

(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 11 月 19 日 (19.11.2020)



(10) 国际公布号
WO 2020/228404 A1

(51) 国际专利分类号:
H04L 12/58 (2006.01)

(21) 国际申请号: PCT/CN2020/079072

(22) 国际申请日: 2020 年 3 月 12 日 (12.03.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910400023.4 2019年5月14日 (14.05.2019) CN

(71) 申请人: 北京达佳互联信息技术有限公司(BEIJING DAJIA INTERNET INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京市海淀区上地西路6号1幢1层101D1-7, Beijing 100085 (CN)。

(72) 发明人: 张晨(ZHANG, Chen); 中国北京市海淀区上地西路6号1幢1层101D1-7, Beijing 100085 (CN)。 郭亮(GUO, Liang); 中国北京市海淀区上

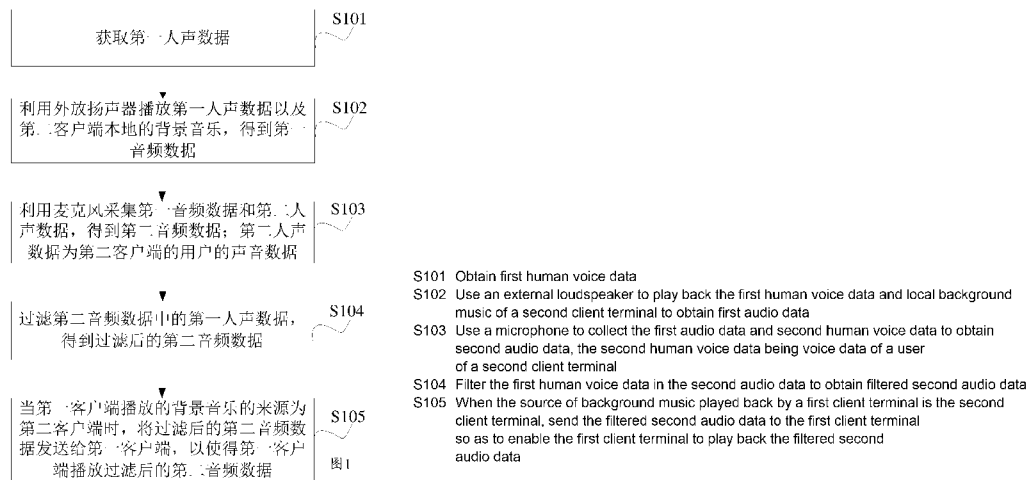
地西路6号1幢1层101D1-7, Beijing 100085 (CN)。
董培(DONG, Pei); 中国北京市海淀区上地西路6号1幢1层101D1-7, Beijing 100085 (CN)。

(74) 代理人: 北京同达信恒知识产权代理有限公司(TDIP & PARTNERS); 中国北京市西城区裕民路18号北环中心A座2002, Beijing 100029 (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, 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。

(54) Title: INSTANT MESSAGING SOUND QUALITY OPTIMIZATION METHOD, APPARATUS AND DEVICE

(54) 发明名称: 即时通讯的音质优化方法、装置及设备



(57) Abstract: The present application relates to an instant messaging sound quality optimization method and apparatus and a device, the method comprising: obtaining first human voice data, the first human voice data being voice data of a user of a first client terminal; using an external loudspeaker to play back the first human voice data and local background music of a second client terminal to obtain first audio data; using a microphone to collect the first audio data and second human voice data to obtain second audio data, the second human voice data being voice data of a user of a second client terminal; filtering the first human voice data in the second audio data to obtain filtered second audio data; when the source of background music played back by the first client terminal is the second client terminal, sending the filtered second audio data to the first client terminal so as to enable the first client terminal to play back the filtered second audio data. By means of the present solution, both echo cancellation and the reduction of non-echo human voice loss may be taken into account in an instant messaging system in which background music is present.

(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))。

(57) 摘要: 本申请关于即时通讯的音质优化方法、装置及设备, 该方法包括: 获取第一人声数据; 第一人声数据为第一客户端的用户的声数据; 利用外放扬声器播放第一人声数据以及第二客户端本地的背景音乐, 得到第一音频数据; 利用麦克风采集第一音频数据和第二人声数据, 得到第二音频数据; 第二人声数据为第二客户端的用户的声数据; 过滤第二音频数据中的第一人声数据, 得到过滤后的第二音频数据; 当第一客户端播放的背景音乐的来源为第二客户端时, 将过滤后的第二音频数据发送给第一客户端, 以使得第一客户端播放过滤后的第二音频数据。通过本方案能够在存在背景音乐的即时通讯系统中, 兼顾回声消除和非回声的人声损耗的减少。

即时通讯的音质优化方法、装置及设备

本申请要求在2019年05月14日提交中国专利局、申请号为201910400023.4、申请名称为“即时通讯的音质优化方法、装置及设备”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。

5

技术领域

本申请涉及即时通讯技术领域，尤其涉及一种即时通讯的音质优化方法、装置及设备。

背景技术

10 即时通讯应用可以支持通讯双方或者多方进行实时的语音交流。在实时的语音交流中，当某一端用户对播放效果要求较高，或者所使用的即时通讯设备无法使用耳机时，该端用户，即近端用户可以使用外放扬声器播放另一端用户，即远端用户的语音。此时，近端用户的麦克在采集近端用户的语音时，外放扬声器播放的远端用户的语音会漏进麦克，即，远端用户的语音与近端用户的语音会一起被近端用户的麦克采集，这将导致远端用户接收到的近端用户的语音中含有近端用户的麦克所采集的远端用户的语音，近端用户的语音中
15 出现远端用户的回声。对此，相关技术会对近端用户的麦克采集的音频数据进行回声消除，即对近端用户的麦克采集的音频数据中的回声进行过滤，得到近端用户的语音，作为目标音频数据发送给远端用户。

但发明人发现，在一些除了通讯者的语音，还存在 BGM（Background Music，背景
20 音乐）的场景中，例如，连麦 K 歌以及连麦短剧表演等等场景中，由于背景音乐在通讯过程中始终存在，并且，背景音乐由外放扬声器播放后，会被采集到近端用户发送给远端用户的音频数据中。此时，为了得到目标音频数据，在按照上述回声消除的方式过滤近端用户发送的音频数据时，需要持续地进行过滤，但持续过滤容易造成过滤过度，导致对无需过滤的非回声的语音即近端用户的人声进行一定程度的过滤，近端用户的人声出现卡顿以及人声忽大忽小等音质损耗的问题。
25

因此，如何在存在背景音乐的即时通讯系统中，兼顾回声消除和减少非回声的人声的音质损耗，是即时通讯技术中亟待解决的问题。

发明内容

30 为克服相关技术中存在的问题，本申请提供一种即时通讯的音质优化方法、装置及设

备。

根据本申请实施例的第一方面，提供一种即时通讯的音质优化方法，应用于第二客户端，所述方法包括：

获取第一人声数据；所述第一人声数据为第一客户端的用户的声音数据；

5 利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；

利用麦克风采集所述第一音频数据和第二人声数据，得到第二音频数据；所述第二人声数据为所述第二客户端的用户的声音数据；

过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

10 当所述第一客户端播放的背景音乐的来源为所述第二客户端时，将所述过滤后的第二音频数据发送给所述第一客户端，以使得所述第一客户端播放所述过滤后的第二音频数据。

根据本申请实施例的第二方面，提供另一种即时通讯的音质优化方法，应用于第一客户端，所述方法包括：

15 发送第一人声数据给第二客户端，以使得所述第二客户端利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；或者，发送第三音频数据给所述第二客户端，以使得所述第二客户端过滤所述第三音频数据中的背景音乐，得到第一人声数据，并利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；其中，所述第一人声数据为所述第一客户端的用户的声音数据；所述第三音频数据为所述第一客户端利用麦克风采集所述第一人声数据和所述第一客户端本地的背景音乐得到的音频数据；

20 接收所述第二客户端发送的第二音频数据；所述第二音频数据为所述第二客户端利用麦克风采集所述第一音频数据和第二人声数据，得到的音频数据；所述第二人声数据为所述第二客户端的用户的声音数据；

过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

25 当所述第一客户端播放的背景音乐的来源为所述第二客户端时，播放所述过滤后的第二音频数据。

根据本申请实施例的第三方面，提供一种即时通讯的音质优化装置，应用于第二客户端，所述装置包括：

30 第一人声获取模块，被配置为获取第一人声数据；所述第一人声数据为第一客户端的用户的声音数据；

第一音频获取模块，被配置为利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；

第二音频获取模块，被配置为利用麦克风采集所述第一音频数据和第二人声数据，得

到第二音频数据；所述第二人声数据为所述第二客户端的用户的声音数据；

过滤模块，被配置为过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

5 发送模块，被配置为在所述第一客户端播放的背景音乐的来源为所述第二客户端时，将所述过滤后的第二音频数据发送给所述第一客户端，以使得所述第一客户端播放所述过滤后的第二音频数据。

根据本申请实施例的第四方面，提供另一种即时通讯的音质优化装置，应用于第一客户端，所述装置包括：

10 发送模块，被配置为发送第一人声数据给第二客户端，以使得所述第二客户端利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；或者，发送第三音频数据给所述第二客户端，以使得所述第二客户端过滤所述第三音频数据中的背景音乐，得到第一人声数据，并利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；其中，所述第一人声数据为所述第一客户端的用户的声音数据；所述第三音频数据为所述第一客户端利用麦克风采集所述第一人声数据和所述第一客户端本地的背景音乐得到的音频数据；

15

接收模块，被配置为接收所述第二客户端发送的第二音频数据；所述第二音频数据为所述第二客户端利用麦克风采集所述第一音频数据和第二人声数据，得到的音频数据；所述第二人声数据为所述第二客户端的用户的声音数据；

20 过滤模块，被配置为过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

播放模块，被配置为在所述第一客户端播放的背景音乐的来源为所述第二客户端时，播放所述过滤后的第二音频数据。

根据本申请实施例的第五方面，提供一种电子设备，应用于第二客户端，该电子设备包括：

25 处理器；

用于存储处理器可执行指令的存储器；

其中，所述处理器被配置为执行：

获取第一人声数据；所述第一人声数据为第一客户端的用户的声音数据；

30 利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；

利用麦克风采集所述第一音频数据和第二人声数据，得到第二音频数据；所述第二人声数据为所述第二客户端的用户的声音数据；

过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

当所述第一客户端播放的背景音乐的来源为所述第二客户端时，将所述过滤后的第二音频数据发送给所述第一客户端，以使得所述第一客户端播放所述过滤后的第二音频数据。

根据本申请实施例的第六方面，提供一种电子设备，应用于第一客户端，所述电子设备包括：

- 5 处理器；
 用于存储所述处理器可执行指令的存储器；
 其中，所述处理器被配置为执行：

10 发送第一人声数据给第二客户端，以使得所述第二客户端利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；或者，发送第三音频数据给所述第二客户端，以使得所述第二客户端过滤所述第三音频数据中的背景音乐，得到第一人声数据，并利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；其中，所述第一人声数据为所述第一客户端的用户的语音数据；所述第三音频数据为所述第一客户端利用麦克风采集所述第一人声数据和所述第一客户端本地的背景音乐得到的音频数据；

15 接收所述第二客户端发送的第二音频数据；所述第二音频数据为所述第二客户端利用麦克风采集所述第一音频数据和第二人声数据，得到的音频数据；所述第二人声数据为所述第二客户端的用户的语音数据；

 过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

20 当所述第一客户端播放的背景音乐的来源为所述第二客户端时，播放所述过滤后的第二音频数据。

 根据本申请实施例的第七方面，提供一种非临时性计算机可读存储介质，包含于电子设备，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行上述第一方面，或者，第二方面所述的即时通讯的音质优化方法的步骤。

25 根据本申请实施例的第八方面，提供一种计算机程序产品，当其在电子设备上运行时，使得电子设备执行上述第一方面，或者，第二方面所述的即时通讯的音质优化方法的步骤。

30 本申请实施例中，在存在背景音乐的即时通讯系统中，由于第一人声数据与背景音乐相比持续的时间较短，因此，过滤第二音频数据中的第一人声数据与传统的对第二音频数据进行持续性回声过滤相比，可以减少对第二音频数据的过度过滤，从而减少对第二音频数据中第二人声数据的过度过滤，减少第二人声的卡顿和忽大忽小等问题，减少对非回声的第二人声的音质的损耗。并且，当第一客户端播放的背景音乐的来源为第二客户端时，过滤后的第二音频数据中的背景音乐可以作为第一客户端播放的背景音乐。因此，当第一客户端播放的背景音乐的来源为第二客户端时，将过滤后的第二音频数据发送给第一客户端进行播放，可避免过滤后的第二音频数据中的背景音乐成为第一客户端的噪声，保证回

声消除的效果。可见，本方案能够在存在背景音乐的即时通讯系统中兼顾回声消除和非回声的人声损耗的减少。

附图说明

- 5 图1是根据一示例性实施例示出的一种即时通讯的音质优化方法的流程图；
图2是根据另一示例性实施例示出的一种即时通讯的音质优化方法的流程图；
图3是根据又一示例性实施例示出的一种即时通讯的音质优化方法的流程图；
图4是根据一示例性实施例示出的一种即时通讯的音质优化装置的框图；
图5是根据另一示例性实施例示出的一种即时通讯的音质优化装置的框图；
10 图6是根据一示例性实施例示出的一种电子设备的框图；
图7是根据另一示例性实施例示出的一种电子设备的框图；
图8是根据又一示例性实施例示出的一种电子设备的框图；
图9是根据再一示例性实施例示出的一种电子设备的框图。

15 具体实施方式

本申请实施例提供的即时通讯的音质优化方法的执行主体可以为即时通讯系统中进行音质优化的电子设备。示例性的，该电子设备可以为进行即时通讯的至少两个客户端中的任一个。举例而言，客户端具体可以是计算机、智能移动终端以及可穿戴式智能终端等等。或者，示例性的，该电子设备可以为即时通讯应用对应的服务器，即与客户端对应的服务器。举例而言，服务器具体可以是台式计算机、云服务器以及笔记本电脑等等。

20 图1是根据一示例性实施例示出的一种即时通讯的音质优化方法的流程图，如图1所示，一种即时通讯的音质优化方法，应用于第二客户端，该方法可以包括以下步骤：

步骤S101，获取第一人声数据；第一人声数据为第一客户端的用户的语音数据。

其中，第一客户端与第二客户端之间进行存在背景音乐的即时通讯，例如，连麦K歌和连麦短剧表演等等。并且，即时通讯系统可以是多种的。示例性的，即时通讯系统可以是直播系统、社交系统、K歌系统等等。

为了便于理解，在后续实施例中，以连麦K歌的应用场景进行示例性说明。在连麦K歌的应用场景中，可以将主播客户端看作第一客户端，与主播进行连麦K歌的连麦歌手客户端看作第二客户端。相应的，主播的人声数据为第一人声数据，连麦歌手的人声数据为第二人声数据。

在存在背景音乐的即时通讯系统中，当第一客户端以不同方式播放背景音乐时，对第

一人声数据的处理也会不同，相应地，第二客户端获取第一人声数据的方式可以是多种的，下面以可选实施例的方式进行具体说明。

在一种可选的实施例中，第二客户端获取第一人声数据的方式可以包括：

当第一客户端利用耳机播放背景音乐时，接收第一客户端发送的第一人声数据。

5 当第一客户端利用耳机播放背景音乐时，第一客户端的麦克风采集到的第一人声数据中不会混入第一客户端播放的第一客户端本地的背景音乐，因此，第一客户端可以将第一人声数据直接发送给第二客户端，第二客户端接收第一客户端发送的第一人声数据，就可以实现对第一人声数据的获取。

10 另外，在本可选实施例中，第一客户端播放的背景音乐的来源可以是多种的。示例性的，第一客户端播放的背景音乐可以是第二客户端发送给第一客户端的，可以是第一客户端本地存储的，或者，可以是第一客户端从即时通讯系统的服务器中下载的。

在另一种可选的实施例中，第二客户端获取第一人声数据的方式可以包括：

15 当第一客户端利用外放扬声器播放背景音乐时，接收第一客户端发送的第一客户端对第三音频数据中的背景音乐过滤得到的第一人声数据；第三音频数据为第一客户端利用麦克风采集第一人声数据和第一客户端播放的第一客户端本地的背景音乐得到的音频数据。

20 如果第一客户端利用外放扬声器播放背景音乐，第一客户端的麦克风采集第一人声数据时，第一客户端播放的第一客户端本地的背景音乐也会被采集，此时，第一客户端的麦克风采集到的是第三音频数据。因此，第一客户端需要对第一音频数据中的背景音乐进行过滤，以得到第一人声数据，并将该第一人声数据发送给第二客户端。第二客户端接收第一客户端发送的第一人声数据，就可以实现对第一人声数据的获取。

在又一种可选的实施例中，第二客户端获取第一人声数据的方式可以包括：

当第一客户端利用外放扬声器播放背景音乐时，接收第一客户端发送的第三音频数据；过滤第三音频数据中的背景音乐，得到第一人声数据。

25 本可选实施例为与上述另一种可选的实施例相似的实施例，区别在于本可选实施例中对第三音频数据中的背景音乐进行过滤的执行主体为第二客户端。第二客户端在接收到第一客户端发送的第三音频数据后，过滤第三音频数据中的背景音乐，就可以得到第一人声数据。

任何在存在背景音乐的即时通讯系统中获取第一人声数据的方式均可用于本申请实施例，在此不作限制。

30 步骤S102，利用外放扬声器播放第一人声数据以及第二客户端本地的背景音乐，得到第一音频数据。

当第二客户端利用外放扬声器播放第一人声数据以及第二客户端本地的背景音乐时，播放的第一人声数据和第二客户端本地的背景音乐会混合在一起，成为第一音频数据。并

且，使用外放扬声器播放会导致后续在步骤S103中利用麦克风采集第二人声数据时，将第一音频数据一并采集，造成第二人声数据中混入了第一音频数据，成为第二音频数据。

其中，第二客户端本地的背景音乐的来源可以是多种的。示例性的，第二客户端播放的背景音乐可以是第二客户端本地存储的，或者，可以是第一客户端从即时通讯系统的服务器中下载的。另外，外放扬声器可以是多种的。示例性的，外放扬声器可以是第二客户端中的扬声器，也可以是与第二客户端连接的音箱等等。

步骤S103，利用麦克风采集第一音频数据和第二人声数据，得到第二音频数据；第二人声数据为第二客户端的用户的语音数据。

步骤S104，过滤第二音频数据中的第一人声数据，得到过滤后的第二音频数据。

步骤S105，当第一客户端播放的背景音乐的来源为第二客户端时，将过滤后的第二音频数据发送给第一客户端，以使得第一客户端播放过滤后的第二音频数据。

在具体应用中，可以利用自适应滤波器过滤第二音频数据中的第一人声数据，得到过滤后的第二音频数据，为了便于理解和合理布局，后续以可选实施例的方式进行具体说明。

在上述步骤S104中，过滤后的第二音频数据是过滤第二音频数据中的第一人声数据得到的，也就是说，过滤后的第二音频数据为包含第二客户端的麦克风采集的第二人声数据和第二客户端的麦克风采集的背景音乐的音频数据。在存在背景音乐的即时通讯系统中，对第一客户端而言，非回声的音频数据为第二人声数据，如果直接将过滤后的第二音频数据作为第一客户端播放的音频数据，过滤后的第二音频数据所包含的背景音乐可能会成为回声。

对此，如果第一客户端播放的背景音乐的来源为第二客户端，过滤后的第二音频数据中包含的背景音乐可以作为第一客户端播放的背景音乐，能够避免过滤后的第二音频数据中的背景音乐成为第一客户端的噪声，保证回声消除的效果。因此，在步骤S105中，可以将过滤后的第二音频数据发送给第一客户端，以使得第一客户端播放过滤后的第二音频数据，实现第一客户端和第二客户端的即时通讯。

另外，对于第一客户端播放的背景音乐的来源不为第二客户端的情况，为了便于理解和合理布局，后续在图3中进行具体说明。

示例性的，主播和连麦歌手进行歌曲S1的连麦K歌，歌曲S1的音乐伴奏BGM1为通讯双方的客户端播放的背景音乐。连麦歌手客户端获取主播的人声数据后，利用外放扬声器播放主播的人声数据和连麦歌手客户端本地的BGM1，得到主播的人声数据和连麦歌手客户端本地的BGM1混合后的第一音频数据。连麦歌手客户端利用麦克风采集连麦歌手歌唱时产生的连麦歌手的人声数据，以及第一音频数据，得到连麦歌手的人声数据和第一音频数据混合后的第二音频数据。过滤第二音频数据中主播的人声数据，得到过滤后的第二音频数据。过滤后的第二音频数据中不再包含主播的人声数据，而是连麦歌手的人声数据和

BGM1。当主播客户端播放的BGM1的来源为连麦歌手客户端时，将过滤后的第二音频数据发送给主播客户端，以使得主播客户端播放过滤后的第二音频数据。此时，主播客户端播放的音频数据为连麦歌手的人声数据和BGM1，不存在回声，因此，实现了回声消除的效果。并且，与BGM1相比，主播的人声数据在连麦K歌过程中的持续时间相对而言较短，因此，过滤第二音频数据中主播的人声数据与传统的对第二音频数据持续进行回声过滤相比，可以减少对第二音频数据的过度过滤，从而减少对第二音频数据中连麦歌手的人声数据的过度过滤，减少连麦歌手的人声的卡顿和忽大忽小等问题，减少对非回声的连麦歌手的人声的音质的损耗。

本申请实施例中，在存在背景音乐的即时通讯系统中，由于第一人声数据与背景音乐相比持续的时间较短，因此，过滤第二音频数据中的第一人声数据与传统的对第二音频数据持续进行回声过滤相比，可以减少对第二音频数据的过度过滤，从而减少对第二音频数据中第二人声数据的过度过滤，减少第二人声的卡顿和忽大忽小等问题，减少对非回声的第二人声的音质的损耗。并且，当第一客户端播放的背景音乐的来源为第二客户端时，过滤后的第二音频数据中的背景音乐可以作为第一客户端播放的背景音乐。因此，当第一客户端播放的背景音乐的来源为第二客户端时，将过滤后的第二音频数据发送给第一客户端进行播放，可以避免过滤后的第二音频数据中的背景音乐成为第一客户端的噪声，保证回声消除的效果。可见，本方案能够在存在背景音乐的即时通讯系统中，兼顾回声消除和非回声的人声损耗的减少。

可选的，上述步骤S104：过滤第二音频数据中的第一人声数据，得到过滤后的第二音频数据，具体可以包括如下步骤：

将第二音频数据和所获取的第一人声数据分别输入自适应滤波器，以使得自适应滤波器按照第一人声数据，模拟第二音频数据中的第一人声数据，得到模拟的第一人声数据，并利用模拟的第一人声数据抵消第二音频数据中的第一人声数据，进而将完成抵消的第二音频数据，作为过滤后的第二音频数据。

在具体应用中，自适应滤波器可以是多种的。不同的自适应滤波器所采用的用于确定自适应滤波器的实际输出是否达到预设的期望输出，即是否收敛的算法不同。举例而言，LMS（Least mean square，最小均方）自适应滤波器采用最小均方算法确定输出是否收敛，RLS（Recursive Least Squares，递推最小二乘）滤波器采用递推最小二乘确定是否收敛。任何自适应滤波器均可用于本申请实施例，在此不作限制。

通过将第二音频数据和所获取的第一人声数据分别输入自适应滤波器，自适应滤波器可以将第一人声数据作为参考信号，模拟第二音频数据所包含的第一人声数据，将第二音频数据和模拟的第一人声数据相减，实现对第二音频数据所包含的第一人声数据的抵消。当然，为了保证过滤后的输出达到期望的输出，在过滤时，自适应滤波器可以判断过滤后

的第二音频数据是否收敛,如果收敛,确定完成对第二音频数据中的第一人声数据的抵消;如果不收敛,可以将过滤后的第二音频数据作为反馈信号,按照反馈信号对自适应滤波器自身的参数进行调整,完成调整后继续进行第一人声数据的抵消,不断循环,直到过滤后的第二音频数据收敛。

5 另外,可以在自适应滤波器后增加残余回声滤波器,提高回声消除的效果。其中,示例性的,残余回声滤波器具体可以是NLP滤波器(与自适应滤波器相似,区别在于将待过滤的信号分成多个子带,针对每个子带进行滤波)。

10 上述可选实施例对第二音频数据中的第一人声数据进行过滤,与对背景音乐和第一人声数据均进行过滤相比,可以减少过滤过程需要处理的数据量,相对而言,可以降低过滤所耗费的时间,提高音质优化的效率。

15 此外,在具体应用中,第一客户端的数量可以为多个,此时,即时通讯的音质优化与图1所示的实施例和可选实施例相似。区别在于,当第一客户端的数量为多个时,第二客户端通过外放扬声器播放的第一人声数据的数量为多个,此时,第二客户端的麦克风采集的第二音频数据中包含多个第一人声数据。对此,需要获取多个第一人声数据,并将多个第一人声数据混合为一个参考信号。将第二音频数据和参考信号分别输入自适应滤波器,以使得自适应滤波器按照参考信号,模拟第二音频数据中作为回声数据的多个第一人声数据,得到模拟的回声数据,并利用模拟的回声数据抵消第二音频数据中的回声数据,完成抵消的第二音频数据即为过滤后的第二音频数据。

20 可选的,在上述利用麦克风采集第一音频数据和第二人声数据,得到第二音频数据得到步骤之后,将第二音频数据和所获取的第一人声数据分别输入自适应滤波器的步骤之前,本申请实施例提供的即时通讯的音质优化方法,还可以包括如下步骤:

对所获取的第一人声数据和第二音频数据进行相关性对比,得到第一人声数据和第二音频数据之间的第一延时;

25 相应的,将第二音频数据和所获取的第一人声数据分别输入自适应滤波器,以使得自适应滤波器按照所输入的第一人声数据,模拟第二音频数据中的第一人声数据,得到模拟的第一人声数据,并利用模拟的第一人声数据抵消第二音频数据中的第一人声数据的,包括:

30 将第二音频数据、所获取的第一人声数据和第一延时分别输入自适应滤波器,以使得自适应滤波器按照第一延时,对第一人声数据和第二音频数据进行对齐,得到对齐后的第一人声数据,按照对齐后的第一人声数据模拟第二音频数据中的第一人声数据,得到模拟的第一人声数据,并利用模拟的第一人声数据抵消第二音频数据中的第一人声数据。

在具体应用中,第二客户端所获取的第一人声数据为第一客户端用户的纯人声数据,第二音频数据是经第二客户端的外放扬声器播放后由第二客户端的麦克风采集得到的。因

此，第二音频数据与第二客户端所获取的第一人声数据之间存在播放以及采集造成的延时，导致输入自适应滤波器的第一人声数据和第二音频数据并不完全对应。举例而言，从第二音频数据开始产生至第30毫秒后，第一人声数据才开始产生。因此，如果直接对第二音频数据中的第一人声数据进行模拟，将出现由第一延时引起的模拟不准确的问题，很可能造成对第二音频数据中的第一人声数据的过滤效果不佳的问题。

对此，可以在上述可选实施例中，对第一人声数据和第二音频数据进行相关性对比，得到第一人声数据和第二音频数据之间的第一延时，进而在过滤第二音频数据中的第一人声数据时，将第一延时输入自适应滤波器，以使得自适应滤波器按照第一延时对第一人声数据和第二音频数据进行对齐，得到对齐后的第一人声数据，从而利用对齐后的第一人声数据过滤第二音频数据中的第一人声数据。与未按照第一延时进行对齐相比，对齐后的第一人声数据与第二音频数据之间不再存在延时，按照对齐后的第一人声数据模拟第二音频数据中的第一音频数据，得到的模拟的第一音频数据相对而言更加准确，因此，可以提升对第二音频数据中的第一人声数据的过滤效果。

其中，示例性的，相关性比对可以是：分别对所获取的第一人声数据和第二音频数据进行频域转换，得到第一人声数据的频带曲线和第二音频数据的频带曲线；在同一频带坐标系中绘制两条频带曲线，将两条频带曲线首次相交位置的时间确定为第一延时。其中，频带坐标系为以频带为纵轴以时间为横轴的二维坐标系。

示例性，按照第一延时对第一人声数据和第二音频数据进行对齐，得到对齐后的第一人声数据，具体可以是：当第一人声数据早于第二音频数据产生时，可以将第一人声数据的频带曲线在时间轴上后移第一延时对应的长度；或者，当第一人声数据晚于第二音频数据产生时，可以将第一人声数据的频带曲线在时间轴上前移第一延时对应的长度；将移动后的第一人声数据的频带曲线，作为对齐后的第一人声数据。当然，也可以对移动后的第一人声数据的频带曲线进行时域变换，将经过时域变换的数据作为对齐后的第一人声数据。

此外，在具体应用中，第一客户端的数量可以为多个。此时，即时通讯的音质优化与上述可选实施例相似。区别在于，当第一客户端的数量为多个时，第二客户端通过外放扬声器播放的第一人声数据的数量为多个，因此，第二客户端的麦克风采集的第二音频数据中包含多个第一人声数据。此时，需要获取多个第一人声数据，并将多个第一人声数据混合为一个参考信号。对参考信号和第二音频数据进行相关性对比，得到参考信号和第二音频数据之间的第三延时；将第二音频数据、参考信号和第三延时分别输入自适应滤波器，以使得自适应滤波器按照第三延时，对参考信号和第二音频数据进行对齐，得到对齐后的参考信号，按照对齐后的参考信号模拟第二音频数据中作为回声数据的多个第一人声数据，得到模拟的回声数据，并利用模拟的回声数据抵消第二音频数据中的回声数据，完成抵消的第二音频数据为过滤后的第二音频数据。

图2是根据另一示例性实施例示出的一种即时通讯的音质优化方法的流程图，如图2所示，该方法可以包括以下步骤：

步骤S201，获取第一人声数据；第一人声数据为第一客户端的用户的语音数据。

步骤S202，利用外放扬声器播放第一人声数据以及第二客户端本地的背景音乐，得到第一音频数据。

步骤S203，利用麦克风采集第一音频数据和第二人声数据，得到第二音频数据；第二人声数据为第二客户端的用户的语音数据。

步骤S201至步骤S203为与步骤S101至步骤S103相同的步骤，在此不再赘述。

步骤S204，对所获取的第一人声数据和第二音频数据进行相关性对比，得到第一人声数据和第二音频数据之间的第一延时。

步骤S205，将第二音频数据、所获取的第一人声数据和第一延时分别输入自适应滤波器，以使得自适应滤波器按照第一延时，对第一人声数据和第二音频数据进行对齐，得到对齐后的第一人声数据，按照对齐后的第一人声数据模拟第二音频数据中的第一人声数据，得到模拟的第一人声数据，并利用模拟的第一人声数据抵消第二音频数据中的第一人声数据。

步骤S206，将完成抵消的第二音频数据，作为过滤后的第二音频数据。当第一客户端播放的背景音乐的来源为第一客户端本地时，执行步骤S207；当第一客户端播放的背景音乐的来源为第二客户端时，执行步骤S208。

步骤S205至步骤S206为与图1中关于按照第一延时得到对齐后的第一人声数据，进而进行过滤的可选实施例中，获取第一延时、对齐后的第一人声数据以及对第二音频数据中的第一人声数据过滤的步骤相似的步骤，区别在于步骤S206对按照第一客户端播放的背景音乐的不同来源，对过滤后的第二音频数据执行不同的处理。对于相同的部分在此不再赘述。

步骤S207，将过滤后的第二音频数据发送给第一客户端，以使得第一客户端对第一客户端本地的背景音乐和过滤后的第二音频数据进行对齐以及叠加，并播放叠加后的音频数据。

当第一客户端播放的背景音乐的来源为第一客户端本地时，第一客户端播放的背景音乐与第一客户端接收的过滤后的第二音频数据之间，存在由过滤后的第二音频数据的传输造成的延时。如果第一客户端直接播放所接收的过滤后的第二音频数据，会出现过滤后的第二音频数据中的背景音乐与第一客户端本地的背景音乐错乱，影响播放效果的问题。

对此，第一客户端可以对过滤后的第二音频数据和第一客户端本地的背景音乐进行延时计算以及对齐处理。经过延时以及对齐处理得到的对齐后的第一客户端本地的背景音乐，与过滤后的第二音频数据之间不再存在延时。因此，第一客户端播放的叠加后的音频数据

中，背景音乐为第一客户端本地的背景音乐和过滤后的第二音频数据中的背景音乐之间的无延时叠加，在避免两种来源的背景音乐出现错乱的同时，实现背景音乐的加强。

在步骤S207中，第一客户端对第一客户端本地的背景音乐和过滤后的第二音频数据进行对齐以及叠加，具体可以包括：第一客户端对第一客户端本地的背景音乐和过滤后的第二音频数据进行相关性对比，得到第一客户端本地的背景音乐和过滤后的第二音频数据之间的第二延时，进而按照第二延时，对第一客户端本地的背景音乐和过滤后的第二音频数据进行对齐，得到对齐后的第一客户端本地的背景音乐，叠加对齐后的第一客户端本地的背景音乐和过滤后的第二音频数据，得到叠加后的音频数据。其中，第二延时以及对齐后的第一客户端本地的背景音乐的获取，与本申请可选实施例第一延时以及对齐后的第一人声数据的获取类似，区别在于步骤S207中第二延时是第一客户端本地的背景音乐和过滤后的第二音频数据之间的，对齐后的第一客户端本地的背景音乐是对第一客户端本地的背景音乐进行调整后得到的。

示例性的，分别对第一客户端本地的背景音乐和过滤后的第二音频数据进行频域转换，得到第一客户端本地的背景音乐的频带曲线，以及过滤后的第二音频数据的频带曲线；在同一频带坐标系中绘制两条频带曲线，将两条频带曲线首次相交位置的时间确定为第二延时。当第一客户端本地的背景音乐早于过滤后的第二音频数据产生时，可以按照第一客户端本地音乐的播放时间轴，对待播放的第一客户端本地的背景音乐，进行时长为第二延时的后退，得到对齐后的第一客户端本地的背景音乐；或者，将第一客户端本地的背景音乐的频带曲线在时间轴上后移第二延时对应的长度，将移动后的第一客户端本地的背景音乐的频带曲线，作为对齐后的第一客户端本地的背景音乐。或者，当第一客户端本地的背景音乐晚于第二音频数据产生时，可以按照第一客户端本地音乐的播放时间轴，对待播放的第一客户端本地的背景音乐的数据，进行时长为第二延时的快进，得到对齐后的第一客户端本地的背景音乐；或者，可以将第一客户端本地的背景音乐的频带曲线在时间轴上前移第一延时对应的长度；将移动后的第一客户端本地的背景音乐的频带曲线，作为对齐后的第一客户端本地的背景音乐。当然，如果过滤后的第二音频数据已经为频域数据，可以直接使用过滤后的第二音频数据，无需进行频域转换。

步骤S208，按照第一延时，对第二客户端本地的背景音乐和过滤后的第二音频数据进行对齐以及叠加，并将叠加后的音频数据发送给第一客户端，以使得第一客户端播放叠加后的音频数据。

当第一客户端播放的背景音乐的来源为第二客户端时，第二客户端本地的背景音乐可以是第二客户端本地存储，或者从服务器中下载的背景音乐。并且，过滤后的第二音频数据是过滤第二客户端利用麦克风采集到的第二音频数据得到的。因此，第二客户端本地的背景音乐与过滤后的第二音频数据之间存在由第二客户端对第二音频数据的采集造成的

延时，也就是第一延时。对此，第二客户端可以按照第一延时，对过滤后的第二音频数据和第二客户端本地的背景音乐进行对齐的处理。经过对齐处理得到的对齐后的第二客户端本地的背景音乐，与过滤后的第二音频数据之间不再存在延时。因此，第一客户端播放的叠加后的音频数据中，背景音乐为第二客户端本地的背景音乐和过滤后的第二音频数据中的背景音乐之间的无延时叠加，在避免两种来源的背景音乐出现错乱的同时，实现背景音乐的加强。

在步骤S208中，按照第一延时，对第二客户端本地的背景音乐和过滤后的第二音频数据进行对齐以及叠加，具体可以包括：按照第一延时，对第二客户端本地的背景音乐和过滤后的第二音频数据进行对齐，得到对齐后的第二客户端本地的背景音乐，叠加对齐后的第二客户端本地的背景音乐和过滤后的第二音频数据，得到叠加后的音频数据。其中，第一延时为上述关于过滤第二音频数据中的第一人声数据的可选实施例中获取的延时，详见上述可选实施例的描述。对齐后的第二客户端本地的背景音乐的获取，步骤S207中对齐后的第一客户端本地的背景音乐的获取类似，区别在于步骤S208中对齐后的第二客户端本地的背景音乐是对第二客户端本地的背景音乐进行调整后得到的。对于相同部分在此不再赘述，详见上述步骤S207的描述。

另外，当第一客户端播放的背景音乐的来源为第二客户端时，第一客户端播放的背景音乐可以是第二客户端所采集的过滤后的第二音频数据中的背景音乐。此时，与步骤S105相同，可以将过滤后的第二音频数据发送给第一客户端，以使得第一客户端播放过滤后的第二音频数据。

在具体应用中，当第一客户端播放的背景音乐的来源为第二客户端时，如果需要加强第一客户端播放的背景音乐，可以执行步骤S208；或者，如果需要减少传输的数据量、提高即时通讯的效率，可以执行将过滤后的第二音频数据发送给第一客户端，以使得第一客户端播放过滤后的第二音频数据。

图3是根据又一示例性实施例示出的一种即时通讯的音质优化方法的流程图，如图3所示，应用于第一客户端，该方法可以包括以下步骤：

步骤S301，发送第一人声数据给第二客户端，以使得第二客户端利用外放扬声器播放第一人声数据以及第二客户端本地的背景音乐，得到第一音频数据；第一人声数据为第一客户端的用户的语音数据。

步骤S302，发送第三音频数据给第二客户端，以使得第二客户端过滤第三音频数据中的背景音乐，得到第一人声数据，并利用外放扬声器播放第一人声数据以及第二客户端本地的背景音乐，得到第一音频数据；第三音频数据为第一客户端利用麦克风采集第一人声数据和第一客户端本地的背景音乐得到的音频数据。

上述步骤S301和步骤S302为并列步骤，分别适用于第一客户端以不同方式播放背景音

乐、以及对第一客户端采集到的数据的不同处理方式。具体的，当第一客户端利用耳机播放背景音乐时，第一客户端的麦克风采集到的第一人声数据中不会混入第一客户端播放的第一客户端本地的背景音乐，因此，可以执行步骤S301。或者，当第一客户端利用外放扬声器播放背景音乐时，第一客户端的麦克风采集第一人声数据时，第一客户端播放的第一客户端本地的背景音乐也会被采集，此时，第一客户端的麦克风采集到的是第三音频数据。因此，第一客户端可以过滤第三音频数据中的背景音乐，得到第一人声数据，并执行步骤S301。或者，当第一客户端的麦克风采集到的是第三音频数据时，可以执行步骤S302，由第二客户端过滤第三音频数据中的背景音乐，得到第一人声数据。

步骤S303，接收第二客户端发送的第二音频数据；第二音频数据为第二客户端利用麦克风采集第一音频数据和第二人声数据，得到的音频数据；第二人声数据为第二客户端的用户的声音数据。

在步骤S303中，第二音频数据与图1相关实施例中的第二音频数据相同，在此不再赘述。

步骤S304，过滤第二音频数据中的第一人声数据，得到过滤后的第二音频数据。

步骤S305，当第一客户端播放的背景音乐的来源为第二客户端时，播放过滤后的第二音频数据。

上述步骤S304至步骤S305与步骤S104至S105为相似步骤。区别在于上述步骤S304至步骤S305的执行主体为第一客户端，无需进行过滤后的第二音频数据的发送。当然，如果执行的是步骤S302，为了保证后续能够执行步骤S304，第一客户端需要利用第三音频数据得到第一人声数据。对于相同部分在此不再赘述。

本申请实施例中，在存在背景音乐的即时通讯系统中，由于第一人声数据与背景音乐相比持续的时间较短，因此，过滤第二音频数据中的第一人声数据与传统的对第二音频数据持续进行回声过滤相比，可以减少对第二音频数据的过度过滤，从而减少对第二音频数据中第二人声数据的过度过滤，减少第二人声的卡顿和忽大忽小等问题，减少对非回声的第二人声的音质的损耗。并且，当第一客户端播放的背景音乐的来源为第二客户端时，过滤后的第二音频数据中的背景音乐可以作为第一客户端播放的背景音乐。因此，当第一客户端播放的背景音乐的来源为第二客户端时，将过滤后的第二音频数据发送给第一客户端进行播放，可以避免过滤后的第二音频数据中的背景音乐成为第一客户端的噪声，保证回声消除的效果。可见，本方案能够在存在背景音乐的即时通讯系统中，兼顾回声消除和非回声的人声损耗的减少。

可选的，在上述步骤S304：过滤第二音频数据中的第一人声数据，得到过滤后的第二音频数据之后，本申请实施例提供的即时通讯的音质优化方法，还可以包括如下步骤：

当第一客户端播放的背景音乐的来源为第一客户端本地时，对第一客户端本地的背景

音乐和过滤后的第二音频数据进行相关性对比，得到第一客户端本地的背景音乐和过滤后的第二音频数据之间的第二延时；

按照第二延时，对第一客户端本地的背景音乐和过滤后的第二音频数据进行对齐，得到对齐后的第一客户端本地的背景音乐，叠加对齐后的第一客户端本地的背景音乐和过滤后的第二音频数据，得到叠加后的音频数据；

播放叠加后的音频数据。

其中，第二延时的获取、对齐后的第一客户端本地的背景音乐的获取以及叠加后的音频数据与步骤S207相似。区别在于在本可选实施例中，执行主体为第一客户端。

另外，当第一客户端播放的背景音乐的来源为第二客户端时，如果需要减少传输的数据量、提高即时通讯的效率，可以执行步骤S305。或者，如果需要加强第一客户端播放的背景音乐，可以执行如下步骤；

对第一人声数据和第二音频数据进行相关性对比，得到第一人声数据和第二音频数据之间的第一延时；

按照第一延时，对接收的第二客户端本地的背景音乐和过滤后的第二音频数据进行对齐，得到对齐后的第二客户端本地的背景音乐，叠加对齐后的第二客户端本地的背景音乐和过滤后的第二音频数据，得到叠加后的音频数据；

播放叠加后的音频数据。

上述步骤与步骤S208相似，区别在于在本可选实施例中，执行主体为第一客户端。由于本可选实施例中第一客户端播放的背景音乐是对过滤后的第二音频数据中的背景乐和所接收的第二客户端本地的背景音乐的叠加，因此，相对于仅存在过滤后的第二音频数据中的背景音乐而言，可以实现背景音乐的加强。

相应于上述方法实施例，本申请实施例还提供一种即时通讯的音质优化装置。

图4是根据一示例性实施例示出的一种即时通讯的音质优化装置框图，应用于第二客户端，该装置可以包括：第一人声获取模块401、第一音频获取模块402、第二音频获取模块403、过滤模块404以及发送模块405。

第一人声获取模块401，被配置为获取第一人声数据；所述第一人声数据为第一客户端的用户的语音数据；

第一音频获取模块402，被配置为利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；

第二音频获取模块403，被配置为利用麦克风采集所述第一音频数据和第二人声数据，得到第二音频数据；所述第二人声数据为所述第二客户端的用户的语音数据；

过滤模块404，被配置为过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

发送模块405,被配置为在所述第一客户端播放的背景音乐的来源为所述第二客户端时,将所述过滤后的第二音频数据发送给所述第一客户端,以使得所述第一客户端播放所述过滤后的第二音频数据。

本申请实施例中,在存在背景音乐的即时通讯系统中,由于第一人声数据与背景音乐相比持续的时间较短,因此,过滤第二音频数据中的第一人声数据与传统的对第二音频数据持续进行回声过滤相比,可以减少对第二音频数据的过度过滤,从而减少对第二音频数据中第二人声数据的过度过滤,减少第二人声的卡顿和忽大忽小等问题,减少对非回声的第二人声的音质的损耗。并且,当第一客户端播放的背景音乐的来源为第二客户端时,过滤后的第二音频数据中的背景音乐可以作为第一客户端播放的背景音乐。因此,当第一客户端播放的背景音乐的来源为第二客户端时,将过滤后的第二音频数据发送给第一客户端进行播放,可以避免过滤后的第二音频数据中的背景音乐成为第一客户端的噪声,保证回声消除的效果。可见,本方案能够在存在背景音乐的即时通讯系统中,兼顾回声消除和非回声的人声损耗的减少。

可选的,所述第一人声获取模块401,被配置为:

在所述第一客户端利用耳机播放背景音乐时,接收所述第一客户端发送的第一人声数据;

或者,在所述第一客户端利用外放扬声器播放背景音乐时,接收所述第一客户端发送的所述第一客户端对第三音频数据中的背景音乐过滤得到的第一人声数据;所述第三音频数据为所述第一客户端利用麦克风采集所述第一人声数据和所述第一客户端播放的第一客户端本地的背景音乐得到的音频数据;

或者,在所述第一客户端利用外放扬声器播放背景音乐时,接收所述第一客户端发送的所述第三音频数据;过滤所述第三音频数据中的背景音乐,得到第一人声数据。

可选的,所述过滤模块404,被配置为:

将所述第二音频数据和所获取的第一人声数据分别输入自适应滤波器,以使得所述自适应滤波器按照所述第一人声数据,模拟所述第二音频数据中的第一人声数据,得到模拟的第一人声数据,并利用所述模拟的第一人声数据抵消所述第二音频数据中的第一人声数据;

将完成抵消的第二音频数据,作为过滤后的第二音频数据。

可选的,所述装置还包括:延时对齐模块406;

所述延时对齐模块406,被配置为在所述第二音频获取模块403利用麦克风采集所述第一音频数据和第二人声数据,得到第二音频数据之后,在所述过滤模块404将所述第二音频数据和所获取的第一人声数据分别输入自适应滤波器之前,对所获取的第一人声数据和所述第二音频数据进行相关性对比,得到所述第一人声数据和所述第二音频数据之间的第

一延时；

所述过滤模块404，被配置为将所述第二音频数据、所获取的第一人声数据和所述第一延时分别输入自适应滤波器，以使得所述自适应滤波器按照所述第一延时，对所述第一人声数据和所述第二音频数据进行对齐，得到对齐后的第一人声数据，按照所述对齐后的第一人声数据模拟所述第二音频数据中的第一人声数据，得到模拟的第一人声数据，并利用所述模拟的第一人声数据抵消所述第二音频数据中的第一人声数据。

可选的，所述发送模块405，被配置为：

在得到所述自适应滤波器输出的过滤后的第二音频数据之后，在所述第一客户端播放的背景音乐的来源为所述第一客户端本地时，将所述过滤后的第二音频数据发送给所述第一客户端，以使得所述第一客户端对所述第一客户端本地的背景音乐和所述过滤后的第二音频数据进行对齐以及叠加，并播放所述叠加后的音频数据；

或者，在所述第一客户端播放的背景音乐的来源为所述第二客户端时，所述延时对齐模块，被配置为按照所述第一延时，对所述第二客户端本地的背景音乐和所述过滤后的第二音频数据进行对齐以及叠加，将叠加后的音频数据发送给所述第一客户端，以使得所述第一客户端播放所述叠加后的音频数据。

图5是根据另一示例性实施例示出的一种即时通讯的音质优化装置框图，应用于第一客户端，该装置可以包括：发送模块501、接收模块502、过滤模块503以及播放模块504。

发送模块501，被配置为发送第一人声数据给第二客户端，以使得所述第二客户端利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；所述第一人声数据为所述第一客户端的用户的语音数据；或者，发送第三音频数据给所述第二客户端，以使得所述第二客户端过滤所述第三音频数据中的背景音乐，得到第一人声数据，并利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；所述第三音频数据为所述第一客户端利用麦克风采集所述第一人声数据和所述第一客户端本地的背景音乐得到的音频数据；

接收模块502，被配置为接收所述第二客户端发送的第二音频数据；所述第二音频数据为所述第二客户端利用麦克风采集所述第一音频数据和第二人声数据，得到的音频数据；所述第二人声数据为所述第二客户端的用户的语音数据；

过滤模块503，被配置为过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

播放模块504，被配置为在所述第一客户端播放的背景音乐的来源为所述第二客户端时，播放所述过滤后的第二音频数据。

本申请实施例中，在存在背景音乐的即时通讯系统中，由于第一人声数据与背景音乐相比持续的时间较短，因此，过滤第二音频数据中的第一人声数据与传统的对第二音频

数据进行持续性回声过滤相比,可以减少对第二音频数据的过度过滤,从而减少对第二音频数据中第二人声数据的过度过滤,减少第二人声的卡顿和忽大忽小等问题,减少对非回声的第二人声的音质的损耗。并且,当第一客户端播放的背景音乐的来源为第二客户端时,过滤后的第二音频数据中的背景音乐可以作为第一客户端播放的背景音乐。因此,当第一客户端播放的背景音乐的来源为第二客户端时,将过滤后的第二音频数据发送给第一客户端进行播放,可以避免过滤后的第二音频数据中的背景音乐成为第一客户端的噪声,保证回声消除的效果。可见,本方案能够在存在背景音乐的即时通讯系统中,兼顾回声消除和非回声的人声损耗的减少。

可选的,所述装置还包括:延时对齐模块505;

所述延时对齐模块505,被配置为在所述过滤模块503过滤所述第二音频数据中的第一人声数据,得到过滤后的第二音频数据之后,当所述第一客户端播放的背景音乐的来源为所述第一客户端本地时,对所述第一客户端本地的背景音乐和所述过滤后的第二音频数据进行相关性对比,得到所述第一客户端本地的背景音乐和所述过滤后的第二音频数据之间的第二延时;

按照所述第二延时,对所述第一客户端本地的背景音乐和所述过滤后的第二音频数据进行对齐,得到对齐后的第一客户端本地的背景音乐,叠加所述对齐后的第一客户端本地的背景音乐和所述过滤后的第二音频数据,得到叠加后的音频数据;

所述播放模块504,被配置为播放所述叠加后的音频数据。

相应于上述方法实施例,本申请实施例还提供一种电子设备。

图6是根据一示例性实施例示出的一种电子设备,该电子设备可以包括:

处理器601;

用于存储处理器可执行指令的存储器602;

其中,处理器601被配置为:执行存储器602上所存放的可执行指令时,实现本申请实施例所提供的任一种应用于第二客户端的即时通讯的音质优化方法的步骤。

可以理解的是,该电子设备为即时通讯系统中的第二客户端。在具体应用中,该电子设备可以为计算机、智能移动终端、平板设备以及服务器等等。

图7是根据另一示例性实施例示出的电子设备700的框图。电子设备700可以是移动电话,计算机,数字广播终端,消息收发设备,游戏控制台,平板设备,健身设备,个人数字助理等。

参照图7,电子设备700可以包括以下一个或多个组件:处理组件702,存储器704,电源组件706,多媒体组件708,音频组件710,输入/输出(I/O)的接口712以及通信组件716。

处理组件702通常控制电子设备700的整体操作,诸如与显示,电话呼叫,数据通信,相机操作和记录操作相关联的操作。处理组件702可以包括一个或多个处理器720来执行指

令，以完成上述的方法的全部或部分步骤。此外，处理组件702可以包括一个或多个模块，便于处理组件702和其他组件之间的交互。例如，处理组件702可以包括多媒体模块，以方便多媒体组件708和处理组件702之间的交互。

存储器704被配置为存储各种类型的数据以支持在设备700的操作。这些数据的示例包括用于在电子设备700上操作的任何应用程序或方法的指令，联系人数据，电话簿数据，消息，图片，视频等。存储器704可以由任何类型的易失性或非易失性存储设备或者它们的组合实现，如SRAM（Static Random Access Memory，静态随机存取存储器），EEPROM（Electrically Erasable Programmable Read Only Memory，电可擦除可编程只读存储器），EPROM（Erasable Programmable Read-Only Memory，可擦除可编程只读存储器），PROM（Programmable Read-Only Memory，可编程只读存储器），ROM，磁存储器，快闪存储器，磁盘或光盘。

电源组件706为装置700的各种组件提供电力。电源组件706可以包括电源管理系统，一个或多个电源，及其他与为装置700生成、管理和分配电力相关联的组件。

多媒体组件708包括在所述设备700和用户之间的提供一个输出接口的屏幕。在一些实施例中，屏幕可以包括液晶显示器（LCD）和触摸面板（TP）。如果屏幕包括触摸面板，屏幕可以被实现为触摸屏，以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界，而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中，多媒体组件708包括一个前置摄像头和/或后置摄像头。当设备700处于操作模式，如拍摄模式或视频模式时，前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

音频组件710被配置为输出和/或输入音频信号。例如，音频组件710包括一个麦克风（MIC），当装置700处于操作模式，如呼叫模式、记录模式和语音识别模式时，麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器704或经由通信组件716发送。在一些实施例中，音频组件710还包括一个扬声器，用于输出音频信号。

I/O接口712为处理组件702和外围接口模块之间提供接口，上述外围接口模块可以是键盘，点击轮，按钮等。这些按钮可包括但不限于：主页按钮、音量按钮、启动按钮和锁定按钮。

通信组件716被配置为便于设备700和其他设备之间有线或无线方式的通信。装置700可以接入基于通信标准的无线网络，如WiFi，运营商网络（如2G、3G、4G或5G），或它们的组合。在一个示例性实施例中，通信组件716经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中，所述通信组件716还可以包括NFC（Near Field Communication，近场通信）模块，以促进短程通信。例如，NFC模块可基于

RFID (Radio Frequency Identification, 射频识别) 技术, IrDA (Infrared Data Association, 红外数据协会) 技术, UWB (Ultra Wideband, 超宽带) 技术, BT (Blue Tooth, 蓝牙) 技术和其他技术来实现。

在示例性实施例中, 电子设备700可以被一个或多个应用ASIC (Application Specific Integrated Circuit, 专用集成电路)、DSP (Digital Signal Processor, 数字信号处理器)、DSPD (Digital Signal Processing Equipment, 数字信号处理设备)、PLD (Programmable Logic Devices, 可编程逻辑器件)、FPGA (Field Programmable Gate Array, 现场可编程门阵列)、控制器、微控制器、微处理器或其他电子元件实现, 用于执行上述应用于第二客户端的即时通讯的音质优化方法。

图8是根据又一示例性实施例示出的一种电子设备, 该电子设备可以包括:

处理器801;

用于存储处理器可执行指令的存储器802;

其中, 处理器801被配置为: 执行存储器802上所存放的可执行指令时, 实现本申请实施例所提供的任一种应用于第一客户端的即时通讯的音质优化方法的步骤。

可以理解的是, 该电子设备为即时通讯系统中的第一客户端。在具体应用中, 该电子设备可以为计算机、智能移动终端、平板设备以及服务器等等。

图9是根据再一示例性实施例示出的电子设备900的框图。电子设备900可以是移动电话, 计算机, 数字广播终端, 消息收发设备, 游戏控制台, 平板设备, 健身设备, 个人数字助理等。

参照图9, 电子设备900可以包括以下一个或多个组件: 处理组件902, 存储器904, 电源组件906, 多媒体组件908, 音频组件910, 输入/输出(I/O)的接口912以及通信组件916。

处理组件902通常控制电子设备900的整体操作, 诸如与显示, 电话呼叫, 数据通信, 相机操作和记录操作相关联的操作。处理组件902可以包括一个或多个处理器920来执行指令, 以完成上述的方法的全部或部分步骤。此外, 处理组件902可以包括一个或多个模块, 便于处理组件902和其他组件之间的交互。例如, 处理组件902可以包括多媒体模块, 以方便多媒体组件908和处理组件902之间的交互。

存储器904被配置为存储各种类型的数据以支持在设备900的操作。这些数据的示例包括用于在电子设备900上操作的任何应用程序或方法的指令, 联系人数据, 电话簿数据, 消息, 图片, 视频等。存储器904可以由任何类型的易失性或非易失性存储设备或者它们的组合实现, 如SRAM (Static Random Access Memory, 静态随机存取存储器), EEPROM (Electrically Erasable Programmable Read Only Memory, 电可擦除可编程只读存储器), EPROM (Erasable Programmable Read-Only Memory, 可擦除可编程只读存储器), PROM (Programmable Read-Only Memory, 可编程只读存储器), ROM, 磁存储器, 快闪存储器,

磁盘或光盘。

电源组件906为装置900的各种组件提供电力。电源组件906可以包括电源管理系统，一个或多个电源，及其他与为装置900生成、管理和分配电力相关联的组件。

多媒体组件908包括在所述设备900和用户之间的提供一个输出接口的屏幕。在一些实施例中，屏幕可以包括液晶显示器（LCD）和触摸面板（TP）。如果屏幕包括触摸面板，屏幕可以被实现为触摸屏，以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界，而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中，多媒体组件908包括一个前置摄像头和/或后置摄像头。当设备900处于操作模式，如拍摄模式或视频模式时，前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

音频组件910被配置为输出和/或输入音频信号。例如，音频组件910包括一个麦克风（MIC），当装置900处于操作模式，如呼叫模式、记录模式和语音识别模式时，麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器904或经由通信组件916发送。在一些实施例中，音频组件910还包括一个扬声器，用于输出音频信号。

I/O接口912为处理组件902和外围接口模块之间提供接口，上述外围接口模块可以是键盘，点击轮，按钮等。这些按钮可包括但不限于：主页按钮、音量按钮、启动按钮和锁定按钮。

通信组件916被配置为便于设备900和其他设备之间有线或无线方式的通信。装置900可以接入基于通信标准的无线网络，如WiFi，运营商网络（如2G、3G、4G或5G），或它们的组合。在一个示例性实施例中，通信组件916经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中，所述通信组件716还可以包括NFC（Near Field Communication，近场通信）模块，以促进短程通信。例如，NFC模块可基于RFID（Radio Frequency Identification，射频识别）技术，IrDA（Infrared Data Association，红外数据协会）技术，UWB（Ultra Wideband，超宽带）技术，BT（Blue Tooth，蓝牙）技术和其他技术来实现。

在示例性实施例中，电子设备900可以被一个或多个应用ASIC（Application Specific Integrated Circuit，专用集成电路）、DSP（Digital Signal Processor，数字信号处理器）、DSPD（Digital Signal Processing Equipment，数字信号处理设备）、PLD（Programmable Logic Devices，可编程逻辑器件）、FPGA（Field Programmable Gate Array，现场可编程门阵列）、控制器、微控制器、微处理器或其他电子元件实现，用于执行上述应用于第一客户端的即时通讯的音质优化方法。

另外，本申请实施例还提供了一种非临时性计算机可读存储介质，包含于电子设备，

当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行本申请实施例中任一所述的应用于第二客户端的即时通讯的音质优化方法的步骤。

在示例性实施例中，一种包括指令的非临时性计算机可读存储介质，例如包括指令的存储器602，上述指令可由处理器601执行以完成上述方法；或者，包括指令的存储器704，上述指令可由电子设备700的处理组件702执行以完成上述应用于第二客户端的即时通讯的音质优化方法。例如，所述非临时性计算机可读存储介质可以是ROM(Read-Only Memory, 只读存储器)、RAM(Random Access Memory, 随机存取存储器)、CD-ROM(Compact Disc Read-Only Memory, 光盘只读存储器)、磁带、软盘和光数据存储设备等。

本申请实施例还提供了另一种非临时性计算机可读存储介质，包含于电子设备，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行本申请实施例中任一所述的应用于第一客户端的即时通讯的音质优化方法的步骤。

在示例性实施例中，一种包括指令的非临时性计算机可读存储介质，例如包括指令的存储器802，上述指令可由处理器801执行以完成上述应用于第一客户端的即时通讯的音质优化方法；或者，包括指令的存储器904，上述指令可由电子设备900的处理组件器902执行以完成上述方法。例如，所述非临时性计算机可读存储介质可以是ROM(Read-Only Memory, 只读存储器)、RAM(Random Access Memory, 随机存取存储器)、CD-ROM(Compact Disc Read-Only Memory, 光盘只读存储器)、磁带、软盘和光数据存储设备等。

在本申请提供的又一实施例中，还提供了一种包含指令的计算机程序产品，当其在电子设备上运行时，使得电子设备执行上述实施例中任一所述的应用于第二客户端的即时通讯的音质优化方法。

在本申请提供的又一实施例中，还提供了一种包含指令的计算机程序产品，当其在电子设备上运行时，使得电子设备执行上述实施例中任一所述的应用于第一客户端的即时通讯的音质优化方法。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时，全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线，例如：同轴电缆、光纤、DSL(Digital Subscriber Line, 数字用户线；或无线，例如：红外线、无线电、微波等方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数

据中心等数据存储设备。所述可用介质可以是磁性介质，例如：软盘、硬盘、磁带；光介质，例如：DVD (Digital Versatile Disc, 数字通用光盘)；或者半导体介质，例如：SSD (Solid State Disk, 固态硬盘) 等。

权 利 要 求

1、一种即时通讯的音质优化方法，应用于第二客户端，所述方法包括：

获取第一人声数据；所述第一人声数据为第一客户端的用户的语音数据；

利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；

利用麦克风采集所述第一音频数据和第二人声数据，得到第二音频数据；所述第二人声数据为所述第二客户端的用户的语音数据；

过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

当所述第一客户端播放的背景音乐的来源为所述第二客户端时，将所述过滤后的第二音频数据发送给所述第一客户端，以使得所述第一客户端播放所述过滤后的第二音频数据。

2、根据权利要求1所述的方法，获取第一人声数据，包括：

当所述第一客户端利用耳机播放背景音乐时，接收所述第一客户端发送的第一人声数据；

或者，当所述第一客户端利用外放扬声器播放背景音乐时，接收所述第一客户端发送的所述第一人声数据；所述第一人声数据为所述第一客户端利用麦克风采集所述第一人声数据和所述第一客户端播放的本地的背景音乐得到的音频数据；

或者，当所述第一客户端利用外放扬声器播放背景音乐时，接收所述第一客户端发送的所述第三音频数据；过滤所述第三音频数据中的背景音乐，得到第一人声数据。

3、根据权利要求1所述的方法，过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据，包括：

将所述第二音频数据和所获取的第一人声数据分别输入自适应滤波器，以使得所述自适应滤波器按照所述第一人声数据，模拟所述第二音频数据中的第一人声数据，得到模拟的第一人声数据，并利用所述模拟的第一人声数据抵消所述第二音频数据中的第一人声数据；

将完成抵消的第二音频数据，作为过滤后的第二音频数据。

4、根据权利要求3所述的方法，在利用麦克风采集所述第一音频数据和第二人声数据，得到第二音频数据之后，将所述第二音频数据和所获取的第一人声数据分别输入自适应滤波器之前，所述方法还包括：

对所获取的第一人声数据和所述第二音频数据进行相关性对比，得到所述第一人声数据和所述第二音频数据之间的第一延时；

将所述第二音频数据和所获取的第一人声数据分别输入自适应滤波器，以使得所述自适应滤波器按照所输入的第一人声数据，模拟所述第二音频数据中的第一人声数据，得到

模拟的第一人声数据，并利用所述模拟的第一人声数据抵消所述第二音频数据中的第一人声数据，包括：

将所述第二音频数据、所获取的第一人声数据和所述第一延时分别输入自适应滤波器，以使得所述自适应滤波器按照所述第一延时，对所述第一人声数据和所述第二音频数据进行对齐，得到对齐后的第一人声数据，按照所述对齐后的第一人声数据模拟所述第二音频数据中的第一人声数据，得到模拟的第一人声数据，并利用所述模拟的第一人声数据抵消所述第二音频数据中的第一人声数据。

5、根据权利要求4所述的方法，在得到所述自适应滤波器输出的过滤后的第二音频数据之后，所述方法还包括：

当所述第一客户端播放的背景音乐的来源为所述第一客户端本地时，将所述过滤后的第二音频数据发送给所述第一客户端，以使得所述第一客户端对所述第一客户端本地的背景音乐和所述过滤后的第二音频数据进行对齐以及叠加，并播放叠加后的音频数据；

或者，当所述第一客户端播放的背景音乐的来源为所述第二客户端时，按照所述第一延时，对所述第二客户端本地的背景音乐和所述过滤后的第二音频数据进行对齐以及叠加，并将叠加后的音频数据发送给所述第一客户端，以使得所述第一客户端播放所述叠加后的音频数据。

6、一种即时通讯的音质优化方法，应用于第一客户端，所述方法包括：

发送第一人声数据给第二客户端，以使得所述第二客户端利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；或者，发送第三音频数据给所述第二客户端，以使得所述第二客户端过滤所述第三音频数据中的背景音乐，得到第一人声数据，并利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；其中，所述第一人声数据为所述第一客户端的用户的语音数据；所述第三音频数据为所述第一客户端利用麦克风采集所述第一人声数据和所述第一客户端本地的背景音乐得到的音频数据；

接收所述第二客户端发送的第二音频数据；所述第二音频数据为所述第二客户端利用麦克风采集所述第一音频数据和第二人声数据，得到的音频数据；所述第二人声数据为所述第二客户端的用户的语音数据；

过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

当所述第一客户端播放的背景音乐的来源为所述第二客户端时，播放所述过滤后的第二音频数据。

7、根据权利要求6所述的方法，在过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据之后，所述方法还包括：

当所述第一客户端播放的背景音乐的来源为所述第一客户端本地时，对所述第一客户

端本地的背景音乐和所述过滤后的第二音频数据进行相关性对比，得到所述第一客户端本地的背景音乐和所述过滤后的第二音频数据之间的第二延时；

按照所述第二延时，对所述第一客户端本地的背景音乐和所述过滤后的第二音频数据进行对齐，得到对齐后的第一客户端本地的背景音乐，叠加所述对齐后的第一客户端本地的背景音乐和所述过滤后的第二音频数据，得到叠加后的音频数据；

播放所述叠加后的音频数据。

8、一种即时通讯的音质优化装置，应用于第二客户端，所述装置包括：

第一人声获取模块，被配置为获取第一人声数据；所述第一人声数据为第一客户端的用户的声音数据；

第一音频获取模块，被配置为利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；

第二音频获取模块，被配置为利用麦克风采集所述第一音频数据和第二人声数据，得到第二音频数据；所述第二人声数据为所述第二客户端的用户的语音数据；

过滤模块，被配置为过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

发送模块，被配置为在所述第一客户端播放的背景音乐的来源为所述第二客户端时，将所述过滤后的第二音频数据发送给所述第一客户端，以使得所述第一客户端播放所述过滤后的第二音频数据。

9、根据权利要求8所述的装置，所述第一人声获取模块，被配置为：

在所述第一客户端利用耳机播放背景音乐时，接收所述第一客户端发送的第一人声数据；

或者，在所述第一客户端利用外放扬声器播放背景音乐时，接收所述第一客户端发送的所述第一客户端对第三音频数据中的背景音乐过滤得到的第一人声数据；所述第三音频数据为所述第一客户端利用麦克风采集所述第一人声数据和所述第一客户端播放的本地的背景音乐得到的音频数据；

或者，在所述第一客户端利用外放扬声器播放背景音乐时，接收所述第一客户端发送的所述第三音频数据；过滤所述第三音频数据中的背景音乐，得到第一人声数据。

10、根据权利要求8所述的装置，所述过滤模块，被配置为：

将所述第二音频数据和所获取的第一人声数据分别输入自适应滤波器，以使得所述自适应滤波器按照所述第一人声数据，模拟所述第二音频数据中的第一人声数据，得到模拟的第一人声数据，并利用所述模拟的第一人声数据抵消所述第二音频数据中的第一人声数据；

将完成抵消的第二音频数据，作为过滤后的第二音频数据。

11、根据权利要求10所述的装置，所述装置还包括：延时对齐模块；

所述延时对齐模块，被配置为在所述第二音频获取模块利用麦克风采集所述第一音频数据和第二人声数据，得到第二音频数据之后，在所述过滤模块将所述第二音频数据和所获取的第一人声数据分别输入自适应滤波器之前，对所获取的第一人声数据和所述第二音频数据进行相关性对比，得到所述第一人声数据和所述第二音频数据之间的第一延时；

所述过滤模块，被配置为将所述第二音频数据、所获取的第一人声数据和所述第一延时分别输入自适应滤波器，以使得所述自适应滤波器按照所述第一延时，对所述第一人声数据和所述第二音频数据进行对齐，得到对齐后的第一人声数据，按照所述对齐后的第一人声数据模拟所述第二音频数据中的第一人声数据，得到模拟的第一人声数据，并利用所述模拟的第一人声数据抵消所述第二音频数据中的第一人声数据。

12、根据权利要求11所述的装置，所述发送模块，被配置为：

在得到所述自适应滤波器输出的过滤后的第二音频数据之后，在所述第一客户端播放的背景音乐的来源为所述第一客户端本地时，将所述过滤后的第二音频数据发送给所述第一客户端，以使得所述第一客户端对所述第一客户端本地的背景音乐和所述过滤后的第二音频数据进行对齐以及叠加，并播放所述叠加后的音频数据；

或者，在所述第一客户端播放的背景音乐的来源为所述第二客户端时，所述延时对齐模块，被配置为按照所述第一延时，对所述第二客户端本地的背景音乐和所述过滤后的第二音频数据进行对齐以及叠加，将叠加后的音频数据发送给所述第一客户端，以使得所述第一客户端播放所述叠加后的音频数据。

13、一种即时通讯的音质优化装置，应用于第一客户端，所述装置包括：

发送模块，被配置为发送第一人声数据给第二客户端，以使得所述第二客户端利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；或者，发送第三音频数据给所述第二客户端，以使得所述第二客户端过滤所述第三音频数据中的背景音乐，得到第一人声数据，并利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；其中，所述第一人声数据为所述第一客户端的用户的语音数据；所述第三音频数据为所述第一客户端利用麦克风采集所述第一人声数据和所述第一客户端本地的背景音乐得到的音频数据；

接收模块，被配置为接收所述第二客户端发送的第二音频数据；所述第二音频数据为所述第二客户端利用麦克风采集所述第一音频数据和第二人声数据，得到的音频数据；所述第二人声数据为所述第二客户端的用户的语音数据；

过滤模块，被配置为过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

播放模块，被配置为在所述第一客户端播放的背景音乐的来源为所述第二客户端时，

播放所述过滤后的第二音频数据。

14、根据权利要求13所述的装置，所述装置还包括：延时对齐模块；

所述延时对齐模块，被配置为在所述过滤模块过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据之后，当所述第一客户端播放的背景音乐的来源为所述第一客户端本地时，对所述第一客户端本地的背景音乐和所述过滤后的第二音频数据进行相关性对比，得到所述第一客户端本地的背景音乐和所述过滤后的第二音频数据之间的第二延时；

按照所述第二延时，对所述第一客户端本地的背景音乐和所述过滤后的第二音频数据进行对齐，得到对齐后的第一客户端本地的背景音乐，叠加所述对齐后的第一客户端本地的背景音乐和所述过滤后的第二音频数据，得到叠加后的音频数据；

所述播放模块，被配置为播放所述叠加后的音频数据。

15、一种电子设备，应用于第二客户端，所述电子设备包括：

处理器；

用于存储所述处理器可执行指令的存储器；

其中，所述处理器被配置为执行：

获取第一人声数据；所述第一人声数据为第一客户端的用户的语音数据；

利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；

利用麦克风采集所述第一音频数据和第二人声数据，得到第二音频数据；所述第二人声数据为所述第二客户端的用户的语音数据；

过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

当所述第一客户端播放的背景音乐的来源为所述第二客户端时，将所述过滤后的第二音频数据发送给所述第一客户端，以使得所述第一客户端播放所述过滤后的第二音频数据。

16、根据权利要求15所述的电子设备，所述处理器被配置为执行：

当所述第一客户端利用耳机播放背景音乐时，接收所述第一客户端发送的第一人声数据；

或者，当所述第一客户端利用外放扬声器播放背景音乐时，接收所述第一客户端发送的所述第一客户端对第三音频数据中的背景音乐过滤得到的第一人声数据；所述第三音频数据为所述第一客户端利用麦克风采集所述第一人声数据和所述第一客户端播放的本地的背景音乐得到的音频数据；

或者，当所述第一客户端利用外放扬声器播放背景音乐时，接收所述第一客户端发送的所述第三音频数据；过滤所述第三音频数据中的背景音乐，得到第一人声数据。

17、根据权利要求15所述的电子设备，所述处理器被配置为执行：

将所述第二音频数据和所获取的第一人声数据分别输入自适应滤波器，以使得所述自适应滤波器按照所述第一人声数据，模拟所述第二音频数据中的第一人声数据，得到模拟的第一人声数据，并利用所述模拟的第一人声数据抵消所述第二音频数据中的第一人声数据；

5 将完成抵消的第二音频数据，作为过滤后的第二音频数据。

18、根据权利要求17所述的电子设备，所述处理器还被配置为执行：

在利用麦克风采集所述第一音频数据和第二人声数据，得到第二音频数据之后，将所述第二音频数据和所获取的第一人声数据分别输入自适应滤波器之前，对所获取的第一人声数据和所述第二音频数据进行相关性对比，得到所述第一人声数据和所述第二音频数据之间的第一延时；以及

10 将所述第二音频数据、所获取的第一人声数据和所述第一延时分别输入自适应滤波器，以使得所述自适应滤波器按照所述第一延时，对所述第一人声数据和所述第二音频数据进行对齐，得到对齐后的第一人声数据，按照所述对齐后的第一人声数据模拟所述第二音频数据中的第一人声数据，得到模拟的第一人声数据，并利用所述模拟的第一人声数据抵消
15 所述第二音频数据中的第一人声数据。

19、根据权利要求18所述的电子设备，所述处理器还被配置为执行：

在得到所述自适应滤波器输出的过滤后的第二音频数据之后，当所述第一客户端播放的背景音乐的来源为所述第一客户端本地时，将所述过滤后的第二音频数据发送给所述第一客户端，以使得所述第一客户端对所述第一客户端本地的背景音乐和所述过滤后的第二
20 音频数据进行对齐以及叠加，并播放叠加后的音频数据；

或者，当所述第一客户端播放的背景音乐的来源为所述第二客户端时，按照所述第一延时，对所述第二客户端本地的背景音乐和所述过滤后的第二音频数据进行对齐以及叠加，并将叠加后的音频数据发送给所述第一客户端，以使得所述第一客户端播放所述叠加后的音频数据。

25 20、一种电子设备，应用于第一客户端，所述电子设备包括：

处理器；

用于存储所述处理器可执行指令的存储器；

其中，所述处理器被配置为执行：

发送第一人声数据给第二客户端，以使得所述第二客户端利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；或者，发送第三音频数据给所述第二客户端，以使得所述第二客户端过滤所述第三音频数据中的背景音乐，
30 得到第一人声数据，并利用外放扬声器播放所述第一人声数据以及所述第二客户端本地的背景音乐，得到第一音频数据；其中，所述第一人声数据为所述第一客户端的用户的语音

数据；所述第三音频数据为所述第一客户端利用麦克风采集所述第一人声数据和所述第一客户端本地的背景音乐得到的音频数据；

接收所述第二客户端发送的第二音频数据；所述第二音频数据为所述第二客户端利用麦克风采集所述第一音频数据和第二人声数据，得到的音频数据；所述第二人声数据为所述
5 所述第二客户端的用户的声音数据；

过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据；

当所述第一客户端播放的背景音乐的来源为所述第二客户端时，播放所述过滤后的第二音频数据。

21、根据权利要求20所述的电子设备，所述处理器还被配置为执行：

在过滤所述第二音频数据中的第一人声数据，得到过滤后的第二音频数据之后，当所述
10 所述第一客户端播放的背景音乐的来源为所述第一客户端本地时，对所述第一客户端本地的背景音乐和所述过滤后的第二音频数据进行相关性对比，得到所述第一客户端本地的背景音乐和所述过滤后的第二音频数据之间的第二延时；

按照所述第二延时，对所述第一客户端本地的背景音乐和所述过滤后的第二音频数据
15 进行对齐，得到对齐后的第一客户端本地的背景音乐，叠加所述对齐后的第一客户端本地的背景音乐和所述过滤后的第二音频数据，得到叠加后的音频数据；

播放所述叠加后的音频数据。

22、一种非临时性计算机可读存储介质，包含于电子设备，当所述存储介质中的指令
20 由电子设备的处理器执行时，使得服务器能够执行权利要求1至7任一项所述的即时通讯的音质优化方法的步骤。

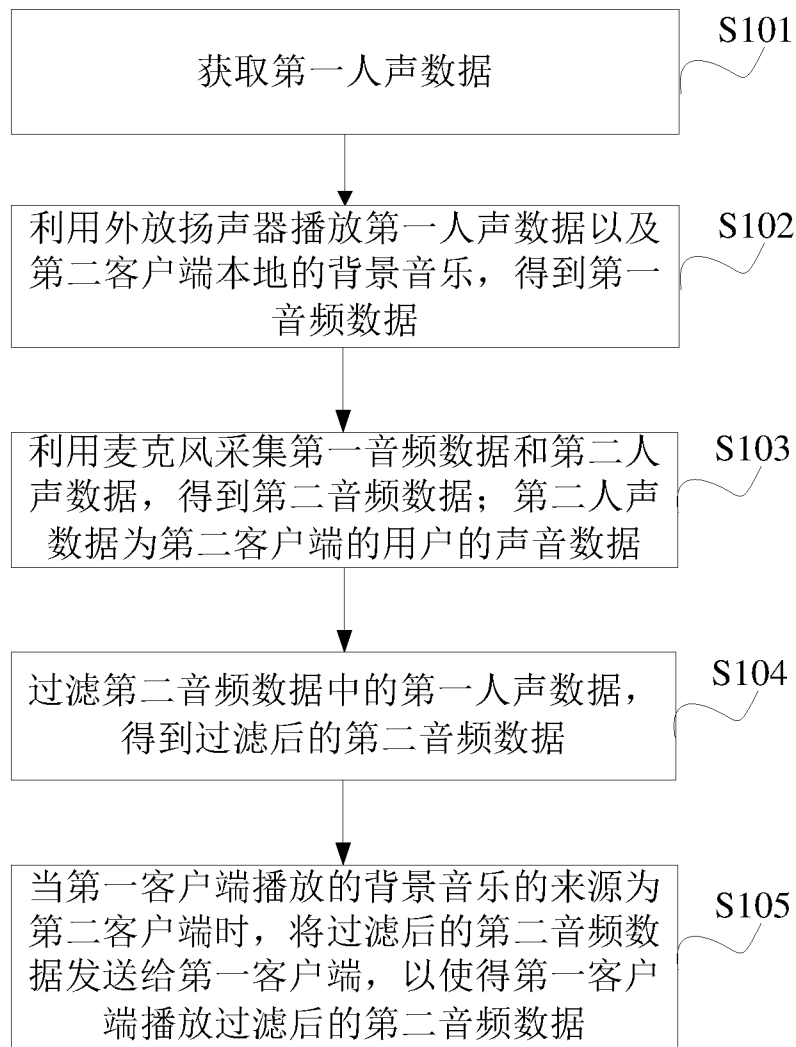


图1

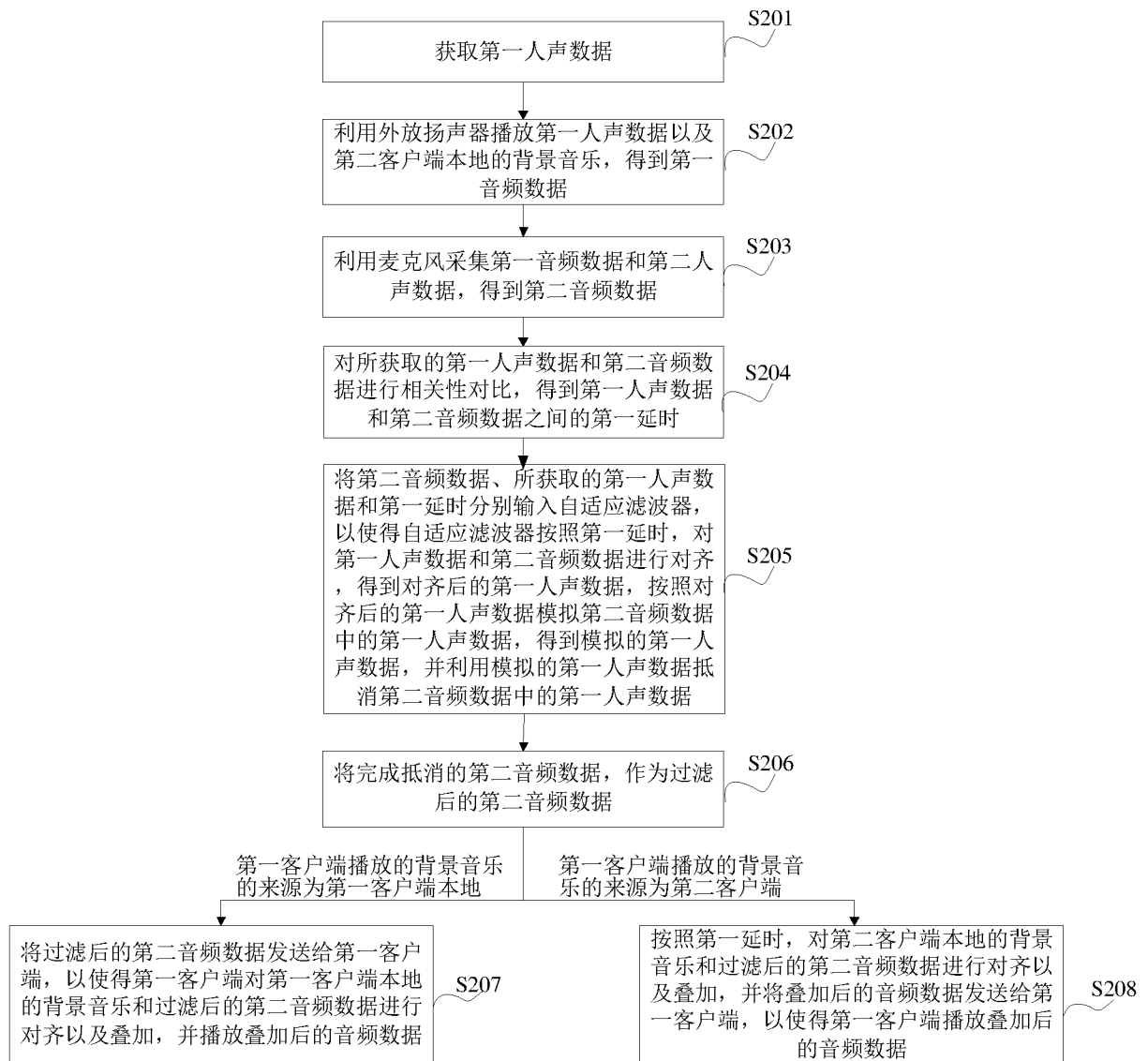


图2

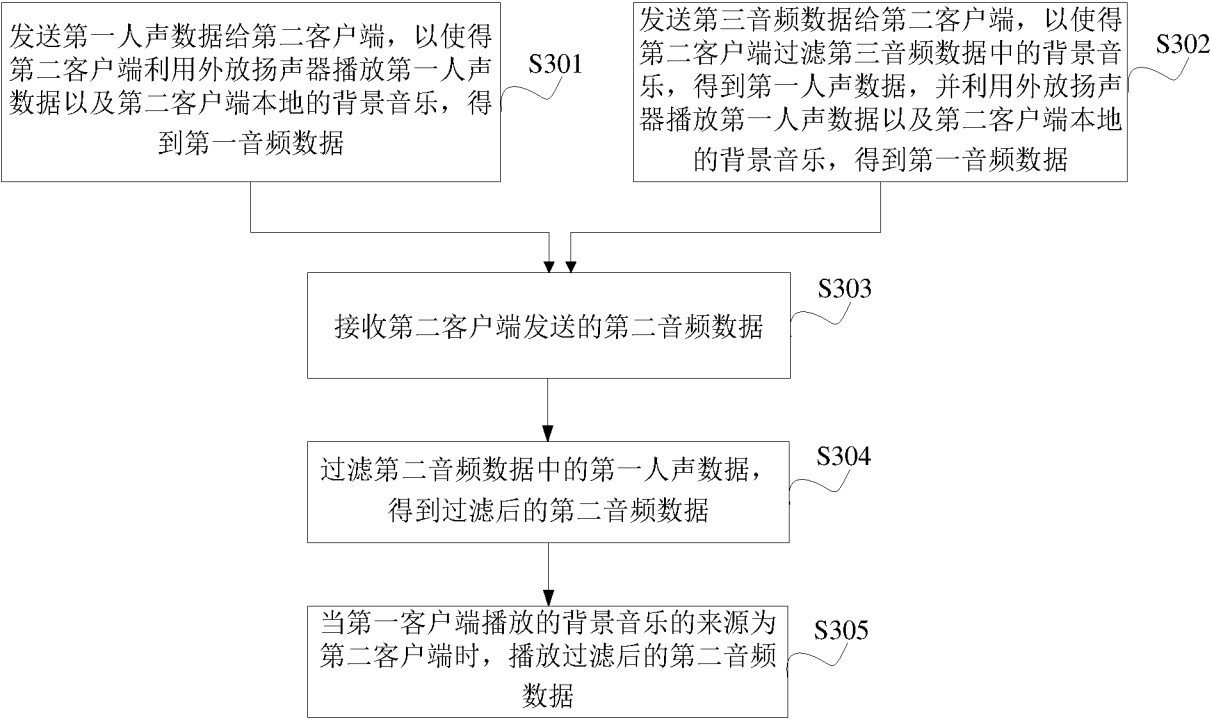


图3

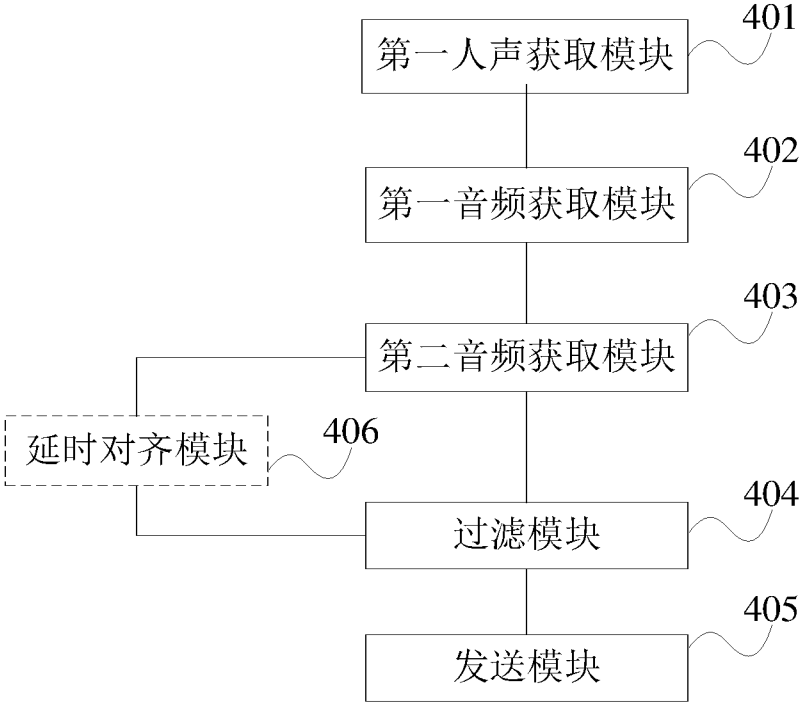


图4

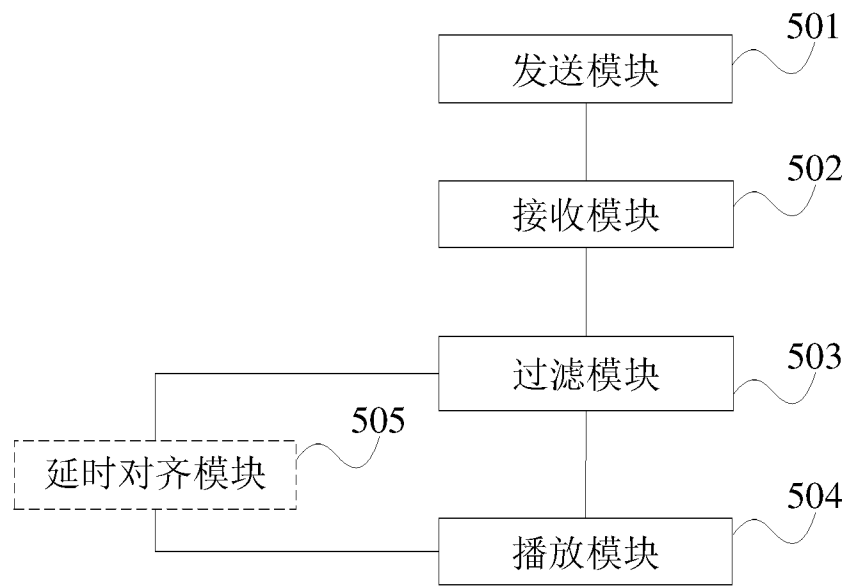


图5

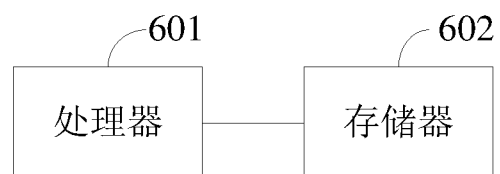


图6

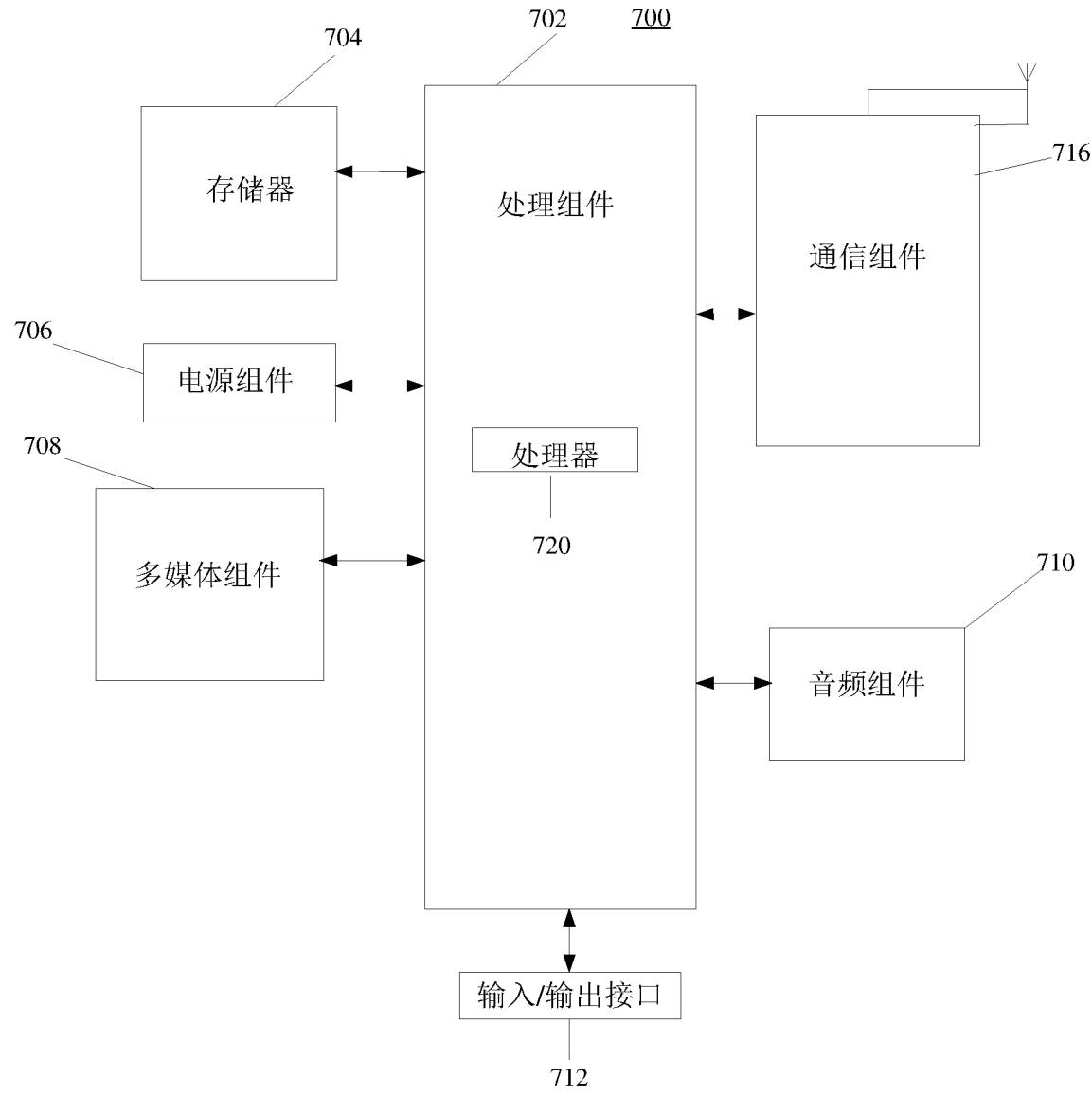


图7

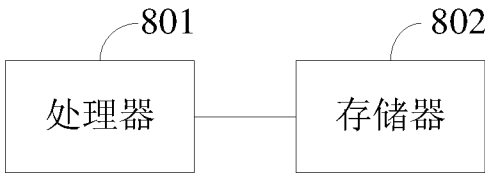


图8

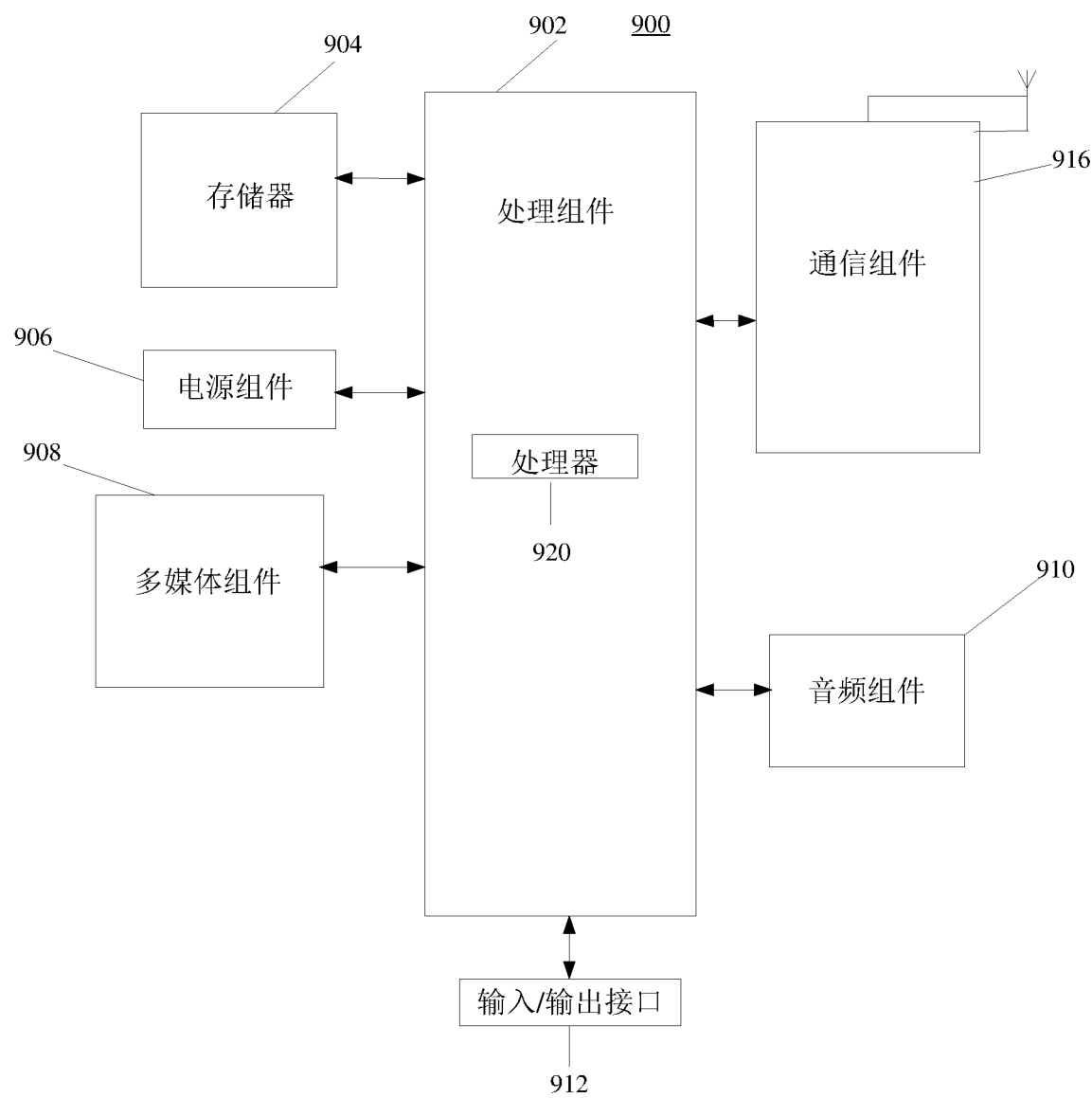


图9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/079072

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/58(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)

H04L

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)

CNPAT; CNKI; EPODOC; WPI; GOOGLE: 语音, 声音, 人声, 聊天, 背景音, 音乐, 过滤, 回声消除, 回音消除, 连麦, 直播, voice, speech, chat, background music, BGM, filter, echo cancell+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110138650 A (BEIJING DAJIA INTERCONNECTION INFORMATION TECHNOLOGY CO., LTD.) 16 August 2019 (2019-08-16) claims 1-10	1-22
A	CN 104464743 A (GUIYANG LONGMASTER INFORMATION & TECHNOLOGY CO., LTD.) 25 March 2015 (2015-03-25) description, paragraphs [0056]-[0060]	1-22
A	CN 109346098 A (WANGSU SCIENCE & TECHNOLOGY CO., LTD.) 15 February 2019 (2019-02-15) entire document	1-22
A	CN 103929531 A (LENOVO (BEIJING) LTD.) 16 July 2014 (2014-07-16) entire document	1-22
A	CN 107743178 A (VIVO COMMUNICATION TECHNOLOGY CO., LTD.) 27 February 2018 (2018-02-27) entire document	1-22
A	US 2014105411 A1 (SANTOS, Peter et al.) 17 April 2014 (2014-04-17) entire document	1-22



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

30 April 2020

Date of mailing of the international search report

27 May 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
Information on patent family members

International application No.

PCT/CN2020/079072

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110138650	A	16 August 2019	None			
CN	104464743	A	25 March 2015	None			
CN	109346098	A	15 February 2019	None			
CN	103929531	A	16 July 2014	None			
CN	107743178	A	27 February 2018	None			
US	2014105411	A1	17 April 2014	WO	2014062842	A1	24 April 2014
				CN	104170011	A	26 November 2014

国际检索报告

国际申请号

PCT/CN2020/079072

A. 主题的分类

H04L 12/58 (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)

H04L

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))

CNPAT; CNKI; EPDOC; WPI; GOOGLE: 语音, 声音, 人声, 聊天, 背景音, 音乐, 过滤, 回声消除, 回音消除, 连麦, 直播, voice, speech, chat, background music, BGM, filter, echo cancell+

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 110138650 A (北京达佳互联信息技术有限公司) 2019年 8月 16日 (2019 - 08 - 16) 权利要求1-10	1-22
A	CN 104464743 A (贵阳朗玛信息技术股份有限公司) 2015年 3月 25日 (2015 - 03 - 25) 说明书第[0056]-[0060]段	1-22
A	CN 109346098 A (网宿科技股份有限公司) 2019年 2月 15日 (2019 - 02 - 15) 全文	1-22
A	CN 103929531 A (联想北京有限公司) 2014年 7月 16日 (2014 - 07 - 16) 全文	1-22
A	CN 107743178 A (维沃移动通信有限公司) 2018年 2月 27日 (2018 - 02 - 27) 全文	1-22
A	US 2014105411 A1 (SANTOS, Peter等) 2014年 4月 17日 (2014 - 04 - 17) 全文	1-22

☐ 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 4月 30日

国际检索报告邮寄日期

2020年 5月 27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

万沙沙

电话号码 86- (10) -53961576

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/079072

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	110138650	A	2019年 8月 16日	无			
CN	104464743	A	2015年 3月 25日	无			
CN	109346098	A	2019年 2月 15日	无			
CN	103929531	A	2014年 7月 16日	无			
CN	107743178	A	2018年 2月 27日	无			
US	2014105411	A1	2014年 4月 17日	WO	2014062842	A1	2014年 4月 24日
				CN	104170011	A	2014年 11月 26日