



(12) 发明专利

(10) 授权公告号 CN 113038224 B

(45) 授权公告日 2023.03.24

(21) 申请号 202110331398.7

H04N 21/439 (2011.01)

(22) 申请日 2021.03.26

H04N 21/442 (2011.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 113038224 A

(56) 对比文件

CN 106411565 A, 2017.02.15

US 2021044867 A1, 2021.02.11

(43) 申请公布日 2021.06.25

CN 106550248 A, 2017.03.29

(73) 专利权人 北京小米移动软件有限公司

CN 107135413 A, 2017.09.05

地址 100085 北京市海淀区西二旗中路33

CN 107959874 A, 2018.04.24

号院6号楼8层018号

杨锐.《远程教学直播系统中媒体同步的研究与实现》.《计算机工程与设计》.2005,全文.

专利权人 北京小米松果电子有限公司

审查员 成聪

(72) 发明人 周岭松 李晓明

(74) 专利代理机构 北京名华博信知识产权代理

有限公司 11453

专利代理师 姜超

(51) Int. Cl.

H04N 21/43 (2011.01)

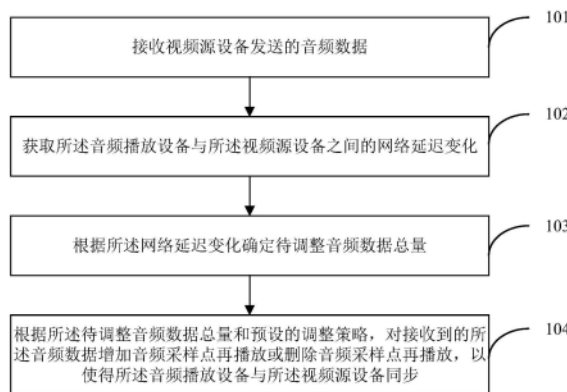
权利要求书3页 说明书11页 附图6页

(54) 发明名称

音画同步方法及装置

(57) 摘要

本公开是关于一种音画同步方法及装置。涉及智能物联网领域,解决了网络传输延迟导致的声音与画面不同步的问题,提升了用户体验。该方法包括:接收视频源设备发送的音频数据,音频数据包括多个音频采样点;获取音频播放设备与视频源设备之间的网络延迟变化,其中,网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差;根据网络延迟变化确定待调整音频数据总量;根据待调整音频数据总量和预设的调整策略,对接收到的音频数据增加音频采样点再播放或删除音频采样点再播放,以使得音频播放设备与视频源设备同步。本公开提供的技术方案适用于智能家居环境中,实现了适应性强的音画同步方案。



1. 一种音画同步方法,其特征在于,应用于音频播放设备,所述音频播放设备与视频源设备通过网络连接,所述方法包括:

接收所述视频源设备发送的音频数据,所述音频数据包括多个音频采样点;

获取所述音频播放设备与所述视频源设备之间的网络延迟变化,其中,所述网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差;

根据所述网络延迟变化确定待调整音频数据总量;

根据所述待调整音频数据总量和预设的调整策略,对接收到的所述音频数据增加音频采样点再播放或删除音频采样点再播放,以使得所述音频播放设备与所述视频源设备同步;

其中,所述调整策略包括:对接收到的所述音频数据在每隔单位时间增加或删除预设数量的音频采样点,直至调整的音频采样点的总量达到所述待调整音频数据总量。

2. 根据权利要求1所述的音画同步方法,其特征在于,采用插值法对单位时间内的音频数据增加或删除预设数量的音频采样点。

3. 根据权利要求1-2任一项所述的音画同步方法,其特征在于,所述获取所述音频播放设备与所述视频源设备之间的网络延迟变化,包括:

在预设的网络延迟监测时间到来时,获取当前时间对应的第一网络延迟时间;

根据所述第一网络延迟时间和上一次获取的第二网络延迟时间,得到所述网络延迟变化。

4. 根据权利要求3所述的音画同步方法,其特征在于,所述在预设的网络延迟监测时间到来时,获取当前时间对应的第一网络延迟时间,包括:

在预设的网络延迟监测时间到来时,

向所述视频源设备发送第一时钟信息,所述第一时钟信息包含第一发送时间,所述第一发送时间用于表示所述音频播放设备发送所述第一时钟信息的时间;

接收所述视频源设备基于所述第一时钟信息所反馈的第二时钟信息,其中,所述第二时钟信息至少包括:第一接收时间和第二发送时间,其中,第一接收时间用于表示所述视频源设备接收到所述第一时钟信息的时间,所述第二发送时间用于表示所述视频源设备发送所述第二时钟信息的时间;

根据所述第一时钟信息和所述第二时钟信息,确定所述第一网络延迟时间。

5. 根据权利要求4所述的音画同步方法,其特征在于,所述根据所述第一时钟信息和所述第二时钟信息,确定所述第一网络延迟时间,包括:根据以下算式确定所述第一网络延迟时间:

第一网络延迟时间 = [(第一接收时间 - 第一发送时间) + (第二接收时间 - 第二发送时间)] / 2。

6. 根据权利要求3所述的音画同步方法,其特征在于,所述视频源设备发送的所述音频数据以报文包进行发送,每个报文包包括至少一个音频帧和所述音频帧对应的划分时间,每个音频帧包括预定数量的音频采样点;其中,所述划分时间指的是将所述音频数据划分成所述音频帧这一划分动作的发生时间。

7. 根据权利要求6所述的音画同步方法,其特征在于,所述方法还包括:

在接收到每个音频帧后,根据每个音频帧的划分时间、预设的与所述视频源设备的约

定时间以及所述第一网络延迟时间,确定播放所述音频帧的时间。

8. 一种音画同步装置,其特征在于,包括:

音频数据接收模块,用于接收视频源设备发送的音频数据,所述音频数据包括多个音频采样点;

延迟变化获取模块,用于获取音频播放设备与所述视频源设备之间的网络延迟变化,其中,所述网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差,所述音频播放设备与视频源设备通过网络连接;

调整音频数据量确定模块,用于根据所述网络延迟变化确定待调整音频数据总量;

调整同步播放模块,用于根据所述待调整音频数据总量和预设的调整策略,对接收到的所述音频数据增加音频采样点再播放或删除音频采样点再播放,以使得所述音频播放设备与所述视频源设备同步;

其中,所述调整策略包括:对接收到的所述音频数据在每隔单位时间增加或删除预设数量的音频采样点,直至调整的音频采样点的总量达到所述待调整音频数据总量。

9. 根据权利要求8所述的音画同步装置,其特征在于,

所述调整同步播放模块包括:

插值处理子模块,用于采用插值法对单位时间内的音频数据增加或删除预设数量的音频采样点。

10. 根据权利要求8或9所述的音画同步装置,其特征在于,所述延迟变化获取模块包括:

第一网络延迟时间获取子模块,用于在预设的网络延迟监测时间到来时,获取当前时间对应的第一网络延迟时间;

变化获取子模块,用于根据所述第一网络延迟时间和上一次获取的第二网络延迟时间,得到所述网络延迟变化。

11. 根据权利要求10所述的音画同步装置,其特征在于,所述第一网络延迟时间获取子模块包括:

第一时钟信息发送单元,用于在预设的网络延迟监测时间到来时,

向所述视频源设备发送第一时钟信息,所述第一时钟信息包含第一发送时间,所述第一发送时间用于表示所述音频播放设备发送所述第一时钟信息的时间;

第二时钟信息接收单元,用于接收所述视频源设备基于所述第一时钟信息所反馈的第二时钟信息,其中,所述第二时钟信息至少包括:第一接收时间和第二发送时间,其中,第一接收时间用于表示所述视频源设备接收到所述第一时钟信息的时间,所述第二发送时间用于表示所述视频源设备发送所述第二时钟信息的时间;

第一网络延迟时间计算单元,用于根据所述第一时钟信息和所述第二时钟信息,确定所述第一网络延迟时间。

12. 根据权利要求11所述的音画同步装置,其特征在于,

所述第一网络延迟时间计算单元,具体用于根据以下算式确定所述第一网络延迟时间:

第一网络延迟时间 = [(第一接收时间 - 第一发送时间) + (第二接收时间 - 第二发送时间)] / 2。

13. 根据权利要求10所述的音画同步装置,其特征在于,所述视频源设备发送的所述音频数据以报文包进行发送,每个报文包包括至少一个音频帧和所述音频帧对应的划分时间,每个音频帧包括预定数量的音频采样点,其中,所述划分时间指的是将所述音频数据划分成所述音频帧这一划分动作的发生时间,

所述装置还包括:

音频帧播放管理模块,用于在接收到每个音频帧后,根据每个音频帧的划分时间、预设的与所述视频源设备的约定时间以及所述第一网络延迟时间,确定播放所述音频帧的时间。

14. 一种计算机装置,其特征在于,包括:

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:

接收视频源设备发送的音频数据,所述音频数据包括多个音频采样点;

获取音频播放设备与所述视频源设备之间的网络延迟变化,其中,所述网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差;

根据所述网络延迟变化确定待调整音频数据总量;

根据所述待调整音频数据总量和预设的调整策略,对接收到的所述音频数据增加音频采样点再播放或删除音频采样点再播放,以使得所述音频播放设备与所述视频源设备同步;

其中,所述调整策略包括:对接收到的所述音频数据在每隔单位时间增加或删除预设数量的音频采样点,直至调整的音频采样点的总量达到所述待调整音频数据总量。

15. 一种非临时性计算机可读存储介质,当所述存储介质中的指令由移动终端的处理器执行时,使得移动终端能够执行一种音画同步方法,所述方法包括:

接收视频源设备发送的音频数据,所述音频数据包括多个音频采样点;

获取音频播放设备与所述视频源设备之间的网络延迟变化,其中,所述网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差;

根据所述网络延迟变化确定待调整音频数据总量;

根据所述待调整音频数据总量和预设的调整策略,对接收到的所述音频数据增加音频采样点再播放或删除音频采样点再播放,以使得所述音频播放设备与所述视频源设备同步;

其中,所述调整策略包括:对接收到的所述音频数据在每隔单位时间增加或删除预设数量的音频采样点,直至调整的音频采样点的总量达到所述待调整音频数据总量。

音画同步方法及装置

技术领域

[0001] 本公开涉及智能物联网领域,尤其涉及一种音画同步方法及装置。

背景技术

[0002] 随着人工智能物联网(AIoT)的发展,跨设备互联同步音乐播放已经在多个产品上线,比如组合立体声功能已经在小爱音箱、amazon echo等平台上线。电视与音箱组合,能为用户提供更好的观影体验。

[0003] 电视和音箱可通过WiFi连接,组成环绕立体声播放系统。在播放视频时,电视播放图像画面的同时,相应的声音则需要由电视和音箱播放出来。但由于音箱播放的音频数据需要由电视经过WiFi等网络传输给音箱,传输延迟造成了音箱播放的声音与电视屏幕播放的图像画面不同步,影响了用户体验。

发明内容

[0004] 为克服相关技术中存在的问题,本公开提供一种音画同步方法及装置。

[0005] 根据本公开实施例的第一方面,提供一种音画同步方法,应用于音频播放设备,音频播放设备与视频源设备通过网络连接,该方法包括:

[0006] 接收视频源设备发送的音频数据,音频数据包括多个音频采样点;

[0007] 获取音频播放设备与视频源设备之间的网络延迟变化,其中,网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差;

[0008] 根据网络延迟变化确定待调整音频数据总量;

[0009] 根据待调整音频数据总量和预设的调整策略,对接收到的音频数据增加音频采样点再播放或删除音频采样点再播放,以使得音频播放设备与视频源设备同步。

[0010] 优选的,调整策略包括:对接收到的音频数据在每隔单位时间增加或删除预设数量的音频采样点,直至调整的音频采样点的总量达到待调整音频数据总量。

[0011] 优选的,采用插值法对单位时间内的音频数据增加或删除预设数量的音频采样点。

[0012] 优选的,获取音频播放设备与视频源设备之间的网络延迟变化,包括:

[0013] 在预设的网络延迟监测时间到来时,获取当前时间对应的第一网络延迟时间;

[0014] 根据第一网络延迟时间和上一次获取的第二网络延迟时间,得到网络延迟变化。

[0015] 优选的,在预设的网络延迟监测时间到来时,获取当前时间对应的第一网络延迟时间,包括:

[0016] 在预设的网络延迟监测时间到来时,

[0017] 向视频源设备发送第一时钟信息,第一时钟信息包含第一发送时间,第一发送时间用于表示音频播放设备发送第一时钟信息的时间;

[0018] 接收视频源设备基于第一时钟信息所反馈的第二时钟信息,其中,第二时钟信息至少包括:第一接收时间和第二发送时间,其中,第一接收时间用于表示视频源设备接收到

第一时钟信息的时间,第二发送时间用于表示视频源设备发送第二时钟信息的时间;

[0019] 根据第一时钟信息和第二时钟信息,确定第一网络延迟时间。

[0020] 优选的,根据第一时钟信息和第二时钟信息,确定第一网络延迟时间,包括:根据以下算式确定第一网络延迟时间:

[0021] 第一网络延迟时间 = $[(\text{第一接收时间} - \text{第一发送时间}) + (\text{第二接收时间} - \text{第二发送时间})] / 2$ 。

[0022] 优选的,视频源设备发送的音频数据以报文包进行发送,每个报文包包括至少一个音频帧和音频帧对应的划分时间,每个音频帧包括预定数量的音频采样点;其中,划分时间指的是将音频数据划分成音频帧这一划分动作的发生时间。

[0023] 优选的,该方法还包括:

[0024] 在接收到每个音频帧后,根据每个音频帧的划分时间、预设的与视频源设备的约定时间以及第一网络延迟时间,确定播放音频帧的时间。

[0025] 根据本公开实施例的第二方面,提供一种音画同步装置,包括:

[0026] 音频数据接收模块,用于接收视频源设备发送的音频数据,音频数据包括多个音频采样点;

[0027] 延迟变化获取模块,用于获取音频播放设备与视频源设备之间的网络延迟变化,其中,网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差,音频播放设备与视频源设备通过网络连接;

[0028] 调整音频数据量确定模块,用于根据网络延迟变化确定待调整音频数据总量;

[0029] 调整同步播放模块,用于根据待调整音频数据总量和预设的调整策略,对接收到的音频数据增加音频采样点再播放或删除音频采样点再播放,以使得音频播放设备与视频源设备同步。

[0030] 优选的,调整策略包括:对接收到的音频数据在每隔单位时间增加或删除预设数量的音频采样点,直至调整的音频采样点的总量达到待调整音频数据总量,

[0031] 调整同步播放模块包括:

[0032] 插值处理子模块,用于采用插值法对单位时间内的音频数据增加或删除预设数量的音频采样点。

[0033] 优选的,延迟变化获取模块包括:

[0034] 第一网络延迟时间获取子模块,用于在预设的网络延迟监测时间到来时,获取当前时间对应的第一网络延迟时间;

[0035] 变化获取子模块,用于根据第一网络延迟时间和上一次获取的第二网络延迟时间,得到网络延迟变化。

[0036] 优选的,第一网络延迟时间获取子模块包括:

[0037] 第一时钟信息发送单元,用于在预设的网络延迟监测时间到来时,

[0038] 向视频源设备发送第一时钟信息,第一时钟信息包含第一发送时间,第一发送时间用于表示音频播放设备发送第一时钟信息的时间;

[0039] 第二时钟信息接收单元,用于接收视频源设备基于第一时钟信息所反馈的第二时钟信息,其中,第二时钟信息至少包括:第一接收时间和第二发送时间,其中,第一接收时间用于表示视频源设备接收到第一时钟信息的时间,第二发送时间用于表示视频源设备发送

第二时钟信息的时间；

[0040] 第一网络延迟时间计算单元,用于根据第一时钟信息和第二时钟信息,确定第一网络延迟时间。

[0041] 优选的,第一网络延迟时间计算单元,具体用于根据以下算式确定第一网络延迟时间:

[0042] 第一网络延迟时间 = [(第一接收时间 - 第一发送时间) + (第二接收时间 - 第二发送时间)] / 2。

[0043] 优选的,视频源设备发送的音频数据以报文包进行发送,每个报文包包括至少一个音频帧和音频帧对应的划分时间,每个音频帧包括预定数量的音频采样点,其中,划分时间指的是将音频数据划分成音频帧这一划分动作的发生时间,

[0044] 优选的,该装置还包括:

[0045] 音频帧播放管理模块,用于在接收到每个音频帧后,根据每个音频帧的划分时间、预设的与视频源设备的约定时间以及第一网络延迟时间,确定播放音频帧的时间。

[0046] 根据本公开实施例的第三方面,提供一种计算机装置,包括:

[0047] 处理器;

[0048] 用于存储处理器可执行指令的存储器;

[0049] 其中,处理器被配置为:

[0050] 接收视频源设备发送的音频数据,音频数据包括多个音频采样点;

[0051] 获取音频播放设备与视频源设备之间的网络延迟变化,其中,网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差;

[0052] 根据网络延迟变化确定待调整音频数据总量;

[0053] 根据待调整音频数据总量和预设的调整策略,对接收到的音频数据增加音频采样点再播放或删除音频采样点再播放,以使得音频播放设备与视频源设备同步。

[0054] 根据本公开实施例的第四方面,提供一种非临时性计算机可读存储介质,当存储介质中的指令由移动终端的处理器执行时,使得移动终端能够执行一种音画同步方法,方法包括:

[0055] 接收视频源设备发送的音频数据,音频数据包括多个音频采样点;

[0056] 获取音频播放设备与视频源设备之间的网络延迟变化,其中,网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差;

[0057] 根据网络延迟变化确定待调整音频数据总量;

[0058] 根据待调整音频数据总量和预设的调整策略,对接收到的音频数据增加音频采样点再播放或删除音频采样点再播放,以使得音频播放设备与视频源设备同步。

[0059] 本公开的实施例提供的技术方案可以包括以下有益效果:考虑到网络传输造成的延时,对于通过网络连接的音频播放设备和视频源设备,首先接收视频源设备发送的包括多个音频采样点的音频数据,然后获取音频播放设备与视频源设备之间的网络延迟变化,再根据网络延迟变化确定待调整音频数据总量,最后根据待调整音频数据总量和预设的调整策略,对接收到的音频数据增加音频采样点再播放或删除音频采样点再播放,以使得音频播放设备与视频源设备同步。解决了网络传输延迟导致的声音与画面不同步的问题,提升了用户体验。

[0060] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

附图说明

[0061] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本发明的实施例,并与说明书一起用于解释本发明的原理。

[0062] 图1是根据一示例性实施例示出的一种音画同步方法的流程图。

[0063] 图2是根据一示例性实施例示出的样条插值增加采样点的原理示意图。

[0064] 图3是根据一示例性实施例示出的又一种音画同步方法的流程图。

[0065] 图4根据一示例性实施例示出的又一种音画同步方法的流程图。

[0066] 图5是根据一示例性实施例示出的音频播放设备与视频源设备进行时钟信息交互的时间戳信息示意图。

[0067] 图6是根据一示例性实施例示出的一种音画同步装置的框图。

[0068] 图7是根据一示例性实施例示出的调整同步播放模块604的结构示意图。

[0069] 图8是根据一示例性实施例示出的延迟变化获取模块602的结构示意图。

[0070] 图9是根据一示例性实施例示出的第一网络延迟时间获取子模块801的结构示意图。

[0071] 图10是根据一示例性实施例示出的一种音画同步装置的框图。

[0072] 图11是根据一示例性实施例示出的一种装置的框图。

具体实施方式

[0073] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本发明相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本发明的一些方面相一致的装置和方法的例子。

[0074] 由于音箱播放的音频数据需要由电视经过WiFi等网络传输给音箱,传输延迟造成了音箱播放的声音与电视屏幕播放的图像画面不同步,影响了用户体验。

[0075] 为了解决上述问题,本公开的一示例性实施例提供了一种音画同步方法。使用该方法消除网络传输延迟的影响,实现音画同步的流程如图1所示,包括:

[0076] 步骤101、接收视频源设备发送的音频数据。

[0077] 本公开中,所述音频播放设备与视频源设备通过网络连接,音频播放设备接收视频源设备发送的音频数据。其中,音频播放设备可通过有线或无线通信方式与所述视频源设备进行数据交互。其中,无线通信方式可以是WiFi、蓝牙等。

[0078] 视频源设备可为电视、手机、平板电脑、笔记本电脑等具有网络通信能力且能够播放视频数据的终端,本领域技术人员应该知晓,本公开所涉及的视频源设备不限于以上所列举的范围。

[0079] 音频播放设备可为音箱或其他具有声音播放模块的电子设备。

[0080] 音频数据包括多个音频采样点。

[0081] 视频源设备发送的音频数据以报文包进行发送,每个报文包包括至少一个音频帧

和音频帧对应的划分时间,每个音频帧包括预定数量的音频采样点;其中,划分时间指的是将音频数据划分成音频帧这一划分动作的发生时间。

[0082] 在接收到每个音频帧后,根据每个音频帧的划分时间、预设的与视频源设备的约定时间以及第一网络延迟时间,确定播放音频帧的时间。

[0083] 步骤102、获取音频播放设备与视频源设备之间的网络延迟变化。

[0084] 本步骤中,网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差。

[0085] 本步骤中,考虑到网络传输会造成延时,因此,获取音频与视频源设备之间的网络延迟。可预设网络延迟监测时间,在该时间到来时对网络延迟进行监测。记录最近一次或多次监测获取的网络延迟时间。

[0086] 本公开实施例所提供的方法,适用于存在一个或多个音频播放设备的场景。

[0087] 对于多个不同的音频播放设备,各音频播放设备可分别对与视频源设备之间的网络延迟进行监测。

[0088] 步骤103、根据网络延迟变化确定待调整音频数据总量。

[0089] 本步骤中,根据网络延迟的变化确定待调整音频数据总量,其中,根据网络延迟变化的时间长度,确定需要调整的音频数据总量。

[0090] 步骤104、根据待调整音频数据总量和预设的调整策略,对接收到的音频数据增加音频采样点再播放或删除音频采样点再播放,以使得音频播放设备与视频源设备同步。

[0091] 本步骤中,对接收到的音频数据在每隔单位时间增加或删除预设数量的音频采样点,直至调整的音频采样点的总量达到待调整音频数据总量。

[0092] 同步可以是音频播放设备播放的音频数据与视频源设备播放的音频数据同步;也可以是音频播放设备播放的音频数据与视频源设备播放的视频数据同步;还可以是音频播放设备播放的音频数据与视频源设备播放的音频数据和视频数据均同步。

[0093] 其中,可采用插值法对单位时间内的音频数据增加或删除预设数量的音频采样点。在网络延迟变长的情况下,在调整时间区间内均匀进行样条插值,添加占据时长总和为变化幅度的至少一个采样点。在网络延迟变短的情况下,在调整时间区间内均匀进行样条插值,删除占据时长总和为变化幅度的至少一个采样点。

[0094] 声音播放依据网络状态自适应延迟时,会遇到声音数据补充和删除的问题。比如随着网络变好,网络延迟变化超过10ms,由30ms变为20ms,那么需要音频快进10ms。如果直接丢弃10ms的数据会引起用户可感知的声音卡顿,导致不好的主观听觉体验。因而需要将10ms数据提前这一操作,在一个较长的调整时间区间内小幅逐渐完成。例如,需要音频数据播放时间向后加速10ms的情况下,将调整时间区间定为10s。这样,就可以在确定需要调整后的10s内逐渐完成播放时间调整,降低对用户听觉感知的影响。在网络延迟变化后需要音频数据播放提前的情况下,在10s时间内,每1s(包含48000采样点)提前1ms(48采样点)。在1s内丢弃1ms,可以每个1000个采样点丢弃一个,这样不会引入明显的失真,也可以通过插值方法(如三次样条插值),将48000个采样点插值成47952个采样点。通过渐进式的方法将音频数据的播放提前而又不引起用户的感知。

[0095] 随着网络变差,网络延迟变化超过10ms的调整门限,由20ms变为30ms,那么需要音频慢放10ms,则可以采样相反的插值操作,即将47952个采样点插值成48000个采样点,渐进

式的补充数据把延迟拉大。图2所示为将19点的数据插值为20点的示例,反之可以将20个点插值为19个采样点。

[0096] 本公开的一示例性实施例还提供了一种音画同步方法,使用该方法,可根据第一网络延迟时间确定网络延迟变化,具体流程如图3所示,包括:

[0097] 步骤301、在预设的网络延迟监测时间到来时,获取当前时间对应的第一网络延迟时间。

[0098] 本步骤可通过在音频播放设备和视频源设备之间的时钟信息交互来获取第一网络延迟时间。时钟信息的交互是实时进行的,可按照固定时长间隔(例如1s)进行周期性的交互操作。

[0099] 具体如图4所示,包括:

[0100] 步骤401、在预设的网络延迟监测时间到来时,向视频源设备发送第一时钟信息。

[0101] 第一时钟信息包含第一发送时间,第一发送时间用于表示音频播放设备发送第一时钟信息的时间。

[0102] 步骤402、接收视频源设备基于第一时钟信息所反馈的第二时钟信息。

[0103] 本步骤中,音频播放设备接收视频源设备反馈的第二时钟信息。

[0104] 其中,第二时钟信息至少包括:第一接收时间和第二发送时间,其中,第一接收时间用于表示视频源设备接收到第一时钟信息的时间,第二发送时间用于表示视频源设备发送第二时钟信息的时间。

[0105] 在本公开另一个实施方式中,第二时钟信息还可携带第一发送时间。

[0106] 图5所示为音频播放设备与视频源设备进行时钟信息交互的原理示意。音频播放设备在第一发送时间发送第一时钟信息,在第二接收时间接收第二时钟信息;视频源设备在第一接收时间接收第一时钟信息,在第二发送时间发送第二时钟信息。

[0107] 步骤403、根据第一时钟信息和第二时钟信息,确定第一网络延迟时间。

[0108] 其中,可根据以下算式确定第一网络延迟时间:

[0109] 第一网络延迟时间 = $[(\text{第一接收时间} - \text{第一发送时间}) + (\text{第二接收时间} - \text{第二发送时间})] / 2$ 。

[0110] 步骤302、根据第一网络延迟时间和上一次获取的第二网络延迟时间,得到网络延迟变化。

[0111] 本步骤中,根据两次对网络延迟的监测结果,得到网络延迟变化。在确定网络延迟变化后,即可根据网络延迟变化对接收到的音频数据的播放进行调整。

[0112] 本公开的一示例性实施例还提供了一种音画同步方法,音频数据一般由视频源设备发送。视频源设备发送的音频数据以报文包进行发送,每个报文包包括至少一个音频帧和音频帧对应的划分时间,每个音频帧包括预定数量的音频采样点;其中,划分时间指的是将音频数据划分成音频帧这一划分动作的发生时间。例如,把划分后得到的每个音频帧的划分时间信息与音频帧数据一起打包成音频报文包;为了降低网络负载,还可以对音频数据进行编码压缩后再打包,从而减少需要传输的数据量。

[0113] 在接收到每个音频帧后,音频播放设备可根据每个音频帧的划分时间、预设的与视频源设备的约定时间以及第一网络延迟时间,确定播放音频帧的时间。播放音频帧的时间可相对于约定时间延迟至少第一网络延迟时间的长度。音频源设备对播放的视频数据进

行解码,得到图像数据和音频数据。可对音频数据进行处理后再向音频播放设备发送。

[0114] 音频帧的延迟播放,可保证播放音频数据时,能够消除网络传输延迟时间的影响。

[0115] 本公开的一示例性实施例还提供了一种音画同步方法,使用该方法可在音频播放设备与视频源设备之间进行系统时间校准,确定音频播放设备相对于视频源设备的系统时间差。

[0116] 可依赖于时钟信息交互过程确定系统时间差。其中,可根据以下表达式计算系统时间差:

[0117] 系统时间差 = [(第一接收时间 - 第一发送时间) - (第二接收时间 - 第二发送时间)]/2。

[0118] 在图5示例中,

[0119] $T_2 - T_1 = \Delta + \tau$

[0120] $T_4 - T_3 = -\Delta + \tau$

[0121] 其中,其中 τ 为视频源设备向音频播放设备发送数据时的第一网络延迟时间, Δ 为音频播放设备相对于视频源设备的系统时间差。

[0122] 第一网络延迟时间可设置为大于等于 $\Delta + \tau$ 。为了提供更好的音画同步体验,通过监测网络的状态来动态设置第一网络延迟时间,网络状态可以通过上述公式中的 τ 来映射。比如 τ 需要消耗10ms时间,考虑到中间运算消耗,可设置播放延迟时间为 $\Delta + 15\text{ms}$ 。实际体验表明,小于200ms都能提供较好的音画同步体验。

[0123] 例如,设置播放延迟时间为 $\Delta + 1\text{s}$,1s大于 τ 。对于视频文件的 T_0 时刻,若视频源设备和所有音频播放设备都在1s后播出。在视频源设备上,只需要根据自己的系统时钟在 $T_0 + 1\text{s}$ 时刻将音频对应的图像和/或音频播出;但在音频播放设备上,则需要考虑系统时间差,在 $T_0 + \Delta + 1\text{s}$ 时刻播放出音频。

[0124] 本公开的一示例性实施例还提供了一种音画同步装置,该装置的结构如图6所示,包括:

[0125] 音频数据接收模块601,用于接收视频源设备发送的音频数据,音频数据包括多个音频采样点;

[0126] 延迟变化获取模块602,用于获取音频播放设备与视频源设备之间的网络延迟变化,其中,网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差,音频播放设备与视频源设备通过网络连接;

[0127] 调整音频数据量确定模块603,用于根据网络延迟变化确定待调整音频数据总量;

[0128] 调整同步播放模块604,用于根据待调整音频数据总量和预设的调整策略,对接收到的音频数据增加音频采样点再播放或删除音频采样点再播放,以使得音频播放设备与视频源设备同步。

[0129] 优选的,调整策略包括:对接收到的音频数据在每隔单位时间增加或删除预设数量的音频采样点,直至调整的音频采样点的总量达到待调整音频数据总量,

[0130] 调整同步播放模块604的结构如图7所示,包括:

[0131] 插值处理子模块701,用于采用插值法对单位时间内的音频数据增加或删除预设数量的音频采样点。

[0132] 优选的,延迟变化获取模块602的结构如图8所示,包括:

[0133] 第一网络延迟时间获取子模块801,用于在预设的网络延迟监测时间到来时,获取当前时间对应的第一网络延迟时间;

[0134] 变化获取子模块802,用于根据第一网络延迟时间和上一次获取的第二网络延迟时间,得到网络延迟变化。

[0135] 优选的,第一网络延迟时间获取子模块801的结构如图9所示,包括:

[0136] 第一时钟信息发送单元901,用于在预设的网络延迟监测时间到来时,

[0137] 向视频源设备发送第一时钟信息,第一时钟信息包含第一发送时间,第一发送时间用于表示音频播放设备发送第一时钟信息的时间;

[0138] 第二时钟信息接收单元902,用于接收视频源设备基于第一时钟信息所反馈的第二时钟信息,其中,第二时钟信息至少包括:第一接收时间和第二发送时间,其中,第一接收时间用于表示视频源设备接收到第一时钟信息的时间,第二发送时间用于表示视频源设备发送第二时钟信息的时间;

[0139] 第一网络延迟时间计算单元903,用于根据第一时钟信息和第二时钟信息,确定第一网络延迟时间。

[0140] 优选的,第一网络延迟时间计算单元903,具体用于根据以下算式确定第一网络延迟时间:

[0141] 第一网络延迟时间 = $[(\text{第一接收时间} - \text{第一发送时间}) + (\text{第二接收时间} - \text{第二发送时间})] / 2$ 。

[0142] 优选的,视频源设备发送的音频数据以报文包进行发送,每个报文包包括至少一个音频帧和音频帧对应的划分时间,每个音频帧包括预定数量的音频采样点,其中,划分时间指的是将音频数据划分成音频帧这一划分动作的发生时间,

[0143] 装置如图10所示,还包括:

[0144] 音频帧播放管理模块605,用于在接收到每个音频帧后,根据每个音频帧的划分时间、预设的与视频源设备的约定时间以及第一网络延迟时间,确定播放音频帧的时间。

[0145] 上述装置可集成于音频播放设备之中,由设备实现相应功能。关于上述实施例中的装置,其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述,此处将不做详细阐述说明。

[0146] 图11是根据一示例性实施例示出的一种用于声音播放的装置1100的框图。例如,装置1100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,医疗设备,健身设备,个人数字助理等。

[0147] 参照图11,装置1100可以包括以下一个或多个组件:处理组件1102,存储器1104,电力组件1106,多媒体组件1108,音频组件1110,输入/输出(I/O)的接口1112,传感器组件1114,以及通信组件1116。

[0148] 处理组件1102通常控制装置1100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1102可以包括一个或多个处理器1120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1102可以包括一个或多个模块,便于处理组件1102和其他组件之间的交互。例如,处理组件1102可以包括多媒体模块,以方便多媒体组件1108和处理组件1102之间的交互。

[0149] 存储器1104被配置为存储各种类型的数据以支持在设备1100的操作。这些数据的

示例包括用于在装置1100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1104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,如静态随机存取存储器 (SRAM),电可擦除可编程只读存储器 (EEPROM),可擦除可编程只读存储器 (EPROM),可编程只读存储器 (PROM),只读存储器 (ROM),磁存储器,快闪存储器,磁盘或光盘。

[0150] 电力组件1106为装置1100的各种组件提供电力。电力组件1106可以包括电源管理系统,一个或多个电源,及其他与为装置1100生成、管理和分配电力相关联的组件。

[0151] 多媒体组件1108包括在装置1100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。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1108包括一个前置摄像头和/或后置摄像头。当设备1100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。

[0152] 音频组件1110被配置为输出和/或输入音频信号。例如,音频组件1110包括一个麦克风 (MIC),当装置1100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1104或经由通信组件1116发送。在一些实施例中,音频组件1110还包括一个扬声器,用于输出音频信号。

[0153] I/O接口1112为处理组件1102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。

[0154] 传感器组件1114包括一个或多个传感器,用于为装置1100提供各个方面的状态评估。例如,传感器组件1114可以检测到设备1100的打开/关闭状态,组件的相对定位,例如组件为装置1100的显示器和小键盘,传感器组件1114还可以检测装置1100或装置1100一个组件的位置改变,用户与装置1100接触的存在或不存在,装置1100方位或加速/减速和装置1100的温度变化。传感器组件1114可以包括接近传感器,被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件1114还可以包括光传感器,如CMOS或CCD图像传感器,用于在成像应用中使用。在一些实施例中,该传感器组件1114还可以包括加速度传感器,陀螺仪传感器,磁传感器,压力传感器或温度传感器。

[0155] 通信组件1116被配置为便于装置1100和其他设备之间有线或无线方式的通信。装置1100可以接入基于通信标准的无线网络,如WiFi,2G或3G,或它们的组合。在一个示例性实施例中,通信组件1116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,通信组件1116还包括近场通信 (NFC) 模块,以促进短程通信。例如,在NFC模块可基于射频识别 (RFID) 技术,红外数据协会 (IrDA) 技术,超宽带 (UWB) 技术,蓝牙 (BT) 技术和其他技术来实现。

[0156] 在示例性实施例中,装置1100可以被一个或多个应用专用集成电路 (ASIC)、数字信号处理器 (DSP)、数字信号处理设备 (DSPD)、可编程逻辑器件 (PLD)、现场可编程门阵列 (FPGA)、控制器、微控制器、微处理器或其他电子元件实现,用于执行上述方法。

[0157] 在示例性实施例中,还提供了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1104,上述指令可由装置1100的处理器1120执行以完成上述方法。例如,非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0158] 一种计算机装置,包括:

[0159] 处理器;

[0160] 用于存储处理器可执行指令的存储器;

[0161] 其中,处理器被配置为:

[0162] 接收视频源设备发送的音频数据,音频数据包括多个音频采样点;

[0163] 获取音频播放设备与视频源设备之间的网络延迟变化,其中,网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差;

[0164] 根据网络延迟变化确定待调整音频数据总量;

[0165] 根据待调整音频数据总量和预设的调整策略,对接收到的音频数据增加音频采样点再播放或删除音频采样点再播放,以使得音频播放设备与视频源设备同步。

[0166] 一种非临时性计算机可读存储介质,当存储介质中的指令由移动终端的处理器执行时,使得移动终端能够执行一种音画同步方法,方法包括:

[0167] 接收视频源设备发送的音频数据,音频数据包括多个音频采样点;

[0168] 获取音频播放设备与视频源设备之间的网络延迟变化,其中,网络延迟变化指的是本次监测获取的网络延时和上一次监测获取的网络延时的时间差;

[0169] 根据网络延迟变化确定待调整音频数据总量;

[0170] 根据待调整音频数据总量和预设的调整策略,对接收到的音频数据增加音频采样点再播放或删除音频采样点再播放,以使得音频播放设备与视频源设备同步。

[0171] 本公开的实施例提供了一种音画同步方法及装置,考虑到网络传输造成的延时,对于通过网络连接的音频播放设备和视频源设备,首先接收视频源设备发送的包括多个音频采样点的音频数据,然后获取音频播放设备与视频源设备之间的网络延迟变化,再根据网络延迟变化确定待调整音频数据总量,最后根据待调整音频数据总量和预设的调整策略,对接收到的音频数据增加音频采样点再播放或删除音频采样点再播放,以使得音频播放设备与视频源设备同步。。解决了网络传输延迟导致的声音与画面不同步的问题,提升了用户体验。

[0172] 对于家庭中有一台电视和两个及以上智能音箱的用户,把电视和智能音箱连接起来,电视和音箱一起播放视频中的声音内容,实现环绕立体声,达到比电视单独播放更好的听音观影体验

[0173] 基于时钟信息交互对网络状态进行监测,并根据网络状态的监测结果动态调整声音播放的延迟,提供更自然的音画同步体验。

[0174] 在调整音频数据的播放提前或推后时,通过样条插值的方法来补充或删除由动态调整引起的数据变化问题,降低数据变化的可感知程度,进一步优化了用户的体验。

[0175] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本发明的其它实施方案。本申请旨在涵盖本发明的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本发明的一般性原理并包括本公开未公开的本技术领域中的公知常识

或惯用技术手段。说明书和实施例仅被视为示例性的，本发明的真正范围和精神由下面的权利要求指出。

[0176] 应当理解的是，本发明并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围进行各种修改和改变。本发明的范围仅由所附的权利要求来限制。

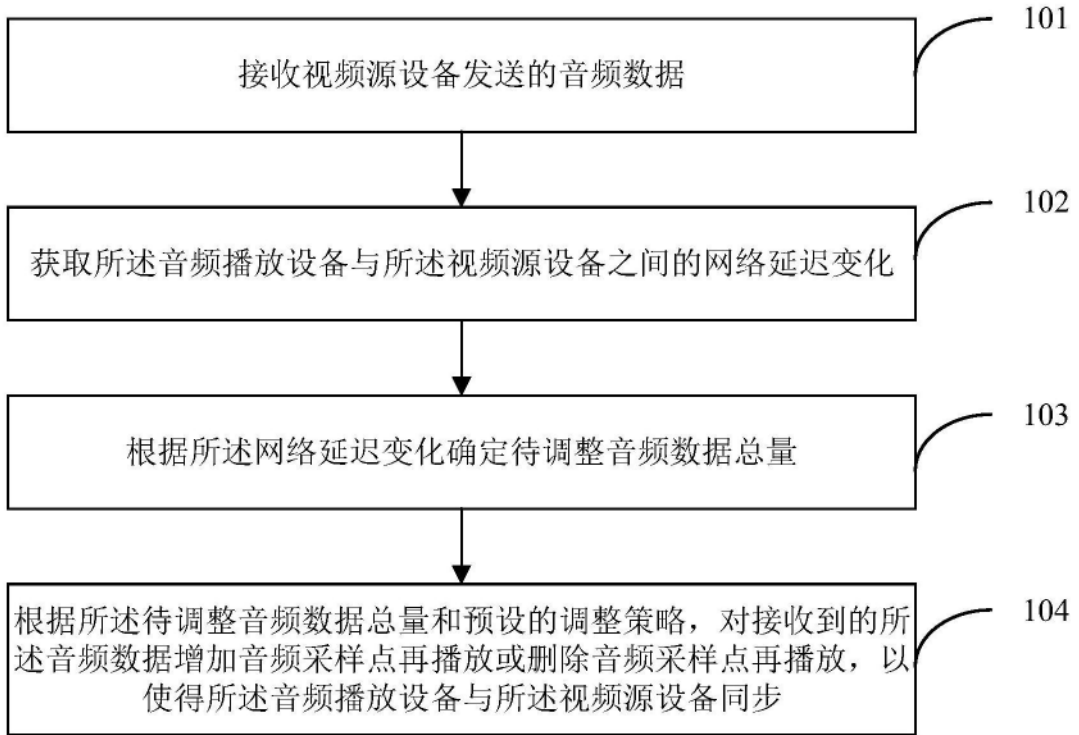


图1

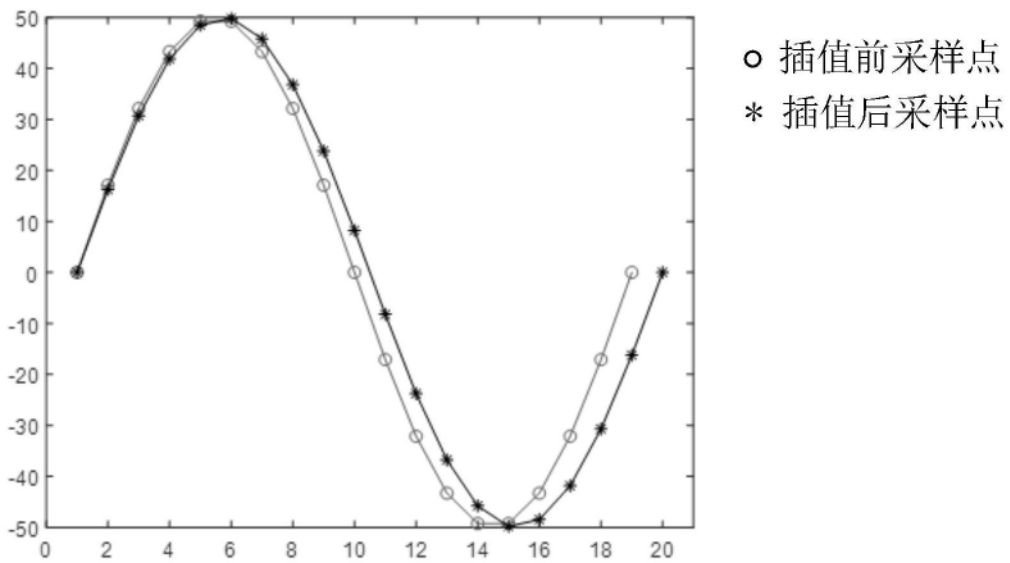


图2

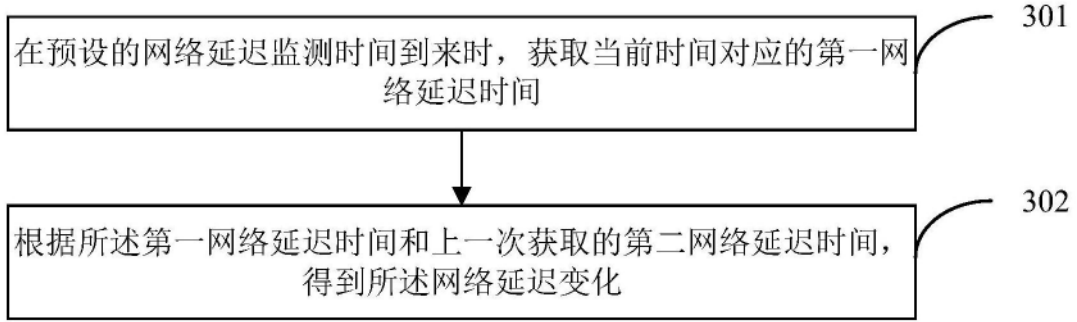


图3

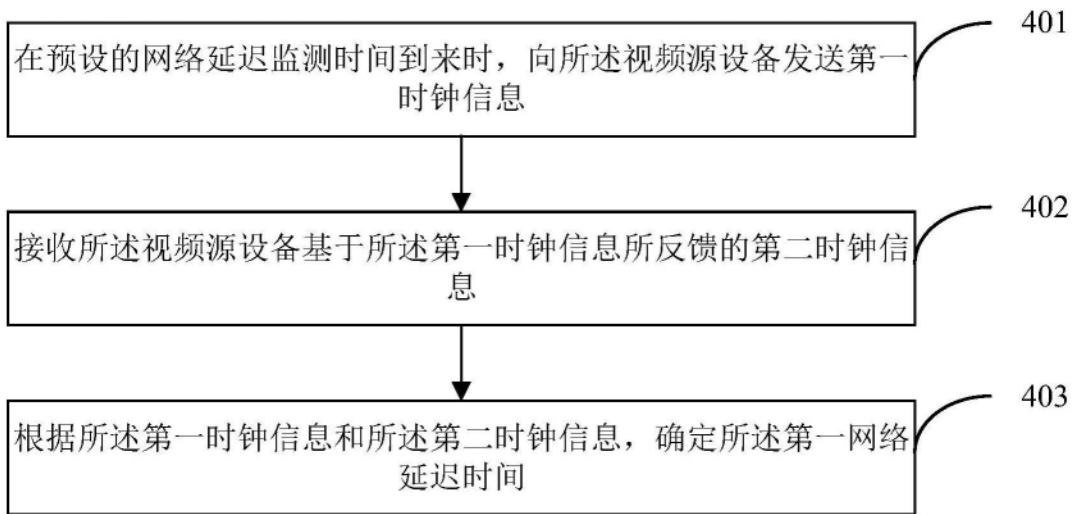


图4

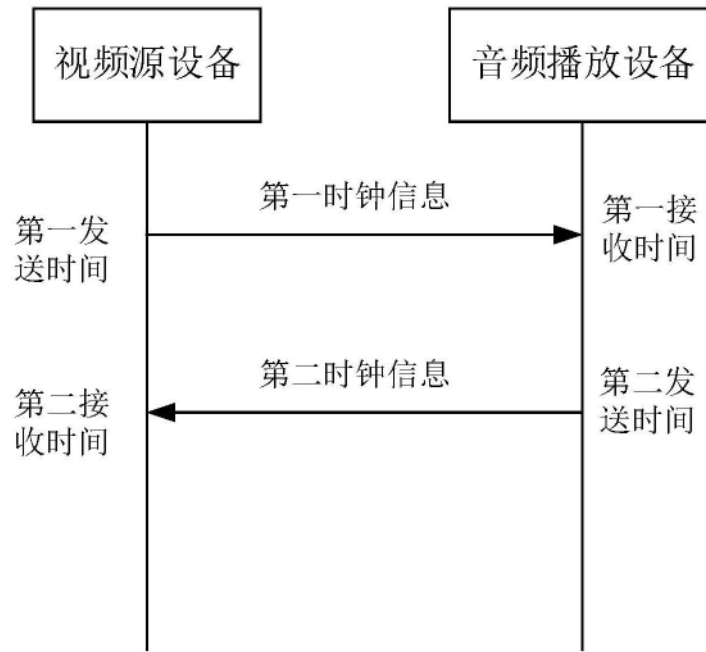


图5

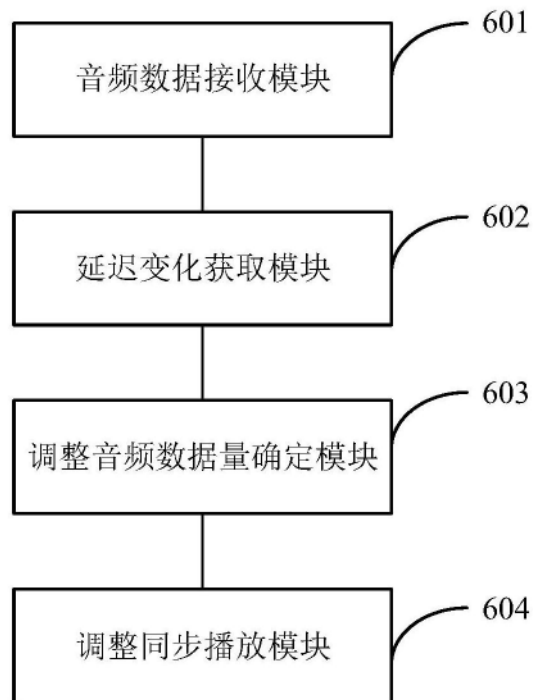


图6

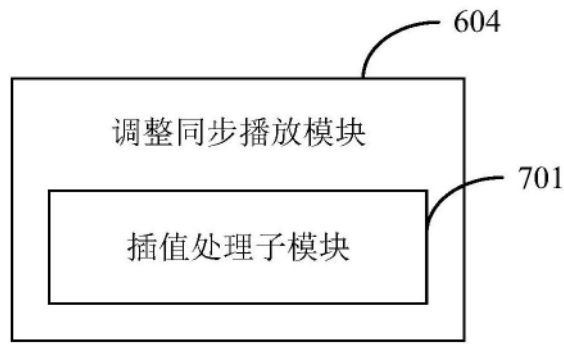


图7

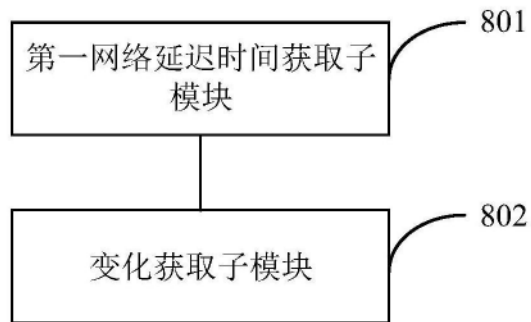


图8

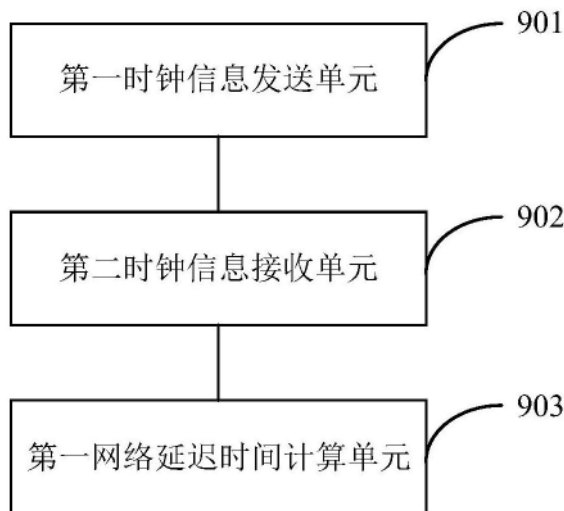


图9

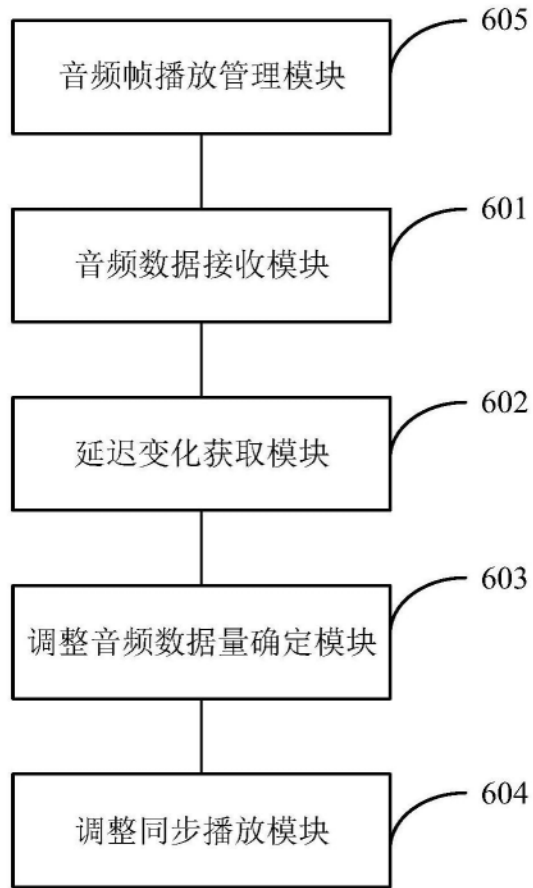


图10

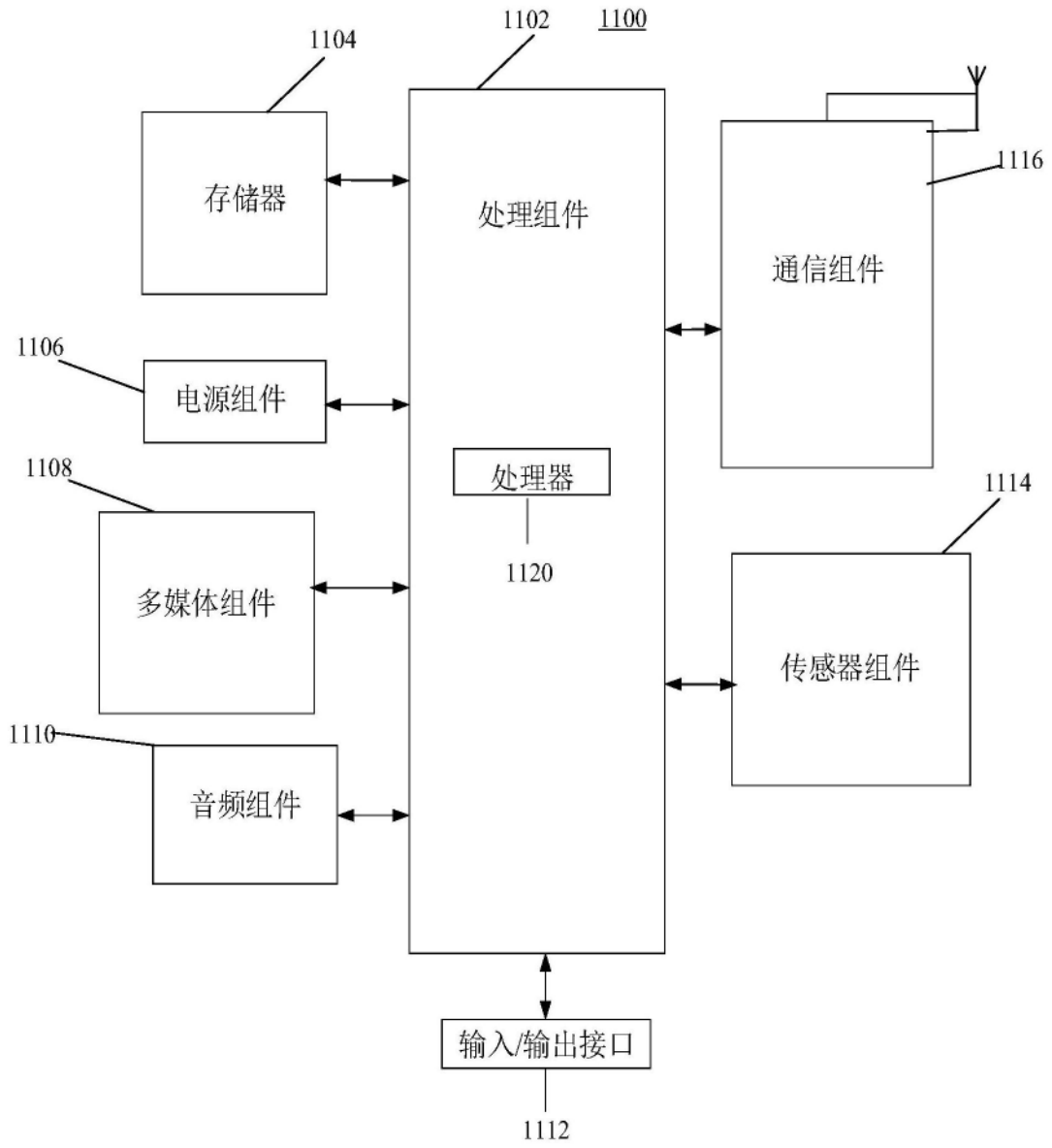


图11