

(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
2012年5月31日 (31.05.2012)



(10) 国际公布号
WO 2012/068960 A1

- (51) 国际专利分类号:
H04S 3/00 (2006.01)
- (21) 国际申请号: PCT/CN2011/082127
- (22) 国际申请日: 2011年11月14日 (14.11.2011)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201010561696.7 2010年11月26日 (26.11.2010) CN
- (71) 申请人 (对除美国外的所有指定国): 华为终端有限公司 (HUAWEI DEVICE CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为基地 B 区 2 号楼, Guangdong 518129 (CN)。
- (72) 发明人; 及
- (75) 发明人/申请人 (仅对美国): 岳中辉 (YUE, Zhonghui) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, 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, 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))。

(54) Title: METHOD AND DEVICE FOR PROCESSING AUDIO IN VIDEO COMMUNICATION

(54) 发明名称: 视频通信中的音频处理方法和装置

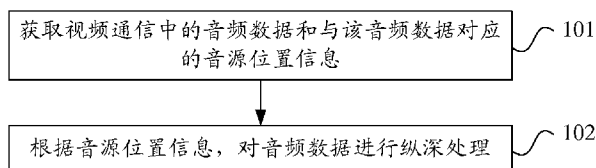


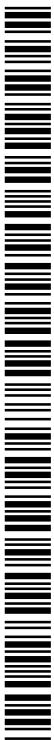
图 1 /Fig.1

101 Acquire audio data in video communication and audio source position information corresponding to the audio data

102 Perform longitudinal-depth processing on the audio data according to the audio source position information

(57) Abstract: Embodiments of the present invention provide a method and a device for processing audio in video communication. The method comprises: acquiring audio data in video communication and audio source position information corresponding to the audio data; and performing longitudinal-depth processing on the audio data according to the audio source position information. In the embodiments of the present invention, audio data in video communication and audio source position information corresponding to the audio data are acquired first, and then longitudinal-depth processing is performed on the audio data according to the acquired audio source position information, to enable the audio data to have a longitudinal depth matching the audio source position information, so that sounds generated by objects at different longitudinal positions in video communication can be distinguished.

[见续页]



WO 2012/068960 A1



(57) 摘要:

本发明实施例提供一种视频通信中的音频处理方法和装置，该方法包括：获取视频通信中的音频数据和与所述音频数据对应的音源位置信息；根据所述音源位置信息，对所述音频数据进行纵深处理。本发明实施例中，先获取视频通信中的音频数据和与该音频数据对应的音源位置信息，然后根据获取的该音源位置信息，对该音频数据进行纵深处理，将该音频数据处理成具有与该音源位置信息相匹配的纵深感，由此可以将视频通信中不同前后位置关系的对象发出的声音分辨出来。

视频通信中的音频处理方法和装置

本申请要求于 2010 年 11 月 26 日提交中国专利局、申请号为 201010561696.7、发明名称为“视频通信中的音频处理方法和装置”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

本发明实施例涉及通信技术领域，尤其涉及一种视频通信中的音频处理方法和装置。

背景技术

视讯会议业务是一种多媒体通信手段，利用电视设备和通信网络召开会议，可以同时实现两地或多地之间的图像、语音、数据的交互功能。视讯会议业务一般由视讯终端设备、传输网络、多点控制单元（Multipoint Control Units，以下简称为：MCU）等几部分组成。视讯终端设备主要包括视频输入/输出设备、音频输入/输出设备、视频编解码器、音频编解码器、信息通信设备及多路复用/信号分线设备等。视讯终端设备的基本功能是将本地摄像机拍摄的图像信号、麦克风拾取的声音信号进行压缩编码，然后发送给传输网络，以传至远方会场；同时，接收远方会场传来的数字信号，经解码后，还原成模拟的图像和声音信号。

视讯会议业务实现了远距离的音视频交流，随着技术的不断进步和发展，出现了可以使远程交流达到面对面的交流效果的网真系统。目前的网真系统利用视讯会议技术实现图像与声音的远程传输，配合整体的外设例如采用尺寸较大的液晶电视机实现“真人大小”，并通过一定的摄像头处理技术实现人们交流时的“眼对眼”，再配上整体的会议室装修方案，使得远程呈现出逼真的效果。

虽然现在的网真系统能实现比较逼真的效果。但是对于新出现的双排或多排网真，由于前后排具有一定的距离，而现有的网真系统仅能实现同一平面的图像与声音方位的对应，即前后排的声音都从同一个平面发出，如果不看图像，则无法分辨出声音来自前排还是后排，由此使得声音的临场感效果不真实。

发明内容

本发明实施例提供一种视频通信中的音频处理方法和装置，以分辨出多排视频通信中不同排的声音。

本发明实施例提供一种视频通信中的音频处理方法，包括：

获取视频通信中的音频数据和与所述音频数据对应的音源位置信息；

根据所述音源位置信息，对所述音频数据进行纵深处理。

本发明实施例还提供一种视频通信中的音频处理方法，包括：

获取视频通信的本端的音频数据和与所述音频数据对应的音源位置信息；

将所述音频数据和所述音源位置信息发送给视频通信控制单元或视频通信的对端，以使所述视频通信控制单元或视频通信的对端，根据所述音源位置信息对所述音频数据进行纵深处理。

本发明实施例提供一种视频通信中的音频处理装置，包括：

第一获取模块，用于获取视频通信中的音频数据和与所述音频数据对应的音源位置信息；

处理模块，用于根据所述第一获取模块获取的所述音源位置信息，对所述音频数据进行纵深处理。

本发明实施例还提供一种视频通信中的音频处理装置，包括：

第二获取模块，用于获取视频通信的本端的音频数据和与所述音频数据对应的音源位置信息；

第二发送模块，用于将所述第二获取模块获取的所述音频数据和所述音源位置信息发送给视频通信控制单元或视频通信的对端，以使所述视频通信控制单元或视频通信的对端，根据所述音源位置信息对所述音频数据进行纵深处理。

本发明实施例的视频通信中的音频处理方法和装置，先获取视频通信中的音频数据和与该音频数据对应的音源位置信息，然后根据获取的该音源位

置信息，对该音频数据进行纵深处理，将该音频数据处理成具有与该音源位置信息相匹配的纵深感，由此可以将视频通信中不同前后位置关系的对象发出的声音分辨出来。

附图说明

为了更清楚地说明本发明实施例的技术方案，下面将对实施例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发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1为本发明视频通信中的音频处理方法实施例一流程图；

图2为本发明视频通信中的音频处理方法实施例二流程图；

图3为本发明视频通信中的音频处理方法实施例三流程图；

图4为图3所示实施例的示意图；

图5为本发明视频通信中的音频处理方法实施例四流程图；

图6为本发明视频通信中的音频处理装置实施例一的示意图；

图7为本发明视频通信中的音频处理装置实施例二的示意图；

图8为本发明视频通信中的音频处理装置实施例三的示意图。

具体实施方式

为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明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发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，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，都属于本发明保护的范围。

图1为本发明视频通信中的音频处理方法实施例一的流程图，如图1所示，该方法包括：

步骤101、获取视频通信中的音频数据和与该音频数据对应的音源位置信息。

本方法各实施例可以应用在具有前后不同位置的多个声音源的视频通信中，例如：双排或多排的视频会议，或者3D视频会议。下面以多排的视频会议的场景为例进行说明，并且该视频会议可以包括两端会话或多端会话，其他场景可以参见该场景中的描述。

音频处理装置获取视频通信中的当前发言人的音频数据，以及与该音频数据对应的音源位置信息；该音源位置信息为该音频数据对应的对象与视频通信中第一排的相对位置信息，也就是说，该音源位置信息为当前发言人距离视频通信中第一排的位置信息，当该发言人位于第一排时，该音源位置信息即为0。

本实施例中的音频处理装置获取的音频数据可以是本端的音频数据，此时音频处理装置为音频数据采集端的设备；该音频处理装置获取的音频数据可以是对端的音频数据，此时音频处理装置为音频数据播放端的设备；此外，该音频处理装置还可以为视频通信的MCU端的设备，用于从音频数据的采集端获取音频数据，将该音频数据处理后发送至音频数据的播放端。

当该音频处理装置获取的音频数据是本端的音频数据时：该音频处理装置可以通过本端的拾音设备（例如麦克风）获取本端当前的音频数据和与该音频数据对应的音源位置信息。

当该音频处理装置获取的音频数据是对端的音频数据时：该音频处理装置通过接收对端（音频数据的采集端）发送的音频数据和音源位置信息，来获取对端的音频数据和音源位置信息，对端发送的音频数据和音源位置信息是由对端的音频处理装置通过拾音设备获取的。

当该音频处理装置为MCU端的设备时：该音频处理装置接收视频通信的一端发送的该端的音频数据和音源位置信息，从而获取到该端的音频数据和音源位置信息。

步骤102、根据音源位置信息，对音频数据进行纵深处理。

纵深感是人耳对距离和深度的感知。距离是人耳对某个特定声源距离远近的感知，深度用于描述对整个声音场景前后距离的感知。纵深感是前方发

出的声音具有远近的层次感。当声音具有纵深感时，人耳通过该声音即可判断出该声音的发生源的远近程度，即用户可以通过该声音分辨出发言者位于视频通信中的前后位置。

音频处理装置获取到音频数据和音源位置信息后，根据该音源位置信息，对该音频数据进行纵深处理，以使该音频数据具有与该音源位置信息相对应的纵深感，从而可以使得用户通过声音即可分辨出该音源数据对应的对象在视频中的前后位置。

音频处理装置对音频数据进行纵深处理之后，若该音频处理装置位于音频数据的采集端或者MCU端，则音频处理装置将纵深处理后的音频数据发送给播放端，以使对端播放该音频数据；若该音频处理装置位于音频数据的播放端，则音频处理装置直接播放该音频数据。

其中，对音频数据进行纵深处理的方式可以包括：（1）使用纵深感控制算法，对声音的响度、直达声与混响声的声能比、高频衰减量等参数进行控制，从而实现纵深处理；（2）使用波前合成（Wave Field Synthesis）技术对声音进行处理，使得处理后的声音具有纵深感。

需要说明的是，使用纵深感控制算法对音频数据进行处理时，该处理过程可以在音频数据的采集端进行，也可以在音频数据的播放端进行；而使用波前合成技术对音频数据进行处理时，该处理过程在音频数据的播放端进行。

本发明实施例，音频处理装置先获取视频通信中的音频数据和与该音频数据对应的音源位置信息，然后根据获取的该音源位置信息，对该音频数据进行纵深处理，将该音频数据处理成具有与该音源位置信息相匹配的纵深感，由此可以将视频通信中不同前后位置关系的对象发出的声音分辨出来。

图2为本发明视频通信中的音频处理方法实施例二的流程图，如图2所示，该方法包括：

步骤201、获取视频通信的本端的音频数据和与音频数据对应的音源位置信息。

本发明实施例应用的场景为：由音频数据的播放端或者视频通信的MCU端对音频数据进行纵深处理。本实施例的执行主体为音频数据的采集端。

视频通信本端的音频处理装置通过拾音设备获取本端当前的音频数据，并通过该拾音设备的标识获取与该音频数据对应的音源位置信息。其中，前后位置不同的与会者对应的拾音设备不同，由此可以根据拾音设备的不同，获取与该拾音设备拾取的音频数据对应的音源位置信息。

步骤202、将音频数据和音源位置信息发送给视频通信控制单元或视频通信的对端，以使视频通信控制单元或视频通信的对端，根据音源位置信息对音频数据进行纵深处理。

音频处理装置将获取的本端的音频数据和音源位置信息发送给视频通信控制单元，例如MCU，然后由MCU根据音源位置信息对音频数据进行纵深处理；或者，音频处理装置将获取的本端的音频数据和音源位置信息发送给视频通信的对端，然后由对端根据音源位置信息对音频数据进行纵深处理。

本发明实施例，音频处理装置获取视频通信中本端的音频数据和与该音频数据对应的音源位置信息，然后将获取到的音频数据和音源位置信息发送出去，以使MCU或视频通信的对端根据该音源位置信息对该音频数据进行纵深处理，得到具有与该音源位置信息相匹配的纵深感的音频数据，从而使得听者可以分辨出视频通信中不同前后位置关系的对象发出的声音。

由上述本发明各实施例的描述可知，本发明各实施例提供的方案中，可以在音频数据的采集端对音频数据进行纵深处理，也可以在MCU端对音频数据进行纵深处理，还可以在音频数据的播放端对音频数据进行纵深处理。下面根据对音频数据进行纵深处理的位置不同，分别对本发明实施例进行详细描述。

图3为本发明视频通信中的音频处理方法实施例三的流程，图4为图3所示实施例的示意图，本实施例为在音频数据的采集端对音频数据进行纵深处

理的情况，如图3所示，该方法包括：

步骤301、通过不同的拾音设备，获取视频通信中具有不同音源位置信息的第一音频数据和与第一音频数据对应的音源位置信息。

其中，第一音频数据为本实施例中视频通信的本端的音频数据。

在视频通信的本端（即音频数据的采集端）可以通过拾音设备拾取本端的音频数据，该拾音设备可以为麦克风，其中通过麦克风拾取音频数据可以有多种方式，只要保证能够对不同排的与会者的音频数据进行标识即可；例如：本端的每一个与会者使用给一个麦克风，每一个麦克风拾取的音频数据都对应音源位置信息，其中，相同排的与会者的音频数据对应的音源位置信息相同；或者每一排的与会者共用一个或几个麦克风，每一排的麦克风拾取的音频数据对应相同的音源位置信息。其中可以根据麦克风的指向性和灵敏度来实现每一排的与会者共用一个或几个麦克风。

在麦克风拾取本端的第一音频数据后，可以对该第一音频数据进行回声抵消、噪声抑制等前处理。音频处理装置获取经过前处理的第一音频数据，并根据该第一音频数据对应的麦克风，获取到该第一音频数据对应的音源位置信息；其中不同排的麦克风，对应不同的音源位置信息。

步骤302、根据音源位置信息，通过纵深感控制算法对第一音频数据进行纵深处理。

纵深感主要与声音的响度、直达声与混响声的声能比这两个因素有关，也与高频衰减量有关。

其中，响度就是音量的大小，响度与距离的平方成正比，距离增加1倍，响度衰减约6dB，由此距离远的声音衰减大，到达人耳的音量就会小。据此，根据前后排的距离，可以计算出声音的衰减量，使得播放前排与后排的声音时，音量会有差别。

要体现声音的纵深感，还必须调节直达声与混响声的声能比；其中可以通过控制前后排声音的延时和混响来调节直达声与混响声的声能比。通过加

延时混响的方法，可以虚拟出声音的空间感。当直达声较多时，听者会感觉声音是从近处传来的；当混响声较多时，会形成较明显的空间感，听者会感觉声音是从远处传来的。

此外，由于高频声波的波长比低频声波的波长短，在遇到房间的障碍物时高频声波的衰减比低频声波的衰减多；由此高频衰减量也是影响纵深感的一个因素。

纵深感控制算法可以为：首先获知房间大小及房间混响时间等房间环境的参数，然后根据该房间环境的参数计算出系统传输函数，再通过控制该系统传输函数的响度、直达声与混响声的声能、高频衰减量这三个因素，就可以控制一个声音的纵深感效果，使得该声音的效果与该声音的音源位置相匹配，例如听者可以分辨出该声音是从1m处传来的，还是从2m处传来的。

其中，该纵深感控制算法可以预先配置在音频处理装置中，使得音频处理装置在每次获取到音频数据及其音源位置信息后，即可根据音源位置信息和该算法对音频数据进行纵深处理，例如：音源位置信息为1m时，可以通过该算法调节响度、延时和混响的比例、高频的衰减量，从而产生具有1m的纵深感效果的音频数据；音源位置信息为0时，表示该音源位于最前排，由此不需要对该音源发出的声音进行纵深处理。

步骤303、将纵深处理后的第一音频数据发送给视频通信的对端，以使对端播放纵深处理后的第一音频数据。

音频处理装置将纵深处理后的第一音频数据发送给视频通信的对端，即该第一音频数据的播放端，使得播放端播放纵深处理后的该第一音频数据。

需要说明的是：当视频通信由MCU控制时，本实施例前述的音频处理装置对音频数据的纵深处理（步骤302）可以在该音频数据的采集端，也可以在MCU端；当视频通信没有MCU控制时，本实施例前述的音频处理装置对音频数据的纵深处理在该音频数据的采集端。当由MCU端对音频数据进行纵深处理时，采集端将步骤301中获取的音频数据和与该音频数据对应的音源位置信

息发送给MCU端，由MCU端对该音频数据进行纵深处理后，再将纵深处理后的音频数据发送给播放端。

然而，当麦克风采集的任两路或多路音频数据通过一路数据流发送给对端时，本实施例前述的对音频数据的纵深处理只能在音频数据的采集端进行处理，并且，在位于采集端的音频处理装置对音频数据进行纵深处理后，还包括混音切换步骤，即上述的步骤303可以被替换为以下步骤：

步骤303'、音频处理装置对纵深处理后的音频数据进行混音切换，然后将切换出的一路或两路数据发送给视频通信的对端。

例如：当系统采用单声道或双声道编码时，将多路经过纵深处理的音频数据混合，然后根据预置的策略切换出一路或两路数据信号，将切换出的数据信号经编码后发送给对端。对端接收并解码得到该数据信号后，直接播放即可得到具有纵深感的声音。

步骤304、对端对接收到的音频数据解码处理后进行播放。

对端先对接收到的音频数据进行解码，然后将解码后的音频数据通过扬声器输出。

下面结合图4对本实施例描述如下：如图4所示，第一端与会者共有两排（前排和后排），第一前排与会者1、第二前排与会者2、第三前排与会者3分别通过前排的第一麦克风M1、第二麦克风M2、第三麦克风M3拾音，前排麦克风的数据无纵深感处理，第一后排与会者4、第二后排与会者5、第三后排与会者6分别通过后排的第四麦克风M4、第五麦克风M5、第六麦克风M6拾音；前后排距离为1.5m，后排麦克风采集的数据统一加入1.5m的纵深感效果；由此，前排麦克风拾取的声音传到第二端进行播放时，第二端与会者感觉声音是从扬声器（例如音箱）发出的声音，而后排麦克风拾取的声音传到第二端进行播放时，第二端与会者感觉声音是从扬声器后面的1.5m处发出来的声音。其中，扬声器的摆放位于同一平面内，可以位于视频显示装置的上方、下方、左侧或右侧。其中，对于第一端而言，第一端为本端，第二端为对端；对于

第二端而言，第二端为本端，第一端为对端。

本发明实施例，音频处理装置先获取视频通信中的音频数据和与该音频数据对应的音源位置信息，然后根据获取的该音源位置信息，对该音频数据进行纵深处理，将该音频数据处理成具有与该音源位置信息相匹配的纵深感的音频数据，然后将处理后的音频数据发送给视频通信的对端进行播放，由此可以使得对端的听者通过声音分辨出发言者在视频通信中的前后位置。

图5为本发明视频通信中的音频处理方法实施例四的流程图，本实施例为在音频数据的播放端对音频数据进行纵深处理的情况，如图5所示，该方法包括：

步骤501、音频数据的采集端通过不同的拾音设备拾取视频通信中具有不同音源位置信息的第二音频数据和与第二音频数据对应的音源位置信息对音频数据。

其中，第二音频数据为本实施例中视频通信的对端的音频数据，本实施例中，音频数据的采集端为对端，音频数据的播放端为本端。

拾音设备对音频数据的拾取过程可以参见图3所示实施例中步骤301中的描述。

拾音设备拾取音频数据后，将拾取的音频数据发送给采集端的音频处理装置，同时采集端的音频处理装置通过拾音设备的标识获取该音频数据对应的音源位置信息。其中，前后位置不同的与会者对应的拾音设备不同，由此可以根据拾音设备的不同，获取与该拾音设备拾取的音频数据对应的音源位置信息。

步骤502、获取视频通信中的第二音频数据和与第二音频数据对应的音源位置信息。

采集端的音频处理装置将获取到的第二音频数据和音源位置信息编码后发送给播放端的音频处理装置。由此播放端的音频处理装置通过解码就获取到该第二音频数据和音源位置信息。

步骤503、播放端的音频处理装置根据音源位置信息，对第二音频数据进

行纵深处理。

具体的，该步骤可以包括步骤a或步骤b。

步骤a、播放端的音频处理装置根据音源位置信息，通过纵深感控制算法对第二音频数据进行纵深处理。

该步骤具体可以参见图3所示实施例中的步骤302。

步骤b、播放端的音频处理装置根据音源位置信息，对第二音频数据进行波前合成处理，以形成纵深处理后的第二音频数据。

波前合成（Wave Field Synthesis）利用Huyghen原理进行声波合成。波前合成技术为：波阵面（Wave Front）上所有的点均可以看作一个新的波源，这些新的波源和原始的波源有同样的速率和波长，在叠加之后这些新的波源在下一个瞬间形成新的波阵面。采用波前合成技术可以对声场进行真实的重现。

波前合成技术可以基于声波动理论并用多个摆放在一个平面上的扬声器形成的扬声器矩阵来产生波前，其中波前为离波源最远的波面；矩阵中的每一个扬声器都被馈给一个和其位置相应的、经瑞利重建积分算得的信号；每个扬声器根据该信号产生声波；每个扬声器产生的声波的叠加在交叠频率以下重建了准确的原声波波前。其中，交叠频率决定于各个扬声器间的距离。用波前技术得到的声场重建在整个聆听空间里保持了原声场的时域和空间性质。

播放端的音频处理装置根据音源位置信息，对第二音频数据进行波前合成处理的过程可以为：

播放端采用扬声器阵列进行放音，可以采用多个（例如6个）扬声器，放在视频显示装置的下方，具体扬声器个数根据算法及实际应用场景确定；音频处理装置根据该音频数据的声源位置信息，将该音频数据经过不同的计算，输出给多个扬声器，使多个扬声器同时放音，这些声音叠加在一起，发出的声音形成一个波阵面，该波阵面可以虚拟出原始音源的位置，由此恢复出具

有纵深感的声源。

步骤504、播放纵深处理后的第二音频数据。

播放端的音频处理装置通过纵深感控制算法对第二音频数据进行纵深处理后，将处理后的第二音频数据播放出来；或者，通过波前合成技术对第二音频数据进行纵深处理时，通过扬声器播放出来的，就是经过处理后的第二音频数据。

本发明实施例，音频处理装置先获取视频通信中对端的音频数据和与该音频数据对应的音源位置信息，然后根据获取的该音源位置信息，对该音频数据进行纵深处理，将该音频数据处理成具有与该音源位置信息相匹配的纵深感强的音频数据，然后将处理后的音频数据进行播放，由此可以使得对端的听者通过声音分辨出发言者在视频通信中的前后位置。

本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

图6为本发明视频通信中的音频处理装置实施例一的示意图，如图6所示，该音频处理装置包括：第一获取模块61和处理模块63。

第一获取模块61用于获取视频通信中的音频数据和与音频数据对应的音源位置信息。

处理模块63用于根据第一获取模块61获取的音源位置信息，对音频数据进行纵深处理。

本实施例中的音频处理装置可以为音频数据采集端的设备，也可以为音频数据播放端的设备，或者，还可以为视频通信的MCU端的设备。本实施例中各个模块的工作流程和工作原理参见上述方法实施例一中的描述，在此不再赘述。

本发明实施例，第一获取模块先获取视频通信中的音频数据和与该音频

数据对应的音源位置信息，然后处理模块根据获取的该音源位置信息，对该音频数据进行纵深处理，将该音频数据处理成具有与该音源位置信息相匹配的纵深感，由此可以将视频通信中不同前后位置关系的对象发出的声音分辨出来。

图7为本发明视频通信中的音频处理装置实施例二的示意图，在图6所示实施例的基础上，如图7所示，处理模块63具体可以包括：第一处理单元631和/或第二处理单元633；进一步，该音频处理装置还可以包括：第一发送模块65和播放模块67。

第一处理单元631用于当音频数据为视频通信的本端或发送端的第一音频数据时，根据第一获取模块获取的音源位置信息，通过纵深感控制算法对第一音频数据进行纵深处理。其中，当所述的音频处理装置位于MCU端时，根据数据流的方向，可以把视频通信的双方称为发送端和接收端。

第二处理单元633用于当音频数据为视频通信的对端的第二音频数据时，根据音源位置信息，通过纵深感控制算法对第二音频数据进行纵深处理，或者第二处理单元633用于当音频数据为视频通信的对端的第二音频数据时，根据音源位置信息，对第二音频数据进行波前合成处理，以形成纵深处理后的第二音频数据。

第一发送模块65用于将第一处理单元631纵深处理后的第一音频数据发送给视频通信的对端，以使对端播放纵深处理后的第一音频数据。

当所述的音频处理装置位于音频数据的播放端时，即在音频数据的播放端进行纵深处理时，该音频处理装置还包括：播放模块67，用于播放第二处理单元633纵深处理后的第二音频数据。

具体的：

当所述的音频处理装置位于音频数据的采集端时，即在音频数据的采集端进行纵深处理时，第一获取模块61具体可以用于：通过不同的拾音设备，获取视频通信中具有不同音源位置信息的音频数据和与音频数据对应的音源

位置信息。

当所述的音频处理装置位于MCU端时，即在MCU端进行纵深处理时，第一获取模块61具体可以用于：接收音频数据的采集端发送的音频数据和与音频数据对应的音源位置信息。

当所述的音频处理装置位于音频数据的播放端时，即在音频数据的播放端进行纵深处理时，第一获取模块61具体可以用于：接收音频数据的采集端发送的音频数据和与音频数据对应的音源位置信息。

本实施例中各个模块的工作流程和工作原理参见上述各方法实施例中的描述，在此不再赘述。

本发明实施例，音频处理装置先获取视频通信中的音频数据和与该音频数据对应的音源位置信息，然后根据获取的该音源位置信息，对该音频数据进行纵深处理，将该音频数据处理成具有与该音源位置信息相匹配的纵深感的音频数据，然后将处理后的音频数据发送给视频通信的对端进行播放，由此可以使得对端的听者通过声音分辨出发言者在视频通信中的前后位置。

图8为本发明视频通信中的音频处理装置实施例三的示意图，如图8所示，该音频处理装置包括：第二获取模块81和第二发送模块83。

本实施例提供的音频处理装置可以应用在以下场景：在视频通信中，由音频数据的播放端或者视频通信的MCU端对音频数据进行纵深处理，本实施例提供的音频处理装置设置在音频数据的采集端。

第二获取模块81用于获取视频通信的本端的音频数据和与音频数据对应的音源位置信息。

第二发送模块83用于将第二获取模块81获取的音频数据和音源位置信息发送给视频通信控制单元或视频通信的对端，以使视频通信控制单元或视频通信的对端，根据音源位置信息对音频数据进行纵深处理。

本实施例中各个模块的工作流程和工作原理参见上述方法实施例二中的描述，在此不再赘述。

本发明实施例，音频处理装置获取视频通信中本端的音频数据和与该音频数据对应的音源位置信息，然后将获取到的音频数据和音源位置信息发送出去，以使MCU或视频通信的对端根据该音源位置信息对该音频数据进行纵深处理，得到具有与该音源位置信息相匹配的纵深感的音频数据，从而使得听者可以分辨出视频通信中不同前后位置关系的对象发出的声音。

最后应说明的是：以上实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：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；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

权利要求书

1、一种视频通信中的音频处理方法，其特征在于，包括：

获取视频通信中的音频数据和与所述音频数据对应的音源位置信息；

根据所述音源位置信息，对所述音频数据进行纵深处理。

2、根据权利要求1所述的视频通信中的音频处理方法，其特征在于，所述音源位置信息为：所述音频数据对应的对象与所述视频通信中第一排的相对位置信息。

3、根据权利要求1或2所述的视频通信中的音频处理方法，其特征在于，所述音频数据为视频通信的本端或发送端的第一音频数据，所述根据所述音源位置信息，对所述音频数据进行纵深处理包括：

根据所述音源位置信息，通过纵深感控制算法对所述第一音频数据进行纵深处理。

4、根据权利要求3所述的视频通信中的音频处理方法，其特征在于，所述根据所述音源位置信息，对所述音频数据进行纵深处理之后还包括：

将纵深处理后的所述第一音频数据发送给视频通信的对端，以使所述对端播放纵深处理后的所述第一音频数据。

5、根据权利要求3所述的视频通信中的音频处理方法，其特征在于，所述获取视频通信中的音频数据和与所述音频数据对应的音源位置信息包括：

通过不同的拾音设备，获取视频通信中具有不同所述音源位置信息的第一音频数据和与所述第一音频数据对应的音源位置信息。

6、根据权利要求3所述的视频通信中的音频处理方法，其特征在于，所述纵深感控制算法包括：通过调节声音的响度、声音的直达声与混响声的声能比、声音的高频衰减量比，对音频数据进行纵深处理。

7、根据权利要求1或2所述的视频通信中的音频处理方法，其特征在于，所述音频数据为视频通信的对端的第二音频数据，所述根据所述音源位置信息，对所述音频数据进行纵深处理包括：

根据所述音源位置信息，通过纵深感控制算法对所述第二音频数据进行纵深处理；或者

根据所述音源位置信息，对所述第二音频数据进行波前合成处理，以形成纵深处理后的所述第二音频数据。

8、根据权利要求7所述的视频通信中的音频处理方法，其特征在于，所述根据所述音源位置信息，对所述音频数据进行纵深处理之后还包括：

播放纵深处理后的所述第二音频数据。

9、一种视频通信中的音频处理方法，其特征在于，包括：

获取视频通信的本端的音频数据和与所述音频数据对应的音源位置信息；

将所述音频数据和所述音源位置信息发送给视频通信控制单元或视频通信的对端，以使所述视频通信控制单元或视频通信的对端，根据所述音源位置信息对所述音频数据进行纵深处理。

10、一种视频通信中的音频处理装置，其特征在于，包括：

第一获取模块，用于获取视频通信中的音频数据和与所述音频数据对应的音源位置信息；

处理模块，用于根据所述第一获取模块获取的所述音源位置信息，对所述音频数据进行纵深处理。

11、根据权利要求10所述的视频通信中的音频处理装置，其特征在于，所述处理模块包括：

第一处理单元，用于当所述音频数据为视频通信的本端或发送端的第一音频数据时，根据所述第一获取模块获取的所述音源位置信息，通过纵深感控制算法对所述第一音频数据进行纵深处理。

12、根据权利要求10或11所述的视频通信中的音频处理装置，其特征在于，所述处理模块包括：

第二处理单元，用于当所述音频数据为视频通信的对端的第二音频数据时，根据所述音源位置信息，通过纵深感控制算法对所述第二音频数据进行纵深处

理，或者用于当所述音频数据为视频通信的对端的第二音频数据时，根据所述音源位置信息，对所述第二音频数据进行波前合成处理，以形成纵深处理后的所述第二音频数据。

13、根据权利要求11所述的视频通信中的音频处理装置，其特征在于，还包括：

第一发送模块，用于将所述处理模块纵深处理后的所述第一音频数据发送给视频通信的对端，以使所述对端播放纵深处理后的所述第一音频数据。

14、根据权利要求12所述的视频通信中的音频处理装置，其特征在于，还包括：

播放模块，用于播放所述处理模块纵深处理后的所述第二音频数据。

15、一种视频通信中的音频处理装置，其特征在于，包括：

第二获取模块，用于获取视频通信的本端的音频数据和与所述音频数据对应的音源位置信息；

第二发送模块，用于将所述第二获取模块获取的所述音频数据和所述音源位置信息发送给视频通信控制单元或视频通信的对端，以使所述视频通信控制单元或视频通信的对端，根据所述音源位置信息对所述音频数据进行纵深处理。

1/3

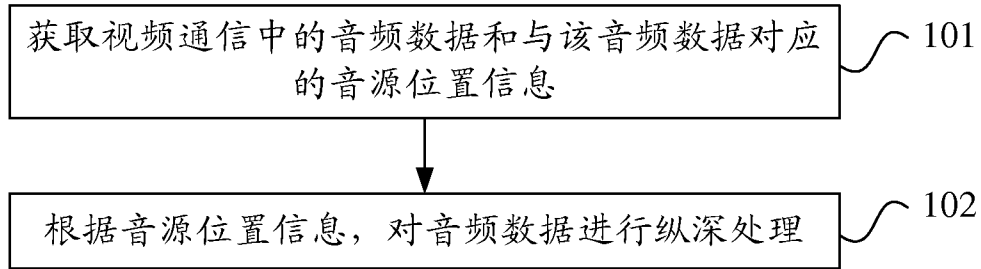


图 1

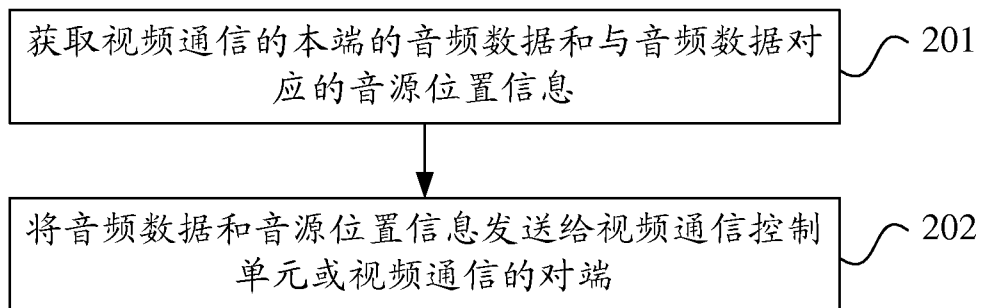


图 2

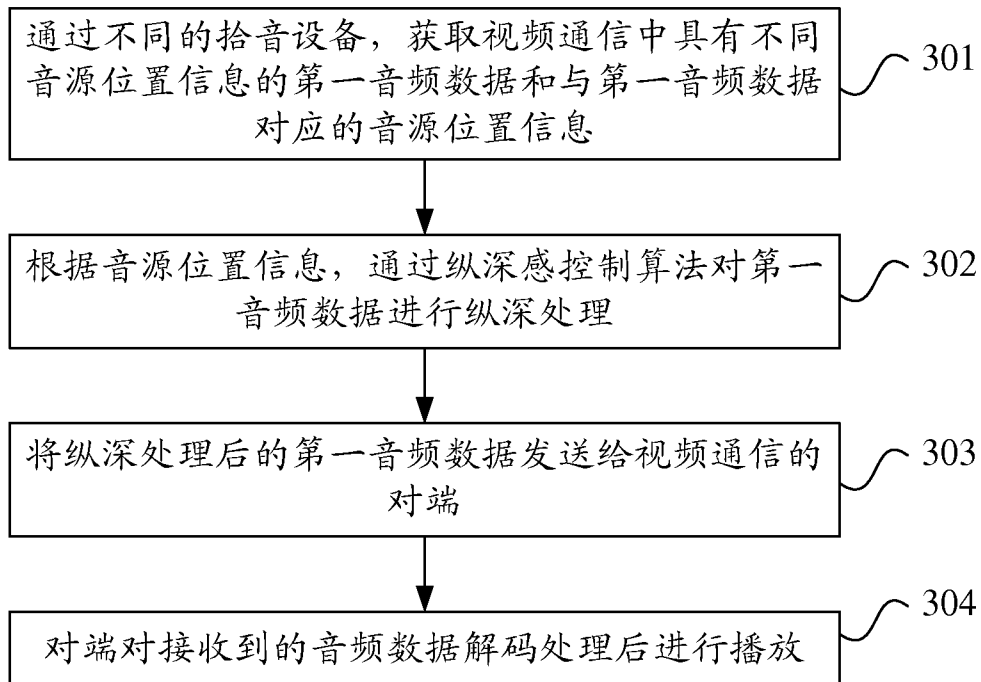


图 3

2/3

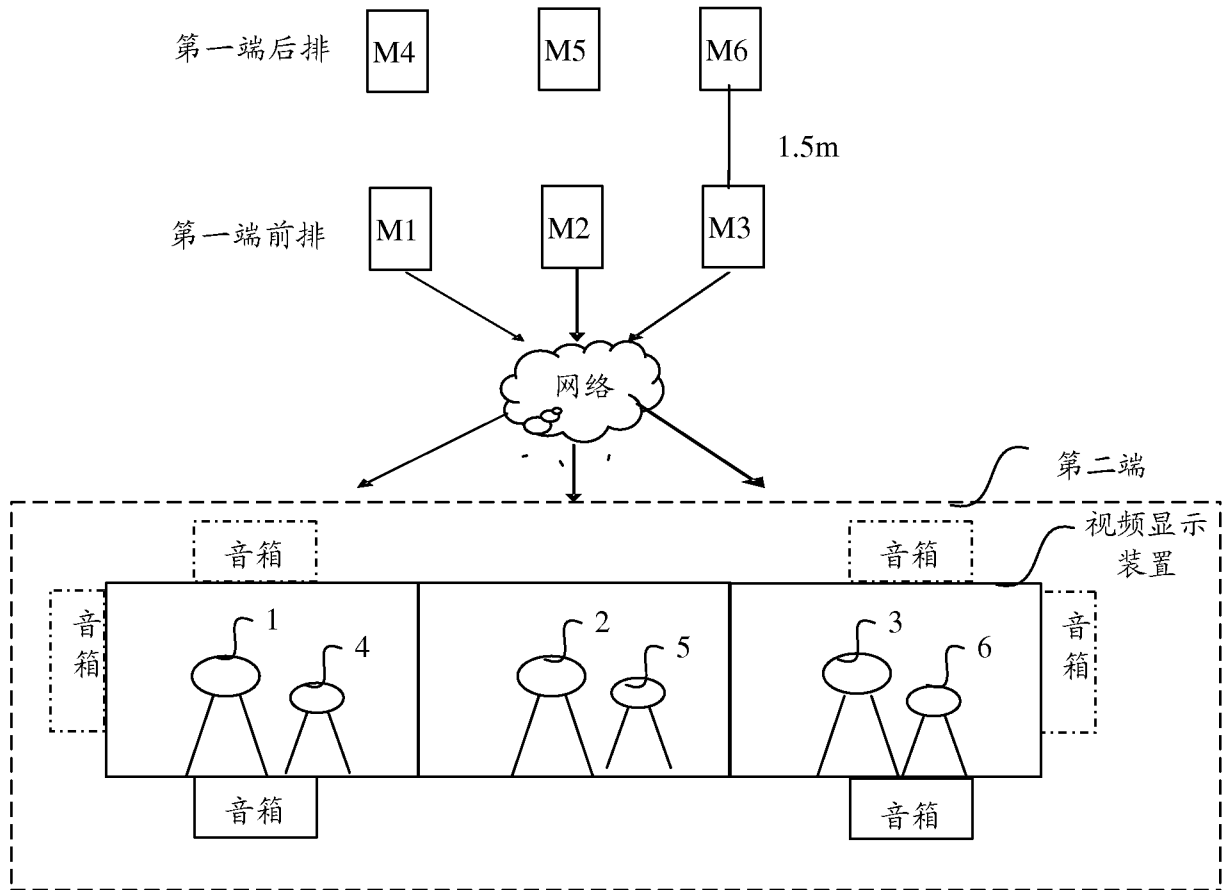


图 4

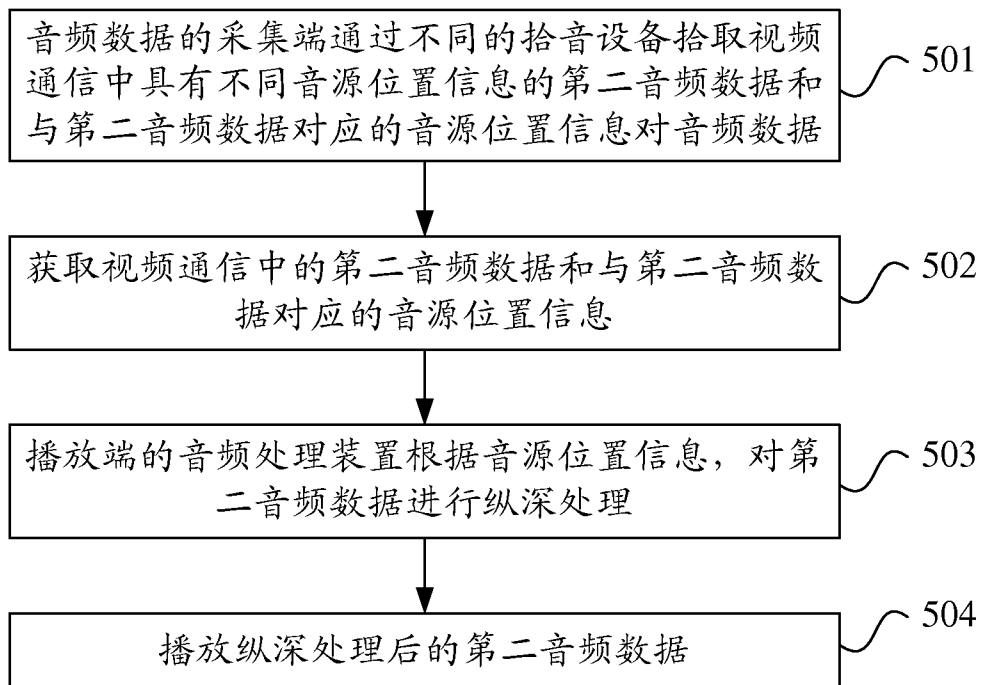


图 5

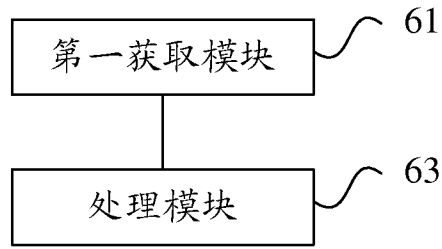


图 6

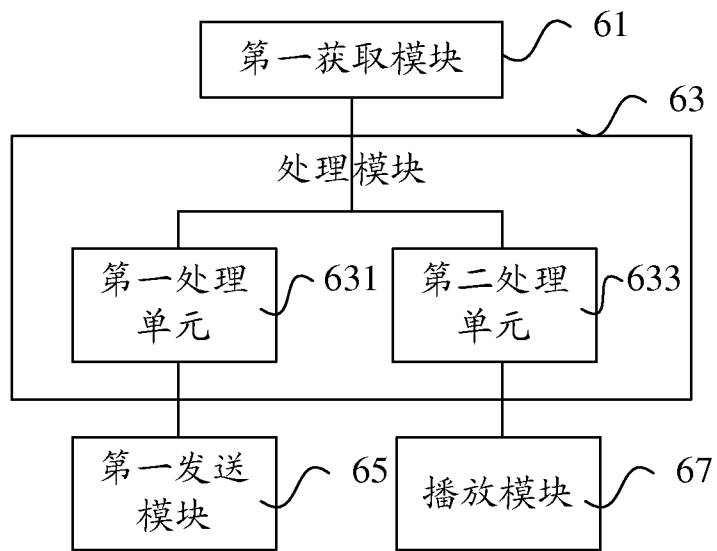


图 7

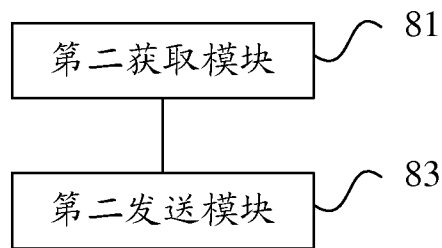


图 8

INTERNATIONAL SEARCH REPORT		International application No. PCT/CN2011/082127
A. CLASSIFICATION OF SUBJECT MATTER		
H04S3/00(2006.01)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)		
IPC: H04S; 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)		
VEN, CNABS, CNKI: audio, video conference, video communication, depth, location, position, distance, row, 3D		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN1929593A (POLYCOM INC) 14 Mar. 2007 (14.03.2007) abstract, claims 1-14, figure 2	1-15
A	CN101350931A (SHENZHEN HUAWEI COMM TECH CO) 21 Jan. 2009 (21.01.2009) the whole document	1-15
A	CN1717955A (THOMSON LICENSING SA) 04 Jan. 2006 (04.01.2006) the whole document	1-15
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“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</p> <p>“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</p> <p>“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</p> <p>“&” document member of the same patent family</p>	
Date of the actual completion of the international search 09 Feb. 2012 (09.02.2012)	Date of mailing of the international search report 23 Feb. 2012 (23.02.2012)	
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10)62019451	Authorized officer GE, Xiaolan Telephone No. (86-10) 62411470	

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2011/082127

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN1929593A	14.03.2007	US2007064094A1	22.03.2007
		IN200600570I2	22.06.2007
		JP2009177827A	06.08.2009
		EP1763241A3	05.03.2008
		US7612793B2	03.11.2009
		JP2007074732A	22.03.2007
		EP1763241A2	14.03.2007
CN101350931A	21.01.2009	EP2323425A1	18.05.2011
		US20110164769A1	07.07.2011
		WO2010022633A1	04.03.2010
CN1717955A	04.01.2006	JP4338647B2	07.10.2009
		WO2004051624A2	17.06.2004
		DE60311522D1	15.03.2007
		DE60311522T2	31.10.2007
		AU2003298146A1	23.06.2004
		US2006167695A1	27.07.2006
		EP1568251B1	24.01.2007
		AU2003298146B2	09.04.2009
		BR0316548A	04.10.2005
		KR20050084083A	26.08.2005
		ZA200503594A	30.08.2006
		VN10008405B	25.06.2010
		EP1568251A2	31.08.2005
		JP2006517356T	20.07.2006
		KR1004249B1	24.12.2010

国际检索报告

国际申请号
PCT/CN2011/082127

A. 主题的分类

H04S3/00(2006.01)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)

IPC: H04S; 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, CNKI: 音频, 声音, 位置, 座位, 纵深, 深度, 前后, 远近, 视频会议, 视频通信, 视频通讯, 三维, 3D, 网真;

VEN: audio, video conference, video communication, depth, location, position, distance, row, 3D

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	CN1929593A (宝利通公司) 14.3 月 2007 (14.03.2007) 摘要, 权利要求 1-17, 附图 2	1-15
A	CN101350931A (深圳华为通信技术有限公司) 21.1 月 2009 (21.01.2009) 全文	1-15
A	CN1717955A (汤姆森许可贸易公司) 04.1 月 2006 (04.01.2006) 全文	1-15

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
09.2 月 2012 (09.02.2012)

国际检索报告邮寄日期
23.2 月 2012 (23.02.2012)

ISA/CN 的名称和邮寄地址:
中华人民共和国国家知识产权局
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
葛晓兰
电话号码: (86-10) 62411470

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2011/082127

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN1929593A	14.03.2007	US2007064094A1	22.03.2007
		IN200600570I2	22.06.2007
		JP2009177827A	06.08.2009
		EP1763241A3	05.03.2008
		US7612793B2	03.11.2009
		JP2007074732A	22.03.2007
		EP1763241A2	14.03.2007
CN101350931A	21.01.2009	EP2323425A1	18.05.2011
		US20110164769A1	07.07.2011
		WO2010022633A1	04.03.2010
CN1717955A	04.01.2006	JP4338647B2	07.10.2009
		WO2004051624A2	17.06.2004
		DE60311522D1	15.03.2007
		DE60311522T2	31.10.2007
		AU2003298146A1	23.06.2004
		US2006167695A1	27.07.2006
		EP1568251B1	24.01.2007
		AU2003298146B2	09.04.2009
		BR0316548A	04.10.2005
		KR20050084083A	26.08.2005
		ZA200503594A	30.08.2006
		VN10008405B	25.06.2010
		EP1568251A2	31.08.2005
		JP2006517356T	20.07.2006
KR1004249B1	24.12.2010		