



(12)发明专利申请

(10)申请公布号 CN 106887241 A

(43)申请公布日 2017.06.23

(21)申请号 201610890946.9

(22)申请日 2016.10.12

(71)申请人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼资本大厦一座四
层847号邮箱

(72)发明人 焦雷 官砚楚 曾晓东 林锋

(74)专利代理机构 北京国昊天诚知识产权代理
有限公司 11315

代理人 黄熊

(51)Int.Cl.

G10L 25/84(2013.01)

G10L 25/21(2013.01)

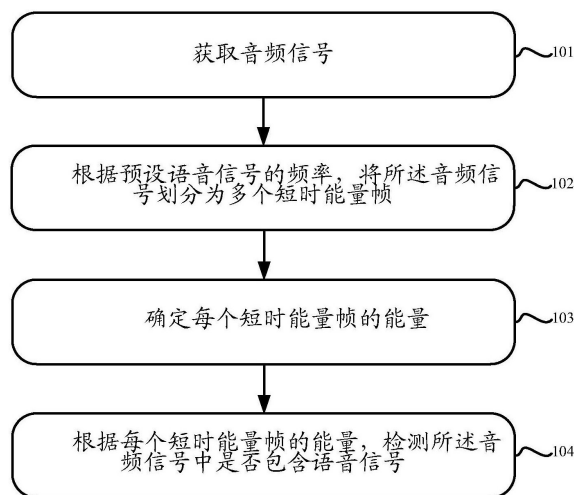
权利要求书2页 说明书9页 附图3页

(54)发明名称

一种语音信号检测方法与装置

(57)摘要

本申请公开了一种语音信号检测方法与装置,用于解决现有技术中的语音信号检测方法存在的处理速度较慢,且耗费资源较多的问题。该方法包括:获取音频信号;根据预设语音信号的频率,将所述音频信号划分为多个短时能量帧;确定每个短时能量帧的能量;根据每个短时能量帧的能量,检测所述音频信号中是否包含语音信号。



1. 一种语音信号检测方法,其特征在于,所述方法包括:
 - 获取音频信号;
 - 根据预设语音信号的频率,将所述音频信号划分为多个短时能量帧;
 - 确定每个短时能量帧的能量;
 - 根据每个短时能量帧的能量,检测所述音频信号中是否包含语音信号。
2. 如权利要求1所述的方法,其特征在于,获取音频信号,具体包括:
 - 获取当前音频信号;
 - 在上一次获取到的音频信号中,截取预设时段的子信号;
 - 将所述当前音频信号和截取的子信号进行拼接,作为获取到的音频信号。
3. 如权利要求1所述的方法,其特征在于,根据预设语音信号的频率,将所述音频信号划分为多个短时能量帧,具体包括:
 - 根据预设语音信号的频率,确定出所述预设语音信号的周期;
 - 按照确定出的周期,将所述音频信号划分为对应的时长均为所述周期的多个短时能量帧。
4. 如权利要求1所述的方法,其特征在于,根据每个短时能量帧的能量,检测所述音频信号中是否包含语音信号,具体包括:
 - 确定能量大于预设阈值的短时能量帧的数量占有所有短时能量帧总数的比率;
 - 判断所述比率是否大于预设比率;
 - 若是,则确定检测到所述音频信号中包含语音信号;
 - 若否,则确定未检测到所述音频信号中包含语音信号。
5. 如权利要求1所述的方法,其特征在于,根据每个短时能量帧的能量,检测所述音频信号中是否包含语音信号,具体包括:
 - 确定能量大于预设阈值的短时能量帧的数量占有所有短时能量帧总数的比率;
 - 判断所述比率是否大于预设比率;
 - 若否,则确定未检测到所述音频信号中包含语音信号;
 - 若是,则当能量大于预设阈值的短时能量帧中存在至少N个连续短时能量帧时,确定检测到所述音频信号中包含语音信号,当能量大于预设阈值的短时能量帧中不存在至少N个连续短时能量帧时,确定未检测到所述音频信号中包含语音信号。
6. 一种语音信号检测装置,其特征在于,所述装置包括:
 - 获取模块,获取音频信号;
 - 划分模块,根据预设语音信号的频率,将所述音频信号划分为多个短时能量帧;
 - 确定模块,确定每个短时能量帧的能量;
 - 检测模块,根据每个短时能量帧的能量,检测所述音频信号中是否包含语音信号。
7. 如权利要求1所述的装置,其特征在于,获取模块:
 - 获取当前音频信号;
 - 在上一次获取到的音频信号中,截取预设时段的子信号;
 - 将所述当前音频信号和截取的子信号进行拼接,作为获取到的音频信号。
8. 如权利要求1所述的装置,其特征在于,划分模块,根据预设语音信号的频率,确定出所述预设语音信号的周期;

按照确定出的周期,将所述音频信号划分为对应的时长均为所述周期的多个短时能量帧。

9.如权利要求1所述的装置,其特征在于,检测模块,确定能量大于预设阈值的短时能量帧的数量占有所有短时能量帧总数的比率;

判断所述比率是否大于预设比率;

若是,则确定检测到所述音频信号中包含语音信号;

若否,则确定未检测到所述音频信号中包含语音信号。

10.如权利要求1所述的装置,其特征在于,检测模块,确定能量大于预设阈值的短时能量帧的数量占有所有短时能量帧总数的比率;

判断所述比率是否大于预设比率;

若否,则确定未检测到所述音频信号中包含语音信号;

若是,则当能量大于预设阈值的短时能量帧中存在至少N个连续短时能量帧时,确定检测到所述音频信号中包含语音信号,当能量大于预设阈值的短时能量帧中不存在至少N个连续短时能量帧时,确定未检测到所述音频信号中包含语音信号。

一种语音信号检测方法与装置

技术领域

[0001] 本申请涉及计算机技术领域,尤其涉及一种语音信号检测方法与装置。

背景技术

[0002] 在实际生活中,人们会经常使用智能设备(例如智能手机、平板电脑等)来发送语音消息。但是人们在使用智能设备发送语音消息时,往往需要点击智能设备屏幕中的开始或结束按钮,才能够完成语音消息的发送,而这些点击操作,会给用户造成诸多不便。

[0003] 若用户无需点击按钮便可完成语音消息的发送,那么智能设备需要一直进行录音或者按照预设周期进行录音,并判断获取到的音频信号中是否包含语音信号,若包含语音信号,便将该语音信号提取出来,然后进行后续处理并发送出去,这样便完成了语音消息的发送。

[0004] 现有技术中,一般采用双门限方法、基于自相关极大值的检测方法或基于小波变换的检测方法等语音信号检测方法,来检测获取到的音频信号中是否包含语音信号。但是这些方法基本都是通过傅里叶变换等复杂的计算,获取音频信息的频率特征,进而根据该频率特征来确定是否包含语音信号的,需要计算较大缓冲数据,内存占用较高,计算量偏大,处理速度较慢,且耗电量较大。

发明内容

[0005] 本申请实施例提供一种语音信号检测方法与装置,用于解决现有技术中的语音信号检测方法存在的处理速度较慢,且耗费资源较多的问题。

[0006] 本申请实施例采用下述技术方案:

[0007] 一种语音信号检测方法,所述方法包括:

[0008] 获取音频信号;

[0009] 根据预设语音信号的频率,将所述音频信号划分为多个短时能量帧;

[0010] 确定每个短时能量帧的能量;

[0011] 根据每个短时能量帧的能量,检测所述音频信号中是否包含语音信号。

[0012] 一种语音信号检测装置,所述装置包括:

[0013] 获取模块,获取音频信号;

[0014] 划分模块,根据预设语音信号的频率,将所述音频信号划分为多个短时能量帧;

[0015] 确定模块,确定每个短时能量帧的能量;

[0016] 检测模块,根据每个短时能量帧的能量,检测所述音频信号中是否包含语音信号。

[0017] 本申请实施例采用的上述至少一个技术方案能够达到以下有益效果:

[0018] 与现有技术中的通过傅里叶变换等复杂计算来确定音频信号中是否包含语音信号的检测方法相比,本申请实施例采用的语音信号检测方法,无需进行傅里叶变换等复杂计算,通过根据预设语音信号的频率,将获取到的音频信号划分为多个短时能量帧,进而确定出每个短时能量帧的能量,并根据每个短时能量帧的能量,便可检测出获取到的音频信

号中是否包含语音信号。因此,本申请实施例提供的语音信号检测方法,能够解决现有技术中的语音信号检测方法存在的处理速度较慢,且耗费资源较多的问题。

附图说明

[0019] 此处所说明的附图用来提供对本申请的进一步理解,构成本申请的一部分,本申请的示意性实施例及其说明用于解释本申请,并不构成对本申请的不当限定。在附图中:

[0020] 图1为本申请实施例提供的一种语音信号检测方法的具体流程图;

[0021] 图2为本申请实施例提供的另一种语音信号检测方法的具体流程图;

[0022] 图3为本申请实施例提供的预设时长的音频信号显示图;

[0023] 图4为本申请实施例提供的一种语音信号检测装置的具体结构示意图。

具体实施方式

[0024] 为使本申请的目的、技术方案和优点更加清楚,下面将结合本申请具体实施例及相应的附图对本申请技术方案进行清楚、完整地描述。显然,所描述的实施例仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,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,都属于本申请保护的范围。

[0025] 以下结合附图,详细说明本申请实施例提供的技术方案。

[0026] 为了解决现有技术中的语音信号检测方法存在的处理速度较慢,且耗费资源较多的问题,本申请实施例提供一种语音信号检测方法。

[0027] 该方法的执行主体,可以但不限于为手机、平板电脑或个人电脑(Personal Computer,PC)等用户终端,或者该些用户终端上运行的应用(application,APP),或者,还可以是服务器等设备。

[0028] 为便于描述,下文以该方法的执行主体为APP为例,对该方法的实施方式进行介绍。可以理解,该方法的执行主体为APP只是一种示例性的说明,并不应理解为对该方法的限定。

[0029] 该方法的具体流程示意图如图1所示,包括下述步骤:

[0030] 步骤101,获取音频信号。

[0031] 上述音频信号,可以为APP通过音频采集设备采集到的音频信号,也可以为APP接收到的音频信号,比如可以是由其他APP或者设备传输的音频信号,本申请实施例对此不进行任何限定。APP在获取到音频信号之后,可以将该音频信号保存在本地。

[0032] 本申请对上述音频信号对应的采样率、时长、格式或声道等也不作任何限制。

[0033] 上述APP可以为任意类型的APP,比如聊天APP或支付APP等,只要该APP可以获取到音频信号,并且可以利用本申请实施例提供的语音信号检测方法对获取到的音频信号进行语音信号的检测即可。

[0034] 步骤102,根据预设语音信号的频率,将所述音频信号划分为多个短时能量帧。

[0035] 上述短时能量帧实际上是步骤101获取到的音频信号中的一部分音频信号。

[0036] 具体的,可以根据预设语音信号的频率,确定出该预设语音信号的周期,按照确定出的周期,将步骤101获取到的音频信号划分为对应的时长均为所述周期的多个短时能量帧。例如,假设该预设语音信号的周期为0.01S,则可根据步骤101获取到的音频信号的时

长,将该音频信号划分为若干个时长均为0.01S的短时能量帧。需要说明的是,在划分步骤101获取到的音频信号时,也可以根据实际情况,根据预设语音信号的频率,将该音频信号划分为至少两个短时能量帧。为了后续描述方便,本申请实施例后文中以将音频信号划分为多个短时能量帧为例进行说明。

[0037] 另外,当步骤101中由该APP自身通过音频采集设备采集音频信号时,由于采集音频信号一般是将实际上是模拟信号的音频信号以一定的采样率采集成数字信号,即脉冲编码调制(Pulse Code Modulation,PCM)格式的音频信号,因此,还可以根据该音频信号的采样率和预设语音信号的频率,将该音频信号划分为多个短时能量帧。

[0038] 具体的,可确定该音频信号的采样率与预设语音信号的频率的比值 m ,再根据该比值 m ,将采集到的数字形式的音频信号中每 m 个采样点划分为一个短时能量帧。若 m 为正整数,则可根据 m 将该音频信号划分为最大数量的短时能量帧;若 m 不为正整数,则可根据按照四舍五入原则转化为正整数的 m ,将该音频信号划分为最大数量的短时能量帧。其中,需要特别说明的是,若步骤101获取到的音频信号包含的采样点数量并非为 m 的整数倍,将该音频信号划分为最大数量的短时能量帧后,可将剩余的采样点丢弃,也可将剩余的采样点也作为一个短时能量帧进行后续处理。其中,上述 m ,用于表示在一个预设语音信号的周期内,步骤101获取到的音频信号包含的采样点数量。

[0039] 例如,若预设语音信号的频率为82HZ,步骤101获取到的音频信号的时长为1S,采样率为16000HZ,那么 $m=16000/82=195.1$ 。其中, m 不是正整数,将195.1按照四舍五入原则转化成正整数195。根据上述音频信号的时长以及采样率,可以确定出该音频信号包含的采样点数量为16000,那么,由于上述音频信号包含的采样点的数量并非是195的整数倍,因此,可以在将该音频信号划分为82个短时能量帧后,将剩余的10个采样点丢弃。其中,上述每个短时能量帧包含的采样点数量均为195。

[0040] 当步骤101获取到的音频信号是接收到的其他APP或设备传输的音频信号时,可以采用上述任一方法将该音频信号划分为多个短时能量帧。需要特别说明的是,上述音频信号的格式可能并非为PCM格式。若采用上述方法根据音频信号的采样率和预设语音信号的频率来划分短时能量帧,便需将接收到的音频信号转化为PCM格式的音频信号,另外,在接收到音频信号时,也需识别出该音频信号的采样率,具体识别出音频信号的采样率的方法均可采用现有技术的方法来识别,这里就不再一一赘述。

[0041] 步骤103,确定每个短时能量帧的能量。

[0042] 在本申请实施例中,当采用上述方法将PCM格式的音频信号划分为若干同样为PCM格式的短时能量帧时,则可以根据短时能量帧中的每个采样点对应的音频信号的幅值,来确定短时能量帧的能量。具体的,可以根据短时能量帧中的每个采样点对应的音频信号的幅值,确定出每个采样点的能量,然后将该些能量相加,将最终得到的能量之和,做为该短时能量帧的能量。

[0043] 例如,可以采用下述公式来确定短时能量帧的能量:
$$\text{能量} = \sum_i^{i+n} (A_i[t])^2$$
。其中, i 表示音频信号的第 i 个采样点; n 为短时能量帧中包含的采样点的数量; $A_i[t]$ 为第 i 个采样点对应的音频信号的幅值,其中,短时能量帧的幅值的取值范围为-32768~32767。

[0044] 另外,本申请实施例中,为了简化计算,节省资源,还可以将采集音频信号时获取

到的幅值除以32768的值,作为短时能量帧的归一化幅值,那么短时能量帧的归一化幅值的取值范围为-1~1。

[0045] 若短时能量帧的格式不为PCM格式,可以根据短时能量帧每一时刻的振幅,确定出计算振幅的函数,针对该函数的平方进行积分,最终得到的积分结果便为该短时能量帧的能量。

[0046] 步骤104,根据每个短时能量帧的能量,检测所述音频信号中是否包含语音信号。

[0047] 具体的,可以采用下述两种方法,来确定是否检测到音频信号中包含语音信号:

[0048] 方法1:确定能量大于预设阈值的短时能量帧的数量占所有短时能量帧总数的比率(后称高能量帧比率),并判断确定出的高能量帧比率是否大于预设比率。若是,则确定检测到所述音频信号中包含语音信号;若否,则确定未检测到音频信号中包含语音信号。

[0049] 其中,可以根据实际需要设置预设阈值以及预设比率的大小,在本申请实施例中,可以将预设阈值设置为2,预设比率设置为20%,若高能量帧比率大于20%,则确定检测到所述音频信号中包含语音信号;否则,则确定未检测到音频信号中包含语音信号。

[0050] 本申请实施例中,之所以可以采用方法1来确定是否检测到音频信号中包含语音信号,是因为在现实生活中,人们说话时,外部环境中多多少少会存在一些噪声,而噪声一般相对于人们说的话来说能量较低。那么若一段音频信号中,存在能量高于预设阈值的短时能量帧,且该些短时能量帧在这一段音频信号中占据一定的比率,便可认为该音频信号中包含语音信号。

[0051] 方法2:为了使得最终检测结果更加准确,可采用方法1提及的方法来确定高能量帧比率,并判断确定出的高能量帧比率是否大于预设比率,若否,则确定未检测到音频信号中包含语音信号;若是,则当能量大于预设阈值的短时能量帧中存在至少N个连续短时能量帧时,确定检测到音频信号中包含语音信号,当能量大于预设阈值的短时能量帧中不存在至少N个连续短时能量帧时,确定未检测到音频信号中包含语音信号。其中,N可以为任意正整数。在本申请实施例中,可以将N设置为10。

[0052] 也就是说,方法2在方法1的基础上,增加了一个判定音频信号中是否包含语音信号的条件:能量大于预设阈值的短时能量帧中是否存在至少N个连续短时能量帧。这样做可以有效降噪。由于在实际生活中,噪音相对于人类所说的话来说能量较低,且信号随机,因此利用方法2,便可以有效排除音频信号中噪声过多的情况,降低外部环境中噪音的影响,达到降噪的作用。

[0053] 需要特别说明的是,本申请实施例提供的上述语音信号检测方法,可适用于检测单声道音频信号、双声道音频信号或多声道音频信号等。其中,通过一个声道来采集的音频信号为单声道音频信号;通过两个声道来采集的音频信号为双声道音频信号,通过多个声道来采集的音频信号为多声道音频信号。

[0054] 在采用如图1所示的方法来检测双声道音频信号和多声道音频信号时,可按照步骤101~104提及的操作,分别针对获取到的每一路声道的音频信号进行检测,最终根据对每一路声道的音频信号的检测结果,判断获取到的音频信号中是否包含语音信号。

[0055] 具体的,若步骤101获取到的音频信号为单声道音频信号,便可针对该音频信号,直接执行步骤101~104中提及的操作,将检测结果作为最终检测结果。

[0056] 若步骤101获取到的音频信号不为单声道音频信号,而为双声道或多声道音频信

号,那么便分别对每一路声道的音频信号按照步骤101~104中的操作进行处理。若检测出每一路声道的音频信号均不包含语音信号,则确定步骤101获取到的音频信号不包含语音信号。若检测出至少一路声道的音频信号包含语音信号,则确定步骤101获取到的音频信号包含语音信号。

[0057] 另外,步骤102中所提及的预设语音信号的频率可以为任意语音的频率,本申请对此不进行任何限定。在实际应用中,可以根据实际情况,针对步骤101获取到的不同的音频信号,设置不同的预设语音信号的频率。需要特别说明的是,不管预设语音的频率是哪一种语音信号的频率,比如女高音的频率,或男低音的频率,只要使得最终划分出来的短时能量帧满足下述条件即可:短时能量帧对应的时长不小于步骤101获取到的音频信号对应的周期。为了达到比较好的检测效果、尽可能节省资源、提高处理速度,本申请实施例中,可以将预设语音信号的频率设置为最小人声频率,即82HZ。因为周期为频率的倒数,若预设语音信号的频率为最小人声频率,那么预设语音信号的周期便为最大人声周期,因此,不管步骤101获取到的音频信号的周期是多大,短时能量帧对应的时长均不小于上述获取到的音频信号的周期。

[0058] 需要特别说明的是,本申请实施例中,之所以要使得短时能量帧对应的时长均不小于步骤101获取到的音频信号的周期,是因为本申请实施例所提供的检测方法,是基于人类所说的话的特点来检测音频信号中是否包含语音信号的。人类所说的话相较于噪声来说,能量较高、较稳定且连续。若短时能量帧对应的时长小于步骤101获取到的音频信号的周期,那么短时能量帧对应的波形中不存在一个完整周期的波形,该短时能量帧的时长便相对较短。这一情况下,即便高能量帧比率大于预设比率,且能量大于预设阈值的短时能量帧中存在至少N个连续短时能量帧,仅仅可以表明音频信号中包含声音信号,却无法表明该声音信号为语音信号。因此,本申请实施例中,步骤101获取到的音频信号的时长应大于一个人声最大周期。

[0059] 另外,本申请实施例提供的语音信号检测方法尤其适用于在无需用户进行任何点击操作,聊天APP便可完成语音消息的发送这一应用场景。那么下面便针对该场景,详细说明本申请实施例提供的语音信号检测方法。其中,这一场景下,该方法的具体流程示意图如图2所示,包括下述步骤:

[0060] 步骤201,实时采集音频信号。

[0061] 若用户希望开启聊天APP之后,无需进行任何点击操作,该APP便可完成语音消息的发送,于是,当用户开启该APP之后,该APP便可开始不间断地针对外部环境进行录音,实时采集音频信号,以尽量避免漏掉用户所说的话。另外,在采集到音频信号之后,可以实时将该音频信号保存在本地。当用户关闭该APP之后,该APP便停止录音。

[0062] 步骤202,实时从采集到的音频信号中截取预设时长的音频信号。

[0063] 若APP一直进行录音,却并非实时进行语音信号的检测,便会导致语音消息的时效性较差。因此,APP可以实时截取步骤201采集到的音频信号中的、预设时长的音频信号,并针对该预设时长的音频信号进行后续检测。

[0064] 其中,可以将当前截取的预设时长的音频信号称为当前音频信号,可以将上一次截取的预设时长的音频信号称为上一次获取到的音频信号。

[0065] 步骤203,根据预设语音信号的频率,将预设时长的音频信号划分为多个短时能量

帧。

[0066] 步骤204,确定每个短时能量帧的能量。

[0067] 步骤205,根据每个短时能量帧的能量,检测预设时长的音频信号中是否包含语音信号。

[0068] 若检测出当前音频信号中包含语音信号,便判断上一次获取到的音频信号中是否包含语音信号,若判断出上一次获取到的音频信号中不包含语音信号,则可将当前音频信号的起始点确定为语音信号的起始点;若判断出上一次获取到的音频信号中包含语音信号,那么当前音频信号的起始点不为语音信号的起始点。

[0069] 若检测出当前音频信号中不包含语音信号,便判断上一次获取到的音频信号中是否包含语音信号,若判断出上一次获取到的音频信号中包含语音信号,则可将上一次获取到的音频信号的终点确定为语音信号的终点;若上一次获取到的音频信号中不包含语音信号,那么当前音频信号或者上一次获取到的音频信号的终点,均不为语音信号的终点。

[0070] 例如,如图3所示,其中A、B、C、D为四段相邻的预设时长的音频信号,A和D中不包含语音信号,B和C中包含语音信号,那么可以将B的开始点确定为语音信号的起始点,可以将C的终点确定为语音信号的终点。

[0071] 有时,当前音频信号刚好为用户一句话的开始或结尾部分,该音频信号中包含的语音信号比较少,这一情况下,APP有可能会误将该音频信号判定为不包含语音信号。那么为了尽量避免误判而导致遗漏掉用户所说的话,可以在检测出当前音频信号中包含语音信号后,判断上一次获取到的音频信号中是否包含语音信号,若判断出上一次获取到的音频信号中不包含语音信号,则可将上一次获取到的音频信号的起始点确定为语音信号的起始点。另外,可以在检测出当前音频信号中不包含语音信号后,判断上一次获取到的音频信号中是否包含语音信号,若判断出上一次获取到的音频信号中包含语音信号,则可将当前音频信号的终点确定为语音信号的终点。沿用上例,可以将A的起始点确定为语音信号的起始点,可以将D的终点确定为语音信号的终点。

[0072] 在APP检测出当前音频信号包含语音信号之后,可以将该音频信号发送给语音识别装置,以使得该语音识别装置可以对该音频信号进行语音处理,获取到语音结果,然后语音识别装置再将该音频信号发送给后续处理装置,最终将该音频信号以语音消息的形式发送出去。其中,为了使得发送出去的语音消息中包含的用户所说的话是完整的句子,APP可以将确定出的语音信号的起始点与终点之间的所有音频信号都发送给语音识别装置之后,向语音识别装置发送音频终止信号,用以告知语音识别装置用户当前所说的这一句话已经完结,以使得语音识别装置将该些音频信号一并发送给后续处理装置,最终将该些音频信号以语音消息的形式发送出去。

[0073] 另外,为了尽量避免误判情况的发生,还可以在获取到当前音频信号之后,在上一次获取到的音频信号中,截取预设时段的子信号,将当前音频信号和截取的子信号进行拼接,作为获取到的音频信号(后称拼接音频信号),并针对该拼接音频信号进行后续语音信号的检测。

[0074] 其中,可以将子信号拼接在当前音频信号之前。预设时段可以为上一次获取到的音频信号的尾部时段,该时段对应的时长可以为任意时长。为了使得最终检测结果更加准确,在本申请实施例中,可以将该预设时段对应的时长设置为不大于拼接音频信号对应的

时长与预设比率的乘积。

[0075] 若在检测出拼接音频信号中包含语音信号后,可判断上一次获取到的拼接音频信号中是否包含语音信号,若判断出上一次获取到的拼接音频信号中不包含语音信号,则可将拼接音频信号的起始点作为语音信号的起始点。若检测出拼接音频信号中不包含语音信号后,可判断上一次获取到的拼接音频信号中是否包含语音信号,若判断出上一次获取到的拼接音频信号中包含语音信号,则可将拼接音频信号的终点作为语音信号的终点。

[0076] 在本申请实施例中,APP除了可以一直不间断的进行录音外,还可以周期性进行录音,本申请实施例对此不进行任何限定。

[0077] 本申请实施例提供的语音信号检测方法,还可以通过语音信号检测装置来实现,该装置的具体结构示意图如图4所示,主要包括下述装置:

[0078] 获取模块41,获取音频信号;

[0079] 划分模块42,根据预设语音信号的频率,将所述音频信号划分为多个短时能量帧;

[0080] 确定模块43,确定每个短时能量帧的能量;

[0081] 检测模块44,根据每个短时能量帧的能量,检测所述音频信号中是否包含语音信号。

[0082] 在一种实施方式中,获取模块41获取当前音频信号;在上一次获取到的音频信号中,截取预设时段的子信号;

[0083] 将所述当前音频信号和截取的子信号进行拼接,作为获取到的音频信号。

[0084] 在一种实施方式中,划分模块42,根据预设语音信号的频率,确定出所述预设语音信号的周期;

[0085] 按照确定出的周期,将所述音频信号划分为对应的时长均为所述周期的多个短时能量帧。

[0086] 在一种实施方式中,检测模块44,确定能量大于预设阈值的短时能量帧的数量占所有短时能量帧总数量的比率;

[0087] 判断所述比率是否大于预设比率;

[0088] 若是,则确定检测到所述音频信号中包含语音信号;

[0089] 若否,则确定未检测到所述音频信号中包含语音信号。

[0090] 在一种实施方式中,检测模块44,确定能量大于预设阈值的短时能量帧的数量占所有短时能量帧总数量的比率;

[0091] 判断所述比率是否大于预设比率;

[0092] 若否,则确定未检测到所述音频信号中包含语音信号;

[0093] 若是,则当能量大于预设阈值的短时能量帧中存在至少N个连续短时能量帧时,确定检测到所述音频信号中包含语音信号,当能量大于预设阈值的短时能量帧中不存在至少N个连续短时能量帧时,确定未检测到所述音频信号中包含语音信号。

[0094] 与现有技术中的通过傅里叶变换等复杂计算来确定音频信号中是否包含语音信号的检测方法相比,本申请实施例采用的语音信号检测方法,无需进行傅里叶变换等复杂计算,通过根据预设语音信号的频率,将获取到的音频信号划分为多个短时能量帧,进而确定出每个短时能量帧的能量,并根据每个短时能量帧的能量,便可检测出获取到的音频信号中是否包含语音信号。因此,本申请实施例提供的语音信号检测方法,能够解决现有技术

中的语音信号检测方法存在的处理速度较慢,且耗费资源较多的问题。

[0095] 本发明是参照根据本发明实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0096] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0097] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0098] 在一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0099] 内存可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0100] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0101] 还需要说明的是,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0102] 本领域技术人员应明白,本申请的实施例可提供为方法、系统或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0103] 以上所述仅为本申请的实施例而已,并不用于限制本申请。对于本领域技术人员

来说,本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等,均应包含在本申请的权利要求范围之内。

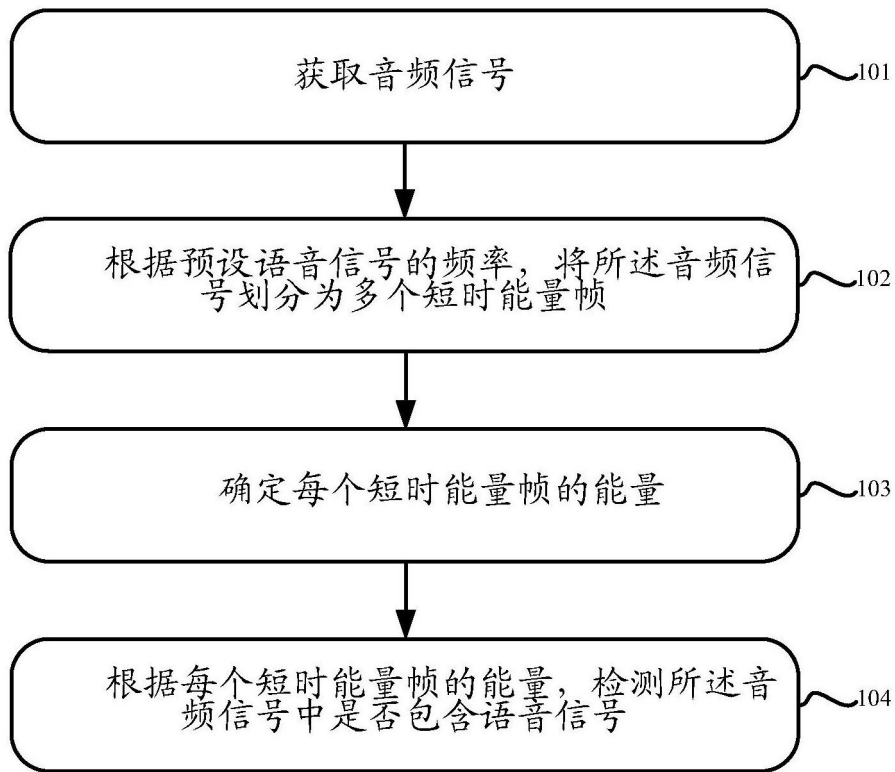


图1

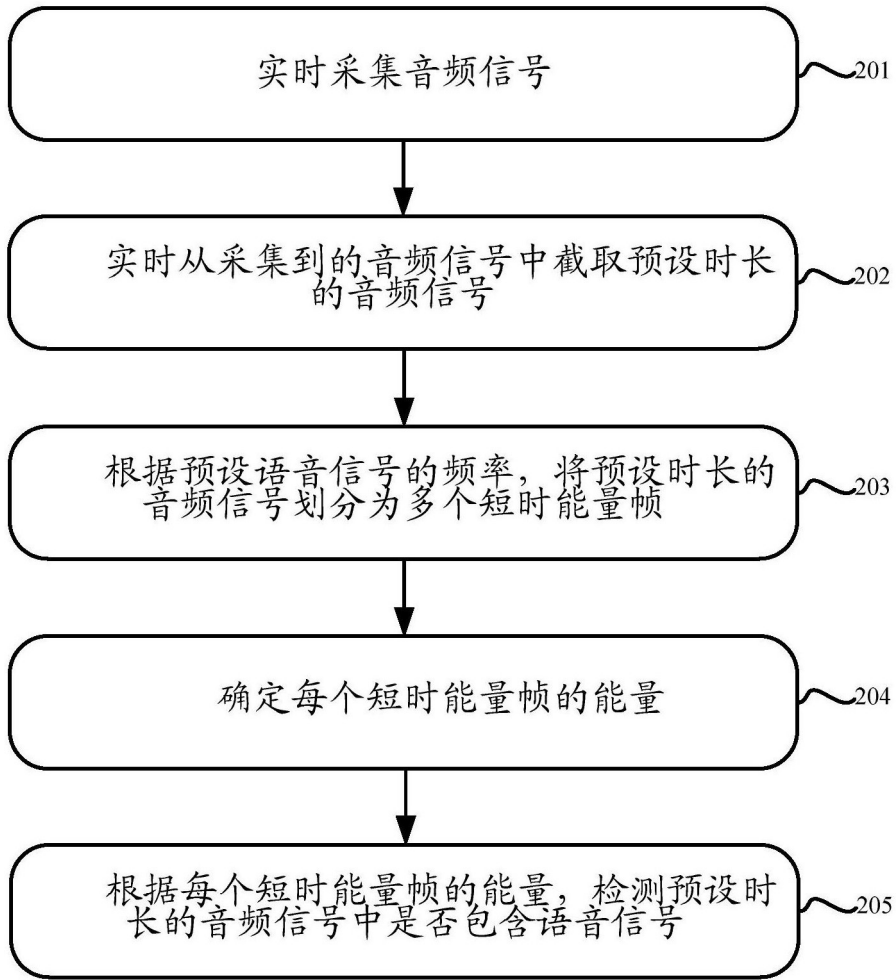


图2



图3

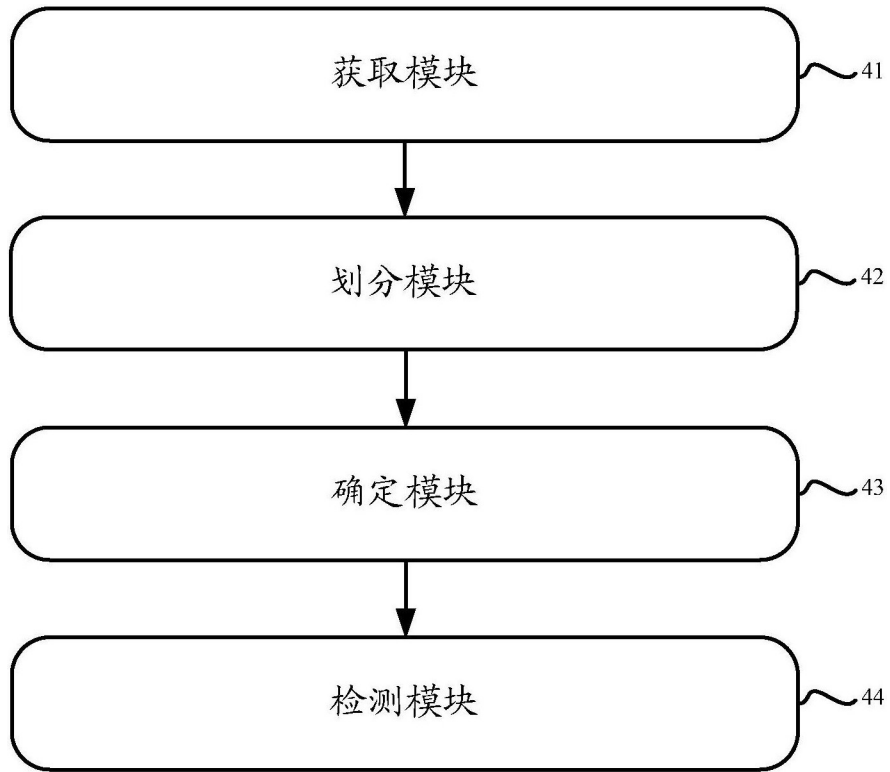


图4

Abstract

The present application discloses a voice signal detection method and apparatus for solving the problem that the processing speed is slow and the resource consumption is high existing in the voice signal detection method in the conventional technology. The method comprises: acquiring an audio signal; dividing the audio signal into a plurality of short-time energy frames according to the frequency of the preset voice signal; determining the energy of each short-term energy frame; detecting whether the voice signal is included in the audio signal based on the energy of each short-time energy frame.