



(12) 发明专利

(10) 授权公告号 CN 110827849 B

(45) 授权公告日 2022.07.26

(21) 申请号 201911099096.0

(22) 申请日 2019.11.11

(65) 同一申请的已公布的文献号
申请公布号 CN 110827849 A

(43) 申请公布日 2020.02.21

(73) 专利权人 广州国音智能科技有限公司
地址 510000 广东省广州市黄埔区黄埔东
路3110号三社大楼二楼2009房

(72) 发明人 陈昊亮 许敏强 杨世清

(74) 专利代理机构 深圳市世纪恒程知识产权代
理事务所 44287
专利代理师 晏波

(51) Int. Cl.
G10L 21/0272 (2013.01)
G10L 25/27 (2013.01)

(56) 对比文件

US 2017/0110121 A1, 2017.04.20

CN 108449497 A, 2018.08.24

CN 108182945 A, 2018.06.19

CN 109326303 A, 2019.02.12

CN 109741754 A, 2019.05.10

KR 10-2007-0109156 A, 2007.11.15

KR 10-2010-0073167 A, 2010.07.01

WO 2019/008580 A1, 2019.01.10

JP 特开2012-211768 A, 2012.11.01

Jessica Rincón-Trujillo et

al., Analysis of Speech Separation Methods based on Deep Learning.《Research in Computing Science》.2019,第21-29页.

周燕.基于Fast ICA的多说话人识别系统.《苏州市职业大学学报》.2011,第22卷(第2期),第10-13页.

审查员 李召卿

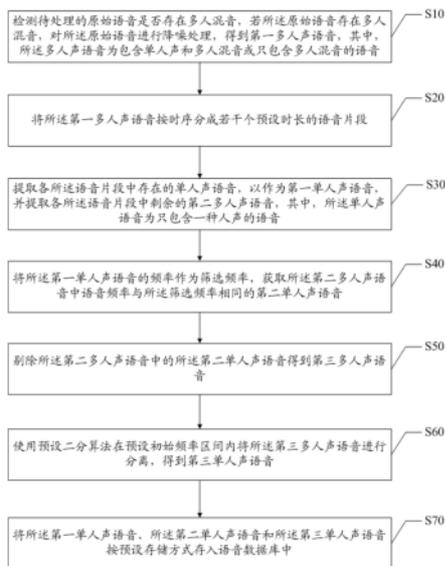
权利要求书2页 说明书10页 附图4页

(54) 发明名称

数据建库的人声分离方法、装置、终端及可读存储介质

(57) 摘要

本发明公开了一种数据建库的人声分离方法,检测原始语音是否存在多人混音,若所述原始语音存在多人混音,对所述原始语音进行降噪处理,得到第一多人声语音,将所述第一多人声语音按时序分成若干个预设时长的语音片段,提取所述语音片段中的第一单人声语音,并提取各所述语音片段中剩余的第二多人声语音,将所述第一单人声语音的频率作为筛选频率,以获取第二单人声语音,剔除所述第二多人声语音中的所述第二单人声语音得到第三多人声语音,使用预设二分算法分离所述第三多人声语音得到第三单人声语音,将所有单人声语音存入语音数据库中。本发明还公开了一种装置、终端及可读存储介质。提高了语音会议记录处理效率,降低了语音会议记录处理的误差。



CN 110827849 B

1. 一种数据建库的人声分离方法,其特征在于,包括:

检测待处理的原始语音是否存在多人混音,若所述原始语音存在多人混音,对所述原始语音进行降噪处理,得到第一多人声语音,其中,所述第一多人声语音为包含单人声和多人混音的语音;

将所述第一多人声语音按时序分成若干个预设时长的语音片段;

提取各所述语音片段中存在的单人声语音,以作为第一单人声语音,并提取各所述语音片段中剩余的第二多人声语音,其中,所述单人声语音为只包含一种人声的语音;

将所述第一单人声语音的频率作为筛选频率,获取所述第二多人声语音中语音频率与所述筛选频率相同的第二单人声语音;

剔除所述第二多人声语音中的所述第二单人声语音得到第三多人声语音;

使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离,得到第三单人声语音;

将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中。

2. 如权利要求1所述的数据建库的人声分离方法,其特征在于,所述使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离,得到第三单人声语音的步骤包括:

确定预设初始频率区间,基于预设二分算法,将所述初始频率区间分为两个子频率区间;

根据所述子频率区间,对所述第三多人声语音进行语音分离,直至提取所有单人声语音,以作为第三单人声语音。

3. 如权利要求2所述的数据建库的人声分离方法,其特征在于,所述根据所述子频率区间,对所述第三多人声语音进行语音分离的步骤具体包括:

根据所述子频率区间,对所述第三多人声语音进行语音分离,得到分离中间结果;

判断所述分离中间结果是否只包含单人声语音;

若所述分离中间结果只包含单人声语音,则将所述分离中间结果作为第三单人声语音;

若所述分离中间结果为多人声语音,则基于预设二分算法对所述子频率区间拆分,形成新的子频率区间,执行所述根据所述子频率区间,对所述第三多人声语音进行语音分离,得到分离中间结果的步骤,直至所述分离中间结果只包含单人声语音,将所述分离中间结果作为第三单人声语音。

4. 如权利要求1所述的数据建库的人声分离方法,其特征在于,所述将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中的步骤之前包括:

为所述第一单人声语音、所述第二单人声语音和所述第三单人声语音建立索引,所述索引为所述第一单人声语音、所述第二单人声语音和所述第三单人声语音的开始时刻和结束时刻。

5. 如权利要求1所述的数据建库的人声分离方法,其特征在于,所述检测待处理的原始语音是否存在多人混音的步骤之后还包括:

若所述原始语音不存在多人混音,则对所述原始语音进行降噪处理,得到纯单人声语

音；

提取所述纯单人声语音中的第四单人声语音，并存入语音数据库。

6. 如权利要求5所述的数据建库的人声分离方法，其特征在于，所述并存入语音数据库中的步骤之前包括：

为所述第四单人声语音建立索引，所述索引为所述第四单人声语音的开始时刻和结束时刻。

7. 如权利要求1至6任一所述的数据建库的人声分离方法，其特征在于，所述对所述原始语音进行降噪处理的步骤包括：

获取反向接入电路的麦克风的负信号语音；

将所述负信号语音与所述原始语音进行相互叠加以消除噪音。

8. 一种数据建库的人声分离装置，其特征在于，包括：

检测模块，用于检测待处理的原始语音是否存在多人对话，若原始语音存在多人对话，对所述原始语音进行降噪处理，得到第一多人声语音，其中，所述第一多人声语音为包含单人声和多人混音的语音；

处理模块，用于将所述第一多人声语音按时序分成若干个预设时长的语音片段；

获取模块，用于提取各所述语音片段中存在的单人声语音，以作为第一单人声语音，并提取各所述语音片段中剩余的第二多人声语音，其中，所述单人声语音为只包含一种人声的语音；

获取模块，还用于将所述第一单人声语音的频率作为筛选频率，获取所述第二多人声语音中语音频率与所述筛选频率相同的第二单人声语音；

获取模块，还用于剔除所述第二多人声语音中的所述第二单人声语音得到第三多人声语音；

获取模块，还用于使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离，得到第三单人声语音；

存储模块，用于将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中。

9. 一种终端，其特征在于，所述终端包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的数据建库的人声分离程序，所述数据建库的人声分离程序被所述处理器执行时实现如权利要求1至7中任一项所述的数据建库的人声分离方法的步骤。

10. 一种可读存储介质，其特征在于，所述可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现如权利要求1至7中任一项所述的数据建库的人声分离方法的步骤。

数据建库的人声分离方法、装置、终端及可读存储介质

技术领域

[0001] 本发明涉及人声分离技术领域,尤其涉及数据建库的人声分离方法、装置、终端及可读存储介质。

背景技术

[0002] 当前会议记录有很多形式,比如语音,以便进行会议回顾。但是,会议过程中会有如讨论、争论等情形,这种多人同时讲话加上其他噪声,给后期的人工处理语音会议记录,也就是将参会人员所说的内容分别记录带来了巨大的难度,需要花费大量的时间来处理,而计算机又难以将参会人员所说的内容分开记录,并且由于人的主观性,会导致处理语音会议记录时出现误差,造成了语音会议记录处理效率低、误差大的问题。

发明内容

[0003] 本发明的主要目的在于提供一种涉及数据建库的人声分离方法、装置、终端及可读存储介质,旨在解决语音会议记录处理过程中出现的效率低、误差大的技术问题。

[0004] 为实现上述目的,本发明提供一种数据建库的人声分离方法,所述数据建库的人声分离方法包括:

[0005] 检测待处理的原始语音是否存在多人混音,若所述原始语音存在多人混音,对所述原始语音进行降噪处理,得到第一多人声语音,其中,所述多人声语音为包含单人声和多人混音或只包含多人混音的语音;

[0006] 将所述第一多人声语音按时序分成若干个预设时长的语音片段;

[0007] 提取各所述语音片段中存在的单人声语音,以作为第一单人声语音,并提取各所述语音片段中剩余的第二多人声语音,其中,所述单人声语音为只包含一种人声的语音;

[0008] 将所述第一单人声语音的频率作为筛选频率,获取所述第二多人声语音中语音频率与所述筛选频率相同的第二单人声语音;

[0009] 剔除所述第二多人声语音中的所述第二单人声语音得到第三多人声语音;

[0010] 使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离,得到第三单人声语音;

[0011] 将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中。

[0012] 可选地,所述使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离,得到第三单人声语音的步骤包括:

[0013] 确定预设初始频率区间,基于预设二分算法,将所述初始频率区间分为两个子频率区间;

[0014] 根据所述子频率区间,对所述第三多人声语音进行语音分离,直至提取所有单人声语音,以作为第三单人声语音。

[0015] 可选地,所述根据所述子频率区间,对所述第三多人声语音进行语音分离的步骤

具体包括：

[0016] 根据所述子频率区间,对所述第三多人声语音进行语音分离,得到分离中间结果;

[0017] 判断所述分离中间结果是否只包含单人声语音;

[0018] 若所述分离中间结果只包含单人声语音,则将所述分离中间结果作为第三单人声语音;

[0019] 若所述分离中间结果为多人声语音,则基于预设二分算法对所述子频率区间拆分,形成新的子频率区间,执行所述根据所述子频率区间,对所述第三多人声语音进行语音分离,得到分离中间结果的步骤,直至所述分离中间结果只包含单人声语音,将所述分离中间结果作为第三单人声语音。

[0020] 可选地,所述将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中的步骤之前包括:

[0021] 为所述第一单人声语音、所述第二单人声语音和所述第三单人声语音建立索引,所述索引为所述第一单人声语音、所述第二单人声语音和所述第三单人声语音的开始时刻和结束时刻。

[0022] 可选地,所述检测待处理的原始语音是否存在多人混音的步骤之后还包括:

[0023] 若所述原始语音不存在多人混音,则对所述原始语音进行降噪处理,得到纯单人声语音;

[0024] 提取所述纯单人声语音中的第四单人声语音,并存入语音数据库。

[0025] 可选地,所述并存入语音数据库中的步骤之前包括:

[0026] 为所述第四单人声语音建立索引,所述索引为所述第四单人声语音的开始时刻和结束时刻。

[0027] 可选地,所述对所述原始语音进行降噪处理的步骤包括:

[0028] 获取反向接入电路的麦克风的负信号语音;

[0029] 将所述负信号语音与所述原始语音进行相互叠加以消除噪音。

[0030] 为实现上述目的,本发明还提供一种数据建库的人声分离装置,所述数据建库的人声分离装置包括以下步骤:

[0031] 检测模块,用于检测待处理的原始语音是否存在多人对话,若原始语音存在多人对话,对所述原始语音进行降噪处理,得到第一多人声语音,其中,所述第一多人声语音为包含单人声和多人混音或只包含多人混音的语音;

[0032] 处理模块,用于将所述第一多人声语音按时序分成若干个预设时长的语音片段;

[0033] 获取模块,用于提取各所述语音片段中存在的单人声语音,以作为第一单人声语音,并提取各所述语音片段中剩余的第二多人声语音,其中,所述单人声语音为只包含一种人声的语音;

[0034] 获取模块,还用于将所述第一单人声语音的频率作为筛选频率,获取所述第二多人声语音中语音频率与所述筛选频率相同的第二单人声语音;

[0035] 获取模块,还用于剔除所述第二多人声语音中的所述第二单人声语音得到第三多人声语音;

[0036] 获取模块,还用于使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离,得到第三单人声语音;

[0037] 存储模块,用于将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中。

[0038] 为实现上述目的,本发明还提供一种终端,所述终端包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的数据建库的人声分离程序,所述数据建库的人声分离程序被所述处理器执行时实现上述的数据建库的人声分离方法的步骤。

[0039] 此外,为实现上述目的,本发明还提供一种计算机存储介质,计算机存储介质上存储有数据建库的人声分离程序,数据建库的人声分离程序被处理器执行时实现上述的数据建库的人声分离方法的步骤。

[0040] 本发明实施例提出的一种数据建库的人声分离方法、装置、终端及可读存储介质,通过检测待处理的原始语音是否存在多人混音,若所述原始语音存在多人混音,对所述原始语音进行降噪处理,得到第一多人声语音,其中,所述多人声语音为包含单人声和多人混音或只包含多人混音的语音,将所述第一多人声语音按时序分成若干个预设时长的语音片段,提取各所述语音片段中存在的单人声语音,以作为第一单人声语音,并提取各所述语音片段中剩余的第二多人声语音,其中,所述单人声语音为只包含一种人声的语音,将所述第一单人声语音的频率作为筛选频率,获取所述第二多人声语音中语音频率与所述筛选频率相同的第二单人声语音,剔除所述第二多人声语音中的所述第二单人声语音得到第三多人声语音,使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离,得到第三单人声语音,将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中。由于会议时间一般比较长,所以将上述第一多人声语音按时序分成若干个预设时长的语音片段,方便对多个语音片段同时处理,提高对语音的处理效率,采用预设二分算法的好处在于循环结构简单,没有复杂的步骤,只需要不断地拆分频率区间就能达到将多人声语音根据参会人员语音频率的不同分离成单人声语音,提高了对语音处理的效率,另外,由于语音都有固定的频率,通过不断缩小频率区间,实现精确区分每一位参会人员的语音,保证语音会议记录处理的准确性。实现了提高语音会议记录处理效率,降低语音会议记录处理的误差的目的。

附图说明

[0041] 图1是本发明实施例方案涉及的硬件运行环境的终端\装置结构示意图;

[0042] 图2为本发明数据建库的人声分离方法第一实施例的流程示意图;

[0043] 图3为本发明数据建库的人声分离方法第一实施例的举例说明示意图;

[0044] 图4为本发明数据建库的人声分离方法第二实施例的流程示意图;

[0045] 图5为本发明数据建库的人声分离方法第二实施例中预设二分算法的一种可能结果的流程示意图;

[0046] 图6为本发明数据建库的人声分离方法第三实施例的流程示意图。

[0047] 本发明目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

具体实施方式

[0048] 应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0049] 本发明实施例的主要解决方案是:检测待处理的原始语音是否存在多人混音,若

所述原始语音存在多人混音,对所述原始语音进行降噪处理,得到第一多人声语音,其中,所述多人声语音为包含单人声和多人混音或只包含多人混音的语音,将所述第一多人声语音按时序分成若干个预设时长的语音片段,提取各所述语音片段中存在的单人声语音,以作为第一单人声语音,并提取各所述语音片段中剩余的第二多人声语音,其中,所述单人声语音为只包含一种人声的语音,将所述第一单人声语音的频率作为筛选频率,获取所述第二多人声语音中语音频率与所述筛选频率相同的第二单人声语音,剔除所述第二多人声语音中的所述第二单人声语音得到第三多人声语音,使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离,得到第三单人声语音,将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中。实现了提高语音会议记录处理效率,降低语音会议记录处理的误差的目的。

[0050] 由于现有技术在处理语音会议记录时,是由人工进行处理的,由于人的主观性,处理会议记录的过程中可能会出现听到的和写下来的不一致,导致会议记录出现误差,并且在进行人声分离的过程中处理速度慢,效率低。

[0051] 本发明提供一种解决方案,使语音会议记录处理过程更高效,误差更小。

[0052] 如图1所示,图1是本发明实施例方案涉及的硬件运行环境的终端结构示意图。

[0053] 本发明实施例终端可以是PC,也可以是智能手机、平板电脑、电子书阅读器、MP3 (Moving Picture Experts Group Audio Layer III,动态影像专家压缩标准音频层面3)播放器、MP4 (Moving Picture Experts Group Audio Layer IV,动态影像专家压缩标准音频层面3)播放器、便携计算机等具有显示功能的可移动式终端设备。

[0054] 如图1所示,该终端可以包括:处理器1001,例如CPU,网络接口1004,用户接口1003,存储器1005,通信总线1002。其中,通信总线1002用于实现这些组件之间的连接通信。用户接口1003可以包括显示屏(Display)、输入单元比如键盘(Keyboard),可选用户接口1003还可以包括标准的有线接口、无线接口。网络接口1004可选的可以包括标准的有线接口、无线接口(如WI-FI接口)。存储器1005可以是高速RAM存储器,也可以是稳定的存储器(non-volatile memory),例如磁盘存储器。存储器1005可选的还可以是独立于前述处理器1001的存储装置。

[0055] 可选地,终端还可以包括摄像头、RF (Radio Frequency,射频)电路,传感器、音频电路、WiFi模块等等。其中,传感器比如光传感器、运动传感器以及其他传感器。具体地,光传感器可包括环境光传感器及接近传感器,其中,环境光传感器可根据环境光线的明暗来调节显示屏的亮度,接近传感器可在移动终端移动到耳边时,关闭显示屏和/或背光。作为运动传感器的一种,重力加速度传感器可检测各个方向上(一般为三轴)加速度的大小,静止时可检测出重力的大小及方向,可用于识别移动终端姿态的应用(比如横竖屏切换、相关游戏、磁力计姿态校准)、振动识别相关功能(比如计步器、敲击)等;当然,移动终端还可配置陀螺仪、气压计、湿度计、温度计、红外线传感器等其他传感器,在此不再赘述。

[0056] 本领域技术人员可以理解,图1中示出的终端结构并不构成对终端的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。

[0057] 如图1所示,作为一种计算机存储介质的存储器1005中可以包括操作系统、网络通信模块、用户接口模块以及数据建库的人声分离程序。

[0058] 在图1所示的终端中,网络接口1004主要用于连接后台服务器,与后台服务器进行

数据通信;用户接口1003主要用于连接客户端(用户端),与客户端进行数据通信;而处理器1001可以用于调用存储器1005中存储的数据建库的人声分离程序,并执行以下操作:

[0059] 检测待处理的原始语音是否存在多人混音,若所述原始语音存在多人混音,对所述原始语音进行降噪处理,得到第一多人声语音,其中,所述多人声语音为包含单人声和多人混音或只包含多人混音的语音;

[0060] 将所述第一多人声语音按时序分成若干个预设时长的语音片段;

[0061] 提取各所述语音片段中存在的单人声语音,以作为第一单人声语音,并提取各所述语音片段中剩余的第二多人声语音,其中,所述单人声语音为只包含一种人声的语音;

[0062] 将所述第一单人声语音的频率作为筛选频率,获取所述第二多人声语音中语音频率与所述筛选频率相同的第二单人声语音;

[0063] 剔除所述第二多人声语音中的所述第二单人声语音得到第三多人声语音;

[0064] 使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离,得到第三单人声语音;

[0065] 将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中。

[0066] 进一步地,使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离,得到第三单人声语音的步骤包括:

[0067] 确定预设初始频率区间,基于预设二分算法,将所述初始频率区间分为两个子频率区间;

[0068] 根据所述子频率区间,对所述第三多人声语音进行语音分离,直至提取所有单人声语音,以作为第三单人声语音。

[0069] 进一步地,根据所述子频率区间,对所述第三多人声语音进行语音分离的步骤具体包括:

[0070] 根据所述子频率区间,对所述第三多人声语音进行语音分离,得到分离中间结果;

[0071] 判断所述分离中间结果是否只包含单人声语音;

[0072] 若所述分离中间结果只包含单人声语音,则将所述分离中间结果作为第三单人声语音;

[0073] 若所述分离中间结果为多人声语音,则基于预设二分算法对所述子频率区间拆分,形成新的子频率区间,执行所述根据所述子频率区间,对所述第三多人声语音进行语音分离,得到分离中间结果的步骤,直至所述分离中间结果只包含单人声语音,将所述分离中间结果作为第三单人声语音。

[0074] 进一步地,将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中的步骤之前包括:

[0075] 为所述第一单人声语音、所述第二单人声语音和所述第三单人声语音建立索引,所述索引为所述第一单人声语音、所述第二单人声语音和所述第三单人声语音的开始时刻和结束时刻。

[0076] 进一步地,检测待处理的原始语音是否存在多人混音的步骤之后还包括:

[0077] 若所述原始语音不存在多人混音,则对所述原始语音进行降噪处理,得到纯单人声语音;

[0078] 提取所述纯单人声语音中的第四单人声语音,并存入语音数据库。

[0079] 进一步地,并存入语音数据库的步骤之前包括:

[0080] 为所述第四单人声语音建立索引,所述索引为所述第四单人声语音的开始时刻和结束时刻。

[0081] 进一步地,对所述原始语音进行降噪处理的步骤包括:

[0082] 获取反向接入电路的麦克风的负信号语音;

[0083] 将所述负信号语音与所述原始语音进行相互叠加以消除噪音。

[0084] 参照图2,在本发明数据建库的人声分离方法第一实施例中,所述数据建库的人声分离方法包括:

[0085] 步骤S10,检测待处理的原始语音是否存在多人对话,若原始语音存在多人对话,对所述原始语音进行降噪处理,得到第一多人声语音,其中,所述第一多人声语音为包含单人声和多人混音或只包含多人混音的语音;

[0086] 检测待处理的原始语音是否存在多人对话,如果待处理的原始语音存在多人对话,那么需要对原始语音进行降噪处理,经过降噪处理得到去除噪声后,得到只包含人声的第一多人声语音,其中,多人声语音是指包含单人声和多人混音或只包含多人混音的语音,降噪技术具体为,除了距离参会人员很近的麦克风采集正信号之外,还有一个离参会人员较远的麦克风,较远的麦克风反向接入电路,它采集的是负信号,在正负信号进行叠加时,由于噪音与较近的麦克风和较远的麦克风之间的距离相等,所以被抵消,而参会人员由于距离两个麦克风距离不同,叠加时会有一个差值,从而不会被抵消,例,噪音正信号强度为1,噪音负信号强度为-1,参会人员语音正信号强度为2,参会人员负信号强度为-0.5,那么在正负信号进行叠加时,噪音消除,得到强度为1.5的参会人员语音。

[0087] 步骤S20,将所述第一多人声语音按时序分成若干个预设时长的语音片段;

[0088] 将第一多人声语音按照时间的先后顺序分成若干个预设时长的语音片段,便于后续对语音的处理,时长可以按需要设置,例,多人声语音时长为500秒,时长可以设置为50秒,则得到10个语音片段。

[0089] 步骤S30,提取各所述语音片段中存在的单人声语音,以作为第一单人声语音,并提取各所述语音片段中剩余的第二多人声语音,其中,所述单人声语音为只包含一种人声的语音;

[0090] 把容易提取的单人声语音先提取出来,并且作为第一单人声语音,然后将所有语音片段中的单人声语音剔除,得到不包含单人声语音的第二多人声语音,其中,单人声语音是指只包含一种人声的语音,将所有语音片段中的单人声语音剔除的方法是,首先将所有提取到的单人声语音的语音信号反相,然后将得到的单人声语音的反相语音信号与所有语音片段中的语音信号进行叠加,就可以得到剔除了所有语音片段中的单人声语音的不包含单人声语音的第二多人声语音。

[0091] 步骤S40,将所述第一单人声语音的频率作为筛选频率,获取所述第二多人声语音中语音频率与所述筛选频率相同的第二单人声语音;

[0092] 会议过程中,有些参会人员存在单独发言的情况,有些参会人员可能只在和其他参会人员一起讨论时发过言,参会人员单独发言的语音就是第一单人声语音,相应的,参会人员一起讨论时的语音就是第二多人声语音,将第一单人声语音的频率作为筛选频率来筛

选第二多人声语音,就可以得到存在单独发言的情况的参会人员在和和其他参会人员一起讨论时发言的语音,将它作为第二单人声语音。

[0093] 步骤S50,剔除所述第二多人声语音中的所述第二单人声语音得到第三多人声语音;

[0094] 将第二多人声语音中的第二单人声语音剔除得到第三多人声语音的方案与上述步骤S30中的将所有语音片段中的单人声语音剔除的方法类似,在此不再赘述。

[0095] 步骤S60,使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离,得到第三单人声语音;

[0096] 使用预设的二分算法,通过不断地拆分预设初始频率区间,使第三多人声语音依据语音频率的不同,分离为单人声语音。

[0097] 参照图3,其中,a、b和c分别代表三个参会人员的语音,根据降噪后的第一多人声语音,第一多人声语音总时长为120s,分成四个语音片段,则每个语音片段的时长为30s,第一个语音片段包含单人声语音a,第二个语音片段包含单人声语音a和多人混音abc,第三个语音片段包含单人声语音a和多人混音ab,第四个语音片段包含多人混音ac,首先,提取四个语音片段中的单人声语音a,作为第一单人声语音,每个语音片段剔除单人声语音a得到包含第二语音片段的多人混音abc、第三语音片段的多人混音ab和第四语音片段的多人混音ac作为第二多人声语音,获取第一单人声语音a的频率,根据第一单人声语音a的频率,获取第二多人声语音abc、ab和ac中与第一单人声语音a的频率相同的单人声语音a作为第二单人声语音,剔除第二多人声语音abc、ab和ac中的第二单人声语音a得到包含多人混音bc、单人声语音b和单人声语音c的第三多人声语音,然后根据预设的二分算法,通过不断地二分频率区间,直到得到单人声语音b和单人声语音c作为第三单人声语音。

[0098] 步骤S70,将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中。

[0099] 将第一单人声语音、第二单人声语音和第三单人声语音的开始时刻和结束时刻作为索引,方便后续根据大致时刻来查找相应时刻的语音,比如,想知道某位参会人员某句话之后的语音内容比如对该参会人员问题的回复,可以根据该参会人员某句话的结束时刻来搜索接下来的语音就可以得到想要的语音,相应地,想知道某位参会人员某句话之前的语音内容比如对该参会人员答复的提问,可以根据该参会人员某句话的开始时刻来搜索之前的语音就可以得到想要的语音。

[0100] 在本实施例中终端检测待处理的原始语音是否存在多人混音,若所述原始语音存在多人混音,对所述原始语音进行降噪处理,得到第一多人声语音,其中,所述多人声语音为包含单人声和多人混音或只包含多人混音的语音,将所述第一多人声语音按时序分成若干个预设时长的语音片段,提取各所述语音片段中存在的单人声语音,以作为第一单人声语音,并提取各所述语音片段中剩余的第二多人声语音,其中,所述单人声语音为只包含一种人声的语音,将所述第一单人声语音的频率作为筛选频率,获取所述第二多人声语音中语音频率与所述筛选频率相同的第二单人声语音,剔除所述第二多人声语音中的所述第二单人声语音得到第三多人声语音,使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离,得到第三单人声语音,将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中。由于会议时间一般比较长,所以将上述第一多人

声语音按时序分成若干个预设时长的语音片段,方便对多个语音片段同时处理,提高对语音的处理效率,采用预设二分算法的好处在于循环结构简单,没有复杂的步骤,只需要不断地拆分频率区间就能达到将多人声语音根据参会人员语音频率的不同分离成单人声语音,提高了对语音处理的效率,另外,由于语音都有固定的频率,通过不断缩小频率区间,实现精确区分每一位参会人员的语音,保证语音会议记录处理的准确性。实现了提高语音会议记录处理效率,降低语音会议记录处理的误差的目的。

[0101] 进一步地,参照图4,在本发明数据建库的人声分离方法第二实施例中,基于第一实施例,所述使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离,得到第三单人声语音的步骤包括:

[0102] 步骤S80,确定预设初始频率区间,基于预设二分算法,将所述初始频率区间分为两个子频率区间;

[0103] 确定预设的初始频率区间,例,该频率区间可以是20-20000Hz,也可以将区间缩小,设置为人的语音频率区间,一般为300-3000Hz,根据预设二分算法,将初始频率均分为两个子频率区间。

[0104] 步骤S90,根据所述子频率区间,对所述第三多人声语音进行语音分离,直至提取所有单人声语音,以作为第三单人声语音。

[0105] 参照图5,N为自然数,由于二分算法的循环次数是根据语音会议记录中混音的复杂程度决定的,所以循环次数是不固定的,因此,N可以为0,也可以为其它的正整数,如1、2等,将第三多人声语音根据其频率所属子频率区间的不同分到各自所属的子频率区间,判断两个子频率区间中的语音是否都是单人声语音,如果两个子频率区间中的语音都是单人声语音,则语音分离过程结束,将基于预设二分算法得到的单人声语音作为第三单人声语音;如果属于两个子频率区间中的语音还有多人声语音,则继续判断属于两个子频率区间中的多人声语音是否有单人声语音,如果有单人声语音,则将单人声语音作为第三单人声语音的一部分并将多人声语音所处的子频率区间继续基于预设二分算法分为新的两个子频率区间,如果没有单人声语音,则将多人声语音所处的子频率区间分别继续分为新的两个子频率区间,然后将多人声语音中的语音根据频率的不同分到各自所属的新子频率区间,判断新子频率区间中的语音是否都是单人声语音,如果新子频率区间中的语音都是单人声语音,则语音分离过程结束,得到单人声语音,并将分离过程得到的所有单人声语音作为第三单人声语音,如果两个子频率区间中的语音还有多人声语音,则继续按照上述预设二分算法进行分离,直到经过语音分离过程得到的所有语音都是单人声语音也就是得到第三单人声语音为止。

[0106] 在本实施例中,通过不断循环缩小频率区间,直到所有子频率区间的语音都为单人声语音为止,因为预设的二分算法循环结构简单,没有复杂的步骤,所以处理过程高效,提高了对语音处理的效率,并且,由于语音都有固定的频率,通过不断缩小频率区间,实现精确区分每一位参会人员的语音,保证了语音会议记录处理的准确性,降低了语音会议记录处理的误差。

[0107] 参照图6,在本发明数据建库的人声分离方法第三实施例中,基于第二实施例,所述检测待处理的原始语音是否存在多人混音的步骤之后还包括:

[0108] 步骤S100,若所述原始语音不存在多人混音,则对所述原始语音进行降噪处理,得

到纯单人声语音；

[0109] 如果原始语音中没有多人混音，只有单人声语音，则对原始语音进行降噪处理，降噪技术具体为，除了距离参会人员很近的麦克风采集正信号之外，还有一个离参会人员较远的麦克风，较远的麦克风反向接入电路，它采集的是负信号，在正负信号进行叠加时，由于噪音与较近的麦克风和较远的麦克风之间的距离相等，所以被抵消，而参会人员由于距离两个麦克风距离不同，叠加时会有一个差值，从而不会被抵消，例，噪音正信号强度为1，噪音负信号强度为-1，参会人员语音正信号强度为2，参会人员负信号强度为-0.5，那么在正负信号进行叠加时，噪音消除，得到强度为1.5的参会人员语音，经过降噪处理得到去除噪声后，得到只包含人声的纯人声语音。

[0110] 步骤S110，提取所述纯单人声语音中的第四单人声语音，并存入语音数据库。

[0111] 提取每个参会人员的语音作为第四单人声语音，并将每个语音的开始时刻和结束时刻作为索引，方便后续根据大致时刻来查找相应时刻的语音，比如，想知道某位参会人员某句话之后的语音内容比如对该参会人员问题的回复，可以根据该参会人员某句话的结束时刻来搜索接下来的语音就可以得到想要的语音，相应地，想知道某位参会人员某句话之前的语音内容比如对该参会人员答复的提问，可以根据该参会人员某句话的开始时刻来搜索之前的语音就可以得到想要的语音。

[0112] 在本实施例中，原始语音不存在多人混音，只包含单人声语音，将每个语音的开始时刻和结束时刻作为索引然后存入数据库，方便后续对指定语音的取用，而不用遍历数据库，然后逐条判断该语音是否是当前需要使用的语音，提高了语音会议记录处理的效率。

[0113] 本发明还提供一种数据建库的人声分离装置，所述数据建库的人声分离装置包括：

[0114] 检测模块，用于检测待处理的原始语音是否存在多人对话，若原始语音存在多人对话，对所述原始语音进行降噪处理，得到第一多人声语音，其中，所述第一多人声语音为包含单人声和多人混音或只包含多人混音的语音；

[0115] 处理模块，用于将所述第一多人声语音按时序分成若干个预设时长的语音片段；

[0116] 获取模块，用于提取各所述语音片段中存在的单人声语音，以作为第一单人声语音，并提取各所述语音片段中剩余的第二多人声语音，其中，所述单人声语音为只包含一种人声的语音；

[0117] 获取模块，还用于将所述第一单人声语音的频率作为筛选频率，获取所述第二多人声语音中语音频率与所述筛选频率相同的第二单人声语音；

[0118] 获取模块，还用于剔除所述第二多人声语音中的所述第二单人声语音得到第三多人声语音；

[0119] 获取模块，还用于使用预设二分算法在预设初始频率区间内将所述第三多人声语音进行分离，得到第三单人声语音；

[0120] 存储模块，用于将所述第一单人声语音、所述第二单人声语音和所述第三单人声语音存入语音数据库中。

[0121] 本发明还提供一种终端，所述终端包括：存储器、处理器以及存储在所述存储器上并可在所述处理器上运行的数据建库的人声分离程序，所述数据建库的人声分离程序被所述处理器执行时实现上述数据建库的人声分离方法各实施例的步骤。

[0122] 本发明还提供一种计算机可读存储介质,所述计算机可读存储介质上存储有数据建库的人声分离程序,所述数据建库的人声分离程序被处理器执行时实现上述数据建库的人声分离方法各实施例的步骤。

[0123] 在本发明数据建库的人声分离装置、终端和计算机可读存储介质的实施例中,包含了上述数据建库的人声分离方法各实施例的全部技术特征,说明书拓展和解释内容与上述数据建库的人声分离方法各实施例基本相同,在此不做累述。

[0124] 需要说明的是,在本文中,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、物品或者系统不仅包括那些要素,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、物品或者系统所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者系统中还存在另外的相同要素。

[0125] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0126] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在如上所述的一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,空调器,或者网络设备等)执行本发明各个实施例所述的方法。

[0127] 以上仅为本发明的优选实施例,并非因此限制本发明的专利范围,凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本发明的专利保护范围内。

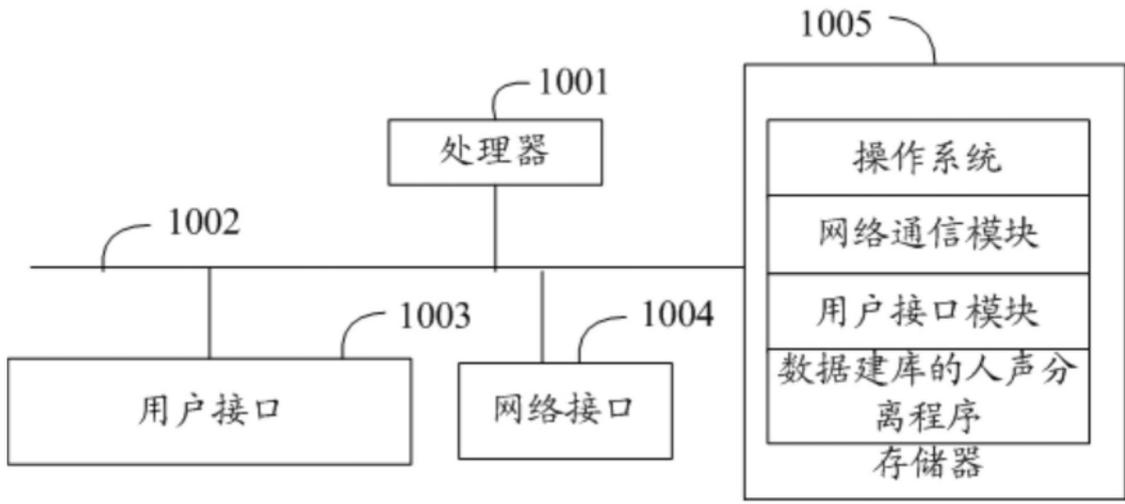


图1

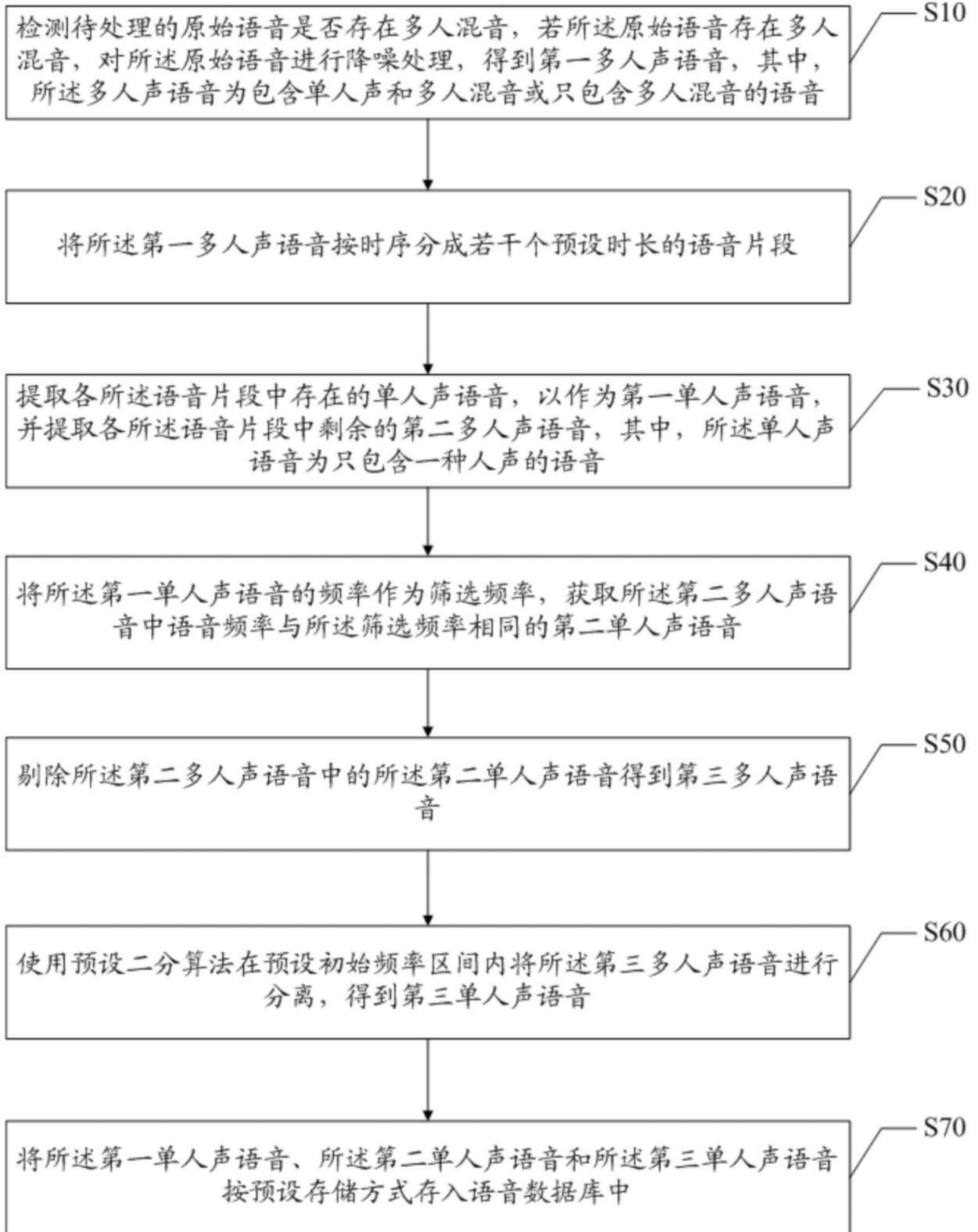


图2



图3

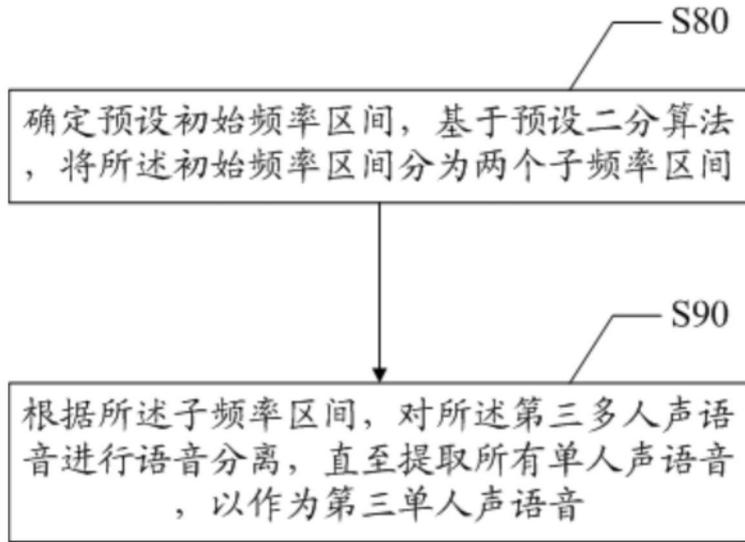


图4

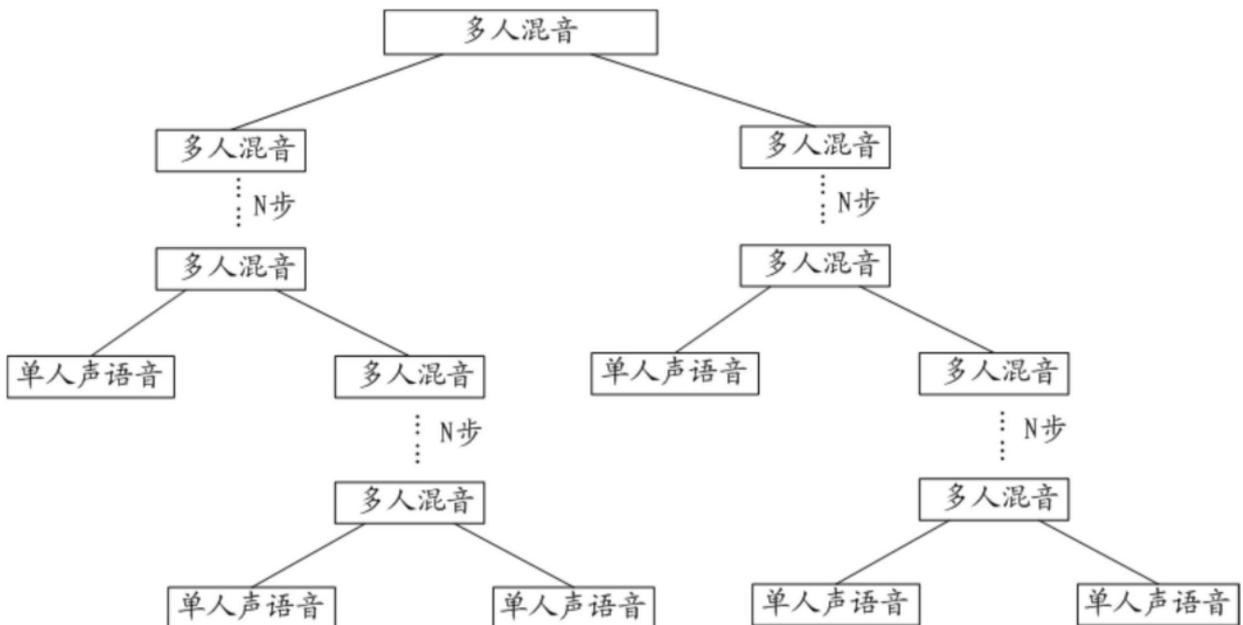


图5

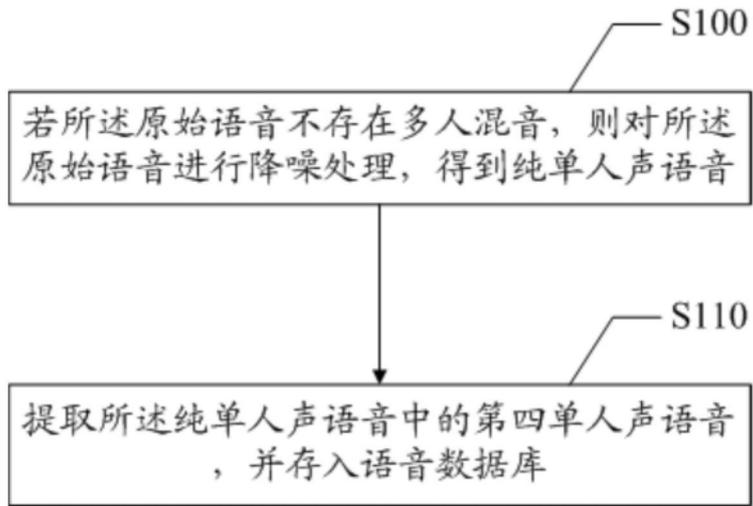


图6