



(12) 发明专利

(10) 授权公告号 CN 109963184 B

(45) 授权公告日 2022. 04. 29

(21) 申请号 201711339869.9

H04N 21/439 (2011.01)

(22) 申请日 2017.12.14

H04N 21/44 (2011.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 109963184 A

(56) 对比文件

CN 102271280 A, 2011.12.07

CN 102271280 A, 2011.12.07

(43) 申请公布日 2019.07.02

CN 103475927 A, 2013.12.25

(73) 专利权人 阿里巴巴集团控股有限公司

CN 103259803 A, 2013.08.21

地址 英属开曼群岛大开曼资本大厦一座四层847号邮箱

CN 107147937 A, 2017.09.08

CN 107484009 A, 2017.12.15

(72) 发明人 唐磊

CN 105245496 A, 2016.01.13

CN 105847857 A, 2016.08.10

(74) 专利代理机构 北京清源汇知识产权代理事务所(特殊普通合伙) 11644

WO 2015099687 A1, 2015.07.02

代理人 冯德魁 窦晓慧

审查员 宋佳慧

(51) Int. Cl.

H04N 21/43 (2011.01)

H04N 21/8547 (2011.01)

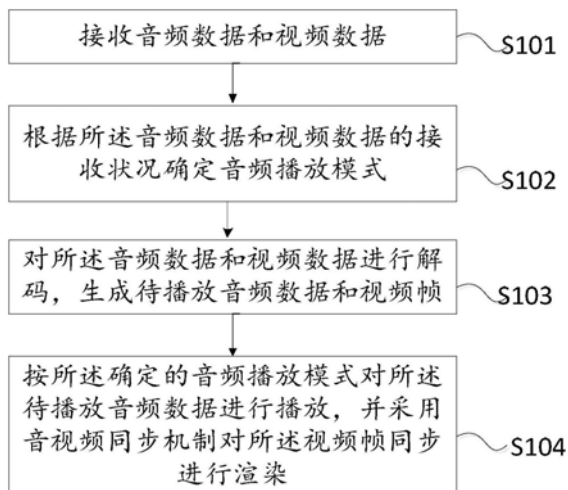
权利要求书3页 说明书16页 附图4页

(54) 发明名称

一种音视频网络播放的方法、装置以及电子设备

(57) 摘要

本申请公开了一种音视频网络播放的方法、装置以及电子设备,所述方法包括:接收音频数据和视频数据;根据所述音频数据和视频数据的接收状况确定音频播放模式;对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧;按所述确定的音频播放模式对所述待播放音频数据进行播放,并采用音视频同步机制对所述视频帧同步进行渲染。本方法通过所述音视频的加速播放加速消耗缓存,抵消音视频数据的累积所造成的播放延时;通过所述音视频的减速播放减速消耗缓存,避免网络抖动所引起的音视频播放的卡顿,在增加抗网络抖动的能力以及解决播放延时的同时,最大限度地保证了音视频的播放效果。



1. 一种音视频网络播放的方法,其特征在于,包括:

接收音频数据和视频数据,将所述音频数据和视频数据存储至接收缓冲区中;

根据所述音频数据和视频数据的接收状况确定音频播放模式,包括:根据所述接收缓冲区的占用比例确定音频播放模式,并在所述确定音频播放模式的过程中对所述接收缓冲区的占用水平做平滑处理,所述平滑处理包括:当接收的所述音频数据和视频数据稍大于所述接收缓冲区的存储阈值,而所述接收缓冲区的占用水平正在减小,则保持现有的音频播放模式不做任何变动,或者,当接收的音频数据和视频数据稍小于所述存储阈值,而接收缓冲区的占用水平正在增加,则保持现有的音频播放模式不做任何变动;

对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧;

按所述确定的音频播放模式对所述待播放音频数据进行播放,并采用音视频同步机制对所述视频帧同步进行渲染,所述音视频同步机制包括:采用基于时间戳的音视频同步技术中的视频时间戳跟踪音频时间戳的方法对音频和视频同步进行播放。

2. 根据权利要求1所述的音视频网络播放的方法,其特征在于,所述音频播放模式包括加速播放、减速播放和正常播放中的一种;所述按所述确定的音频播放模式对所述待播放音频数据进行播放的方法包括:

声音变速不变调的方法。

3. 根据权利要求2所述的音视频网络播放的方法,其特征在于,所述声音变速不变调的方法为同步波形叠加法,所述按所述确定的音频播放模式对所述待播放音频数据进行播放,包括如下步骤:

对所述待播放音频数据进行分析,获得该待播放音频数据的特征参数信息;

对所述待播放音频数据进行语音活动检测,用以检测出所述待播放音频数据为有用声音或者为背景噪音;

若所述待播放音频数据为有用声音,则对所述待播放音频数据进行降采样,并且,对所述待播放音频数据进行分解以及合成;若所述待播放音频数据为背景噪音,则利用上一音频数据对该背景噪音进行合并;

输出所述待播放音频数据。

4. 根据权利要求3所述的音视频网络播放的方法,其特征在于,所述对所述待播放音频数据进行降采样,包括:

若所述待播放音频数据的采样率大于1khz,则降采样到1khz。

5. 根据权利要求1所述的音视频网络播放的方法,其特征在于,所述音频数据和视频数据的接收状况,包括:

接收的音频数据和视频数据的数量;和/或

所述音频数据和视频数据的接收时间。

6. 根据权利要求5所述的音视频网络播放的方法,其特征在于,当所述音频数据和视频数据的接收状况为接收的音频数据和视频数据的数量时,所述根据所述音频数据的接收状况确定音频播放模式,包括如下方式中的一种:

接收的音频数据和视频数据的数量大于预定阈值,确定音频播放模式为加速播放模式;

接收的音频数据和视频数据的数量小于预定阈值,确定音频播放模式为减速播放模

式；

接收的音频数据和视频数据的数量在预定阈值范围内，确定音频播放模式为正常播放模式。

7. 根据权利要求1-6任一项所述的音视频网络播放的方法，其特征在于，所述音视频网络播放的方法可应用于音视频网络直播中。

8. 一种音视频网络播放的装置，其特征在于，包括：

音视频数据接收单元：用于接收音频数据和视频数据，将所述音频数据和视频数据存储至接收缓冲区中；

音频播放模式确定单元：用于根据所述音频数据和视频数据的接收状况确定音频播放模式，包括：根据所述接收缓冲区的占用比例确定音频播放模式，并在所述确定音频播放模式的过程中对所述接收缓冲区的占用水平做平滑处理，所述平滑处理包括：当接收的所述音频数据和视频数据稍大于所述接收缓冲区的存储阈值，而所述接收缓冲区的占用水平正在减小，则保持现有的音频播放模式不做任何变动，或者，当接收的音频数据和视频数据稍小于所述存储阈值，而接收缓冲区的占用水平正在增加，则保持现有的音频播放模式不做任何变动；

音视频数据解码单元：用于对所述音频数据和视频数据进行解码，生成待播放音频数据和视频帧；

音视频播放单元：用于按所述确定的音频播放模式对所述待播放音频数据进行播放，并采用音视频同步机制对所述视频帧同步进行渲染，所述音视频同步机制包括：采用基于时间戳的音视频同步技术中的视频时间戳跟踪音频时间戳的方法对音频和视频同步进行播放。

9. 根据权利要求8所述的音视频网络播放的装置，其特征在于，所述音频播放模式包括加速播放、减速播放和正常播放中的一种；所述按所述确定的音频播放模式对所述待播放音频数据进行播放的方法包括：

声音变速不变调的方法。

10. 一种电子设备，其特征在于，包括：

处理器；以及

存储器，用于存储一种用于音视频网络播放的程序，该设备通电并通过所述处理器运行所述用于音视频网络播放的程序后，执行下述步骤：

接收音频数据和视频数据，将所述音频数据和视频数据存储至接收缓冲区中；

根据所述音频数据和视频数据的接收状况确定音频播放模式，包括：根据所述接收缓冲区的占用比例确定音频播放模式，并在所述确定音频播放模式的过程中对所述接收缓冲区的占用水平做平滑处理，所述平滑处理包括：当接收的所述音频数据和视频数据稍大于所述接收缓冲区的存储阈值，而所述接收缓冲区的占用水平正在减小，则保持现有的音频播放模式不做任何变动，或者，当接收的音频数据和视频数据稍小于所述存储阈值，而接收缓冲区的占用水平正在增加，则保持现有的音频播放模式不做任何变动；

对所述音频数据和视频数据进行解码，生成待播放音频数据和视频帧；

按所述确定的音频播放模式对所述待播放音频数据进行播放，并采用音视频同步机制对所述视频帧同步进行渲染，所述音视频同步机制包括：采用基于时间戳的音视频同步技

术中的视频时间戳跟踪音频时间戳的方法对音频和视频同步进行播放。

## 一种音视频网络播放的方法、装置以及电子设备

### 技术领域

[0001] 本发明涉及音视频播放领域,具体涉及一种音视频网络播放的方法。本发明同时涉及一种音视频网络播放的装置以及电子设备。本申请另外涉及一种音视频网络播放的方法、一种音视频网络播放的装置、以及一种电子设备。

### 背景技术

[0002] 随着信息化技术的高速发展,多媒体技术已经成为人们生活中一种非常重要的交流手段和信息获取手段,多媒体中的音频和视频为人们提供了巨大的信息量,音视频网络播放技术作为现代流媒体播放技术的重要组成部分,其播放效果的好坏直接影响人们摄取信息的质量,并且影响人们的主观感受。

[0003] 影响音视频网络播放效果的因素较多,其中最主要的影响因素为网络传输过程的不稳定所导致的播放延时或网络抖动,现有的音视频播放器,如果因网络拥塞、定时漂移或路由变更而产生抖动,则可能导致迟迟接收不到音视频包,原播放器缓存中的音视频数据将被消耗殆尽,因此出现画面静止、声音停止等音视频卡顿现象;当网络由不好转到好的情况时,则会在同一时间接收到大量的音视频数据,造成大量的音视频数据在播放器的缓存区域进行累积,从而导致播放延时增大。

[0004] 现有的处理办法主要是通过预设缓存阈值来适应网络状况的变化,例如,设定一定的缓存阈值,当接收的音视频数据超过该阈值时,直接丢弃超出部分的音视频数据。

[0005] 然而,上述处理方法存在一定的缺陷,主要为:丢弃超出部分的音视频数据将造成后续播放的音视频不完整,直接影响视频播放的连贯度、以及使音频出现破音,从而造成音视频播放效果较差。

### 发明内容

[0006] 本申请提供一种音视频网络播放的方法,以解决现有的音视频网络播放中因网络状况的变化所导致的音视频卡顿或播放延时增大的问题,以及现有的为了适应网络状况的变化所采取的丢弃音视频数据所造成的音视频播放效果差的问题,本申请同时提供一种音视频网络播放的装置以及电子设备。本申请另外提供一种音视频网络播放的方法、装置、以及电子设备。

[0007] 本申请提供一种音视频网络播放的方法,包括:

[0008] 接收音频数据和视频数据;

[0009] 根据所述音频数据和视频数据的接收状况确定音频播放模式;

[0010] 对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧;

[0011] 按所述确定的音频播放模式对所述待播放音频数据进行播放,并采用音视频同步机制对所述视频帧同步进行渲染。

[0012] 优选的,所述音频播放模式包括加速播放、减速播放和正常播放中的一种;所述按所述确定的音频播放模式对所述待播放音频数据进行播放的方法包括:

- [0013] 声音变速不变调的方法。
- [0014] 可选的,所述声音变速不变调的方法为同步波形叠加法,所述按所述确定的音频播放模式对所述待播放音频数据进行播放,包括如下步骤:
- [0015] 对所述待播放音频数据进行分析,获得该待播放音频数据的特征参数信息;
- [0016] 对所述待播放音频数据进行语音活动检测,用以检测出所述待播放音频数据为有用声音或者为背景噪音;
- [0017] 若所述待播放音频数据为有用声音,则对所述待播放音频数据进行降采样,并且,对所述待播放音频数据进行分解以及合成;若所述待播放音频数据为背景噪音,则利用上一音频数据对该背景噪音进行合并;
- [0018] 输出所述待播放音频数据。
- [0019] 可选的,所述对所述待播放音频数据进行降采样,包括:
- [0020] 若所述待播放音频数据的采样率大于1khz,则降采样到1khz。
- [0021] 优选的,所述音视频同步机制包括:
- [0022] 采用基于时间戳的音视频同步技术中的视频时间戳跟踪音频时间戳的方法对音频和视频同步进行播放。
- [0023] 可选的,所述音频数据和视频数据的接收状况,包括:
- [0024] 接收的音频数据和视频数据的数量;和/或
- [0025] 所述音频数据和视频数据的接收时间。
- [0026] 可选的,当所述音频数据和视频数据的接收状况为接收的音频数据和视频数据的数量时,所述根据所述音频数据的接收状况确定音频播放模式,包括如下方式中的一种:
- [0027] 接收的音频数据和视频数据的数量大于预定阈值,确定音频播放模式为加速播放模式;
- [0028] 接收的音频数据和视频数据的数量小于预定阈值,确定音频播放模式为减速播放模式;
- [0029] 接收的音频数据和视频数据的数量在预定阈值范围内,确定音频播放模式为正常播放模式。
- [0030] 可选的,所述接收的音频数据和视频数据预先存储于抖动缓冲区Jitter buffer中。
- [0031] 可选的,所述音视频网络播放的方法可应用于音视频网络直播中。
- [0032] 本申请提供一种音视频网络播放的装置,包括:
- [0033] 音视频数据接收单元:用于接收音频数据和视频数据;
- [0034] 音频播放模式确定单元:用于根据所述音频数据和视频数据的接收状况确定音频播放模式;
- [0035] 音视频数据解码单元:用于对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧;
- [0036] 音视频播放单元:用于按所述确定的音频播放模式对所述待播放音频数据进行播放,并采用音视频同步机制对所述视频帧同步进行渲染。
- [0037] 优选的,所述音频播放模式包括加速播放、减速播放和正常播放中的一种;所述按所述确定的音频播放模式对所述待播放音频数据进行播放的方法包括:

- [0038] 声音变速不变调的方法。
- [0039] 优选的,所述音视频同步机制包括:
- [0040] 采用基于时间戳的音视频同步技术中的视频时间戳跟踪音频时间戳的方法对音频和视频同步进行播放。
- [0041] 本申请提供一种电子设备,包括:
- [0042] 处理器;以及
- [0043] 存储器,用于存储一种用于音视频网络播放的程序,该设备通电并通过所述处理器运行所述用于音视频网络播放的程序后,执行下述步骤:
- [0044] 接收音频数据和视频数据;
- [0045] 根据所述音频数据和视频数据的接收状况确定音频播放模式;
- [0046] 对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧;
- [0047] 按所述确定的音频播放模式对所述待播放音频数据进行播放,并采用音视频同步机制对所述视频帧同步进行渲染。
- [0048] 本申请另外提供一种音视频网络播放的方法,包括:
- [0049] 接收音频数据和视频数据;
- [0050] 根据所述音频数据和视频数据的接收状况,确定视频播放模式;
- [0051] 对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧;
- [0052] 按所述确定的视频播放模式,对所述视频帧进行播放,并采用音视频同步机制对所述待播放音频数据同步进行播放。
- [0053] 优选的,所述视频播放模式包括加速播放、减速播放和正常播放中的一种;所述按所述确定的视频播放模式,对所述视频帧进行播放,包括:
- [0054] 对所述视频帧进行识别,获得识别结果;
- [0055] 按所述确定的视频播放模式、结合所述识别结果对所述视频帧进行加速播放、减速播放、或正常播放。
- [0056] 可选的,对所述视频帧进行识别的方法包括:
- [0057] 采用基于机器学习的方法。
- [0058] 可选的,所述音视频同步机制包括:
- [0059] 采用基于时间戳的音视频同步技术中的音频时间戳跟踪视频时间戳的方法对音频和视频同步进行播放。
- [0060] 可选的,所述对所述待播放音频数据同步进行播放,其播放方式包括:
- [0061] 采用声音变速不变调方法。
- [0062] 本申请还提供一种音视频网络播放的装置,包括:
- [0063] 音视频数据接收单元:用于接收音频数据和视频数据;
- [0064] 视频播放模式确定单元:用于根据所述音频数据和视频数据的接收状况,确定视频播放模式;
- [0065] 音视频数据解码单元:用于对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧;
- [0066] 音视频播放单元:用于按所述确定的视频播放模式对所述视频帧进行播放,并采用音视频同步机制对所述待播放音频数据同步进行播放。

- [0067] 本申请还提供一种电子设备,包括:
- [0068] 处理器;以及
- [0069] 存储器,用于存储一种用于音视频网络播放的程序,该设备通电并通过所述处理器运行所述用于音视频网络播放的程序后,执行下述步骤:
- [0070] 接收音频数据和视频数据;
- [0071] 根据所述音频数据和视频数据的接收状况,确定视频播放模式;
- [0072] 对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧;
- [0073] 按所述确定的视频播放模式,对所述视频帧进行播放,并采用音视频同步机制对所述待播放音频数据同步进行播放。
- [0074] 与现有技术相比,本发明具有以下优点:
- [0075] 本申请通过对音视频数据的接收状况进行分析,以判定当前的网络状况,据此动态调整音频的播放模式,并采用音视频同步机制对视频同步渲染,从而达到根据网络状况动态调整音视频缓存大小,通过所述音视频的加速播放加速消耗缓存,抵消音视频数据的累积所造成的播放延时;通过所述音视频的减速播放减速消耗缓存,避免网络抖动所引起的音视频播放的卡顿。相比于现有技术,本申请在增加抗网络抖动的能力以及解决播放延时的同时,最大限度地保证了音视频的播放效果。
- [0076] 进一步的,本申请采用声音变速不变调的方法进行加速或减速播放音频,在不影响语音声调的情况下,控制声音的播放速度,使得在加速消耗或减速消耗缓存的同时,对所述音频的播放效果的影响降低到最小;并且,本申请采用基于时间戳的视频时间戳跟踪音频时间戳的音视频同步方法播放所述音频和视频,使得音视频的同步播放更加灵活以及科学。

### 附图说明

- [0077] 图1是本申请第一实施例提供的一种音视频网络播放的方法的流程图;
- [0078] 图2是本申请第一实施例提供的对待播放音频数据进行播放的流程图;
- [0079] 图3是本申请第二实施例提供的一种音视频网络播放的装置的示意图;
- [0080] 图4是本申请第三实施例提供的一种电子设备的示意图;
- [0081] 图5为本申请第四实施例提供的一种音视频网络播放的方法的流程图;
- [0082] 图6为本申请第四实施例提供的按所述确定的视频播放模式对所述视频帧进行播放的流程图;
- [0083] 图7为本申请第五实施例提供的一种音视频网络播放的装置的示意图;
- [0084] 图8为本申请第六实施例提供的一种电子设备的示意图。

### 具体实施方式

[0085] 在下面的描述中阐述了很多具体细节以便于充分理解本发明。但是本发明能够以很多不同于在此描述的其它方式来实施,本领域技术人员可以在不违背本发明内涵的情况下做类似推广,因此本发明不受下面公开的具体实施的限制。

[0086] 本申请提供一种音视频网络播放的方法,该方法主要应用于对经网络传输的音视频数据进行播放,所述经网络传输的音视频数据可为预先存储于服务器端的音视频,也可



为网络直播中的音视频;本申请还提供一种音视频网络播放的装置以及电子设备,以下提供实施例对所述方法、装置以及电子设备进行详细说明。

[0087] 本申请第一实施例提供一种音视频网络播放的方法,请参看图1理解该实施例,图1为本实施例提供的音视频网络播放的方法的流程图。以下结合图1介绍该实施例。

[0088] 如图1所示,本实施例提供的音视频网络播放的方法包括如下步骤:

[0089] S101,接收音频数据和视频数据。

[0090] 本步骤用于接收经网络传输后的音视频数据,并对所述音视频数据进行预处理。由于本申请所提供的方法为音视频网络播放的方法,因此,需首先接收音视频的数据信息作为音视频网络播放的基础信息,之后才能进行后续操作。

[0091] 所述音频数据和视频数据指的是将语音和视频经采集、编码压缩后,按特定的网络传输协议进行打包和网络传输后获得的音频数据文件和视频数据文件。所述音频数据和视频数据的传输过程一般为:将所述视频数据和音频数据通过实时流媒体传输协议传输至专用的流媒体服务器,常用的实时流媒体传输协议为实时传输协议/实时传输控制协议(RTP/RTCP协议),流媒体服务器将所述视频数据和音频数据进行存储或实时分发给接收端,接收端接收所述音频数据和视频数据,并作进一步处理后进行实时呈现。

[0092] 接收音频数据和视频数据,主要包括:读取音视频数据和对音视频数据进行分流,所述读取音视频数据指的是将多媒体数据按照流的方式读入并判断数据流的类型,以确定所述多媒体数据为音频或视频,根据解读数据流的头文件判断所述视频数据或音频数据的压缩格式,以确定所述数据流为何种视频、音频格式的文件;所述对音视频数据进行分流,指的是按照所述已判断出的文件格式中存储音频块和视频块的规则,依靠数据流的头部信息来判断所述音频和视频数据在该段数据块中的位置,继而对音频和视频进行分离,分离后的音频数据和视频数据分别通过各自的数据通道进入其对应的缓存区域进行预存储。

[0093] 需要说明的是,在对所述音频数据和视频数据进行分离的同时,将其各自的时间戳一同分离出来,随所述音频或视频数据一同传送,所述时间戳是在编码环节设定的,它标识了图像数据进行渲染以及音频数据进行播放的时间、或为预定的播放策略提供依据。

[0094] 所述将分离后的音频数据和视频数据进行预存储,指的是将所述音视频数据分别存储至预设的缓冲区域中,使之成为缓存数据,用于消除网络延时抖动导致的乱序,保证音频数据流和视频数据流的连续性,即保证媒体内同步。例如,本实施例中,所述音频数据和所述视频数据经分离后分别存储至预设的接收缓冲区(英文名称为Jitter buffer)中,该缓冲区设置有门限值,当所述音频数据或视频数据在其各自的接收缓冲区中累积的数值达到所述设定的门限值时,音频数据或视频数据按预定的输出方式从所述接收缓冲区中输出。

[0095] S102,根据所述音频数据和视频数据的接收状况确定音频播放模式。

[0096] 本步骤的作用在于根据上述接收到的音频数据和视频数据,预先确定音频的播放模式,所述音频播放模式包括如下播放模式中的一种:加速播放、减速播放、正常播放,通过确定的加速或减速播放模式来加速消耗或减慢消耗接收的数据流,从而达到降低播放延时或降低网络抖动带来的影响。

[0097] 所述音频数据和视频数据的接收状况包括接收的音频数据和视频数据的数量和/或接收时间,所述音频数据和视频数据的数量以及接收时间均可反应当前网络状况的好

坏。

[0098] 在接收端,通过对所述音频数据和视频数据的到达时间进行监测,可准确得出网络的时延抖动,对网络状况的变化提供准确依据。在本实施例中,通过对所述接收的音频数据和视频数据的数量进行监测来实现对网络状况的判定,所述接收的音频数据和视频数据的数量可通过对所述接收缓冲区中的数据量的占用比例进行周期性监测而获得,所述接收缓冲区占用比例的变化,动态反映了网络的当前状况。该过程与实时传输协议/实时传输控制协议(RTP/RTCP协议)的发送端与接收端的反馈控制过程相似,所述RTP/RTCP协议的发送端与接收端的反馈控制过程为:接收端通过RTP包的接收状况判定网络状况,根据网络状况调整缓存容量以及阈值以适应网络状况的变化,以及接收端有规律地发送RTCP包给发送端以告知抖动、丢包等信息,发送端根据所述RTCP包中包含的反映网络状况的信息动态调整音频编码和视频编码的压缩率,实现动态调整音视频数据量的目的。而本方法则为通过动态监测所述接收的音频数据和视频数据的数量的多少来确定后续的播放策略,该过程具体为:首先,对所述音频数据和视频数据的接收缓冲区预先设定存储阈值(接收缓冲区的占用比例),当所述接收的音频数据和视频数据大于所述预定阈值,即当前接收的音频数据和视频数据的数量超出所述预定阈值的上限,则表明当前网络状况较好,对应的,设定音频播放模式为加速播放;当所述接收的音频数据和视频数据小于所述存储阈值,即当前接收的音频数据和视频数据的数量低于所述预定阈值的下限,则表明当前网络状况较差,对应的,设定音频播放模式为减速播放;当所述接收的音频数据和视频数据处于所述预定阈值的范围内,则设定音频播放模式为正常播放。

[0099] 设定所述音频的播放模式除了确定加速或减速播放外,还需初步确定加速或减速播放的幅度,即具体的加速或减速播放的速率,所述进行加速或减速播放的幅度与所述接收缓冲区的占用比例相对应,其可根据所述接收缓冲区的占用比例实时进行调整。并且,对所述接收缓冲区预先设定的存储阈值也可根据所述音频数据和视频数据的类型、应用场景、反馈的播放效果等因素而实时进行调整,例如,可提高或降低所述预定阈值的上限或下限,以满足实际播放需求。

[0100] 需要说明的是,为了避免因缓存状态的临时波动而对上述音频播放模式的调整过于频繁,可对所述接收缓冲区的占用水平做平滑处理,具体的可通过设定缓存变化率实现对所述接收缓冲区的缓存占用趋势的判断,例如,当所述接收的音频数据和视频数据稍大于所述存储阈值,而缓存变化率显示当前的接收缓冲区的占用水平正在减小,或者,当所述接收的音频数据和视频数据稍小于所述存储阈值,而缓存变化率显示当前的接收缓冲区的占用水平正在增加,则保持现有的播放模式不做任何变动。

[0101] 所述动态监测接收的音频数据量和视频数据量,可按照预设监测周期对所述接收缓冲区的占用水平进行监测,也可按照每播放一个音视频数据单元就对所述接收缓冲区监测一次的方式进行监测。在按预设监测周期对所述接收缓冲区进行监测时,所述监测周期的取值较为敏感,若取值偏小,则可能出现因为对网络抖动估计不正确而过度调整,导致接收缓冲区的占用水平出现震荡;若取值偏大,则可能调整不及时而出现延误。本实施例中采用每播放一个音视频数据单元就对所述接收缓冲区监测一次的方式进行监测。

[0102] 确定音频的播放模式后,需对所述接收缓冲区中的音频数据和视频数据的输出方式进行相应调整,使所述接收缓冲区中的音频数据和视频数据的输出状况与所述播放模式

相符。常用的输出方式为采用到达事件驱动的方式或定时读取的方式，所述定时读取的方式指的是从所述接收缓冲区中读取的音频数据和视频数据是按照固定的时钟节拍进行的，本实施例中所述音频数据和视频数据均采用到达事件驱动的方式进行输出，具体为：当所述播放模式为加速播放时，加速输出所述音频数据和视频数据；当所述播放模式为正常播放或减速播放时，则按原有的音视频读取方式进行输出。

[0103] 确定所述音频的播放模式后，将所述音频的播放模式信息作为控制信息实时发送至后续的播放环节。

[0104] S103,对所述音频数据和视频数据进行解码，生成待播放音频数据和视频帧。

[0105] 本步骤的作用在于对上述经预存储之后输出的音频数据和视频数据进行解码，以生成待播放音频数据和视频帧。

[0106] 对于接收的视频数据，在网络传输前都需按照预定的视频编解码标准进行编码，例如，采用H264视频编解码标准对视频图像帧进行量化处理、预测编码、变换编码以及熵编码等压缩途径来消除时间冗余和空间冗余，以减少所述视频图像帧在网络传输或存储过程中的数据量，对应的，在接收端需采用H264视频编解码标准进行解码，用以还原视频图像。同样的，对于接收的音频数据，在网络传输前和传输后也需分别按照预定的音频编解码标准进行编码以及解码，例如，可采用算法简单、容易实现的G.711音频编解码标准对音频数据进行编码以及解码，所述G.711音频编解码标准是基于PCM编码技术实现的。相对于视频数据，音频数据所占的数据量较小，其编码和解码过程也相对较为简单。

[0107] 上述步骤S101中的接收音频数据和视频数据的环节已确定出所述音频数据和视频数据的压缩格式，本步骤根据所述压缩格式以及对应的视频编解码标准和音频编解码标准，利用音频数据或视频数据的头部文件中所包含解码所需的关键参数，如输入输出所需缓存、选定的编解码器、帧率等信息，对所述视频数据和音频数据分别进行解码，以获取对应的视频帧以及待播放音频数据，所述待播放音频数据和视频帧存储于各自的播放缓存中，用于后续的输出播放。

[0108] S104,按所述确定的音频播放模式对所述待播放音频数据进行播放，并采用音视频同步机制对所述视频帧同步进行渲染。

[0109] 在步骤S102中根据音频数据和视频数据的接收状况确定了音频播放模式，本步骤用于按照所述音频的播放模式信息中的内容对所述待播放音频数据进行播放，并采用音视频同步机制对所述视频帧同步进行渲染，以此完成音视频的完整播放过程。

[0110] 所述音频的播放模式信息包括：加速播放及其初步预定的播放速率、或减速播放及其初步预定的播放速率、或正常播放，本申请采用声音变速不变调的方式对所述待播放音频数据进行播放，所述声音变速不变调指的是在保持原来语音信号相对强弱、频率、音调和音色等特征不变的情况下，使语音信号播放的速度放慢或加快。声音变速不变调算法包括时域法、频域法和参量法，其中，所述时域法包括剪切法、同步波形叠加法（英文简称为SOLA）、固定同步波形叠加法（英文简称为SOLA FS）、波形相似叠加法（英文简称为WSOLA）等。

[0111] 本实施例采用同步波形叠加法SOLA对所述待播放音频数据进行播放，该方法利用声音的短时平稳特性，把语音时长按一定的规整比例进行拉伸或压缩，用以实现对语音的加速和减速播放，该过程如图2所示，图2为对待播放音频数据进行播放的流程图，对待播放音频数据进行播放包括如下步骤：

[0112] S1041,对所述待播放音频数据进行分析,获得该待播放音频数据的特征参数信息。

[0113] 本步骤用于对所述经解码后的音频数据进行分析,获得音频缓存数据的单双声道、采样率等音频的特征参数信息。

[0114] S1042,对所述待播放音频数据进行语音活动检测(VAD检测),用以检测出所述待播放音频数据为有用声音或者为背景噪音。

[0115] 本步骤的目的在于通过语音活动检测方法对所述待播放音频数据的内容进行分类,所述有用声音指的是音频播放中的主要目标声音,例如,在淘宝直播中,所述有用声音主要为主播说话的声音;所述背景噪音为辅助声音或者可被滤除的干扰声音。

[0116] S1043,若所述待播放音频数据为有用声音,则对所述待播放音频数据进行降采样,并且,对该待播放音频数据进行分解以及合成;若所述待播放音频数据为背景噪音,则利用上一音频数据对该背景噪音进行合并。

[0117] 本步骤用于完成声音变速不变调的主体工作。若所述音频数据为有用声音,则根据该音频数据的所述特征参数信息中的采样率确定是否将该音频数据进行降采样,由于声音具有短时周期性,声音的基因周期一般在1kHz以下,因此,在本实施例中,若所述待播放音频数据的采样率大于1kHz,则需降采样到1kHz。

[0118] 所述对所述待播放音频数据进行分解以及合成,该过程可实现对音频的加快或减慢操作。所述分解阶段用于完成对所述待播放音频数据的分帧任务,分解后的所有帧用于合成变速语音。该分帧过程具体为:将所述音频数据按照间隔 $S_a$ 分割为长度固定的彼此重叠的帧,帧长为 $N$ , $N > S_a$ 。在完成分帧任务后,需通过高通滤波器滤除低频的背景噪声,以防止背景噪声的低频干扰。所述合成阶段用于将上述分解后的所有帧按照预定的合成间隔 $S_s$ 进行合并,该过程需首先确定所述合成间隔 $S_s$ ,所述合成间隔 $S_s$ 由规整因子 $\alpha$ 所决定,即 $S_s = \alpha S_a$ , $\alpha > 1$ 表示对语音进行拉伸,即对应于音频的减速播放, $\alpha < 1$ 表示对语音进行压缩,即对应于音频的加速播放,语音播放速度可按所述规整因子 $\alpha$ 的大小精确改变, $\alpha$ 的取值根据上述确定的播放模式进行确定。在合成第 $m$ 帧信号时,如果 $m = 1$ ,则直接将其写入输出信号,对于 $m > 1$ 的分解帧,在该帧应插入的位置(合成位置 $S_s \times m$ )处的左右附近 $[-km, km]$ 区域范围内滑动( $-km$ 和 $km$ 选择在 $-N/2 \sim N/2$ 之间),以重新找一个插入位置点插入第 $m$ 帧,该插入点为合并时的拼接点,该点需满足使第 $m$ 帧信号与已合成的信号在重叠区的互相关系数最大(即波形最相似),该过程可通过互相关系数的计算公式实现。然后重复该合成过程,直至所有的分解帧均被插入后实现合并,从而获得被规整(改变播放速度)的语音信号。通过上述方式,实现了语音信号长度的规整,并保证了非重叠区语音信号基频不变,从而实现语音播放时的变速不变调的效果。

[0119] 如果为背景噪音,则利用上一音频数据的信息对该背景噪音直接进行合并,使得在对该背景噪声进行加速或减慢操作时,所使用的参数继续沿用上一音频数据的参数,具体为:将上一音频数据的最强相关性位置与相关性最合适的长度,与当前音频数据的对应位置以及对应的前后长度进行合并或增加声音数据,并且同时保证声音的连贯性。

[0120] S1044,输出所述待播放音频数据。

[0121] 在对所述待播放音频数据经过上述步骤的处理后,已完成了对所述音频进行变速不变调的设定,本步骤用于对所述完成变速不变调设定的待播放音频数据进行输出播放。

[0122] 在按所述播放模式对所述音频数据进行加速或减速播放时,需将所述音频数据的真实播放速率反馈给步骤S102中的确定音频播放模式的阶段,用以校准和调整播放速率,结合所述语音的变速不变调算法实现对音频播放速率的精准控制,以制定更加精准的音频播放模式。

[0123] 在按上述方法实现所述待播放音频数据的变速不变调播放的同时,需采用音视频同步机制对所述视频帧同步进行渲染,以达到音视频同步播放的效果。

[0124] 之所以采用音视频同步机制,其原因在于音频数据和视频数据在发送、传输以及接收过程均可能发生不同步现象,一般情况下,音频数据和视频数据不同步的原因可分为以下两种:一种是终端处理数据引起的,发送端在处理采样、编码、打包等环节和接收端在处理解包、解码、播放等环节时,由于音频和视频的数据量以及编解码算法各不相同而引起时间差;另一种是传输过程中由于受到网络实时传输带宽、传输距离、网络节点的处理速度以及不同的地理环境等因素的影响所引起的网络时延和抖动、数据丢弃、时钟偏移等,从而引起音频网络播放和视频网络播放的失步。

[0125] 在本实施例中,需进行音视频同步机制的原因除了上述影响因素外,最主要的因素在于本方法所设定的采用声音变速不变调方法对所述待播放音频数据进行了加速或减速播放,若想实现同一采样时间点的视频数据和音频数据能够在播放环节同时呈现,则必须使用音视频同步机制。

[0126] 所述音视频同步机制指的是采用特定方法维系音频数据和视频数据之间的关联性,使所述音频数据和视频数据的偏移和抖动(不同步)维持在一定的误差范围内,从而进行同步播放,所述误差范围也称同步容限。常用的音视频同步技术主要包括:基于时间戳的同步技术、基于同步信道的同步技术、基于同步标记的同步技术、基于多路复用的同步技术、基于反馈的同步技术以及基于嵌入式编码的同步技术等。其中,所述基于时间戳的同步技术应用最为广泛,所述时间戳分为绝对时间戳和相对时间戳,所述绝对时间戳指的是在发送端对每一个音频数据和视频数据添加代表真实时间的的时间戳,接收端根据该时间信息准确的计算出每一个音频或视频呈现的起止时间;所述相对时间戳一般为发送端和接收端预先确定好时间戳的步长,接收端将该时间戳映射到本地时钟上,然后根据该系统本地时间戳来决定音频和视频的播放。

[0127] 然而,本实施例采用所述基于时间戳的同步技术中的视频时间戳跟踪音频时间戳的音视频同步方法,该方法同样为应用相对时间戳,所不同的是,该方法不依赖于本地系统时间戳,而是采用:从媒体的时间戳参考主媒体的时间戳,由于人的视觉灵敏度远低于听觉灵敏度,因此本实施例选用音频作为主媒体类型,视频作为从媒体类型,所述视频时间戳跟踪音频时间戳即为视频时间戳参考音频时间戳进行音视频网络播放,具体为:在发送端,所述音频数据和视频数据经同步采集以及编码后,在发送环节对相同时间点采集的音频数据和视频数据分别打上一样的时间戳,之后将具有相同时间戳的音频数据包和视频数据包同时发送到网络,在接收端,所述音频数据和视频数据经读取、分流以及解码等环节后,每一个音频数据和视频帧都有其对应的的时间戳,作为主媒体的音频数据在经所述变速不变调算法进行加速播放、减速播放或正常播放时,作为从媒体的视频帧需将自身的时间戳与所述音频数据的时间戳进行比较,如果视频时间戳落后于音频时间戳,则调整频率以加快所述视频帧渲染速度;如果视频时间戳与音频时间戳在同步容限内,则立即渲染所述视频帧;如

果视频时间戳超前于音频时间戳,则调整频率以减慢所述视频帧的渲染速度,以此实现音频与视频的同步播放。

[0128] 本申请所提供的音视频网络播放的方法,其尤其适用于音视频网络直播中,音视频网络直播对实时性和播放质量均有较高要求,例如,在淘宝直播中,在商品展示环节,如果出现由于网络延时和抖动所导致的播放中断、音视频丢弃等情况,可能造成关键信息的丢失,并且影响音视频播放的流畅性,因此对直播效果产生较大影响。而使用本申请所提供的音视频网络播放的方法,考虑到网络直播中音频播放相较于视频播放具有更高的敏感性,通过根据网络状况预先动态调整音频的播放速度、并使用音视频同步机制对视频同步进行渲染,可抵消网络延时和抖动带来的不利影响,从而保证音视频信息的相对完整性和音视频播放的流畅性,使得网络状况的变化对直播效果的影响最小化。

[0129] 本申请第二实施例提供一种音视频网络播放装置,请参考图3理解该实施例,图3为所述音视频网络播放装置的示意图。

[0130] 如图3所示,所述音视频网络播放装置包括:

[0131] 音视频数据接收单元201:用于接收音频数据和视频数据;

[0132] 音频播放模式确定单元202:用于根据所述音频数据和视频数据的接收状况确定音频播放模式;

[0133] 音视频数据解码单元203:用于对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧;

[0134] 音视频播放单元204:用于按所述确定的音频播放模式对所述待播放音频数据进行播放,并采用音视频同步机制对所述视频帧同步进行渲染。

[0135] 优选的,所述音频播放模式包括如下播放模式中的一种:加速播放、减速播放、正常播放;所述按所述确定的音频播放模式对所述待播放音频数据进行播放的方法包括:

[0136] 声音变速不变调的方法。

[0137] 优选的,所述声音变速不变调的方法为同步波形叠加法,所述按所述确定的音频播放模式对所述待播放音频数据进行播放,包括如下步骤:

[0138] 对所述待播放音频数据进行分析,获得该待播放音频数据的特征参数信息;

[0139] 对所述待播放音频数据进行语音活动检测,用以检测出所述待播放音频数据为有用声音或者为背景噪音;

[0140] 若所述待播放音频数据为有用声音,则对所述待播放音频数据进行降采样,并且,对所述待播放音频数据进行分解以及合成;若所述待播放音频数据为背景噪音,则利用上一音频数据对该背景噪音进行合并;

[0141] 输出所述待播放音频数据。

[0142] 优选的,所述对所述待播放音频数据进行降采样,包括:

[0143] 若所述待播放音频数据的采样率大于1khz,则降采样到1khz。

[0144] 优选的,所述音视频同步机制包括:采用基于时间戳的音视频同步技术中的视频时间戳跟踪音频时间戳的方法对音频和视频同步进行播放。

[0145] 优选的,所述音频数据和视频数据的接收状况,包括:

[0146] 接收的音频数据和视频数据的数量;和/或

[0147] 所述音频数据和视频数据的接收时间。

[0148] 优选的,当所述音频数据和视频数据的接收状况为接收的音频数据和视频数据的数量时,所述根据所述音频数据的接收状况确定音频播放模式,包括如下方式中的一种:

[0149] 接收的音频数据和视频数据的数量大于预定阈值,确定音频播放模式为加速播放模式;

[0150] 接收的音频数据和视频数据的数量小于预定阈值,确定音频播放模式为减速播放模式;

[0151] 接收的音频数据和视频数据的数量在预定阈值范围内,确定音频播放模式为正常播放模式。

[0152] 优选的,所述接收的音频数据和视频数据预先存储于抖动缓冲区Jitter buffer中。

[0153] 优选的,所述音视频网络播放的装置可应用于音视频网络直播中。

[0154] 本申请第三实施例提供一种电子设备,请参看图4,图4为该实施例的示意图。由于设备实施例基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0155] 本实施例提供的电子设备包括:处理器301以及存储器302,所述存储器302用于存储一种用于音视频网络播放的程序,该设备通电并通过所述处理器运行所述用于音视频网络播放的程序后,执行下述步骤:

[0156] 接收音频数据和视频数据;

[0157] 根据所述音频数据和视频数据的接收状况确定音频播放模式;

[0158] 对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧;

[0159] 按所述确定的音频播放模式对所述待播放音频数据进行播放,并采用音视频同步机制对所述视频帧同步进行渲染。

[0160] 可选的,所述音频播放模式包括加速播放、减速播放和正常播放中的一种;所述按所述确定的音频播放模式对所述待播放音频数据进行播放的方法包括:

[0161] 声音变速不变调的方法。

[0162] 可选的,所述声音变速不变调的方法为同步波形叠加法,所述按所述确定的音频播放模式对所述待播放音频数据进行播放,包括如下步骤:

[0163] 对所述待播放音频数据进行分析,获得该待播放音频数据的特征参数信息;

[0164] 对所述待播放音频数据进行语音活动检测,用以检测出所述待播放音频数据为有用声音或者为背景噪音;

[0165] 若所述待播放音频数据为有用声音,则对所述待播放音频数据进行降采样,并且,对所述待播放音频数据进行分解以及合成;若所述待播放音频数据为背景噪音,则利用上一音频数据对该背景噪音进行合并;

[0166] 输出所述待播放音频数据。

[0167] 可选的,所述对所述待播放音频数据进行降采样,包括:

[0168] 若所述待播放音频数据的采样率大于1khz,则降采样到1khz。

[0169] 可选的,所述音视频同步机制包括:

[0170] 采用基于时间戳的音视频同步技术中的视频时间戳跟踪音频时间戳的方法对音频和视频同步进行播放。

- [0171] 可选的,所述音频数据和视频数据的接收状况,包括:
- [0172] 接收的音频数据和视频数据的数量;和/或
- [0173] 所述音频数据和视频数据的接收时间。
- [0174] 可选的,当所述音频数据和视频数据的接收状况为接收的音频数据和视频数据的数量时,所述根据所述音频数据的接收状况确定音频播放模式,包括如下方式中的一种:
- [0175] 接收的音频数据和视频数据的数量大于预定阈值,确定音频播放模式为加速播放模式;
- [0176] 接收的音频数据和视频数据的数量小于预定阈值,确定音频播放模式为减速播放模式;
- [0177] 接收的音频数据和视频数据的数量在预定阈值范围内,确定音频播放模式为正常播放模式。
- [0178] 可选的,所述接收的音频数据和视频数据预先存储于抖动缓冲区Jitter buffer中。
- [0179] 可选的,所述音视频网络播放的方法可应用于音视频网络直播中。
- [0180] 本申请第四实施例提供一种音视频网络播放的方法,该方法与本申请第一实施例提供的方法所达到的技术效果一致,均为根据网络状况动态调整音视频缓存大小,通过对音频数据和视频数据的加速或减速播放来加速消耗或减缓消耗缓存,以抵消音频数据和视频数据的累积所造成的播放延时,以及避免网络抖动所引起的音频和视频播放的卡顿。本实施例与所述第一实施例的实现流程基本一致,仅在实现方式上略有差别,本实施例与所述第一实施例的相似处请参考第一实施例的相关内容进行理解,在此不再赘述。
- [0181] 请参考图5理解该实施例,图5为本实施例提供的音视频网络播放的方法的流程图。如图5所示,所述音视频网络播放的方法包括如下步骤:
- [0182] S201,接收音频数据和视频数据。
- [0183] 本步骤与第一实施例中的步骤S101的内容一致,请参考步骤S101理解本步骤。
- [0184] S202,根据所述音频数据和视频数据的接收状况,确定视频播放模式。
- [0185] 本步骤与第一实施例中的步骤S102基本一致,均为根据音频数据和视频数据的接收状况所反映出的当前网络状况设定播放模式,区别之处仅在于,本实施例中所设定的播放模式为视频播放模式,相应的,所述视频播放模式为加速播放、或减速播放,或正常播放。请参考步骤S102理解本步骤。
- [0186] S203,对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧。
- [0187] 本步骤与第一实施例的步骤S103的内容一致,请参考步骤S103理解本步骤。
- [0188] S204,按所述确定的视频播放模式,对所述视频帧进行播放,并采用音视频同步机制对所述待播放音频数据同步进行播放。
- [0189] 本步骤与第一实施例的步骤S104的区别之处在于:第一实施例的步骤S104首先播放的为待播放音频数据,而本步骤首先播放的是视频帧;并且,本实施例和第一实施例虽然都采用基于时间戳的音视频同步技术,然而,第一实施例的步骤S104的音视频同步机制指的是视频时间戳跟踪音频时间戳,而本步骤中所述音视频同步机制为音频时间戳跟踪视频时间戳。
- [0190] 本实施例中,所述按所述确定的视频播放模式,对所述视频帧进行播放的过程请



参考图6,图6为本实施例提供的按所述确定的视频播放模式对所述视频帧进行播放的流程图,如图6所示,该过程包括如下步骤:

[0191] S204-1,对所述视频帧进行识别,获得识别结果。

[0192] 本步骤的作用在于对所述视频帧按其重要性进行差异化识别,以获得所述视频帧的重要性等级。所述差异化识别的识别依据为:站在用户角度来说,在一段视频中,不同的视频画面,其重要性有所不同,例如,对于球赛播放来说,进球画面的重要性明显高于赛事的其它播放画面,如场外状况画面或观众席画面;再例如,在新闻播放、娱乐综艺播放或电影播放等播放场景中,观众对于人物面部特写、明星动态或其他刺激性画面的关注度往往较高,因此,可认为该类视频画面为重要性最高的视频画面,依次类推,可获得重要性较高和重要性不高的视频画面。

[0193] 本实施例中,所述对所述视频帧进行识别的方法为:采用基于机器学习的方法对所述视频帧进行识别,具体为:将所述视频帧输入视频图像集识别模型中进行匹配,以识别出所述视频帧的重要性等级。

[0194] 所述视频图像集识别模型的构建过程一般包括如下流程:

[0195] 第一,获取视频图像集基准数据。所述图像集基准数据为各种视频播放场景的采集画面。

[0196] 第二,对所述视频图像集基准数据进行预处理,以获取视频图像训练集和视频图像测试集。该过程为对所述视频播放场景的采集画面按其重要性进行等级划分以及等级标记,并将所述进行等级标记后的所述视频播放场景的采集画面分为视频图像训练集和视频图像测试集,所述视频图像训练集被用于训练生成视频图像集识别模型,所述视频图像测试集被用于评估所述视频图像集识别模型的识别性能。

[0197] 第三,对所述视频图像训练集进行训练,以生成视频图像集初始识别模型,同时,利用所述视频图像测试集对所述视频图像集初始识别模型的识别效果进行评估,根据评估结果确定所述视频图像集识别模型、或进一步改善所述视频图像集初始识别模型。所述对所述视频图像训练集进行训练的方式可采用基于机器学习的分类方法中的决策树、贝叶斯、关联规则学习、以及神经网络等算法中的任意一种实现。

[0198] 上述构建视频图像集识别模型的方法以及将所述视频帧输入视频图像集识别模型中进行匹配的方法均为现有技术内容,具体实现方式请参考现有技术的相关知识,在此不作过多说明。

[0199] 经上述过程的识别匹配后,可获得所述视频帧的重要性等级标记信息。

[0200] S204-2,按所述确定的视频播放模式、结合所述识别结果对所述视频帧进行加速播放、减速播放、或正常播放。

[0201] 在所述步骤S202中已确定了所述视频播放模式为加速播放、或减速播放、或正常播放,本步骤的作用在于,在按照所述已确定的视频播放模式所规定的播放模式信息对视频进行播放的基础上,对所述视频帧的播放速度按所述步骤S204-1所获得的所述视频帧的重要性等级标记信息进行调节,使得所述视频帧所在的视频段落按照上述播放模式信息所要求的播放速率进行播放的同时,所述视频帧可作加快或减慢调整,以使所述视频帧的呈现效果更符合当前的播放场景和观众的观看感受。

[0202] 所述对所述视频帧的播放速度按所述步骤S204-1所获得的所述视频帧的重要性

等级标记信息进行调节,具体为:当所述视频帧的重要性等级为高时,该视频帧的播放速度应适当减慢,以使该视频帧可清楚呈现,起到了类似“画面特写”的功效;当所述视频帧的重要性等级为低时,该视频帧的播放速度应适当加快,在加速播放的同时,起到了滤除多余画面或者对次要画面进行快进播放的作用。

[0203] 需要说明的是,本实施例中,上述对所述视频帧的加快或减慢操作,是相对于所述播放模式信息的加速播放或减速播放而言的,即,在所述播放模式所确定的加速播放或减速播放的播放速率的基础上进行加快或减慢,在一段视频中,有需要加快的视频帧,也有需要减慢或正常播放的视频帧,所述加快或减慢播放的结果为,最终所述视频段落的播放速率应与所述播放模式所确定的播放速率相一致。

[0204] 本实施例中,所述采用音视频同步机制对所述待播放音频数据同步进行播放,其实现过程为:基于时间戳的同步技术中的音频时间戳跟踪视频时间戳的音视频同步方法,其采用从媒体的时间戳参考主媒体的时间戳,本实施例从视频画面的重要性等级出发,将视频作为主媒体,音频作为从媒体,所述音频时间戳跟踪视频时间戳即为音频时间戳参考视频时间戳进行音视频网络播放,具体为:在发送端,所述音频数据和视频数据经同步采集以及编码后,在发送环节对相同时间点采集的音频数据和视频数据分别打上一样的时间戳,之后将具有相同时间戳的音频数据包和视频数据包同时发送到网络,在接收端,所述音频数据和视频数据经读取、分流以及解码等环节后,每一个音频数据和视频帧都有其对应的时间戳,作为主媒体的视频帧在进行加速播放、减速播放或正常播放时,作为从媒体的音频数据需将自身的时间戳与所述视频帧的时间戳进行比较,如果音频时间戳落后于视频时间戳,则加快播放所述音频数据;如果视频时间戳与音频时间戳在同步容限内,则立即播放所述音频数据;如果音频时间戳超前于视频时间戳,则减慢播放所述音频数据,以此实现音频与视频的同步播放。本实施例中,对所述音频数据的播放采用声音变速不变调的方法,有关所述声音变速不变调的说明,请参考第一实施例的相关内容,在此不再赘述。

[0205] 与上述第四实施例提供的音视频网络播放的方法相对应,本申请第五实施例提供一种音视频网络播放的装置,请参考图7理解该实施例,图7为本实施例提供的装置示意图。

[0206] 如图7所示,所述音视频网络播放的装置包括:

[0207] 音视频数据接收单元401:用于接收音频数据和视频数据;

[0208] 视频播放模式确定单元402:用于根据所述音频数据和视频数据的接收状况,确定视频播放模式;

[0209] 音视频数据解码单元403:用于对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧;

[0210] 音视频播放单元404:用于按所述确定的视频播放模式对所述视频帧进行播放,并采用音视频同步机制对所述待播放音频数据同步进行播放。

[0211] 可选的,所述视频播放模式包括加速播放、减速播放和正常播放中的一种;所述按所述确定的视频播放模式,对所述视频帧进行播放,包括:

[0212] 对所述视频帧进行识别,获得识别结果;

[0213] 按所述确定的视频播放模式、结合所述识别结果对所述视频帧进行加速播放、减速播放、或正常播放。

[0214] 可选的,对所述视频帧进行识别的方法包括:

[0215] 采用基于机器学习的方法。

[0216] 可选的,所述音视频同步机制包括:

[0217] 采用基于时间戳的音视频同步技术中的音频时间戳跟踪视频时间戳的方法对音频和视频同步进行播放。

[0218] 可选的,所述对所述待播放音频数据同步进行播放,其播放方式包括:

[0219] 采用声音变速不变调方法。

[0220] 本申请第六实施例提供一种电子设备,请参看图8,图8为该实施例的示意图。由于设备实施例基本类似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。下述描述的设备实施例仅仅是示意性的。

[0221] 本实施例提供的电子设备包括:处理器501以及存储器502,所述存储器502用于存储一种用于音视频网络播放的程序,该设备通电并通过所述处理器运行所述用于音视频网络播放的程序后,执行下述步骤:

[0222] 接收音频数据和视频数据;

[0223] 根据所述音频数据和视频数据的接收状况,确定视频播放模式;

[0224] 对所述音频数据和视频数据进行解码,生成待播放音频数据和视频帧;

[0225] 按所述确定的视频播放模式,对所述视频帧进行播放,并采用音视频同步机制对所述待播放音频数据同步进行播放。

[0226] 可选的,所述视频播放模式包括加速播放、减速播放和正常播放中的一种;所述按所述确定的视频播放模式,对所述视频帧进行播放,包括:

[0227] 对所述视频帧进行识别,获得识别结果;

[0228] 按所述确定的视频播放模式、结合所述识别结果对所述视频帧进行加速播放、减速播放、或正常播放。

[0229] 可选的,对所述视频帧进行识别的方法包括:

[0230] 采用基于机器学习的方法。

[0231] 可选的,所述音视频同步机制包括:

[0232] 采用基于时间戳的音视频同步技术中的音频时间戳跟踪视频时间戳的方法对音频和视频同步进行播放。

[0233] 可选的,所述对所述待播放音频数据同步进行播放,其播放方式包括:

[0234] 采用声音变速不变调方法。

[0235] 本发明虽然以较佳实施例公开如上,但其并不是用来限定本发明,任何本领域技术人员在不脱离本发明的精神和范围内,都可以做出可能的变动和修改,因此本发明的保护范围应当以本发明权利要求所界定的范围为准。

[0236] 在一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0237] 内存可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0238] 1、计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数

据。计算机的存储介质的例子包括,但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器 (CD-ROM)、数字多功能光盘 (DVD) 或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括非暂存电脑可读媒体 (transitory media),如调制的数据信号和载波。

[0239] 2、本领域技术人员应明白,本申请的实施例可提供为方法、系统或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

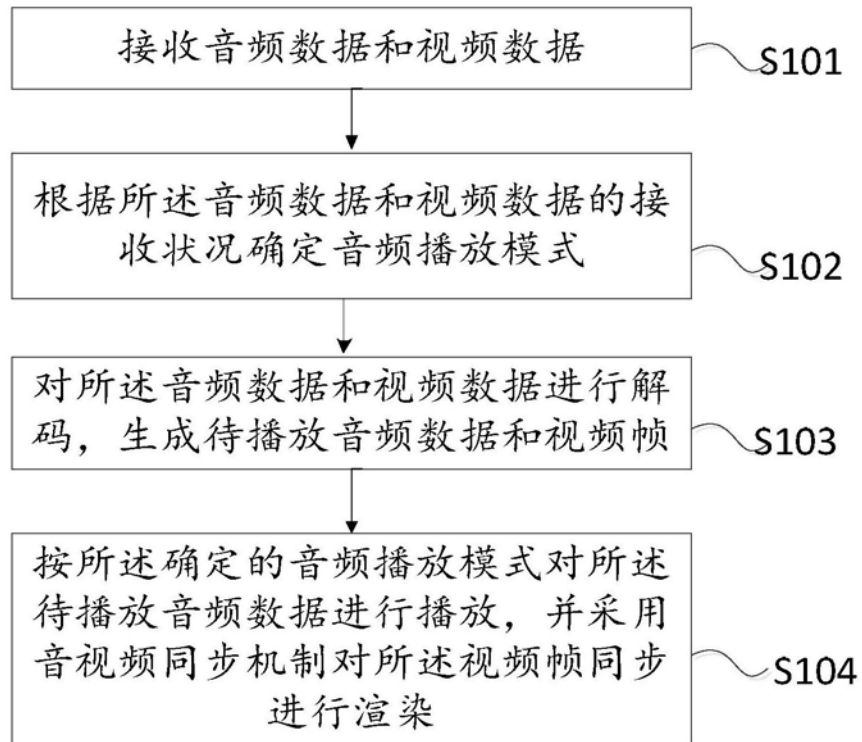


图1

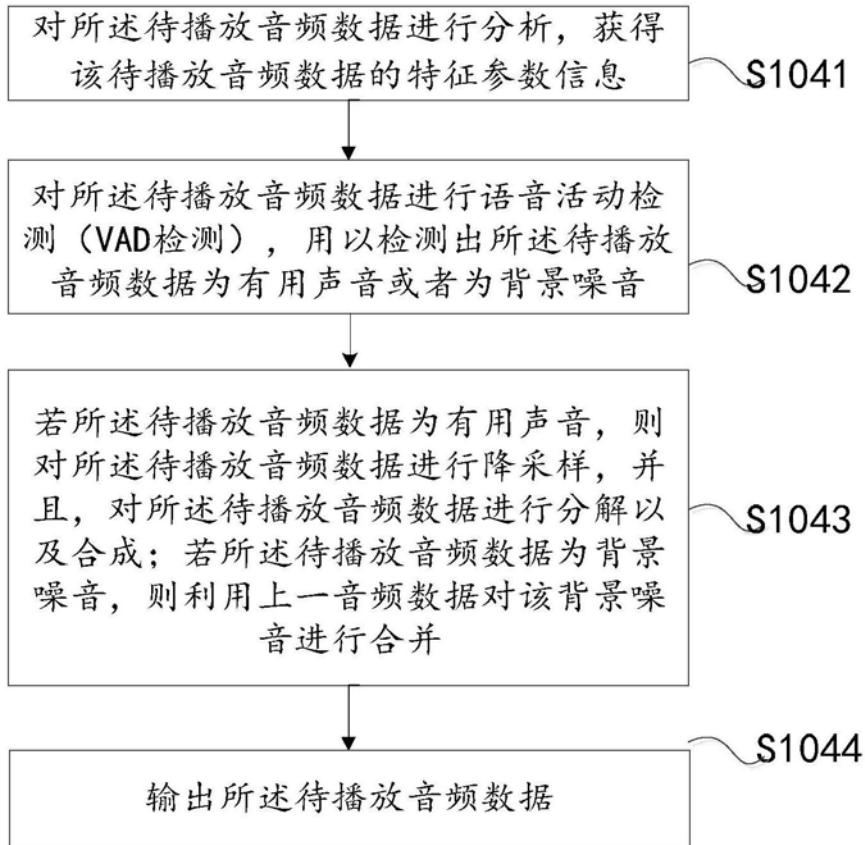


图2

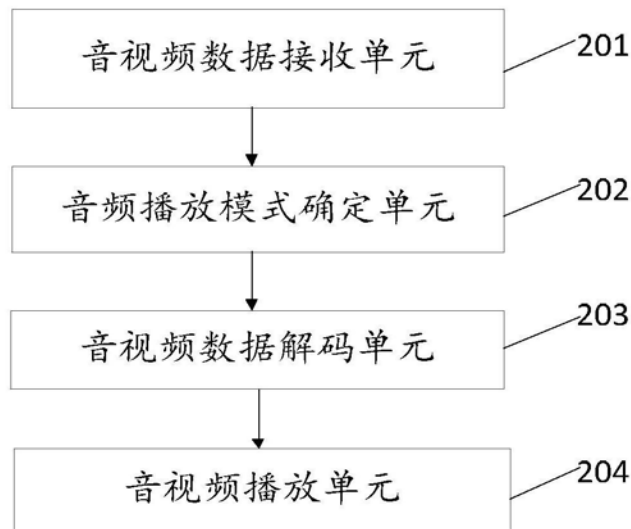


图3

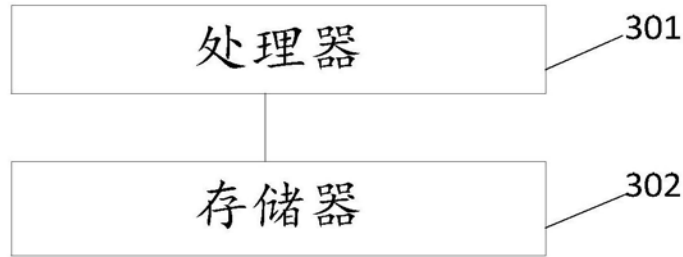


图4

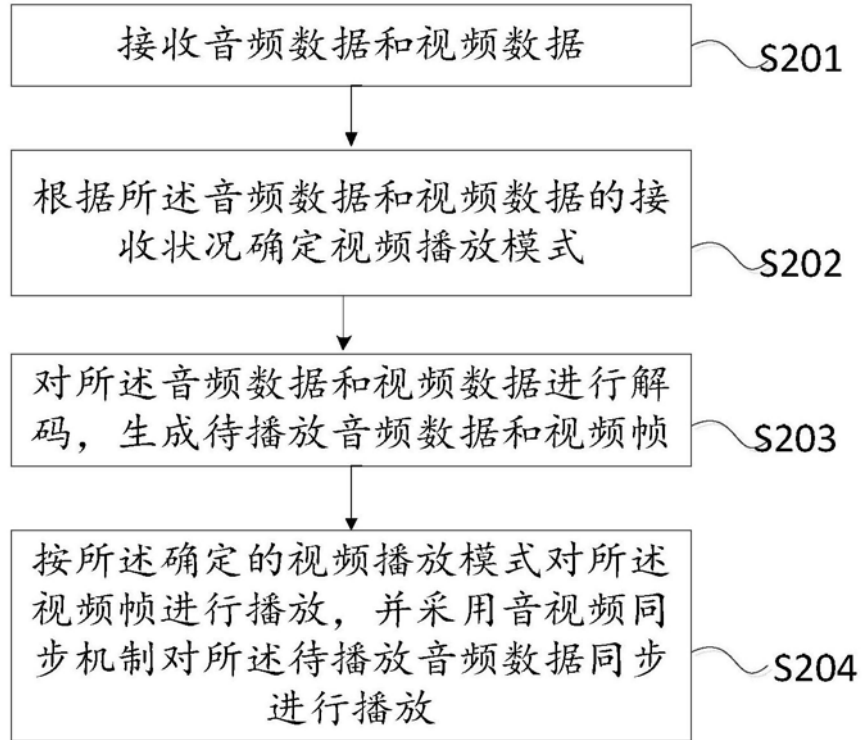


图5

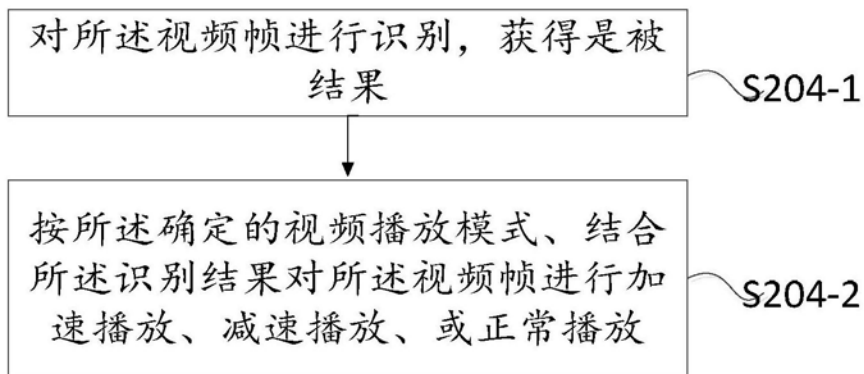


图6

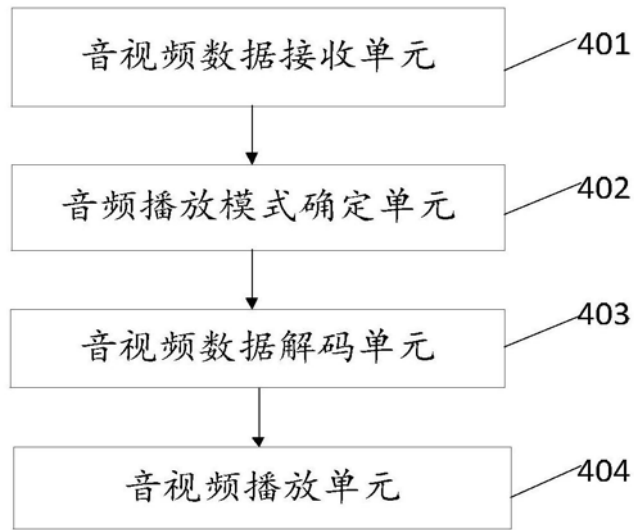


图7

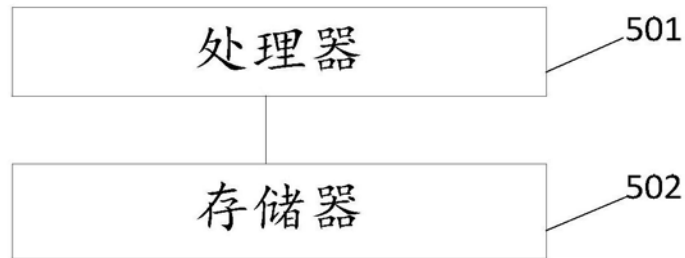


图8