(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 110517678 B
(45)授权公告日 2022.04.08

(21)申请号 201910804779.5
(22)申请日 2019.08.28
(65)同一申请的已公布的文献号
申请公布号 CN 110517678 A
(43)申请公布日 2019.11.29
(73)专利权人 南昌保莱科技有限公司
地址 330029 江西省南昌市高新技术产业开发区京东大道1189号创新工场
(72)发明人 邹珺 熊阿伟
(51)Int.Cl.
G10L 15/22(2006.01)
H04N 7/18(2006.01)
G06V 40/16(2022.01)
G06V 40/10(2022.01)

(56)对比文件
CN 109979036 A,2019.07.05
CN 109658925 A,2019.04.19
CN 109767774 A,2019.05.17
CN 110010125 A,2019.07.12
CN 108337362 A,2018.07.27
WO 2014159581 A1,2014.10.02
US 2018158449 A1,2018.06.07
Jun″ichi Ido,等.Interaction of receptionist ASKA using vision and speech information.《IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems 2003》.2003,第335-340页.
郑志辉,等.基于语音实现人机对话的空调控制器研究开发.《2 0 1 8年中国家用电器技术大会 论文集》.2018,第331-335页.

审查员 周家行

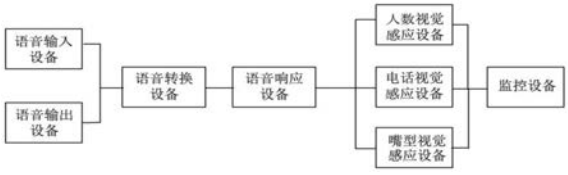
权利要求书2页 说明书6页 附图7页

(54)发明名称

一种基于视觉感应的AI语音应答响应系统

(57)摘要

本发明涉及一种基于视觉感应的AI语音应答响应系统,包括语音输出设备,语音输入设备,语音转换设备,语音响应设备;人数视觉感应设备,电话视觉感应设备,嘴型视觉感应设备,耳塞视觉感应设备,监控设备;用户通过语音输入设备输入语音,语音转换设备对输入的语音进行模拟信号和数字信号的转换,语音响应设备进行判断是否是特定语音,是特定语音则进行语音响应,通过语音输出设备进行AI对话模式;不是特定语音则为其他语音响应;则启动监控设备,这时就要根据嘴型视觉感应设备,电话视觉感应设备,人数视觉感应设备产生的信息来判断是否响应,只有当三者都判断为是时,通过语音输出设备进行AI对话模式。



CN 110517678 B

1.一种基于视觉感应的AI语音应答响应系统,其特征在于,包括语音输出设备,语音输入设备,语音转换设备,语音响应设备;人数视觉感应设备,电话视觉感应设备,嘴型视觉感应设备,监控设备;

监控设备,安装在需要响应的区域,对该区域进行实时监控;

语音输出设备,与语音转换设备相连,是产生语音的输出设备;

语音输入设备,与语音转换设备相连,将人的语音信息直接输入到计算机的人机接口设备;

语音转换设备,与语音输入设备和语音输出设备相连,输入的语音进行模拟信号和数字信号的转换,把语音输入设备输入的语音的特征信息作数字化处理后记录在计算机中;或者把计算机的信息转换为语音的特征信息输出;

嘴型视觉感应设备,与语音响应设备,监控设备相连,对监控设备对监控区域拍摄的视频,进行判断视频中的人的嘴型是否静止,人的嘴型是静止,则不响应;

人数视觉感应设备,与语音响应设备,监控设备相连,对监控设备对监控区域拍摄的视频,进行判断,视频中有几个人,两个或两个以上的人就判断为是,就不响应;

电话视觉感应设备,与语音响应设备,监控设备相连,对监控设备对监控区域拍摄的视频,进行判断视频中的人是否手持电话或戴耳机,人手持电话或戴耳机,则不响应;

语音响应设备,与语音输入设备,语音转换设备相连,对语音产生响应的设备,语音响应分为两种,一种为特定语音响应,一种为特定语音响应,一种为其他语音响应;特定语音响应,就是只要语音响应设备接收到特定语音就产生响应,通过语音输出设备进行对话模式;其他语音响应,是除了语音响应设备接收到特定语音的其他语音,则启动监控设备,这时就要根据嘴型视觉感应设备,电话视觉感应设备,人数视觉感应设备产生的信息来判断是否响应,只有当人数视觉感应设备,电话视觉感应设备,嘴型视觉感应设备,都判断为是时,通过语音输出设备进行AI对话模式;

语音响应的工作流程是,用户通过语音输入设备输入语音,语音转换设备对输入的语音进行模拟信号和数字信号的转换,语音响应设备进行判断是否是特定语音,是特定语音则进行语音响应,通过语音输出设备进行AI对话模式;不是特定语音则为其他语音响应;

其他语音响应的工作流程是,由嘴型视觉感应设备,对监控设备对监控区域拍摄的视频,进行判断视频中的人的嘴型是否静止,人的嘴型是静止则不响应,人的嘴型不静止则,人的嘴型不静止则,由人数视觉感应设备,对监控设备对监控区域拍摄的视频,进行判断,视频中有几个人,两个或两个以上的人就判断为是,就不响应,一个人就由电话视觉感应设备,对监控设备对监控区域拍摄的视频,进行判断视频中的人是否手持电话,人手持电话则不响应,人不手持电话则,进行语音响应,通过语音输出设备进行AI对话模式。

2.根据权利要求1中的所述一种基于视觉感应的AI语音应答响应系统,其特征在于,在监控设备中,能360°旋转摄像头,对响应区域进行全景视频监控。

3.根据权利要求1中的所述一种基于视觉感应的AI语音应答响应系统,其特征在于,在语音输出设备中,设置了锥盆式扬声器,使用的振膜材料在纸浆材料中掺入羊毛、蚕丝、碳纤维材料。

4.根据权利要求1中的所述一种基于视觉感应的AI语音应答响应系统,其特征在于,在语音输出设备中,设置了分频器,分频器为功率分频器也称无源式后级分频器,是在功率功

放之后进行分频的;它主要包含电感、电阻、电容无源组件,组成滤波器网络,把各频段的音频信号分别送到相应频段的扬声器中去重放。

5.根据权利要求1中的所述一种基于视觉感应的AI语音应答响应系统,其特征在于,在嘴型视觉感应设备中,设置人脸识别系统,在识别的人脸区域内通过设定特定色彩的阈值,检测到嘴唇的区域,通过视频的上一帧和下一帧的对比,嘴唇的边界不重合,则人的嘴型不是静止的。

6.根据权利要求1中的所述一种基于视觉感应的AI语音应答响应系统,其特征在于,在人数视觉感应设备中,设置计数器,计数器为1,则响应,计数器大于1,则不响应。

7.根据权利要求1中的所述一种基于视觉感应的AI语音应答响应系统,其特征在于,在电话视觉感应设备中,设置移动电话和固定电话的三维模型库,通过识别人的手,进而通过三维模型库比对人的手中物体,进而判断是否为电话。

8.根据权利要求1中的所述一种基于视觉感应的AI语音应答响应系统,其特征在于,在电话视觉感应设备中,设置蓝牙耳机和普通耳机的三维模型库,通过识别人的耳朵,进而通过三维模型库比对人的耳朵上戴的物体,进而判断是否为耳机。

一种基于视觉感应的AI语音应答响应系统

技术领域

[0001] 本发明涉及一种人工智能语音应答响应系统，具体的说是一种基于视觉感应的AI语音应答响应系统。

背景技术

[0002] 智能音箱，是一个音箱升级的产物，是家庭消费者用语音进行上网的一个工具，比如点播歌曲、上网购物，或是了解天气预报，它也可以对智能家居设备进行控制，比如打开窗帘、设置冰箱温度、提前让热水器升温等。

[0003] 智能音箱实际上都属于智能语音技术，其核心非常简要——要让机器在语音对话这一环节拥有近似于人的能力，智能音箱成为小家电一般的存在，渗入人们的日常生活空间，但是目前的智能语音技术的应答响应系统，对于模拟人的日常习惯和行为方面表现并不尽如人意。

[0004] 目前的智能语音技术的应答响应系统，需要使用者说出一个特定的词语，智能音箱通过这个特定的词语，进行应答响应，这个特定的词语通常是智能音箱的名称。而人们在日常对话中，人与人面对面对话时，很少说对方的名称，再进行对话，这就不符合人的日常习惯和行为，这是现有技术的不足之处。

发明内容

[0005] 为了解决现有技术中的智能检索功能，本发明采取的技术方案是，一种基于视觉感应的AI语音应答响应系统，其特征在于，包括语音输出设备，语音输入设备，语音转换设备，语音响应设备；人数视觉感应设备，电话视觉感应设备，嘴型视觉感应设备，监控设备。

[0006] 本发明还可以说是一种智能语音交互平台，监控设备，安装在需要响应的区域，对该区域进行实时监控。

[0007] 在监控设备中，能360°旋转摄像头，对响应区域进行全景视频监控。

[0008] 本发明还可以说是一种AI语音判定对话系统，语音输出设备，与语音转换设备相连，是产生语音的输出设备。

[0009] 在语音输出设备中，设置了电动式扬声器，利用音圈与恒定磁场之间的相互作用力使振膜振动而发声。

[0010] 在语音输出设备中，设置了锥盆式扬声器，使用的振膜材料在纸浆材料中或掺入羊毛、蚕丝、碳纤维材料，以增加其刚性、内阻尼及防水性能。

[0011] 在语音输出设备中，设置了分频器，分频器为功率分频器也称无源式后级分频器，是在功率功放之后进行分频的。它主要包含电感、电阻、电容无源组件，组成滤波器网络，把各频段的音频信号分别送到相应频段的扬声器中去重放。

[0012] 本发明还可以说是一种人工智能语音应答响应交互平台，语音输入设备，与语音转换设备相连，将人的语音信息直接输入到计算机的人机接口设备。

[0013] 本发明还可以说是一种AI语音技术应答响应系统，语音转换设备，与语音输入设

备和语音输出设备相连，输入的语音进行模拟信号和数字信号的转换，把语音输入设备输入的语音的特征信息(频率、周期、声调等变化)作数字化处理后记录在计算机中；或者把计算机的信息转换为语音的特征信息输出。

[0014] 嘴型视觉感应设备，与语音响应设备，监控设备相连，对监控设备对监控区域拍摄的视频，进行判断视频中的人的嘴型是否静止，人的嘴型是静止，则不响应。

[0015] 在嘴型视觉感应设备中，设置人脸识别系统，在识别的人脸区域内通过设定特定色彩的阈值，检测到嘴唇的区域，通过视频的上一帧和下一帧的对比，嘴唇的边界不重合，则人的嘴型不是静止的。

[0016] 人数视觉感应设备，与语音响应设备，监控设备相连，对监控设备对监控区域拍摄的视频，进行判断，视频中有几个人，两个或两个以上的人就判断为是，就不响应。

[0017] 在人数视觉感应设备中，设置计数器，计数器为1，则响应，计数器大于1，则不响应。

[0018] 电话视觉感应设备，与语音响应设备，监控设备相连，对监控设备对监控区域拍摄的视频，进行判断视频中的人是否手持电话或戴耳机，人手持电话或戴耳机，则不响应。

[0019] 在电话视觉感应设备中，设置移动电话和固定电话的三维模型库，通过识别人的手，进而通过三维模型库比对人的手中物体，进而判断是否为电话。

[0020] 在电话视觉感应设备中，设置蓝牙耳机和普通耳机的三维模型库，通过识别人的耳朵，进而通过三维模型库比对人的耳朵上戴的物体，进而判断是否为耳机。

[0021] 语音响应设备，与语音输入设备，语音转换设备相连，对语音产生响应的设备，语音响应分为两种，一种为特定语音响应，一种为其他语音响应。特定语音响应，就是只要语音响应设备接收到特定语音就产生响应，通过语音输出设备进行对话模式；其他语音响应，是除了语音响应设备接收到特定语音的其他语音，则启动监控设备，这时就要根据嘴型视觉感应设备，电话视觉感应设备，人数视觉感应设备产生的信息来判断是否响应，只有当人数视觉感应设备，电话视觉感应设备，嘴型视觉感应设备，都判断为是时，通过语音输出设备进行AI对话模式。

[0022] 语音响应的工作流程是，用户通过语音输入设备输入语音，语音转换设备对输入的语音进行模拟信号和数字信号的转换，语音响应设备进行判断是否是特定语音，是特定语音则进行语音响应，通过语音输出设备进行AI对话模式；不是特定语音则为其他语音响应；

[0023] 其他语音响应的工作流程是，由嘴型视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人的嘴型是否静止，人的嘴型是静止则不响应，人的嘴型不静止则，由人数视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断，视频中有几个人，两个或两个以上的人就判断为是，就不响应，一个人就由电话视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人是否手持电话或戴耳机，人手持电话或戴耳机则不响应，人不手持电话或戴耳机则，进行语音响应，通过语音输出设备进行AI对话模式。

附图说明

[0024] 图1为本发明的整体结构示意图。

[0025] 图2为本发明的语音响应的工作流图。

[0026] 图3为本发明的其他语音响应的实施例一工作流程图。

[0027] 图4为本发明的其他语音响应的实施例二工作流程图。

[0028] 图5为本发明的其他语音响应的实施例三工作流程图。

[0029] 图6为本发明的其他语音响应的实施例四工作流程图。

[0030] 图7为本发明的其他语音响应的实施例五工作流程图。

[0031] 图8为本发明的其他语音响应的实施例六工作流程图。

具体实施方式

[0032] 下面将参照附图对本发明的智能检索的监控平台系统的实施方案进行详细说明。

[0033] 实施例一

[0034] 为了解决现有技术中的智能检索功能，本发明采取的技术方案是，一种基于视觉感应的AI语音应答响应系统，其特征在于，包括语音输出设备，语音输入设备，语音转换设备，语音响应设备；人数视觉感应设备，电话视觉感应设备，嘴型视觉感应设备，监控设备。

[0035] 监控设备，安装在需要响应的区域，对该区域进行实时监控。

[0036] 在监控设备中，能360°旋转摄像头，对响应区域进行全景视频监控。

[0037] 语音输出设备，与语音转换设备相连，是产生语音的输出设备。

[0038] 在语音输出设备中，电动式扬声器，利用音圈与恒定磁场之间的相互作用力使振膜振动而发声。

[0039] 在语音输出设备中，设置了电动式扬声器，利用音圈与恒定磁场之间的相互作用力使振膜振动而发声。

[0040] 在语音输出设备中，设置了锥盆式扬声器，使用的振膜材料以纸浆材料为主，或掺入羊毛、蚕丝、碳纤维材料，以增加其刚性、内阻尼及防水性能。

[0041] 在语音输出设备中，设置了分频器，分频器为功率分频器也称无源式后级分频器，是在功率功放之后进行分频的。它主要包含电感、电阻、电容无源组件，组成滤波器网络，把各频段的音频信号分别送到相应频段的扬声器中去重放。

[0042] 语音输入设备，与语音转换设备相连，将人的语音信息直接输入到计算机的人机接口设备。

[0043] 语音转换设备，与语音输入设备和语音输出设备相连，输入的语音进行模拟信号和数字信号的转换，把语音输入设备输入的语音的特征信息(频率、周期、声调等变化)作数字化处理后记录在计算机中；或者把计算机的信息转换为语音的特征信息输出。

[0044] 嘴型视觉感应设备，与语音响应设备，监控设备相连，对监控设备对监控区域拍摄的视频，进行判断视频中的人的嘴型是否静止，人的嘴型是静止，则不响应。

[0045] 在嘴型视觉感应设备中，设置人脸识别系统，在识别的人脸区域内通过设定特定色彩的阈值，检测到嘴唇的区域，通过视频的上一帧和下一帧的对比，嘴唇的边界不重合，则人的嘴型不是静止的。

[0046] 在嘴型视觉感应设备中，设置人脸识别系统，通过矩形边缘对比，忽略边框内部的图像识别。

[0047] 这个主要是为了嘴型视觉感应设备排除电视机中的人脸。由于电视机为矩形边框，因此将电视机中的人脸进行忽略，以免误将电视机中的人脸进行识别。

[0048] 人数视觉感应设备，与语音响应设备，监控设备相连，对监控设备对监控区域拍摄的视频，进行判断，视频中有几个人，两个或两个以上的人就判断为是，就不响应。

[0049] 在人数视觉感应设备中，设置计数器，计数器为1，则响应，计数器大于1，则不响应。

[0050] 电话视觉感应设备，与语音响应设备，监控设备相连，对监控设备对监控区域拍摄的视频，进行判断视频中的人是否手持电话，人手持电话，则不响应。

[0051] 在电话视觉感应设备中，设置移动电话和固定电话的三维模型库，通过识别人的手，进而通过三维模型库比对人的手中物体，进而判断是否为电话。

[0052] 在电话视觉感应设备中，设置蓝牙耳机和普通耳机的三维模型库，通过识别人的耳朵，进而通过三维模型库比对人的耳朵上戴的物体，进而判断是否为耳机。

[0053] 为了判断使用者是否是通过耳机进行打电话的，识别使用者的是否带有耳机。

[0054] 语音响应设备，与语音输入设备，语音转换设备相连，对语音产生响应的设备，语音响应分为两种，一种为特定语音响应，一种为其他语音响应。特定语音响应，就是只要语音响应设备接收到特定语音就产生响应，通过语音输出设备进行对话模式；其他语音响应，是除了语音响应设备接收到特定语音的其他语音，则启动监控设备，这时就要根据嘴型视觉感应设备，电话视觉感应设备，人数视觉感应设备产生的信息来判断是否响应，只有当人数视觉感应设备，电话视觉感应设备，嘴型视觉感应设备，都判断为是时，语音响应设备进行响应，并通过语音输出设备进行AI对话模式。

[0055] 语音响应的工作流程是，用户通过语音输入设备输入语音，语音转换设备对输入的语音进行模拟信号和数字信号的转换，语音响应设备进行判断是否是特定语音，是特定语音则进行语音响应，通过语音输出设备进行AI对话模式；不是特定语音则为其他语音响应；

[0056] 其他语音响应的工作流程是，由嘴型视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人的嘴型是否静止，人的嘴型是静止则不响应，人的嘴型不静止则，由人数视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断，视频中有几个人，两个或两个以上的人就判断为是，就不响应，一个人就由电话视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人是否手持电话或戴耳机，人手持电话或戴耳机则不响应，人不手持电话或戴耳机则，进行语音响应，通过语音输出设备进行AI对话模式。

[0057] 实施例二

[0058] 其他语音响应的工作流程是，由嘴型视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人的嘴型是否静止，人的嘴型是静止则不响应，人的嘴型不静止则，由电话视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人是否手持电话或戴耳机，人手持电话或戴耳机则不响应，人不手持电话或戴耳机则，由人数视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断，视频中有几个人，两个或两个以上的人就判断为是，就不响应，一个人就进行语音响应，通过语音输出设备进行AI对话模式。

[0059] 实施例三

[0060] 其他语音响应的工作流程是，由人数视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中有几个人，两个或两个以上的人就判断为是，就不响应，，再由嘴型视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人的嘴型是否静止，

人的嘴型是静止则不响应，人的嘴型不静止则，由电话视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人是否手持电话或戴耳机，人手持电话或戴耳机则不响应，人不手持电话或戴耳机则进行语音响应，通过语音输出设备进行AI对话模式。

[0061] 实施例四

[0062] 其他语音响应的工作流程是，由人数视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中有几个人，两个或两个以上的人就判断为是，就不响应，再由电话视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人是否手持电话或戴耳机，人手持电话或戴耳机则不响应，人不手持电话或戴耳机，再由嘴型视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人的嘴型是否静止，人的嘴型是静止则不响应，人的嘴型不静止，则进行语音响应，通过语音输出设备进行AI对话模式。

[0063] 实施例五

[0064] 其他语音响应的工作流程是，由电话视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人是否手持电话或戴耳机，人手持电话或戴耳机则不响应，人不手持电话或戴耳机则，由人数视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断，视频中有几个人，两个或两个以上的人就判断为是，就不响应，一个人就由嘴型视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人的嘴型是否静止，人的嘴型是静止则不响应，人的嘴型不静止则，进行语音响应，通过语音输出设备进行AI对话模式。

[0065] 实施例六

[0066] 其他语音响应的工作流程是，由电话视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人是否手持电话或戴耳机，人手持电话或戴耳机则不响应，人不手持电话或戴耳机则，由嘴型视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断视频中的人的嘴型是否静止，人的嘴型是静止则不响应，人的嘴型不静止，再由人数视觉感应设备，对监控设备对监控区域拍摄的视频，进行判断，视频中有几个人，两个或两个以上的人就判断为是，就不响应，一个人就则，进行语音响应，通过语音输出设备进行AI对话模式。

[0067] 通过嘴型视觉感应设备，用来判断人的嘴型不变化，说明使用者并没有说话，声音的来源可能是来自电视，收音机，其他噪音，则不响应；人的嘴型发生变化则，说明声音为使用者的声音，但是有可能是和其他人说话，因此再通过人数视觉感应设备，判断视频中的人数，如果是两个或两个以上的人，就说明可能是两个人之间在对话，则不响应；如果是一个人，就说明这个人很可能是对智能应答系统说话，但是有可能在打电话或戴耳机；因此再通过电话视觉感应设备，判断视频中的人是否手持电话或戴耳机，人手持电话或戴耳机，则说明他在打电话或戴耳机，则不响应，如果人不手持电话或戴耳机则，则说明他是在和智能语音应答系统说话，则进行语音响应，通过语音输出设备进行AI对话模式。

[0068] 本发明的目的是为了使得智能语音应答系统能更合理的模仿人的行为习惯，将智能语音应答系统作为一个“人”来看，他应该在什么情况下做出应答反应才能更加人性化。智能语音应答系统经过嘴型视觉感应设备，电话视觉感应设备，人数视觉感应设备的判断是否使用者在与智能语音应答系统对话，而不需要特定词语作为呆板的指令。当然也有例外，比如说如果使用者是自言自语呢。首先，这种情况很少，再者，如果将智能语音应答系统

作为一个“人”,A和B两个人呆在一起,A自言自语,另外B也很可能会认为A是在和自己说话,这正是人的行为习惯。

[0069] 以上所述,仅为本发明较佳的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,根据本发明的技术方案及其发明构思加以等同替换或改变,都应涵盖在本发明的保护范围之内。

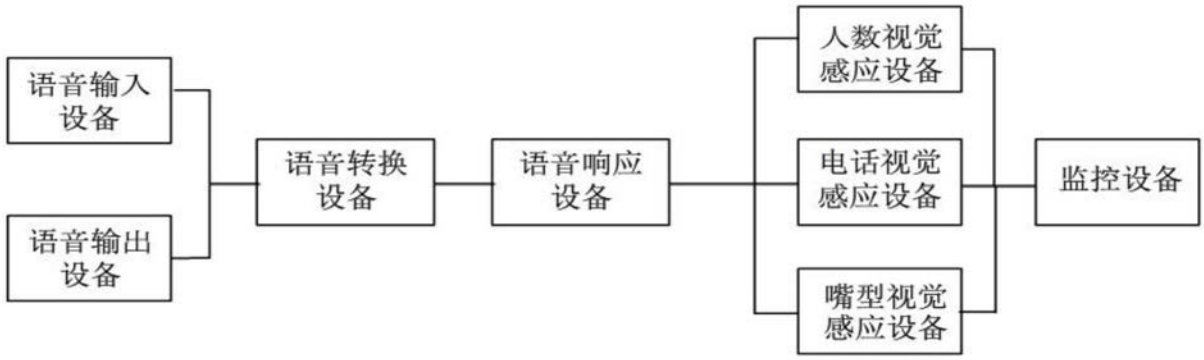

语音输入设备
语音输出设备
语音转换设备
语音响应设备
人数视觉感应设备
电话视觉感应设备
嘴型视觉感应设备
监控设备

图1

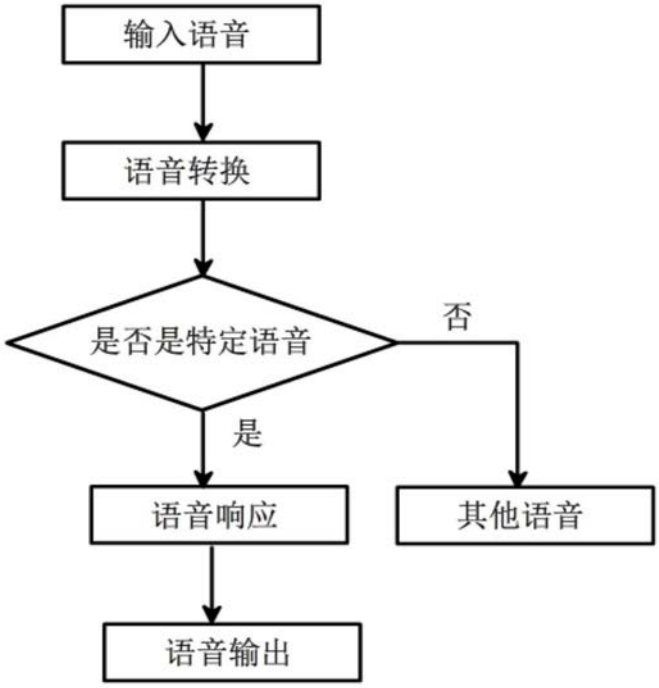

输入语音
语音转换
是否是特定语音
否
是
语音响应
其他语音
语音输出

图2

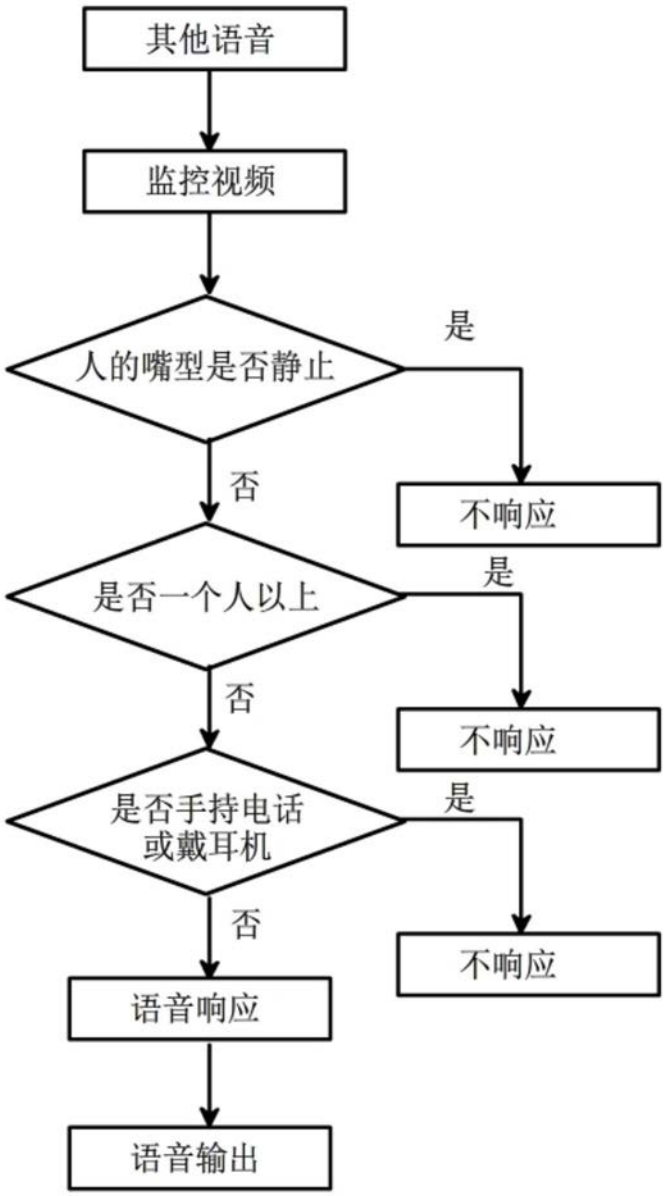

其他语音
监控视频
人的嘴型是否静止
是
否
不响应
是否一个人以上
是
否
不响应
是否手持电话
或戴耳机
是
否
不响应
语音响应
语音输出

图3

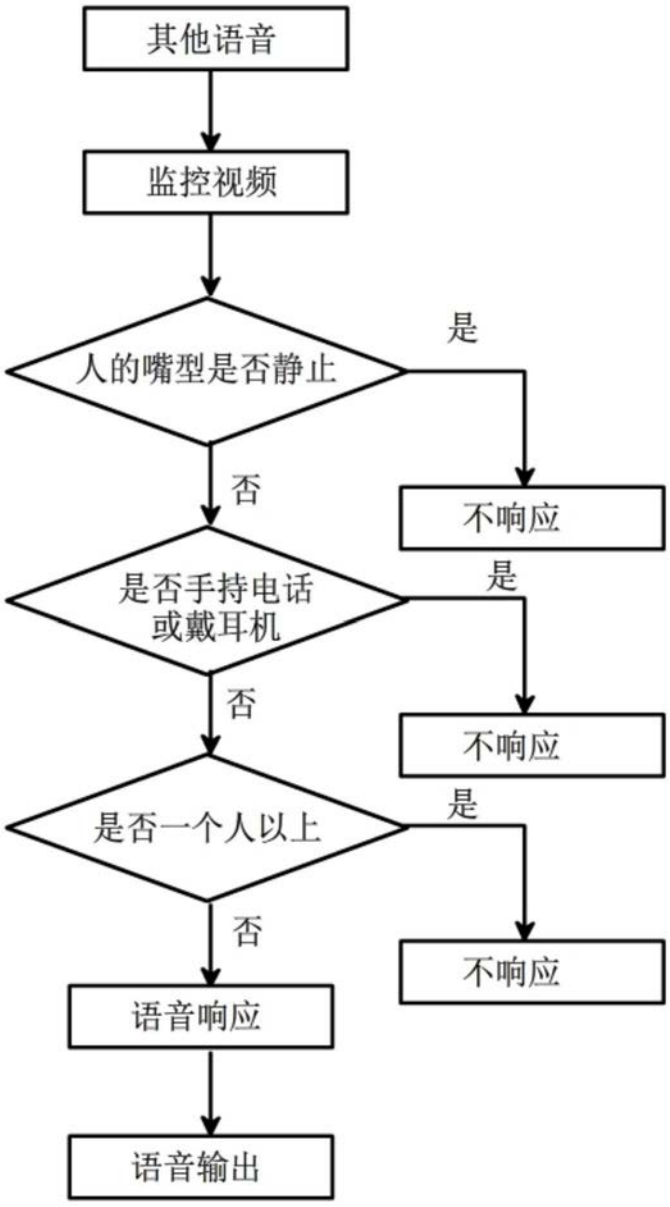

其他语音
监控视频
人的嘴型是否静止
是
否
不响应
是否手持电话
或戴耳机
是
否
不响应
是否一个人以上
是
否
不响应
语音响应
语音输出

图4

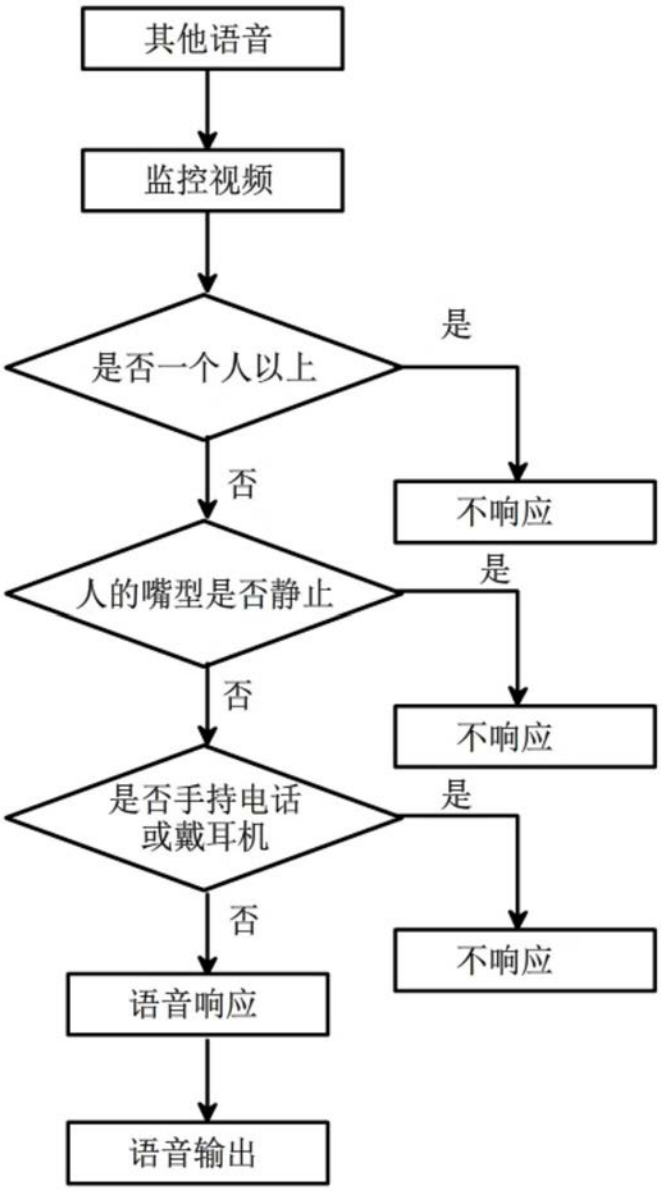

其他语音
监控视频
是否一个人以上
是
否
不响应
人的嘴型是否静止
是
否
不响应
是否手持电话
或戴耳机
是
否
不响应
语音响应
语音输出

图5

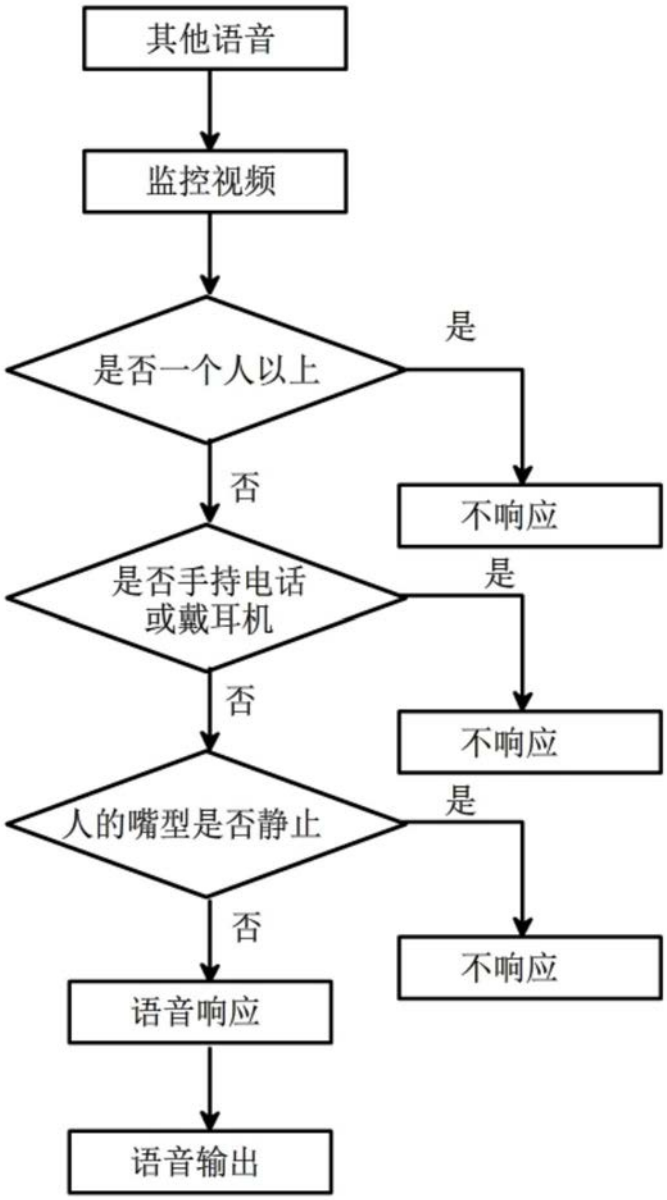

其他语音
监控视频
是否一个人以上
是
否
不响应
是否手持电话
或戴耳机
是
否
不响应
人的嘴型是否静止
是
否
不响应
语音响应
语音输出

图6

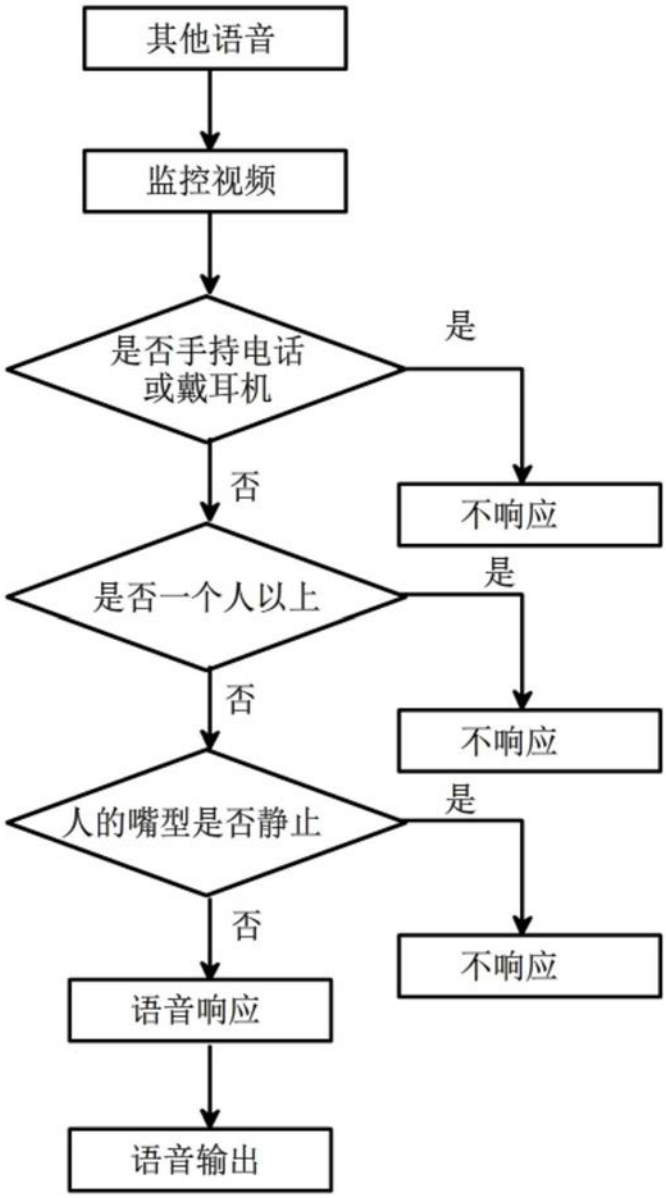

其他语音
监控视频
是否手持电话
或戴耳机
是
否
不响应
是否一个人以上
是
否
不响应
人的嘴型是否静止
是
否
不响应
语音响应
语音输出

图7

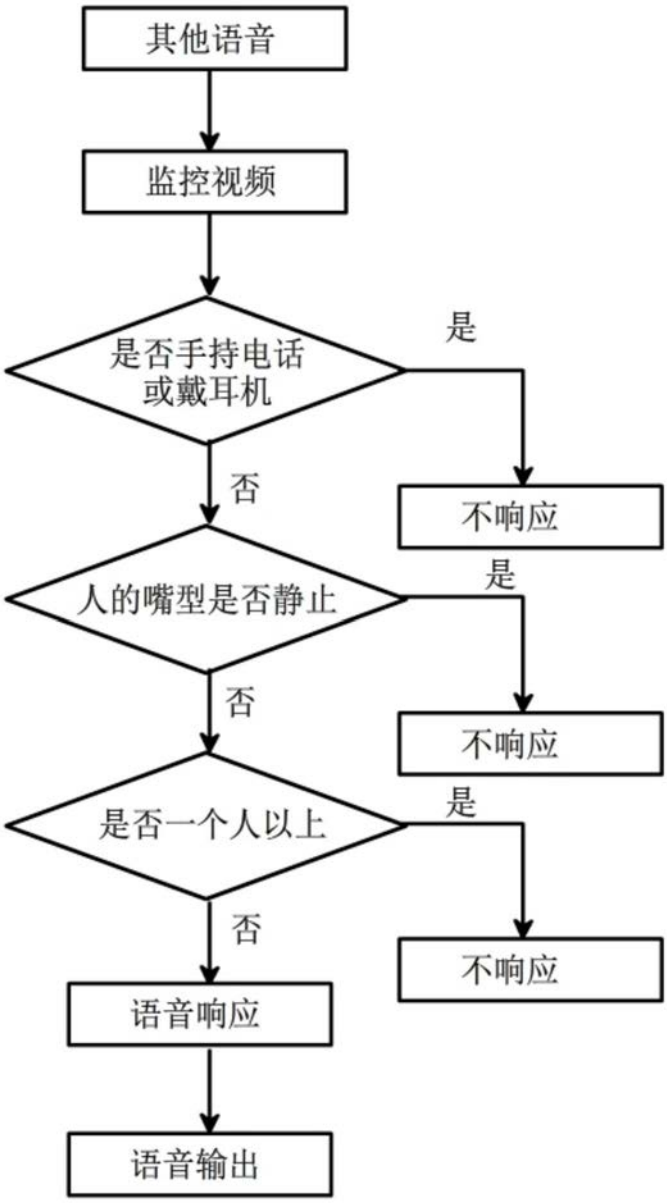

其他语音
监控视频
是否手持电话
或戴耳机
是
否
不响应
人的嘴型是否静止
是
否
不响应
是否一个人以上
是
否
不响应
语音响应
语音输出

图8