非语音标签。

[0121] 在一个实施例中，提供了一种计算机设备，该计算机设备可以是终端，其内部结构图可以如图3所示。例如，该计算机设备可以是终端，终端可以但不限于是各种是智能手机、平板电脑、笔记本电脑、个人计算机和便携式可穿戴设备等具有输入语音的功能的设备。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和语音输入装置。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种语音端点检测方法。该计算机设备的语音输入装置可以包括麦克风，还可以包括外接的耳机等。

[0122] 本领域技术人员可以理解，图3中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的服务器的限定，具体的服务器可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

[0123] 在一个实施例中，提供了一种计算机设备，包括存储器和处理器，存储器中存储有计算机程序，该处理器执行计算机程序时实现以下步骤：获取带噪语音信号，提取带噪语音信号对应的声学特征和频谱特征；对声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量；获取分类器，将声学特征向量和频谱特征向量输入至分类器，得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量；对添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到对应的语音信号；根据语音信号的时序确定语音信号对应的起始点和终止点。

[0124] 在一个实施例中，处理器执行计算机程序时还实现以下步骤：将所述带噪语音信号转换为带噪语音频谱；对所述带噪语音频谱进行时域分析和/或频域分析和/或变换域分析，得到所述带噪语音信号对应的声学特征。

[0125] 在一个实施例中，处理器执行计算机程序时还实现以下步骤：将带噪语音信号转换为带噪语音频谱，根据带噪语音频谱计算带噪语音幅度谱；根据带噪语音幅度谱对带噪语音频谱进行动态噪声估计，得到噪声幅度谱；根据带噪语音幅度谱和噪声幅度谱估计纯净语音信号的语音幅度谱；利用带噪语音幅度谱、噪声幅度谱和语音幅度谱生成带噪语音信号对应的频谱特征。

[0126] 在一个实施例中，处理器执行计算机程序时还实现以下步骤：提取所述声学特征和所述频谱特征中当前帧的前后预设数量帧；通过利用当前帧的前后预设数量帧计算当前帧对应的均值矢量和/或方差矢量；对计算当前帧对应的均值矢量和/或方差矢量后的声学特征和频谱特征进行对数域转换，得到转换后的声学特征向量和频谱特征向量。

[0127] 在一个实施例中，处理器执行计算机程序时还实现以下步骤：获取添加语音类别标签的带噪语音数据，通过对带噪语音数据进行训练，得到初始分类器；获取第一验证集，第一验证集中包括多个第一语音数据；将多个第一语音数据输入至分类器，得到多个第一语音数据对应的类别概率；对多个第一语音数据对应的类别概率进行筛选，对选出的第一语音数据添加类别标签，得到添加类别标签的验证集；利用添加类别标签的验证集和训练集进行训练，得到验证分类器；获取第二验证集，第二验证集中包括多个第二语音数据；将多个第二语音数据输入至验证分类器，得到多个第二语音数据对应的类别概率；当多个第二语音数据对应的类别概率达到预设概率值时，得到所需的分类器。

[0128] 在一个实施例中，处理器执行计算机程序时还实现以下步骤：将声学特征向量和频谱特征向量作为分类器的输入，得到声学特征向量和频谱特征向量对应的决策值；当决策值为第一阈值时，对声学特征向量或频谱特征向量添加语音标签；当决策值为第二阈值时，对声学特征向量或频谱特征向量添加非语音标签。

[0129] 在一个实施例中，提供了一种计算机可读存储介质，其上存储有计算机程序，计算机程序被处理器执行时实现以下步骤：获取带噪语音信号，提取带噪语音信号对应的声学特征和频谱特征；对声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量；获取分类器，将声学特征向量和频谱特征向量输入至分类器，得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量；对添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到对应的语音信号；根据语音信号的时序确定语音信号对应的起始点和终止点。

[0130] 在一个实施例中，计算机程序被处理器执行时还实现以下步骤：将所述带噪语音信号转换为带噪语音频谱；对所述带噪语音频谱进行时域分析和/或频域分析和/或变换域分析，得到所述带噪语音信号对应的声学特征。

[0131] 在一个实施例中，计算机程序被处理器执行时还实现以下步骤：将带噪语音信号转换为带噪语音频谱，根据带噪语音频谱计算带噪语音幅度谱；根据带噪语音幅度谱对带噪语音频谱进行动态噪声估计，得到噪声幅度谱；根据带噪语音幅度谱和噪声幅度谱估计纯净语音信号的语音幅度谱；利用带噪语音幅度谱、噪声幅度谱和语音幅度谱生成带噪语音信号对应的频谱特征。

[0132] 在一个实施例中，计算机程序被处理器执行时还实现以下步骤：提取所述声学特征和所述频谱特征中当前帧的前后预设数量帧；通过利用当前帧的前后预设数量帧计算当前帧对应的均值矢量和/或方差矢量；对计算当前帧对应的均值矢量和/或方差矢量后的声学特征和频谱特征进行对数域转换，得到转换后的声学特征向量和频谱特征向量。

[0133] 在一个实施例中，计算机程序被处理器执行时还实现以下步骤：获取添加语音类别标签的带噪语音数据，通过对带噪语音数据进行训练，得到初始分类器；获取第一验证集，第一验证集中包括多个第一语音数据；将多个第一语音数据输入至分类器，得到多个第一语音数据对应的类别概率；对多个第一语音数据对应的类别概率进行筛选，对选出的第

一语音数据添加类别标签，得到添加类别标签的验证集；利用添加类别标签的验证集和训练集进行训练，得到验证分类器；获取第二验证集，第二验证集中包括多个第二语音数据；将多个第二语音数据输入至验证分类器，得到多个第二语音数据对应的类别概率；当多个第二语音数据对应的类别概率达到预设概率值时，得到所需的分类器。

[0134] 在一个实施例中，计算机程序被处理器执行时还实现以下步骤：将声学特征向量和频谱特征向量作为分类器的输入，得到声学特征向量和频谱特征向量对应的决策值；当决策值为第一阈值时，对声学特征向量或频谱特征向量添加语音标签；当决策值为第二阈值时，对声学特征向量或频谱特征向量添加非语音标签。

[0135] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一非易失性计算机可读取存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0136] 以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

[0137] 以上所述实施例仅表达了本申请的几种实施方式，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，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

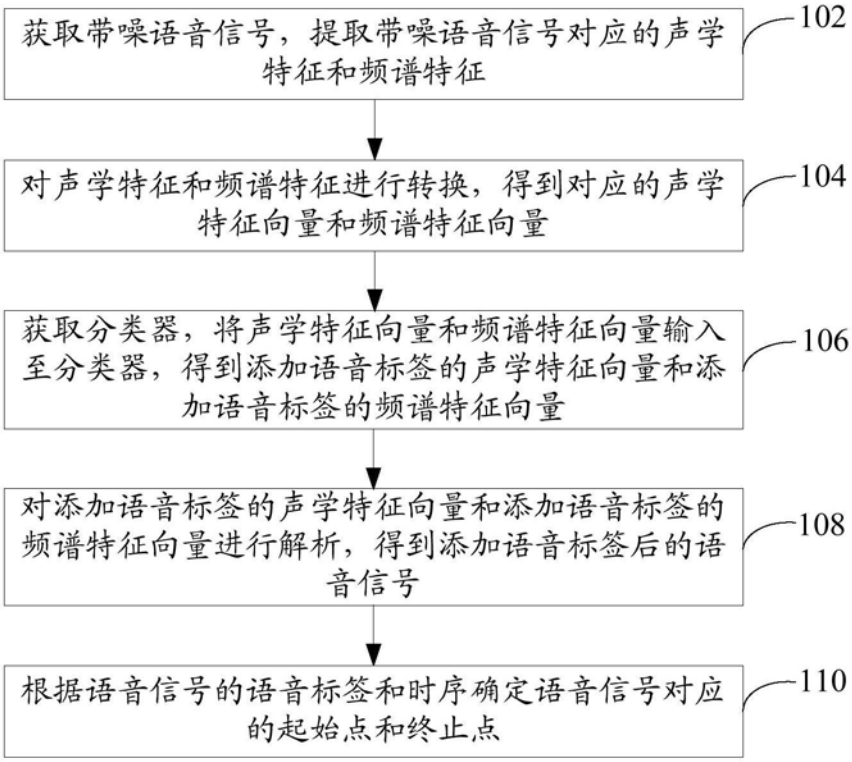

获取带噪语音信号，提取带噪语音信号对应的声学特征和频谱特征
102
对声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量
104
获取分类器，将声学特征向量和频谱特征向量输入至分类器，得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量
106
对添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到添加语音标签后的语音信号
108
根据语音信号的语音标签和时序确定语音信号对应的起始点和终止点
110

图1

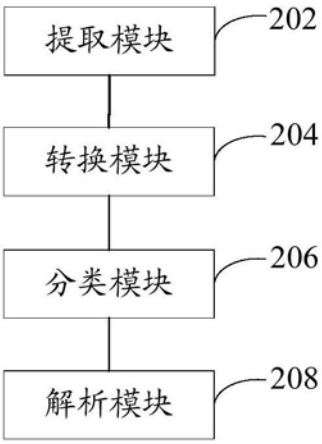

提取模块
202
转换模块
204
分类模块
206
解析模块
208

图2

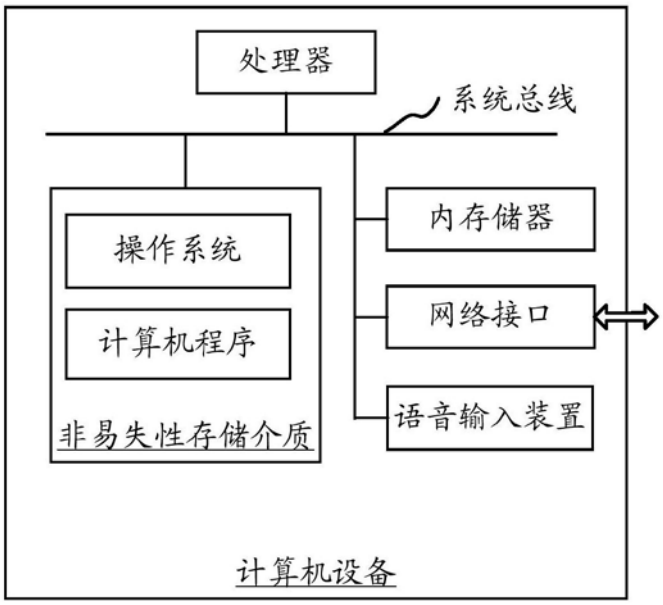


图3