(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
2022 年 11 月 17 日 (17.11.2022)

WIPO | PCT

(10) 国际公布号
WO 2022/237464 A1

(51) 国际专利分类号:
G10L 13/02 (2013.01) *H04N 21/4788* (2011.01)
H04N 21/439 (2011.01)

(21) 国际申请号: PCT/CN2022/087491

(22) 国际申请日: 2022 年 4 月 18 日 (18.04.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110523959.3 2021年5月13日 (13.05.2021) CN

(71) 申请人: 北京字节跳动网络技术有限公司 (**BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.**) [CN/CN]; 中国北京市石景山区实兴大街30号院3号楼2层B-0035房间, Beijing 100041 (CN)。

(72) 发明人: 陈映宜 (**CHEN, Yingyi**); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100086 (CN)。

(74) 代理人: 上海光栅知识产权代理有限公司 (**LINKER IP LLC**); 中国上海市长宁区延安西路889号702室, Shanghai 201109 (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, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC,

(54) **Title:** AUDIO SYNTHESIS METHOD AND APPARATUS, AND DEVICE, MEDIUM AND PROGRAM PRODUCT

(54) **发明名称**: 音频合成方法、装置、设备、介质及程序产品

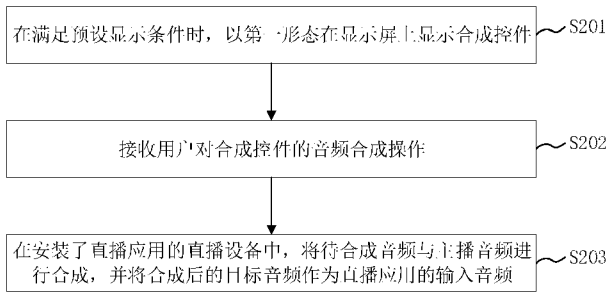


图 2

S201 When a preset display condition is met, display, in a first form, a synthesis control on a display screen
S202 Receive an audio synthesis operation of a user on the synthesis control
S203 In a live streaming device in which a live streaming application is installed, synthesize audio to be subjected to synthesis with streamer audio, and take target audio obtained after synthesis as input audio of the live streaming application

(57) **Abstract:** Provided in the embodiments of the present disclosure are an audio synthesis method and apparatus, and a device, a medium and a program product. The method comprises: firstly, when a preset display condition is met, displaying, in a first form, a synthesis control on a display screen; then receiving an audio synthesis operation of a user on the synthesis control; and finally, in a live streaming device in which a live streaming application is installed, synthesizing audio to be subjected to synthesis with initial audio, and taking target audio obtained after synthesis as input audio of the live streaming application. In the embodiments of the present disclosure, by means of the method, the technical problem in the prior art of, when there are requirements for high-quality live streaming, it only being possible to add synthesized audio by relying on a sophisticated professional device and a professional technician team is solved; and the technical effect of a streamer being able to conveniently and quickly add, on a live streaming terminal, synthesized audio to a live-streaming audio stream is achieved, thereby reducing the costs and technical requirements for live streaming, and improving the use experience of users.

(57) **摘要**: 本公开实施例提供的音频合成方法、装置、设备、介质及程序产品，在该方法中，首先在满足预设显示条件时，以第一形态在显示屏上显示合成控件，然后接收用户对合成控件的音频合成操作，最后在安装了直播应用的直播设备中，将待合成音频与初始音频进行合成，并将合成后的目标音频作为直播应用的输入音频。本公开实施例通过上述方法，解决了现有技术在面对高质量的直播需求时，存在只能够依赖复杂的专业设备和专业技术人员团队来添加合成音频的技术问题。达到了主播在直播终端上即可便捷快速地在直播音频流中添加合成音频的技术效果，降低了直播的成本和技术门槛，提高了用户的使用体验感。

WO 2022/237464 A1

[见续页]

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))。

音频合成方法、装置、设备、介质及程序产品

相关申请交叉引用

本申请要求于 2021 年 5 月 13 日提交中国专利局、申请号为 202110523959.3、发明名称为“音频合成方法、装置、设备、介质及程序产品”的中国专利申请的优先权，其全部内容通过引用并入本文。

技术领域

本公开实施例涉及广播电视及互联网直播技术领域，尤其涉及一种音频合成方法、装置、设备、介质及程序产品。

背景技术

随着社会信息化的不断深入发展，人们对于信息获取的及时性需求也不断提高。直播技术能够使得人们突破空间的限制，实时获取到直播现场的各类信息。而近年来兴起的网络直播，也成为了人们展示自我的一个平台，获得了极大的社会热度。手机直播更是将降低了直播的技术门槛，让不具备专业技术的普通人也能简单地进行直播。然而，随着观众对于直播内容及质量的要求不断提高，围绕着手机这类移动直播设备，衍生出了大量的周边辅助产品。

目前，一个高质量的手机直播，所需要的硬件设备包括：台式电脑、声卡、若干中间转换器/衔接器、手机、耳机、麦克风等，除了主播外，还需要配备专门的技术人员来负责各种直播效果的实时添加。

即现有技术在面对高质量的直播需求时，存在只能够依赖复杂的专业设备和专业技术人员团队来添加合成音频的技术问题。这无疑又提高了直播的技术门槛，增加了主播们的直播成本。

发明内容

本公开实施例提供一种音频合成方法、装置、设备、介质及程序产品，以解决现有技术在面对高质量的直播需求时，存在只能够依赖复杂的专业设备和专业技术人员团队来添加合成音频的技术问题。

第一方面，本公开实施例提供一种音频合成方法，包括：

在满足预设显示条件时，以第一形态在显示屏上显示合成控件，第一形态对显示屏上显示内容的视觉影响满足预设视觉要求；

接收用户对合成控件的音频合成操作；

在安装了直播应用的直播设备中，将待合成音频与初始音频进行合成，并将合成后的目标音频作为直播应用的输入音频，其中，待合成音频为音频合成操作所指示的音频，初始音频为直播设备通过音频采集设备接收到主播所发出的，和/或，直播环境中的声音信号。

第二方面，本公开实施例提供一种音频合成装置，包括：

显示模块，用于在满足预设显示条件时，以第一形态在显示屏上显示合成控件，第一形态对显示屏上显示内容的视觉影响满足预设视觉要求；

接收模块，用于接收用户对合成控件的音频合成操作；

处理模块，用于在安装了直播应用的直播设备中，将待合成音频与初始音频进行合成，并将合成后的目标音频作为直播应用的输入音频，其中，待合成音频为音频合成操作所指示的音频，初始音频为直播设备通过音频采集设备接收到主播所发出的，和/或，直播环境中的声音信号。

第三方面，本公开实施例提供一种电子设备，包括：

至少一个处理器和存储器；

所述存储器存储计算机程序；

所述至少一个处理器执行所述存储器存储的计算机程序，使得所述至少一个处理器执行如上第一方面以及第一方面各种可能的设计所述的音频合成方法。

第四方面，本公开实施例提供一种直播一体机，包括第三方面所提供的任意一种可能的电子设备。

第五方面，本公开实施例提供一种计算机可读存储介质，所述计算机可读存储介质中存储有计算机程序，当处理器执行所述计算机程序时，实现如上第一方面以及第一方面各种可能的设计所述的音频合成方法。

第六方面，本公开实施例提供一种计算机程序产品，包括计算机程序，该计算机程序被处理器执行时，实现如上第一方面以及第一方面各种可能的设计所述的音频合成方法。

第七方面，本公开实施例提供一种计算机程序，所述计算机程序被处理器执行时，实现如上述第一方面以及第一方面各种可能的设计所述的音频合成方法。

本公开实施例提供了音频合成方法、装置、设备、介质及程序产品，在该方法中，首先在满足预设显示条件时，以第一形态在显示屏上显示合成控件，然后接收用户对合成控件的音频合成操作，最后在安装了直播应用的直播设备中，将待合成音频与初始音频进行合成，并将合成后的目标音频作为直播应用的输入音频。本公开实施例通过上述方法，解决了现有技术在面对高质量的直播需求时，存在只能够依赖复杂的专业设备和专业技术人员团队来添加合成音频的技术问题。达到了主播在直播终端上即可便捷快速地在直播音频流中添加合成音频的技术效果，降低了直播的成本和技术门槛，提高了用户的使用体验感。

附图说明

为了更清楚地说明本公开实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图是本公开的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 为本公开实施例提供的现有室内主播使用的直播设备套件结构示意图；

图 2 为本公开实施例提供的音频合成方法的流程示意图一；

图 3 为本公开实施例提供的音频合成方法的流程示意图二；

图 4a-4c 为本公开实施例提供的在屏幕上显示合成控件及用户交互操作示意图；

图 5 为本公开实施例提供的音频合成方法的流程示意图三；

图 6a-6b 为本公开实施例提供的通过合成控件进行弹窗设置的应用场景示意图；

图 7 为本公开实施例提供的音频合成方法的流程示意图四；

图 8 为本公开实施例提供的一种声音设置界面示意图；

图 9 为本公开实施例提供的一种直播一体化设备外接快捷键键盘的场景示意图；

图 10a-10c 为本公开实施例提供的一种合成控件在待合成音频播放时显示播放进度的示意图；

图 11 为本公开实施例提供的音频合成装置的结构框图；

图 12 为本公开实施例提供的电子设备的结构示意图。

具体实施方式

为使本公开实施例的目的、技术方案和优点更加清楚，下面将结合本公开实施例中的附图，对本公开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公开一部分实施例，而不是全部的实施例。基于本公开中的实施例，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，包括但不限于对多个实施例的组合，都属于本公开保护的范围。

本公开的说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”、“第四”等（如果存在）是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本公开的实施例例如能够以除了在这里图示或描述的那些以外的顺序实施。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

随着移动互联网技术的不断发展，人们的日常生活已经高度信息化，而直播技术解决了人们在不同地理空间上异地接收直播地信息的一种高效手段。近年来兴起的网络直播，更是给人们的娱乐生活增加了很多乐趣，收到了人们的追捧。大量主播纷纷加入了网络直播的行列。而通过移动设备来进行直播，以其简便性受到了很多主播的青睐。但是，随着观众对于直播质量要求的不断提高，主播在进行直播的时候，需要进行音频合成，如添加“入场音”、“暖场笑声”、“鼓掌声”等等音频效果到直播应用中。

面对这种需求，对于大多数主播来说，只能够通过购置一套专业的直播设备，包括：台式电脑、专业声卡、若干个中间转换器/衔接器等专业直播辅助设备来实现直播中的音频合成需求。

图 1 为本公开实施例提供的现有室内主播使用的直播设备套件结构示意图。如图 1 所示，直播设备套件包括：移动终端 101、中间衔接器 102、中间转换器 103、声卡 104、台式电脑 105。

其中，移动终端 101 中安装有直播应用。

中间衔接器 102 用于解决移动终端 101 在长时间直播时，充电接口被数据线占用而无法充电的问题。

中间转换器 103 用于将台式电脑 105 中的特效数据转接到移动终端 101 中，并且连接耳机后可以给客户提供耳返功能，还能够解决声卡 104 只能单路输出而不能反向接收直播应用输出的音频数据的问题。

声卡 104 与麦克风连接，实现音频模拟信号转换为数字信号，输入台式电脑 105 中进行各种音效处理。

但是，图 1 所示的方式不但造成直播成本的提高，需要配置多个辅助设备，而绝大多数主播不具备专业技术，无法自己去配置这一套设备，因此还需要配备专门的技术人员团队来构建、运营和维护这套专业直播设备，这些都使得本已通过手机等移动终端降低了的直播门槛，又被提高了起来。

并且，由于连接了多个设备，使得整套直播设备的稳定性容易受到各种因素的影响，任何一个设备的接插件插口出现问题，或者线路由于某种原因断路，都会使得系统无法工作。

基于上述技术问题，本申请的发明构思旨在：

通过一个直播终端直接让主播在直播的时候通过屏幕或者按键快速合成所需要的音频到直播的音频流中。使得直播无需多个设备，仅用一个小型设备就能完成音频合成，也无需专业技术人员团队参与运维，主播一人即可在直播时实时添加合成音频。

下面结合附图对本申请提供的音频合成方法进行详细介绍。

参考图 2，图 2 为本公开实施例提供的音频合成方法的流程示意图一。本实施例的方法应用在直播一体式终端中，该音频合成方法包括：

S201：在满足预设显示条件时，以第一形态在显示屏上显示合成控件。

在本步骤中，预设显示条件包括：检测到摄像设备开启指令、检测到直播终端按预设运动方式进行运动、接收到预设语音指令和接收到预设按钮开关开启指令中的至少一种。

在一种可能的设计中，合成控件包括悬浮窗控件，以第一形态在显示屏上显示合成控件，包括：

根据预设透明度、预设大小以及预设形状在显示屏的边缘显示悬浮窗控件。

在另一种可能的设计中，第一形态对显示屏上显示内容的视觉影响满足预设视觉要求，包括：音频合成控件对直播界面的遮挡率小于或等于预设遮挡阈值，和/或，音频合成控件的综合透明度大于或等于预设透明阈值；

其中，综合透明度是利用预设模型，根据音频合成控件中各个显示元素的透明度来确定的。具体的，某个控件上各个显示元素的透明度先乘以其对应的预设权重值，然后将各个乘积相加即为综合权重值。

例如：一个按钮控件，其背景色的透明度与按钮控件上的文本的透明度分别乘以对应的权重值，再求两者的和，即为按钮控件的综合透明度。

在一种可能的设计中，所述显示屏包括：触控屏和非触控屏。

在本实施例中，直播一体式终端上包括至少一个显示屏。在一种可能的设计中，直播一体式终端可以通过扩展接口连接多个显示屏，扩展接口包括无线接口和有线接口。

S202：接收用户对合成控件的音频合成操作。

在本步骤中，对合成控件的音频合成操作包括：单击、双击、摇一摇、按预设路径滑动等

在本实施例中，当显示屏为触控屏时，具体的，实时检测所述用户利用预设操作物靠近

或触碰到所述触控屏上的所述合成控件。

预设操作物包括：用户的身体任意部位，触控物（如触控笔）、热感应物（如发热棒）、电磁感应物（电磁感应笔），RFID（Radio Frequency Identification，射频识别）标签钥匙等。

触控物直接触控合成控件。

热感应物只要靠近触控屏所对应的预设感应区域，则安装在直播一体式终端上的传感器，感应到热源靠近预设感应区域时，确认接收到用户的音频合成操作。

电磁感应物只要靠近触控屏所对应的预设感应区域，则安装在直播一体式终端上的传感器，感应到有磁场靠近预设感应区域时，确认接收到用户的音频合成操作。

RFID 标签钥匙只要靠近触控屏所对应的预设感应区域，则安装在直播一体式终端上的读卡器或读卡电路，识别 RFID 标签中包含的待合成音频信息，即接收到了音频合成操作。

S203：在安装了直播应用的直播设备中，将待合成音频与初始音频进行合成，并将合成后的目标音频作为直播应用的输入音频。

在本步骤中，待合成音频为音频合成操作所指示的音频，初始音频为直播设备通过音频采集设备接收到主播所发出的，和/或，直播环境中的声音信号。

直播应用将目标音频编码后，通过网络传输给直播平台服务器，再由直播平台服务发送给各个观众。

需要说明的是，待合成音频包括：笑声、出场/入场声、鸟叫声（如乌鸦叫声）、掌声、打耳光声、枪声等等。

需要说明的是，在安装了直播应用的直播设备包括：直播摄像设备（用于新闻媒体直播的专业摄像机）、个人电脑（Personal Computer，简称 PC）、移动终端（手机、平板电脑）、网络直播一体化设备、直播载具（如直播车辆、直播购物车）等。

本公开实施例提供的音频合成方法，该方法首先在满足预设显示条件时，以第一形态在显示屏上显示合成控件，然后接收用户对合成控件的音频合成操作，最后在安装了直播应用的直播设备中，将待合成音频与初始音频进行合成，并将合成后的目标音频作为直播应用的输入音频。本公开实施例通过，解决了现有技术在面对高质量的直播需求时，存在只能够依赖复杂的专业设备和专业技术人员团队来添加合成音频的技术问题。达到了主播在直播终端上即可便捷快速地在直播音频流中添加合成音频的技术效果，降低了直播的成本和技术门槛，提高了用户的使用体验感。

参考图3，图3为本公开实施例提供的音频合成方法的流程示意图二。本实施例的方法应用在直播一体式终端中，该音频合成方法包括：

S301：检测到摄像设备开启指令时，以第一形态在显示屏上显示合成控件。

在本步骤中，第一形态对所述显示屏上显示内容的视觉影响满足预设视觉要求。

在本实施例中，用户在启动直播一体式终端上的直播应用后，点击开始直播，则直播应用开启直播一体式终端上的摄像头。而当直播一体式终端检测到摄像头被打开后，则在显示屏的边缘显示合成控件。

在一种可能的设计中，合成控件包括悬浮窗控件，则以第一形态在显示屏上显示合成控件，包括：

根据预设透明度、预设大小以及预设形状在显示屏的边缘显示悬浮窗控件。

图 4a-4c 为本公开实施例提供的在屏幕上显示合成控件及用户交互操作示意图。如图 4a

所示，合成控件为悬浮窗 41，在检测到摄像头打开后，以一个小悬浮窗的形态显示合成控件，并且悬浮窗 41 设置为底色透明度 50%以上，使得悬浮窗 41 对直播界面上其它内容的视觉影响降低到满足预设要求。

在一种可能的设计中，合成控件包括多个子控件，如图 4b 所示，合成控件包括：弹窗设置子控件 411、多个与待合成音频对应的音频子控件 412、收起控件 413。在检测到摄像头打开后，直接显示所有的子控件。

在另一种可能的设计中，合成控件包括多个子控件并，但是第一形态并不展示所有的子控件，如图 4c 所示，第一形态只显示部分与待合成音频对应的音频子控件 412，弹窗设置子控件 411 以及展开控件 414。

S302：获取用户对合成控件的状态设置操作。

在本步骤中，状态设置操作包括：改变合成控件显示形态的操作以及改变合成控件所在位置的操作。

改变合成控件显示形态的操作，包括：展开/收起操作、调换子控件顺序或位置的操作、压缩、放大等等。

改变合成控件所在位置的操作，包括：将合成控件拖拽到显示屏的其它位置。

需要说明的是，在一种可能的设计中，合成控件始终需要靠边显示，当用户拖拽合成控件松开后，计算合成控件到显示屏四个边缘的距离，将合成控件吸附到最小距离的边缘附近显示。

S303：根据状态设置操作切换合成控件的形态，和/或，显示位置。

例如，如图 4a-4c 所示，用户在图 4c 中点击了展开控件 414 后，合成控件从图 4c 的形态转换成图 4b 的形态，用户再点击收起控件 413 后，合成控件从图 4b 的形态转换成图 4a 的形态。

S304：实时检测用户利用预设操作物靠近或触碰到触控屏上的合成控件。

在本实施例中，显示屏包括至少一个触控屏，预设操作物包括：用户的身体任意部位，触控物（如触控笔）、热感应物（如发热棒）、电磁感应物（电磁感应笔），RFID 标签钥匙等。

在一种可能的设计中，合成控件包括多个与待合成音频对应的子控件，如图 4b 所示，则本步骤具体包括：

若检测到所述预设操作物靠近或触碰到所述子控件，则将所述子控件切换为被选中形态；

确定与所述子控件对应的合成操作指令。

S305：在安装了直播应用的直播设备中，将待合成音频与初始音频进行合成，并将合成后的目标音频作为直播应用的输入音频。

本步骤的具体原理与名词解释介绍参考 S203，在此不再赘述。

本公开实施例提供的音频合成方法，该方法首先在满足预设显示条件时，以第一形态在显示屏上显示合成控件，然后接收用户对合成控件的音频合成操作，最后在安装了直播应用的直播设备中，将待合成音频与初始音频进行合成，并将合成后的目标音频作为直播应用的输入音频。本公开实施例通过，解决了现有技术在面对高质量的直播需求时，存在只能够依赖复杂的专业设备和专业技术人员团队来添加合成音频的技术问题。达到了主播在直播终端上即可便捷快速地在直播音频流中添加合成音频的技术效果，降低了直播的成本和技术门

槛，提高了用户的使用体验感。

参考图 5，在上述两个实施例的基础上，本公开提供的音频合成方法还能够利用合成控件在直播时，通过弹窗设置的形式对直播一体式设备的各项音频控制参数进行设置，下面以具体实施例进行举例说明。

图 5 为本公开实施例提供的音频合成方法的流程示意图三。本实施例的方法应用在直播一体式终端中，该音频合成方法包括：

S501：获取用户对合成控件的弹窗设置操作。

在本实施例中，如图 4b 或图 4c 中所示的弹窗设置子控件 411，用户在点击了该控件后，直播一体式设备获取到弹窗设置操作的控制指令。

S502：响应于弹窗设置操作，在显示屏上叠加显示设置窗口。

在本实施例中，设置窗口包括声音设置选项卡，该声音设置选项卡所对应的显示区域内，包括：耳返开关控件、闪避开关控件、扬声器选择控件、麦克风选择控件、连麦音量设置控件、音效选择控件、音效音量控件等控件中的至少一个控件。

S503：响应于用户对各个控件的操作指令，对应设置各项音频控制参数。

在本实施例中，具体的：

S5031、耳返开关控件的开关指令，设置耳返效果开关状态所对应的控制参数，并对应切换耳返开关控件的开关显示状态。

和/或，

S5032、响应于用户对闪避开关控件的开关指令，设置闪避效果开关状态所对应的控制参数，并对应切换闪避开关控件的开关显示状态。

和/或，

S5033、响应于用户对扬声器选择控件的选择指令，设置直播设备中各个扬声器的工作状态，并对应切换扬声器选择控件的扬声器选中状态。

和/或，

S5034、响应于用户对麦克风选择控件的选择指令，设置直播设备中各个麦克风的工作状态，并对应切换麦克风选择控件的麦克风选中状态。

和/或，

S5035、响应于用户对连麦音量设置控件的第一音量设置操作，设置连麦时对方音量的大小控制参数，并对应改变连麦音量设置控件的显示状态；

和/或，

S5036、响应于用户对音效选择控件的选择指令，设置目标音频合成时的音效控制参数，并对应切换音效选择控件的音效选中状态。

和/或，

S5037、响应于用户对音效音量控件的第二音量设置操作，设置音效对应的音量大小控制参数，并对应改变音效音量控件的显示状态。

在一种可能的设计中，设置窗口包括快捷键设置选项卡，该快捷键设置选项卡包括：已选择快捷键阵列以及待选快捷键阵列，则本步骤还包括：

S5038、响应于用户对已选择快捷键阵列的删除指令，更新已选择快捷键阵列以及待选快捷键阵列的展示状态。

和/或，

S5039、响应于用户对待选快捷键阵列的选择指令，更新已选择快捷键阵列以及待选快捷键阵列的展示状态。

需要说明的是，S5031-S5039 并没有先后顺序的要求，用户可以同时设置几个参数。

图 6a-6b 为本公开实施例提供的通过合成控件进行弹窗设置的应用场景示意图。声音设置选项卡的一种实现方式如图 6a 所示，用户点击了弹窗设置子控件 411 后，在显示屏下方弹出设置窗口，在设置窗口的声音设置选项卡 421 上，用户通过拨动耳返或闪避控件的开关来设置开启或关闭耳返效果或闪避效果。通过点击对应的控件来选择扬声器或麦克风。通过拨动音量条来设置各类音量大小。通过点击对应的音效控件来切换音效模式。

需要说明的是，选择控件除了图 6a 的每种选项单独一个按钮控件外，还可以是通过勾选，或者是通过下拉选择控件来选择。音量条也可以设置为开口弧线、封闭圆环、旋钮等形式。

音效快捷键选项卡 422 的一种实现方式如图 6b 所示，方框中显示的是已经被选中显示的待合成音频所对应的快捷键，在本实施例中，最多允许添加 8 个快捷键，点击已被选中的快捷键右侧的“垃圾箱”状按钮即删除按钮，即可设置不显示此快捷键。

下方以阵列形式排布着多个待合成音频选项，点击右侧的“+”号图标，即可将此待合成音频添加到上方的处于显示状态的快捷键方框中，并且，选中后，切换该选择控件的背景色，并且将“+”改为“√”表明，此选项已被选择。

此外，待合成音频选项还可以通过网络接口从云端服务器或互联网服务器中下载，如图 6b 中“口哨”，“尖叫”对应的待合成音频，在直播一体式终端中没有预先存储有的，则只需点击控件右侧的下载按钮，即可从服务器下载到本地存储中。

值得说明的是，在一种可能的设计中，只要云端服务器或互联网服务器中更新了待合成音频选项列表，则在音效快捷键选项卡 422 中也进行同步更新，用户看到有新选项后，可以点击下载按钮下载。当然，用户也可以设置自动下载，或批量下载，一次性将所有更新的待合成音频都下载到本地存储中。

需要说明的是，音效快捷键的设置也可以采用下拉框选择，或者是其它的选择方式，本领域技术人员可以根据实际情况进行选用，本申请不做限定。

本公开实施例提供的音频合成方法，在图 2 和图 3 所示实施例的基础上，获取用户对合成控件的弹窗设置操作，然后响应于弹窗设置操作，在显示屏上叠加显示设置窗口，再响应于用户对各个控件的操作指令，对应设置各项音频控制参数，解决了主播在直播时无法简便快捷地更改各项声音相关的设置参数，需要专业技术人员辅助的技术问题。使得声音设置和合成控件的快捷键设置简单方便，无需改变视线方向，影响直播效果，也无需暂停直播来设置，提高了用户对于直播一体式终端的使用体验感。

在上述三个实施例的基础上，本公开提供的音频合成方法还能够在直播前，在直播一体式终端上，通过专门的声音设置界面对合成控件的显示状态以及各项声音控制参数，下面以具体实施例来对具体步骤进行说明介绍。

图 7 为本公开实施例提供的音频合成方法的流程示意图四。本实施例的方法应用在直播一体式终端中，该音频合成方法包括：

S701：在没有开启摄像头或者直播开启之前，获取用户的声音设置操作。

在本步骤中，直播一体式终端检测到用户点击声音设置按钮，或者是通过预设的直播应用开始时的设置指令，进行了声音设置界面的开启操作。

S702：响应于用户的声音设置操作，在显示屏上显示声音设置界面。

在本实施例中，声音设置界面包括：耳返开关控件、闪避开关控件、扬声器选择控件、麦克风选择控件、连麦音量设置控件、音效选择控件、音效音量控件、合成控件（如悬浮窗）显示开关控件、已选择快捷键阵列以及待选快捷键阵列等控件中的至少一个控件。

S703：根据用户对声音设置界面上各个参数控件的设置指令，对应设置各项音频控制参数。

在本实施例中，具体的：

S7031、耳返开关控件的开关指令，设置耳返效果开关状态所对应的控制参数，并对应切换耳返开关控件的开关显示状态。

和/或，

S7032、响应于用户对闪避开关控件的开关指令，设置闪避效果开关状态所对应的控制参数，并对应切换闪避开关控件的开关显示状态。

和/或，

S7033、响应于用户对扬声器选择控件的选择指令，设置直播设备中各个扬声器的工作状态，并对应切换扬声器选择控件的扬声器选中状态。

和/或，

S7034、响应于用户对麦克风选择控件的选择指令，设置直播设备中各个麦克风的工作状态，并对应切换麦克风选择控件的麦克风选中状态。

和/或，

S7035、响应于用户对连麦音量设置控件的第一音量设置操作，设置连麦时对方音量的大小控制参数，并对应改变连麦音量设置控件的显示状态；

和/或，

S7036、响应于用户对音效选择控件的选择指令，设置目标音频合成时的音效控制参数，并对应切换音效选择控件的音效选中状态。

和/或，

S7037、响应于用户对音效音量控件的第二音量设置操作，设置音效对应的音量大小控制参数，并对应改变音效音量控件的显示状态。

和/或，

S7038、响应于用户对已选择快捷键阵列的删除指令，更新已选择快捷键阵列以及待选快捷键阵列的展示状态。

和/或，

S7039、响应于用户对待选快捷键阵列的选择指令，更新已选择快捷键阵列以及待选快捷键阵列的展示状态。

和/或，

S70310、响应于用户对合成控件（如悬浮窗）显示开关的开关指令，对应设置合成控件是否在显示屏上显示。

需要说明的是，S7031-S70310 并没有先后顺序的要求，用户可以同时设置几个参数。

具体的，一种可能的实现方式可以参考图 8。

图 8 为本公开实施例提供的一种声音设置界面示意图。如图 8 所示，用户在声音设置界面上，通过扬声器选择控件，用户可以选择用“有线耳机”或者是直播一体设备自带的扬声器，即“本机扬声器”；通过麦克风选择控件，用户可以选择“有线耳麦”、“本机麦克风”、“卡农麦克风”；通过耳返开关控件，用户通过向右/向左拨动开关控件，开启或关闭耳返效果，并且在耳返开关控件上，还包括提示文本，如“本机扬声器不支持耳返功能”，提示用户耳返功能与扬声器选择直接是具备关联设置的，以避免用户参数选择不当，影响使用效果；同理，闪避开关控件和悬浮窗开关控件也具备类似的操作，具体内容参见图 8。

图 8 中的美声选项即为音效美化效果的选择控件，通过单击右侧展开图标“>”，即可展开音效列表，供用户选择对应的音效美化效果，在本公开的各个实施例中，音效美化效果包括：原声、KTV、磁性、老唱片、空灵、剧场等音效特性，每种音效对应着多个音效设置参数，通过选择特定音效，可以快速设置各个参数，降低了音效设置的专业门槛，使得不具备专业技术的主播也可以进行简单方便的音效设置。

图 8 的最下方为悬浮窗即合成控件的快捷键选择阵列，用户只需要选中各个按钮控件，则设置显示对应的待合成音频的快捷键，便于用户在直播开始后，在直播界面中通过快捷键直接进行音频合成。

本公开实施例提供的音频合成方法，在图 2、图 3 和图 5 所示实施例的基础上，通过在没有开启摄像头或者直播开启之前，获取用户的声音设置操作，然后响应于用户的声音设置操作，在显示屏上显示声音设置界面，再根据用户对声音设置界面上各个参数控件的设置指令，对应设置各项音频控制参数，解决了主播无法简便快捷地更改各项声音相关的设置参数，需要专业技术人员辅助的技术问题。使得声音设置和合成控件的快捷键设置简单方便，提高了用户对于直播一体式终端的使用体验感。

在一种可能的设计中，在本公开的各个实施例的基础上，直播一体机即直播一体化设备通过外接接口连接至少一个快捷键键盘，步骤 S202：接收用户对合成控件的音频合成操作，包括；

接收快捷键键盘发送的操作指令，操作指令是快捷键键盘响应于用户的操作而确定的；

根据操作指令确定对应的待合成音频。

值得说明的是，外接接口包括：无线接口以及有线接口。

图 9 为本公开实施例提供的一种直播一体化设备外接快捷键键盘的场景示意图。如图 9 所示，直播一体机 91 通过有线接口或无线接口连接快捷键键盘 92，用于实现本公开各个实施例的音频合成方法的合成控件 911 中的多个子控件与键盘上的按键呈一一对应关系。

在一种可能的设计中，快捷键键盘上各个按键与显示屏上各个子控件的对应关系是可以任意设置的。快捷键键盘包括：机械按钮式键盘、触控屏等。用户点击快捷键相当于在显示屏上点击了对应的子控件，同时，子控件切换显示为被选中的状态。

还需要说明的是，子控件在被点击选中后，可以用条形进度条或扇形进度图的形式，同步显示待合成音频的播放进度。

图 10a-10c 为本公开实施例提供的一种合成控件在待合成音频播放时显示播放进度的示意图。如图 10a 所示，在用户点击了“笑声 1”子控件后，在子控件上显示播放“笑声 1”所对应的音频的播放进度条。

如图10b所示，在用户点击了“笑声1”子控件后，子控件的背景颜色或背景图案以旋转渐变的形式来显示“笑声1”所对应的音频的播放进度条。

如图10c所示，在用户点击了“笑声1”子控件后，子控件的背景颜色或背景图案以从左至右、从右至左、从上至下、从下至上、从左下至右上、从右下至左上等方式中的任意一种渐变的形式来显示“笑声1”所对应的音频的播放进度条。

这样主播就能够知道待合成音频的播放进度，进一步的，可以通过再次点击子控件结束播放，或者是点击其它子控件混合播放多个待合成音频，增加了直播音频合成的操作灵活性，提升直播时的现场音频效果，提高用户体验。

对应于上文实施例的音频合成方法，图11为本公开实施例提供的音频合成装置1100的结构框图。为了便于说明，仅示出了与本公开实施例相关的部分。参照图11，装置包括：

显示模块1101，用于在满足预设显示条件时，以第一形态在显示屏上显示合成控件，所述第一形态对所述显示屏上显示内容的视觉影响满足预设视觉要求；

接收模块1102，用于接收用户对所述合成控件的音频合成操作；

处理模块1103，用于在安装了直播应用的直播设备中，将待合成音频与初始音频进行合成，并将合成后的目标音频作为所述直播应用的输入音频，其中，所述待合成音频为所述音频合成操作所指示的音频，所述初始音频为所述直播设备通过音频采集设备接收到主播所发出的，和/或，直播环境中的声音信号。

在本公开的一个实施例中，所述显示屏包括触控屏，所述接收模块1102，用于实时检测所述用户利用预设操作物靠近或触碰到所述触控屏上的所述合成控件。

在本公开的一个实施例中，所述合成控件包括与所述待合成音频对应的子控件，所述接收模块1102，具体用于：

若检测到所述预设操作物靠近或触碰到所述子控件，则将所述子控件切换为被选中形态；

确定与所述子控件对应的合成操作指令。

在本公开的一个实施例中，所述预设显示条件包括：检测到摄像设备开启指令、检测到直播终端按预设运动方式进行运动、接收到预设语音指令和接收到预设按钮开关开启指令中的至少一种。

在本公开的一个实施例中，所述合成控件包括悬浮窗控件，所述显示模块1101，用于根据预设透明度、预设大小以及预设形状在显示屏的边缘显示所述悬浮窗控件。

在本公开的一个实施例中，所述接收模块1102，还用于获取所述用户对所述合成控件的状态设置操作；

所述显示模块1101，还用于根据所述状态设置操作切换所述合成控件的形态，和/或，显示位置。

在本公开的一个实施例中，所述显示模块1101，还用于响应于所述用户对所述合成控件的弹窗设置操作，在显示屏上叠加显示设置窗口；

所述处理模块1103，还用于根据所述用户对所述设置窗口上各个参数控件的设置指令，对应设置各项音频控制参数。

在本公开的一个实施例中，所述设置窗口包括声音设置选项卡，所述声音设置选项卡包括：耳返开关控件，和/或，闪避开关控件，所述处理模块1103，具体用于：

响应于所述用户对所述耳返开关控件的开关指令，设置耳返效果开关状态所对应的控制

参数，并对应切换所述耳返开关控件的开关显示状态；

和/或，

响应于所述用户对所述闪避开关控件的开关指令，设置闪避效果开关状态所对应的控制参数，并对应切换所述闪避开关控件的开关显示状态。

在本公开的一个实施例中，所述设置窗口包括声音设置选项卡，所述声音设置选项卡包括：扬声器选择控件，所述处理模块 1103，具体用于：

响应于所述用户对所述扬声器选择控件的选择指令，设置所述直播设备中各个扬声器的工作状态，并对应切换所述扬声器选择控件的扬声器选中状态。

在本公开的一个实施例中，所述设置窗口包括声音设置选项卡，所述声音设置选项卡包括：麦克风选择控件，所述处理模块 1103，具体用于：

响应于所述用户对所述麦克风选择控件的选择指令，设置所述直播设备中各个麦克风的工作状态，并对应切换所述麦克风选择控件的麦克风选中状态。

在本公开的一个实施例中，所述设置窗口包括声音设置选项卡，所述声音设置选项卡包括：连麦音量设置控件，所述处理模块 1103，具体用于：

响应于所述用户对所述连麦音量设置控件的第一音量设置操作，设置连麦时对方音量的大小控制参数，并对应改变所述连麦音量设置控件的显示状态。

在本公开的一个实施例中，所述设置窗口包括声音设置选项卡，所述声音设置选项卡包括：音效选择控件，和/或，音效音量控件，所述处理模块 1103，具体用于：

响应于所述用户对所述音效选择控件的选择指令，设置所述目标音频合成时的音效控制参数，并对应切换所述音效选择控件的音效选中状态；

和/或，

响应于所述用户对所述音效音量控件的第二音量设置操作，设置音效对应的音量大小控制参数，并对应改变所述音效音量控件的显示状态。

在本公开的一个实施例中，所述设置窗口包括快捷键设置选项卡，所述快捷键设置选项卡包括：已选择快捷键阵列以及待选快捷键阵列，所述处理模块 1103，具体用于：

响应于所述用户对所述已选择快捷键阵列的删除指令，更新所述已选择快捷键阵列以及所述待选快捷键阵列的展示状态；

和/或，

响应于所述用户对所述待选快捷键阵列的选择指令，更新所述已选择快捷键阵列以及所述待选快捷键阵列的展示状态。

在本公开的一个实施例中，在所述在满足预设显示条件时，以第一形态在显示屏上显示合成控件之前，所述显示模块 1101，还用于响应于所述用户的声音设置操作，在显示屏上显示声音设置界面；

所述处理模块 1103，还用于根据所述用户对所述声音设置界面上各个参数控件的设置指令，对应设置各项音频控制参数。

在本公开的一个实施例中，所述声音设置界面包括：耳返开关控件，和/或，闪避开关控件，所述处理模块 1103，还用于：

响应于所述用户对所述耳返开关控件的开关指令，设置耳返效果开关状态所对应的控制参数，并对应切换所述耳返开关控件的开关显示状态；

和/或，

响应于所述用户对所述闪避开关控件的开关指令，设置闪避效果开关状态所对应的控制参数，并对应切换所述闪避开关控件的开关显示状态。

在本公开的一个实施例中，所述声音设置界面包括：扬声器选择控件，所述处理模块 1103，还用于：

响应于所述用户对所述扬声器选择控件的选择指令，设置所述直播设备中各个扬声器的工作状态，并对应切换所述扬声器选择控件的扬声器选中状态。

在本公开的一个实施例中，所述声音设置界面包括：麦克风选择控件，所述处理模块 1103，还用于：

响应于所述用户对所述麦克风选择控件的选择指令，设置所述直播设备中各个麦克风的工作状态，并对应切换所述麦克风选择控件的麦克风选中状态。

在本公开的一个实施例中，所述声音设置界面包括：连麦音量设置控件，所述处理模块 1103，还用于：

响应于所述用户对所述连麦音量设置控件的第一音量设置操作，设置连麦时对方音量的大小控制参数，并对应改变所述连麦音量设置控件的显示状态。

在本公开的一个实施例中，所述声音设置界面包括：音效选择控件，和/或，音效音量控件，所述处理模块 1103，还用于：

响应于所述用户对所述音效选择控件的选择指令，设置所述目标音频合成时的音效控制参数，并对应切换所述音效选择控件的音效选中状态；

和/或，

响应于所述用户对所述音效音量控件的第二音量设置操作，设置音效对应的音量大小控制参数，并对应改变所述音效音量控件的显示状态。

在本公开的一个实施例中，所述声音设置界面包括：已选择快捷键阵列以及待选快捷键阵列，所述处理模块 1103，还用于：

响应于所述用户对所述已选择快捷键阵列的删除指令，更新所述已选择快捷键阵列以及所述待选快捷键阵列的展示状态；

和/或，

响应于所述用户对所述待选快捷键阵列的选择指令，更新所述已选择快捷键阵列以及所述待选快捷键阵列的展示状态。

在本公开的一个实施例中，所述声音设置界面包括：合成控件显示开关，所述处理模块 1103，还用于：

响应于所述用户对所述合成控件显示开关的开关指令，对应设置所述合成控件是否在显示屏上显示。

在本公开的一个实施例中，所述音频合成装置 1100 通过外接接口连接至少一个快捷键键盘，所述接收用户对所述合成控件的音频合成操作，包括；

所述接收模块 1102，还用于接收所述快捷键键盘发送的操作指令，所述操作指令是所述快捷键键盘响应于所述用户的操作而确定的；

所述处理模块 1103，还用于根据所述操作指令确定对应的所述待合成音频。

本实施例提供的装置 1100，可用于执行上述方法实施例的技术方案中用户端相关的步

骤，其实现原理和技术效果类似，本实施例此处不再赘述。

参考图 12，其示出了适于用来实现本公开实施例的电子设备的结构示意图，该电子设备 1200 可以为终端设备或服务器。其中，终端设备可以包括但不限于诸如移动电话、笔记本电脑、数字广播接收器、个人数字助理（Personal Digital Assistant，简称 PDA）、平板电脑（Portable Android Device，简称 PAD）、便携式多媒体播放器（Portable Media Player，简称 PMP）、车载终端（例如车载导航终端）等等的移动终端以及诸如数字 TV、台式计算机等等的固定终端。图 12 示出的电子设备仅仅是一个示例，不应对本公开实施例的功能和使用范围带来任何限制。

如图 12 所示，电子设备 1200 可以包括处理装置（例如中央处理器、图形处理器等）1201，其可以根据存储在只读存储器（Read Only Memory，简称 ROM）1202 中的程序或者从存储装置 1208 加载到随机访问存储器（Random Access Memory，简称 RAM）1203 中的程序而执行各种适当的动作和处理。在 RAM 1203 中，还存储有电子设备 1200 操作所需的各种程序和数据。处理装置 1201、ROM 1202 以及 RAM 1203 通过总线 1204 彼此相连。输入/输出（I/O）接口 1205 也连接至总线 1204。

通常，以下装置可以连接至 I/O 接口 1205：包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置 12012；包括例如液晶显示器（Liquid Crystal Display，简称 LCD）、扬声器、振动器等的输出装置 1207；包括例如磁带、硬盘等的存储装置 1208；以及通信装置 1209。通信装置 1209 可以允许电子设备 1200 与其他设备进行无线或有线通信以交换数据。虽然图12示出了具有各种装置的电子设备1200，但是应理解的是，并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。

特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括承载在计算机可读介质上的计算机程序，该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信装置 1209 从网络上被下载和安装，或者从存储装置 1208 被安装，或者从 ROM 1202 被安装。在该计算机程序被处理装置 1201 执行时，执行本公开实施例的方法中限定的上述功能。

需要说明的是，本公开上述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（Erasable Programmable Read-Only Memory，简称 EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（Compact Disc Read-Only Memory，简称 CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开中，计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读信号介质可以发送、传播或者传输用

于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：电线、光缆、射频（Radio Frequency，简称 RF）等等，或者上述的任意合适的组合。

上述计算机可读介质可以是上述电子设备中所包含的；也可以是单独存在，而未装配入该电子设备中。

上述计算机可读介质承载有一个或者多个程序，当上述一个或者多个程序被该电子设备执行时，使得该电子设备执行上述实施例所示的方法。

可以以一种或多种程序设计语言或其组合来编写用于执行本公开的操作的计算机程序代码，上述程序设计语言包括面向对象的程序设计语言—诸如 Java、Smalltalk、C++，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网（Local Area Network，简称 LAN）或广域网（Wide Area Network，简称 WAN）—连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

附图中的流程图和框图，图示了按照本公开各种实施例的系统、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，该模块、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，两个接连地表示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。也要注意的是，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操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。

描述于本公开实施例中所涉及到的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。其中，单元的名称在某种情况下并不构成对该单元本身的限定，例如，第一获取单元还可以被描述为“获取至少两个网际协议地址的单元”。

本文中以上描述的功能可以至少部分地由一个或多个硬件逻辑部件来执行。例如，非限制性地，可以使用的示范类型的硬件逻辑部件包括：现场可编程门阵列（Field Programmable Gate Array，简称 FPGA）、专用集成电路（Application Specific Intergrated Circuit，简称 ASIC）、专用标准产品（Application Specific Standard Product，简称 ASSP）、片上系统（System on a Chip，简称 SOC）、复杂可编程逻辑设备（Complex Programmable Logic Device，简称 CPLD）等等。

在本公开的上下文中，机器可读介质可以是有形的介质，其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦除可编程只读存储器

（EPROM 或快闪存储器）、光纤、便捷式紧凑盘只读存储器（CD-ROM）、光学储存设备、磁储存设备、或上述内容的任何合适组合。

本公开实施例中还提供了一种计算机程序产品，包括计算机程序，其特征在于，该计算机程序被处理器执行时实现上述各实施例中的方法。

本公开实施例中还提供了一种直播一体机或直播一体化设备，包括图 12 所对应的电子设备。还需要说明的是，该直播一体机的控制电路包括：主控模块和音频处理模块，其中，主控模块上安装了直播应用，音频处理模块用于将待合成音频合成到初始音频中，以形成目标音频。待合成音频可以存储在音频处理模块中，也可以存储在主控模块中，还可以存储在直播一体机的存储器中。

以上描述仅为本公开的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本公开中所涉及的公开范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述公开构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本公开中公开的（但不限于）具有类似功能的技术特征进行互相替换而形成的技术方案。

此外，虽然采用特定次序描绘了各操作，但是这不应当理解为要求这些操作以所示出的特定次序或以顺序次序执行来执行。在一定环境下，多任务和并行处理可能是有利的。同样地，虽然在上面论述中包含了若干具体实现细节，但是这些不应当被解释为对本公开的范围的限制。在单独的实施例的上下文中描述的某些特征还可以组合地实现在单个实施例中。相反地，在单个实施例的上下文中描述的各种特征也可以单独地或以任何合适的子组合的方式实现在多个实施例中。

尽管已经采用特定于结构特征和/或方法逻辑动作的语言描述了本主题，但是应当理解所附权利要求书中所限定的主题未必局限于上面描述的特定特征或动作。相反，上面所描述的特定特征和动作仅仅是实现权利要求书的示例形式。

权 利 要 求 书

1、一种音频合成方法，其特征在于，包括：

在满足预设显示条件时，以第一形态在显示屏上显示合成控件，所述第一形态对所述显示屏上显示内容的视觉影响满足预设视觉要求；

接收用户对所述合成控件的音频合成操作；

在安装了直播应用的直播设备中，将待合成音频与初始音频进行合成，并将合成后的目标音频作为所述直播应用的输入音频，其中，所述待合成音频为所述音频合成操作所指示的音频，所述初始音频为所述直播设备通过音频采集设备接收到主播所发出的，和/或，直播环境中的声音信号。

2、根据权利要求 1 所述的音频合成方法，其特征在于，所述显示屏包括触控屏，所述接收用户对所述合成控件的音频合成操作，包括：

实时检测所述用户利用预设操作物靠近或触碰到所述触控屏上的所述合成控件。

3、根据权利要求 2 所述的音频合成方法，其特征在于，所述合成控件包括与所述待合成音频对应的子控件，所述实时检测所述用户利用预设操作物靠近或触碰到所述触控屏上的所述合成控件，包括：

若检测到所述预设操作物靠近或触碰到所述子控件，则将所述子控件切换为被选中形态；

确定与所述子控件对应的合成操作指令。

4、根据权利要求 1 至 3 中任意一项所述的音频合成方法，其特征在于，所述预设显示条件包括：检测到摄像设备开启指令、检测到直播终端按预设运动方式进行运动、接收到预设语音指令和接收到预设按钮开关开启指令中的至少一种。

5、根据权利要求 1 至 4 中任意一项所述的音频合成方法，其特征在于，所述合成控件包括悬浮窗控件，所述以第一形态在显示屏上显示合成控件，包括：

根据预设透明度、预设大小以及预设形状在显示屏的边缘显示所述悬浮窗控件。

6、根据权利要求 1 至 5 中任意一项所述的音频合成方法，其特征在于，还包括：

获取所述用户对所述合成控件的状态设置操作；

根据所述状态设置操作切换所述合成控件的形态，和/或，显示位置。

7、根据权利要求 1 至 6 中任意一项所述的音频合成方法，其特征在于，还包括：

响应于所述用户在直播时对所述合成控件的弹窗设置操作，在显示屏上叠加显示设置窗口；

和/或，响应于所述用户在直播前触发的声音设置操作，在显示屏上显示声音设置界面；

根据所述用户对所述设置窗口或所述声音设置界面上各个参数控件的设置指令，对应设置各项音频控制参数。

8、根据权利要求 7 所述的音频合成方法，其特征在于，所述设置窗口包括声音设置选项卡，所述声音设置选项卡，和/或，所述声音设置界面中包括：耳返开关控件，和/或，闪避开关控件，所述根据所述用户对所述设置窗口或所述声音设置界面上各个参数控件的设置指令，对应设置各项音频控制参数，包括：

响应于所述用户对所述耳返开关控件的开关指令，设置耳返效果开关状态所对应的控制参数，并对应切换所述耳返开关控件的开关显示状态；

和/或，

响应于所述用户对所述闪避开关控件的开关指令，设置闪避效果开关状态所对应的控制参数，并对应切换所述闪避开关控件的开关显示状态。

9、根据权利要求 7 或 8 所述的音频合成方法，其特征在于，所述设置窗口包括声音设置选项卡，所述声音设置选项卡，和/或，所述声音设置界面中包括：扬声器选择控件，所述根据所述用户对所述设置窗口或所述声音设置界面上各个参数控件的设置指令，对应设置各项音频控制参数，包括：

响应于所述用户对所述扬声器选择控件的选择指令，设置所述直播设备中各个扬声器的工作状态，并对应切换所述扬声器选择控件的扬声器选中状态。

10、根据权利要求 7 至 9 中任意一项所述的音频合成方法，其特征在于，所述设置窗口包括声音设置选项卡，所述声音设置选项卡，和/或，所述声音设置界面中包括：麦克风选择控件，所述根据所述用户对所述设置窗口或所述声音设置界面上各个参数控件的设置指令，对应设置各项音频控制参数，包括：

响应于所述用户对所述麦克风选择控件的选择指令，设置所述直播设备中各个麦克风的工作状态，并对应切换所述麦克风选择控件的麦克风选中状态。

11、根据权利要求 7 至 10 中任意一项所述的音频合成方法，其特征在于，所述设置窗口包括声音设置选项卡，所述声音设置选项卡，和/或，所述声音设置界面中包括：连麦音量设置控件，所述根据所述用户对所述设置窗口或所述声音设置界面上各个参数控件的设置指令，对应设置各项音频控制参数，包括：

响应于所述用户对所述连麦音量设置控件的第一音量设置操作，设置连麦时对方音量的大小控制参数，并对应改变所述连麦音量设置控件的显示状态。

12、根据权利要求 7 至 11 中任意一项所述的音频合成方法，其特征在于，所述设置窗口包括声音设置选项卡，所述声音设置选项卡，和/或，所述声音设置界面中包括：音效选择控件，和/或，音效音量控件，所述根据所述用户对所述设置窗口或所述声音设置界面上各个参数控件的设置指令，对应设置各项音频控制参数，包括：

响应于所述用户对所述音效选择控件的选择指令，设置所述目标音频合成时的音效控制参数，并对应切换所述音效选择控件的音效选中状态；

和/或，

响应于所述用户对所述音效音量控件的第二音量设置操作，设置音效对应的音量大小控制参数，并对应改变所述音效音量控件的显示状态。

13、根据权利要求 7 至 12 中任意一项所述的音频合成方法，其特征在于，所述设置窗口包括快捷键设置选项卡，所述快捷键设置选项卡，和/或，所述声音设置界面中包括：已选择快捷键阵列以及待选快捷键阵列，所述根据所述用户对所述设置窗口或所述声音设置界面上各个参数控件的设置指令，对应设置各项音频控制参数，包括：

响应于所述用户对所述已选择快捷键阵列的删除指令，更新所述已选择快捷键阵列以及所述待选快捷键阵列的展示状态；

和/或，

响应于所述用户对所述待选快捷键阵列的选择指令，更新所述已选择快捷键阵列以及所述待选快捷键阵列的展示状态。

14、根据权利要求 1 至 13 中任意一项所述的音频合成方法，其特征在于，所述方法应用

于直播一体机，所述直播一体机包括：主控模块以及音频处理模块，所述直播应用安装于所述主控模块中，所述音频处理模块用于合成所述目标音频。

15、根据权利要求 14 所述的音频合成方法，其特征在于，所述直播一体机通过外接接口连接至少一个快捷键键盘，所述接收用户对所述合成控件的音频合成操作，包括；

接收所述快捷键键盘发送的操作指令，所述操作指令是所述快捷键键盘响应于所述用户的操作而确定的；

根据所述操作指令确定对应的所述待合成音频。

16、一种音频合成装置，其特征在于，包括：

显示模块，用于在满足预设显示条件时，以第一形态在显示屏上显示合成控件，所述第一形态对所述显示屏上显示内容的视觉影响满足预设视觉要求；

接收模块，用于接收用户对所述合成控件的音频合成操作；

处理模块，用于在安装了直播应用的直播设备中，将待合成音频与初始音频进行合成，并将合成后的目标音频作为所述直播应用的输入音频，其中，所述待合成音频为所述音频合成操作所指示的音频，所述初始音频为所述直播设备通过音频采集设备接收到主播所发出的，和/或，直播环境中的声音信号。

17、一种电子设备，其特征在于，包括：

至少一个处理器和存储器；

所述存储器用于存储计算机程序；

所述至少一个处理器执行所述存储器存储的所述计算机程序，使得所述至少一个处理器执行如权利要求 1 至 15 中任意一项所述的音频合成方法。

18、一种直播一体机，其特征在于，包括：权利要求 17 所述的电子设备。

19、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质中存储有计算机程序，当处理器执行所述计算机程序时，实现如权利要求 1 至 15 中任意一项所述的音频合成方法。

20、一种计算机程序产品，包括计算机程序，其特征在于，所述计算机程序被处理器执行时实现权利要求 1 至 15 中任意一项所述的音频合成方法。

21、一种计算机程序，其特征在于，所述计算机程序被处理器执行时实现权利要求 1 至 15 中任意一项所述的音频合成方法。

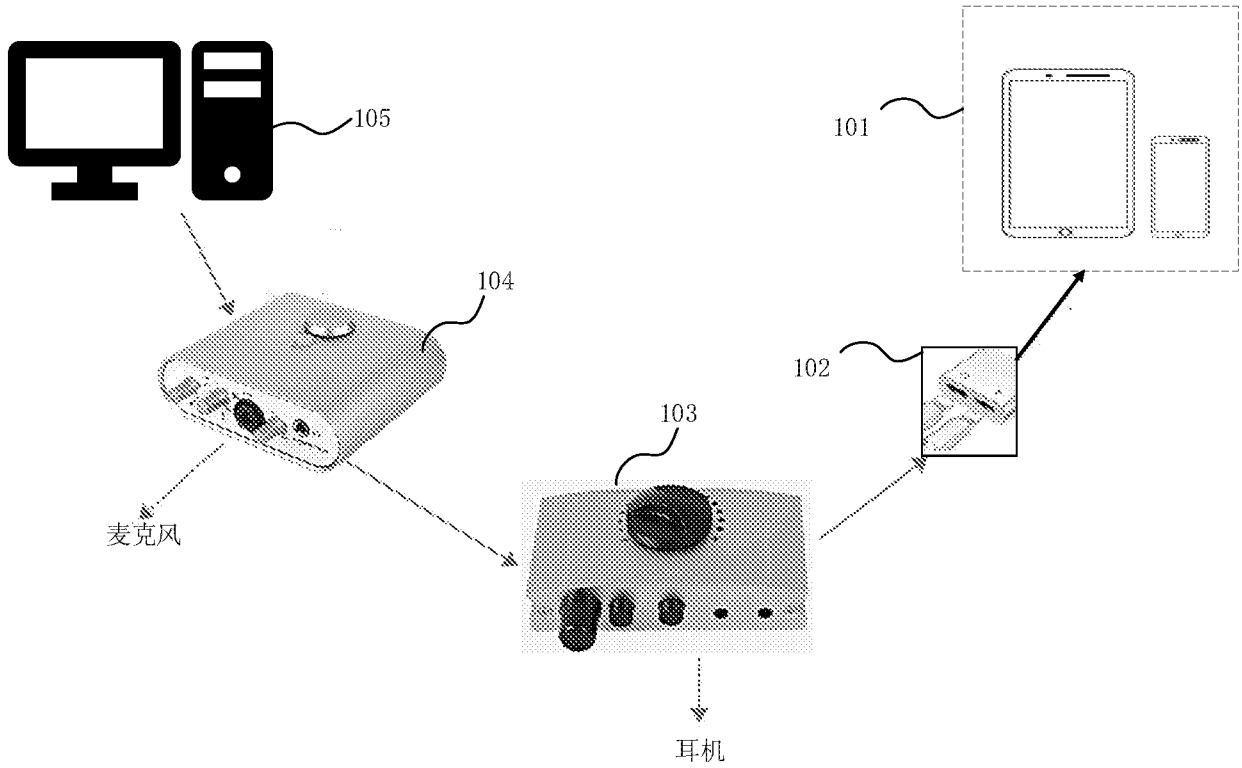

105
101
104
102
103
麦克风
耳机

图 1

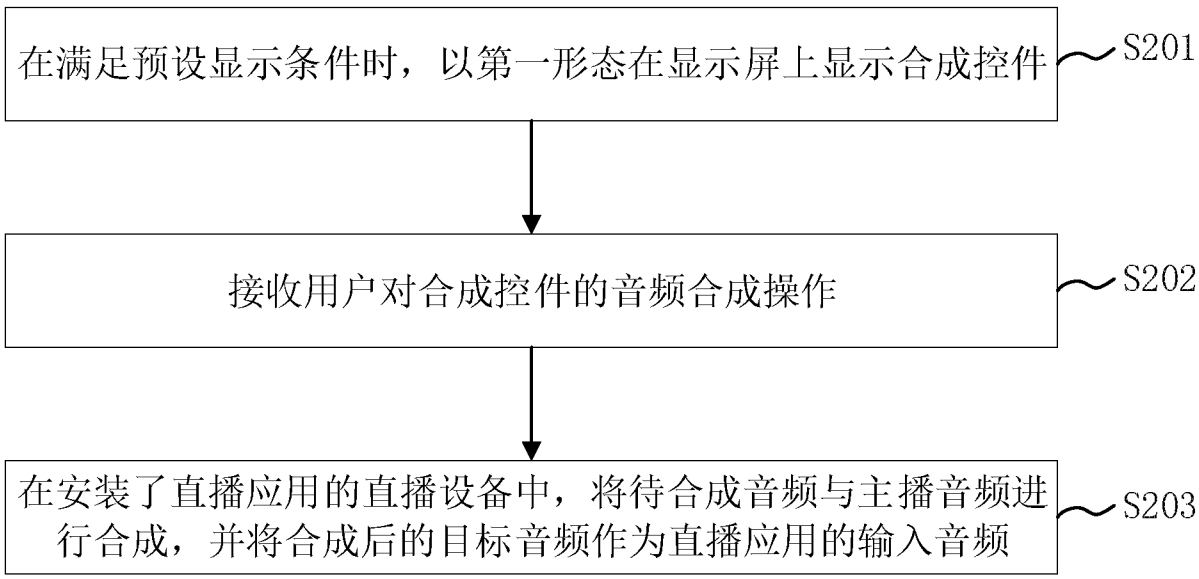

在满足预设显示条件时，以第一形态在显示屏上显示合成控件
S201
接收用户对合成控件的音频合成操作
S202
在安装了直播应用的直播设备中，将待合成音频与主播音频进行合成，并将合成后的目标音频作为直播应用的输入音频
S203

图 2

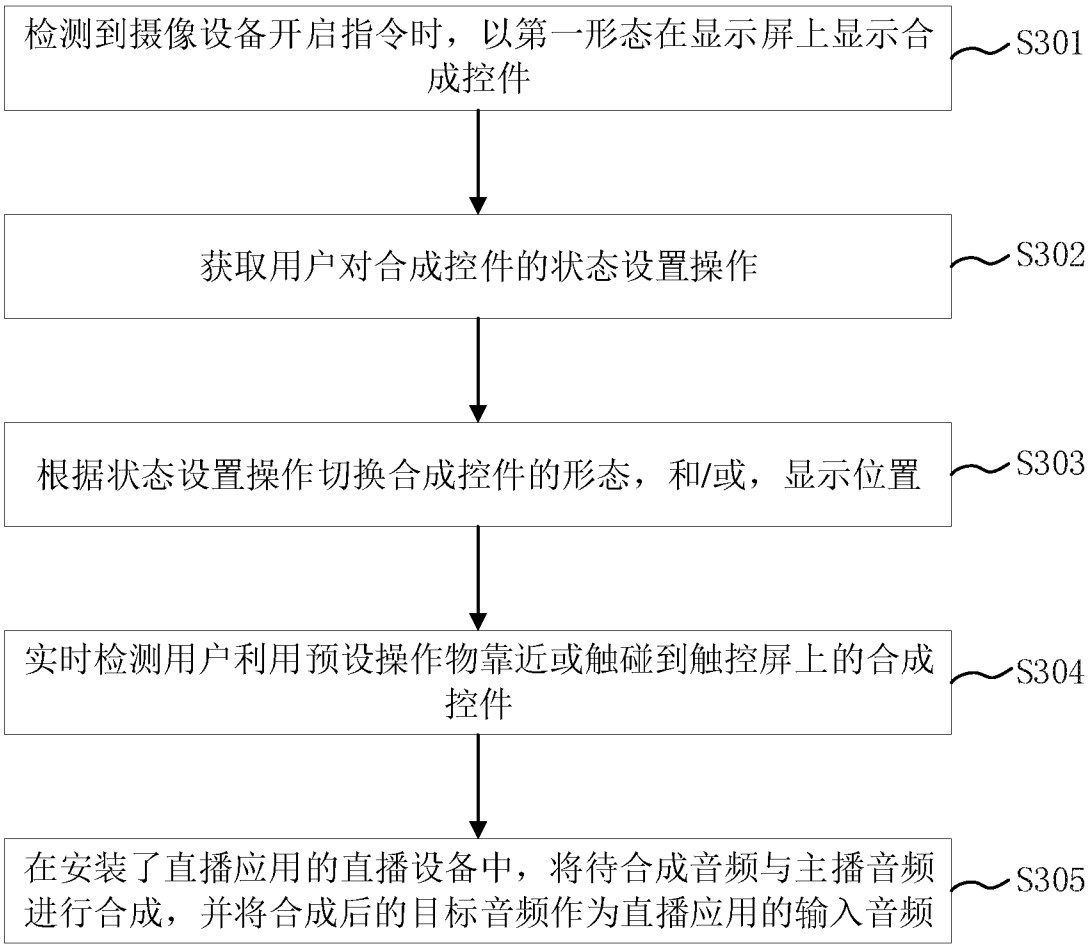

检测到摄像设备开启指令时，以第一形态在显示屏上显示合成控件
S301
获取用户对合成控件的状态设置操作
S302
根据状态设置操作切换合成控件的形态，和/或，显示位置
S303
实时检测用户利用预设操作物靠近或触碰到触控屏上的合成控件
S304
在安装了直播应用的直播设备中，将待合成音频与主播音频进行合成，并将合成后的目标音频作为直播应用的输入音频
S305

图 3

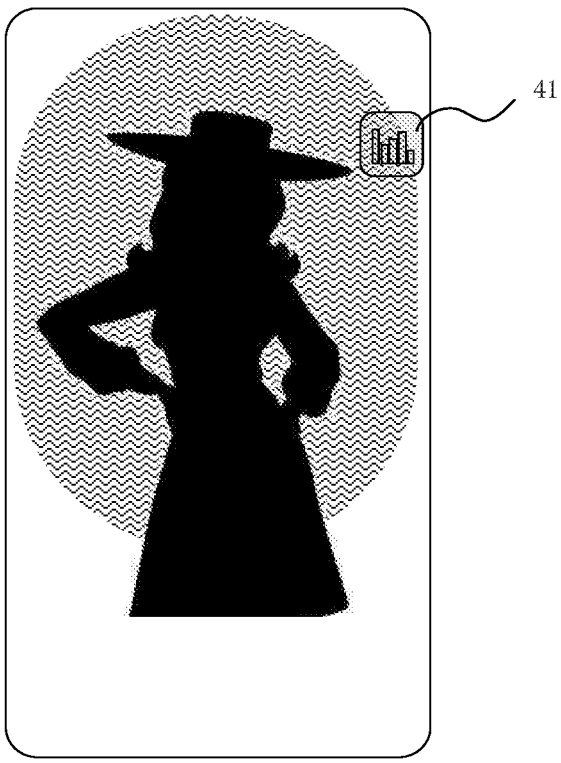

41

图 4a

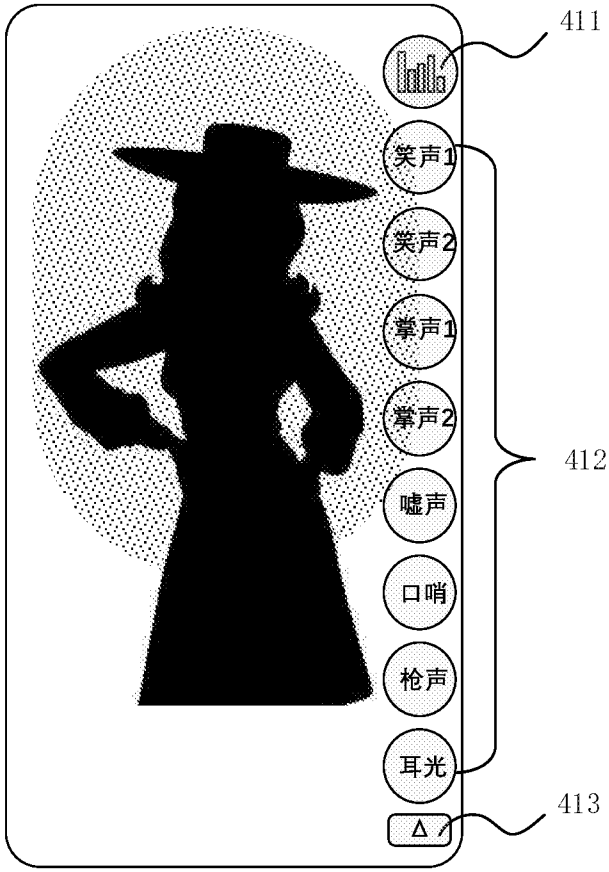


图 4b

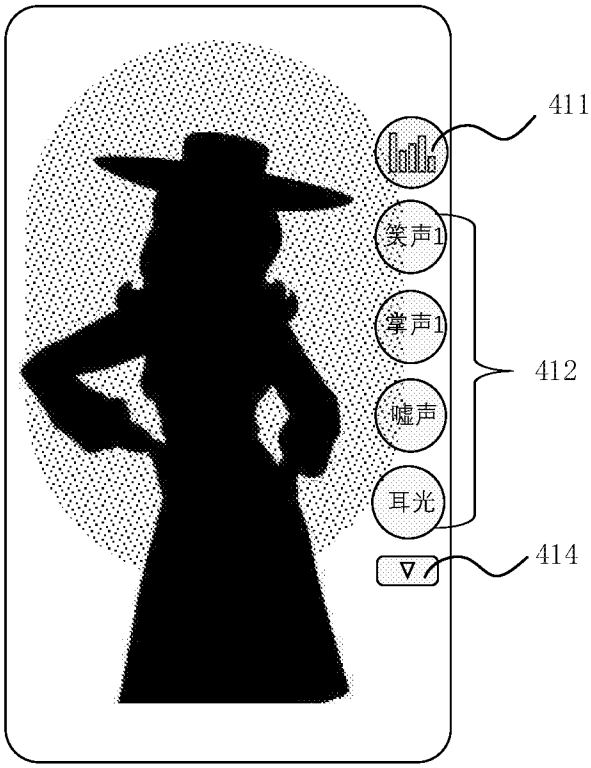


图 4c


获取用户对合成控件的弹窗设置操作
S501
响应于弹窗设置操作，在显示屏上叠加显示设置窗口
S502
响应于用户对各个控件的操作指令，对应设置各项音频控制参数
S5031、耳返开关控件的开关指令，设置耳返效果开关状态所对应的控制参数，并对应切换耳返开关控件的开关显示状态
S5032、响应于用户对闪避开关控件的开关指令，设置闪避效果开关状态所对应的控制参数，并对应切换闪避开关控件的开关显示状态
S5033、响应于用户对扬声器选择控件的选择指令，设置直播设备中各个扬声器的工作状态，并对应切换扬声器选择控件的扬声器选中状态
S5034、响应于用户对麦克风选择控件的选择指令，设置直播设备中各个麦克风的工作状态，并对应切换麦克风选择控件的麦克风选中状态
S5035、响应于用户对连麦音量设置控件的第一音量设置操作，设置连麦时对方音量的大小控制参数，并对应改变连麦音量设置控件的显示状态
S503
S5036、响应于用户对音效选择控件的选择指令，设置目标音频合成时的音效控制参数，并对应切换音效选择控件的音效选中状态
S5037、响应于用户对音效音量控件的第二音量设置操作，设置音效对应的音量大小控制参数，并对应改变音效音量控件的显示状态
S5038、响应于用户对已选择快捷键阵列的删除指令，更新已选择快捷键阵列以及待选快捷键阵列的展示状态
S5039、响应于用户对待选快捷键阵列的选择指令，更新已选择快捷键阵列以及待选快捷键阵列的展示状态

图 5

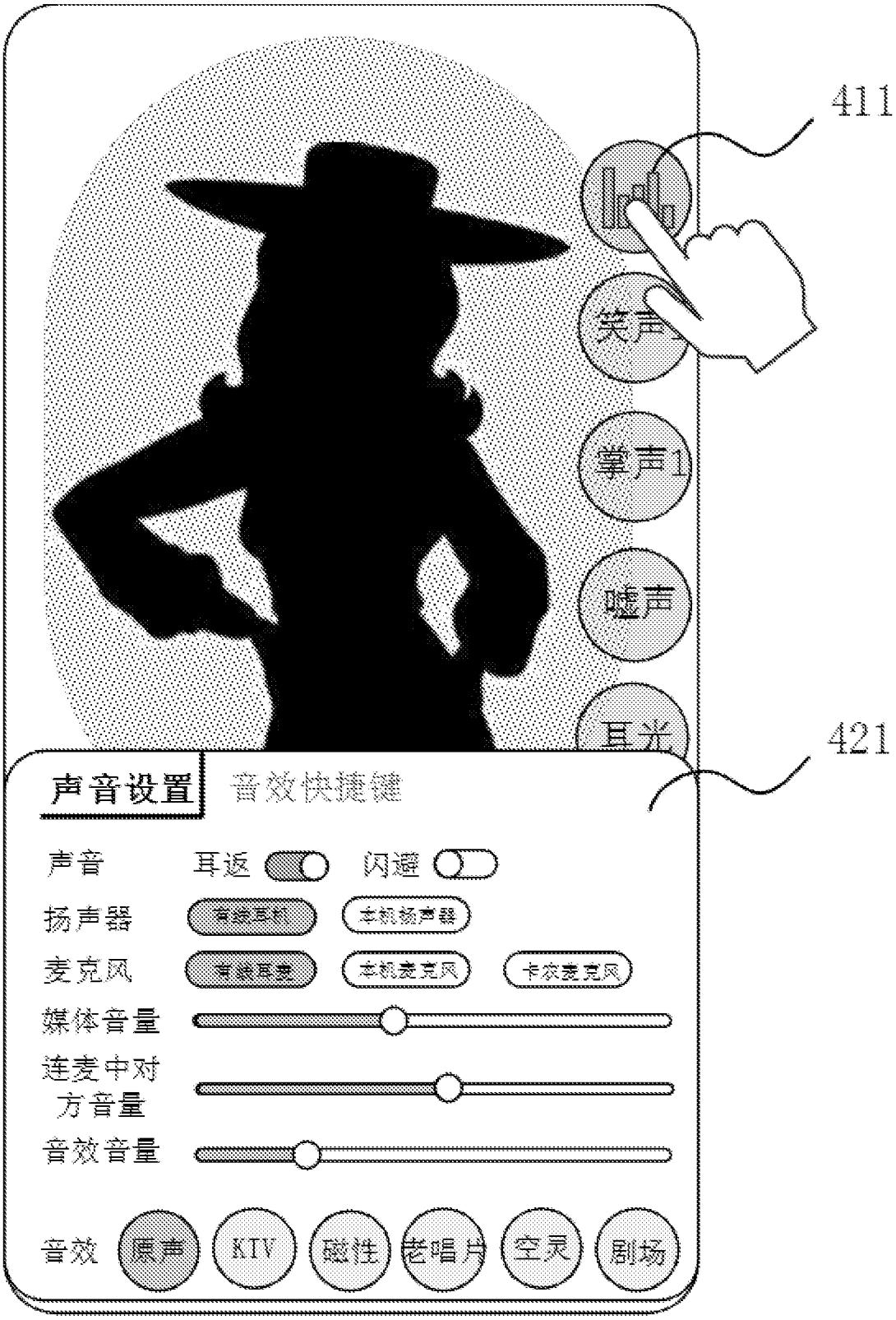

411
笑声
掌声1
嘘声
耳光
421
声音设置
音效快捷键
声音
耳返
闪避
扬声器
有线耳机
主机扬声器
麦克风
有线耳麦
主机麦克风
卡农麦克风
媒体音量
连麦中对方音量
音效音量
音效
原声
KTV
磁性
老唱片
空灵
剧场

图 6a

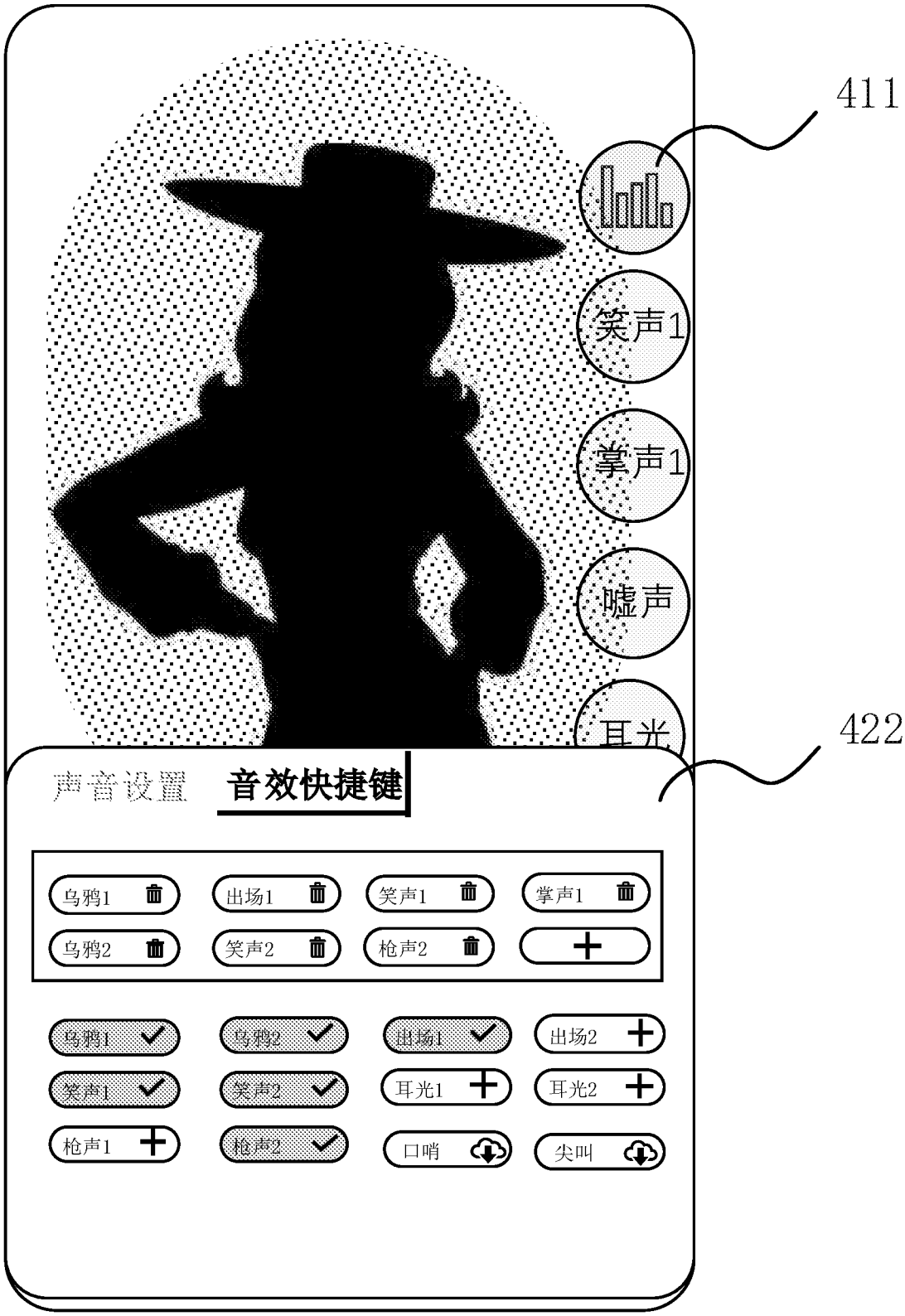

411
笑声1
掌声1
嘘声
耳光
422
声音设置
音效快捷键
乌鸦1
出场1
笑声1
掌声1
乌鸦2
笑声2
枪声2
乌鸦1
乌鸦2
出场1
出场2
笑声1
笑声2
耳光1
耳光2
枪声1
枪声2
口哨
尖叫

图 6b

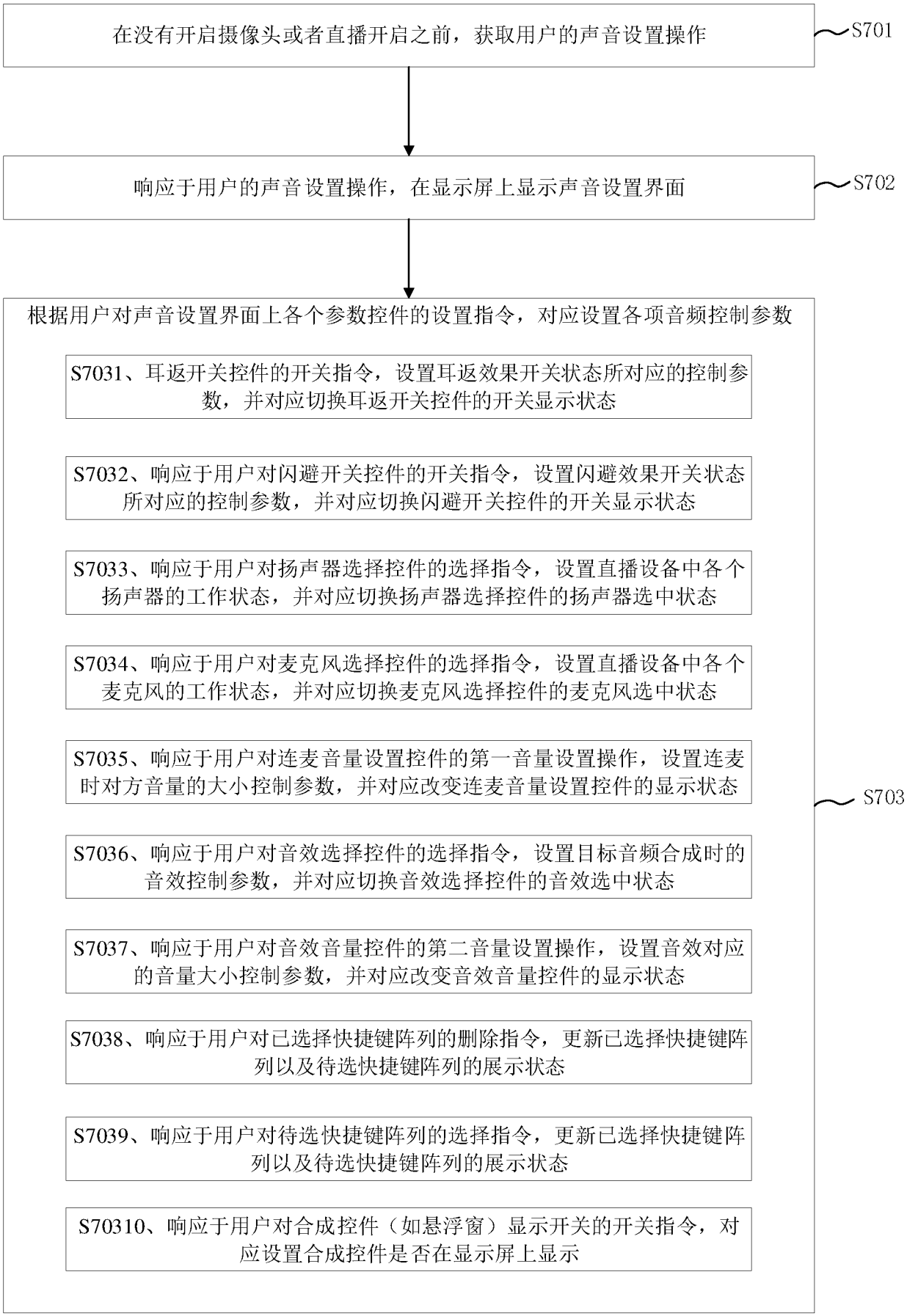

在没有开启摄像头或者直播开启之前，获取用户的声音设置操作
S701
响应于用户的声音设置操作，在显示屏上显示声音设置界面
S702
根据用户对声音设置界面上各个参数控件的设置指令，对应设置各项音频控制参数
S7031、耳返开关控件的开关指令，设置耳返效果开关状态所对应的控制参数，并对应切换耳返开关控件的开关显示状态
S7032、响应于用户对闪避开关控件的开关指令，设置闪避效果开关状态所对应的控制参数，并对应切换闪避开关控件的开关显示状态
S7033、响应于用户对扬声器选择控件的选择指令，设置直播设备中各个扬声器的工作状态，并对应切换扬声器选择控件的扬声器选中状态
S7034、响应于用户对麦克风选择控件的选择指令，设置直播设备中各个麦克风的工作状态，并对应切换麦克风选择控件的麦克风选中状态
S7035、响应于用户对连麦音量设置控件的第一音量设置操作，设置连麦时对方音量的大小控制参数，并对应改变连麦音量设置控件的显示状态
S7036、响应于用户对音效选择控件的选择指令，设置目标音频合成时的音效控制参数，并对应切换音效选择控件的音效选中状态
S7037、响应于用户对音效音量控件的第二音量设置操作，设置音效对应的音量大小控制参数，并对应改变音效音量控件的显示状态
S7038、响应于用户对已选择快捷键阵列的删除指令，更新已选择快捷键阵列以及待选快捷键阵列的展示状态
S7039、响应于用户对待选快捷键阵列的选择指令，更新已选择快捷键阵列以及待选快捷键阵列的展示状态
S70310、响应于用户对合成控件（如悬浮窗）显示开关的开关指令，对应设置合成控件是否在显示屏上显示
S703

图 7


声音设置
扬声器
有线耳机
麦克风
有线耳麦
卡农麦克风
耳返
闪避
悬浮窗
音效
笑声1
笑声2
乌鸦1
乌鸦2
出场1
出场2
耳光1
耳光2
枪声1
枪声2
嘘声1
嘘声2

图 8

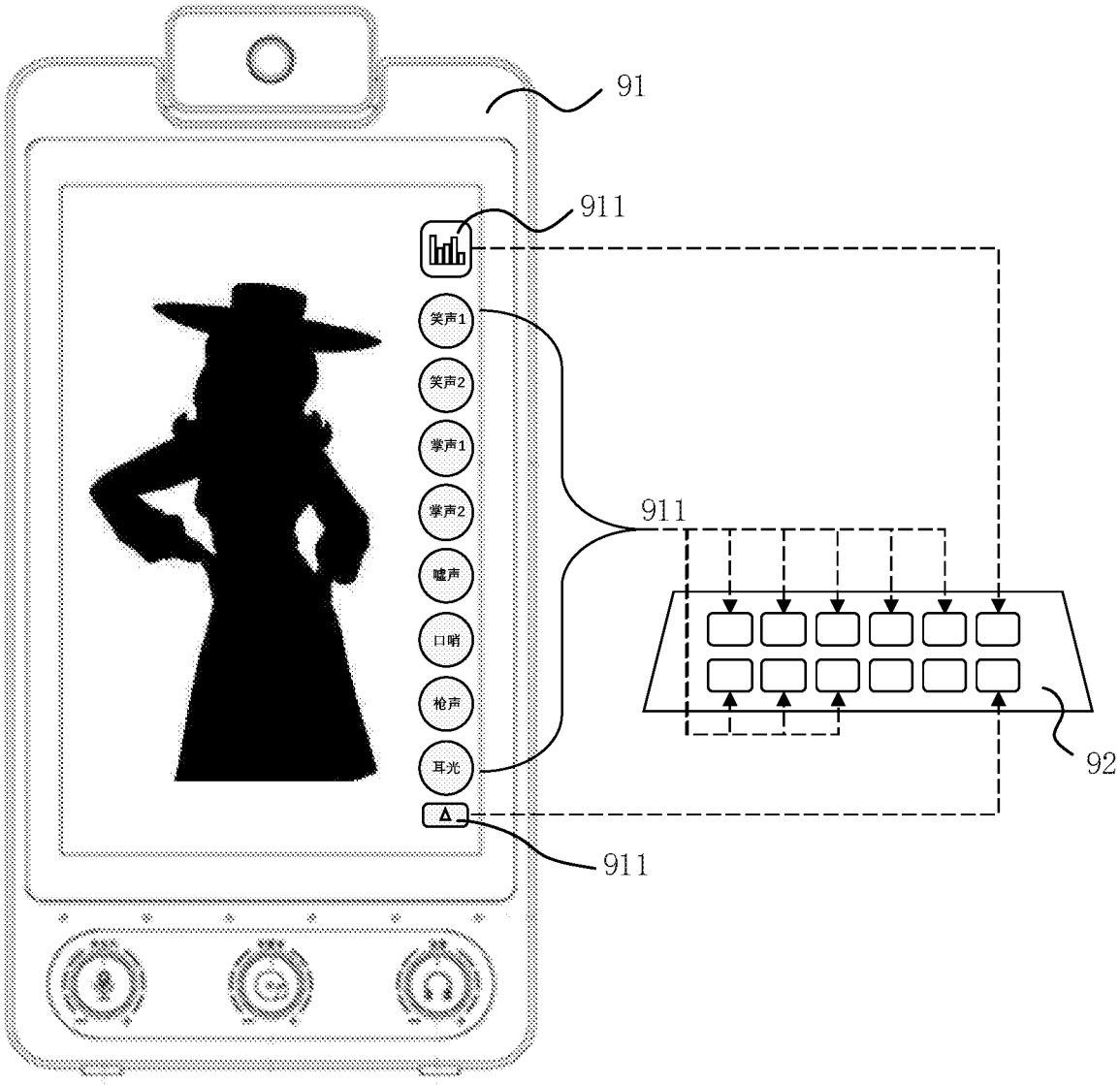


图 9



图 10a

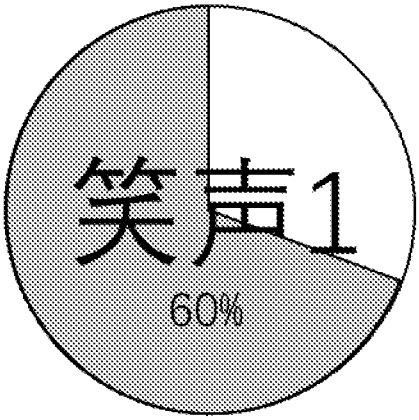


图 10b

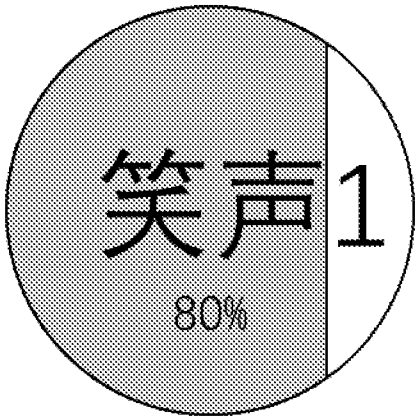


图 10c

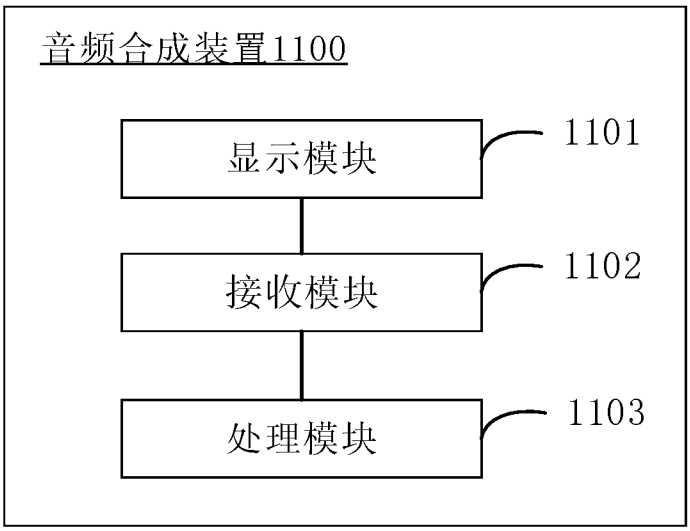


图 11

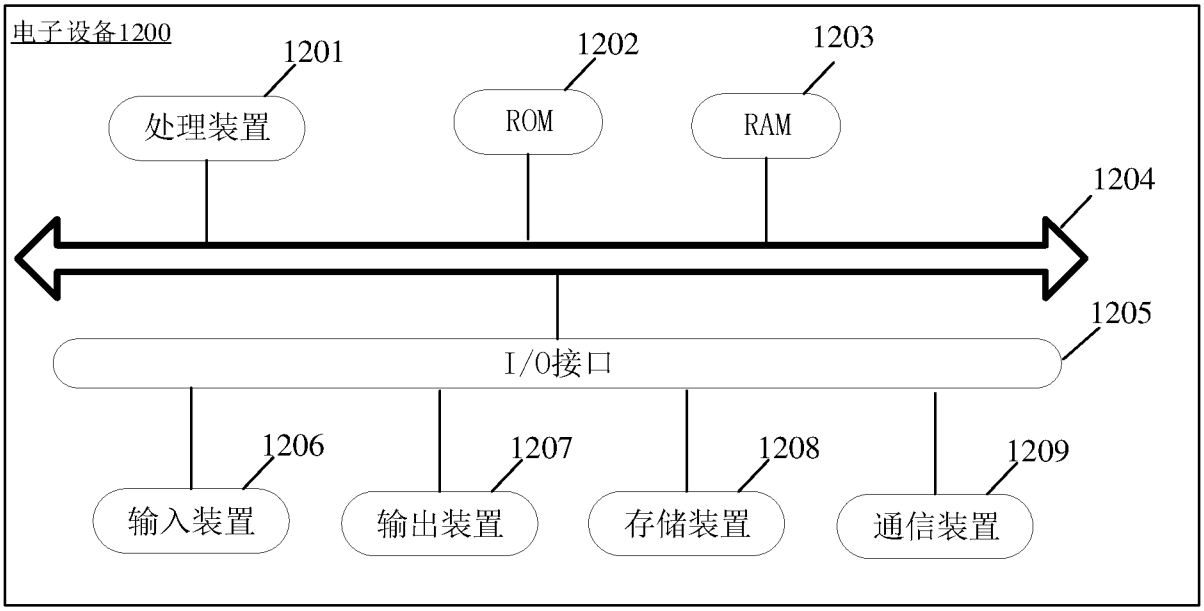

电子设备1200
1201
处理装置
1202
ROM
1203
RAM
1204
1205
I/O接口
1206
输入装置
1207
输出装置
1208
存储装置
1209
通信装置

图 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2022/087491

A. CLASSIFICATION OF SUBJECT MATTER

G10L 13/02(2013.01)i; H04N 21/439(2011.01)i; H04N 21/4788(2011.01)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; H04N

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; WOTXT; USTXT; EPTXT; CNKI: 北京字节跳动, 陈映宜, 直播, 音频, 音效, 合成, 混合, 叠加, 控件, 按钮, 按键, 透视, 透明, 显示, live streaming, direct seeding, live coverage, live telecast+, audio, voice, sound, mix+, overdub+, control+, button?, key+, lucency, transparency, perspectivity, display

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113112986 A (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) 13 July 2021 (2021-07-13) description paragraphs [0039]-[0321], and figures 1-12	1-21
Y	CN 112423009 A (ZHUHAI GREE ELECTRIC APPLIANCES INC.) 26 February 2021 (2021-02-26) description, paragraphs [0068]-[0175], and figures 1-9	1-21
Y	CN 111309214 A (NETEASE (HANGZHOU) NETWORK CO., LTD.) 19 June 2020 (2020-06-19) description, paragraphs [0051]-[0092], and figures 2-5	1-21
A	CN 111510738 A (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) 07 August 2020 (2020-08-07) entire document	1-21
A	CN 111880694 A (VIVO COMMUNICATION TECHNOLOGY CO., LTD.) 03 November 2020 (2020-11-03) entire document	1-21

☑ 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	Date of mailing of the international search report
20 May 2022	**22 June 2022**
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/CN2022/087491

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 112347273 A (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) 09 February 2021 (2021-02-09) entire document	1-21

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2022/087491

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 113112986 A	13 July 2021	None	
CN 112423009 A	26 February 2021	None	
CN 111309214 A	19 June 2020	None	
CN 111510738 A	07 August 2020	None	
CN 111880694 A	03 November 2020	WO 2022017410 A1	27 January 2022
CN 112347273 A	09 February 2021	None	

国际检索报告

国际申请号
PCT/CN2022/087491

A. 主题的分类

G10L 13/02(2013.01)i; H04N 21/439(2011.01)i; H04N 21/4788(2011.01)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; H04N

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;WOTXT;USTXT;EPTXT;CNKI:北京字节跳动，陈映宜，直播，音频，音效，合成，混合，叠加，控件，按钮，按键，透视，透明，显示，live streaming，direct seeding，live coverage，live telecast+，audio，voice，sound，mix+，overdub+，control+，button?，key+，lucency，transparency，perspectivity，display

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 113112986 A (北京字节跳动网络技术有限公司) 2021年7月13日 (2021-07-13) 说明书第[0039]-[0321]段，图1-12	1-21
Y	CN 112423009 A (珠海格力电器股份有限公司) 2021年2月26日 (2021-02-26) 说明书第[0068]-[0175]段，图1-9	1-21
Y	CN 111309214 A (网易杭州网络有限公司) 2020年6月19日 (2020-06-19) 说明书第[0051]-[0092]段，图2-5	1-21
A	CN 111510738 A (北京字节跳动网络技术有限公司) 2020年8月7日 (2020-08-07) 全文	1-21
A	CN 111880694 A (维沃移动通信有限公司) 2020年11月3日 (2020-11-03) 全文	1-21
A	CN 112347273 A (北京字节跳动网络技术有限公司) 2021年2月9日 (2021-02-09) 全文	1-21

☐ 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	国际检索报告邮寄日期
2022年5月20日	2022年6月22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	受权官员 马燕玲 电话号码 （86-512）88997418

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2022/087491

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 113112986 A	2021年7月13日	无	
CN 112423009 A	2021年2月26日	无	
CN 111309214 A	2020年6月19日	无	
CN 111510738 A	2020年8月7日	无	
CN 111880694 A	2020年11月3日	WO 2022017410 A1	2022年1月27日
CN 112347273 A	2021年2月9日	无	