



(12)发明专利申请

(10)申请公布号 CN 110730360 A

(43)申请公布日 2020.01.24

(21)申请号 201911023915.3

(22)申请日 2019.10.25

(71)申请人 北京达佳互联信息技术有限公司
地址 100085 北京市海淀区上地西路6号1
幢1层101D1-7

(72)发明人 王聪

(74)专利代理机构 北京柏杉松知识产权代理事
务所(普通合伙) 11413
代理人 项京 高莺然

(51)Int.Cl.

H04N 21/233(2011.01)

H04N 21/439(2011.01)

H04N 21/44(2011.01)

G10L 15/26(2006.01)

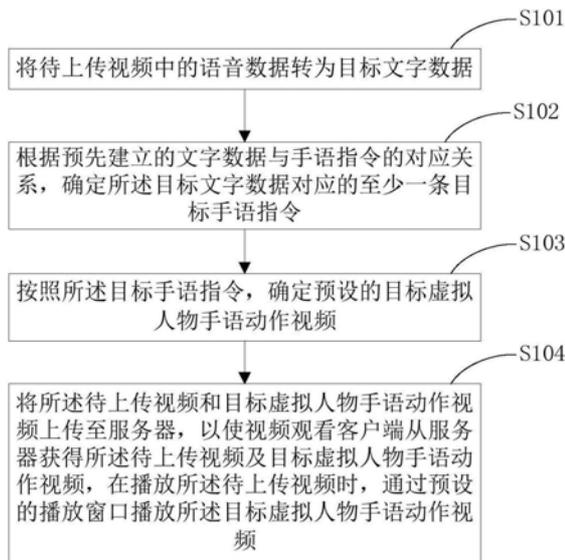
权利要求书2页 说明书21页 附图8页

(54)发明名称

视频上传、播放的方法、装置、客户端设备及
存储介质

(57)摘要

本公开关于视频上传、播放的方法、装置、设备
及存储介质,视频上传的方法应用于视频生成
客户端,可以包括:将待上传视频中的语音数据
转为目标文字数据;根据预先建立的文字数据与
手语指令的对应关系,确定目标文字数据对应
的至少一条目标手语指令;按照目标手语指令,
确定预设的目标虚拟人物手语动作视频;将待
上传视频和目标虚拟人物手语动作视频上传至
服务器,以使视频观看客户端,在播放待上传
视频时,通过预设的播放窗口播放目标虚拟人
物手语动作视频。可见,在本公开实施例中,听
障人士可以在观看直播或短视频等形式的视频
时,观看到目标虚拟人物手语动作视频,能够
有效获取信息,提高了听障人士的观看体验。



1. 一种视频上传的方法,其特征在于,应用于视频生成客户端设备,包括:
 - 将待上传视频中的语音数据转为目标文字数据;
 - 根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;
 - 按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频;
 - 将所述待上传视频和目标虚拟人物手语动作视频上传至服务器,以使视频观看客户端从服务器获得所述待上传视频及目标虚拟人物手语动作视频,在播放所述待上传视频时,通过预设的播放窗口播放所述目标虚拟人物手语动作视频。
2. 根据权利要求1所述的方法,其特征在于,所述按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频的步骤,包括:
 - 根据待上传视频中语音数据对应的播放时间,确定各个目标手语指令与所述待上传视频中视频帧画面的对应关系;
 - 基于各个目标手语指令、各个目标手语指令与所述待上传视频中视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频。
3. 一种视频播放的方法,其特征在于,应用于视频观看客户端设备,包括:
 - 从服务器获得待播放视频及目标虚拟人物手语动作视频;所述目标虚拟人物手语动作视频,为:视频生成客户端,将待上传视频中的语音数据转为目标文字数据;根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;按照所述目标手语指令和预设的虚拟人物确定的;
 - 在播放所述待播放视频时显示预设的播放窗口;
 - 在所述播放窗口中播放所述目标虚拟人物手语动作视频。
4. 一种视频播放的方法,其特征在于,应用于视频观看客户端设备,包括:
 - 将待播放视频中的语音数据转为目标文字数据;
 - 根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;
 - 按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频;
 - 在播放所述待播放视频时,显示预设的播放窗口;
 - 在所述播放窗口中播放所述目标虚拟人物手语动作视频。
5. 一种视频上传的装置,其特征在于,应用于视频生成客户端设备,包括:
 - 第一目标文字数据转换单元,被配置为将待上传视频中的语音数据转为目标文字数据;
 - 第一目标手语指令确定单元,被配置为根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;
 - 第一手语动作视频确定单元,被配置为按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频;
 - 上传单元,被配置为将所述待上传视频和目标虚拟人物手语动作视频上传至服务器,以使视频观看客户端从服务器获得所述待上传视频及目标虚拟人物手语动作视频,在播放所述待上传视频时,通过预设的播放窗口播放所述目标虚拟人物手语动作视频。
6. 根据权利要求5所述的装置,其特征在于,所述第一手语动作视频确定单元,包括:

第一对应关系确定模块,被配置为根据待上传视频中语音数据对应的播放时间,确定各个目标手语指令与所述待上传视频中视频帧画面的对应关系;

第一手语动作视频生成模块,被配置为基于各个目标手语指令、各个目标手语指令与所述待上传视频中视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频。

7. 一种视频播放的装置,其特征在于,应用于视频观看客户端设备,包括:

第一手语动作视频获得单元,被配置为从服务器获得待播放视频及目标虚拟人物手语动作视频;所述目标虚拟人物手语动作视频,为:视频生成客户端,将待上传视频中的语音数据转为目标文字数据;根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;按照所述目标手语指令和预设的虚拟人物确定的;

第一播放窗口显示单元,被配置为在播放所述待播放视频时显示预设的播放窗口;

第一手语动作视频播放单元,被配置为在所述播放窗口中播放所述目标虚拟人物手语动作视频。

8. 一种视频播放的装置,其特征在于,应用于视频观看客户端设备,包括:

第二目标文字数据转换单元,被配置为将待播放视频中的语音数据转为目标文字数据;

第二目标手语指令确定单元,被配置为根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;

第二手语动作视频确定单元,被配置为按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频;

第二播放窗口显示单元,被配置为在播放所述待播放视频时,显示预设的播放窗口;

第二手语动作视频播放单元,被配置为在所述播放窗口中播放所述目标虚拟人物手语动作视频。

9. 一种客户端设备,其特征在于,包括处理器、通信接口、存储器和通信总线,其中,处理器,通信接口,存储器通过通信总线完成相互间的通信;

存储器,用于存放计算机程序;

处理器,用于执行存储器上所存放的程序时,实现权利要求1-2任一所述的方法步骤;或实现权利要求3所述的方法步骤;或实现权利要求4所述的方法步骤。

10. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质内存储有计算机程序,所述计算机程序被处理器执行时实现权利要求1-2任一所述的方法步骤;或实现权利要求3所述的方法步骤;或实现权利要求4所述的方法步骤。

视频上传、播放的方法、装置、客户端设备及存储介质

技术领域

[0001] 本公开涉及视频上传、视频播放的技术领域,尤其涉及视频上传、视频播放的方法、装置、客户端设备及存储介质。

背景技术

[0002] 随着视频自媒体的迅速发展,视频直播和短视频等形式的自媒体成为传递信息的主要工具。但是听障人士由于患有残疾性听力损失,在观看视频时,无法有效的从视频直播和短视频等形式的自媒体中获取信息,听障人士的观看体验较低。

发明内容

[0003] 本公开提供视频上传、视频播放的方法、装置、客户端设备及存储介质,以至少解决相关技术中听障人士无法有效的从直播和短视频等形式的自媒体中获取信息,听障人士的观看体验较低的问题。本公开的技术方案如下:

[0004] 根据本公开实施例的第一方面,提供一种视频上传的方法,应用于视频生成客户端设备,包括:

[0005] 将待上传视频中的语音数据转为目标文字数据;

[0006] 根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;

[0007] 按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频;

[0008] 将所述待上传视频和目标虚拟人物手语动作视频上传至服务器,以使视频观看客户端从服务器获得所述待上传视频及目标虚拟人物手语动作视频,在播放所述待上传视频时,通过预设的播放窗口播放所述目标虚拟人物手语动作视频。

[0009] 可选的,所述按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频的步骤,包括:

[0010] 根据待上传视频中语音数据对应的播放时间,确定各个目标手语指令与所述待上传视频中视频帧画面的对应关系;

[0011] 基于各个目标手语指令、各个目标手语指令与所述待上传视频中视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频。

[0012] 可选的,所述将待上传视频中的语音数据转为目标文字数据的步骤,包括:

[0013] 获取待上传视频;

[0014] 从所述待上传视频中获取语音数据作为目标语音数据;

[0015] 将所述目标语音数据转为目标文字数据。

[0016] 可选的,在所述从所述待上传视频中获取语音数据作为目标语音数据的步骤之前,还包括:

[0017] 判断当前是否为上传虚拟人物手语动作视频的上传状态;所述上传状态,包括上传虚拟人物手语动作视频状态,和非上传虚拟人物手语动作视频状态,由用户预先设置;

- [0018] 如果是,则执行所述从所述待上传视频中获取语音数据作为目标语音数据的步骤。
- [0019] 可选的,所述从所述待上传视频中获取语音数据作为目标语音数据的步骤,包括:
- [0020] 将所述待上传视频中的音频数据实时转换为非压缩的波形文件,输入到预先训练好的神经网络模型;
- [0021] 获取所述神经网络模型预测出的目标波形文件;所述神经网络模型,用于对非压缩的波形文件进行切割,得到语音数据中各个语句对应的目标波形文件作为目标语音数据。
- [0022] 可选的,所述将所述目标语音数据转为目标文字数据的步骤,包括:
- [0023] 将所述目标波形文件,按照预设的切分规则切分成各个目标帧文件;
- [0024] 对所述各个目标帧文件提取声学特征,获得所述各个目标帧文件的各个目标多维向量;
- [0025] 基于预设的多维向量与文字数据之间的对应关系,将所述各个目标多维向量转换为目标文字数据。
- [0026] 可选的,所述文字数据与手语指令的对应关系的建立过程包括:
- [0027] 获得基于手语规则设置的每个手语词汇对应的手语指令;
- [0028] 获得设置的虚拟人物形象以及每个手语指令对应的虚拟人物形象手部的动作轨迹;
- [0029] 将各个手语词汇对应的文字数据与手语指令对应保存。
- [0030] 可选的,所述根据待上传视频中语音数据对应的播放时间,确定各个目标手语指令与所述待上传视频中视频帧画面的对应关系的步骤,包括:
- [0031] 根据待上传视频中语音数据对应的播放时间,确定各个目标手语指令对应的虚拟人物形象手部的动作轨迹与所述待上传视频中视频帧画面的对应关系。
- [0032] 可选的,所述基于各个目标手语指令、各个目标手语指令与所述待上传视频中视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频的步骤,包括:
- [0033] 基于所述虚拟人物形象、所述目标手语指令对应的虚拟人物形象手部的动作轨迹、所述虚拟人物形象手部的动作轨迹与所述待上传视频中视频帧画面的对应关系,生成目标虚拟人物手语动作视频。
- [0034] 根据本公开实施例的第二方面,提供一种视频播放的方法,应用于视频观看客户端设备,包括:
- [0035] 从服务器获得待播放视频及目标虚拟人物手语动作视频;所述目标虚拟人物手语动作视频,为:视频生成客户端,将待上传视频中的语音数据转为目标文字数据;根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;按照所述目标手语指令和预设的虚拟人物确定的;
- [0036] 在播放所述待播放视频时显示预设的播放窗口;
- [0037] 在所述播放窗口中播放所述目标虚拟人物手语动作视频。
- [0038] 可选的,所述方法还包括:
- [0039] 检测是否接收到关闭所述目标虚拟人物手语动作视频的指令;
- [0040] 如果是,则关闭所述预设的播放窗口,停止播放所述目标虚拟人物手语动作视频。

- [0041] 根据本公开实施例的第三方面,提供一种视频播放的方法,应用于视频观看客户端设备,包括:
- [0042] 将待播放视频中的语音数据转为目标文字数据;
- [0043] 根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;
- [0044] 按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频;
- [0045] 在播放所述待播放视频时,显示预设的播放窗口;
- [0046] 在所述播放窗口中播放所述目标虚拟人物手语动作视频。
- [0047] 可选的,所述按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频的步骤,包括:
- [0048] 根据待播放视频中语音数据对应的播放时间,确定各个目标手语指令与所述待播放视频中视频帧画面的对应关系;
- [0049] 基于各个目标手语指令、各个目标手语指令与所述待播放视频中视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频。
- [0050] 可选的,所述将待播放视频中的语音数据转为目标文字数据的步骤,包括:
- [0051] 获得待播放视频;
- [0052] 从所述待播放视频中获取语音数据作为目标语音数据;
- [0053] 将所述目标语音数据转为目标文字数据。
- [0054] 可选的,在所述从所述待播放视频中获取语音数据作为目标语音数据的步骤之前,还包括:
- [0055] 判断当前是否为播放虚拟人物手语动作视频的播放状态;所述播放状态,包括播放虚拟人物手语动作视频状态,和非播放虚拟人物手语动作视频状态;
- [0056] 如果是,则执行所述从所述待播放视频中获取语音数据作为目标语音数据的步骤。
- [0057] 可选的,所述从所述待播放视频中获取语音数据作为目标语音数据的步骤,包括:
- [0058] 将所述待播放视频中的音频数据实时转换为非压缩的波形文件,输入到预先训练好的神经网络模型;
- [0059] 获取所述神经网络模型预测出的目标波形文件;所述神经网络模型,用于对非压缩的波形文件进行切割,得到语音数据中各个语句对应的目标波形文件作为目标语音数据。
- [0060] 可选的,所述将所述目标语音数据转为目标文字数据的步骤,包括:
- [0061] 将所述目标波形文件,按照预设的切分规则切分成各个目标帧文件;
- [0062] 对所述各个目标帧文件提取声学特征,获得所述各个目标帧文件的各个目标多维向量;
- [0063] 基于预设的多维向量与文字数据之间的对应关系,将所述各个目标多维向量转换为目标文字数据。
- [0064] 可选的,所述文字数据与手语指令的对应关系的建立过程包括:
- [0065] 获得基于手语规则设置的每个手语词汇对应的手语指令;
- [0066] 获得设置的虚拟人物形象以及每个手语指令对应的虚拟人物形象手部的动作轨

迹；

[0067] 将各个手语词汇对应的文字数据与手语指令对应保存。

[0068] 可选的,所述根据待播放视频中语音数据对应的播放时间,确定各个目标手语指令与所述待播放视频中视频帧画面的对应关系的步骤,包括:

[0069] 根据待播放视频中语音数据对应的播放时间,确定各个目标手语指令对应的虚拟人物形象手部的动作轨迹与所述待播放视频中视频帧画面的对应关系。

[0070] 可选的,所述基于各个目标手语指令、各个目标手语指令与所述待播放视频中视频帧画面的对应关系及预设的虚拟人物形象,生成目标虚拟人物手语动作视频的步骤,包括:

[0071] 基于所述虚拟人物形象、所述目标手语指令对应的虚拟人物形象手部的动作轨迹、所述虚拟人物形象手部的动作轨迹与所述待播放视频中视频帧画面的对应关系,生成目标虚拟人物手语动作视频。

[0072] 根据本公开实施例的第四方面,提供一种视频上传的装置,应用于视频生成客户端设备,包括:

[0073] 第一目标文字数据转换单元,被配置为将待上传视频中的语音数据转为目标文字数据;

[0074] 第一目标手语指令确定单元,被配置为根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;

[0075] 第一手语动作视频确定单元,被配置为按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频;

[0076] 上传单元,被配置为将所述待上传视频和目标虚拟人物手语动作视频上传至服务器,以使视频观看客户端从服务器获得所述待上传视频及目标虚拟人物手语动作视频,在播放所述待上传视频时,通过预设的播放窗口播放所述目标虚拟人物手语动作视频。

[0077] 可选的,所述第一手语动作视频确定单元,包括:

[0078] 第一对应关系确定模块,被配置为根据待上传视频中语音数据对应的播放时间,确定各个目标手语指令与所述待上传视频中视频帧画面的对应关系;

[0079] 第一手语动作视频生成模块,被配置为基于各个目标手语指令、各个目标手语指令与所述待上传视频中视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频。

[0080] 可选的,所述第一目标文字数据转换单元,包括:

[0081] 视频获取模块,被配置为获取待上传视频;

[0082] 第一目标语音数据获取模块,被配置为从所述待上传视频中获取语音数据作为目标语音数据;

[0083] 第一目标文字数据转换模块,被配置为将所述目标语音数据转为目标文字数据。

[0084] 可选的,所述装置,还包括:

[0085] 第一判断模块,被配置为在所述第一目标语音数据获取模块目标语音数据获取模块从所述待上传视频中获取语音数据作为目标语音数据之前,判断当前是否为上传虚拟人物手语动作视频的上传状态;所述上传状态,包括上传虚拟人物手语动作视频状态,和非上传虚拟人物手语动作视频状态,由用户预先设置;如果是,则执行所述从所述待上传视频中

获取语音数据作为目标语音数据的步骤。

[0086] 可选的,所述第一目标语音数据获取模块,具体被配置为:

[0087] 将所述待上传视频中的音频数据实时转换为非压缩的波形文件,输入到预先训练好的神经网络模型;

[0088] 获取所述神经网络模型预测出的目标波形文件;所述神经网络模型,用于对非压缩的波形文件进行切割,得到语音数据中各个语句对应的目标波形文件作为目标语音数据。

[0089] 可选的,所述第一目标文字数据转换模块,具体被配置为:

[0090] 将所述目标波形文件,按照预设的切分规则切分成各个目标帧文件;

[0091] 对所述各个目标帧文件提取声学特征,获得所述各个目标帧文件的各个目标多维向量;

[0092] 基于预设的多维向量与文字数据之间的对应关系,将所述各个目标多维向量转换为目标文字数据。

[0093] 可选的,所述装置,还包括:第一对应关系建立单元;

[0094] 所述第一对应关系建立单元,具体被配置为:

[0095] 获得基于手语规则设置的每个手语词汇对应的手语指令;

[0096] 获得设置的虚拟人物形象以及每个手语指令对应的虚拟人物形象手部的动作轨迹;

[0097] 将各个手语词汇对应的文字数据与手语指令对应保存。

[0098] 可选的,所述第一对应关系确定模块,具体被配置为:

[0099] 根据待上传视频中语音数据对应的播放时间,确定各个目标手语指令对应的虚拟人物形象手部的动作轨迹与所述待上传视频中视频帧画面的对应关系。

[0100] 可选的,所述第一手语动作视频生成模块,具体被配置为:

[0101] 基于所述虚拟人物形象、所述目标手语指令对应的虚拟人物形象手部的动作轨迹、所述虚拟人物形象手部的动作轨迹与所述待上传视频中视频帧画面的对应关系,生成目标虚拟人物手语动作视频。

[0102] 根据本公开实施例的第五方面,提供一种视频播放的装置,应用于视频观看客户端设备,包括:

[0103] 第一手语动作视频获得单元,被配置为从服务器获得待播放视频及目标虚拟人物手语动作视频;所述目标虚拟人物手语动作视频,为:视频生成客户端,将待上传视频中的语音数据转为目标文字数据;根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;按照所述目标手语指令和预设的虚拟人物确定的;

[0104] 第一播放窗口显示单元,被配置为在播放所述待播放视频时显示预设的播放窗口;

[0105] 第一手语动作视频播放单元,被配置为在所述播放窗口中播放所述目标虚拟人物手语动作视频。

[0106] 可选的,所述装置还包括:

[0107] 第一手语动作视频关闭指令检测单元,被配置为检测是否接收到关闭所述目标虚

拟人物手语动作视频的指令；如果是，则关闭所述预设的播放窗口，停止播放所述目标虚拟人物手语动作视频。

[0108] 根据本公开实施例的第六方面，提供一种视频播放的装置，应用于视频观看客户端设备，包括：

[0109] 第二目标文字数据转换单元，被配置为将待播放视频中的语音数据转为目标文字数据；

[0110] 第二目标手语指令确定单元，被配置为根据预先建立的文字数据与手语指令的对应关系，确定所述目标文字数据对应的至少一条目标手语指令；

[0111] 第二手语动作视频确定单元，被配置为按照所述目标手语指令，确定预设的目标虚拟人物手语动作视频；

[0112] 第二播放窗口显示单元，被配置为在播放所述待播放视频时，显示预设的播放窗口；

[0113] 第二手语动作视频播放单元，被配置为在所述播放窗口中播放所述目标虚拟人物手语动作视频。

[0114] 可选的，所述第二手语动作视频确定单元，包括：

[0115] 第二对应关系确定模块，被配置为根据待播放视频中语音数据对应的播放时间，确定各个目标手语指令与所述待播放视频中视频帧画面的对应关系；

[0116] 第二手语动作视频生成模块，被配置为基于各个目标手语指令、各个目标手语指令与所述待播放视频中视频帧画面的对应关系及预设的虚拟人物，生成目标虚拟人物手语动作视频。

[0117] 可选的，所述第二目标文字数据转换单元，包括：

[0118] 视频获得模块，被配置为获得待播放视频；

[0119] 第二目标语音数据获取模块，被配置为从所述待播放视频中获取语音数据作为目标语音数据；

[0120] 第二目标文字数据转换模块，被配置为将所述目标语音数据转为目标文字数据。

[0121] 可选的，所述装置，还包括：

[0122] 第二判断模块，被配置为在所述第二目标语音数据获取模块从所述待播放视频中获取语音数据作为目标语音数据之前，判断当前是否为播放虚拟人物手语动作视频的播放状态；所述播放状态，包括播放虚拟人物手语动作视频状态，和非播放虚拟人物手语动作视频状态；如果是，则执行所述从所述待播放视频中获取语音数据作为目标语音数据的步骤。

[0123] 可选的，所述第二目标语音数据获取模块，具体被配置为：

[0124] 将所述待播放视频中的音频数据实时转换为非压缩的波形文件，输入到预先训练好的神经网络模型；

[0125] 获取所述神经网络模型预测出的目标波形文件；所述神经网络模型，用于对非压缩的波形文件进行切割，得到语音数据中各个语句对应的目标波形文件作为目标语音数据。

[0126] 可选的，所述第二目标文字数据转换模块，具体被配置为：

[0127] 将所述目标波形文件，按照预设的切分规则切分成各个目标帧文件；

[0128] 对所述各个目标帧文件提取声学特征，获得所述各个目标帧文件的各个目标多维

向量；

[0129] 基于预设的多维向量与文字数据之间的对应关系，将所述各个目标多维向量转换为目标文字数据。

[0130] 可选的，所述装置，还包括：第二对应关系建立单元；

[0131] 所述第二对应关系建立单元，具体被配置为：

[0132] 获得基于手语规则设置的每个手语词汇对应的手语指令；

[0133] 获得设置的虚拟人物形象以及每个手语指令对应的虚拟人物形象手部的动作轨迹；

[0134] 将各个手语词汇对应的文字数据与手语指令对应保存。

[0135] 可选的，所述第二对应关系确定模块，具体被配置为：

[0136] 根据待播放视频中语音数据对应的播放时间，确定各个目标手语指令对应的虚拟人物形象手部的动作轨迹与所述待播放视频中视频帧画面的对应关系。

[0137] 可选的，所述第二手语动作视频生成模块，具体被配置为：

[0138] 基于所述虚拟人物形象、所述目标手语指令对应的虚拟人物形象手部的动作轨迹、所述虚拟人物形象手部的动作轨迹与所述待播放视频中视频帧画面的对应关系，生成目标虚拟人物手语动作视频。

[0139] 根据本公开实施例的第七方面，提供了一种电子设备，包括处理器、通信接口、存储器和通信总线，其中，处理器，通信接口，存储器通过通信总线完成相互间的通信；

[0140] 存储器，用于存放计算机程序；

[0141] 处理器，用于执行存储器上所存放的程序时，实现上述视频上传任一所述的方法步骤；或实现上述视频播放任一所述的方法步骤。

[0142] 根据本公开实施例的第八方面，提供了一种计算机可读存储介质，所述计算机可读存储介质内存储有计算机程序，所述计算机程序被处理器执行时实现上述任一视频上传的方法的步骤；或实现上述任一视频播放的方法的步骤。

[0143] 根据本公开实施例的第九方面，提供了一种包含指令的计算机程序产品，当其在计算机上运行时，使得计算机执行上述任一视频上传的方法；或执行上述任一视频播放的方法。

[0144] 本公开的实施例提供的技术方案至少带来以下有益效果：应用本公开实施例，实现了在播放待上传视频时，通过预设的播放窗口播放目标虚拟人物手语动作视频。因此，应用本公开实施例，听障人士可以在观看直播或短视频等形式的视频时，观看到目标虚拟人物手语动作视频，能够有效的获取信息，提高了听障人士的观看体验。

[0145] 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

[0146] 此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本公开的实施例，并与说明书一起用于解释本公开的原理，并不构成对本公开的不当限定。

[0147] 图1a是根据一示例性实施例示出的一种视频上传的方法，应用于视频生成客户端的流程图；

[0148] 图1b为图1a所示实施例中步骤S101的一种具体流程图；

[0149] 图2是根据一示例性实施例示出的一种建立文字数据与手语指令的对应关系的过程的流程图；

[0150] 图3是根据一示例性实施例示出的另一种视频上传的方法，应用于视频生成客户端的流程图；

[0151] 图4是根据一示例性实施例示出的一种视频播放的方法，应用于视频观看客户端的流程图；

[0152] 图5是根据一示例性实施例示出的另一种视频播放的方法，应用于视频观看客户端的流程图；

[0153] 图6是根据一示例性实施例示出的再一种视频播放的方法，应用于视频观看客户端的流程图；

[0154] 图7是根据一示例性实施例示出的一种视频上传的装置，应用于视频生成客户端的装置的框图；

[0155] 图8是根据一示例性实施例示出的一种视频播放的装置，应用于视频观看客户端的装置的框图；

[0156] 图9是根据一示例性实施例示出的再一种视频播放的装置，应用于视频观看客户端的装置的框图；

[0157] 图10是根据一示例性实施例示出的一种客户端设备的框图；

[0158] 图11是根据一示例性实施例示出的一种用于视频上传或视频播放的设备的框图；

[0159] 图12是根据一示例性实施例示出的另一种用于视频上传或视频播放的设备的框图。

具体实施方式

[0160] 为了使本领域普通人员更好地理解本公开的技术方案，下面将结合附图，对本公开实施例中的技术方案进行清楚、完整地描述。

[0161] 需要说明的是，本公开的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本公开的实施例能够以除了在这里图示或描述的那些以外的顺序实施。以下示例性实施例中所描述的实施方式并不代表与本公开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、本公开的一些方面相一致的装置和方法的例子。

[0162] 为了解决相关技术中听障人士无法有效的从直播和短视频等形式的自媒体中获取信息，听障人士的观看体验较低的问题，本公开实施例提供了视频上传、视频播放的方法、装置、设备及存储介质。本公开实施例所提供的视频上传、视频播放的方法，可以应用于任意需要视频上传、视频播放的客户端设备，如：电脑或移动终端等，在此不做具体限定。

[0163] 为了听障人士可以在观看直播或短视频等形式的视频时，观看到目标虚拟人物手语动作视频，本发明实施例提供了两种实现方式。

[0164] 第一种：在待上传视频生成过程中，同时生成目标虚拟人物手语动作视频，将待上传视频和目标虚拟人物手语动作视频都上传到服务器。这样，在视频播放时，可以从服务器

获得所述待上传视频及目标虚拟人物手语动作视频,在播放所述待上传视频时,通过预设的播放窗口播放所述目标虚拟人物手语动作视频。

[0165] 第二种,不在视频生成的过程同时生成目标虚拟人物手语动作视频,而仅在视频播放时,可以在从服务器获得待播放视频后,生成目标虚拟人物手语动作视频,在播放所述待播放视频时,通过预设的播放窗口播放所述目标虚拟人物手语动作视频。

[0166] 以下分别进行详细说明。

[0167] 首先,对第一种实现方式进行详细说明。

[0168] 图1a是根据一示例性实施例示出的一种视频上传的方法的流程图,如图1a所示,该视频上传的方法,应用于视频生成客户端设备,包括以下步骤:

[0169] 步骤S101,将待上传视频中的语音数据转为目标文字数据。

[0170] 步骤S102,根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令。

[0171] 可实施的,可以预先按照建立文字数据与手语指令的对应关系的过程,建立文字数据与手语指令的对应关系。

[0172] 步骤S103,按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频。

[0173] 可实施的,在一种具体的实施方式中,可以先根据待上传视频中语音数据对应的播放时间,确定各个目标手语指令与所述待上传视频中视频帧画面的对应关系;再基于各个目标手语指令、各个目标手语指令与所述待上传视频中视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频。

[0174] 步骤S104,将所述待上传视频和目标虚拟人物手语动作视频上传至服务器,以使视频观看客户端从服务器获得所述待上传视频及目标虚拟人物手语动作视频,在播放所述待上传视频时,通过预设的播放窗口播放所述目标虚拟人物手语动作视频。

[0175] 可见,应用本公开实施例,实现了在播放待上传视频时,通过预设的播放窗口播放目标虚拟人物手语动作视频。因此,应用本公开实施例,听障人士可以在观看直播或短视频等形式的视频时,观看到目标虚拟人物手语动作视频,能够有效的获取信息,提高了听障人士的观看体验。

[0176] 具体的,图1a所示实施例中步骤S101的具体实现过程,可以参见图1b,图1b为图1a所示实施例中步骤S101的一种具体流程图,包括:

[0177] 步骤S1011,获取待上传视频。

[0178] 可实施的,视频生成客户端设备可以获取待上传视频,该待上传视频可以是直播过程中的待上传视频,也可以是预先录制好的短视频,也可以是其它视频,在这里不做具体限定。

[0179] 步骤S1012,从所述待上传视频中获取语音数据作为目标语音数据。

[0180] 可实施的,可以将所述待上传视频中的音频数据实时转换为非压缩的波形文件,输入到预先训练好的神经网络模型;获取所述神经网络模型预测出的目标波形文件;所述神经网络模型,用于对非压缩的波形文件进行切割,得到语音数据中各个语句对应的目标波形文件作为目标语音数据。

[0181] 其中,神经网络模型的训练过程具体可以包括:

[0182] 将样本波形文件输入到待训练的当前神经网络模型;获取当前神经网络模型预测

出的各个目标波形文件;根据人工预先标注的目标波形文件和当前神经网络模型预测的各个目标波形文件,及损失函数计算损失值;

[0183] 根据损失值判断当前神经网络模型是否收敛;如果收敛,则当前神经网络模型为训练完成的神经网络模型;如果未收敛,则调整当前神经网络模型的网络参数,返回所述将样本波形文件输入到待训练的当前神经网络模型的步骤。

[0184] 其中,人工预先标注的目标波形文件可以为:将样本波形文件中的静音时间超过预设阈值的位置,作为当前切割位置,在该当前切割位置对样本波形文件进行切割,获取当前切割位置至上一次切割位置的文件作为目标波形文件;将当前切割位置作为上一次切割位置,继续对样本波形文件进行切割,直到样本波形文件切割完成。

[0185] 步骤S1013,将所述目标语音数据转为目标文字数据。

[0186] 可实施的,可以将所述目标波形文件,按照预设的切分规则切分成各个目标帧文件;对所述各个目标帧文件提取声学特征,获得所述各个目标帧文件的各个目标多维向量;基于预设的多维向量与文字数据之间的对应关系,将所述各个目标多维向量转换为目标文字数据。

[0187] 可见,应用本实施例,可以将待上传视频中的语音数据转为目标文字数据,进而实现了在播放待上传视频时,通过预设的播放窗口播放目标虚拟人物手语动作视频。因此,应用本公开实施例,听障人士可以在观看直播或短视频等形式的视频时,观看到目标虚拟人物手语动作视频,能够有效的获取信息,提高了听障人士的观看体验。

[0188] 可实施的,上述实施例中提到的建立文字数据与手语指令的对应关系的过程,可以如图2所示。

[0189] 图2是根据一示例性实施例示出的一种建立文字数据与手语指令的对应关系的流程的流程图,可以包括:

[0190] 步骤S201,获得基于手语规则设置的每个手语词汇对应的手语指令。

[0191] 可实施的,可以基于手语语法规则,设置每个手语词汇对应的手语指令。

[0192] 步骤S202,获得设置的虚拟人物形象以及每个手语指令对应的虚拟人物形象手部的动作轨迹。

[0193] 可实施的,可以预先设置虚拟人物形象,将每个手语指令分别对应虚拟人物形象手部的各个动作轨迹。

[0194] 步骤S203,将各个手语词汇对应的文字数据与手语指令对应保存。

[0195] 实际应用中,可以针对每个手语指令,编写对应的处理程序,该处理程序用来控制虚拟人物形象的手部动作。例如:对于“我”这个词语,设置对应的手语指令是“生成“我”字的手语”。对应的处理程序中,就包含控制虚拟人物手部的动作,形成“我”字的手语视频。

[0196] 可见,应用本实施例,可以将各个手语词汇对应的文字数据与手语指令对应保存,建立文字数据与手语指令的一一对应关系。后续可以基于各个目标手语指令、各个目标手语指令与视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频,以方便听障人士获取信息,提高听障人士的观看体验。

[0197] 图3是根据一示例性实施例示出的另一种视频上传的方法,应用于视频生成客户端的流程图,可以包括:

[0198] 步骤S301,获取待上传视频。

[0199] 可实施的,视频生成客户端设备可以获取待上传视频,该待上传视频可以是直播过程中的待上传视频,也可以是预先录制好的短视频,也可以是其它视频,在这里不做具体限定。

[0200] 步骤S302,判断当前是否为上传虚拟人物手语动作视频的上传状态。

[0201] 所述上传状态,包括上传虚拟人物手语动作视频状态,和非上传虚拟人物手语动作视频状态,由用户预先设置。

[0202] 可实施的,可以判断预先设置在视频生成客户端视频显示界面的菜单栏中的上传虚拟人物手语动作视频的按钮是否被按下,如果被按下,则为上传虚拟人物手语动作视频的上传状态。技术人员可以根据实际情况将上传虚拟人物手语动作视频的设置其它合适位置,在此对上传虚拟人物手语动作视频的设置的位置不做具体限定。

[0203] 如果判断的结果为否,即为非上传虚拟人物手语动作视频状态,则执行步骤S303;如果判断的结果为是,即为上传虚拟人物手语动作视频状态,则执行步骤S304。

[0204] 步骤S303,将待上传视频上传至服务器。

[0205] 可实施的,如果判断结果为非上传虚拟人物手语动作视频状态时,则可以只将待上传视频上传。结束操作。

[0206] 步骤S304,将所述待上传视频中的音频数据实时转换为非压缩的波形文件,输入到预先训练好的神经网络模型。

[0207] 可实施的,视频生成客户端可以是直播应用中的主播客户端,也可以是短视频应用中录制短视频的客户端;相应的,视频观看客户端可以是直播应用中的观众客户端,也可以是短视频应用中观看短视频的客户端。

[0208] 直播或录制短视频时可以将所述待上传视频中的音频数据实时转换为非压缩的波形文件,输入到预先训练好的神经网络模型;其中,如果是短视频时,也可以将录制好的短视频作为待上传视频,将其对应的音频文件输入到预先训练好的神经网络模型。

[0209] 可实施的,神经网络模型的训练过程可以和上述步骤S102中神经网络模型的训练过程相同,这里不再详细说明。

[0210] 步骤S305,获取所述神经网络模型预测出的目标波形文件;所述神经网络模型,用于对非压缩的波形文件进行切割,得到语音数据中各个语句对应的目标波形文件作为目标语音数据。

[0211] 步骤S306,将所述目标波形文件,按照预设的切分规则切分成各个目标帧文件。

[0212] 可实施的,可以先将所述目标波形文件的首段和尾段的静音切除,获得切除静音后的目标波形文件;再对切除静音后的目标波形文件,按照预设的时间,进行分帧,获得各个目标帧文件。可实施的,每帧的长度可以为25毫秒。

[0213] 步骤S307,对所述各个目标帧文件提取声学特征,获得所述各个目标帧文件的各个目标多维向量。

[0214] 步骤S308,基于预设的多维向量与文字数据之间的对应关系,将所述各个目标多维向量转换为目标文字数据。

[0215] 可实施的,可以用相关技术中的隐马尔可夫模型,构建状态网络,从状态网络中寻找与目标多维向量最匹配的路径,获得目标文字数据。

[0216] 步骤S309,根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字

数据对应的至少一条目标手语指令。

[0217] 可实施的,建立文字数据与手语指令的对应关系的过程可以如上述图2所示。

[0218] 步骤S310,根据待上传视频中语音数据对应的播放时间,确定各个目标手语指令与所述待上传视频中视频帧画面的对应关系。

[0219] 可实施的,可以根据待上传视频中语音数据对应的播放时间,确定各个目标手语指令对应的虚拟人物形象手部的动作轨迹与所述待上传视频中视频帧画面的对应关系。具体的,可以在将待上传视频对应的音频数据实时转换为非压缩的波形文件的过程中,记录每段语音数据在视频中对应的起始帧的时间戳信息,并根据记录确定待上传视频中语音数据对应的播放时间。

[0220] 步骤S311,基于各个目标手语指令、各个目标手语指令与所述待上传视频中视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频。

[0221] 在一种具体的实施方式中,可以基于所述虚拟人物形象、所述目标手语指令对应的虚拟人物形象手部的动作轨迹、所述虚拟人物形象手部的动作轨迹与所述待上传视频中视频帧画面的对应关系,生成目标虚拟人物手语动作视频。

[0222] 步骤S312,将所述待上传视频和目标虚拟人物手语动作视频上传至服务器,以使视频观看客户端从服务器获得所述待上传视频及目标虚拟人物手语动作视频,在播放所述待上传视频时,通过预设的播放窗口播放所述目标虚拟人物手语动作视频。

[0223] 可实施的,该预设的播放窗口可以为预设的同步播放窗口。

[0224] 可见,应用本公开实施例,实现了在播放待上传视频时,通过预设的播放窗口播放目标虚拟人物手语动作视频。因此,应用本公开实施例,听障人士可以在观看直播或短视频等形式的视频时,观看到目标虚拟人物手语动作视频,能够有效的获取信息,提高了听障人士的观看体验。

[0225] 而且,在本实施例中,视频生成客户端可以判断当前是否为上传虚拟人物手语动作视频的上传状态,如果为上传状态,则开始执行从所述待上传视频中获取语音数据作为目标语音数据的步骤,如果不是上传虚拟人物手语动作视频的上传状态,则将待上传视频上传。这样,视频生成客户端的用户可以选择是否生成并上传目标虚拟人物手语动作视频,提高了视频生成客户端的用户体验。

[0226] 与图1a所示的视频上传实施例对应,图4是根据一示例性实施例示出的一种视频播放的方法,应用于视频观看客户端的流程图,如图4所示,可以包括:

[0227] 步骤S401,从服务器获得待播放视频及目标虚拟人物手语动作视频。

[0228] 所述目标虚拟主播手语动作视频,为:视频生成客户端,将待上传视频中的语音数据转为目标文字数据;根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;按照所述目标手语指令和预设的虚拟人物确定的。

[0229] 步骤S402,在播放所述待播放视频时显示预设的播放窗口。

[0230] 步骤S403,在所述播放窗口中播放所述目标虚拟人物手语动作视频。

[0231] 可实施的,还可以在播放过程中,检测是否接收到关闭所述目标虚拟人物手语动作视频的指令;如果是,则关闭所述预设的播放窗口,停止播放所述目标虚拟人物手语动作视频;如果否,则在播放所述待上传视频时播放所述目标虚拟人物手语动作视频。

[0232] 可见,应用本公开实施例,实现了在播放待上传视频时,通过预设的播放窗口播放

目标虚拟人物手语动作视频。因此,应用本公开实施例,听障人士可以在观看直播或短视频等形式的视频时,观看到目标虚拟人物手语动作视频,能够有效的获取信息,提高了听障人士的观看体验。

[0233] 下面,对第二种实现方式进行详细说明。

[0234] 图5是根据一示例性实施例示出的另一种视频播放的方法,应用于视频观看客户端的流程图,可以包括:

[0235] 步骤S501,将待播放视频中的语音数据转为目标文字数据。

[0236] 可实施的,在视频观看客户端将待播放视频中的语音数据转为目标文字数据的具体实现过程可以参照上述图1b。

[0237] 步骤S502,根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令。

[0238] 可实施的,可以预先按照上述图2所述的建立文字数据与手语指令的对应关系的过程,建立文字数据与手语指令的对应关系。

[0239] 步骤S503,按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频。

[0240] 在一种具体的实施方式中,可以根据待播放视频中语音数据对应的播放时间,确定各个目标手语指令与所述待播放视频中视频帧画面的对应关系;基于各个目标手语指令、各个目标手语指令与所述待播放视频中视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频。

[0241] 步骤S504,在播放所述待播放视频时,显示预设的播放窗口。

[0242] 步骤S505,在所述播放窗口中播放所述目标虚拟人物手语动作视频。

[0243] 可见,应用本公开实施例,实现了在播放待播放视频时,通过预设的播放窗口播放目标虚拟人物手语动作视频。因此,应用本公开实施例,听障人士可以在观看直播或短视频等形式的视频时,观看到目标虚拟人物手语动作视频,能够有效的获取信息,提高了听障人士的观看体验。

[0244] 图6是根据一示例性实施例示出的再一种视频播放的方法,应用于视频观看客户端的流程图,可以包括:

[0245] 步骤S601,获得待播放视频。

[0246] 步骤S602,判断当前是否为播放虚拟人物手语动作视频的播放状态。

[0247] 所述播放状态,包括播放虚拟人物手语动作视频状态,和非播放虚拟人物手语动作视频状态。

[0248] 如果判断的结果为否,即不是播放虚拟人物手语动作视频的播放状态,则执行步骤 S603;如果判断的结果为是,即为播放虚拟人物手语动作视频的播放状态,则执行步骤 S604;

[0249] 步骤S603,播放待播放视频。

[0250] 步骤S604,将所述待播放视频中的音频数据实时转换为非压缩的波形文件,输入到预先训练好的神经网络模型。

[0251] 可实施的,神经网络模型的训练过程具体可以参照上述图1b中步骤S1012中的神经网络模型的训练过程。

[0252] 步骤S605,获取所述神经网络模型预测出的目标波形文件;所述神经网络模型,用

于对非压缩的波形文件进行切割,得到语音数据中各个语句对应的目标波形文件作为目标语音数据。

[0253] 步骤S606,将所述目标波形文件,按照预设的切分规则切分成各个目标帧文件。

[0254] 实际应用中,步骤S606可以与图3中的步骤S306相同,这里不再详细说明。

[0255] 步骤S607,对所述各个目标帧文件提取声学特征,获得所述各个目标帧文件的各个目标多维向量。

[0256] 步骤S608,基于预设的多维向量与文字数据之间的对应关系,将所述各个目标多维向量转换为目标文字数据。

[0257] 步骤S609,根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令。

[0258] 可实施的,建立文字数据与手语指令的对应关系的过程可以如上述图2所示。

[0259] 步骤S610,根据待播放视频中语音数据对应的播放时间,确定各个目标手语指令与所述待播放视频中视频帧画面的对应关系。

[0260] 实际应用中,步骤S610可以与图3中的步骤S310实现方式相同,这里不再详细说明。

[0261] 步骤S611,基于各个目标手语指令、各个目标手语指令与所述待播放视频中视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频。

[0262] 实际应用中,步骤S611可以与图3中的步骤S311实现方式相同,这里不再详细说明。

[0263] 步骤S612,在播放所述待播放视频时,显示预设的播放窗口。

[0264] 可实施的,该播放窗口可以为同步播放窗口。

[0265] 步骤S613,在所述播放窗口中播放所述目标虚拟人物手语动作视频。

[0266] 可实施的,可以在同步播放窗口中播放目标虚拟人物手语动作视频。

[0267] 可见,应用本公开实施例,实现了在播放待播放视频时,通过预设的播放窗口播放目标虚拟人物手语动作视频。因此,应用本公开实施例,听障人士可以在观看直播或短视频等形式的视频时,观看到目标虚拟人物手语动作视频,能够有效的获取信息,提高了听障人士的观看体验。

[0268] 而且,在本实施例中,视频观看客户端可以判断当前是否为播放虚拟人物手语动作视频的播放状态,如果是,则开始执行从所述待播放视频中获取语音数据作为目标语音数据的步骤;如果没检测到该指令,则播放待播放视频。这样,视频观看客户端的用户可以选择是否在播放所述待播放视频时播放所述目标虚拟人物手语动作视频,听力正常的用户就可以选择不播放目标虚拟人物手语动作视频,听障人士可以选择播放目标虚拟人物手语动作视频,提高了视频观看客户端的不同用户的用户体验。

[0269] 图7是根据一示例性实施例示出的一种视频上传的装置,应用于视频生成客户端的装置的框图。参照图7,该装置包括:

[0270] 第一目标文字数据转换单元701,被配置为将待上传视频中的语音数据转为目标文字数据;

[0271] 第一目标手语指令确定单元702,被配置为根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;

- [0272] 第一手语动作视频确定单元703,被配置为按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频;
- [0273] 上传单元704,被配置为将所述待上传视频和目标虚拟人物手语动作视频上传至服务器,以使视频观看客户端从服务器获得所述待上传视频及目标虚拟人物手语动作视频,在播放所述待上传视频时,通过预设的播放窗口播放所述目标虚拟人物手语动作视频。
- [0274] 可选的,所述第一手语动作视频确定单元,包括:
- [0275] 第一对应关系确定模块,被配置为根据待上传视频中语音数据对应的播放时间,确定各个目标手语指令与所述待上传视频中视频帧画面的对应关系;
- [0276] 第一手语动作视频生成模块,被配置为基于各个目标手语指令、各个目标手语指令与所述待上传视频中视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频。
- [0277] 可选的,所述第一目标文字数据转换单元,包括:
- [0278] 视频获取模块,被配置为获取待上传视频;
- [0279] 第一目标语音数据获取模块,被配置为从所述待上传视频中获取语音数据作为目标语音数据;
- [0280] 第一目标文字数据转换模块,被配置为将所述目标语音数据转为目标文字数据。
- [0281] 可选的,所述装置,还包括:
- [0282] 第一判断模块,被配置为在所述第一目标语音数据获取模块目标语音数据获取模块从所述待上传视频中获取语音数据作为目标语音数据之前,判断当前是否为上传虚拟人物手语动作视频的上传状态;所述上传状态,包括上传虚拟人物手语动作视频状态,和非上传虚拟人物手语动作视频状态,由用户预先设置;如果是,则执行所述从所述待上传视频中获取语音数据作为目标语音数据的步骤。
- [0283] 可选的,所述第一目标语音数据获取模块,具体被配置为:
- [0284] 将所述待上传视频中的音频数据实时转换为非压缩的波形文件,输入到预先训练好的神经网络模型;
- [0285] 获取所述神经网络模型预测出的目标波形文件;所述神经网络模型,用于对非压缩的波形文件进行切割,得到语音数据中各个语句对应的目标波形文件作为目标语音数据。
- [0286] 可选的,所述第一目标文字数据转换模块,具体被配置为:
- [0287] 将所述目标波形文件,按照预设的切分规则切分成各个目标帧文件;
- [0288] 对所述各个目标帧文件提取声学特征,获得所述各个目标帧文件的各个目标多维向量;
- [0289] 基于预设的多维向量与文字数据之间的对应关系,将所述各个目标多维向量转换为目标文字数据。
- [0290] 可选的,所述装置,还包括:第一对应关系建立单元;
- [0291] 所述第一对应关系建立单元,具体被配置为:
- [0292] 获得基于手语规则设置的每个手语词汇对应的手语指令;
- [0293] 获得设置的虚拟人物形象以及每个手语指令对应的虚拟人物形象手部的动作轨迹;

[0294] 将各个手语词汇对应的文字数据与手语指令对应保存。

[0295] 可选的,所述第一对应关系确定模块,具体被配置为:

[0296] 根据待上传视频中语音数据对应的播放时间,确定各个目标手语指令对应的虚拟人物形象手部的动作轨迹与所述待上传视频中视频帧画面的对应关系。

[0297] 可选的,所述第一手语动作视频生成模块,具体被配置为:

[0298] 基于所述虚拟人物形象、所述目标手语指令对应的虚拟人物形象手部的动作轨迹、所述虚拟人物形象手部的动作轨迹与所述待上传视频中视频帧画面的对应关系,生成目标虚拟人物手语动作视频。

[0299] 可见,应用本公开实施例,实现了在播放待上传视频时,通过预设的播放窗口播放目标虚拟人物手语动作视频。因此,应用本公开实施例,听障人士可以在观看直播或短视频等形式的视频时,观看到目标虚拟人物手语动作视频,能够有效的获取信息,提高了听障人士的观看体验。

[0300] 图8是根据一示例性实施例示出的一种视频播放的装置,应用于视频观看客户端的装置的框图。参照图8,该装置包括:

[0301] 第一手语动作视频获得单元801,被配置为从服务器获得待播放视频及目标虚拟人物手语动作视频;所述目标虚拟人物手语动作视频,为:视频生成客户端,将待上传视频中的语音数据转为目标文字数据;根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;按照所述目标手语指令和预设的虚拟人物确定的;

[0302] 第一播放窗口显示单元802,被配置为在播放所述待播放视频时显示预设的播放窗口;

[0303] 第一手语动作视频播放单元803,被配置为在所述播放窗口中播放所述目标虚拟人物手语动作视频。

[0304] 可选的,所述装置还包括:

[0305] 第一手语动作视频关闭指令检测单元,被配置为检测是否接收到关闭所述目标虚拟人物手语动作视频的指令;如果是,则关闭所述预设的播放窗口,停止播放所述目标虚拟人物手语动作视频。

[0306] 可见,应用本公开实施例,实现了在播放待上传视频时,通过预设的播放窗口播放目标虚拟人物手语动作视频。因此,应用本公开实施例,听障人士可以在观看直播或短视频等形式的视频时,观看到目标虚拟人物手语动作视频,能够有效的获取信息,提高了听障人士的观看体验。

[0307] 图9是根据一示例性实施例示出的再一种视频播放的装置,应用于视频观看客户端的装置的框图。参照图9,该装置包括:

[0308] 第二目标文字数据转换单元901,被配置为将待播放视频中的语音数据转为目标文字数据;

[0309] 第二目标手语指令确定单元902,被配置为根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;

[0310] 第二手语动作视频确定单元903,被配置为按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频;

- [0311] 第二播放窗口显示单元904,被配置为在播放所述待播放视频时,显示预设的播放窗口;
- [0312] 第二手语动作视频播放单元905,被配置为在所述播放窗口中播放所述目标虚拟人物手语动作视频。
- [0313] 可选的,所述第二手语动作视频确定单元,包括:
- [0314] 第二对应关系确定模块,被配置为根据待播放视频中语音数据对应的播放时间,确定各个目标手语指令与所述待播放视频中视频帧画面的对应关系;
- [0315] 第二手语动作视频生成模块,被配置为基于各个目标手语指令、各个目标手语指令与所述待播放视频中视频帧画面的对应关系及预设的虚拟人物,生成目标虚拟人物手语动作视频。
- [0316] 可选的,所述第二目标文字数据转换单元,包括:
- [0317] 视频获得模块,被配置为获得待播放视频;
- [0318] 第二目标语音数据获取模块,被配置为从所述待播放视频中获取语音数据作为目标语音数据;
- [0319] 第二目标文字数据转换模块,被配置为将所述目标语音数据转为目标文字数据。
- [0320] 可选的,所述装置,还包括:
- [0321] 第二判断模块,被配置为在所述第二目标语音数据获取模块从所述待播放视频中获取语音数据作为目标语音数据之前,判断当前是否为播放虚拟人物手语动作视频的播放状态;所述播放状态,包括播放虚拟人物手语动作视频状态,和非播放虚拟人物手语动作视频状态;如果是,则执行所述从所述待播放视频中获取语音数据作为目标语音数据的步骤。
- [0322] 可选的,所述第二目标语音数据获取模块,具体被配置为:
- [0323] 将所述待播放视频中的音频数据实时转换为非压缩的波形文件,输入到预先训练好的神经网络模型;
- [0324] 获取所述神经网络模型预测出的目标波形文件;所述神经网络模型,用于对非压缩的波形文件进行切割,得到语音数据中各个语句对应的目标波形文件作为目标语音数据。
- [0325] 可选的,所述第二目标文字数据转换模块,具体被配置为:
- [0326] 将所述目标波形文件,按照预设的切分规则切分成各个目标帧文件;
- [0327] 对所述各个目标帧文件提取声学特征,获得所述各个目标帧文件的各个目标多维向量;
- [0328] 基于预设的多维向量与文字数据之间的对应关系,将所述各个目标多维向量转换为目标文字数据。
- [0329] 可选的,所述装置,还包括:第二对应关系建立单元;
- [0330] 所述第二对应关系建立单元,具体被配置为:
- [0331] 获得基于手语规则设置的每个手语词汇对应的手语指令;
- [0332] 获得设置的虚拟人物形象以及每个手语指令对应的虚拟人物形象手部的动作轨迹;
- [0333] 将各个手语词汇对应的文字数据与手语指令对应保存。
- [0334] 可选的,所述第二对应关系确定模块,具体被配置为:

[0335] 根据待播放视频中语音数据对应的播放时间,确定各个目标手语指令对应的虚拟人物形象手部的动作轨迹与所述待播放视频中视频帧画面的对应关系。

[0336] 可选的,所述第二手语动作视频生成模块,具体被配置为:

[0337] 基于所述虚拟人物形象、所述目标手语指令对应的虚拟人物形象手部的动作轨迹、所述虚拟人物形象手部的动作轨迹与所述待播放视频中视频帧画面的对应关系,生成目标虚拟人物手语动作视频。

[0338] 可见,应用本公开实施例,实现了在播放待上传视频时,通过预设的播放窗口播放目标虚拟人物手语动作视频。因此,应用本公开实施例,听障人士可以在观看直播或短视频等形式的视频时,观看到目标虚拟人物手语动作视频,能够有效的获取信息,提高了听障人士的观看体验。

[0339] 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。

[0340] 本发明实施例还提供了一种客户端设备,如图10所示,包括处理器1001、通信接口1002、存储器1003和通信总线1004,其中,处理器1001,通信接口1002,存储器1003 通过通信总线1004完成相互间的通信,

[0341] 存储器1003,用于存放计算机程序;

[0342] 处理器1001,用于执行存储器1003上所存放的程序时,实现如下视频生成客户端的步骤:

[0343] 将待上传视频中的语音数据转为目标文字数据;

[0344] 根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;

[0345] 按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频;

[0346] 将所述待上传视频和目标虚拟人物手语动作视频上传至服务器,以使视频观看客户端从服务器获得所述待上传视频及目标虚拟人物手语动作视频,在播放所述待上传视频时,通过预设的播放窗口播放所述目标虚拟人物手语动作视频。

[0347] 或,实现如下视频观看客户端的步骤:

[0348] 从服务器获得待播放视频及目标虚拟人物手语动作视频;所述目标虚拟人物手语动作视频,为:视频生成客户端,将待上传视频中的语音数据转为目标文字数据;根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;按照所述目标手语指令和预设的虚拟人物确定的;

[0349] 在播放所述待播放视频时显示预设的播放窗口;

[0350] 在所述播放窗口中播放所述目标虚拟人物手语动作视频。

[0351] 或,实现如下视频观看客户端的步骤:

[0352] 将待播放视频中的语音数据转为目标文字数据;

[0353] 根据预先建立的文字数据与手语指令的对应关系,确定所述目标文字数据对应的至少一条目标手语指令;

[0354] 按照所述目标手语指令,确定预设的目标虚拟人物手语动作视频;

[0355] 在播放所述待播放视频时,显示预设的播放窗口;

[0356] 在所述播放窗口中播放所述目标虚拟人物手语动作视频。

[0357] 可见,应用本公开实施例,实现了在播放视频时,通过预设的播放窗口播放目标虚拟人物手语动作视频。因此,应用本公开实施例,听障人士可以在观看直播或短视频等形式的视频时,观看到目标虚拟人物手语动作视频,能够有效的获取信息,提高了听障人士的观看体验。

[0358] 图11是根据一示例性实施例示出的一种用于视频上传或视频播放的装置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。

[0359] 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,电源组件1106,多媒体组件1108,音频组件1110,输入/输出(I/O)接口1112,传感器组件1114,以及通信组件1116。

[0360] 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。

[0361] 存储器1104被配置为存储各种类型的数据以支持在装置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。

[0362] 电源组件1106为装置1100的各种组件提供电力。电源组件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。

[0363] 多媒体组件1108包括在所述装置1100和用户之间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。

[0364] 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。

[0365] 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。

[0366] 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,例如所述组件为装置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。

[0367] 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,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,所述通信组件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。

[0368] 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。

[0369] 在示例性实施例中,还提供了一种包括指令的存储介质,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。可选地,存储介质可以是非临时性计算机可读存储介质,例如,所述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。

[0370] 图12是根据一示例性实施例示出的一种用于视频上传或视频播放的装置1200的框图。例如,装置1200可以被提供为一服务器。参照图12,装置1200包括处理组件1222,其进一步包括一个或多个处理器,以及由存储器1232所代表的存储器资源,用于存储可由处理组件1222的执行的指令,例如应用程序。存储器1232中存储的应用程序可以包括一个或一个以上的每一个对应于一组指令的模块。此外,处理组件1222被配置为执行指令,以执行上述视频上传或视频播放的方法。

[0371] 装置1200还可以包括一个电源组件1226被配置为执行装置1200的电源管理,一个有线或无线网络接口1250被配置为将装置1200连接到网络,和一个输入输出(I/O)接口1258。装置1200可以操作基于存储在存储器1232的操作系统,例如Windows Server™,Mac OS X™,Unix™,Linux™,FreeBSD™或类似……。

[0372] 在示例性实施例中,还提供了一种存储介质,所述存储介质内存储有计算机程序,当所述计算机程序被处理器执行时,实现上述任一所述的视频上传的方法;或实现上述任一所述的视频播放的方法。

[0373] 可选地,该存储介质可以是非临时性计算机可读存储介质,示例性的,所述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。

[0374] 在示例性实施例中,还提供了一种计算机程序产品,当其在计算机上运行时,使得计算机执行上述任一所述的视频上传的方法;或执行上述任一所述的视频播放的方法。

[0375] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本公开的其

它实施方案。本申请旨在涵盖本公开的任何变型、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公开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,本公开的真正范围和精神由下面的权利要求指出。

[0376] 应当理解的是,本公开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。本公开的范围仅由所附的权利要求来限制。

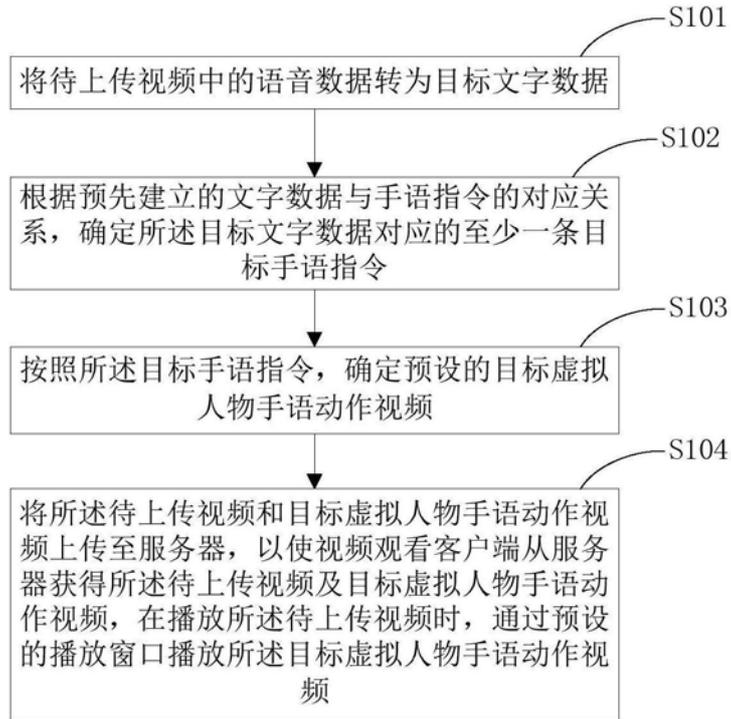


图1a

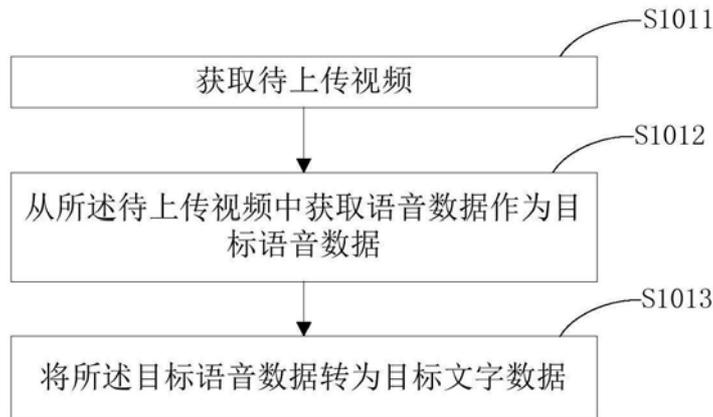


图1b

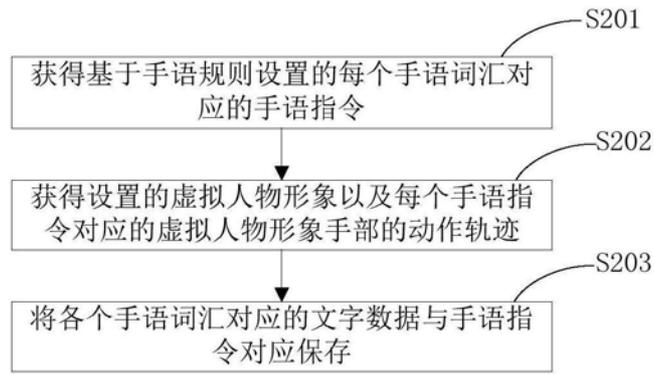


图2

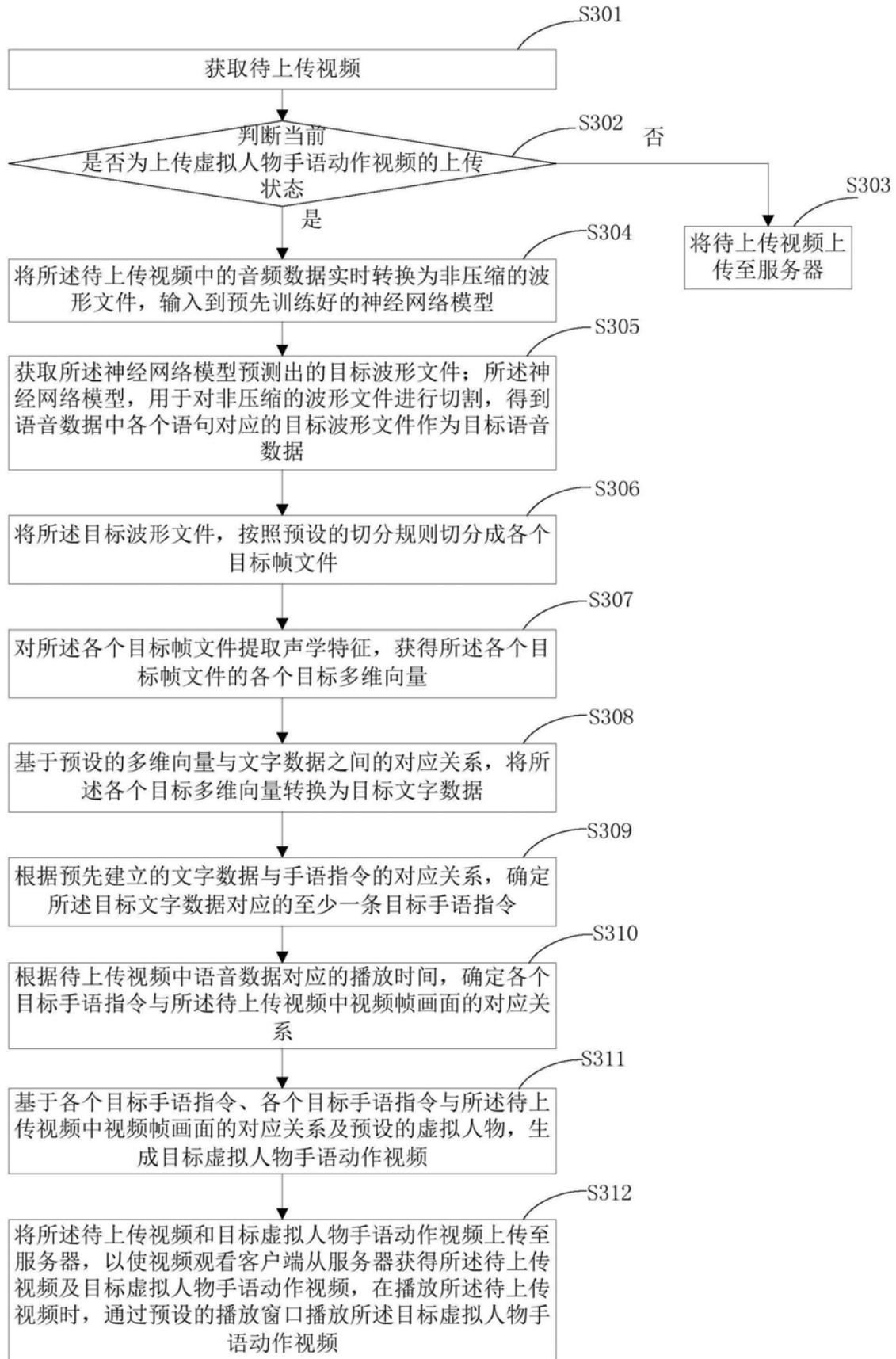


图3

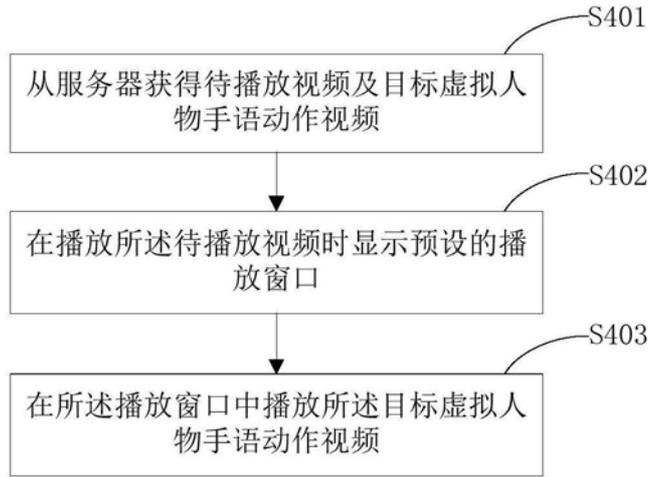


图4

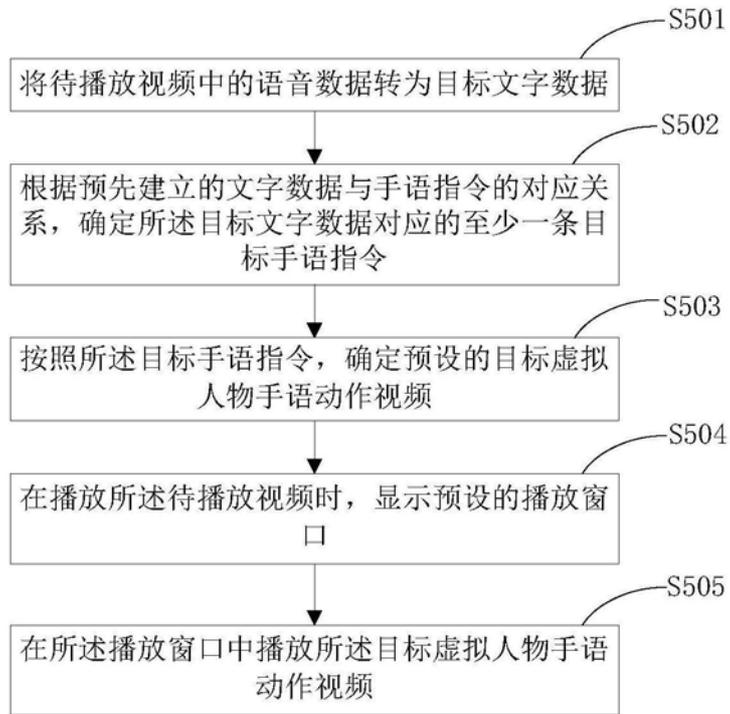


图5

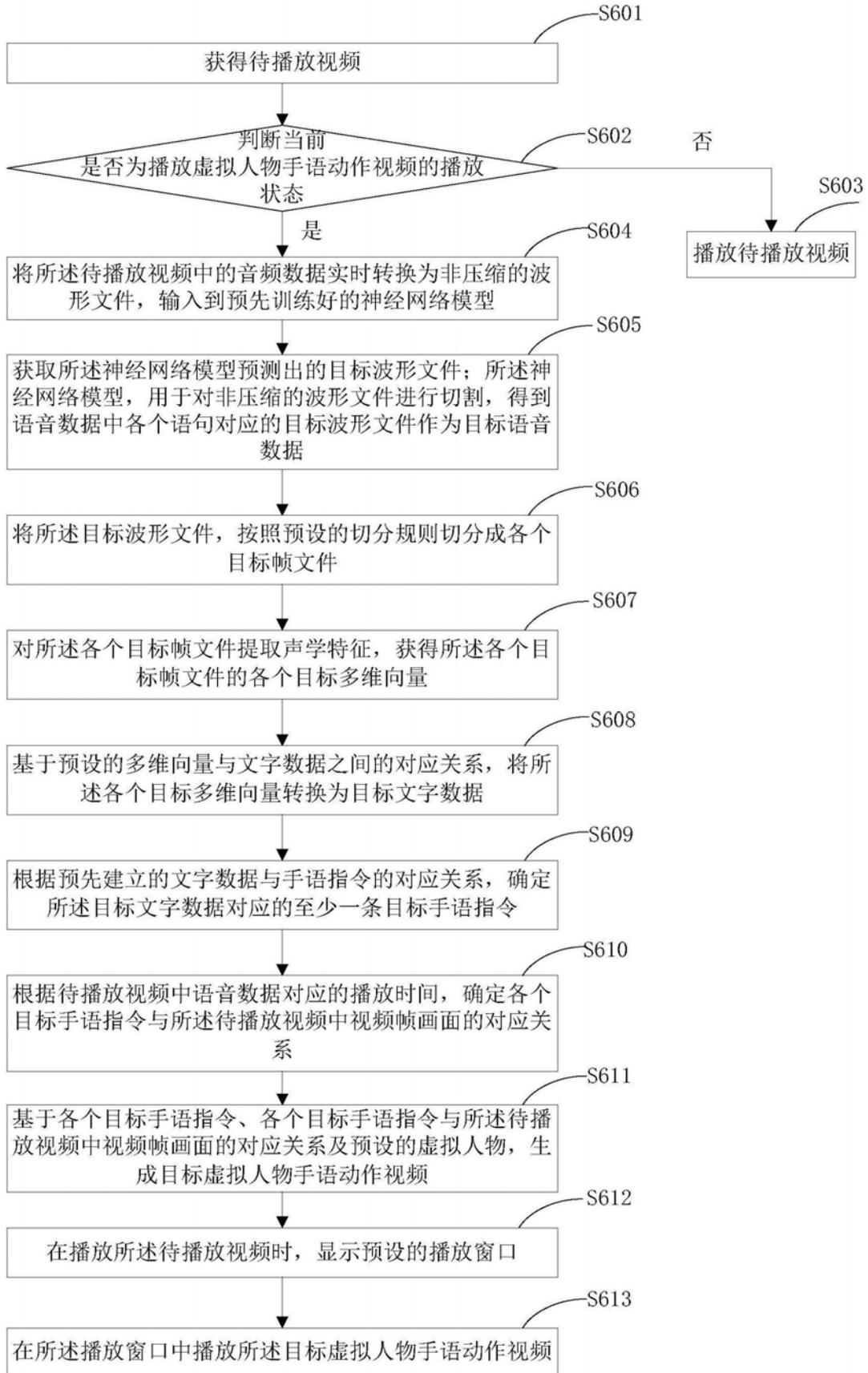


图6

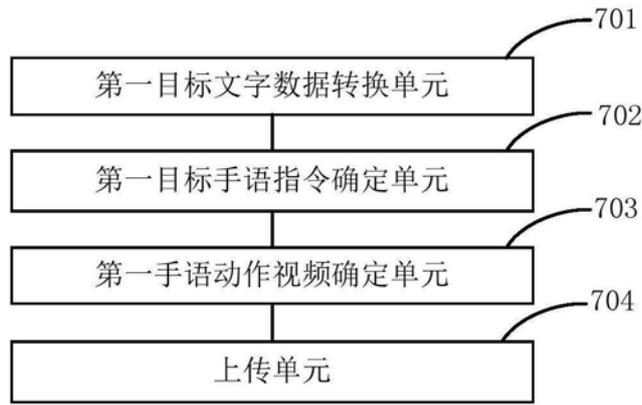


图7



图8

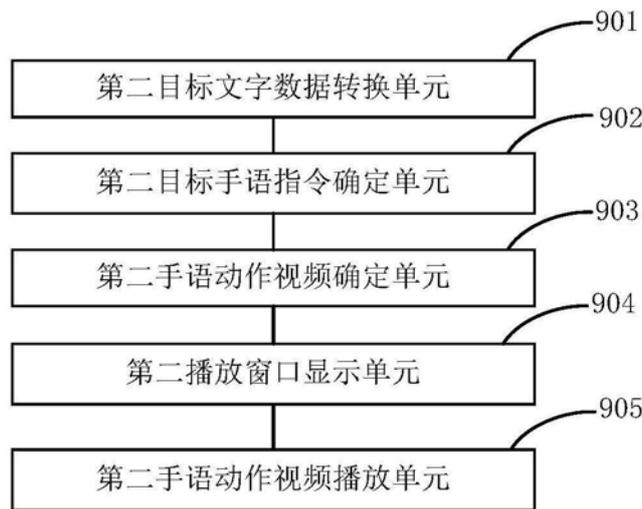


图9

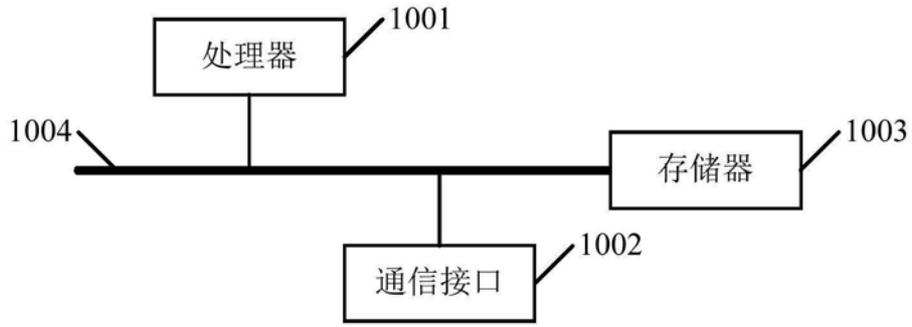


图10

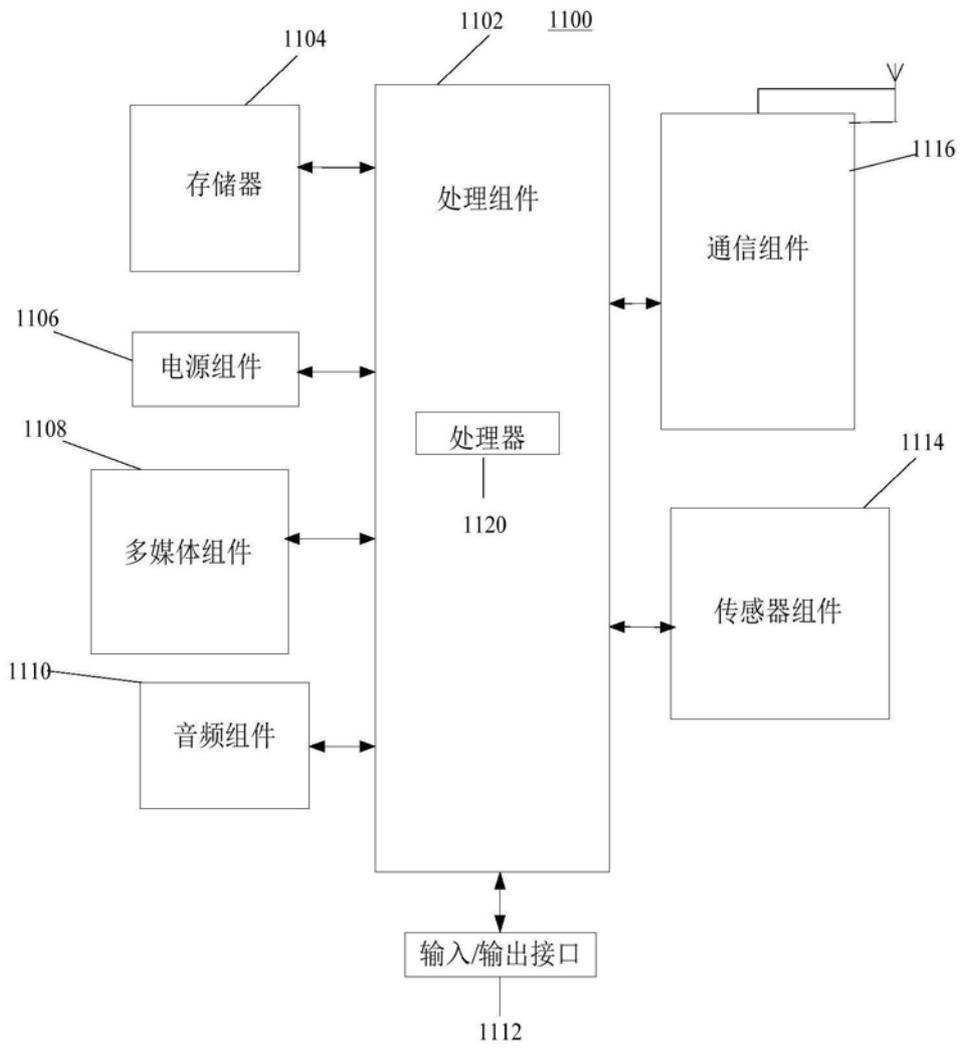


图11

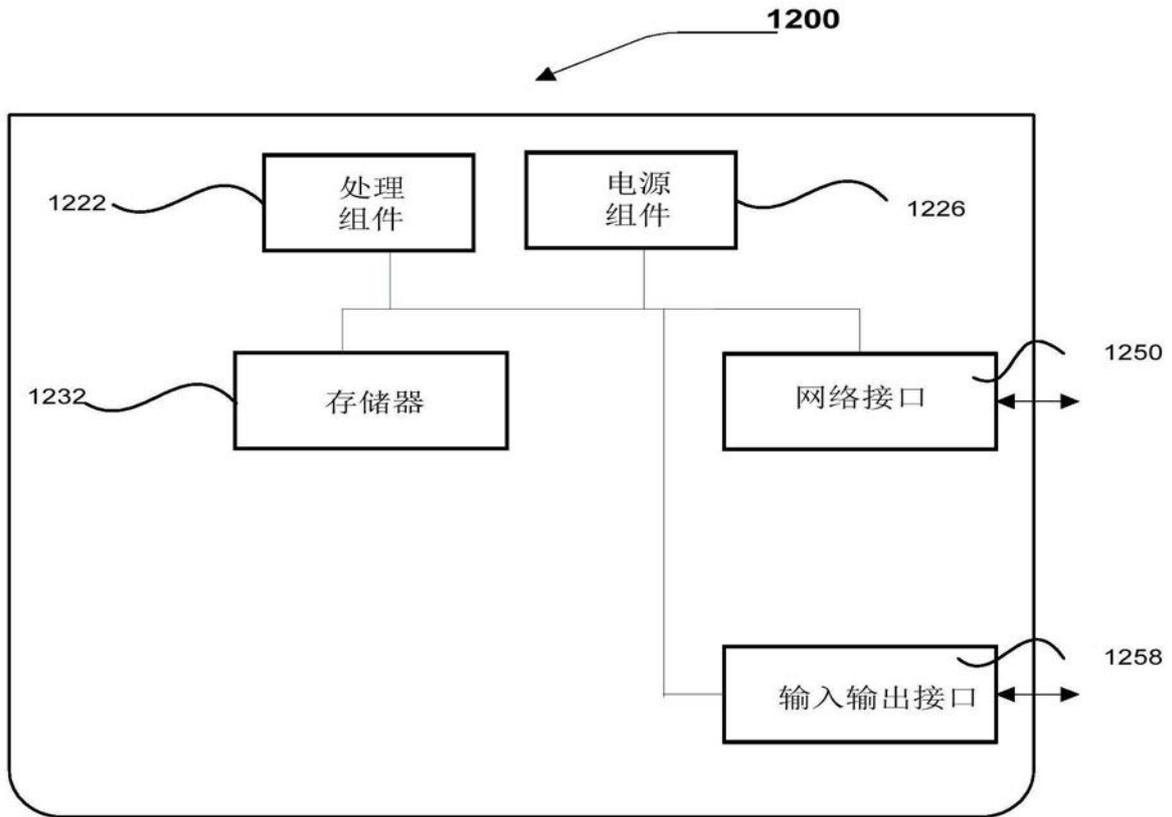


图12