

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 10 月 8 日 (08.10.2020)



(10) 国际公布号
WO 2020/200081 A1

- (51) 国际专利分类号:
G10L 21/10 (2013.01)
- (21) 国际申请号: PCT/CN2020/081626
- (22) 国际申请日: 2020 年 3 月 27 日 (27.03.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910250929.2 2019年3月29日 (29.03.2019) CN
201910252003.7 2019年3月29日 (29.03.2019) CN
- (71) 申请人: 广州虎牙信息科技有限公司 (GUANGZHOU HUYA INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国广东省广州市番禺区钟村街(汉溪商业中心)泽溪街13号1401, Guangdong 511400 (CN)。
- (72) 发明人: 徐子豪 (XU, Zihao); 中国广东省广州市番禺区钟村街(汉溪商业中心)泽溪街13号1401, Guangdong 511400 (CN)。 吴昊 (WU,

Hao); 中国广东省广州市番禺区钟村街(汉溪商业中心)泽溪街13号1401, Guangdong 511400 (CN)。 马明参 (MA, Mingshen); 中国广东省广州市番禺区钟村街(汉溪商业中心)泽溪街13号1401, Guangdong 511400 (CN)。 李政 (LI, Zheng); 中国广东省广州市番禺区钟村街(汉溪商业中心)泽溪街13号1401, Guangdong 511400 (CN)。 周志颖 (ZHOU, Zhiying); 中国广东省广州市番禺区钟村街(汉溪商业中心)泽溪街13号1401, Guangdong 511400 (CN)。

- (74) 代理人: 北京超凡宏宇专利代理事务所(特殊普通合伙) (CHOFN INTELLECTUAL PROPERTY); 中国北京市海淀区北四环西路68号左岸工社1215-1218室, Beijing 100080 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,

(54) **Title:** LIVE STREAMING CONTROL METHOD AND APPARATUS, LIVE STREAMING DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 一种直播控制方法、装置、直播设备及存储介质

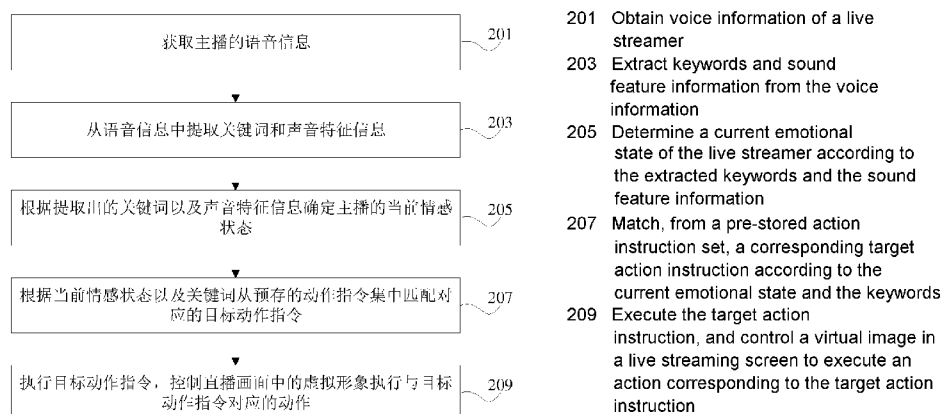


图 4

(57) **Abstract:** Embodiments of the present application relate to the technical field of Internet, and provide a live streaming control method and apparatus, a live streaming device, and a storage medium. Voice information of a live streamer is obtained, and the voice information is analyzed and processed, so that according to the processing result, a virtual image in a live streaming screen is controlled to execute an action matching the voice information, so as to improve the precision of controlling the virtual image and enable the virtual image in the live streaming screen and the live streaming content of the live streamer to have a high matching degree.

(57) **摘要:** 本申请实施例提供了一种直播控制方法、装置、直播设备及存储介质, 涉及互联网技术领域, 通过获取主播的语音信息, 并对该语音信息进行分析处理, 从而根据处理的结果, 控制直播画面中的虚拟形象执行与该语音信息相匹配的动作, 以提高对虚拟形象进行控制时的精度, 从而使直播画面中的虚拟形象和主播的直播内容具有较高的契合度。



WO 2020/200081 A1

JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK,
LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国 (除另有指明，要求每一种可提供的地区
保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告 (条约第21条(3))。

一种直播控制方法、装置、直播设备及存储介质

相关申请的交叉引用

本申请要求于 2019 年 3 月 29 日提交中国专利局的申请号为 201910250929.2、名称为“直播控制方法、装置、直播设备及可读存储介质”的中国专利申请的优先权，以及要求于 2019 年 3 月 29 日提交中国专利局的申请号为 201910252003.7、名称为“虚拟形象控制方法、虚拟形象控制装置和电子设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及互联网技术领域，具体而言，提供一种直播控制方法、装置、直播设备及存储介质。

背景技术

随着互联网技术的快速发展，直播已成为一种广受欢迎的网络互动方式。主播可以通过电子设备进行直播，观众可以通过电子设备观看直播。

在一些直播方案中，为了增加直播的趣味性，以及为了满足一些主播不愿在直播画面中出现的需求，可以通过在直播画面中展示主播的虚拟形象，通过该虚拟形象与观众互动。但是，在一些其他的采用虚拟形象进行直播的方案中，控制虚拟形象的方式比较单一。

发明内容

本申请的目的在于提供一种直播控制方法、装置、直播设备及存储介质，能够使直播画面中的虚拟形象和主播的直播内容具有较高的契合度。

为实现上述目的中的至少一个目的，本申请采用的技术方案如下：

本申请实施例提供了一种直播控制方法，应用于直播设备，所述方法包括：

获取主播的语音信息；

从所述语音信息中提取关键词和声音特征信息；

根据提取出的所述关键词以及所述声音特征信息确定所述主播的当前情感状态；

根据所述当前情感状态以及所述关键词从预存的动作指令集中匹配对应的目标动作指令；

执行所述目标动作指令，控制直播画面中的虚拟形象执行与所述目标动作指令对应的动作。

可选地，作为一种可能的实现方式，所述预存的动作指令集包括通用指令集以及与所述主播的当前虚拟形象对应的定制指令集，所述通用指令集存储有被配置成控制各个虚拟形象的通用动作指令，所述定制指令集存储有被配置成控制所述当前虚拟形象的定制动作指令。

可选地，作为一种可能的实现方式，所述根据所述当前情感状态以及所述关键词从预存的动作指令集中匹配动作指令的步骤，包括：

在所述预存的动作指令集中存在与所述当前情感状态和所述关键词关联的第一动作指令的情况下，将所述第一动作指令作为所述目标动作指令；

在所述预存的动作指令集中不存在所述第一动作指令的情况下，从所述预存的动作指令集中获取与所述当前情感状态对应的第二动作指令以及与所述关键词关联的第三动作指令；

根据所述第二动作指令和所述第三动作指令确定所述目标动作指令。

可选地，作为一种可能的实现方式，所述根据所述第二动作指令和所述第三动作指令确定所述目标动作指令的步骤，包括：

检测所述第二动作指令和所述第三动作指令是否存在联动关系；

若存在联动关系，则按照该联动关系指示的动作执行顺序对所述第二动作指令和所述第三动作指令进行合并，得到所述目标动作指令；

若不存在联动关系，则根据所述第二动作指令和所述第三动作指令各自的预设优先级

从所述第二动作指令和所述第三动作指令中选择一个作为所述目标动作指令。

可选地，作为一种可能的实现方式，所述方法还包括：

针对从所述语音信息中提取出的每个关键词，统计包含该关键词的目标语音信息的数量，以及根据最新获取的第一数量个所述目标语音信息确定的所述第一数量个目标动作指令；

若所述目标语音信息的数量达到第二数量，且所述第一数量个目标动作指令为相同指令，则在所述直播设备的内存中缓存该关键词与所述相同指令的对应关系；其中，所述第一数量不超过所述第二数量；

所述根据所述当前情感状态以及所述关键词从预存的动作指令集中匹配动作指令的步骤，包括：

从缓存的对应关系中查找是否存在所述关键词命中的对应关系；

若存在，则将所述命中的对应关系中记录的指令确定为所述目标动作指令；

若不存在，再执行根据所述当前情感状态以及所述关键词从预存的动作指令集中匹配动作指令的步骤。

可选地，作为一种可能的实现方式，所述方法还包括：

每间隔第一预设时长清空所述内存中缓存的所述对应关系。

可选地，作为一种可能的实现方式，针对所述预存的动作指令集中的每个动作指令，所述直播设备记录有该动作指令的最新执行时间；

所述执行所述目标动作指令的步骤，包括：

获取当前时间，判断所述当前时间与所述目标动作指令的最新执行时间的间隔是否超过第二预设时长；

若超过所述第二预设时长，再执行所述目标动作指令；

若没有超过所述第二预设时长，则从所述预存的动作指令集中查找与所述目标动作指令存在近似关系的其他动作指令替换所述目标动作指令，并执行替换后的目标动作指令。

本申请实施例还提供了一种直播控制方法，应用于直播设备，所述直播设备被配置成对直播画面中展示的虚拟形象进行控制，所述方法包括：

获取主播的语音信息；

对所述语音信息进行语音分析处理，得到对应的语音参数；

根据预设的参数转换算法将所述语音参数转换为控制参数，并根据该控制参数对所述虚拟形象的口型进行控制。

可选地，作为一种可能的实现方式，所述对所述语音信息进行语音分析处理，得到对应的语音参数的步骤，包括：

将所述语音信息进行分段处理，并提取分段后每一段语音信息中设定时长内的语音片段；

对提取的每个语音片段分别进行语音分析处理，得到每个语音片段对应的语音参数。

可选地，作为一种可能的实现方式，所述将所述语音信息进行分段处理，并提取分段后每一段语音信息中设定时长内的语音片段的步骤，包括：

按照每间隔设定时长提取所述语音信息中该设定时长内的语音片段。

可选地，作为一种可能的实现方式，所述将所述语音信息进行分段处理，并提取分段后每一段语音信息中设定时长内的语音片段的步骤，包括：

按照所述语音信息的连续性对该语音信息进行分段处理，并提取分段后每一段语音信息中设定时长内的语音片段。

可选地，作为一种可能的实现方式，所述对提取的每个语音片段进行语音分析处理，得到每个语音片段对应的语音参数的步骤，包括：

提取每个语音片段的振幅信息；

针对每个语音片段，根据该语音片段的振幅信息计算得到该语音片段对应的语音参数。

可选地，作为一种可能的实现方式，所述根据该语音片段的振幅信息计算得到该语音片段对应的语音参数的步骤，包括：

根据该语音片段的帧长信息和所述振幅信息按照归一化算法进行计算，得到该语音片段对应的语音参数。

可选地，作为一种可能的实现方式，所述控制参数包括所述虚拟形象的上下嘴唇之间的唇间距以及嘴角角度二者中的至少一个。

可选地，作为一种可能的实现方式，当所述控制参数包括所述唇间距时，该唇间距根据所述语音参数和预设的与所述虚拟形象对应的最大唇间距按照预设的参数转换算法计算得到；

当所述控制参数包括所述嘴角角度时，该嘴角角度根据所述语音参数和预设的与所述虚拟形象对应的最大嘴角角度按照预设的参数转换算法计算得到。

可选地，作为一种可能的实现方式，当所述控制参数包括所述唇间距时，所述最大唇间距根据预先获取的所述主播的唇间距设置；

当所述控制参数包括所述嘴角角度时，所述最大嘴角角度根据预先获取的所述主播的嘴角角度设置。

本申请实施例还提供了一种直播控制方法，应用于直播设备，所述方法包括：

获取主播的语音信息；

从所述语音信息中提取关键词和声音特征信息；

根据提取出的所述关键词以及所述声音特征信息确定所述主播的当前情感状态；

根据所述当前情感状态以及所述关键词从预存的动作指令集中匹配对应的目标动作指令；

执行所述目标动作指令，控制直播画面中的虚拟形象执行与所述目标动作指令对应的动作；

对所述语音信息进行语音分析处理，得到对应的语音参数；

根据预设的参数转换算法将所述语音参数转换为控制参数，并根据该控制参数对所述虚拟形象的口型进行控制。

本申请实施例还提供了一种直播设备，包括存储器、处理器及存储在所述存储器中并在所述处理器中被执行的机器可执行指令，所述机器可执行指令被所述处理器执行时上述的直播控制方法。

本申请实施例还提供了一种可读存储介质，其上存储有机器可执行指令，所述机器可执行指令被执行时实现上述的直播控制方法。

附图说明

图1为本申请实施例提供的直播系统的一种框架示意图；

图2为本申请实施例提供的直播界面的一种示意图；

图3为本申请实施例提供的直播设备的一种方框示意图；

图4为本申请实施例提供的直播控制方法的一种流程示意图；

图5为图4所示步骤207的子步骤的一种示意图；

图6为图4所示步骤207的子步骤的又一种示意图；

图7为图6所示步骤207-9的子步骤的一种示意图；

图8为本申请实施例提供的直播控制方法的另一种流程示意图；

图9为图8中步骤303包括的子步骤的一种流程示意图；

图10为图9中步骤303-3包括的子步骤的一种流程示意图；

图11为本申请实施例提供的20帧语音数据的示意图；

图12为本申请实施例提供的虚拟形象的唇间距和嘴角角度的示意图。

图中：11-直播服务器；12-第一终端设备；13-第二终端设备；100-直播设备；110-存储器；120-处理器。

具体实施方式

为使本申请实施例的目的、技术方案和效果等更加清楚，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。通常在此处附图中描述和示出的本申请实施例的组件可以以各种不同的配置来布置和设计。

因此，以下对在附图中提供的本申请的实施例的详细描述并非旨在限制要求保护的本申请的范围，而是仅仅表示本申请的选定一些实施例。基于本申请中的实施例，本领域普通技术人员在没有作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。

应注意到：相似的标号和字母在下面的附图中表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一步定义和解释。

请参照图 1，图 1 是本申请实施例提供的直播系统的一种示意图。该直播系统可以包括通过网络通信连接的直播服务器 11 和终端设备。其中，该终端设备可以是，但不限于，智能手机、个人数字助理、平板电脑、个人计算机（Personal Computer，PC）、笔记本电脑、虚拟实现终端、增强现实终端等。

在一些可能的实施例中，终端设备与直播服务器 11 的通信方式可以有多种。比如，终端设备中可以安装有客户端（例如，应用程序），该客户端可以与直播服务器 11 进行通信，进而使用直播服务器 11 提供的直播服务。

又比如，终端设备可以通过运行于第三方应用的程序与直播服务器 11 建立通信连接，进而使用直播服务器提供的直播服务。

再比如，终端设备可以通过浏览器登录到直播服务器 11，从而使用直播服务器 11 提供的直播服务。

在一种可能的实施例中，根据使用用户的不同，终端设备可以分为主播侧的第一终端设备 12 以及观众侧的第二终端设备 13。值得说明的是，当第一终端设备 12 的使用用户从主播变为观众时，该第一终端设备 12 也可以作为第二终端设备 13；当第二终端设备 13 的使用用户从观众变为主播时，该第二终端设备 13 也可以作为第一终端设备 12。

其中，第一终端设备 12 可以设置有音频采集器件，该音频采集器件可以被配置成采集主播的语音信息。音频采集器件可以内置于第一终端设备 12，也可以外接于第一终端设备 12，本申请实施例对于音频采集器件的配置方式不进行限定。

在一种可能的场景中，当主播采用虚拟形象进行直播时，第一终端设备 12 可以根据虚拟形象及采集的语音信息生成视频流，并发送给直播服务器 11 后，经由直播服务器 11 将该视频流发送给第二终端设备 13，实现基于虚拟形象的直播（如图 2 所示）。

在另一种可能的场景中，第一终端设备 12 可以将采集的语音信息直接发送给直播服务器 11，由直播服务器 11 根据该虚拟形象及语音信息生成视频流，并将该视频流发送给第二终端设备 13，实现基于虚拟形象的直播。

请参照图 3，图 3 是本申请实施例提供的直播设备 100 的一种方框示意图，在一些可能的实施例中，直播设备 100 可以是图 1 中示出的直播服务器 11 或第一终端设备 12。直播设备 100 可以包括存储器 110 以及处理器 120，存储器 110 和处理器 120 可以经由系统总线相互连接，以实现数据传输。存储器 110 可以存储有机器可执行指令，处理器 120 通过读取和执行机器可执行指令，可以实现本申请实施例下文描述的直播控制方法。

值得说明的是，图 3 所示的结构仅为示意。直播设备 100 还可以包括比图 3 所示更多或更少的组件，例如，当直播设备 100 为第一终端设备 12 时，直播设备 100 还包括上述的音频采集器件。或者，直播设备 100 可以具有与图 3 所示完全不同的配置。

请参照图 4，图 4 是本申请实施例提供的直播控制方法的一种流程示意图，该直播控制方法可以由图 3 所示的直播设备 100 执行。下面对该直播控制方法的各个步骤进行示意性描述。

步骤 201，获取主播的语音信息。

在一些可能的实施例中，当直播设备 100 是例如图 1 中的第一终端设备 12 时，则直播设备 100 可以通过音频采集器件（如，内置麦克风或外接话筒等）实时采集主播的语音信息。而在另一些可能的实施例中，当直播设备 100 是例如图 1 中的直播服务器 11 时，则直播设备 100 可以接收第一终端设备 12 采集并发送的语音信息，例如，从第一终端设备 12 推流的视频流中获得语音信息。

步骤 203，从语音信息中提取关键词和声音特征信息。

在一些可能的实施例中，直播设备 100 获取到主播的语音信息后，可以并行地从语音信息中提取关键词和声音特征信息，也可以按照指定的先后顺序，依次提取关键词和声音特征信息。可以理解的是，本申请实施例对于提取关键词和声音特征信息的先后顺序没有限制。

其中，在一些可能的场景中，上述的声音特征信息可以是音调信息、振幅信息、频率信息、低频信号图谱等；本申请实施例对提取声音特征信息的具体算法没有限制，只要能够提取出相应的声音特征信息即可。

另外，在一些可能的实施例中，直播设备 100 从语音信息中提取关键词的方式可以有多种。例如，可以基于预设的关键词库从该语音信息中提取关键词。该关键词库可以包括：预设的被配置成指示主播的情感状态的关键词，比如，“开心”、“高兴”、“快乐”、“悲伤”、“难过”、“发愁”、“兴奋”、“哈哈”、“哭”等；以及预设的被配置成指示主播的待执行动作的关键词，比如，“再见”（可以被配置成指示挥手等动作）、“兴奋”（可以被配置成指示手舞足蹈等动作）、“敬礼”、“转身”等。其中，可以理解的是，该关键词库可以存储在直播设备 100 中，也可以存储在第三方服务器中。

在一些可能的实施例中，直播设备 100 可以识别上述的语音信息，并检测识别结果中是否包含有关键词库中的关键词；当检测到识别结果中包含有该关键词库中的关键词时，直播设备 100 则可以提取出该关键词。

另外，在一些其他可能的实施例中，直播设备 100 还可以通过神经网络模型对该语音信息对应的语句进行分词，得到多个词语。针对得到的每个词语，通过该神经网络模型对每一词语进行识别，以获得每个词语的类型，即：识别每个词语是否指示情感状态或是否指示动作；当该词语指示情感状态或者是指示动作时，直播设备 100 则可以将该词语作为提取出的关键词。

步骤 205，根据提取出的关键词以及声音特征信息确定主播的当前情感状态。

在一些可能的场景中，直播设备 100 或与直播设备 100 通信的第三方服务器中可以存储有多个对应关系，比如可以是不同关键词与不同情感状态的对应关系，也可以是不同声音特征信息与不同情感状态的对应关系。

于是，在一些可能的实施例中，直播设备 100 可以根据该对应关系，以及提取出的关键词和声音特征信息确定主播的当前情感状态。

值得说明的是，在一些可能的场景中，对于同一语音信息提取的关键词和声音特征信息，当基于关键词确定的情感状态与基于声音特征信息确定的情感状态是相反的两种情感状态（如，“高兴”和“悲伤”）时，直播设备 100 可以基于该语音信息的低频信号图谱确定主播发音时的生理参数信息（例如，肌肉紧张程度、是否兴奋等），并依据该生理参数信息确定主播的心理状态信息，从而可以依据该生理参数信息从两种情感状态中选择一种作为主播的当前情感状态。

另外，在一些其他的实施例中，直播设备 100 还可以通过神经网络模型来实现步骤 205。例如，可以获取多个主播的多条语音信息；从每条语音信息中提取关键词和声音特征信息形成一个样本，并将主播发出该条语音时的实际情感状态标注到该样本中，从而形成一个样本集；再采用该样本集对预先建立的神经网络模型进行训练，得到一训练好的神经网络模型。亦或者，该神经网络模型可以包括第一神经网络子模型以及第二神经网络子模型，

第一神经网络子模型可以被配置成识别关键词，第二神经网络子模型可以被配置成识别声音状态，第一神经网络子模型和第二神经网络子模型可以并行识别。

如此，在执行步骤 205 时，直播设备 100 可以将提取出的关键词和声音特征信息输入训练好的神经网络模型，即可获得主播的当前情感状态。

值得说明的是，上述两种实施方式仅为示例，在本申请实施例其他一些可能的实现方式中，步骤 205 还可以通过其他方式实现，本申请实施例对于步骤 205 的实现方式不进行限定。

步骤 207，根据当前情感状态以及关键词从预存的动作指令集中匹配对应的目标动作指令。

在一些可能的场景中，预存的动作指令集可以存储在直播设备 100 或者与直播设备 100 通信连接的第三方服务器中。对应地，直播设备 100 或与直播设备 100 通信连接的第三方服务器还可以存储有该预存的动作指令集中的各个动作指令与情感状态和关键词的关联关系。

在一些可能的场景中，动作指令可以分为两类，一类是可以应用于各种虚拟形象的动作指令，在此可以称为“通用动作指令”；另一类是仅能应用于一些特定虚拟形象的动作指令，通过该动作指令可以实现特定的直播特效，在此可以将这种动作指令称为“定制动作指令”。

对应地，该预存的动作指令集可以包括存储有通用动作指令的通用指令集以及存储有定制动作指令的定制指令集。在一种可能的实施例中，当主播使用特定虚拟形象时，第一终端设备 12 可以下载并保存该特定虚拟形象对应的定制指令集。在另一种可能的实施例中，可以针对该定制指令集设置收费服务，当主播选用该特定虚拟形象且支付相应的费用时，第一终端设备 12 可以下载并保存该特定虚拟形象对应的定制指令集。

可选地，作为一种可能的实现方式，步骤 207 可以通过以下过程实现：

在预存的动作指令集中存在与当前情感状态和关键词关联的第一动作指令的情况下，将第一动作指令作为目标动作指令；

在预存的动作指令集中不存在第一动作指令的情况下，从预存的动作指令集中获取与当前情感状态对应的第二动作指令以及与关键词关联的第三动作指令；

根据第二动作指令和第三动作指令确定目标动作指令。

其中，第一动作指令既可以与当前情感状态关联，又可以与关键词关联，其与主播的说话内容匹配程度较高，因此在存在第一动作指令的情况下可以直接将第一动作指令作为目标动作指令。

示例性地，直播设备 100 可以通过不同的执行逻辑实现上述步骤 207 的过程。例如，作为一种可能的实现方式，步骤 207 的过程可以通过图 5 所示的步骤实现。

步骤 207-1，从预存的动作指令集中查找是否存在与当前情感状态和关键词关联的第一动作指令。若是，则执行步骤 207-2；若否，则执行步骤 207-3。

在一种可能的实现方式中，直播设备 100 可以以当前情感状态和关键词为检索索引，查找对应的动作指令，查找到的动作指令即为第一动作指令。

步骤 207-2，将第一动作指令作为目标动作指令。

步骤 207-3，分别从预存的动作指令集中查找是否存在与当前情感状态关联的第二动作指令以及与关键词关联的第三动作指令。

其中，在一种可能的实现方式中，直播设备 100 可以以当前情感状态为检索索引从预存的动作指令集中查找动作指令，查找到的动作指令即为第二动作指令。直播设备 100 可以以关键词为检索索引从预存的动作指令集中查找动作指令，查找到的动作指令即为第三动作指令。

步骤 207-4，若存在第二动作指令和第三动作指令，则根据第二动作指令和第三动作指令确定目标动作指令。

又如，在另一种可能的实施例中，上述步骤 207 的过程还可以通过图 6 所示的步骤实现。

步骤 207-6，从预存的动作指令集中查找是否存在与当前情感状态关联的第二动作指令以及与关键词关联的第三动作指令。

步骤 207-7，判断第二动作指令和第三动作指令是否为相同指令。若是，则执行步骤 207-8；若否，则执行步骤 207-9。

步骤 207-8，将该相同指令作为目标动作指令。

其中，当第二动作指令和第三动作指令为相同指令时，该相同指令可以充当本申请实施例中的第一动作指令。

步骤 207-9，根据第二动作指令和第三动作指令确定目标动作指令。

在一些可能的实现方式中，直播设备 100 在执行根据第二动作指令和第三动作指令确定目标动作指令的步骤（例如，上述的步骤 207-4 或步骤 207-9）时，可以通过例如如图 7 所示的步骤实现。

步骤 207-9a，检测第二动作指令和第三动作指令是否存在联动关系。若是，则执行步骤 207-9b；若否，则执行步骤 207-9c。

在一些可能的实施例中，直播设备 100 可以存储有预存的动作指令集的各个动作指令间的关联关系。该关联关系的记录方式可以有多种，本申请实施例对该关联关系的记录方式不进行限制。例如，该关联关系可以以一条数据记录的形式进行保存，每条数据记录包括相应动作指令的标识信息以及被配置成指示关联关系类型的标志位。

例如，一条数据记录 a 可以被配置成表示动作指令 1 和 2 的关联关系，则数据记录 a 可以包括动作指令 1 和 2 各自的标识信息（例如，预设的编号信息）。关联关系类型例如可以是联动关系或近似关系，比如，当标志位为 1 时，可以表示数据记录中记录的动作指令间具有联动关系；当标志位为 0 时，可以表示数据记录中记录的动作指令间具有近似关系。应当理解，上述利用 0 和 1 分别对联动关系和近似关系进行表示仅为示意，在其他一些可能的实施例中，联动关系和近似关系也可以用其他值或者字符进行表示，本申请实施例对联动关系和近似关系各自的标识方式不进行限制。

其中，在一些可能的实施例中，具有联动关系的至少两个动作指令可以按照一定顺序合并成一个动作指令，比如，当实现“大笑”的动作指令和实现“跳舞”的动作指令具有联动关系时，这两个动作指令可以合并成一个动作指令，可以通过合并的动作指令一次性控制主播的虚拟形象进行“大笑”和“跳舞”。

可选地，在一些可能的实施例中，对于具有联动关系的至少两个动作指令，可以在相应的数据记录中设置该至少两个动作指令中各个动作指令的执行顺序。

具有近似关系的至少两个动作指令是指被配置为实现类似动作的指令，比如被配置为实现“大笑”的动作指令和被配置为实现“微笑”的动作指令可以认为是近似的动作指令，可以建立这“大笑”和“微笑”两个动作指令的近似关系。

基于以上配置，直播设备 100 可以查找同时记录有第二动作指令和第三动作指令的标识信息的第一数据记录。如果查找到，则根据第一数据记录中的标志位的值确定第二动作指令和第三动作指令的关联关系类型，若该标志位的值指示的关联关系类型为联动关系，则可以确定第二动作指令和第三动作指令之间存在联动关系。如果该标志位的值指示的关联关系不是联动关系，或者没有查找到第一数据记录，则可以确定第二动作指令和第三动作指令之间不存在联动关系。

步骤 207-9b，按照该联动关系指示的动作执行顺序对第二动作指令和第三动作指令进行合并，得到目标动作指令。

其中，在一些可能的实施例中，第一数据记录中设置的执行顺序可以充当联动关系指示的动作执行顺序。

步骤 207-9c，根据第二动作指令和第三动作指令各自的预设优先级从第二动作指令和

第三动作指令中选择一个作为目标动作指令。

在一些可能的实施例中，可以针对预存的动作指令集中的各个动作指令分别设置优先级。如此，直播设备 100 可以根据实际需要第二动作指令和第三动作指令中选取优先级较高的一个或者优先级较低的一个作为目标动作指令。如果第二动作指令和第三动作指令的优先级相同，直播设备 100 则可以随机选取一个作为目标动作指令。

可选地，在一些可能的实施例中，为了提高动作指令的匹配速度，该直播控制方法还可以包括以下步骤。

第一，针对从语音信息中提取出的每个关键词，统计包含该关键词的目标语音信息的数量，以及根据最新获取的第一数量个目标语音信息确定的第一数量个目标动作指令。

第二，若目标语音信息的数量达到第二数量，且第一数量个目标动作指令为相同指令，则在直播设备的内存中缓存该关键词与相同指令的对应关系。

其中，第一数量不超过第二数量。

下面通过一个示例对上述两个步骤进行阐述。假设：

第一数量为 2，第二数量为 3；

第一次获取到语音信息 1，从中提取出关键词 aa、bb 和 cc，且按照图 4 所示步骤，根据语音信息 1 确定了目标动作指令 t2；

第二次获取到语音信息 2，从中提取出关键词 aa 和 dd，且按照图 4 所示步骤，根据语音信息 2 确定了目标动作指令 t1；

第三次获取到语音信息 3，从中提取出关键词 bb，且按照图 4 所示步骤，根据语音信息 3 确定目标动作指令 t3；

第四次获取到语音信息 4，从中提取出关键词 aa 和 bb，按照图 4 所示步骤，根据语音信息 4 确定目标动作指令 t1；

第五次获取到语音信息 5，从中提取出关键词 cc，且按照图 4 所示步骤，根据语音信息 5 确定目标动作指令 t2。

在以上示例中，针对关键词 aa，对应的目标语音信息有语音信息 1、语音信息 2 和语音信息 4，即；包含关键词 aa 的目标语音信息的数量为 3，达到了第二数量 3。其中，分别基于语音信息 1、语音信息 2 和语音信息 4 确定的目标动作指令中有两个相同，均为 t1，即到达了第一数量 2。因此，可以建立关键词 aa 和动作指令 t1 的对应关系，并将该对应关系缓存到直播设备 100 的内存中。当下一次再获取到包含关键词 aa 的语音信息时，可以直接将动作指令 t1 确定为目标动作指令。

基于以上描述，在执行步骤 207-3 之后，可以先从缓存的对应关系中查找是否存在关键词命中的对应关系；若存在，则将命中的对应关系中记录的指令确定为目标动作指令；若不存在，再执行步骤 207-4。

考虑到主播在不同时间段内使用同一关键词表达的含义可能发生改变，因此，直播设备 100 可以每间隔第一预设时长清空内存中缓存的对应关系。如此，可以确保直播设备 100 中缓存的对应关系与主播最近的用词习惯相适应。

请再次参照图 4，在确定目标动作指令后，直播设备 100 可以执行步骤 209。

步骤 209，执行目标动作指令，控制直播画面中的虚拟形象执行与目标动作指令对应的动作。

在一些可能的实施例中，直播设备 100 可以根据目标动作指令对虚拟形象进行处理，从而生成相应的直播视频流，并将直播视频流直接或间接地发送给第二终端设备 13。

可选地，在一些可能的实施例中，为了增加趣味性，避免主播的虚拟形象在短时间执行重复的动作，在执行步骤 209 之前，可以先执行以下步骤。

首先，获取当前时间，判断当前时间与目标动作指令的最新执行时间的间隔是否超过第二预设时长；若超过第二预设时长，则执行步骤 209。

其中，针对预存的动作指令集中的每个动作指令，直播设备 100 可以记录有该动作指

令的最新执行时间。值得说明的是，对于未被执行过的动作指令，直播设备 100 记录的最新执行时间可以为空，或为预设的默认值。

然后，若没有超过第二预设时长，直播设备 100 则从预存的动作指令集中查找与目标动作指令存在近似关系的其他动作指令替换目标动作指令，并执行替换后的目标动作指令。

其中，直播设备 100 可以从存储的数据记录中查找包含有目标动作指令的标识信息的第二数据记录，再从查找到的第二数据记录中获取不同于目标动作指令的标识的其他标识信息，采用其他标识信息指示的动作指令来替换目标动作指令。

另外，针对一些特定的场景，还可以通过对虚拟形象的一些特定部位进行控制，从而使虚拟形象的一些特定部位能够执行一些与语音信息相对应的动作，以提高对虚拟对象的控制精度。

比如，请参阅图 8，图 8 是本申请实施例提供的另一种直播控制方法的一种流程示意图，该直播控制方法能够对直播画面中展示的虚拟形象进行控制。其中，该直播控制方法有关的流程所定义的方法步骤可以由上述的直播设备 100 实现。下面将对图 8 所示的具体流程进行示例性阐述。

步骤 301，获取主播的语音信息。

在一些可能的实施例中，直播设备 100 可以通过语音采集设备（比如手机的话筒或连接的麦克风等），实时获取主播的语音信息。比如，在一种可能的示例中，若直播设备 100 为主播使用的终端设备，可以直接通过连接的麦克风或自带的话筒等语音采集设备获取主播的语音信息。又比如，在另一种可能的示例中，若直播设备 100 为后台服务器，主播使用的终端设备获取到主播的语音信息之后，可以将该语音信息发送至该后台服务器。

步骤 303，对语音信息进行语音分析处理，得到对应的语音参数。

在一些可能的实施例中，直播设备 100 通过步骤 301 获取到语音信息之后，可以对该语音信息进行分析处理，以得到对应的语音参数。

其中，在一些可能的实施例中，为保证分析得到的语音参数具有较高的精确度，在执行步骤 303 之前，还可以对语音信息进行预处理，该预处理的方式可以举例介绍如下。

首先，直播设备 100 可通过重采样的方式将获得语音信息转换为窄带的语音信息；然后，通过带通滤波器对获得语音信息进行滤波处理得到频率属于该带通滤波器的通带的语音信息，其中，该带通滤波器的通带一般是基于人的声音的基频和共振峰确定；最后，通过音频降噪算法对获得的用户的音频数据进行噪声滤除处理。

需要说明的是，考虑到人的声音的基频一般属于（90，600）Hz，可以设置一个截止频率为 60Hz 的高通滤波器；然后，就基频和共振峰（共振峰可以包括第一共振峰、第二共振峰以及第三共振峰）可以知道，人的声音的主要频率一般处于 3kHz 以下；因此，可以设置一个截止频率为 3kHz 的低通滤波器。也就是说，前述的带通滤波器可以由一个截止频率为 60Hz 的高通滤波器和一个截止频率为 3kHz 的低通滤波器组成，以使频率不属于（60，3000）Hz 的语音信息能够被有效地滤除，有效地避免了环境噪声对语音分析处理造成干扰的问题。

步骤 305，根据预设的参数转换算法将语音参数转换为控制参数，并根据该控制参数对虚拟形象的口型进行控制。

在一些可能的实施例中，直播设备 100 通过步骤 303 得到语音参数之后，可以基于预设的参数转换算法将该语音信息转换为对应的控制参数，然后，基于该控制参数对虚拟形象的口型进行控制。

通过上述方法，直播设备 100 可以基于主播的语音信息对虚拟形象的口型进行控制，使得在直播时播放的语音信息与虚拟形象的口型具有较高的一致性，从而提高了虚拟对象的控制精度，能够有效地提高了用户的体验。并且，由于虚拟形象的口型是基于语音信息确定的，也就是说，不同的语音信息对应的口型不同，基于口型的变化还可以提高虚拟形象直播时的灵动性，以提高直播的趣味性。

需要说明的是，直播设备 100 执行步骤 303 以对语音信息进行分析处理的具体方式不

受限制，可以根据实际应用需求进行选择。例如，结合图 9，作为一种可能的实施例，步骤 303 可以包括步骤 303-1 和步骤 303-3，其中，步骤 303 包括的步骤内容可以为如下所述。

步骤 303-1，将语音信息进行分段处理，并提取分段后每一段语音信息中设定长度内的语音片段。

在一种可能的实施例中，直播设备 100 可以基于预设规则对语音信息进行分段处理，得到至少一段语音信息。然后，针对每一段语音信息提取该段语音信息中设定长度内的语音片段，得到至少一个语音片段。

其中，该设定长度既可以是在时间上的长度，例如，可以是 1s、2s、3s 等；也可以是在其它维度上的长度，例如，可以是基于对应的字数（如 2 个字、3 个字、4 个字等）。

步骤 303-3，对提取的每个语音片段分别进行语音分析处理，得到每个语音片段对应的语音参数。

在一种可能的实施例中，直播设备 100 通过步骤 303-3 得到至少一个语音片段之后，可以对每个语音片段分别进行语音分析处理，得到每个语音片段对应的语音参数。对应地，直播设备 100 对每个语音片段进行分析处理后，可以得到至少一个语音参数。

其中，直播设备 100 在执行步骤 303-1 以进行语音信息分段处理的具体方不受限制，可以根据实际应用需求进行选择。比如，作为一种可能的实现方式，步骤 303-1 可以为：按照每间隔设定长度提取语音信息中该设定长度内的语音片段。例如，获取的语音信息可以为 1s、设定长度可以为 0.2s，对应地，通过分段处理后，可以得到长度为 0.2s 的 5 个语音片段。又例如，获取的语音信息可以为 20 个字、设定的长度可以为 5 个字，对应地，通过分段处理后，可以得到长度为 5 个字的 4 个语音片段。

并且，作为另一种可能的实现方式，直播设备 100 可以基于语音信息的连续性进行分段处理。比如，步骤 303-1 可以为：按照语音信息的连续性对该语音信息进行分段处理，并提取分段后每一段语音信息中设定长度内的语音片段。

也就是说，在获取到语音信息之后，直播设备 100 可以对该语音信息进行识别，以判断该段语音信息中是否存在停顿（进行停顿判断的方式可以是，对语音信息的波形进行分析，若波形存在间断且间断的时长大于预设时长，可以判定存在停顿）。例如，若语音信息为“今天的直播结束了，我们明天……”，那么，通过对该语音信息进行识别，可以判断出在“，”的位置出现了停顿，因此，可以得到一段语音信息，为“今天的直播结束了”。然后，可以在该段语音信息中提取长度为设定长度的语音片段。其中，该设定长度的具体大小不受限制，可以根据实际应用需求进行选择。

例如，作为一种可能的实现方式，设定长度可以小于对应段语音信息的长度（例如，一段语音信息的长度为 0.8s，设定长度可以为 0.6s；或者说，一段语音信息的长度为 8 个字，设定长度可以为 6 个字），以使得到的语音片段的数据量小于主播的语音信息的数据量，从而降低执行步骤 303-3 和步骤 305 时数据的处理量或运算量，进而有效地保证虚拟形象的直播具有较高的实时性。并且，由于数据的处理量降低，还可以降低对直播设备 100 处理性能的要求，从而提高该虚拟形象控制方法的适应能力。

需要说明的是，直播设备 100 在基于连续性对语音进行分段处理时，针对各段语音信息可以配置不同的设定长度。比如，作为一种可能的实现方式，可以基于各段语音信息的长度，分别配置对应的设定长度。例如，若一段语音信息的长度为 0.6s（或 6 个字），配置的设定时长可以为 0.3s（或 3 个字）；若一段语音信息的长度为 0.4s（或 4 个字），配置的设定时长可以为 0.2s（或 2 个字）。又例如，若一段语音信息的长度为 0.6s（或 6 个字），配置的设定时长可以为 0.5s（或 5 个字）；若一段语音信息的长度为 0.4s（或 4 个字），配置的设定时长可以为 0.3s（或 3 个字）。

并且，在设定长度的长度配置之后，该设定长度的起始位置（如起始时刻或起始字）或截止位置（截止时刻或截止字）可以不受限制，可以根据实际应用需求进行配置。例如，作为一种可能的实现方式，可以在一段语音信息中，提取起始时刻或截止时刻为任意时刻

的一个语音片段。

又例如，在另一种可能的实现方式中，可以在一段语音信息中，提取截止时刻为该段语音信息的截止时刻的一个语音片段。例如，一段语音信息的起始时刻为“15h: 40min: 10.23s”、截止时刻为“15h: 40min: 10.99s”，设定时长为 0.50s，那么提取的语音片段的截止时刻为“15h: 40min: 10.99s”、起始时刻为“15h: 40min: 10.49s”。

如此，通过上述设置，可以保证每一段语音信息在靠近截止时刻的内容和对应的口型具有一致性，使得在降低数据的运算量的同时，还能使观众不容易发现存在语音信息与口型不对应的情形，更逼真的还原主播的口型，以有效地保证观众具有较高的体验度。例如，在上述的示例“今天的直播结束了，我们明天.....”中，若能保证“结束了”对应的语音和口型具有一致性，就能使观众不容易发现甚至于可以忽略“今天的直播”对应的语音和口型存在不一致的问题，进而使得观众认为直播时播放的语音与虚拟形象的口型具有较高的一致性。

需要说明的是，直播设备 100 在基于连续性对语音信息进行分段处理时，在检测到有停顿之后，若再次检测到主播的语音信息，如在上述的示例“今天的直播结束了，我们明天.....”中，在停顿之后，又再次检测到“我们明天.....”的语音信息，此时，为使虚拟形象的口型与播放的语音更加一致，还可以在该语音信息中提取预设长度的语音片段。例如，可以以该语音信息的起始时刻为起始时刻提取预设长度（如 0.4s）的一语音片段，或者，可以以该语音信息的第一个字为起始字提取预设长度（如 2 两个字）的一语音片段。

也就是说，针对得到的每一段语音信息，可以分别得到该段语音信息中首部和尾部的两个语音片段，并基于该两个语音片段对虚拟形象的口型进行控制。例如，在上述的示例“今天的直播结束了”示例中，可以提取“今天”和“结束了”两个语音片段，使得该两个语音片段对应的内容和口型具有一致性，以使观众认为“今天的直播结束了”对应的内容和口型都具有一致性。

需要说明的是，在上述示例中，若针对一段语音信息分别提取两个语音片段，该两个语音片段的长度可以相同，也可以不同。其中，在该两个语音片段的长度不同时，可以是尾部的语音片段的长度大于首部的语音片段的长度。

另外，直播设备 100 执行步骤 303-3 以进行语音分析处理的具体方式也不受限制，可以根据实际应用需求进行选择。例如，作为一种可能的实现方式，直播设备 100 可以基于语音信息中的振幅信息和/或频率信息进行分析处理。

比如，直播设备 100 在执行步骤 133 时，可以基于振幅信息进行语音分析处理。示例性地，结合图 10，作为一种可能的实现方式，步骤 303-3 可以包括步骤 303-3a 和步骤 303-3b，具体内容可以为如下所述。

步骤 303-3a，提取每个语音片段的振幅信息。

在一种可能的实施例，直播设备 100 通过步骤 303-1 得到至少一个语音片段之后，可以先提取每一个语音片段的振幅信息。

步骤 303-3b，针对每个语音片段，根据该语音片段的振幅信息计算得到该语音片段对应的语音参数。

在一种可能的实施例，通过步骤 303-1 得到每个语音片段的振幅信息之后，直播设备 100 可以基于该振幅信息计算每个语音片段对应的语音参数。其中，该语音参数可以是 (0, 1) 区间的任意一个值，也就是说，直播设备 100 可以基于归一化算法对得到的振幅信息进行处理，以得到对应的语音参数。

示例性地，直播设备 100 可以根据语音片段的帧长信息和振幅信息按照归一化算法进行计算，得到该语音片段对应的语音参数。

其中，需要说明的是，基于提取语音片段的方式不同，得到的语音片段的长度一般不同，计算每个语音片段的语音参数的方式一般也可以不同。例如，若一个语音片段较长，可以针对该语音片段中的每一帧语音数据分别计算得到一个语音参数；若一个语音片段较短，可以作为一帧语音数据，因此，可以基于该帧语音数据计算得到一个语音参数，作为

该语音片段对应的语音参数。

也就是说，直播设备 100 可以针对每一个帧语音数据，基于该帧语音数据的帧长信息和振幅信息按照归一化算法，可以计算得到一个属于 $(0, 1)$ 区间的数值，该数值可以作为该帧语音数据对应的语音参数。例如，在上述的示例“今天的直播结束了，我们明天……”中，直播设备 100 可以提取 20 帧语音数据，然后，对每一帧语音数据的振幅进行归一化计算，得到 20 个数值，作为该 12 帧语音数据对应的 20 个语音参数（如图 11 所示）。

其中，需要说明的是，归一化算法的具体内容不受限制，可以根据实际应用需求进行选择。例如，直播设备 100 可以先计算一帧语音数据中各个时刻的振幅信息的平方和，然后，基于该帧语音数据的帧长计算该帧语音数据的振幅信息的平方和均值，并对该平方和均值进行开平方运算得到对应的语音参数。

可选地，直播设备 100 执行步骤 305 以将语音参数转换为控制参数的具体方式也不受限制，可以根据实际应用需求进行选择。也就是说，在一些可能的实现方式中，参数转换算法的具体内容不受限制。例如，根据控制参数的具体内容不同，参数转换算法的具体内容也可以不同。

示例性地，作为一种可能的实现方式，控制参数可以包括，但不限于虚拟形象的上下嘴唇之间的唇间距以及嘴角角度二者中的至少一个。

比如，当控制参数包括唇间距时，该唇间距可以根据语音参数和预设的与虚拟形象对应的最大唇间距按照预设的参数转换算法计算得到。当控制参数包括嘴角角度时，该嘴角角度可以根据语音参数和预设的与虚拟形象对应的最大嘴角角度按照预设的参数转换算法计算得到。

例如，若最大唇间距为 5cm、归一化后的语音参数为 0.5，控制参数可以包括 $0.5 \times 5 = 2.5\text{cm}$ ，也就是说，此时可以控制虚拟形象的上下嘴唇之间的唇间距为 2.5cm（如图所示的 h）。同理，若最大嘴角角度为 120° 、归一化后的语音参数为 0.5，控制参数可以包括 $0.5 \times 120 = 60^\circ$ ，也就是说，此时可以控制虚拟形象的嘴角角度为 60° （如图 12 所示的 a）。

可选地，在控制参数包括唇间距时，最大唇间距的具体数值可以不受限制，可以根据实际应用需求进行设置。比如，作为一种可能的实现方式，最大唇间距可以基于主播的唇间距设置。

例如，针对 A 主播，经过测试，若该主播的最大唇间距为 5cm，可以将该主播对应的虚拟形象的最大唇间距设置为 5cm；针对 B 主播，经过测试，若该主播的最大唇间距为 6cm，可以将该主播对应的虚拟形象的最大唇间距设置为 6cm。

同理，在控制参数包括嘴角角度时，最大嘴角角度的具体数值可以不受限制，可以根据实际应用需求进行设置。比如，作为一种可能的实现方式，最大嘴角角度可以基于主播的嘴角角度设置。

例如，针对 A 主播，经过测试，若该主播的最大嘴角角度为 120° ，可以将该主播对应的虚拟形象的最大嘴角角度设置为 120° ；针对 B 主播，经过测试，若该主播的最大嘴角角度为 135° ，可以将该主播对应的虚拟形象的最大嘴角角度设置为 135° 。

可见，通过上述设置，可以使得虚拟形象的口型可以与对应主播的实际口型具有较高的一致性，从而在直播时实现更为逼真的形象展示。并且，由于不同的主播一般具有不同的最大唇间距和最大嘴角角度，因此，不同的主播对应的虚拟形象的最大唇间距和最大嘴角角度也会不同，使得观众在观看不同的主播对应的虚拟形象的直播时，可以看到不同的口型（最大唇间距和/或最大嘴角角度不同），使得虚拟形象具有更高的灵动性，进而提高直播的趣味性。

另外，本申请实施例还提供了一种直播控制方法，该直播控制方法可以包括例如图 4 以及图 8 各自所述方法的所有步骤，使得在利用该直播控制方法对虚拟形象进行控制时，不仅能控制直播画面中的虚拟形象执行与目标动作指令对应的动作，还能够对虚拟形象的口型进行控制，从而提高对虚拟形象进行控制的精度。

其中，需要说明的是，为描述的方便与简洁，前述直播控制方法的各个具体流程与步骤的实现方式，请具体参阅上述针对例如图 4 以及图 8 的实施例，本申请实施例对此不再进行赘述。

并且，在本申请实施例所提供的一些示意性实施例中，应该理解到，所揭露的方法和流程等，也可以通过其它的方式实现。以上所描述的方法实施例仅仅是示意性的，例如，附图中的流程图和框图显示了根据本申请实施例的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，这些模块、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方式中，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，两个连续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，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动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。

另外，在本申请实施例中的各功能模块可以集成在一起形成一个独立的部分，也可以是各个模块单独存在，也可以两个或两个以上模块集成形成一个独立的部分。

这些功能如果以软件功能模块的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请实施例提供的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案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一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，电子设备，或者网络设备等）执行本申请实施例提供的方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（ROM, Read-Only Memory）、随机存取存储器（RAM, Random Access Memory）、磁碟或者光盘等各种可以存储程序代码的介质。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、物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、物品或者设备中还存在另外的相同要素。

最后应说明的是：以上所述仅为本申请的部分实施例而已，并不用于限制本申请，尽管参照前述实施例对本申请进行了详细的说明，对于本领域的技术人员来说，其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换。凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

工业实用性

通过获取主播的语音信息，从而控制直播画面中的虚拟形象执行与语音信息相匹配的动作，以提高对虚拟形象控制的精度。

权利要求书

- 1、一种直播控制方法，其特征在于，应用于直播设备，所述方法包括：
获取主播的语音信息；
从所述语音信息中提取关键词和声音特征信息；
根据提取出的所述关键词以及所述声音特征信息确定所述主播的当前情感状态；
根据所述当前情感状态以及所述关键词从预存的动作指令集中匹配对应的目标动作指令；
执行所述目标动作指令，控制直播画面中的虚拟形象执行与所述目标动作指令对应的动作。
- 2、根据权利要求1所述的直播控制方法，其特征在于，所述预存的动作指令集包括通用指令集以及与所述主播的当前虚拟形象对应的定制指令集，所述通用指令集存储有被配置成控制各个虚拟形象的通用动作指令，所述定制指令集存储有被配置成控制所述当前虚拟形象的定制动作指令。
- 3、根据权利要求1或2所述的直播控制方法，其特征在于，所述根据所述当前情感状态以及所述关键词从预存的动作指令集中匹配动作指令的步骤，包括：
在所述预存的动作指令集中存在与所述当前情感状态和所述关键词关联的第一动作指令的情况下，将所述第一动作指令作为所述目标动作指令；
在所述预存的动作指令集中不存在所述第一动作指令的情况下，从所述预存的动作指令集中获取与所述当前情感状态对应的第二动作指令以及与所述关键词关联的第三动作指令；
根据所述第二动作指令和所述第三动作指令确定所述目标动作指令。
- 4、根据权利要求3所述的直播控制方法，其特征在于，所述根据所述第二动作指令和所述第三动作指令确定所述目标动作指令的步骤，包括：
检测所述第二动作指令和所述第三动作指令是否存在联动关系；
若存在联动关系，则按照该联动关系指示的动作执行顺序对所述第二动作指令和所述第三动作指令进行合并，得到所述目标动作指令；
若不存在联动关系，则根据所述第二动作指令和所述第三动作指令各自的预设优先级从所述第二动作指令和所述第三动作指令中选择一个作为所述目标动作指令。
- 5、根据权利要求1或2所述的直播控制方法，其特征在于，所述方法还包括：
针对从所述语音信息中提取出的每个关键词，统计包含该关键词的目标语音信息的数量，以及根据最新获取的第一数量个所述目标语音信息确定的所述第一数量个目标动作指令；
若所述目标语音信息的数量达到第二数量，且所述第一数量个目标动作指令为相同指令，则在所述直播设备的内存中缓存该关键词与所述相同指令的对应关系；其中，所述第一数量不超过所述第二数量；
所述根据所述当前情感状态以及所述关键词从预存的动作指令集中匹配动作指令的步骤，包括：
从缓存的对应关系中查找是否存在所述关键词命中的对应关系；
若存在，则将所述命中的对应关系中记录的指令确定为所述目标动作指令；
若不存在，再执行根据所述当前情感状态以及所述关键词从预存的动作指令集中匹配动作指令的步骤。
- 6、根据权利要求5所述的直播控制方法，其特征在于，所述方法还包括：
每间隔第一预设时长清空所述内存中缓存的所述对应关系。
- 7、根据权利要求1或2所述的直播控制方法，其特征在于，针对所述预存的动作指令集中的每个动作指令，所述直播设备记录有该动作指令的最新执行时间；

所述执行所述目标动作指令的步骤，包括：

获取当前时间，判断所述当前时间与所述目标动作指令的最新执行时间的间隔是否超过第二预设时长；

若超过所述第二预设时长，再执行所述目标动作指令；

若没有超过所述第二预设时长，则从所述预存的动作指令集中查找与所述目标动作指令存在近似关系的其他动作指令替换所述目标动作指令，并执行替换后的目标动作指令。

8、一种直播控制方法，其特征在于，应用于直播设备，所述直播设备被配置成对直播画面中展示的虚拟形象进行控制，所述方法包括：

获取主播的语音信息；

对所述语音信息进行语音分析处理，得到对应的语音参数；

根据预设的参数转换算法将所述语音参数转换为控制参数，并根据该控制参数对所述虚拟形象的口型进行控制。

9、根据权利要求8所述的直播控制方法，其特征在于，所述对所述语音信息进行语音分析处理，得到对应的语音参数的步骤，包括：

将所述语音信息进行分段处理，并提取分段后每一段语音信息中设定时长内的语音片段；

对提取的每个语音片段分别进行语音分析处理，得到每个语音片段对应的语音参数。

10、根据权利要求9所述的直播控制方法，其特征在于，所述将所述语音信息进行分段处理，并提取分段后每一段语音信息中设定时长内的语音片段的步骤，包括：

按照每间隔设定时长提取所述语音信息中该设定时长内的语音片段。

11、根据权利要求9所述的直播控制方法，其特征在于，所述将所述语音信息进行分段处理，并提取分段后每一段语音信息中设定时长内的语音片段的步骤，包括：

按照所述语音信息的连续性对该语音信息进行分段处理，并提取分段后每一段语音信息中设定时长内的语音片段。

12、根据权利要求9所述的直播控制方法，其特征在于，所述对提取的每个语音片段进行语音分析处理，得到每个语音片段对应的语音参数的步骤，包括：

提取每个语音片段的振幅信息；

针对每个语音片段，根据该语音片段的振幅信息计算得到该语音片段对应的语音参数。

13、根据权利要求12所述的直播控制方法，其特征在于，所述根据该语音片段的振幅信息计算得到该语音片段对应的语音参数的步骤，包括：

根据该语音片段的帧长信息和所述振幅信息按照归一化算法进行计算，得到该语音片段对应的语音参数。

14、根据权利要求8-13任意一项所述的直播控制方法，其特征在于，所述控制参数包括所述虚拟形象的上下嘴唇之间的唇间距以及嘴角角度二者中的至少一个。

15、根据权利要求14所述的直播控制方法，其特征在于，当所述控制参数包括所述唇间距时，该唇间距根据所述语音参数和预设的与所述虚拟形象对应的最大唇间距按照预设的参数转换算法计算得到；

当所述控制参数包括所述嘴角角度时，该嘴角角度根据所述语音参数和预设的与所述虚拟形象对应的最大嘴角角度按照预设的参数转换算法计算得到。

16、根据权利要求14所述的直播控制方法，其特征在于，当所述控制参数包括所述唇间距时，所述最大唇间距根据预先获取的所述主播的唇间距设置；

当所述控制参数包括所述嘴角角度时，所述最大嘴角角度根据预先获取的所述主播的嘴角角度设置。

17、一种直播控制方法，其特征在于，应用于直播设备，所述方法包括：

获取主播的语音信息；

从所述语音信息中提取关键词和声音特征信息；

根据提取出的所述关键词以及所述声音特征信息确定所述主播的当前情感状态；
根据所述当前情感状态以及所述关键词从预存的动作指令集中匹配对应的目标动作指令；

执行所述目标动作指令，控制直播画面中的虚拟形象执行与所述目标动作指令对应的动作；

对所述语音信息进行语音分析处理，得到对应的语音参数；

根据预设的参数转换算法将所述语音参数转换为控制参数，并根据该控制参数对所述虚拟形象的口型进行控制。

18、一种直播设备，其特征在于，包括存储器、处理器及存储在所述存储器中并在所述处理器中被执行的机器可执行指令，所述机器可执行指令被所述处理器执行时实现权利要求 1-17 中任意一项所述的直播控制方法。

19、一种可读存储介质，其上存储有机器可执行指令，其特征在于，所述机器可执行指令被执行时实现权利要求 1-17 中任意一项所述的直播控制方法。

附图

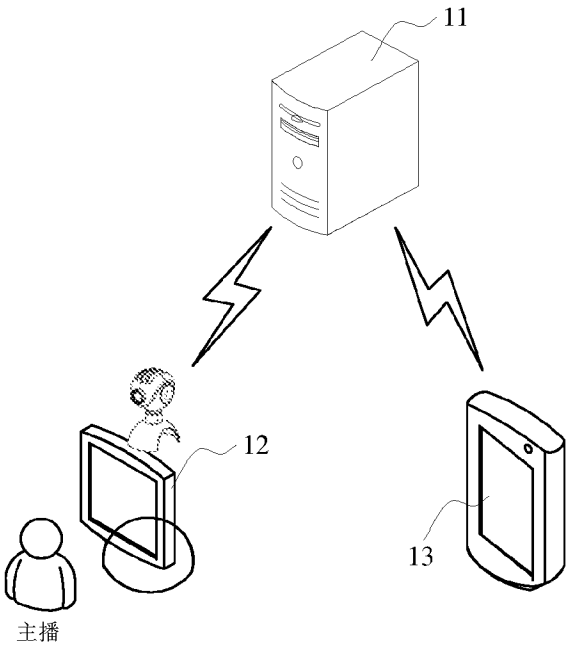


图 1

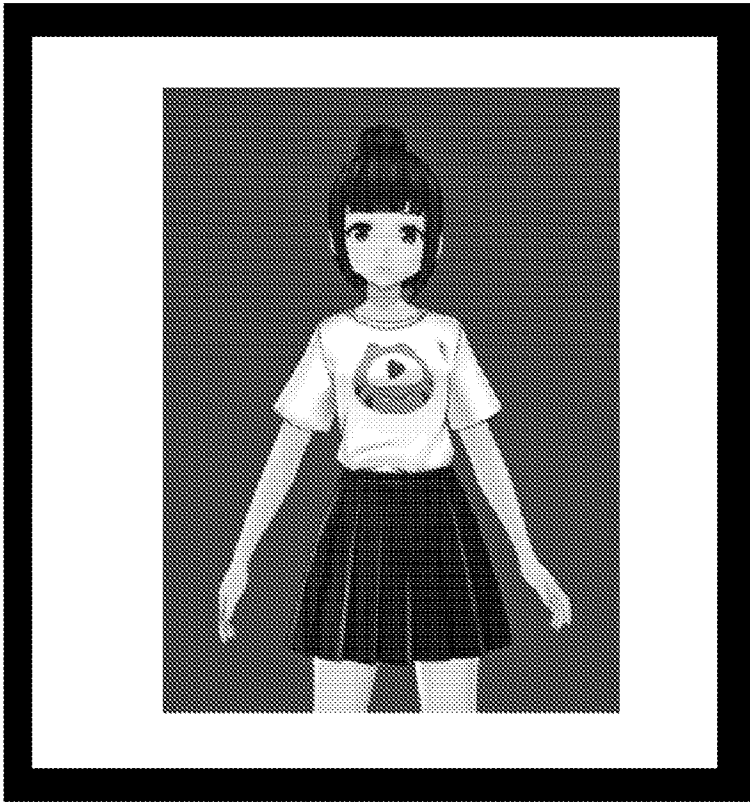


图 2

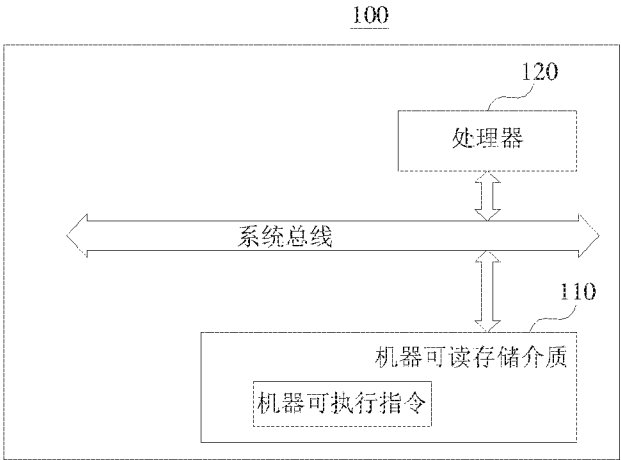


图 3

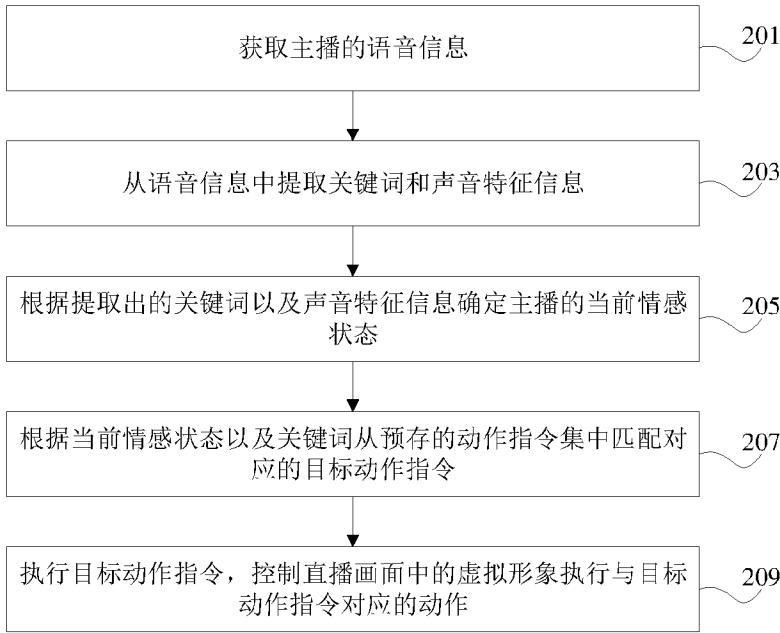


图 4

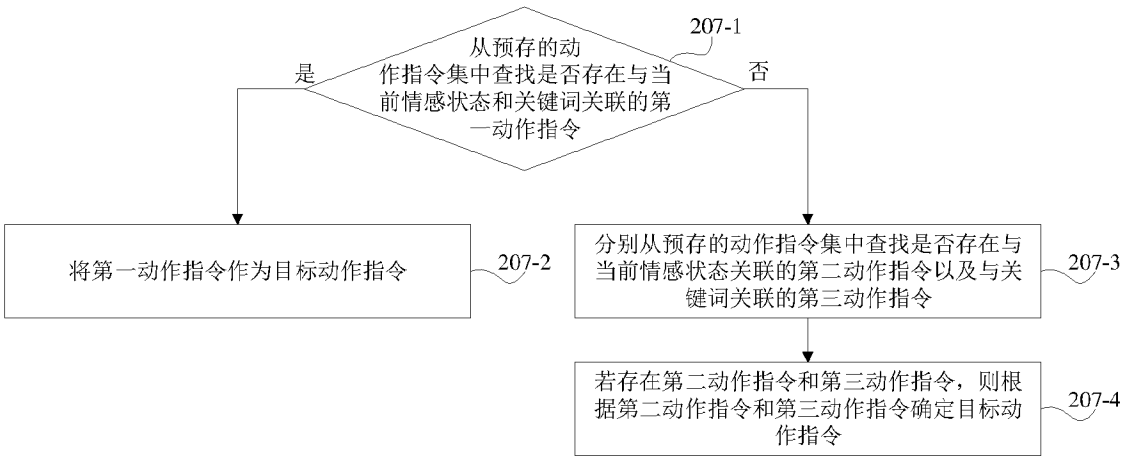
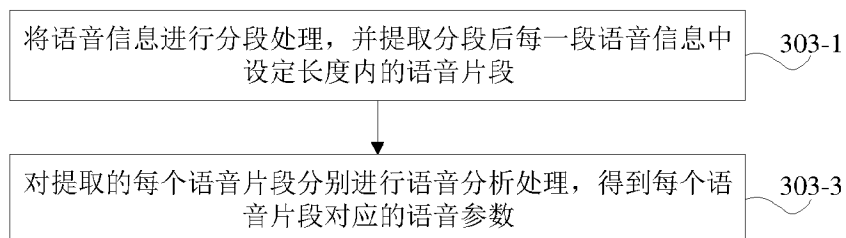
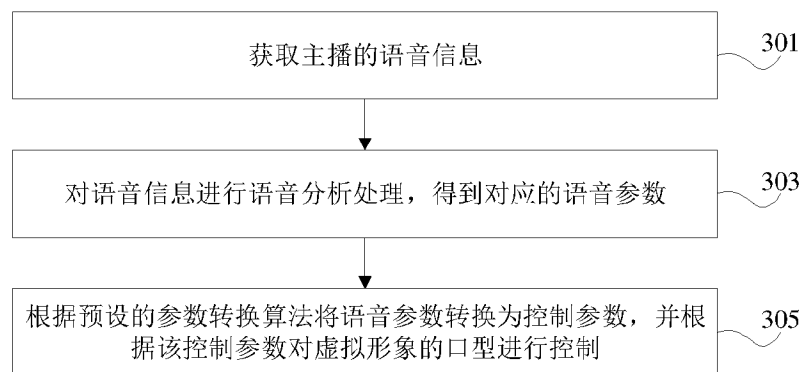
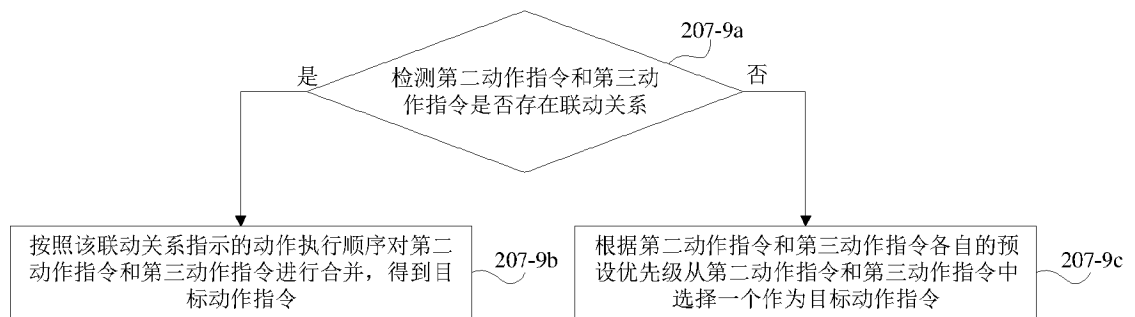
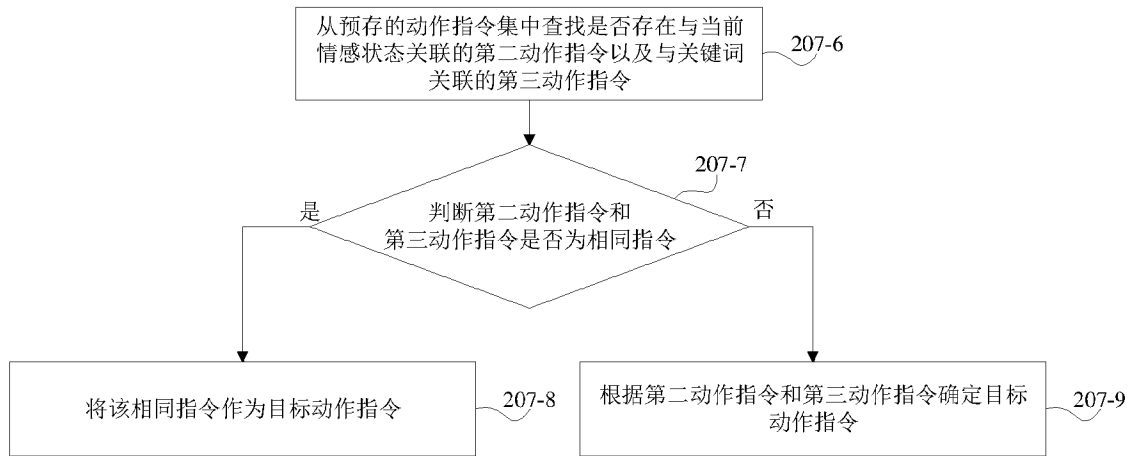


图 5



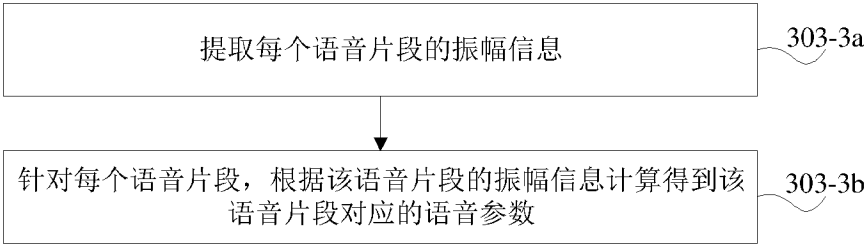


图 10

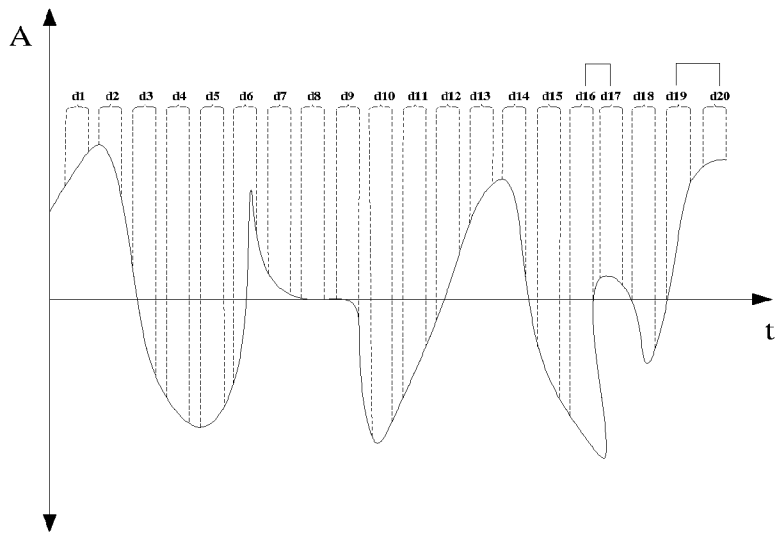


图 11

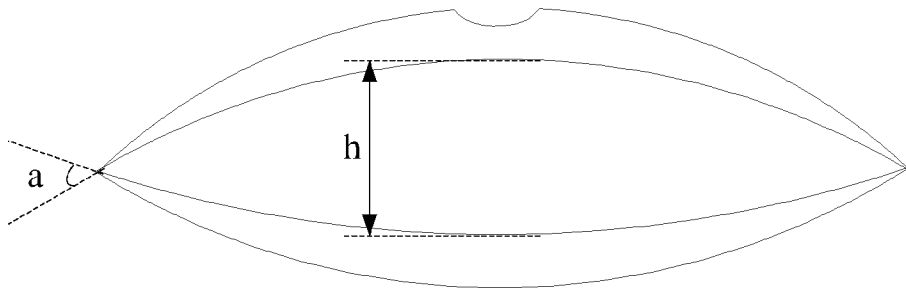


图 12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/081626

A. CLASSIFICATION OF SUBJECT MATTER

G10L 21/10(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, VEN, EPTXT, USTXT, WOTXT, CNKI: 直播, 主播, 即时通讯, 虚拟, 情感, 表情, 情绪, 动作, 语音, 声音, 音频, 提取, 采集, 分析, 特征, 口型, 嘴型; Live broadcast, virtual, facial expression, sentiment, sound, voice, audio, feature

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 102568023 A (MICROSOFT CORPORATION) 11 July 2012 (2012-07-11) description paragraphs [0020]-[0089] and figures 1-9	1-3, 7-19
X	CN 109410297 A (CHONGQING IQIYI SMART TECH CO., LTD.) 01 March 2019 (2019-03-01) description paragraphs [0098]-[0203] and figures 1-7	1-3, 7-19
X	CN 109326151 A (BEIJING INTELLIGENT YOUXUE TECHNOLOGY CO., LTD.) 12 February 2019 (2019-02-12) description paragraphs [0047]-[0090] and figures 1-5	1-3, 7-19
A	CN 107995442 A (BEIJING QIHOO TECHNOLOGY CO., LTD.) 04 May 2018 (2018-05-04) entire document	1-19
A	CN 107610205 A (ZHUHAI KINGSOFT NETWORK GAME TECHNOLOGY CO., LTD. et al.) 19 January 2018 (2018-01-19) entire document	1-19
A	CN 107920256 A (GUANGZHOU KUGOU COMPUTER TECHNOLOGY CO., LTD.) 17 April 2018 (2018-04-17) entire document	1-19



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

23 June 2020

Date of mailing of the international search report

30 June 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/081626**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109788345 A (GUANGZHOU HUYA INFORMATION TECHNOLOGY CO., LTD.) 21 May 2019 (2019-05-21) entire document	1-19
PX	CN 109872724 A (GUANGZHOU HUYA INFORMATION TECHNOLOGY CO., LTD.) 11 June 2019 (2019-06-11) entire document	1-19

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/081626

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	102568023	A	11 July 2012	US	2012130717	A1	24 May 2012
CN	109410297	A	01 March 2019	None			
CN	109326151	A	12 February 2019	None			
CN	107995442	A	04 May 2018	None			
CN	107610205	A	19 January 2018	None			
CN	107920256	A	17 April 2018	CN	107920256	B	10 January 2020
CN	109788345	A	21 May 2019	CN	109788345	B	10 March 2020
CN	109872724	A	11 June 2019	None			

国际检索报告

国际申请号

PCT/CN2020/081626

A. 主题的分类

G10L 21/10(2013.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G10L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS, CNTXT, VEN, EPTXT, USTXT, WOTXT, CNKI:直播, 主播, 即时通讯, 虚拟, 情感, 表情, 情绪, 动作, 语音, 声音, 音频, 提取, 采集, 分析, 特征, 口型, 嘴型; Live broadcast, virtual, facial expression, sentiment, sound, voice, audio, feature

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 102568023 A (微软公司) 2012年 7月 11日 (2012 - 07 - 11) 说明书第[0020]-[0089]段及图1-9	1-3, 7-19
X	CN 109410297 A (重庆爱奇艺智能科技有限公司) 2019年 3月 1日 (2019 - 03 - 01) 说明书第[0098]-[0203]段及图1-7	1-3, 7-19
X	CN 109326151 A (北京智能优学科技有限公司) 2019年 2月 12日 (2019 - 02 - 12) 说明书第[0047]-[0090]段及图1-5	1-3, 7-19
A	CN 107995442 A (北京奇虎科技有限公司) 2018年 5月 4日 (2018 - 05 - 04) 全文	1-19
A	CN 107610205 A (珠海金山网络游戏科技有限公司等) 2018年 1月 19日 (2018 - 01 - 19) 全文	1-19
A	CN 107920256 A (广州酷狗计算机科技有限公司) 2018年 4月 17日 (2018 - 04 - 17) 全文	1-19
PX	CN 109788345 A (广州虎牙信息科技有限公司) 2019年 5月 21日 (2019 - 05 - 21) 全文	1-19

☒ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 6月 23日

国际检索报告邮寄日期

2020年 6月 30日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

陈德锋

电话号码 86-(010)-62411501

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 109872724 A (广州虎牙信息科技有限公司) 2019年 6月 11日 (2019 - 06 - 11) 全文	1-19

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/081626

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	102568023	A	2012年 7月 11日	US	2012130717	A1	2012年 5月 24日
CN	109410297	A	2019年 3月 1日	无			
CN	109326151	A	2019年 2月 12日	无			
CN	107995442	A	2018年 5月 4日	无			
CN	107610205	A	2018年 1月 19日	无			
CN	107920256	A	2018年 4月 17日	CN	107920256	B	2020年 1月 10日
CN	109788345	A	2019年 5月 21日	CN	109788345	B	2020年 3月 10日
CN	109872724	A	2019年 6月 11日	无			