(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 109658920 B
(45)授权公告日 2020.10.09

(21)申请号 201811550079.X

(22)申请日 2018.12.18

(65)同一申请的已公布的文献号
申请公布号 CN 109658920 A

(43)申请公布日 2019.04.19

(73)专利权人 百度在线网络技术(北京)有限公司
地址 100085 北京市海淀区上地十街10号百度大厦三层

(72)发明人 李超

(74)专利代理机构 北京英赛嘉华知识产权代理有限责任公司 11204
代理人 王达佐 马晓亚

(51)Int.Cl.
G10L 15/06(2013.01)
G10L 15/04(2013.01)
G10L 19/005(2013.01)
G10L 19/24(2013.01)
G10L 25/30(2013.01)
G10L 25/87(2013.01)

(56)对比文件
CN 108922513 A,2018.11.30
CN 106531190 A,2017.03.22
CN 108847238 A,2018.11.20
CN 107799126 A,2018.03.13
Florian Eyben等.REAL-LIFE VOICE ACTIVITY DETECTION WITH LSTM RECURRENT NEURAL.《 2013 IEEE International Conference on Acoustics, Speech and Signal Processing》.2013,
Gyeowoon Jung等.DNN-GRU Multiple Layers for VAD in PC Game Cafe.《2018 IEEE International Conference on Consumer Electronics - Asia》.2018,

审查员 张岩

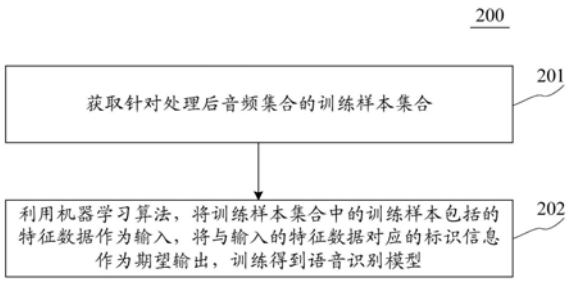
权利要求书3页 说明书14页 附图7页

(54)发明名称

用于生成模型的方法和装置

(57)摘要

本申请实施例公开了用于生成模型的方法和装置,以及用于检测语音的方法和装置。该用于生成模型的方法的一具体实施方式包括:获取针对处理后音频集合的训练样本集合,其中,处理后音频集合包括对处理前音频执行音质劣化处理得到的音频,训练样本与处理后音频一一对应,训练样本包括处理后音频的特征数据和标识信息,标识信息用于指示处理后音频中是否包括语音音频;利用机器学习算法,将训练样本集合中的训练样本包括的特征数据作为输入,将与输入的特征数据对应的标识信息作为期望输出,训练得到语音识别模型。该实施方式丰富了模型的训练方式,有助于提高语音端点识别的准确度。



CN 109658920 B

1.一种用于生成模型的方法,包括:

获取针对处理后音频集合的训练样本集合,其中,所述处理后音频集合包括对处理前音频执行音质劣化处理得到的音频,所述音质劣化处理包括丢帧处理和置零处理中的至少一项,置零处理为将处理前音频的属性的属性值设置为零的处理,训练样本与处理后音频一一对应,训练样本包括处理后音频的特征数据和标识信息,标识信息用于指示处理后音频中是否包括语音音频;

利用机器学习算法,将所述训练样本集合中的训练样本包括的特征数据作为输入,将与输入的特征数据对应的标识信息作为期望输出,训练得到语音识别模型。

2.根据权利要求1所述的方法,其中,在所述音质劣化处理包括丢帧处理的情况下,所述处理后音频集合包括丢帧音频,丢帧音频是通过针对处理前音频集合中的处理前音频,执行如下处理得到的:

对处理前音频进行丢帧处理,得到丢帧音频作为处理后音频。

3.根据权利要求1所述的方法,其中,在所述音质劣化处理包括置零处理的情况下,所述处理后音频集合包括置零音频,置零音频是通过针对处理前音频集合中的处理前音频执行如下处理得到的:

对处理前音频进行置零处理,得到置零音频作为处理后音频。

4.根据权利要求1所述的方法,其中,在所述处理后音频集合包括丢帧音频和置零音频的情况下,所述处理后音频集合包括的丢帧音频的数量与所述处理后音频集合中的处理后音频的数量之比为预先确定的第一数值,所述处理后音频集合包括的置零音频的数量与所述处理后音频集合中的处理后音频的数量之比为预先确定的第二数值,其中,所述第一数值是音频集合中丢帧音频的数量与音频集合中音频的数量的比值,所述第二数值是音频集合中置零音频的数量与音频集合中音频的数量的比值。

5.根据权利要求1所述的方法,其中,在所述处理后音频集合包括丢帧音频和置零音频的情况下,所述处理后音频集合是通过针对处理前音频集合中的处理前音频,执行如下处理步骤得到的:

随机生成第一随机数和第二随机数,其中,所述第一随机数和所述第二随机数均为0到1之间的数;

响应于确定第一随机数大于等于预先确定的第一数值,并且,第二随机数小于预先确定的第二数值,将该处理前音频的所述属性的属性值设置为零,其中,所述第一数值用于表征音频中出现丢帧音频的概率,所述第二数值用于表征音频中出现置零音频的概率;

响应于确定第一随机数小于所述第一数值,并且,第二随机数大于等于所述第二数值,对该处理前音频进行丢帧处理。

6.根据权利要求1所述的方法,其中,所述属性为幅值。

7.根据权利要求1-6之一所述的方法,其中,所述语音识别模型为具有门控循环单元的循环神经网络模型。

8.一种用于检测语音的方法,包括:

获取目标音频,其中,所述目标音频包括语音音频;

针对所述目标音频包括的至少一个音频帧中的音频帧,将该音频帧的特征数据输入至预先训练的语音识别模型,得到标识信息,其中,所述语音识别模型是按照如权利要求1-7

之一所述的方法训练得到的；

基于所得到的标识信息集合，生成所述目标音频的语音端点检测结果。

9.一种用于生成模型的装置，包括：

第一获取单元，被配置成获取针对处理后音频集合的训练样本集合，其中，所述处理后音频集合包括对处理前音频执行音质劣化处理得到的音频，所述音质劣化处理包括丢帧处理和置零处理中的至少一项，置零处理为将处理前音频的属性的属性值设置为零的处理，训练样本与处理后音频一一对应，训练样本包括处理后音频的特征数据和标识信息，标识信息用于指示处理后音频中是否包括语音音频；

训练单元，被配置成利用机器学习算法，将所述训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型。

10.根据权利要求9所述的装置，其中，在所述音质劣化处理包括丢帧处理的情况下，所述处理后音频集合包括丢帧音频，丢帧音频是通过针对处理前音频集合中的处理前音频，执行如下处理得到的：

对处理前音频进行丢帧处理，得到丢帧音频作为处理后音频。

11.根据权利要求9所述的装置，其中，在所述音质劣化处理包括置零处理的情况下，所述处理后音频集合包括置零音频，置零音频是通过针对处理前音频集合中的处理前音频执行如下处理得到的：

对处理前音频进行置零处理，得到置零音频作为处理后音频。

12.根据权利要求9所述的装置，其中，在所述处理后音频集合包括丢帧音频和置零音频的情况下，所述处理后音频集合包括的丢帧音频的数量与所述处理后音频集合中的处理后音频的数量之比为预先确定的第一数值，所述处理后音频集合包括的置零音频的数量与所述处理后音频集合中的处理后音频的数量之比为预先确定的第二数值，其中，所述第一数值是音频集合中丢帧音频的数量与音频集合中音频的数量的比值，所述第二数值是音频集合中置零音频的数量与音频集合中音频的数量的比值。

13.根据权利要求9所述的装置，其中，在所述处理后音频集合包括丢帧音频和置零音频的情况下，所述处理后音频集合是通过针对处理前音频集合中的处理前音频，执行如下处理步骤得到的：

随机生成第一随机数和第二随机数，其中，所述第一随机数和所述第二随机数均为0到1之间的数；

响应于确定第一随机数大于等于预先确定的第一数值，并且，第二随机数小于预先确定的第二数值，将该处理前音频的所述属性的属性值设置为零，其中，所述第一数值用于表征音频中出现丢帧音频的概率，所述第二数值用于表征音频中出现置零音频的概率；

响应于确定第一随机数小于所述第一数值，并且，第二随机数大于等于所述第二数值，对该处理前音频进行丢帧处理。

14.根据权利要求9所述的装置，其中，所述属性为幅值。

15.根据权利要求9-14之一所述的装置，其中，所述语音识别模型为具有门控循环单元的循环神经网络模型。

16.一种用于检测语音的装置，包括：

第二获取单元，被配置成获取目标音频，其中，所述目标音频包括语音音频；

输入单元，被配置成针对所述目标音频包括的至少一个音频帧中的音频帧，将该音频帧的特征数据输入至预先训练的语音识别模型，得到标识信息，其中，所述语音识别模型是按照如权利要求1-7之一所述的方法训练得到的；

生成单元，被配置成基于所得到的标识信息集合，生成所述目标音频的语音端点检测结果。

17.一种电子设备，包括：

一个或多个处理器；

存储装置，其上存储有一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求1-8中任一所述的方法。

18.一种计算机可读介质，其上存储有计算机程序，其中，所述程序被处理器执行时实现如权利要求1-8中任一所述的方法。

用于生成模型的方法和装置

技术领域

[0001] 本申请实施例涉及计算机技术领域，具体涉及用于生成模型的方法和装置。

背景技术

[0002] 语音交互中很重要的一点是能够在音频中，判断语音的起点和终点在音频中的位置。现有技术中，通常采用语音活动检测(Voice Activity Detection，VAD)来进行语音的端点检测。语音活动检测，又称语音端点检测、语音边界检测，是指在噪声环境中检测语音的存在与否。通常，语音活动检测可以用于语音编码、语音增强等语音处理系统中，起到降低语音编码速率、节省通信带宽、减少移动设备能耗、提高识别率等作用。

发明内容

[0003] 本申请实施例提出了用于生成模型的方法和装置，以及用于检测语音的方法和装置。

[0004] 第一方面，本申请实施例提供了一种用于生成模型的方法，该方法包括：获取针对处理后音频集合的训练样本集合，其中，处理后音频集合包括对处理前音频执行音质劣化处理得到的音频，训练样本与处理后音频一一对应，训练样本包括处理后音频的特征数据和标识信息，标识信息用于指示处理后音频中是否包括语音音频；利用机器学习算法，将训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型。

[0005] 在一些实施例中，音质劣化处理包括丢帧处理，处理后音频集合包括丢帧音频，丢帧音频是通过针对处理前音频集合中的处理前音频，执行如下处理得到的：对处理前音频进行丢帧处理，得到丢帧音频作为处理后音频。

[0006] 在一些实施例中，音质劣化处理包括置零处理，处理后音频集合包括置零音频，置零处理为将处理前音频的属性的属性值设置为零的处理，置零音频是通过针对处理前音频集合中的处理前音频执行如下处理得到的：对处理前音频进行置零处理，得到置零音频作为处理后音频。

[0007] 在一些实施例中，处理后音频集合包括丢帧音频和置零音频，处理后音频集合包括的丢帧音频的数量与处理后音频集合中的处理后音频的数量之比为预先确定的第一数值，处理后音频集合包括的置零音频的数量与处理后音频集合中的处理后音频的数量之比为预先确定的第二数值，其中，第一数值是音频集合中丢帧音频的数量与音频集合中音频的数量的比值，第二数值是音频集合中置零音频的数量与音频集合中音频的数量的比值。

[0008] 在一些实施例中，处理后音频集合包括丢帧音频和置零音频，处理后音频集合是通过针对处理前音频集合中的处理前音频，执行如下处理步骤得到的：随机生成第一随机数和第二随机数，其中，第一随机数和第二随机数均为0到1之间的数；响应于确定第一随机数大于等于预先确定的第一数值，并且，第二随机数小于预先确定的第二数值，将该处理前音频的属性的属性值设置为零，其中，第一数值用于表征音频中出现丢帧音频的概率，第二

数值用于表征音频中出现置零音频的概率;响应于确定第一随机数小于第一数值,并且,第二随机数大于等于第二数值,对该处理前音频进行丢帧处理。

[0009] 在一些实施例中,上述属性为幅值。

[0010] 在一些实施例中,语音识别模型为具有门控循环单元的循环神经网络模型。

[0011] 第二方面,本申请实施例提供了一种用于生成模型的装置,该装置包括:第一获取单元,被配置成获取针对处理后音频集合的训练样本集合,其中,处理后音频集合包括对处理前音频执行音质劣化处理得到的音频,训练样本与处理后音频一一对应,训练样本包括处理后音频的特征数据和标识信息,标识信息用于指示处理后音频中是否包括语音音频;训练单元,被配置成利用机器学习算法,将训练样本集合中的训练样本包括的特征数据作为输入,将与输入的特征数据对应的标识信息作为期望输出,训练得到语音识别模型。

[0012] 在一些实施例中,音质劣化处理包括丢帧处理,处理后音频集合包括丢帧音频,丢帧音频是通过针对处理前音频集合中的处理前音频,执行如下处理得到的:对处理前音频进行丢帧处理,得到丢帧音频作为处理后音频。

[0013] 在一些实施例中,音质劣化处理包括置零处理,处理后音频集合包括置零音频,置零处理为将处理前音频的属性的属性值设置为零的处理,置零音频是通过针对处理前音频集合中的处理前音频执行如下处理得到的:对处理前音频进行置零处理,得到置零音频作为处理后音频。

[0014] 在一些实施例中,处理后音频集合包括丢帧音频和置零音频,处理后音频集合包括的丢帧音频的数量与处理后音频集合中的处理后音频的数量之比为预先确定的第一数值,处理后音频集合包括的置零音频的数量与处理后音频集合中的处理后音频的数量之比为预先确定的第二数值,其中,第一数值是音频集合中丢帧音频的数量与音频集合中音频的数量的比值,第二数值是音频集合中置零音频的数量与音频集合中音频的数量的比值。

[0015] 在一些实施例中,处理后音频集合包括丢帧音频和置零音频,处理后音频集合是通过针对处理前音频集合中的处理前音频,执行如下处理步骤得到的:随机生成第一随机数和第二随机数,其中,第一随机数和第二随机数均为0到1之间的数;响应于确定第一随机数大于等于预先确定的第一数值,并且,第二随机数小于预先确定的第二数值,将该处理前音频的属性的属性值设置为零,其中,第一数值用于表征音频中出现丢帧音频的概率,第二数值用于表征音频中出现置零音频的概率;响应于确定第一随机数小于第一数值,并且,第二随机数大于等于第二数值,对该处理前音频进行丢帧处理。

[0016] 在一些实施例中,上述属性为幅值。

[0017] 在一些实施例中,语音识别模型为具有门控循环单元的循环神经网络模型。

[0018] 第三方面,本申请实施例提供了一种用于检测语音的方法,该方法包括:获取目标音频,其中,目标音频包括语音音频;针对目标音频包括的至少一个音频帧中的音频帧,将该音频帧的特征数据输入至预先训练的语音识别模型,得到标识信息,其中,语音识别模型是如上述用于生成模型的方法中任一实施例的方法训练得到的;基于所得到的标识信息集合,生成目标音频的语音端点检测结果。

[0019] 第四方面,本申请实施例提供了一种用于生成模型的装置,该装置包括:第二获取单元,被配置成获取目标音频,其中,目标音频包括语音音频;输入单元,被配置成针对目标音频包括的至少一个音频帧中的音频帧,将该音频帧的特征数据输入至预先训练的语音识

别模型，得到标识信息，其中，语音识别模型是如上述用于生成模型的方法中任一实施例的方法训练得到的；生成单元，被配置成基于所得到的标识信息集合，生成目标音频的语音端点检测结果。

[0020] 第五方面，本申请实施例提供了一种电子设备，包括：一个或多个处理器；存储装置，其上存储有一个或多个程序，当上述一个或多个程序被上述一个或多个处理器执行，使得该一个或多个处理器实现如上述用于生成模型的方法中任一实施例的方法，或者，使得该一个或多个处理器实现如上述用于生成信息的方法中任一实施例的方法。

[0021] 第六方面，本申请实施例提供了一种计算机可读介质，其上存储有计算机程序，该程序被处理器执行时实现如上述用于生成模型的方法中任一实施例的方法，或者，该程序被处理器执行时实现如上述用于生成信息的方法中任一实施例的方法。

[0022] 本申请实施例提供的用于生成模型的方法和装置，通过获取针对处理后音频集合的训练样本集合，其中，处理后音频集合包括对处理前音频执行音质劣化处理得到的音频，训练样本与处理后音频一一对应，训练样本包括处理后音频的特征数据和标识信息，标识信息用于指示处理后音频中是否包括语音音频，然后，利用机器学习算法，将训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型，从而丰富了模型的训练方式，有助于提高语音端点识别的准确度。

附图说明

[0023] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述，本申请的其它特征、目的和优点将会变得更明显：

[0024] 图1是本申请的一个实施例可以应用于其中的示例性系统架构图；

[0025] 图2是根据本申请的用于生成模型的方法的一个实施例的流程图；

[0026] 图3A是根据本申请的用于生成模型的方法的一个实施例的处理前音频的波形示意图；

[0027] 图3B和图3C是针对图3A的处理前音频进行置零处理的操作示意图；

[0028] 图3D和图3E是针对图3A的处理前音频进行丢帧处理的操作示意图；

[0029] 图4是根据本申请的用于生成模型的方法的一个应用场景的示意图；

[0030] 图5是根据本申请的用于生成模型的方法的又一个实施例的流程图；

[0031] 图6是根据本申请的用于生成模型的装置的一个实施例的结构示意图；

[0032] 图7是根据本申请的用于检测语音的方法的一个实施例的流程图；

[0033] 图8是根据本申请的用于检测语音的装置的一个实施例的结构示意图；

[0034] 图9是适于用来实现本申请实施例的电子设备的计算机系统的结构示意图。

具体实施方式

[0035] 下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与有关发明相关的部分。

[0036] 需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相

互组合。下面将参考附图并结合实施例来详细说明本申请。

[0037] 图1示出了可以应用本申请实施例的用于生成模型的方法或用于生成模型的装置，或者，用于检测语音的方法或用于检测语音的装置的实施例的示例性系统架构100。

[0038] 如图1所示，系统架构100可以包括终端设备101、102、103，网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型，例如有线、无线通信链路或者光纤电缆等等。

[0039] 用户可以使用终端设备101、102、103通过网络104与服务器105交互，以接收或发送消息等。终端设备101、102、103上可以安装有各种通讯客户端应用，例如语音识别类应用、网页浏览器应用、购物类应用、搜索类应用、即时通信工具、邮箱客户端、社交平台软件等。

[0040] 终端设备101、102、103可以是硬件，也可以是软件。当终端设备101、102、103为硬件时，可以是具有音频传输功能的各种电子设备，包括但不限于智能手机、平板电脑、电子书阅读器、MP3播放器(Moving Picture Experts Group Audio Layer III，动态影像专家压缩标准音频层面3)、MP4(Moving Picture Experts Group Audio Layer IV，动态影像专家压缩标准音频层面4)播放器、膝上型便携计算机和台式计算机等等。当终端设备101、102、103为软件时，可以安装在上述所列举的电子设备中。其可以实现成多个软件或软件模块(例如用来提供分布式服务的软件或软件模块)，也可以实现成单个软件或软件模块。在此不做具体限定。

[0041] 服务器105可以是提供各种服务的服务器，例如对终端设备101、102、103发送的音频提供支持的后台服务器。后台服务器可以对接收到的音频进行音频特征提取等处理，并生成处理结果(例如提取的音频特征)。

[0042] 需要说明的是，本申请实施例所提供的用于生成模型的方法可以由服务器105执行，也可以由终端设备101、102、103执行，相应地，用于生成模型的装置可以设置于服务器105中，也可以设置于终端设备101、102、103中。此外，本申请实施例所提供的用于检测语音的方法可以由服务器105执行，也可以由终端设备101、102、103执行，相应地，用于检测语音的装置可以设置于服务器105中，也可以设置于终端设备101、102、103中。在这里，上述用于生成模型的方法与用于检测语音的方法的执行主体可以相同，也可以不同。

[0043] 需要说明的是，服务器可以是硬件，也可以是软件。当服务器为硬件时，可以实现成多个服务器组成的分布式服务器集群，也可以实现成单个服务器。当服务器为软件时，可以实现成多个软件或软件模块(例如用来提供分布式服务的软件或软件模块)，也可以实现成单个软件或软件模块。在此不做具体限定。

[0044] 应该理解，图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端设备、网络和服务器。例如，当用于生成模型方法运行于其上的电子设备不需要与其他电子设备进行数据传输时，该系统架构可以仅包括用于生成模型方法运行于其上的电子设备。

[0045] 继续参考图2，示出了根据本申请的用于生成模型的方法的一个实施例的流程200。该用于生成模型的方法，包括以下步骤：

[0046] 步骤201，获取针对处理后音频集合的训练样本集合。

[0047] 在本实施例中，用于生成模型的方法的执行主体(例如图1所示的服务器或终端设

备)可以通过有线连接方式或者无线连接方式从其他电子设备或者本地获取针对处理后音频集合的训练样本集合。其中,处理后音频集合包括对处理前音频执行音质劣化处理得到的音频。训练样本与处理后音频一一对应,训练样本包括处理后音频的特征数据和标识信息。标识信息用于指示处理后音频中是否包括语音音频。

[0048] 上述处理后音频可以是各种音频,例如,上述处理后音频可以包括但不限于以下任意一项:带噪音的语音音频、包括背景音和前景音的音频、包括静音和非静音的音频等等。该音频可以是任意长度的音频,例如,一句话;也可以是音频帧,其中,音频帧的长度可以是预先设置的,例如帧长可以是32毫秒、30毫秒等等。上述处理前音频可以包括但不限于以下任意一项:带噪音的语音音频、包括背景音和前景音的音频、包括静音和非静音的音频等等。

[0049] 在这里,上述音质劣化处理可以是对处理前音频的保真度进行降低的处理。经过音质劣化处理后得到的处理后音频,相对于未经音质劣化处理的处理前音频的保真度存在降低。上述保真度可以包括但不限于以下至少一项:清晰程度、不失真程度、再现平面声象的程度等等。可以理解,上述音质劣化处理可以是在处理前音频的音频信号中加入信号,或者,删除信号,或者,对信号进行调整的处理。作为示例,上述音质劣化处理可以包括但不限于以下任一项:置零处理,丢帧处理,加入杂音处理等等。在这里,对处理前音频进行置零处理所得到的处理后音频可以是置零音频,对处理前音频进行丢帧处理所得到的处理后音频可以是丢帧音频。

[0050] 上述丢帧音频可以是对处理前音频进行丢帧处理得到的音频。具体地,可以采用现有的各种方式对处理前音频进行丢帧处理,以得到丢帧音频。

[0051] 上述置零音频可以是将处理前音频的以下任意一项属性的属性值设置为零之后得到的:幅值、频率、振幅、音调等等。

[0052] 在本实施例的一些可选的实现方式中,上述属性可以为幅值。即,上述置零音频可以是将处理前音频包括的一帧或多帧音频帧的幅值设置为零之后得到的音频。

[0053] 上述特征数据可以包括但不限于音频的以下至少一项特征的数据:幅值、帧率、过零率、短时能量等。

[0054] 作为示例,请参考图3A-图3E。图3A是根据本申请的用于生成模型的方法的一个实施例的处理前音频的波形示意图。图3B和图3C是针对图3A的处理前音频进行置零处理的操作示意图。图3D和图3E是针对图3A的处理前音频进行丢帧处理的操作示意图。

[0055] 如图3B所示,如果上述执行主体或者与上述执行主体通信连接的其他电子设备对处理前音频包括的音频帧301进行置零操作,那么,上述执行主体或者与上述执行主体通信连接的其他电子设备可以将处理前音频包括的音频帧301的属性(例如幅值)的属性值设置为零,从而得到处理后音频(如图3C所示)。在此场景下,所得到的处理后音频为置零音频。

[0056] 下面请参考图3D,如果上述执行主体或者与上述执行主体通信连接的其他电子设备对处理前音频包括的音频帧302进行丢帧操作,那么,上述执行主体或者与上述执行主体通信连接的其他电子设备可以从处理前音频中删除(即丢弃)音频帧302,从而得到处理后音频(如图3E所示)。在此场景下,所得到的处理后音频为丢帧音频。可以理解,对处理前音频进行丢帧处理,所得到的处理后音频中将不包括所删除的音频帧的任何信息。

[0057] 可以理解,处理后音频集合包括以下至少一项:丢帧音频、置零音频。作为示例,上

述处理后音频集合可以包括未处理的音频和丢帧音频;也可以包括未处理的音频和置零音频;还可以包括未处理的音频、丢帧音频和置零音频等等。其中,上述未处理的音频即处理前音频,即可以将处理前音频直接确定为上述处理后音频集合中的处理后音频。

[0058] 在本实施例的一些可选的实现方式中,音质劣化处理包括丢帧处理,处理后音频集合包括丢帧音频,丢帧音频是通过针对处理前音频集合中的处理前音频,执行如下处理得到的:对处理前音频进行丢帧处理,得到丢帧音频作为处理后音频。

[0059] 在本实施例的一些可选的实现方式中,音质劣化处理包括置零处理,处理后音频集合包括置零音频,置零处理为将处理前音频的属性的属性值设置为零的处理,置零音频是通过针对处理前音频集合中的处理前音频执行如下处理得到的:对处理前音频进行置零处理,得到置零音频作为处理后音频。

[0060] 在本实施例的一些可选的实现方式中,处理后音频集合包括丢帧音频和置零音频。处理后音频集合包括的丢帧音频的数量与处理后音频集合中的处理后音频的数量之比为预先确定的第一数值。处理后音频集合包括的置零音频的数量与处理后音频集合中的处理后音频的数量之比为预先确定的第二数值。其中,第一数值是音频集合中丢帧音频的数量与音频集合中音频的数量的比值,第二数值是音频集合中置零音频的数量与音频集合中音频的数量的比值。

[0061] 在这里,上述第一数值可以是技术人员通过大量的统计计算而确定的、由于硬件设备出现故障,或者,网络信号差等非人为原因导致的音频中出现丢帧音频的概率(即音频集合中丢帧音频的数量与音频集合中音频的数量的比值),上述第二数值可以是技术人员通过大量的统计计算而确定的、由于硬件设备出现故障,或者,网络信号差等非人为原因导致的音频中出现置零音频的概率(即音频集合中置零音频的数量与音频集合中音频的数量的比值)。作为示例,人为原因可以是人员通过电子设备操作所导致的音频帧中出现置零音频或丢帧音频。上述非人为原因可以包括除上述人为原因之外的任何原因所导致的音频帧中出现置零音频或丢帧音频。

[0062] 在本实施例的一些可选的实现方式中,处理后音频集合包括丢帧音频和置零音频,处理后音频集合是通过针对处理前音频集合中的处理前音频,执行如下处理步骤得到的:

[0063] 第一步骤,随机生成第一随机数和第二随机数。其中,第一随机数和第二随机数均为0到1之间的数。

[0064] 在这里,第一随机数和第二随机数中的第一、第二仅用作区分随机数,并不构成对随机数的特殊限定。上述第一随机数和第二随机数可以相等,也可以不等。

[0065] 第二步骤,响应于确定第一随机数大于等于预先确定的第一数值,并且,第二随机数小于预先确定的第二数值,将该处理前音频的属性的属性值设置为零。其中,第一数值用于表征音频中出现丢帧音频的概率。第二数值用于表征音频中出现置零音频的概率。

[0066] 在这里,上述第一数值和第二数值中的第一、第二,仅用作区分数值,并不构成对数值的特殊限定。上述第一数值和第二数值可以相等,也可以不等。

[0067] 第三步骤,响应于确定第一随机数小于第一数值,并且,第二随机数大于等于第二数值,对该处理前音频进行丢帧处理。

[0068] 可选的,由于硬件设备出现故障,或者,网络信号差等原因,也可以导致丢帧音频

或者置零音频的出现，因而，上述处理后音频集合也可以包括上述执行主体直接获取的、上述非人为因素导致并生成的丢帧音频或者置零音频。

[0069] 步骤202，利用机器学习算法，将训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型。

[0070] 在本实施例中，上述执行主体可以利用机器学习算法，将步骤201所获取到的训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型。

[0071] 具体地，上述执行主体可以利用机器学习算法，将步骤201获取到的训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，对初始模型(例如循环神经网络，卷积神经网络)进行训练，针对每次训练输入的特征数据，可以得到实际输出。其中，实际输出是初始模型实际输出的，用于表征标识信息。然后，上述执行主体可以采用梯度下降法，基于实际输出和期望输出，调整初始模型的参数，将每次调整参数后得到的模型作为下次训练的初始模型，并在满足预设的训练结束条件的情况下，结束训练，从而训练得到语音识别模型。

[0072] 需要说明的是，这里预设的训练结束条件可以包括但不限于以下至少一项：训练时间超过预设时长；训练次数超过预设次数；计算所得的差异(例如损失函数的函数值)小于预设差异阈值。

[0073] 在本实施例的一些可选的实现方式中，上述初始模型也可以是具有门控循环单元的循环神经网络模型，由此，上述语音识别模型可以为具有门控循环单元的循环神经网络模型。

[0074] 在这里，采用具有门控循环单元的循环神经网络模型作为初始语音识别模型，训练得到的语音识别模型，相对其他模型作为初始语音识别模型，训练得到的语音识别模型而言，可以具有更快的计算效率。

[0075] 继续参见图4，图4是根据本实施例的用于生成模型的方法的应用场景的一个示意图。在图4的应用场景中，服务器401首先获取针对处理后音频集合的训练样本集合4001。其中，处理后音频集合包括对处理前音频执行音质劣化处理得到的音频。训练样本与处理后音频一一对应，训练样本包括处理后音频的特征数据和标识信息。标识信息用于指示处理后音频中是否包括语音音频。然后，服务器401利用机器学习算法，将训练样本集合中的训练样本包括的特征数据作为初始模型4002(例如具有门控循环单元的循环神经网络模型)的输入，将与输入的特征数据对应的标识信息作为初始模型4002的期望输出，训练得到语音识别模型4003。

[0076] 本申请的上述实施例提供的方法，通过获取针对处理后音频集合的训练样本集合，其中，处理后音频集合包括对处理前音频执行音质劣化处理得到的音频，训练样本与处理后音频一一对应，训练样本包括处理后音频的特征数据和标识信息，标识信息用于指示处理后音频中是否包括语音音频，然后，利用机器学习算法，将训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型，从而采用包括音质劣化处理得到的音频的特征数据及对应的标识信息的训练样本，来训练语音识别模型，丰富了模型的训练方式，此外，采用训练得到的语音识别模型，可以提高语音端点检测的准确度。

[0077] 进一步参考图5，其示出了用于生成模型的方法的又一个实施例的流程500。该用于生成模型的方法的流程500，包括以下步骤：

[0078] 步骤501，从处理前音频集合中选取未被选取过的处理前音频。之后，执行步骤502。

[0079] 在本实施例中，用于生成模型的方法的执行主体(例如图1所示的服务器或终端设备)可以从处理前音频集合中选取未被选取过的处理前音频。

[0080] 上述处理前音频可以是各种音频，例如，上述处理前音频可以包括但不限于以下任意一项：带噪音的语音音频、包括背景音和前景音的音频、包括静音和非静音的音频等等。该音频可以是任意长度的音频，例如，一句话；也可以是音频帧，其中，音频帧的长度可以是预先设置的，例如帧长可以是32毫秒、30毫秒等等。

[0081] 步骤502，随机生成第一随机数和第二随机数。之后，执行步骤503。

[0082] 在本实施例中，上述执行主体可以随机生成第一随机数和第二随机数。其中，第一随机数和第二随机数均为0到1之间的数。

[0083] 在这里，第一随机数和第二随机数中的第一、第二仅用作区分随机数，并不构成对随机数的特殊限定。上述第一随机数和第二随机数可以相等，也可以不等。

[0084] 步骤503，确定第一随机数是否大于等于预先确定的第一数值。之后，若是，则执行步骤505；若否，则执行步骤504。

[0085] 在本实施例中，上述执行主体可以确定第一随机数是否大于等于预先确定的第一数值。其中，第一数值是音频集合中丢帧音频的数量与音频集合中音频的数量的比值。

[0086] 步骤504，确定第二随机数小于预先确定的第二数值。之后，若是，则执行步骤508；若否，则执行步骤506。

[0087] 在本实施例中，上述执行主体可以确定第二随机数小于预先确定的第二数值。其中，第二数值是音频集合中置零音频的数量与音频集合中音频的数量的比值。

[0088] 步骤505，确定第二随机数小于预先确定的第二数值。之后，若是，则执行步骤507；若否，则执行步骤508。

[0089] 在本实施例中，上述执行主体可以确定第二随机数小于预先确定的第二数值。其中，第二数值是音频集合中置零音频的数量与音频集合中音频的数量的比值。

[0090] 在这里，上述第一数值和第二数值中的第一、第二，仅用作区分数值，并不构成对数值的特殊限定。上述第一数值和第二数值可以相等，也可以不等。

[0091] 步骤506，将该处理前音频的属性的属性值设置为零。之后，执行步骤508。

[0092] 在本实施例中，上述执行主体可以将该处理前音频的属性的属性值设置为零。例如，上述属性可以为幅值。

[0093] 步骤507，对该处理前音频进行丢帧处理。之后，执行步骤508。

[0094] 在本实施例中，上述执行主体可以对该处理前音频进行丢帧处理。

[0095] 步骤508，得到处理后音频。

[0096] 在本实施例中，上述执行主体可以得到处理后音频。

[0097] 可以理解，该步骤508得到的处理后音频可以是以下任一项：对处理前音频进行丢帧处理后得到的音频，对处理前音频进行置零处理(即将处理前音频的属性的属性值设置为零)后得到的音频，处理前音频。在这里，当第一随机数大于等于预先确定的第一数值，并

且，第二随机数小于预先确定的第二数值时，上述执行主体可以将对处理前音频进行丢帧处理后得到的音频确定为该步骤所得到的处理后音频；当第一随机数小于预先确定的第一数值，并且，第二随机数大于等于预先确定的第二数值时，上述执行主体可以将对处理前音频的属性的属性值设置为零后得到的音频确定为该步骤所得到的处理后音频；当第一随机数小于预先确定的第一数值，并且，第二随机数小于等于预先确定的第二数值时，上述执行主体可以将处理前音频确定为该步骤所得到的处理后音频；当第一随机数大于等于预先确定的第一数值，并且，第二随机数大于等于预先确定的第二数值时，上述执行主体可以将处理前音频确定为该步骤所得到的处理后音频。

[0098] 步骤509，确定处理前音频集合中，是否存在未被选取过的处理前音频。之后，若是，则执行步骤501；若否，则执行步骤510。

[0099] 在本实施例中，上述执行主体可以确定处理前音频集合中，是否存在未被选取过的处理前音频。

[0100] 步骤510，获取针对处理后音频集合的训练样本集合。之后，执行步骤511。

[0101] 在本实施例中，步骤510与图2对应实施例中的步骤201基本一致，这里不再赘述。

[0102] 步骤511，利用机器学习算法，将训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型。

[0103] 在本实施例中，步骤511与图2对应实施例中的步骤202基本一致，这里不再赘述。

[0104] 从图5中可以看出，与图2对应的实施例相比，本实施例中的用于生成模型的方法的流程500突出了得到处理后音频的步骤。由此，本实施例描述的方案用于训练语音识别模型的训练样本中包括的特征数据指示的丢帧音频、置零音频的占总的训练样本集合的数量的比值分别为音频中出现丢帧音频的概率，以及音频中出现置零音频的概率，因而，训练得到的语音识别模型可以更准确地确定音频中是否包含语音音频，以及语音音频在音频中的位置。

[0105] 进一步参考图6，作为对上述各图所示方法的实现，本申请提供了一种用于生成模型的装置的一个实施例，该装置实施例与图2所示的方法实施例相对应，除下面所记载的特征外，该装置实施例还可以包括与图2所示的方法实施例相同或相应的特征。该装置具体可以应用于各种电子设备中。

[0106] 如图6所示，本实施例的用于生成模型的装置600包括：第一获取单元601和训练单元602。其中，第一获取单元601被配置成获取针对处理后音频集合的训练样本集合，其中，处理后音频集合包括对处理前音频执行音质劣化处理得到的音频，训练样本与处理后音频一一对应，训练样本包括处理后音频的特征数据和标识信息，标识信息用于指示处理后音频中是否包括语音音频；训练单元602被配置成利用机器学习算法，将训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型。

[0107] 在本实施例中，用于生成模型的装置600的第一获取单元601可以通过有线连接方式或者无线连接方式从其他电子设备或者本地获取针对处理后音频集合的训练样本集合。其中，处理后音频集合包括对处理前音频执行音质劣化处理得到的音频。训练样本与处理后音频一一对应，训练样本包括处理后音频的特征数据和标识信息。标识信息用于指示处理后音频中是否包括语音音频。

[0108] 在这里，上述音质劣化处理可以是对处理前音频的保真度进行降低的处理。经过音质劣化处理后得到的处理后音频，相对于未经音质劣化处理的处理前音频的保真度存在降低。上述保真度可以包括但不限于以下至少一项：清晰程度、不失真程度、再现平面声象的程度等等。可以理解，上述音质劣化处理可以是在处理前音频的音频信号中加入信号，或者，删除信号，或者，对信号进行调整的处理。作为示例，上述音质劣化处理可以包括但不限于以下任一项：置零处理，丢帧处理，加入杂音处理等等。在这里，对处理前音频进行置零处理所得到的处理后音频可以是置零音频，对处理前音频进行丢帧处理所得到的处理后音频可以是丢帧音频。

[0109] 上述置零音频可以是将处理前音频的以下任意一项属性的属性值设置为零之后得到的：幅值、频率、振幅、音调等等。

[0110] 上述特征数据可以包括但不限于音频的以下至少一项特征的数据：幅值、帧率、过零率、短时能量等其他音频特征。

[0111] 在本实施例中，上述训练单元602可以利用机器学习算法，将训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型。

[0112] 在本实施例的一些可选的实现方式中，音质劣化处理包括丢帧处理，处理后音频集合包括丢帧音频，丢帧音频是通过针对处理前音频集合中的处理前音频，执行如下处理得到的：对处理前音频进行丢帧处理，得到丢帧音频作为处理后音频。

[0113] 在本实施例的一些可选的实现方式中，音质劣化处理包括置零处理，处理后音频集合包括置零音频，置零处理为将处理前音频的属性的属性值设置为零的处理，置零音频是通过针对处理前音频集合中的处理前音频执行如下处理得到的：对处理前音频进行置零处理，得到置零音频作为处理后音频。

[0114] 在本实施例的一些可选的实现方式中，处理后音频集合包括丢帧音频和置零音频，处理后音频集合包括的丢帧音频的数量与处理后音频集合中的处理后音频的数量之比为预先确定的第一数值，处理后音频集合包括的置零音频的数量与处理后音频集合中的处理后音频的数量之比为预先确定的第二数值，其中，第一数值是音频集合中丢帧音频的数量与音频集合中音频的数量的比值，第二数值是音频集合中置零音频的数量与音频集合中音频的数量的比值。

[0115] 在本实施例的一些可选的实现方式中，处理后音频集合包括丢帧音频和置零音频，处理后音频集合是通过针对处理前音频集合中的处理前音频，执行如下处理步骤得到的：

[0116] 第一步骤，随机生成第一随机数和第二随机数。其中，第一随机数和第二随机数均为0到1之间的数。

[0117] 第二步骤，响应于确定第一随机数大于等于预先确定的第一数值，并且，第二随机数小于预先确定的第二数值，将该处理前音频的属性的属性值设置为零，其中，第一数值用于表征音频中出现丢帧音频的概率，第二数值用于表征音频中出现置零音频的概率

[0118] 第三步骤，响应于确定第一随机数小于第一数值，并且，第二随机数大于等于第二数值，对该处理前音频进行丢帧处理。

[0119] 在本实施例的一些可选的实现方式中，上述属性可以为幅值。

[0120] 在本实施例的一些可选的实现方式中，语音识别模型为具有门控循环单元的循环神经网络模型。

[0121] 本申请的上述实施例提供的装置，通过第一获取单元601获取针对处理后音频集合的训练样本集合，其中，处理后音频集合包括对处理前音频执行音质劣化处理得到的音频，训练样本与处理后音频一一对应，训练样本包括处理后音频的特征数据和标识信息，标识信息用于指示处理后音频中是否包括语音音频，然后，训练单元602利用机器学习算法，将训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型，从而采用包括音质劣化处理得到的音频的特征数据及对应的标识信息的训练样本，来训练语音识别模型，丰富了模型的训练方式，此外，采用训练得到的语音识别模型，可以提高语音端点检测的准确度。

[0122] 继续参考图7，示出了根据本申请的用于检测语音的方法的一个实施例的流程700。该用于检测语音的方法，包括以下步骤：

[0123] 步骤701，获取目标音频。

[0124] 在本实施例中，用于检测语音的方法的执行主体(例如图1所示的服务器或终端设备)可以通过有线连接方式或者无线连接方式从其他电子设备或者本地，获取目标音频。其中，上述目标音频可以是包括语音音频的各种音频。

[0125] 步骤702，针对目标音频包括的至少一个音频帧中的音频帧，将该音频帧的特征数据输入至预先训练的语音识别模型，得到标识信息。

[0126] 在本实施例中，针对目标音频包括的至少一个音频帧中的音频帧，上述执行主体可以将该音频帧的特征数据输入至预先训练的语音识别模型，得到标识信息。其中，上述语音识别模型可以是上述执行主体或者与上述执行主体通信连接的电子设备按照如图2所示的用于生成模型的方法中的任一实施例所描述的方法训练得到的。

[0127] 在这里，上述音频帧可以具有预先确定的帧长。例如，该音频帧可以为帧长32毫米的音频帧，也可以是帧长30毫米的音频帧，等等。

[0128] 上述标识信息可以用于指示音频帧中是否包括语音音频，也可以用于指示音频帧中包含语音音频的概率。

[0129] 可以理解，通常，按照上述训练方式得到的语音识别模型，在实际使用的过程中，可以输出音频帧中包含语音音频的概率，进而，上述执行主体可以通过比较所得到的概率与预设概率阈值之间的大小关系，从而确定音频帧中是否包括语音音频。

[0130] 步骤703，基于所得到的标识信息集合，生成目标音频的语音端点检测结果。

[0131] 在本实施例中，上述执行主体可以基于所得到的标识信息集合，生成目标音频的语音端点检测结果。

[0132] 上述语音端点检测结果可以用于指示上述目标音频中包含的语音音频的起始位置和终止位置。

[0133] 作为示例，上述执行主体可以首先确定标识信息集合中的标识信息指示的、目标音频包括的音频帧序列中的首个和最后一个包括语音音频的音频帧，以及将所确定的首个包括语音音频的音频帧，确定为目标音频中包含的语音音频的起始位置，将所确定的最后一个包括语音音频的音频帧，确定为目标音频中包含的语音音频的终止位置，从而得到了语音端点检测结果。

[0134] 可选的，上述执行主体还可以直接将标识信息集合确定为语音端点检测结果。例如，如果上述目标音频由10帧音频帧组成。其中，第2帧至第9帧音频帧包括语音音频，第1帧和第10帧音频帧不包括语音音频。那么，上述执行主体可以生成标识信息序列{0,1,1,1,1,1,1,1,1,0}，其中，上述标识信息序列中的第一个标识信息是目标音频包括的第一个音频帧对应的标识信息，上述标识信息序列中的第2个标识信息是目标音频包括的第2个音频帧对应的标识信息，以此类推。“0”可以表征不包括语音音频，“1”可以表征包括语音音频。由此，上述执行主体可以将标识信息序列{0,1,1,1,1,1,1,1,1,0}直接确定为语音端点检测结果。再此应用场景下，通过该语音端点检测结果，可以确定目标音频由10帧音频帧组成。其中，第2帧至第9帧音频帧包括语音音频，第1帧和第10帧音频帧不包括语音音频。

[0135] 本申请的上述实施例提供的方法，通过获取目标音频，其中，目标音频包括语音音频，然后，针对目标音频包括的至少一个音频帧中的音频帧，将该音频帧的特征数据输入至预先训练的语音识别模型，得到标识信息，其中，语音识别模型是如上述用于生成模型的方法中任一实施例的方法训练得到的，最后，基于所得到的标识信息集合，生成目标音频的语音端点检测结果，由此，将语音识别模型应用于语音端点检测，从而提高了语音端点检测的准确程度，丰富了语音端点检测的方式。

[0136] 进一步参考图8，作为对上述各图所示方法的实现，本申请提供了一种用于检测语音的装置的一个实施例，该装置实施例与图7所示的方法实施例相对应，除下面所记载的特征外，该装置实施例还可以包括与图7所示的方法实施例相同或相应的特征。该装置具体可以应用于各种电子设备中。

[0137] 如图8所示，本实施例的用于检测语音的装置800包括：第二获取单元801、输入单元802和生成单元803。其中，第二获取单元801被配置成获取目标音频，其中，目标音频包括语音音频；输入单元802被配置成针对目标音频包括的至少一个音频帧中的音频帧，将该音频帧的特征数据输入至预先训练的语音识别模型，得到标识信息，其中，语音识别模型是如上述用于生成模型的方法中任一实施例的方法训练得到的；生成单元803被配置成基于所得到的标识信息集合，生成目标音频的语音端点检测结果。

[0138] 在本实施例中，用于检测语音的装置800的第二获取单元801可以通过有线连接方式或者无线连接方式从其他电子设备或者本地，获取目标音频。

[0139] 上述目标音频可以是包括语音音频的各种音频。

[0140] 在本实施例中，针对第二获取单元801获取到的目标音频包括的至少一个音频帧中的音频帧，上述输入单元802可以将该音频帧的特征数据输入至预先训练的语音识别模型，得到标识信息。其中，上述语音识别模型可以是上述执行主体或者与上述执行主体通信连接的电子设备按照如图2所示的用于生成模型的方法中的任一实施例所描述的方法训练得到的。

[0141] 在这里，上述音频帧可以具有预先确定的帧长。例如，该音频帧可以为帧长32毫米的音频帧，也可以是帧长30毫米的音频帧，等等。

[0142] 上述标识信息可以用于指示音频帧中是否包括语音音频，也可以用于指示音频帧中包含语音音频的概率。

[0143] 在本实施例中，基于输入单元802所得到的标识信息集合，上述生成单元803可以生成目标音频的语音端点检测结果。其中，上述语音端点检测结果可以用于指示上述目标

音频中包含的语音音频的起始位置和终止位置。

[0144] 本申请的上述实施例提供的装置，通过第二获取单元801获取目标音频，其中，目标音频包括语音音频，然后，输入单元802针对目标音频包括的至少一个音频帧中的音频帧，将该音频帧的特征数据输入至预先训练的语音识别模型，得到标识信息，其中，语音识别模型是如上述用于生成模型的方法中任一实施例的方法训练得到的，最后，生成单元803基于所得到的标识信息集合，生成目标音频的语音端点检测结果，由此，将语音识别模型应用于语音端点检测，从而提高了语音端点检测的准确程度，丰富了语音端点检测的方式。

[0145] 下面参考图9，其示出了适于用来实现本申请实施例的电子设备的计算机系统900的结构示意图。图9示出的电子设备仅仅是一个示例，不应对本申请实施例的功能和使用范围带来任何限制。

[0146] 如图9所示，计算机系统900包括中央处理单元(CPU)901，其可以根据存储在只读存储器(ROM)902中的程序或者从存储部分908加载到随机访问存储器(RAM)903中的程序而执行各种适当的动作和处理。在RAM 903中，还存储有系统900操作所需的各种程序和数据。CPU901、ROM 902以及RAM 903通过总线904彼此相连。输入/输出(I/O)接口905也连接至总线904。

[0147] 以下部件连接至I/O接口905：包括键盘、鼠标等的输入部分906；包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分907；包括硬盘等的存储部分908；以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分909。通信部分909经由诸如因特网的网络执行通信处理。驱动器910也根据需要连接至I/O接口905。可拆卸介质911，诸如磁盘、光盘、磁光盘、半导体存储器等等，根据需要安装在驱动器910上，以便于从其上读出的计算机程序根据需要被安装入存储部分908。

[0148] 特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括承载在计算机可读介质上的计算机程序，该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信部分909从网络上被下载和安装，和/或从可拆卸介质911被安装。在该计算机程序被中央处理单元(CPU)901执行时，执行本申请的方法中限定的上述功能。

[0149] 需要说明的是，本申请所述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本申请中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本申请中，计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读介质可以发送、传播或者传输用于

由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：无线、电线、光缆、RF等等，或者上述的任意合适的组合。

[0150] 可以以一种或多种程序设计语言或其组合来编写用于执行本申请的操作的计算机程序代码，所述程序设计语言包括面向目标的程序设计语言—诸如Python、Java、Smalltalk、C++，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码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)。

[0151] 附图中的流程图和框图，图示了按照本申请各种实施例的系统、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、或代码的一部分，该模块、程序段、或代码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。也应当注意，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，两个接连地表示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。也要注意的是，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操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。

[0152] 描述于本申请实施例中所涉及到的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中，例如，可以描述为：一种处理器包括第一获取单元和训练单元。其中，这些单元的名称在某种情况下并不构成对该单元本身的限定，例如，第一获取单元还可以被描述为“获取针对处理后音频集合的训练样本集合的单元”。

[0153] 作为另一方面，本申请还提供了一种计算机可读介质，该计算机可读介质可以是上述实施例中描述的电子设备中所包含的；也可以是单独存在，而未装配入该电子设备中。上述计算机可读介质承载有一个或者多个程序，当上述一个或者多个程序被该电子设备执行时，使得该电子设备：获取针对处理后音频集合的训练样本集合，其中，处理后音频集合包括对处理前音频执行音质劣化处理得到的音频，训练样本与处理后音频一一对应，训练样本包括处理后音频的特征数据和标识信息，标识信息用于指示处理后音频中是否包括语音音频；利用机器学习算法，将训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型。

[0154] 以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本申请中所涉及的发明范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

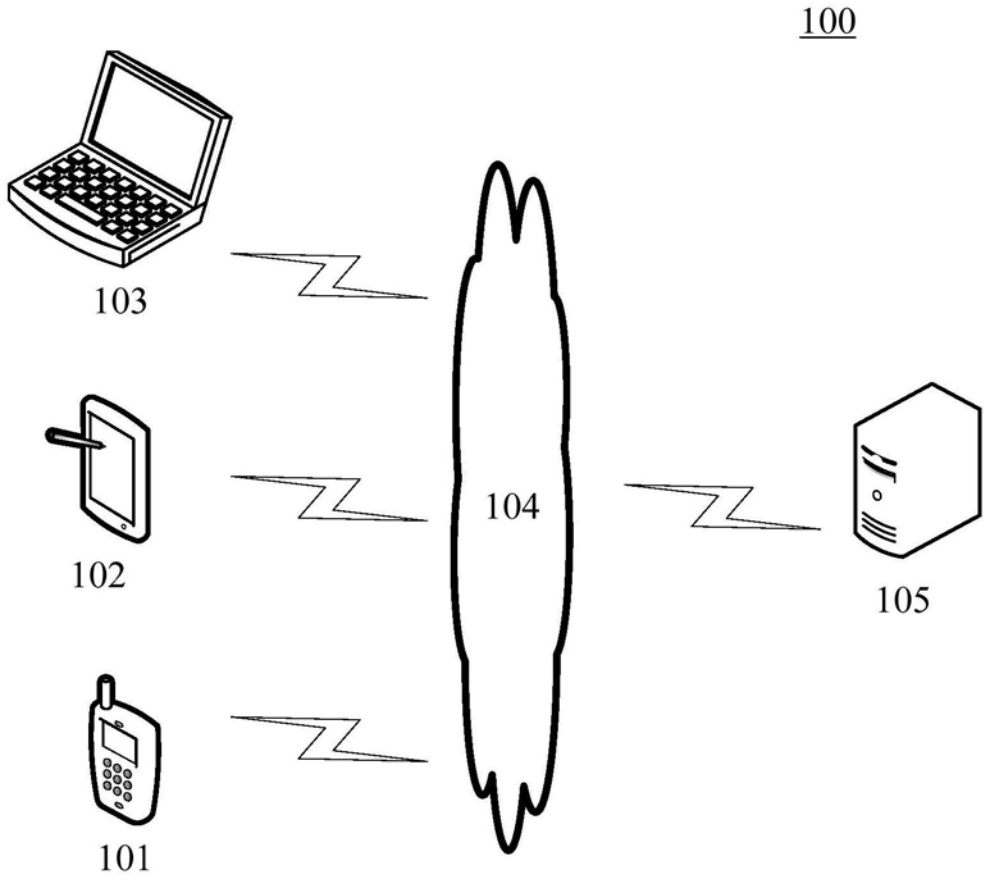

100
103
102
101
104
105

图1

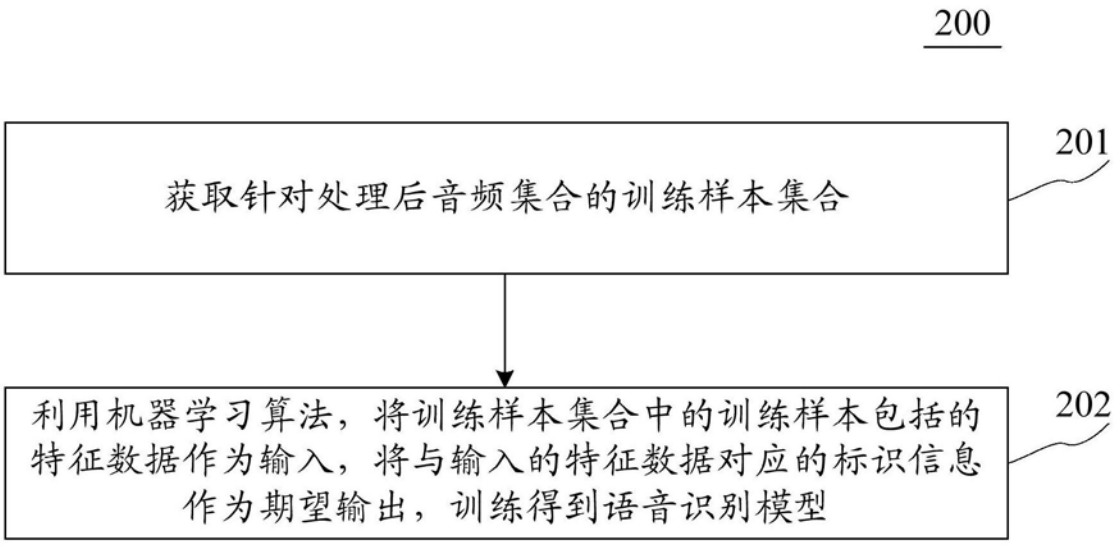

200
获取针对处理后音频集合的训练样本集合
201
利用机器学习算法，将训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型
202

图2

图3A

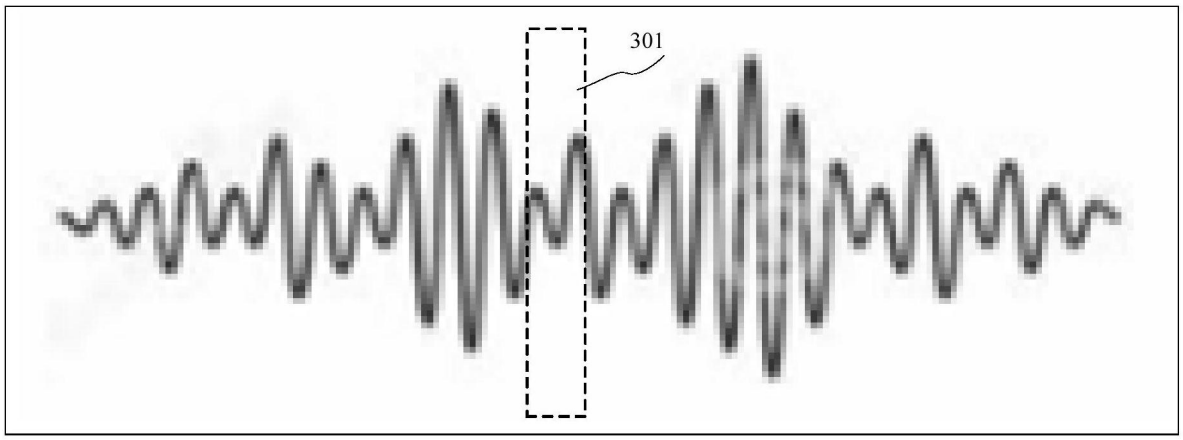


图3B

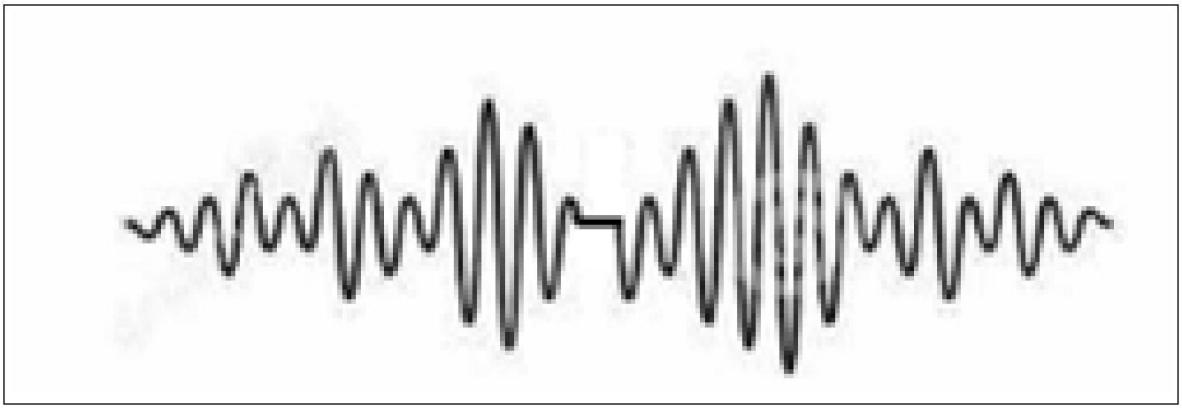


图3C

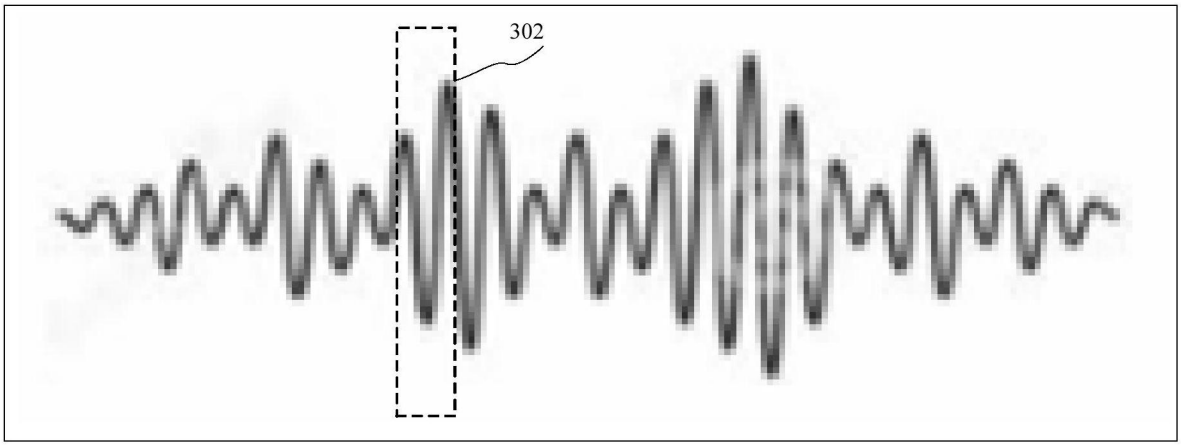


图3D

图3E

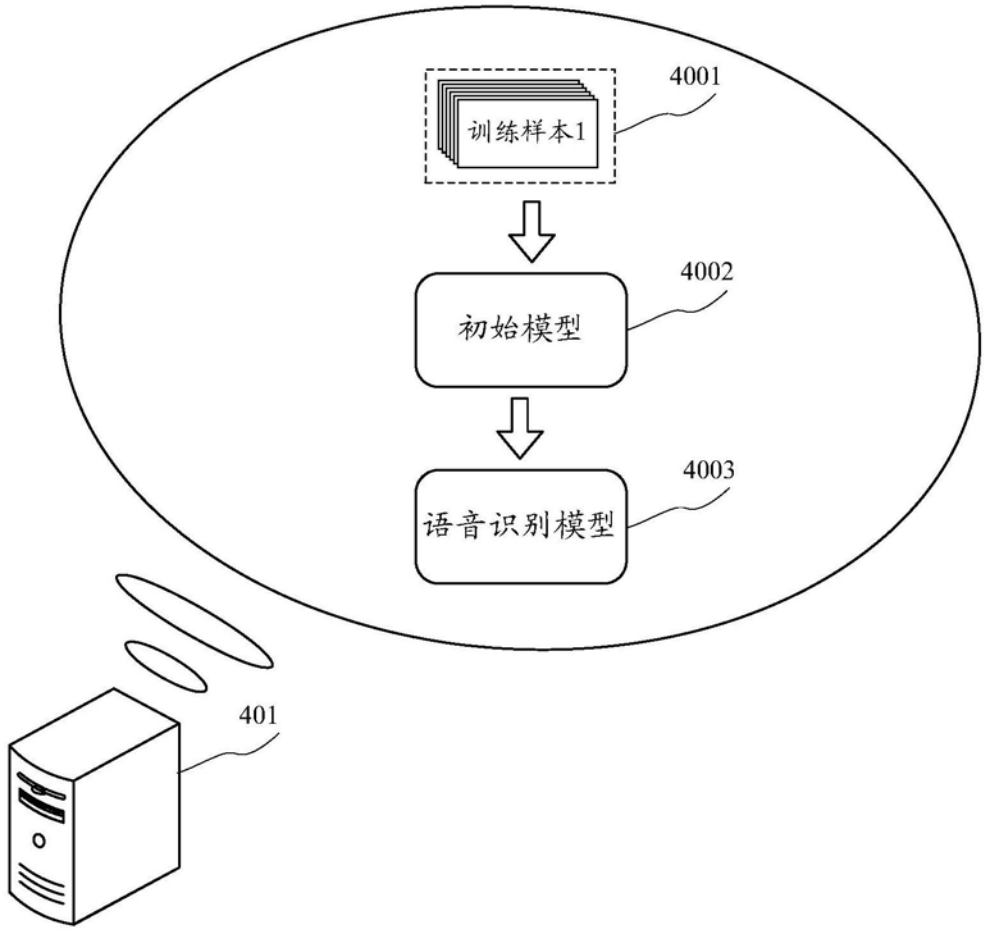


图4

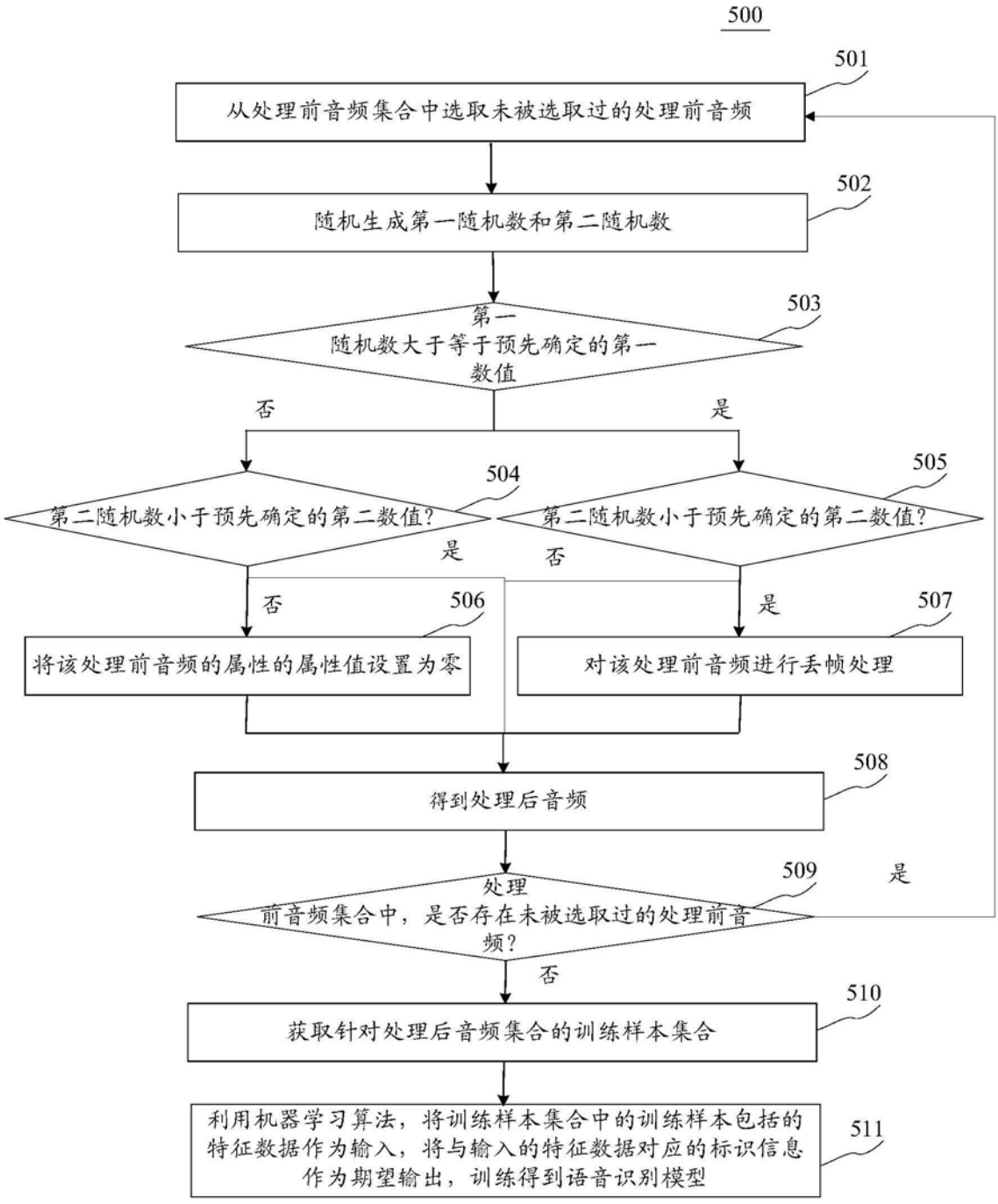

500
501
从处理前音频集合中选取未被选取过的处理前音频
502
随机生成第一随机数和第二随机数
503
第一随机数大于等于预先确定的第一数值
否
是
504
第二随机数小于预先确定的第二数值?
505
第二随机数小于预先确定的第二数值?
是
否
否
506
将该处理前音频的属性的属性值设置为零
是
507
对该处理前音频进行丢帧处理
508
得到处理后音频
509
处理前音频集合中，是否存在未被选取过的处理前音频?
是
否
510
获取针对处理后音频集合的训练样本集合
511
利用机器学习算法，将训练样本集合中的训练样本包括的特征数据作为输入，将与输入的特征数据对应的标识信息作为期望输出，训练得到语音识别模型

图5

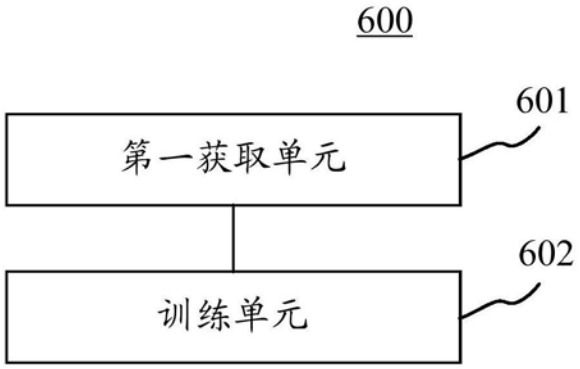

600
601
第一获取单元
602
训练单元

图6

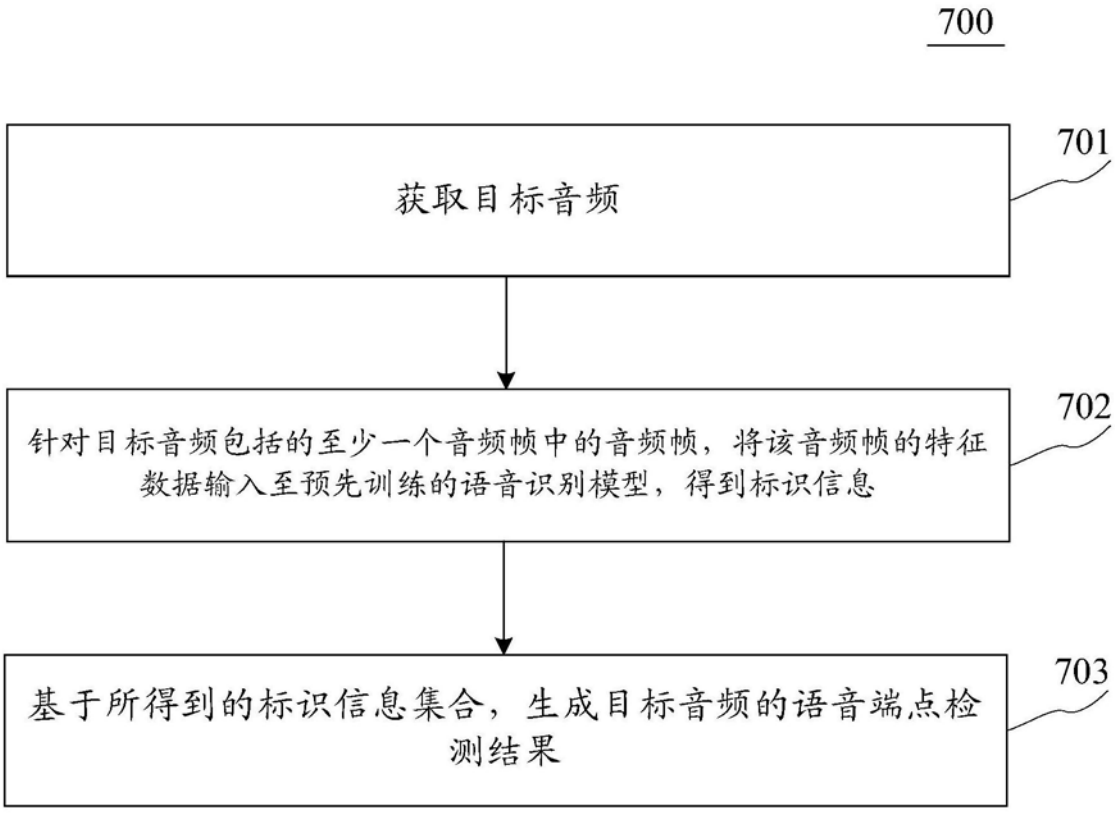

700
701
获取目标音频
702
针对目标音频包括的至少一个音频帧中的音频帧，将该音频帧的特征数据输入至预先训练的语音识别模型，得到标识信息
703
基于所得到的标识信息集合，生成目标音频的语音端点检测结果

图7

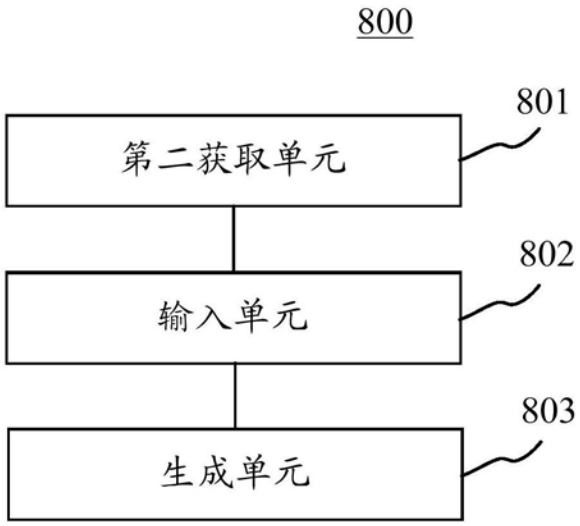

800
801
第二获取单元
802
输入单元
803
生成单元

图8

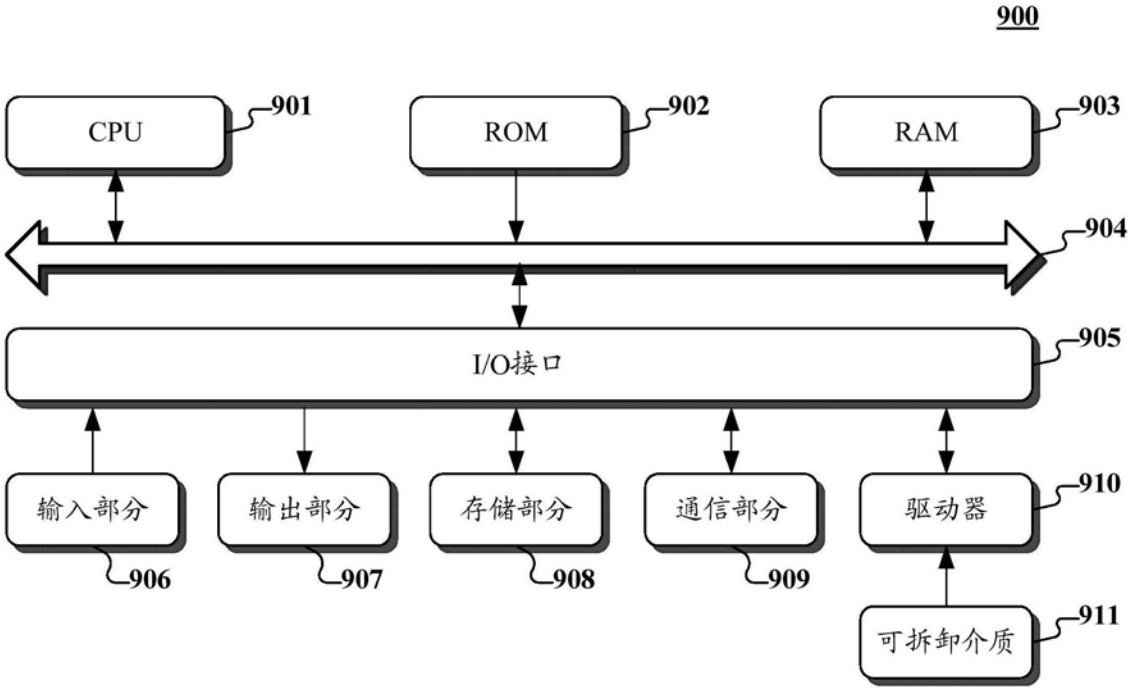

900
CPU
901
ROM
902
RAM
903
904
I/O接口
905
输入部分
906
输出部分
907
存储部分
908
通信部分
909
驱动器
910
可拆卸介质
911

图9