(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 109074803 A
(43)申请公布日 2018.12.21

(21)申请号 201780029259.0

(22)申请日 2017.12.04

(66)本国优先权数据
201710170345.5 2017.03.21 CN

(85)PCT国际申请进入国家阶段日
2018.11.15

(86)PCT国际申请的申请数据
PCT/CN2017/114415 2017.12.04

(87)PCT国际申请的公布数据
WO2018/171257 EN 2018.09.27

(71)申请人 北京嘀嘀无限科技发展有限公司
地址 100193 北京市海淀区东北旺路西路8号院34号楼

(72)发明人 贺利强 李晓辉 万广鲁

(74)专利代理机构 成都七星天知识产权代理有限公司 51253
代理人 袁春晓

(51)Int.Cl.
G10L 13/033(2013.01)
G10L 13/08(2013.01)

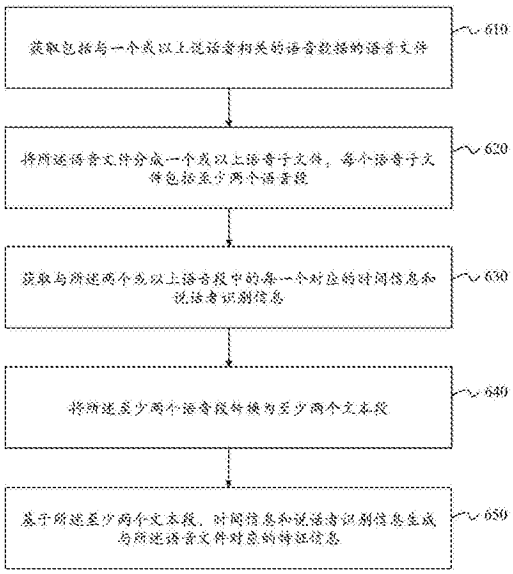
权利要求书5页 说明书22页 附图10页

(54)发明名称

语音信息处理系统和方法

(57)摘要

提供了一种使用语音识别方法生成用户行为的系统和方法。所述方法可以包括:获取包括与一个或以上说话者相关的语音数据的音频文件(610),以及将所述音频文件分成一个或以上音频子文件,每个音频子文件包括至少两个语音段(620)。所述一个或以上音频子文件中的每一个可以与所述一个或以上说话者中的一个对应。所述方法可以进一步包括:获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息(630),并将所述至少两个语音段转换为至少两个文本段(640)。所述至少两个语音段中的每一个可以与所述至少两个文本段中的一个对应。所述方法可以进一步包括:基于所述至少两个文本段、时间信息和说话者识别信息生成第一特征信息(650)。



CN 109074803 A

1.一种语音识别系统,包括:

至少一个存储设备,存储用于语音识别的一组指令;以及

与所述至少一个存储设备通信的至少一个处理器,其中,当执行所述一组指令时,所述至少一个处理器用于:

获取包括与一个或以上说话者相关的语音数据的音频文件;

将所述音频文件分成一个或以上音频子文件,每个所述音频子文件包括至少两个语音段,其中,所述一个或以上音频子文件中的每一个与所述一个或以上说话者中的一个对应;

获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息;

将所述至少两个语音段转换为至少两个文本段,其中,所述至少两个语音段中的每一个与所述至少两个文本段中的一个对应;以及

基于所述至少两个文本段、所述时间信息和所述说话者识别信息生成第一特征信息。

2.根据权利要求1所述的系统,其特征在于,将一个或以上麦克风安装在至少一个车厢中。

3.根据权利要求1所述的系统,其特征在于,从单通道获取所述音频文件,以及为了将所述音频文件分成一个或以上音频子文件,逻辑电路用于执行语音分离,所述语音分离包括计算听觉场景分析或盲源分离中的至少一个。

4.根据权利要求1所述的系统,其特征在于,与所述至少两个语音段中的每一个对应的时间信息包括所述语音段的起始时间和持续时间。

5.根据权利要求1所述的系统,其特征在于,所述至少一个处理器进一步用于:

获取初始模型;

获取一个或以上用户行为,每个用户行为与所述一个或以上说话者中的一个对应;以及

通过基于所述一个或以上用户行为和所述生成的第一特征信息训练所述初始模型来生成用户行为模型。

6.根据权利要求5所述的系统,其特征在于,所述至少一个处理器进一步用于:

获取第二特征信息;以及

基于所述第二特征信息执行所述用户行为模型以生成一个或以上用户行为。

7.根据权利要求1所述的系统,其特征在于,所述至少一个处理器用于:

在将所述音频文件分成一个或以上音频子文件之前,去除所述音频文件中的噪音。

8.根据权利要求1所述的系统,其特征在于,所述至少一个处理器用于:

在将所述音频文件分成一个或以上音频子文件之后,去除所述一个或以上音频子文件中的噪音。

9.根据权利要求1所述的系统,其特征在于,所述至少一个处理器进一步用于:

在将所述至少两个语音段中的每一个转换为文本段之后,将所述至少两个文本段中的每一个切分为词语。

10.根据权利要求1所述的系统,其特征在于,为了基于所述至少两个文本段、所述时间信息和所述说话者识别信息生成所述第一特征信息,所述至少一个处理器用于:

基于所述文本段的时间信息对所述至少两个文本段进行排序;以及

通过用所述相应的说话者识别信息标记每个所述排序的文本段来生成所述第一特征

信息。

11.根据权利要求1所述的系统，其特征在于，所述至少一个处理器进一步用于：

获取所述一个或以上说话者的位置信息；以及

基于所述至少两个文本段、所述时间信息、所述说话者识别信息和所述位置信息生成所述第一特征信息。

12.一种在计算设备上实现的方法，所述计算设备具有存储用于语音识别的一组指令的至少一个存储设备，以及与所述至少一个存储设备通信的至少一个处理器，所述方法包括：

获取包括与一个或以上说话者相关的语音数据的音频文件；

将所述音频文件分成一个或以上音频子文件，每个所述音频子文件包括至少两个语音段，其中，所述一个或以上音频子文件中的每一个与所述一个或以上说话者中的一个对应；

获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息；

将所述至少两个语音段转换为至少两个文本段，其中，所述至少两个语音段中的每一个与所述至少两个文本段中的一个对应；以及

基于所述至少两个文本段、所述时间信息和所述说话者识别信息生成第一特征信息。

13.根据权利要求12所述的方法，其特征在于，将一个或以上麦克风安装在至少一个车厢中，所述方法还包括：

获取所述至少一个车厢的位置信息；以及

基于所述至少两个文本段、所述时间信息、所述说话者识别信息和所述至少一个车厢的位置信息生成所述第一特征信息。

14.根据权利要求12所述的方法，其特征在于，从单通道获取所述音频文件，以及将所述音频文件分成一个或以上音频子文件进一步包括执行语音分离，所述语音分离包括计算听觉场景分析或盲源分离。

15.根据权利要求12所述的方法，其特征在于，与所述至少两个语音段中每一个对应的所述时间信息包括所述语音段的起始时间和持续时间。

16.根据权利要求12所述的方法，进一步包括：

获取初始模型；

获取一个或以上用户行为，每个用户行为与所述一个或以上说话者中的一个对应；以及

通过基于所述一个或以上用户行为和所述生成的第一特征信息训练所述初始模型来生成用户行为模型。

17.根据权利要求16所述的方法，进一步包括：

获取第二特征信息；以及

基于所述第二特征信息执行所述用户行为模型以生成一个或以上用户行为。

18.根据权利要求12所述的方法，进一步包括：

在将所述音频文件分成一个或以上音频子文件之前，去除所述音频文件中的噪音。

19.根据权利要求12所述的方法，进一步包括：

在将所述音频文件分成一个或以上音频子文件之后，去除所述一个或以上音频子文件中的噪音。

20.根据权利要求12所述的方法，进一步包括：

在将所述至少两个语音段中的每一个转换为文本段之后，将所述至少两个文本段中的每一个切分为词语。

21.根据权利要求12所述的方法，其特征在于，基于所述至少两个文本段、所述时间信息和所述说话者识别信息生成所述第一特征信息进一步包括：

基于所述文本段的时间信息对所述至少两个文本段进行排序；以及

通过用所述相应的说话者识别信息标记每个所述排序的文本段来生成所述第一特征信息。

22.根据权利要求12所述的方法，进一步包括：

获取所述一个或以上说话者的位置信息；以及

基于所述至少两个文本段、所述时间信息、所述说话者识别信息和所述位置信息生成所述第一特征信息。

23.一种非暂时性计算机可读介质，包括用于语音识别的至少一组指令，其中，当由电子终端的至少一个处理器执行时，所述至少一组指令指示所述至少一个处理器执行以下动作：

获取包括与一个或以上说话者相关的语音数据的音频文件；

将所述音频文件分成一个或以上音频子文件，每个所述音频子文件包括至少两个语音段，其中，所述一个或以上音频子文件中的每一个与所述一个或以上说话者中的一个对应；

获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息；

将所述至少两个语音段转换为至少两个文本段，其中，所述至少两个语音段中的每一个与所述至少两个文本段中的一个对应；以及

基于所述至少两个文本段、所述时间信息和所述说话者识别信息生成第一特征信息。

24.一种在计算设备上实现的系统，所述计算设备具有存储用于语音识别的一组指令的至少一个存储设备，以及与所述至少一个存储设备通信的至少一个处理器，所述系统包括：

音频文件获取模块，用于获取包括与一个或以上说话者相关的语音数据的音频文件；

音频文件分离模块，用于将所述音频文件分成一个或以上音频子文件，每个所述音频子文件包括至少两个语音段，其中，所述一个或以上音频子文件中的每一个与一个或以上说话者中的一个对应；

信息获取模块，用于获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息；

语音转换模块，用于将所述至少两个语音段转换为至少两个文本段，其中，所述至少两个语音段中的每一个所述至少两个文本段中的一个对应；以及

特征信息生成模块，用于基于所述至少两个文本段、所述时间信息和所述说话者识别信息生成第一特征信息。

25.一种语音识别系统，包括：

一个总线；

连接到所述总线的至少一个输入端口；

连接到所述输入端口的一个或以上麦克风，所述一个或以上麦克风中的每一个用于检

测来自一个或以上说话者中至少一个的语音，并生成相应说话者的语音数据到所述输入端口；

连接到所述总线的至少一个存储设备，存储用于语音识别的一组指令；以及

与所述至少一个存储设备通信的逻辑电路，其中，当执行所述一组指令时，所述逻辑电路用于：

获取包括与一个或以上说话者相关的语音数据的音频文件；

将所述音频文件分成一个或以上音频子文件，每个所述音频子文件包括至少两个语音段，其中，所述一个或以上音频子文件中的每一个与所述一个或以上说话者中的一个对应；

获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息；

将所述至少两个语音段转换为至少两个文本段，其中，所述至少两个语音段中的每一个与所述至少两个文本段中的一个对应；以及

基于所述至少两个文本段、所述时间信息和所述说话者识别信息生成第一特征信息。

26.根据权利要求25所述的系统，其特征在于，将所述一个或以上麦克风安装在至少一个车厢中。

27.根据权利要求25所述的系统，其特征在于，从单通道获取所述音频文件，以及为了将所述音频文件分成一个或以上音频子文件，所述逻辑电路用于执行语音分离，所述语音分离包括计算听觉场景分析或盲源分离中的至少一个。

28.根据权利要求25所述的系统，其特征在于，与所述至少两个语音段中的每一个对应的时间信息包括所述语音段的起始时间和持续时间。

29.根据权利要求25所述的系统，其特征在于，所述逻辑电路进一步用于：

获取初始模型；

获取一个或以上用户行为，每个用户行为与所述一个或以上说话者中的一个对应；以及

通过基于所述一个或以上用户行为和所述生成的第一特征信息训练所述初始模型来生成用户行为模型。

30.根据权利要求29所述的系统，其特征在于，所述逻辑电路进一步用于：

获取第二特征信息；以及

基于所述第二特征信息执行所述用户行为模型以生成一个或以上用户行为。

31.根据权利要求25所述的系统，其特征在于，所述逻辑电路用于：

在将所述音频文件分成一个或以上音频子文件之前，去除所述音频文件中的噪音。

32.根据权利要求25所述的系统，其特征在于，所述逻辑电路用于：

在将所述音频文件分成一个或以上音频子文件之后，去除所述一个或以上音频子文件中的噪音。

33.根据权利要求25所述的系统，其特征在于，所述逻辑电路进一步用于：

在将所述至少两个语音段中的每一个转换为文本段之后，将所述至少两个文本段中的每一个切分为词语。

34.根据权利要求25所述的系统，其特征在于，为了基于所述至少两个文本段、所述时间信息和所述说话者识别信息生成所述第一特征信息，所述逻辑电路用于：

基于所述文本段的时间信息对所述至少两个文本段进行排序；以及

通过用所述相应的说话者识别信息标记每个所述排序的文本段来生成所述第一特征信息。

35.根据权利要求25所述的系统，其特征在于，所述逻辑电路进一步用于：

获取所述一个或以上说话者的位置信息；以及

基于所述至少两个文本段、所述时间信息、所述说话者识别信息和所述位置信息生成所述第一特征信息。

语音信息处理系统和方法

交叉引用

[0001] 本申请要求于2017年3月21日提交的中国专利申请No.201710170345.5的优先权，其全部内容通过引用并入本文。

技术领域

[0002] 本申请涉及语音信息处理，尤其涉及使用语音识别方法处理语音信息以生成用户行为的方法和系统。

背景技术

[0003] 语音信息处理(例如，语音识别方法)已广泛用于日常生活中。对于在线按需服务，用户可以通过将语音信息输入电子设备(例如，移动电话)来简单地提出他/她的请求。例如，用户(例如，乘客)可以通过他/她的终端(例如，移动电话)的麦克风以语音数据的形式提出服务请求。相应地，另一个用户(例如，司机)可以通过他/她的终端(例如，移动电话)的麦克风以语音数据的形式回复该服务请求。在一些实施例中，与说话者相关的语音数据可以反映说话者的行为，可以用于生成用户行为模型，该用户行为模型可以架起语音文件和与该语音文件中用户对应的用户行为之间的连接。但是，机器或计算机可能无法直接理解语音数据。因此，期望提供一种生成适合训练用户行为模型的特征信息的新语音信息处理方法。

发明内容

[0004] 本申请的一个方面提供了一种语音识别系统。所述语音识别系统可以包括总线、连接到总线的至少一个输入端口、连接到输入端口的一个或以上麦克风、连接到总线的至少一个存储设备以及与所述至少一个存储设备通信的逻辑电路。所述一个或以上麦克风中的每一个可以用于检测来自所述一个或以上说话者中至少一个的语音，并生成相应说话者的语音数据到输入端口。所述至少一个存储设备可以存储用于语音识别的一组指令。当执行所述一组指令时，逻辑电路可以用于获取包括与所述一个或以上说话者相关的语音数据的音频文件，并将所述音频文件分成一个或以上音频子文件，每个音频子文件包括至少两个语音段。所述一个或以上音频子文件中的每一个可以与所述一个或以上说话者中的一个对应。逻辑电路还可以用于获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息，并将所述至少两个语音段转换为至少两个文本段。所述至少两个语音段中的每一个可以与所述至少两个文本段中的一个对应。逻辑电路还可以用于基于所述至少两个文本段、时间信息和说话者识别信息生成第一特征信息。

[0005] 在一些实施例中，可以将所述一个或以上麦克风安装在至少一个车厢中。

[0006] 在一些实施例中，可以从单通道获取所述音频文件，以及为了将所述音频文件分成一个或以上音频子文件，所述逻辑电路可以用于执行语音分离，所述语音分离包括计算听觉场景分析或盲源分离中的至少一个。

[0007] 在一些实施例中，与所述至少两个语音段中的每一个对应的时间信息可以包括所述语音段的起始时间和持续时间。

[0008] 在一些实施例中，逻辑电路可以进一步用于获取初始模型、获取一个或以上用户行为，每个用户行为与所述一个或以上说话者中的一个对应，并基于所述一个或以上用户行为和所述生成的第一特征信息训练所述初始模型来生成用户行为模型。

[0009] 在一些实施例中，逻辑电路还可以用于获取第二特征信息，并基于所述第二特征信息执行所述用户行为模型以生成一个或以上用户行为。

[0010] 在一些实施例中，逻辑电路还可以用于在将音频文件分成一个或以上音频子文件之前，去除所述音频文件中的噪音。

[0011] 在一些实施例中，逻辑电路还可以用于在将音频文件分成一个或以上音频子文件之后，去除所述一个或以上音频子文件中的噪音。

[0012] 在一些实施例中，逻辑电路还可以用于在将至少两个语音段中的每一个转换为文本段之后，将所述至少两个文本段中的每一个切分为词语。

[0013] 在一些实施例中，为了基于所述至少两个文本段、时间信息和说话者识别信息生成第一特征信息，逻辑电路可以用于基于所述文本段的时间信息对至少两个文本段进行排序，以及通过用所述相应的说话者识别信息标记每个所述排序的文本段来生成所述第一特征信息。

[0014] 在一些实施例中，逻辑电路还可以用于获取一个或以上说话者的位置信息，并基于所述至少两个文本段、时间信息、说话者识别信息和所述位置信息生成第一特征信息。

[0015] 本申请的另一方面提供了一种方法。所述方法可以在计算设备上实现，该计算设备具有至少一个存储设备，该存储设备存储用于语音识别的一组指令，以及与所述至少一个存储设备通信的逻辑电路。所述方法可以包括获取包括与一个或以上说话者相关的语音数据的音频文件，以及将该音频文件分成一个或以上音频子文件，每个音频子文件包括至少两个语音段。所述一个或以上音频子文件中的每一个可以与所述一个或以上说话者中的一个对应。所述方法还可以包括获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息，并将所述至少两个语音段转换为至少两个文本段。所述至少两个语音段中的每一个可以与所述至少两个文本段中的一个对应。所述方法还可以包括基于所述至少两个文本段、时间信息和说话者识别信息第一特征信息。

[0016] 本申请的另一方面提供了一种非暂时性计算机可读介质。所述非暂时性计算机可读介质可以包括用于语音识别的至少一组指令。当由电子终端的逻辑电路执行时，所述至少一组指令可以指示逻辑电路执行获取包括与一个或以上说话者相关的语音数据的音频文件，并将所述音频文件分成一个或以上音频子文件，以及每个所述音频子文件包括至少两个语音段的动作。所述一个或以上音频子文件中的每一个可以与所述一个或以上说话者中的一个对应。所述至少一组指令还可以指示逻辑电路执行获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息，以及将所述至少两个语音段转换为至少两个文本段的动作。所述至少两个语音段中的每一个可以与所述至少两个文本段中的一个对应。所述至少一组指令还可以指示逻辑电路执行基于所述至少两个文本段、时间信息和说话者识别信息生成第一特征信息的动作。

[0017] 本申请的另一方面提供了一种系统。所述系统可以在计算设备上实现，该计算设

备具有至少一个存储设备，该存储设备存储用于语音识别的一组指令，以及与所述至少一个存储设备通信的逻辑电路。所述系统可以包括音频文件获取模块、音频文件分离模块、信息获取模块、语音转换模块和特征信息生成模块。音频文件获取模块可以用于获取包括与一个或以上说话者相关的语音数据的音频文件。信息获取模块可以用于将所述音频文件分成一个或以上音频子文件，每个所述音频子文件包括至少两个语音段。所述一个或以上音频子文件中的每一个可以与所述一个或以上说话者中的一个对应。信息获取模块可以用于获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息。语音转换模块可以用于将所述至少两个语音段转换为至少两个文本段。所述至少两个语音段中的每一个可以与所述至少两个文本段中的一个对应。特征信息生成模块可以用于基于所述至少两个文本段、时间信息和说话者识别信息生成第一特征信息。

[0018] 本申请的一部分附加特性可以在下面的描述中进行说明。通过对以下描述和相应附图的研究或者对实施例的生产或操作的了解，本申请的一部分附加特性对于本领域技术人员是明显的。本申请的特征可以通过对以下描述的具体实施例的各种方面的方法、手段和组合的实践或使用得以实现和达到。

附图说明

[0019] 本申请将通过示例性实施例进一步描述。这些示例性实施例将参考附图详细描述。附图不是按比例绘制的。这些实施例并非限制性的，在这些实施例中相同的附图标记在附图的若干视图中表示类似的结构，其中：

[0020] 图1是根据本申请的一些实施例所示的按需服务系统的示例性框图；

[0021] 图2是根据本申请的一些实施例所示的计算设备的示例性硬件和/或软件组件的示意图；

[0022] 图3是根据本申请的一些实施例所示的示例性设备的示意图；

[0023] 图4是根据本申请的一些实施例所示的示例性处理引擎的框图；

[0024] 图5是根据本申请的一些实施例所示的音频文件分离模块的示例性框图；

[0025] 图6是根据本申请的一些实施例所示的用于生成语音文件对应的特征信息的示例性过程的流程图；

[0026] 图7是根据本申请的一些实施例所示的与双通道语音文件相对应的示例性特征信息的示意图；

[0027] 图8是根据本申请的一些实施例所示的用于生成与语音文件相对应的特征信息的示例性过程的流程图；

[0028] 图9是根据本申请的一些实施例所示的用于与生成语音文件相对应的特征信息的示例性过程的流程图；

[0029] 图10是根据本申请的一些实施例所示的用于生成用户行为模型的示例性过程的流程图；以及

[0030] 图11是根据本申请的一些实施例所示的用于执行用户行为模型以生成用户行为的示例性过程的流程图。

具体实施方式

[0031] 以下描述是为了使本领域的普通技术人员能够实施和利用本申请，并且该描述是在特定的应用场景及其要求的环境下提供的。对于本领域的普通技术人员来讲，显然可以对所公开的实施例作出各种改变，并且在不偏离本申请的原则和范围的情况下，本申请中所定义的普遍原则可以适用于其他实施例和应用场景。因此，本申请并不限于所描述的实施例，而应该被给予与权利要求一致的最广泛的范围。

[0032] 这里使用的术语仅用于描述特定示例实施例的目的，而不是限制性的。如这里所使用的，单数形式“一”、“一个”和“该”也可以包括复数形式，除非上下文另有明确说明。还应当理解，如在本说明书中所示，术语“包括”、“包含”仅提示存在所述特征、整体、步骤、操作、元件和/或部件，但并不排除存在或添加一个或以上其他特征、整体、步骤、操作、元件、部件和/或其组合的情况本申请中所使用的术语仅用于描述特定的示例性实施例，并不限制本申请的范围。

[0033] 根据以下对附图的描述，本申请的这些和其他的特征、特点以及相关结构元件的功能和操作方法，以及部件组合和制造经济性，可以变得更加显而易见，这些附图都构成本申请说明书的一部分。然而，应当理解的是，附图仅仅是为了说明和描述的目的，并不旨在限制本申请的范围。应当理解的是，附图并不是按比例绘制的。

[0034] 本申请中使用了流程图用来说明根据本申请的一些实施例的系统所执行的操作。应当理解的是，流程图中的操作可以不按顺序执行。相反，可以按照倒序或同时处理各种步骤。此外，可以向流程图添加一个或多个其他操作。也可以从流程图中删除一个或多个操作。

[0035] 此外，虽然本申请中公开的系统和方法主要涉及评估用户终端，但是还应当理解的是，这仅是一个示例性实施例。本申请的系统和方法可以应用于任何其他类型的按需服务平台的用户。本申请的系统或方法可以应用于不同环境的路径规划系统，包括陆地、海洋、航空航天等或其任任意组合。运输系统涉及的车辆可以包括出租车、私家车、挂车、公共汽车、火车、动车、高铁、地铁、船舶、飞机、宇宙飞船、热气球、无人驾驶车辆等或其任意组合。运输系统还可以包括用于管理及/或分配的任何运输系统，例如，用于发送和/或接收快递的系统。本申请的系统和方法的应用场景还可以包括网页、浏览器插件、客户端、客户系统、内部分析系统、人工智能机器人等或其任意组合。

[0036] 可以通过嵌入在无线设备(例如，客运终端、司机终端等)中的定位技术来获取本申请中的服务起点。本申请中使用的定位技术可以包括全球定位系统(GPS)、全球导航卫星系统(GLONASS)、罗盘导航系统(COMPASS)、伽利略定位系统、准天顶卫星系统(QZSS)、无线保真(WiFi)定位技术等或其任意组合。一种或多种上述定位技术可以在本申请中互换使用。例如，基于GPS的方法和基于WiFi的方法可以一起用作定位技术以定位无线设备。

[0037] 本申请一方面涉及语音信息处理的系统和/或方法。语音信息处理可以指生成与语音文件相对应的特征信息。例如，语音文件可以由车载记录系统记录。语音文件可以是与乘客和司机之间的对话有关的双通道语音文件。可以将语音文件分为两个语音子文件，子文件A和子文件B。子文件A可以对应于乘客，子文件B可以对应于司机。对于至少两个语音段中的每一个，可以获取与该语音段对应的时间信息和说话者识别信息。时间信息可以包括起始时间和/或持续时间(或结束时间)。可以将至少两个语音段转换为至少两个文本段。然后，可以基于至少两个文本段、时间信息和说话者识别信息生成与双通道语音文件相对应

的特征信息。生成的特征信息可以进一步用于训练用户行为模型。

[0038] 应当注意的是，本解决方案依赖于收集在线系统注册的用户终端的使用数据(例如，语音数据)，是一种仅在后互联网时代扎根的新的数据收集装置形式。它提供了仅在后互联网时代才能提出的用户终端的详细信息。在前互联网时代，不可能收集例如与旅行路线、出发地点、目的地等相关联的语音数据的用户终端的信息。然而，在线按需服务允许在线平台通过分析与司机和乘客相关的语音数据实时和/或基本实时地监测成千上万的用户终端的行为，然后基于用户终端的行为和/或语音数据提供更好的服务方案。因此，本解决方案深入并旨在解决仅在后互联网时代发生的问题。

[0039] 图1是根据本申请的一些实施例所示的按需服务系统的示例性框图。例如，按需服务系统100可以是用于运输服务的在线运输服务平台，例如呼叫出租车服务、专车服务、快车服务、拼车服务、公交服务、代驾和班车服务。按需服务系统100可以包括服务器110、网络120、乘客终端130、司机终端140和存储器150。服务器110可以包括处理引擎112。

[0040] 服务器110可以用于处理与服务请求有关的信息和/或数据。例如，服务器110可以基于语音文件确定特征信息。在一些实施例中，所述服务器110可以是单个服务器，或服务器组。所述服务器组可以是集中式的或分布式的(例如服务器110可以是分布式系统)。在一些实施例中，服务器110可以是本地的或远程的。例如，服务器110可以经由网络120访问存储在乘客终端130、司机终端140和/或存储器150中的信息和/或数据。又例如，服务器110可以直接与乘客终端130、司机终端140和/或存储器150连接以访问存储的信息和/或数据。在一些实施例中，服务器110可以在云平台上实施。仅作为示例，云平台可以包括私有云、公共云、混合云、社区云、分布云、内部云、多层云等或其任意组合。在一些实施例中，服务器110可以在图2中所示的有一个或以上组件的计算设备200上实现。

[0041] 在一些实施例中，服务器110可以包括处理引擎112。处理引擎112可以处理与服务请求有关的信息和/或数据，以执行本申请中描述的服务器110的一个或以上功能。例如，处理引擎112可以获取音频文件。音频文件可以是语音文件(也称为第一语音文件)，其包括与司机和乘客相关的语音数据(例如，他们之间的对话)。处理引擎112可以从乘客终端130和/或司机终端140获取语音文件。又例如，处理引擎112可以被配置为用于确定与语音文件相对应的特征信息。生成的特征信息可以用于训练用户行为模型。然后，处理引擎112可以将新的语音文件(也称为第二语音文件)输入到训练好的用户行为模型中，并生成与新的语音文件中的说话者相对应的用户行为。在一些实施例中，处理引擎112可以包括一个或以上处理引擎(例如，单核处理器或是多核处理器)。仅作为示例，处理引擎112可包括中央处理器(CPU)、特定应用集成电(ASIC)、特定应用指令集处理器(ASIP)、图形处理器(GPU)、物理运算处理单元(PPU)、数字信号处理器(DSP)、现场可以程序门阵列(FPGA)、可以程序逻辑装置(PLD)、控制器、微控制器单元、精简指令集计算机(RISC)、微处理器等或其任意组合。

[0042] 网络120可以促进信息和/或数据的交换。在一些实施例中，按需服务系统100的一个或以上组件(例如，服务器110、乘客终端130、司机终端140或存储器150)可以经由网络120将信息和/或数据发送到按需服务系统100的其他组件。例如，服务器110可以经由网络120从乘客终端130获得服务请求。在一些实施例中，网络120可以为任意形式的有线或无线网络，或其任意组合。仅作为示例，网络120可以包括电缆网络、有线网络、光纤网络、电信网络，内部网络、互联网、局域网络(LAN)、广域网络(WAN)、无线局域网络(WLAN)、城域网

(MAN)，公共开关电话网络(PSTN)、蓝牙网络、ZigBee网络、近场通信(NFC)网络等，或其任意组合。在一些实施例中，网络120可以包括一个或多个网络接入点。例如，网络120可以包括有线或无线网络接入点，如基站和/或互联网交换点120-1、120-2、……，通过该网络交换点，按需服务系统100的一个或以上部件可以连接到网络120以交换数据和/或信息。

[0043] 乘客可以使用乘客终端130来请求按需服务。例如，乘客终端130的用户可以使用乘客终端130为他/她自己或另一用户的发送服务请求，或者从服务器110接收服务和/或信息或指令。司机可以使用司机终端140回复按需服务。例如，司机终端140的用户可以使用司机终端140接收来自乘客终端130的服务请求和/或来自服务器110的信息或指令。在一些实施例中，术语“用户”和“乘客终端”可以互换使用，术语“用户”和“司机终端”可以互换使用。在一些实施例中，用户(例如，乘客)可以通过他/她的终端(例如，乘客终端130)的麦克风以语音数据的形式发起服务请求。相应地，另一个用户(例如，司机)可以通过他/她的终端(例如，司机终端140)的麦克风以语音数据的形式回复服务请求。司机(或乘客)的麦克风可以与他/她的终端的输入端口连接。

[0044] 在一些实施例中，乘客终端130乘客终端130可以包括移动设备130-1、平板计算机130-2、膝上型计算机130-3、车载设备130-4等或其任意组合。在一些实施例中，移动设备130-1可以包括智能家居设备、可穿戴设备、智能移动设备、虚拟现实设备、增强现实设备等，或其任意组合。在一些实施例中，智能家居设备可以包括智能照明设备、智能电器控制设备、智能监控设备、智能电视、智能摄像机、对讲机等，或其任意组合。在一些实施例中，该可穿戴设备可包括智能手镯、智能鞋袜、智能眼镜、智能头盔、智能手表、智能衣服、智能背包、智能配件等或其任意组合。在一些实施例中，智能移动设备可以包括智能电话、个人数字助理(PDA)、游戏设备、导航设备、销售点(POS)等，或其任意组合。在一些实施例中，虚拟现实设备和/或增强型虚拟现实设备可以包括虚拟现实头盔、虚拟现实眼镜、虚拟现实眼罩、增强现实头盔、增强现实眼镜、增强现实眼罩等，或其任意组合。例如，虚拟现实设备和/或增强现实设备可以包括Google Glass、Oculus Rift、Hololens或Gear VR等。在一些实施例中，车载设备130-4可以包括车载计算机、车载电视等。在一些实施例中，乘客终端130可以是具有定位技术的设备，用于定位乘客终端130的用户(例如，司机)位置。

[0045] 在一些实施例中，司机终端140可以是与乘客终端130类似或相同的设备。在一些实施例中，司机终端140可以是具有用于定位服务提供者和/或司机终端140的位置的定位技术的设备。在一些实施例中，乘客终端130和/或司机终端140可以与其他定位设备通信以确定服务请求者、乘客终端130、服务提供者和/或司机终端140的位置。在一些实施例中，乘客终端130和/或司机终端140可以将定位信息发送到服务器110。

[0046] 存储器150可以存储数据和/或指令。在一些实施例中，存储器150可以存储从乘客终端130和/或司机终端140获得的数据。在一些实施例中，存储器150可以储存服务器110用来执行或使用来完成本申请中描述的示例性方法的数据及/或指令。在一些实施例中，存储器150可包括大容量储存器、可移动储存器、易失性读写储存器、只读存储器(ROM)等或其任意组合。示例性的大容量储存器可以包括磁盘、光盘、固态磁盘等。示例性可移动存储器可以包括闪存驱动器、软盘、光盘、存储卡、压缩盘、磁带等。示例性易失性读写存储器可以包括随机存取内存(RAM)。示例性RAM可包括动态随机存取存储器(DRAM)、双倍数据速率同步动态随机存取存储器(DDR SDRAM)、静态随机存取存储器(SRAM)、晶闸管随机存取存储器

(T-RAM)和零电容随机存取存储器(Z-RAM)等。示例性只读存储器可以包括掩模型只读存储器(MROM)、可编程只读存储器(PROM)、可擦除可编程只读存储器(PEROM)、电可擦除可编程只读存储器(EEPROM)、光盘只读存储器(CD-ROM)和数字多功能磁盘只读存储器等。在一些实施例中,存储器150可以在云平台上实现。仅作为示例,云平台可以包括私有云、公共云、混合云、社区云、分布云、内部云、多层云等,或其任意组合。

[0047] 在一些实施例中,存储器150可以连接到网络120以与按需服务系统100的一个或以上组件(例如,服务器110、乘客终端130、司机终端140)通信。按需服务系统100的一个或以上组件可以经由网络120访问存储在存储器150中的数据和/或指令。在一些实施例中,存储器150可以直接连接到按需服务系统100的一个或以上组件(例如,服务器110、乘客终端130、司机终端140)或与之通信。在一些实施例中,存储器150可以是服务器110的一部分。

[0048] 在一些实施例中,按需服务系统100的一个或以上组件(例如,服务器110、乘客终端130、司机终端140)可以具有访问存储器150的许可。在一些实施例中,当满足一个或以上条件时,按需服务系统100的一个或以上组件可以读取和/或修改与乘客、司机和/或公众有关的信息。例如,服务器110可以在服务完成之后读取和/或修改一个或以上用户的信息。作为另一示例,司机终端140可以在从乘客终端130接收服务请求时访问与乘客有关的信息,但是司机终端140不可以修改乘客的相关信息。

[0049] 在一些实施例中,可以通过请求服务来实现按需服务系统100的一个或以上组件的信息交换。服务请求的对象可以为任何产品。在一些实施方案中,产品可以是有形产品或非物质产品。有形产品可包括食品、药品、商品、化学产品、电器、服装、汽车、房屋、奢侈品等,或其任何组合。非物质产品可以包括服务产品、金融产品、知识产品、互联网产品等,或其任何组合。互联网产品可以包括个人主机产品、网站产品、移动互联网产品、商业主机产品、嵌入式产品等或上述举例的任意组合。移动互联网产品可以用于移动终端的软件、程序、系统等的软件或上述举例的任意组合。移动终端可以包括平板计算机、膝上型计算机、移动电话、个人数字助理(PDA)、智能手表、POS装置、车载计算机、车载电视、可穿戴装置等或其任意组合。例如,产品可以是在计算机或移动电话上使用的任何软件和/或应用。所述软件和/或应用程序可以与社交、购物、交通、娱乐、学习、投资等或上述举例的任意组合有关。在一些实施例中,所述与运输有关系统软件和/或应用程序可以包括出行软件和/或应用程序、车辆调度软件和/或应用程序、地图软件和/或应用程序等。在车辆调度软件和/或应用程序中,交通工具可以包括马、马车、人力车(例如独轮手推车、自行车、三轮车)、汽车(例如,出租车、公共汽车、私家车)、火车、地铁、船舶、飞行器(例如,飞机、直升机、航天飞机、火箭、热气球)等其任意组合。

[0050] 本领域普通技术人员应当理解,当执行按需服务系统100的元件(或组件)时,该元件可以通过电信号和/或电磁信号执行。例如,当乘客终端130处理诸如输入语音数据、识别或选择对象的任务时,乘客终端130可以操作在其处理器中的逻辑电路以执行这些任务。当乘客终端130向服务器110发送服务请求时,服务器110的处理器可以生成编码该请求的电信号。然后,服务器110的处理器可以将电信号发送到输出端口。如果乘客终端130经由有线网络与服务器110通信,则输出端口可以物理连接至某一电缆,其进一步将电信号传输给服务器110的输入端口。如果乘客终端130经由无线网络与服务器110通信,则乘客终端130的输出端口可以是一个或以上天线,其将电信号转换为电磁信号。类似地,司机终端140可以

通过其处理器中的逻辑电路的操作来处理任务，并且经由电信号或电磁信号从服务器110接收指令和/或服务请求。在诸如乘客终端130、司机终端140和/或服务器110的电子设备内，当其处理器处理指令、发出指令和/或执行动作时，该指令和/或动作通过电信号执行。例如，当处理器从存储介质(例如，存储器150)检索或保存数据时，它可以将电信号发送到存储介质的读/写设备，其可以在存储介质中读取或写入结构化数据。该结构化数据可以以电信号的形式经由电子装置的总线传输至处理器。如本申请所示的，电信号是指一个电信号、一系列电信号和/或至少两个不连续的电信号。

[0051] 图2是根据本申请的一些实施例所示的计算设备的示例性硬件和/或软件组件的示意图。在一些实施例中，服务器110和/或乘客终端130和/或司机终端140可以在计算设备200上实现。例如，处理引擎112可以在计算设备200上实施并执行本申请所公开的处理引擎112的功能。

[0052] 计算设备200可用于实现如本申请所述的按需服务系统的任何组件。为了方便起见，图中仅示出了一台计算机，本领域普通技术人员在提交本申请时将理解，本申请所描述的按需服务有关的计算机功能可以在多个类似平台上以分布式的方式实现，以分担处理负载。

[0053] 例如，计算设备200可以包括与网络相连接通信端口250，以促进数据通信。计算设备200还可以包括中央处理器220，可以以一个或以上处理器(例如，逻辑电路)的形式执行程序指令。示例性计算机平台可以包括内部通信总线210、不同类型的程序存储器和数据存储器(例如，硬盘270、只读存储器(ROM 230)、随机存储器(RAM)240)，以适用于计算机处理和/或通信的各种数据文件。示例性计算机平台还可以包括存储在ROM 230、RAM 240和/或其他类型的非暂时储存介质中的程序指令，以由处理器220执行。本申请的方法和/或流程可以以程序指令的方式实现。计算设备200还包括输入/输出组件(I/O)260，用来支持计算机和其他组件之间的输入/输出和电源280，用于为计算设备200或其元件提供电力。计算设备200也可以通过网络通信接收编程和数据。

[0054] 处理器220(例如，逻辑电路)可以执行计算机指令(例如，程序代码)以及根据本申请描述的技术执行处理引擎112的功能。例如，处理器220可以包括接口电路220-a和处理电路220-b。接口电路220-a可以被配置为从总线210接收电信号，其中，电信号编码用于处理电路的结构化数据和/或指令。处理电路220-b可以进行逻辑计算，然后确定结论、结果和/或编码为电子信号的指令。然后，接口电路220-a可以经由总线210从处理电路220-b发出电信号。在一些实施例中，一个或以上麦克风可以与输入/输出组件260或其输入端口(图2中未示出)连接。所述一个或以上麦克风中的每一个被配置为检测来自一个或以上说话者中至少一个的语音，并生成对应说话者的语音数据到输入/输出组件260或其输入端口。

[0055] 为了方便说明，在计算设备200中仅描述了一个处理器220。然而，应当注意的是计算设备200也可以包括至少两个处理器，因此本申请中描述的由一个处理器执行的操作和/或方法步骤也可以由多个处理器共同地或单独执行。例如，如果在本申请中，计算设备200的处理器执行步骤A和步骤B，应当理解的是，步骤A和步骤B也可以由计算设备200的两个不同的CPU和/或处理器共同地或独立地执行(例如，第一处理器执行步骤A、第二处理器执行步骤B，或者第一和第二处理器共同地执行步骤A和步骤B)。

[0056] 图3是根据本申请的一些实施例所示的移动设备的示例性硬件和/或软件组件的

示意图。乘客终端130或司机终端140可以在移动设备300上实现。该设备可以是移动设备，如乘客或司机的移动手机。如图3所示，移动设备300可以包括通信平台310、显示器320、图形处理单元(GPU)330、中央处理单元(CPU)340、输入/输出(I/O)350、内存360和存储器390。在一些实施例中，任何其他合适的组件，包括但不限于系统总线或控制器(未示出)，也可包括在移动设备300内。在一些实施例中，移动操作系统370(例如，iOS™、Android™、Windows Phone™等)和一个或以上应用程序380可从储存器390下载至内存360以便由CPU 340执行。应用程序380可以包括浏览器或任何其他合适的移动应用程序，用于从服务器110接收和呈现与线上按需服务相关的信息或其他信息，以及将与线上按需服务相关的信息或其他信息发送至服务器110。用户与信息流的交互可以经由输入/输出单元(I/O)350实现，并且经由网络120提供给处理引擎112和/或按需服务系统100的其他组件在一些实施例中，设备300可以包括用于捕获语音信息的设备，例如，麦克风315。

[0057] 图4是根据本申请的一些实施例所示的用于生成与语音文件相对应的特征信息的示例性处理引擎的框图。处理引擎112可以与存储器(例如，存储器150、乘客终端130或司机终端140)进行通信，并且可以执行存储在存储介质中的指令。在一些实施例中，处理引擎112可以包括音频文件获取模块410、音频文件分离模块420、信息获取模块430、语音转换模块440、特征信息生成模块450、模型训练模块460以及用户行为确定模块470。

[0058] 音频文件获取模块410可以用于获取音频文件。在一些实施例中，音频文件可以是包括与一个或以上说话者相关的语音数据的语音文件。在一些实施例中，可以将一个或以上麦克风安装在至少一个车厢中(例如，出租车、私家车、公共汽车、火车、动车、高铁、地铁、船只、飞行器、飞船、热气球、潜水艇)，以用于检测来自所述一个或以上说话者中的至少一个说话者的语音，并生成相应说话者的语音数据。例如，定位系统(例如，全球定位系统(GPS))可以在至少一个车厢或安装在其上的一个或以上麦克风上实现。定位系统可以获取车辆(或其中的说话者)的位置信息。位置信息可以是相对位置(例如，车辆或说话者彼此对应的相对方位和距离)或绝对位置(例如，纬度和经度)。又例如，可以将至少两个麦克风安装在每个车厢中，可以从量级上对由该至少两个麦克风记录的音频文件(或声音信号)进行相互整合和/或比较，以获取车厢中说话者的位置信息。

[0059] 在一些实施例中，可以将所述一个或以上麦克风安装在商店、道路或房屋中以检测来自一个或以上说话者中的语音并生成与所述一个或以上说话者相对应的语音数据。在一些实施例中，可以将所述一个或以上麦克风安装在车辆或车辆的配件(例如，摩托车头盔)上。一个或以上摩托车骑手可以通过安装在其头盔上的麦克风彼此进行交谈。该麦克风可以检测来自摩托车骑手的语音并生成该相应的摩托车骑手的语音数据。在一些实施例中，每个摩托车可以有司机和一个或以上乘客，每个乘客佩戴安装有麦克风的摩托车头盔。安装在每个摩托车头盔上的麦克风之间是连接的，安装在不同摩托车头盔上的麦克风之间也可以相互连接。可以手动(例如，通过按下按钮或设置参数)或者自动地(例如，当两个摩托车彼此靠近时通过自动建立蓝牙™连接)建立和终止头盔之间的连接。在一些实施例中，可以将所述一个或以上麦克风安装在特定位置以监测附近的声音(语音)。例如，可以将所述一个或以上麦克风安装在重建站点以监测建筑工人的重建噪音和声音。

[0060] 在一些实施例中，语音文件可以是多通道语音文件。可以从至少两个通道获取该多通道语音文件。所述至少两个通道中的每一个通道可以包括与一个或以上说话者中的一

个说话者相关的语音数据。在一些实施例中，多通道语音文件可以由具有至少两个通道的语音获取设备生成，例如电话录音系统。该至少两个通道中的每一个可以与一个用户终端(例如，乘客终端130或司机终端140)对应。在一些实施例中，所有说话者的用户终端可以同时收集语音数据，并且可以记录与语音数据有关的时间信息。该所有说话者的用户终端可以将相应的语音数据发送到电话录音系统。然后，电话录音系统可以基于接收的语音数据生成多通道语音文件。

[0061] 在一些实施例中，该语音文件可以是单通道语音文件。可以从单通道获取单通道语音文件。具体地，与一个或以上说话者相关的语音数据可以由仅有一个通道的语音获取设备收集，例如车载麦克风、道路监视器等。例如，在打车服务期间，在司机搭载乘客之后，车载麦克风可以记录司机和乘客之间的对话。

[0062] 在一些实施例中，语音获取设备可以存储在各种场景中生成的至少两个语音文件。对于特定场景，音频文件获取模块410可以从至少两个语音文件中选择一个或以上对应的语音文件。例如，在打车服务期间，音频文件获取模块410可以从所述至少两个语音文件中选择包含与打车服务相关的词汇的一个或以上语音文件，例如“车牌号”、“出发地点”、“目的地”、“驾驶时间”等。在一些实施例中，语音获取设备可以在特定场景中收集语音数据。例如，该语音获取设备(例如，电话录音系统)可以与打车应用程序连接。语音获取设备可以在司机和乘客使用打车应用程序时收集与司机和乘客相关的语音数据。在一些实施例中，收集的语音文件(例如，多通道语音文件和/或单通道语音文件)可以存储在存储器150中。音频文件获取模块410可以从存储器150获取该语音文件。

[0063] 音频文件分离模块420可以用于将语音文件(或音频文件)分成一个或以上语音子文件(或音频子文件)。所述一个或以上语音子文件中的每一个可以包括与一个或以上说话者中的一个说话者相对应的至少两个语音段。

[0064] 对于多通道语音文件，与一个或以上说话者中每一个相关的语音数据可以独立地分布在所述一个或以上通道的一个通道中。音频文件分离模块420可以将多通道语音文件分成与所述一个或以上通道相关的一个或以上语音子文件。

[0065] 对于单通道语音文件，与一个或以上说话者相关的语音数据可以收集到单通道中。音频文件分离模块420可以通过执行语音分离将该单通道语音文件分成一个或以上语音子文件。在一些实施例中，该语音分离可以包括盲源分离(BSS)法、计算听觉场景分析(CASA)法等。

[0066] 在一些实施例中，语音转换模块440可以首先基于语音识别方法将语音文件转换为文本文件。该语音识别方法可以包括但不限于特征参数匹配算法、隐马尔可夫模型(HMM)算法、人工神经网络(ANN)算法等。然后，分离模块420可以基于语义分析方法将文本文件分成一个或以上文本子文件。该语义分析方法可以包括基于字符匹配的分词方法(例如，最大匹配算法、全切分算法、统计语言模型算法)、基于序列注释的分词方法(例如，POS标记)、基于深度学习的分词方法(例如，隐马尔可夫模型算法)等。在一些实施例中，所述一个或以上文本子文件中的每一个可以与一个或以上说话者中的一个说话者对应。

[0067] 信息获取模块430可以用于获取与所述至少两个语音段中的每一个相对应的时间信息和说话者识别信息。在一些实施例中，与所述至少两个语音段中的每一个对应的时间信息可包括起始时间和/或持续时间(或结束时间)。在一些实施例中，起始时间和/或持续

时间可以是绝对时间(例如,1分20秒、3分40秒)或相对时间(例如,语音文件的整个时间长度的20%)。具体地,至少两个语音段的起始时间和/或持续时间可以反映语音文件中的至少两个语音段的序列。在一些实施例中,说话者识别信息可以是能够区分一个或以上说话者的信息。该说话者识别信息可以包括姓名、ID号或其它对于该一个或以上说话者唯一的信息。在一些实施例中,每个语音子文件中的语音段可以与相同的说话者对应。信息获取模块430可以为每个语音子文件中的语音段确定说话者的说话者识别信息。

[0068] 语音转换模块440可以用于将至少两个语音段转换为至少两个文本段。所述至少两个语音段中的每一个可以与至少两个文本段中的一个文本段对应。语音转换模块440可以基于语音识别方法将至少两个语音段转换为至少两个文本段。在一些实施例中,语音识别方法可包括特征参数匹配算法、隐马尔可夫模型(HMM)算法、人工神经网络(ANN)算法等或其任意组合。在一些实施例中,语音转换模块440可以基于孤立词识别、关键词定位或连续语音识别将至少两个语音段转换为至少两个文本段。例如,转换后的文本段可以包括词语、短语等。

[0069] 特征信息生成模块450可以用于基于至少两个文本段、时间信息和说话者识别信息生成与语音文件相对应的特征信息。生成的特征信息可以包括至少两个文本段和说话者识别信息(如图7所示)。在一些实施例中,特征信息生成模块450可以基于文本段的时间信息,更具体地,基于文本段的起始时间,对至少两个文本段进行排序。特征信息生成模块450可以用对应的说话者识别信息标记至少两个排序的文本段中的每一个。然后,特征信息生成模块450可以生成与该语音文件相对应的特征信息。在一些实施例中,特征信息生成模块450可以基于所述一个或以上说话者的说话者识别信息对至少两个文本段进行排序。例如,如果两个说话者同时说话,特征信息生成模块450可以基于两个说话者的说话者识别信息对至少两个文本段进行排序。

[0070] 模型训练模块460可以用于通过基于一个或以上用户行为和对应于样本语音文件的特征信息训练初始模型来生成用户行为模型。所述特征信息可以包括一个或以上说话者的至少两个文本段和说话者识别信息。可以通过分析语音文件获取一个或以上用户行为。语音文件的分析可以由用户或系统100执行。例如,用户可以收听打车服务的语音文件,并且可以将一个或以上用户行为确定为:“司机迟到了20分钟”、“乘客携带了一个大行李”、“下雪了”、“司机通常快速驾驶”等。可以在训练初始模型之前获取所述一个或以上用户行为。所述一个或以上用户行为中的每一个可以与一个或以上说话者中的一个说话者对应。与说话者相关的至少两个文本段可以反映说话者的行为。例如,如果与司机相关的文本段是“你要去哪里”,司机的行为可以包括向乘客询问目的地。又例如,如果与乘客相关的文本段是“人民路”,乘客的行为可以包括回复司机的问题。在一些实施例中,处理器220可以生成如图6中所描述的特征信息,然后将其发送到模型训练模块460。在一些实施例中,模型训练模块460可以从存储器150获取特征信息。可以从处理器220获取或者可以从外部设备(例如,处理设备)获取从存储器150获取的特征信息。在一些实施例中,该特征信息和一个或以上用户行为可以构成训练样本。

[0071] 模型训练模块460还可以用于获取初始模型。该初始模型可以包括一个或以上分类器。每个分类器可以有与分类器的权重相关的初始参数,在训练初始模型时,可以更新分类器的初始参数。所述初始模型可以将特征信息作为输入,并且可以基于该特征信息确定

内部输出。模型训练模块460可以将一个或以上用户行为作为期望输出。模型训练模块460可以训练初始模型以最小化损失函数。在一些实施例中，模型训练模块460可以在损失函数中将内部输出与期望输出进行比较。例如，内部输出可以对应内部分数，期望输出可以对应期望分数。内部分数和期望分数可以相同或不同。损失函数可以与内部分数和期望分数之间的差值相关。具体地，当内部输出与期望输出相同时，内部分数与期望分数相同，损失函数最小(例如，零)。损失函数可以包括但不限于0-1损失、感知器损失、铰链损失、对数损失、平方损失、绝对损失和指数损失。损失函数的最小化可以是迭代的。当损失函数的值小于预定阈值时，可以终止损失函数最小化的迭代。所述预定阈值可以基于各种因素来设置，包括训练样本的数量、模型的准确度等。模型训练模块460可以在损失函数最小化期间迭代地调整初始模型的初始参数。在损失函数最小化之后，可以更新初始模型中的分类器的初始参数并生成训练好的用户行为模型。

[0072] 用户行为确定模块470可以用于基于与语音文件相对应的特征信息来执行用户行为模型以生成一个或以上用户行为。语音文件对应的特征信息可以包括至少两个文本段和一个或以上说话者的说话者识别信息。在一些实施例中，处理器220可以生成如图6中所描述的特征信息。并将其发送到用户行为确定模块470。在一些实施例中，用户行为确定模块470可以从存储器150获取特征信息。可以从处理器220获取或者可以从外部设备(例如，处理设备)获取从存储器150获取的特征信息。可以由模型训练模块460训练用户行为模型。

[0073] 用户行为确定模块470可以将特征信息输入到该用户行为模型中。该用户行为模型可以基于输入的特征信息输出一个或以上用户行为。

[0074] 应当注意，上述生成语音文件对应的特征信息的处理引擎的描述是出于说明的目的而提供的，并不旨在限制本申请的范围。对于本领域普通技术人员，可以在本申请的指导下进行多种变化和修改。然而，那些变化和修改不脱离本申请的范围。例如，一些模块可以安装在与其他模块分开的不同设备中。仅作为示例，特征信息生成模块450可以在一个设备中，其他模块可以在不同的设备中。又例如，音频文件分离模块420和信息获取模块430可以集成为一个模块，用于将语音文件分成一个或以上语音子文件，每个语音子文件包括至少两个语音段，并获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息。

[0075] 图5是根据本申请的一些实施例所示的音频文件分离模块的示例性框图。音频文件分离模块420可以包括去噪单元510和分离单元520。

[0076] 在将语音文件分成一个或以上语音子文件之前，去噪单元510可以用于去除语音文件中的噪音以生成去噪语音文件。可以使用去噪方法来去除噪音，包括但不限于语音激活检测(VAD)。VAD可以去除语音文件中的噪音，从而可以呈现保留在语音文件中的语音段。在一些实施例中，VAD还可以确定每个语音段的起始时间和/或持续时间(或结束时间)。

[0077] 在一些实施例中，在将语音文件分离成一个或以上语音子文件之后，去噪单元510可以用于去除该一个或以上语音子文件中的噪音。可以使用去噪方法去除噪音，包括但不限于VAD。VAD可以去除一个或以上语音子文件中每一个中的噪音。VAD还可以确定一个或以上语音子文件中每个语音子文件中的至少两个语音段中的每个语音段的起始时间和/或持续时间(或结束时间)。

[0078] 在去除语音文件中的噪音之后，分离单元520可以用于将去噪语音文件分成一个

或以上去噪语音子文件。对于多通道去噪语音文件，分离单元520可以将多通道去噪语音文件分成相对于通道的一个或以上去噪语音子文件。对于单通道去噪语音文件，分离单元520可以通过执行语音分离将单通道去噪语音文件分离成一个或以上去噪语音子文件。

[0079] 在一些实施例中，在去除语音文件中的噪音之前，分离单元520可以用于将语音精细分成一个或以上语音子文件。对于多通道语音文件，分离单元520可以将该多通道语音文件分成相对于通道的一个或以上语音子文件。对于单通道语音文件，分离单元520可以通过执行语音分离将单通道语音文件分离成一个或以上语音子文件。

[0080] 图6是根据本申请的一些实施例所示的用于生成语音文件对应的特征信息的示例性过程的流程图。在一些实施例中，过程600可以在如图1所示的按需服务系统100中实现。例如，过程600可以以指令的形式存储在存储器150和/或其他存储器(例如，ROM 230、RAM 240)中，由服务器110(例如，服务器110中的处理引擎112、服务器110中的处理引擎112的处理器220、服务器110的逻辑电路和/或服务器110的相应模块)调用和/或执行。本申请以服务器110的模块执行该指令为例。

[0081] 步骤610，音频文件获取模块410可以获取音频文件。在一些实施例中，音频文件可以是包括一个或以上说话者相关的语音数据的语音文件。在一些实施例中，可以将一个或以上麦克风安装在至少一个车厢中(例如，出租车、私家车、公共汽车、火车、动车、高铁、地铁、船只、飞行器、飞船、热气球、潜水艇)，以用于检测来自所述一个或以上说话者中至少一个说话者的语音，并产生相应说话者的语音数据。例如，如果将麦克风安装在汽车中(也称为车载麦克风)，该麦克风可以记录汽车中说话者(例如，司机和乘客)的语音数据。在一些实施例中，可以将所述一个或以上麦克风安装在商店、道路或房屋中以检测来自其中的一个或以上说话者的语音并生成与所述一个或以上说话者相对应的语音数据。例如，如果顾客在商店购物，商店中的麦克风可以记录顾客和店员之间的语音数据。又如，如果一个或以上游客访问一个景点，他(她)们之间的谈话可以通过安装在景区中的麦克风来检测。然后该麦克风可以生成与游客相关的语音数据。该语音数据可用于分析游客的行为及其对景点的看法。在一些实施例中，可以将所述一个或以上麦克风安装在车辆或车辆的配件(例如，摩托车头盔)上。例如，摩托车骑手可以通过安装在其头盔上的麦克风彼此进行交谈。麦克风可记录摩托车骑手之间的谈话并产生相应摩托车骑手的语音数据。在一些实施例中，所述一个或以上麦克风可以安装在特定位置以监测附近的声音。例如，可以将所述一个或以上麦克风安装在重建站点中以监测建筑工人的重建噪音和声音。又例如，如果将麦克风安装在房屋中，该麦克风可以检测家庭成员之间的语音并生成与家庭成员相关的语音数据。该语音数据可用于分析家庭成员的习惯。在一些实施例中，麦克风可以检测房屋中的非人类声音，例如车辆、宠物等的声音。

[0082] 在一些实施例中，语音文件可以是多通道语音文件。可以从至少两个通道获取该多通道语音文件。所述至少两个通道中的每一个可以包括与一个或以上说话者中的一个说话者相关的语音数据。在一些实施例中，多通道语音文件可以由具有至少两个通道的语音获取设备生成，例如电话录音系统。例如，如果说话者A和说话者B两个说话者彼此通话，可以分别由说话者A的移动电话和说话者B的移动电话收集说话者A和说话者B的语音数据。可以将与说话者A相关的语音数据发送到电话录音系统的一个通道，可以将与说话者B相关的语音数据发送到电话录音系统的另一个通道。可以由电话录音系统生成包括与说话者A和

说话者B相关的语音数据的多通道语音文件。在一些实施例中，语音获取设备可以存储在各种场景中生成的多通道语音文件。对于特定场景，音频文件获取模块410可以从至少两个多通道语音文件中选择一个或以上相应的多通道语音文件。例如，在打车服务期间，音频文件获取模块410可以从至少两个语音文件中选择包含与打车服务相关的词汇的一个或以上语音文件，例如“车牌号”、“出发地点”、“目的地”、“驾驶时间”等。在一些实施例中，可以在特定场景中使用语音获取设备(例如，电话录音系统)。例如，电话录音系统可以与打车应用程序连接。电话录音系统可以在司机和乘客使用打车应用程序时收集与司机和乘客相关的语音数据。

[0083] 在一些实施例中，语音文件可以是单通道语音文件。可以从单通道获取单通道语音文件。具体地，与一个或以上说话者相关的语音数据可以由仅有一个通道的语音获取设备收集，例如车载麦克风、道路监视器等。例如，在打车服务期间，在司机搭载乘客之后，车载麦克风可以记录司机和乘客之间的对话。在一些实施例中，语音获取设备可以存储在各种场景中生成的单通道语音文件。对于特定场景，音频文件获取模块410可以从至少两个单通道语音文件中选择一个或以上对应的单通道语音文件。例如，在打车服务期间，音频文件获取模块410可以从至少两个单通道语音文件中选择包含与打车服务相关的词汇的一个或以上单通道语音文件，例如“车牌号”、“出发地点”、“目的地”、“驾驶时间”等。在一些实施例中，语音获取设备(例如，车载麦克风)可以在特定场景中收集语音数据。例如，可以将麦克风安装在已在打车应用程序上注册的司机的汽车中。该车载麦克风可以在司机和乘客使用打车应用程序时记录与司机和乘客相关的语音数据。

[0084] 在一些实施例中，收集的语音文件(例如，多通道语音文件和/或单通道语音文件)可以存储在存储器150中。音频文件获取模块410可以从存储器150或语音获取设备的存储器获取该语音文件。

[0085] 步骤620，音频文件分离模块420可以将语音文件(或音频文件)分成一个或以上语音子文件(或音频子文件)，每个语音子文件包括至少两个语音段。所述一个或以上语音子文件中的每一个语音子文件可以与所述一个或以上说话者中的一个说话者相对应。例如，语音文件可以包括与三个说话者(例如，说话者A、说话者B和说话者C)相关的语音数据。音频文件分离模块420可以将该语音文件分成三个语音子文件(例如，子文件A、子文件B和子文件C)。子文件A可以包括与说话者A相关的至少两个语音段；子文件B可以包括与说话者B相关的至少两个语音段；子文件C可以包括与说话者C相关的至少两个语音段。

[0086] 对于多通道语音文件，与一个或以上说话者中每一个说话者相关的语音数据可以独立地分布在所述一个或以上通道的一个通道中。音频文件分离模块420可以将多通道语音文件分成与所述一个或以上通道相关的一个或以上语音子文件。

[0087] 对于单通道语音文件，可以将与一个或以上说话者相关的语音数据收集到单通道中。音频文件分离模块420可以通过执行语音分离将该单通道语音文件分成一个或以上语音子文件。在一些实施例中，所述语音分离可以包括盲源分离(BSS)方法、计算听觉场景分析(CASA)方法等。BSS是仅基于观测到的信号数据而不知道源信号和传输信道的参数来恢复源信号的独立成分的过程。BSS方法可以包括基于独立分量分析(ICA)的BBS方法、基于信号稀疏度的BSS方法等。CASA是基于使用人类听觉感知建立的模型将混合语音数据分离为物理声源的过程。CASA可以包括数据驱动的CASA、图式驱动的CASA等。

[0088] 在一些实施例中，首先，语音转换模块440可以基于语音识别方法将语音文件转换为文本文件。语音识别方法可以包括但不限于特征参数匹配算法、隐马尔可夫模型(HMM)算法、人工神经网络(ANN)算法等。然后，分离模块420可以基于语义分析方法将文本文件分成一个或以上文本子文件。该语义分析方法可以包括基于字符匹配的分词方法(例如，最大匹配算法、全切分算法、统计语言模型算法)、基于序列注释的分词方法(例如，POS标记)、基于深度学习的分词方法(例如，隐马尔可夫模型算法)等。在一些实施例中，所述一个或以上文本子文件中的每一个可以与一个或以上说话者中的一个说话者对应。

[0089] 在步骤630中，信息获取模块430可以获取至少两个语音段中的每一个对应的时间信息和说话者识别信息。在一些实施例中，所述至少两个语音段中的每一个对应的时间信息可以包括起始时间和/或持续时间(或结束时间)。在一些实施例中，起始时间和/或持续时间可以是绝对时间(例如，1分20秒)或相对时间(例如，语音文件完整时长的20％)。具体地，至少两个语音段的起始时间和/或持续时间可以反映语音文件中的至少两个语音段的序列。在一些实施例中，说话者识别信息是能够区分一个或以上说话者的信息。该说话者识别信息可以包括姓名、ID号或其它对于该一个或以上说话者唯一的信息。在一些实施例中，每个语音子文件中的语音段可以对应相同的说话者(例如，对应于说话者A的子文件A)。信息获取模块430可以为每个语音子文件中的语音段确定说话者的说话者识别信息。

[0090] 在步骤640中，语音转换模块440可以将至少两个语音段转换成至少两个文本段。所述至少两个语音段中的每一个可以与所述至少两个文本段中的一个文本段相对应。语音转换模块440可以基于语音识别方法将至少两个语音段转换成至少两个文本段。语音识别方法可以包括特征参数匹配算法、隐马尔科夫模型(HMM)算法、人工神经网络(ANN)算法等或其任意组合。特征参数匹配算法可以包括将要识别的语音数据的特征参数与语音模板中的语音数据的特征参数进行比较。例如，语音转换模块440可以将语音文件中至少两个语音段的特征参数与语音模板中的语音数据的特征参数进行比较。语音转换模块440可以基于该比较将至少两个语音段转换为至少两个文本段。HMM算法可以从可观测的参数确定过程的隐含参数，并使用该隐含参数将至少两个语音段转换为至少两个文本段。语音转换模块440可以基于ANN算法精确地将至少两个语音段转换为至少两个文本段。在一些实施例中，语音转换模块440可以基于孤立词识别、关键词定位或连续语音识别将至少两个语音段转换为至少两个文本段。例如，转换后的文本段可以包括词语、短语等。

[0091] 在步骤650中，特征信息生成模块450可以基于至少两个文本段、时间信息和说话者识别信息生成与语音文件相对应的特征信息。所述生成的特征信息可以包括至少两个文本段和说话者识别信息。在一些实施例中，特征信息生成模块450可以基于文本段的时间信息对至少两个文本段进行排序，更具体地，基于文本段的起始时间对至少两个文本段进行排序。特征信息生成模块450可以用对应的说话者识别信息标记至少两个排序的文本段中的每一个。然后，特征信息生成模块450可以生成与该语音文件相对应的特征信息。在一些实施例中，特征信息生成模块450可以基于所述一个或以上说话者的说话者识别信息对至少两个文本段进行排序。例如，如果两个说话者同时说话，则特征信息生成模块450可以基于两个说话者的说话者识别信息对至少两个文本段进行排序。

[0092] 应当注意，上述用于确定与语音文件相对应的特征信息的过程是出于说明的目的而提供的，并不旨在限制本申请的范围。对于本领域普通技术人员，可以在本申请的指导下

进行多种变化和修改。然而，那些变化和修改不脱离本申请的范围。在一些实施例中，在将至少两个语音段转换为至少两个文本段之后，可以将至少两个文本段中的每一个切分成词语或短语。

[0093] 图7是根据本申请的一些实施例所示的与双通道语音文件相对应的示例性特征信息的示意图。如图7所示，该语音文件是包括与说话者A和说话者B相关的语音数据的双通道语音文件M。音频文件分离模块420可以将双通道语音文件M分成两个语音子文件，每个语音子文件包括至少两个语音段(图7中未示出)。语音转换模块440可以将至少两个语音段转换为至少两个文本段。两个语音子文件可以分别对应两个文本子文件(例如，文本子文件721和文本子文件722)。如图7所示，文本子文件721包括与说话者A相关的两个文本段(例如，第一文本段721-1和第二文本段721-2)。T_{11}和T_{12}是第一文本段721-1的起始时间和结束时间，T_{13}和T_{14}是第二文本段721-2的起始时间和结束时间。类似地，文本子文件722包括与说话者B相关的两个文本段(例如，第三文本段722-1和第四文本段722-2)。在一些实施例中，文本段可以切分成词语。例如，可以将第一文本段切分为三个词语(例如，w_1，w_2和w_3)。说话者识别信息C_1可以表示说话者A，说话者识别信息C_2可以表示说话者B。特征信息生成模块450可以基于文本段的起始时间(例如T_{11}、T_{21}、T_{13}和T_{23})对两个文本子文件中的文本段(例如，第一文本段721-1、第二文本段721-2、第三文本段722-1和第四文本段722-2)进行排序。然后，特征信息生成模块450可以通过用对应的说话者识别信息(例如，C_1或C_2)标记每个排序的文本段来生成与双通道语音文件M相对应的特征信息。生成的特征信息可以表示为“$w_1\_C_1w_2\_C_1w_3\_C_1w_1\_C_2w_2\_C_2w_3\_C_2w_4\_C_1w_5\_C_1w_4\_C_2w_5\_C_2$”。

[0094] 表1和表2示出了与说话者A和说话者B相关的示例性文本信息(即，文本段)和时间信息。特征信息生成模块450可以基于该时间信息对文本信息进行排序。然后，特征信息生成模块450可以通过相应的说话者识别信息标记排序的文本信息。说话者识别信息C1可以表示说话者A，说话者识别信息C_2可以表示说话者B。生成的特征信息可以表示为“今天_C_1天气_C_1很好_C_1是_C_2今天_C_2天气_C_2很好_C_2去_C_1旅行_C_1好_C_2”。

表1

说话者	文字信息	时间信息
说话者 A	“今天”“天气”“很好”	[1.02s, 3.46s]
	“去”“旅行”	[8.63s, 10.86s]

表2

说话者	文字信息	时间信息
说话者 B	“是”“今天”“天气”“很好”	[4.02s, 7.50s]
	“好”	[11.02s, 14.56s]

[0095] 应当注意的是，以上对生成与双通道语音文件相对应的特征信息的描述是出于说

明的目的，并不旨在限制本申请的范围。对于本领域普通技术人员，可以在本申请的指导下进行多种变化和修改。然而，那些变化和修改不脱离本申请的范围。在该实施例中，可以将文本段切分为词语。在其他实施例中，可以将文本段切分为字符或短语。

[0096] 图8是根据本申请的一些实施例所示的用于生成与语音文件相对应的特征信息的示例性过程的流程图。在一些实施例中，过程800可以在如图1所示的按需服务系统100中实现。例如，过程800可以以指令的形式存储在存储器150和/或其他存储器(例如，ROM 230、RAM 240)中，并且由服务器110(例如，服务器110中的处理引擎112、服务器110中的处理引擎112的处理器220、服务器110的逻辑电路和/或服务器110的相应模块)调用和/或执行。本申请以服务器110的模块执行所述指令为例。

[0097] 在步骤810中，音频文件获取模块410可以获取包括与一个或以上说话者相关的语音数据的语音文件。在一些实施例中，该语音文件可以是从至少两个通道获取的多通道语音文件。所述至少两个通道中的每一个可以包括与一个或以上说话者中的一个说话者相关的语音数据。在一些实施例中，语音文件可以是从单通道获取的单通道语音文件。可以将一个或以上说话者相关语音数据收集到该单通道语音文件中。语音文件的获取可以结合图6的描述，此处不再重复。

[0098] 在步骤820中，音频文件分离模块420(例如，去噪单元510)可以去除语音文件中的噪音以生成去噪语音文件。可以使用去噪方法来去除噪音，包括但不限于语音激活检测(VAD)。VAD可以去除语音文件中的噪音，从而可以呈现保留在语音文件中的语音段。VAD还可以确定每个语音段的起始时间和/或持续时间(或结束时间)。因此，去噪语音文件可以包括与一个或以上说话者相关的语音段、语音段的时间信息等。

[0099] 在步骤830中，音频文件分离模块420(例如，分离单元520)可以将去噪语音文件分成一个或以上去噪语音子文件。所述一个或以上去噪语音子文件中的每一个可以包括与一个或以上说话者中的一个说话者相关的至少两个语音段。对于多通道去噪语音文件，分离单元520可以将该多通道去噪语音文件分成相对于通道的一个或以上去噪语音子文件。对于单通道去噪语音文件，分离单元520可以通过执行语音分离将该单通道去噪语音文件分离成一个或以上去噪语音子文件。所述语音分离可结合图6中的描述，此处不再重复。

[0100] 在步骤840中，信息获取模块430可以获取至少两个语音段中的每一个对应的时间信息和说话者识别信息。在一些实施例中，与所述至少两个语音段中的每一个对应的时间信息可包括起始时间和/或持续时间(或结束时间)。在一些实施例中，起始时间和/或持续时间可以是绝对时间(例如，1分20秒)或相对时间(例如，语音文件完整时长的20％)。说话者识别信息是能够区分一个或以上说话者的信息。该说话者识别信息可以包括姓名、ID号或其它对于该一个或以上说话者唯一的信息。时间信息和说话者识别信息的获取可以结合图6的描述，此处不再重复。

[0101] 在步骤850中，语音转换模块440可以将至少两个语音段转换为至少两个文本段。所述至少两个语音段中的每一个可以与至少两个文本段中的一个文本段对应。所述转换可以结合图6的描述，此处不再重复。

[0102] 在步骤860中，特征信息生成模块450可以基于至少两个文本段、时间信息和说话者识别信息生成与语音文件相对应的特征信息。生成的特征信息可以包括至少两个文本段和说话者识别信息(如图7所示)。特征信息的生成可以结合图6的描述，此处不再重复。

[0103] 图9是根据本申请的一些实施例所示的用于生与成语音文件相对应的特征信息的示例性过程的流程图。在一些实施例中，过程900可以在如图1所示的按需服务系统100中实现。例如，过程900可以以指令的形式存储在存储器150和/或其他存储器(例如，ROM 230、RAM 240)中，并由服务器110(例如，服务器110中的处理引擎112、服务器110中的处理引擎112的处理器220、服务器110的逻辑电路和/或服务器110的相应模块)调用和/或执行。本申请以服务器110的模块执行所述指令为例。

[0104] 在步骤910中，音频文件获取模块410可以获取包括一个或以上说话者相关的语音数据的语音文件。在一些实施例中，该语音文件可以是从至少两个通道获取的多通道语音文件。所述至少两个通道中的每一个可以包括与一个或以上说话者中的一个说话者相关的语音数据。在一些实施例中，语音文件可以是从单通道获取的单通道语音文件。可以将与一个或以上说话者相关的语音数据收集到该单通道语音文件中。语音文件的获取可以结合图6的描述，此处不再重复。

[0105] 在步骤920中，音频文件分离模块420(例如，分离单元520)可以将去噪语音文件分成一个或以上去噪语音子文件。所述一个或以上语音子文件中的每一个可以包括与一个或以上说话者中的一个说话者相关的至少两个语音段。对于多通道语音文件，分离单元520可以将该多通道语音文件分成相对于通道的一个或以上语音子文件。对于单通道语音文件，分离单元520可以通过执行语音分离将单通道语音文件分离成一个或以上语音子文件。所述语音分离可结合图6中的描述，此处不再重复。

[0106] 在步骤930中，音频文件分离模块420(例如，去噪单元510)可以去除一个或以上语音子文件中的噪音。可以使用去噪方法来去除噪音，包括但不限于语音激活检测(VAD)。VAD可以去除一个或以上语音子文件中的每一个中的噪音。VAD还可以确定一个或以上语音子文件中每一个中的至少两个语音段中的每一个的起始时间和/或持续时间(或结束时间)。

[0107] 在步骤940中，信息获取模块430可以获取至少两个语音段中的每一个对应的时间信息和说话者识别信息。在一些实施例中，与所述至少两个语音段中的每一个对应的时间信息可包括起始时间和/或持续时间(或结束时间)。在一些实施例中，起始时间和/或持续时间可以是绝对时间(例如，1分20秒)或相对时间(例如，语音文件完整时长的20％)。说话者识别信息是能够区分一个或以上说话者的信息。该说话者识别信息可以包括姓名、ID号或其它对于该一个或以上说话者唯一的信息。时间信息和说话者识别信息的获取可以结合图6的描述，此处不再重复。

[0108] 在步骤950中，语音转换模块440可以将至少两个语音段转换为至少两个文本段。所述至少两个语音段中的每一个可以与所述至少两个文本段中的一个文本段相对应。所述转换可以结合图6的描述，此处不再重复。

[0109] 在步骤960中，特征信息生成模块450可以基于至少两个文本段、时间信息和说话者识别信息生成与语音文件相对应的特征信息。生成的特征信息可以包括至少两个文本段和说话者识别信息(如图7所示)。特征信息的生成可以结合图6的描述，此处不再重复。

[0110] 应当注意的是，以上对生成与语音文件对应的特征信息过程的描述是出于说明的目的而提供的，并不旨在限制本申请的范围。对于本领域普通技术人员，可以在本申请的指导下进行多种变化和修改。然而，那些变化和修改不脱离本申请的范围。例如，可以按顺序或同时执行该过程中的一些步骤。又例如，该过程中的一些步骤可以分解为至少两个步骤。

[0111] 图10是根据本申请的一些实施例所示的用于生成用户行为模型的示例性过程的流程图。在一些实施例中，过程1000可以在如图1所示的按需服务系统100中实现。例如，过程1000可以以指令的形式存储在存储器150和/或其他存储器(例如，ROM 230、RAM 240)中，由服务器110(例如，服务器110中的处理引擎112、服务器110中的处理引擎112的处理器220、服务器110的逻辑电路和/或服务器110的相应模块)调用和/或执行。本申请以服务器110的模块执行所述指令为例。

[0112] 在步骤1010中，模型训练模块460可以获取初始模型。在一些实施例中，初始模型可包括一个或以上分类器。每个分类器可以有与分类器的权重相关的初始参数。

[0113] 初始模型可以包括排序支持向量机(SVM)模型、梯度提升决策树(GBDT)模型、LambdaMART模型、自适应增强模型、循环神经网络模型、卷积网络模型、隐马尔可夫模型、感知器神经网络模型、Hopfield网络模型、自组织映射(SOM)或学习矢量量化(LVQ)等或其任意组合。循环神经网络模型可以包括长短期记忆(LSTM)神经网络模型、分层循环神经网络模型、双向循环神经网络模型、二阶循环神经网络模型、完全循环网络模型、回声状态网络模型、多时间尺度循环神经网络(MTRNN)模型等。

[0114] 在步骤1020中，模型训练模块460可以获取一个或以上用户行为，每个用户行为与一个或以上说话者中的一个说话者对应。可以通过分析一个或以上说话者的样本语音文件来获取一个或以上用户行为。在一些实施例中，所述一个或以上用户行为可以与特定场景相关。例如，在打车服务期间，一个或以上用户行为可以包括与司机相关的行为、与乘客相关的行为等。对于司机，该行为可以包括向乘客询问出发地点、目的地等。对于乘客，该行为可以包括向司机询问到达时间、车牌号码等。又例如，在购物服务期间，一个或以上用户行为可以包括与销售员相关的行为、与顾客相关的行为等。对于销售员，该行为可以包括询问客户他/她正在寻找的产品、付款方式等。对于顾客，该行为可以包括询问销售人员的价格、使用方法等。在一些实施例中，模型训练模块460可以从存储器150获取所述一个或以上用户行为。

[0115] 在步骤1030中，模型训练模块460可以获取与样本语音文件相对应的特征信息。该特征信息可以与一个或以上说话者相关的一个或以上用户行为对应。与样本语音文件对应的特征信息可以包括一个或以上说话者的至少两个文本段和说话者识别信息。与说话者相关的至少两个文本段可以反映该说话者的行为。例如，如果与司机相关的文本段是“你要去哪里”，司机的行为可以包括向乘客询问目的地。又例如，如果与乘客相关的文本段是“人民路”，乘客的行为可能包括回复司机的问题。在一些实施例中，处理器220可以如图6所述生成与样本语音文件相对应的特征信息并将其发送到模型训练模块460。在一些实施例中，模型训练模块460可以从存储器150获取所述特征信息。可以从处理器220获取或者可以从外部设备(例如，处理设备)获取从存储器150获取的特征信息。

[0116] 在步骤1040中，模型训练模块460可以基于一个或以上用户行为和特征信息通过训练初始模型来生成用户行为模型。所述一个或以上分类器中的每一个可以具有与分类器的权重相关的初始参数。可以在初始模型的训练期间调整与分类器的权重相关的初始参数。

[0117] 特征信息和一个或以上用户行为可以构成训练样本。初始模型可以将特征信息作为输入，并可以基于该特征信息确定内部输出。模型训练模块460可以将一个或以上用户行

为作为期望输出。模型训练模块460可以训练初始模型以最小化损失函数。模型训练模块460可以在损失函数中将内部输出与期望输出进行比较。例如，内部输出可以对应内部分数，期望输出可以对应期望分数。损失函数可以与内部分数和期望分数之间的差值相关。具体地，当内部输出与期望输出相同时，内部分数与期望分数相同，损失函数最小(例如，零)。损失函数的最小化可以是迭代的。当损失函数的值小于预定阈值时，可以终止损失函数最小化的迭代。所述预定阈值可以基于各种因素来设置，包括训练样本的数量、模型的准确度等。模型训练模块460可以在损失函数最小化期间迭代地调整初始模型的初始参数。在损失函数最小化之后，可以更新初始模型中的分类器的初始参数并生成训练好的用户行为模型。

[0118] 图11是根据本申请的一些实施例所示的用于执行用户行为模型以生成用户行为的示例性过程的流程图。在一些实施例中，过程1100可以在如图1所示的按需服务系统100中实现。例如，过程1100可以以指令的形式存储在存储器150和/或其他存储器(例如，ROM 230、RAM 240)中，并由服务器110(例如，服务器110中的处理引擎112、服务器110中的处理引擎112的处理器220、服务器110的逻辑电路和/或服务器110的相应模块)调用和/或执行。本申请以服务器110的模块执行所述指令为例。

[0119] 在步骤1110中，用户行为确定模块470可以获取与语音文件相对应的特征信息。该语音文件可以是包括至少两个说话者之间的对话的语音文件。该语音文件可以与本申请中其他地方描述的示例性语音文件不同。语音文件对应的特征信息可以包括至少两个文本段和一个或以上说话者的说话者识别信息。在一些实施例中，处理器220可以生成如图6所描述的特征信息然后将其发送到用户行为确定模块470。在一些实施例中，用户行为确定模块470可以从存储器150获取特征信息。可以从处理器220获取或者可以从外部设备(例如，处理设备)获取从存储器150获取的特征信息。

[0120] 在步骤1120中，用户行为确定模块470可以获取用户行为模型。在一些实施例中，用户行为模型可以在过程1000中由模型训练模块460训练。

[0121] 用户行为模型可以包括排序支持向量机(SVM)模型、梯度提升决策树(GBDT)模型、LambdaMART模型、自适应增强模型、循环神经网络模型、卷积网络模型、隐马尔可夫模型、感知器神经网络模型、Hopfield网络模型、自组织映射(SOM)或学习矢量量化(LVQ)等或其任意组合。循环神经网络模型可以包括长短期记忆(LSTM)神经网络模型、分层循环神经网络模型、双向循环神经网络模型、二阶循环神经网络模型、完全循环网络模型、回声状态网络模型、多时间尺度循环神经网络(MTRNN)模型等。

[0122] 在步骤1130，用户行为确定模块470可以基于特征信息执行用户行为模型以生成一个或以上用户行为。用户行为确定模块470可以将特征信息输入到用户行为模型中。用户行为模型可以基于该一个或以上输入的特征信息确定一个或以上用户行为。

[0123] 上文已对基本概念做了描述，显然，对于阅读此申请后的本领域的普通技术人员来说，上述发明公开仅作为示例，并不构成对本申请的限制。虽然此处并没有明确说明，本领域技术人员可能会对本申请进行各种修改、改进和修正。该类修改、改进和修正在本申请中被建议，所以该类修改、改进、修正仍属于本申请示范实施例的精神和范围。

[0124] 此外，本申请使用了特定词语来描述本申请的实施例。例如“一个实施例”、“一实施例”、和/或“一些实施例”意指与本申请至少一个实施例相关的某一特征、结构或特性。因

此，应当强调并注意的是，本说明书中在不同位置两次或多次提及的“一实施例”或“一个实施例”或“一替代性实施例”并不一定系指同一实施例。此外，本申请的一个或多个实施例中的某些特征、结构或特性可以进行适当的组合。

[0125] 此外，本领域的普通技术人员可以理解，本申请的各方面可以通过若干具有可专利性的种类或情况进行说明和描述，包括任何新的和有用的过程、机器、产品或物质的组合，或对其任何新的和有用的改良。相应地，本申请的各个方面可以完全由硬件执行、可以完全由软件(包括韧体、常驻软件、微代码等)执行、也可以由硬件和软件组合执行。以上硬件或软件均可被称为“单元”、“模块”或“系统”。此外，本申请公开的各方面可以采取体现在一个或以上计算机可读介质中的计算机程序产品的形式，其中计算机可读程序代码包含在其中。

[0126] 非暂时性计算机可读信号介质可能包含一个内含有计算机程序编码的传播数据信号，例如在基带上或作为载波的一部分。此类传播信号可以有多种形式，包括电磁形式、光形式等或任何合适的组合形式。计算机可读信号介质可以是除计算机可读存储介质之外的任何计算机可读介质，该介质可以通过连接至一个指令执行系统、装置或设备以实现通讯、传播或传输供使用的程序。位于计算机可读信号介质上的程序编码可以通过任何合适的介质进行传播，包括无线电、电缆、光纤电缆、RF、或类似介质等或其任意组合

[0127] 本申请各方面操作所需的计算机程序码可以用一种或多种程序语言的任意组合编写，包括面向对象程序设计，如Java、Scala、Smalltalk、Eiffel、JADE、Emerald、C++、C#、VB.NET，Python或类似的常规程序编程语言，如"C"编程语言，Visual Basic，Fortran 2003，Perl，COBOL 2002，PHP，ABAP，动态编程语言如Python，Ruby和Groovy或其它编程语言。程序代码可以完全在用户计算机上运行、或作为独立的软件包在用户计算机上运行、或部分在用户计算机上运行部分在远程计算机上运行、或完全在远程计算机或服务器上运行。在后种情况下，远程计算器可以通过任何网络形式与用户计算器连接，例如，局域网络(LAN)或广域网(WAN)，或连接至外部计算器(例如通过因特网)，或在云端计算环境中，或作为服务使用如软件即服务(SaaS)。本申请各部分操作所需的计算机程序编码可以用任意一种或以上程序语言编写，包括面向主体编程语言如Java、Scala、Smalltalk、Eiffel、JADE、Emerald、C++、C#、VB.NET、Python等，常规程序化编程语言如C语言、Visual Basic、Fortran 2003、Perl、COBOL 2002、PHP、ABAP，动态编程语言如Python、Ruby和Groovy，或其他编程语言等。该程序编码可以完全在用户计算机上运行、或作为独立的软件包在用户计算机上运行、或部分在用户计算机上运行部分在远程计算机运行、或完全在远程计算机或服务器上运行。在后种情况下，远程计算机可以通过任何网络形式与用户计算机连接，比如局域网(LAN)或广域网(WAN)，或连接至外部计算机(例如通过因特网)，或在云计算环境中，或作为服务使用如软件即服务(SaaS)。

[0128] 此外，除非权利要求中明确说明，本申请所述处理元素和序列的顺序、数字字母的使用、或其他名称的使用，并非用于限定本申请流程和方法的顺序。尽管上述披露中通过各种示例讨论了一些目前认为有用的发明实施例，但应当理解的是，该类细节仅起到说明的目的，附加的权利要求并不仅限于披露的实施例，相反，权利要求旨在覆盖所有符合本申请实施例实质和范围的修正和等价组合。例如，虽然以上所描述的系统组件可以通过硬件设备实现，但是也可以只通过软件的解决方案得以实现，如在现有的服务器或移动设备上安

装所描述的系统。

[0129] 同理，应当注意的是，为了简化本申请披露的表述，从而帮助对一个或以上发明实施例的理解，前文对本申请实施例的描述中，有时会将多种特征归并至一个实施例、附图或对其的描述中。然而，这种披露方法并不意味着本申请对象所需要的特征比权利要求中提及的特征多。相反，实施例的特征要少于上述披露的单个实施例的全部特征。

[0130] 在一些实施方案中，用于描述和要求本申请的某些实施方案的表达数量、性质等的数字应理解为在某些情况下由术语“约”，“近似”或“基本上”修饰。例如，除非另有说明，否则“约”、“近似”或“基本上”可表示其描述的值的±20％变化。因此，在一些实施例中，说明书和所附权利要求书中列出的数值参数是近似值，其可以根据特定实施方案寻求获得的所需性质而变化。在一些实施例中，数值参数应根据报告的有效数字的数量并通过应用普通的舍入技术来解释。尽管阐述本申请的一些实施方案的宽范围的数值范围和参数是近似值，但具体实施例中列出的数值尽可能精确地报告。

[0131] 除了与其相关的任何起诉文件历史、与本文件不一致或相冲突的任何相同的起诉文件历史、或者对于现在或稍后与本文件相关的权利要求的最宽范围可能具有限制性影响的任何起诉文件历史，本文引用的每个专利、专利申请、专利申请的出版物和其他材料，例如文章、书籍、说明书、出版物、文件、物品和/或类似物，在此通过引用整体并入本文。举例来说，如果与任何所包含的材料或与本文件有关内容的相关术语的描述、定义和/或使用存在任何不一致或冲突，以本文件中的术语为准。

[0132] 最后，应当理解的是，本申请公开的实施例是对本申请实施例的原理的说明。可采用的其他修改可以在本申请的范围内。因此，作为示例而非限制，可以根据本文的指导利用本申请的实施例的替代配置。因此，本申请的实施例不限于精确地如上所示和所述的实施例。

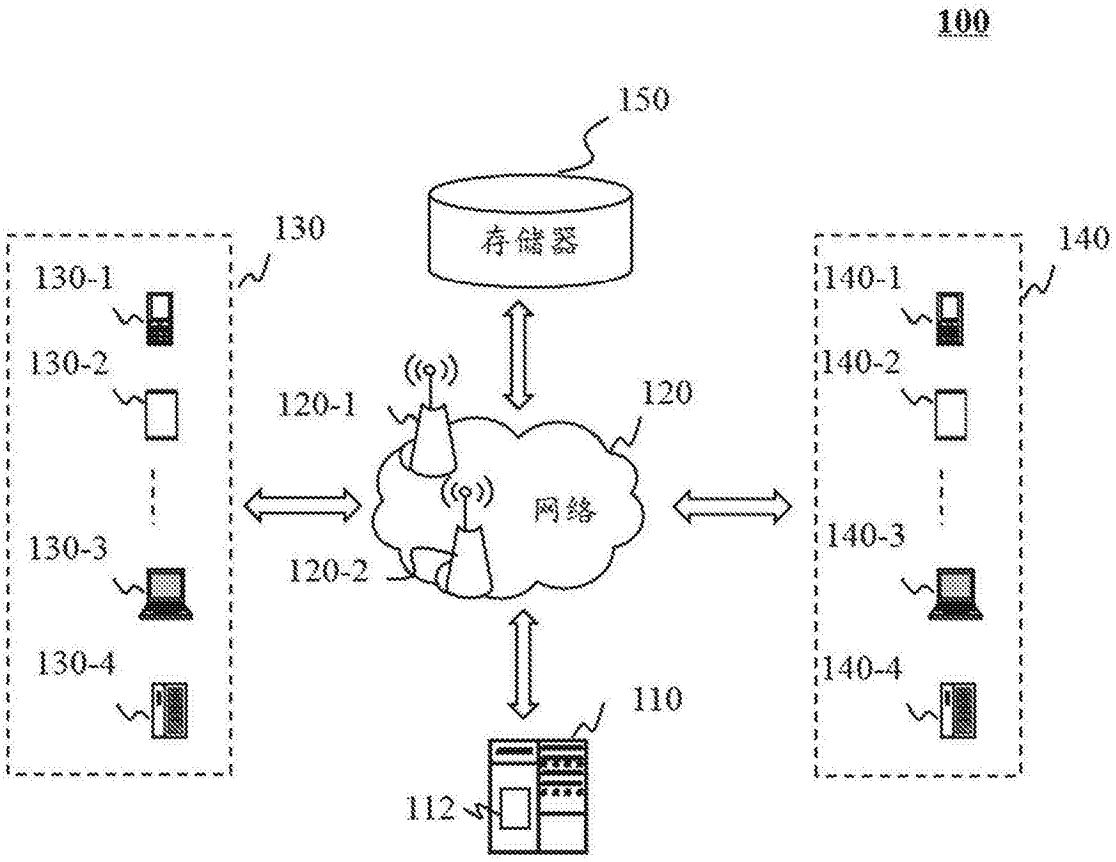

100
150
存储器
130
130-1
130-2
130-3
130-4
120-1
120
网络
120-2
110
112
140
140-1
140-2
140-3
140-4

图1

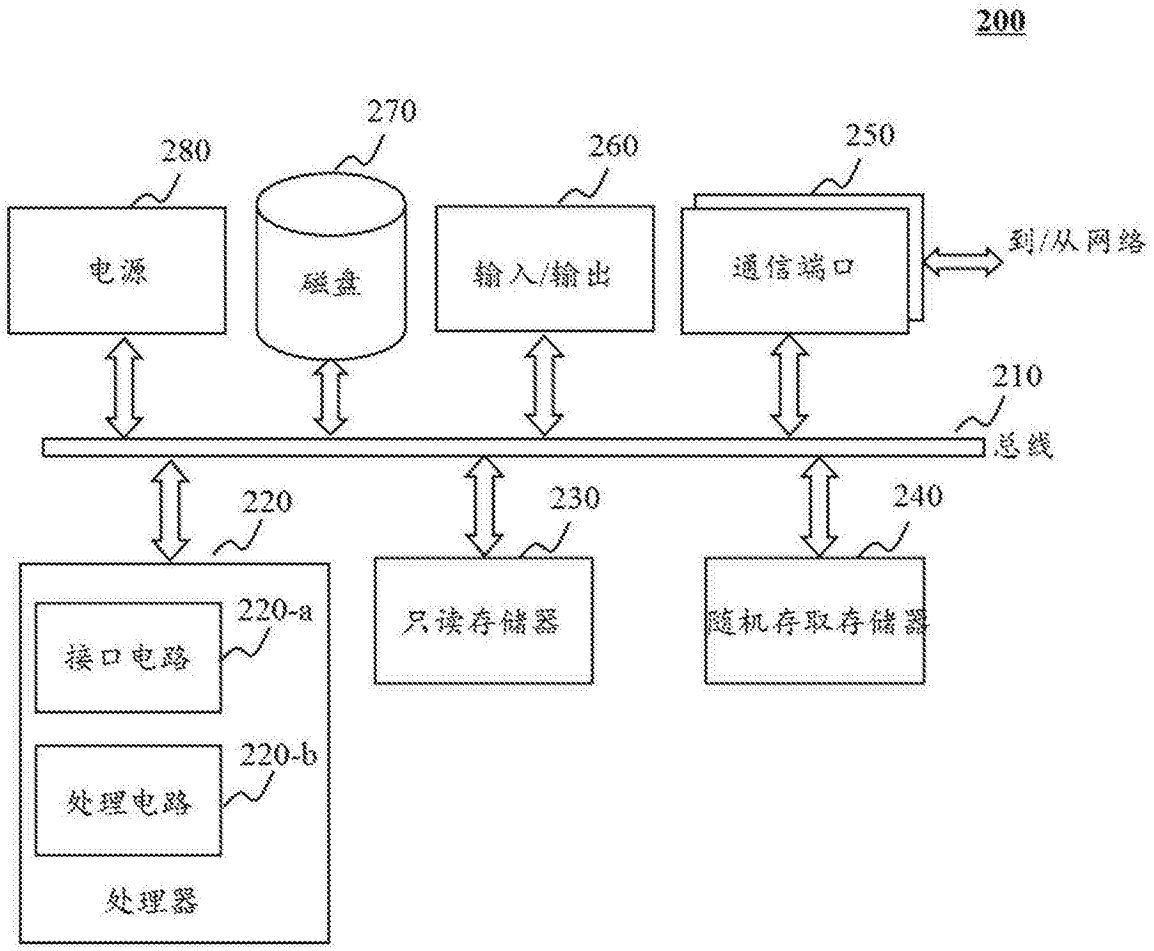

200
270
280
260
250
电源
磁盘
输入/输出
通信端口
到/从网络
210
总线
220
230
240
220-a
接口电路
只读存储器
随机存取存储器
220-b
处理电路
处理器

图2

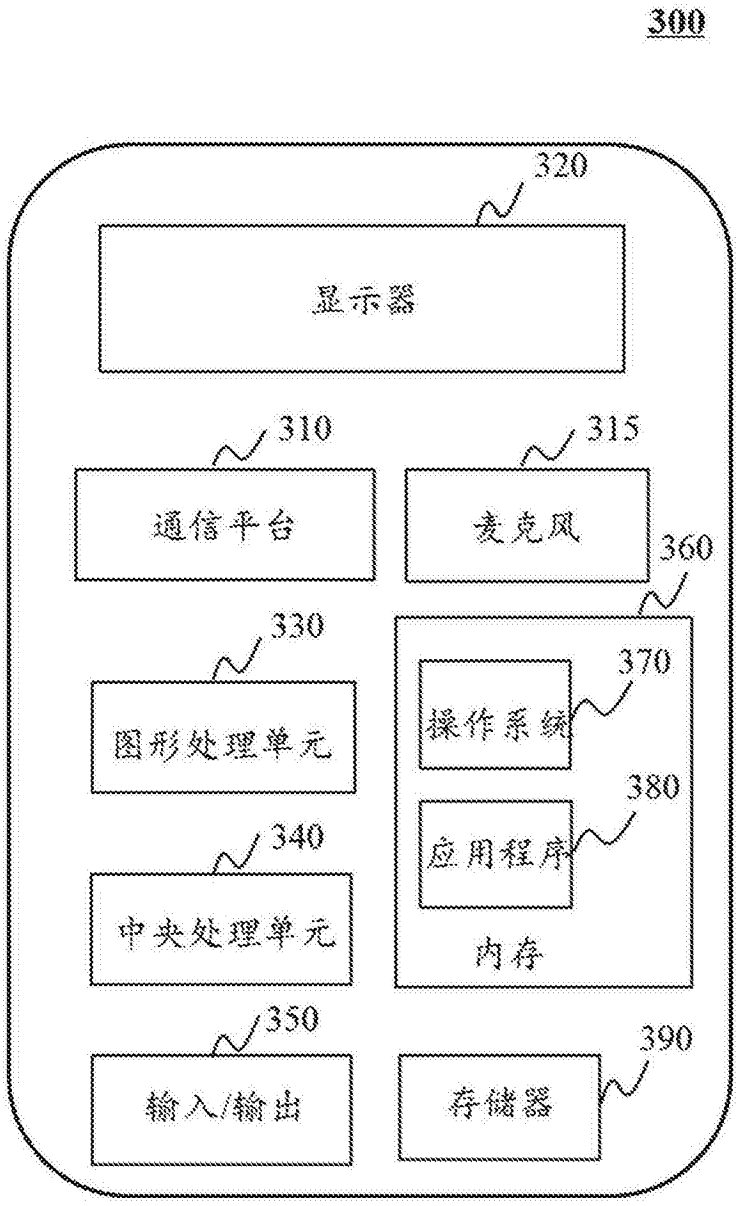

300
320
显示器
310
通信平台
315
麦克风
360
330
图形处理单元
370
操作系统
380
应用程序
340
中央处理单元
内存
350
输入/输出
390
存储器

图3

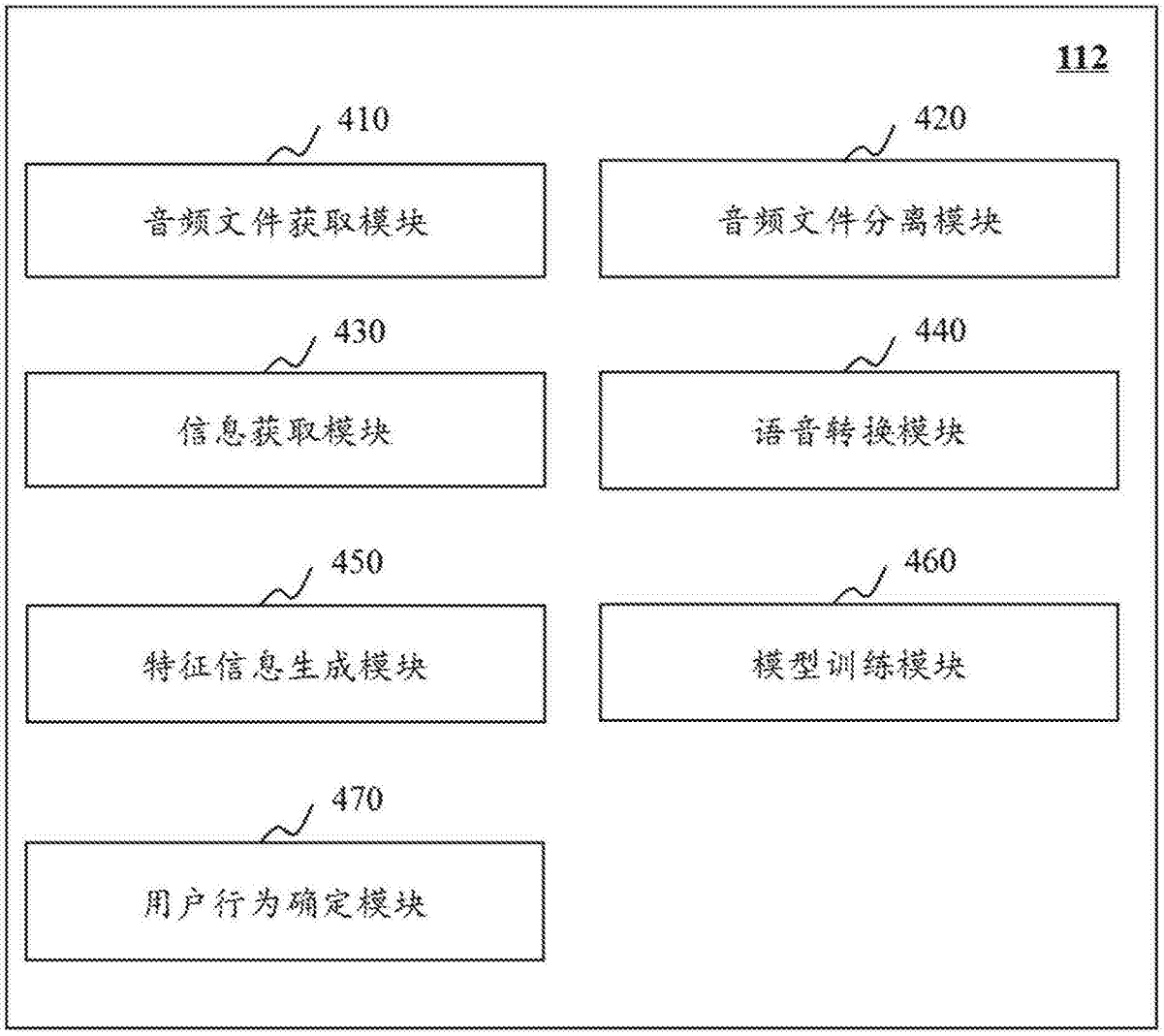

112
410
音频文件获取模块
420
音频文件分离模块
430
信息获取模块
440
语音转换模块
450
特征信息生成模块
460
模型训练模块
470
用户行为确定模块

图4

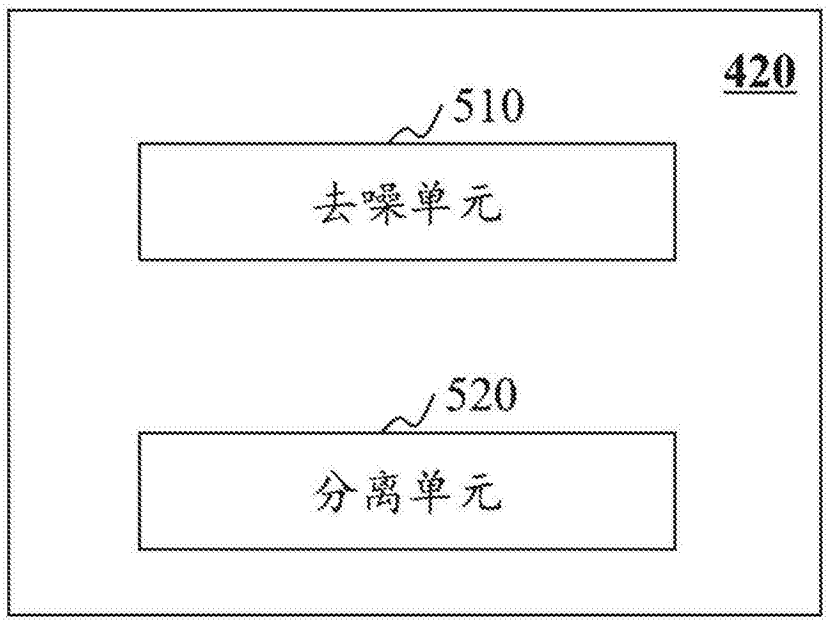


图5

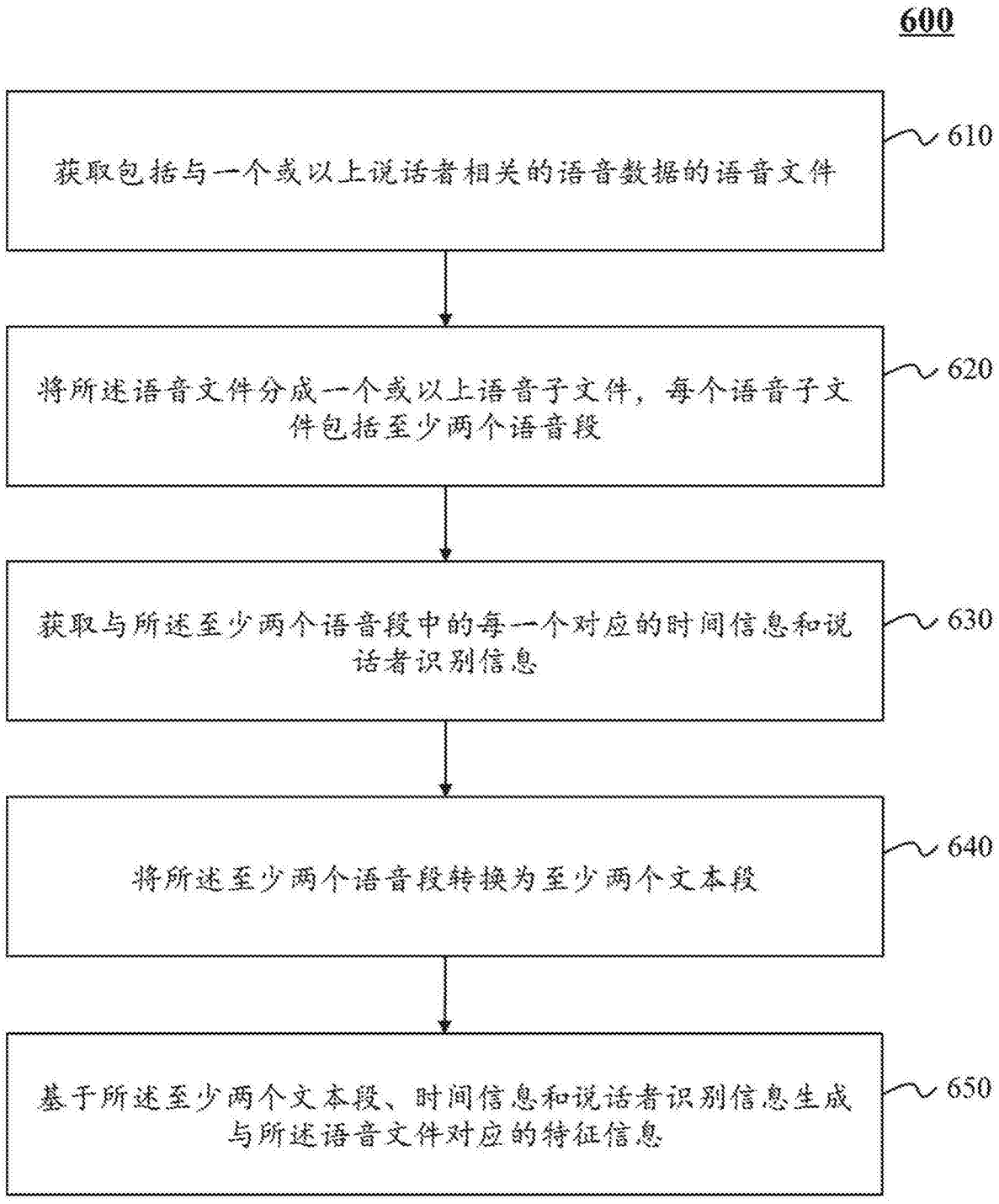

600
获取包括与一个或以上说话者相关的语音数据的语音文件
610
将所述语音文件分成一个或以上语音子文件，每个语音子文件包括至少两个语音段
620
获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息
630
将所述至少两个语音段转换为至少两个文本段
640
基于所述至少两个文本段、时间信息和说话者识别信息生成与所述语音文件对应的特征信息
650

图6

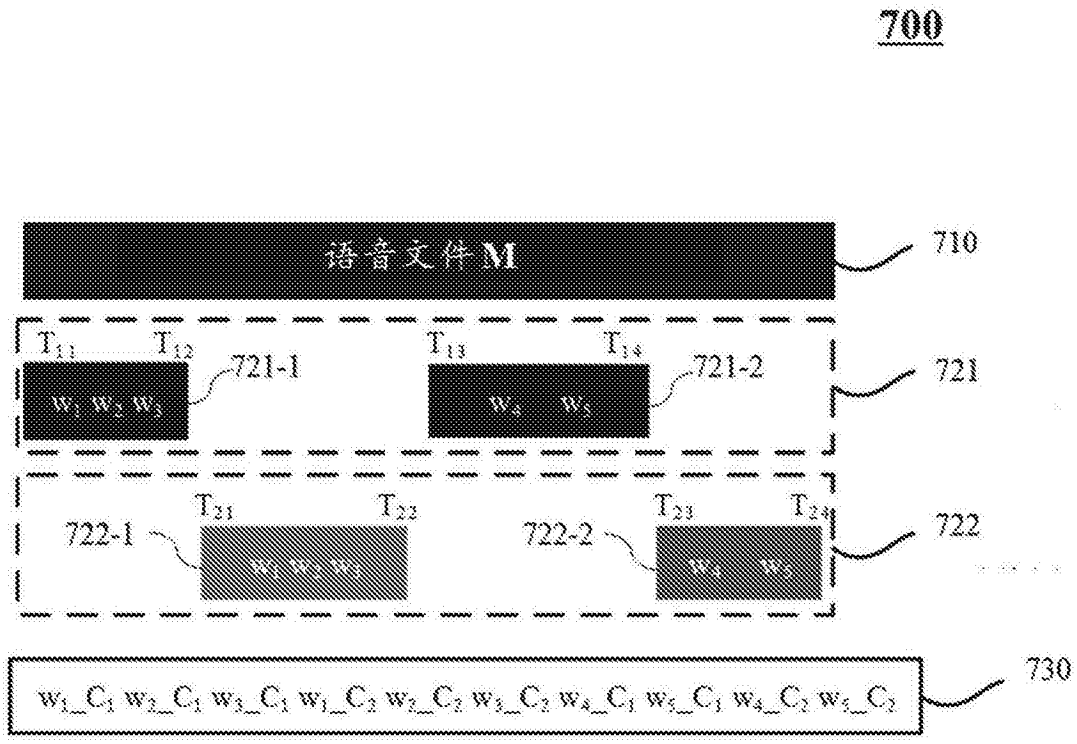


图7

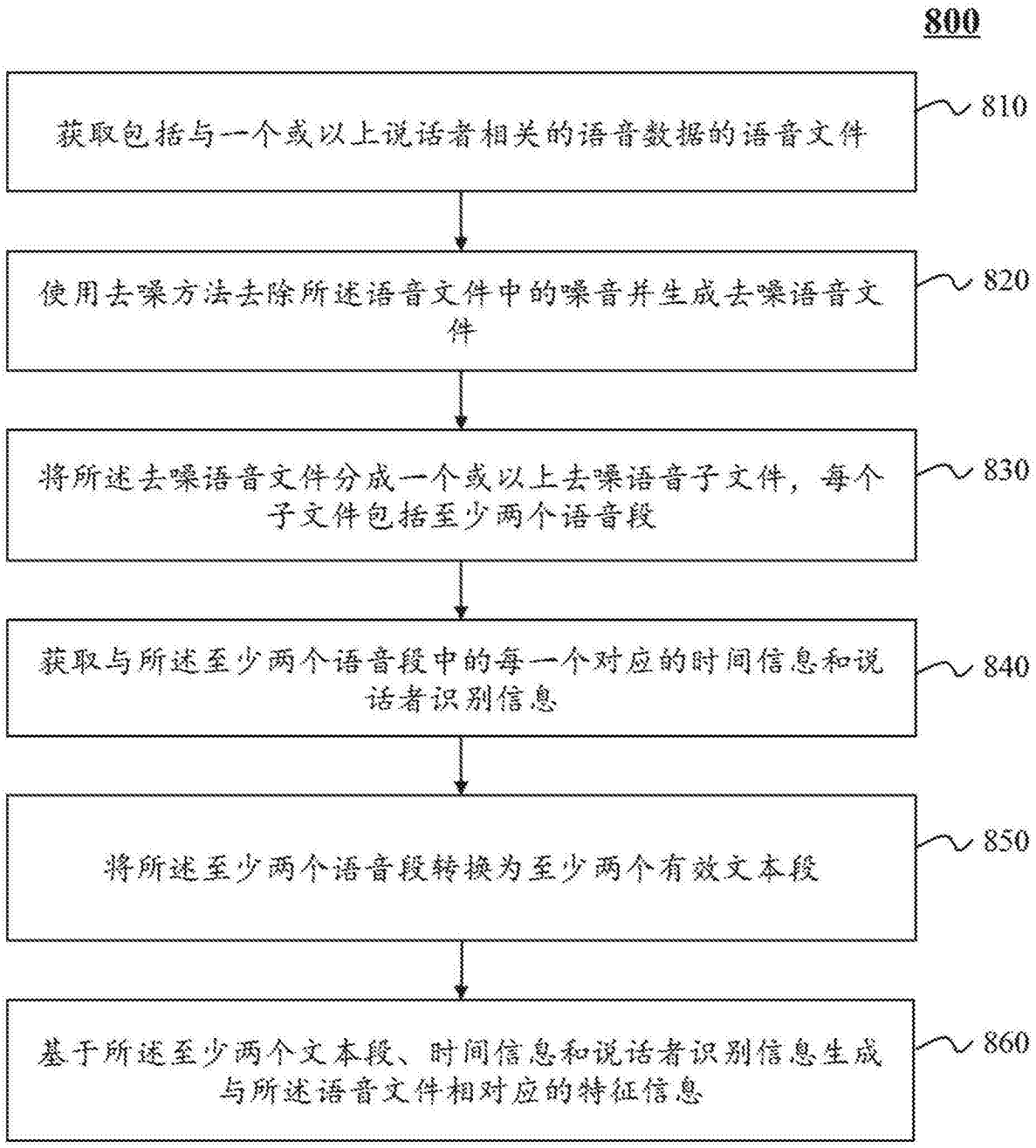

800
获取包括与一个或以上说话者相关的语音数据的语音文件
810
使用去噪方法去除所述语音文件中的噪音并生成去噪语音文件
820
将所述去噪语音文件分成一个或以上去噪语音子文件，每个子文件包括至少两个语音段
830
获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息
840
将所述至少两个语音段转换为至少两个有效文本段
850
基于所述至少两个文本段、时间信息和说话者识别信息生成与所述语音文件相对应的特征信息
860

图8

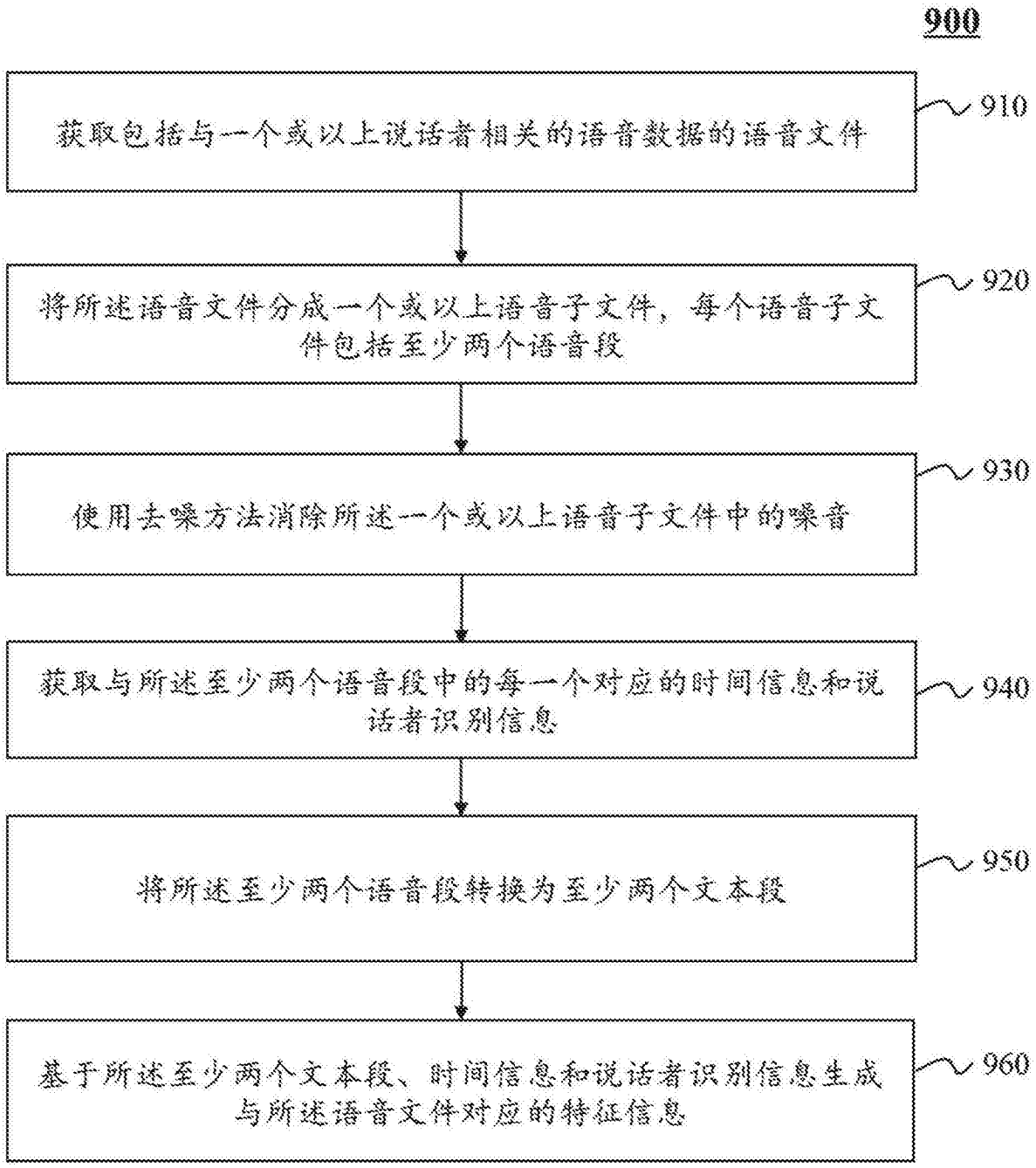

900
获取包括与一个或以上说话者相关的语音数据的语音文件
910
将所述语音文件分成一个或以上语音子文件，每个语音子文件包括至少两个语音段
920
使用去噪方法消除所述一个或以上语音子文件中的噪音
930
获取与所述至少两个语音段中的每一个对应的时间信息和说话者识别信息
940
将所述至少两个语音段转换为至少两个文本段
950
基于所述至少两个文本段、时间信息和说话者识别信息生成与所述语音文件对应的特征信息
960

图9

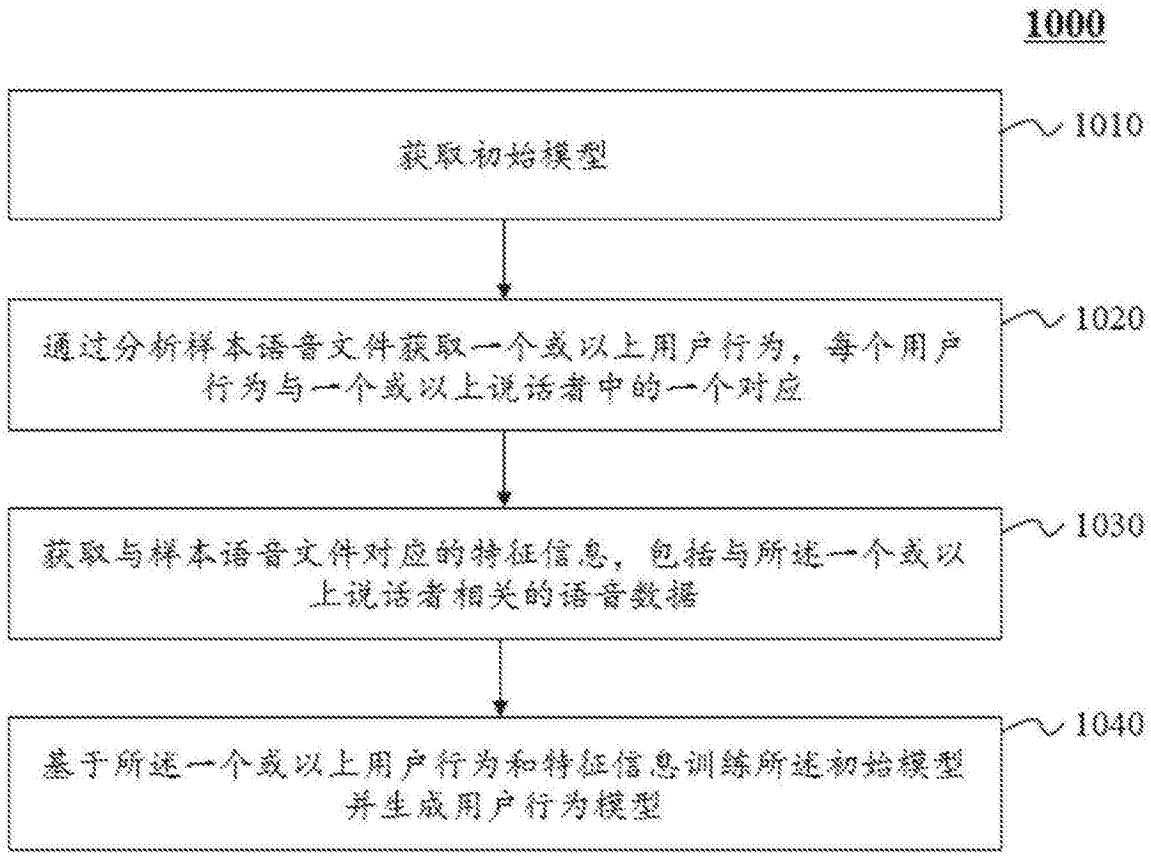

1000
获取初始模型
1010
通过分析样本语音文件获取一个或以上用户行为，每个用户行为与一个或以上说话者中的一个对应
1020
获取与样本语音文件对应的特征信息，包括与所述一个或以上说话者相关的语音数据
1030
基于所述一个或以上用户行为和特征信息训练所述初始模型并生成用户行为模型
1040

图10

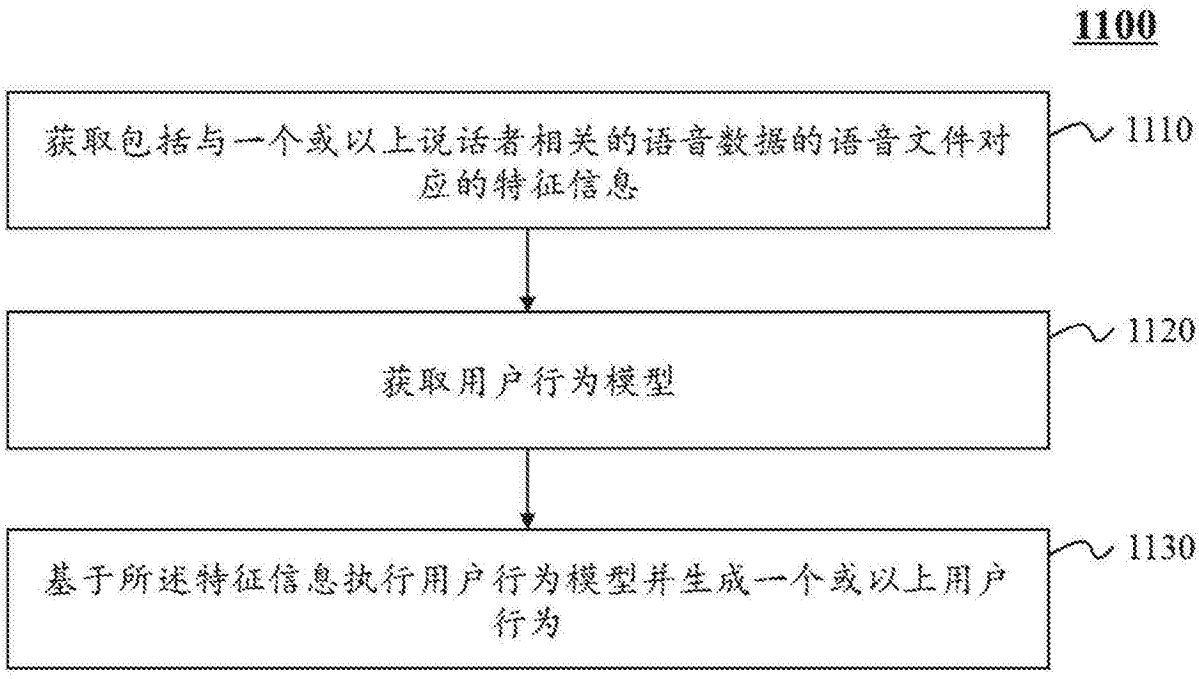

1100
获取包括与一个或以上说话者相关的语音数据的语音文件对应的特征信息
1110
获取用户行为模型
1120
基于所述特征信息执行用户行为模型并生成一个或以上用户行为
1130

图11