



(12)发明专利申请

(10)申请公布号 CN 113409778 A

(43)申请公布日 2021.09.17

(21)申请号 202010183403.X

(22)申请日 2020.03.16

(71)申请人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼资本大厦一座四
层847号邮箱

(72)发明人 徐贤仲

(74)专利代理机构 北京展翼知识产权代理事务
所(特殊普通合伙) 11452

代理人 张阳

(51) Int. Cl.

G10L 15/22(2006.01)

G10L 15/26(2006.01)

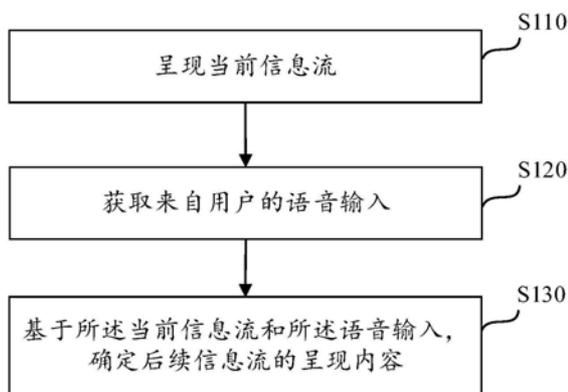
权利要求书4页 说明书10页 附图3页

(54)发明名称

语音交互方法、系统和终端

(57)摘要

公开了一种语音交互方法、系统和终端。所述语音交互方法包括：呈现当前信息流；获取来自用户的语音输入；基于所述当前信息流和所述语音输入，确定后续信息流的呈现内容。所述信息流可以是包括含剧情分支的信息流，或是包括可操控虚拟化身的信息流。由此，本发明提供一种用户能够通过语音交互主动影响内容走向的方案。用户通过语音输入，能够决定当前信息流的后续走向，尤其是能够通过语音输入决定剧情类游戏的剧情分支，从而增强用户的沉浸感和参与感，提升游戏的可玩性。



1. 一种语音交互方法,包括:
呈现当前信息流;
获取来自用户的语音输入;以及
基于所述当前信息流和所述语音输入,确定后续信息流的呈现内容。
2. 如权利要求1所述的方法,其中,所述信息流包括含剧情分支的信息流。
3. 如权利要求2所述的方法,其中,基于所述当前信息流和所述语音输入,确定后续信息流的呈现内容包括:
基于所述信息流的当前分支和所述语音输入,确定后续信息流的分支走向。
4. 如权利要求2所述的方法,其中,所述信息流包括如下的至少一项:
具有剧情分支的游戏;
具有剧情分支的剧集;以及
具有剧情分支的小说。
5. 如权利要求2所述的方法,其中,呈现当前信息流包括:
向用户呈现用于触发不同剧情分支的多个选项,
获取来自用户的语音输入包括:
获取用户对所述多个选项中一个选项的语音选择,并且
基于所述当前信息流和所述语音输入,确定后续信息流的呈现内容包括:
基于所述语音选择,触发与被选选项相对应的剧情分支。
6. 如权利要求5所述的方法,其中,呈现当前信息流包括:
向用户呈现用于触发不同剧情分支的交互点提示,
获取来自用户的语音输入包括:
获取用户对所述交互点提示的语音指令,并且
基于所述当前信息流和所述语音指令,确定后续信息流的呈现内容包括:
基于所述语音选择,触发所述交互点相对应的剧情分支或分支选项。
7. 如权利要求2所述的方法,其中,获取来自用户的语音输入包括:
获取所述语音输入的文本信息;
获取所述语音输入的语音属性信息,并且
确定后续信息流的呈现内容包括:
基于所述文本信息和所述语音属性信息,确定后续信息流的呈现内容。
8. 如权利要求7所述的方法,还包括:
基于所述文本信息和所述语音属性信息,确定后续信息流的呈现方式。
9. 如权利要求7所述的方法,其中,所述语音属性信息包括如下至少一项:
所述语音输入相对应当前信息流的起始时刻;
所述语音输入的持续时间;
所述语音输入的情感信息;
所述语音输入的语调信息;以及
所述语音输入对应的用户身份。
10. 如权利要求2所述的方法,还包括:
获取所述语音输入生成时的环境信息,并且

确定后续信息流的呈现内容包括：

基于所述环境信息，确定后续信息流的呈现内容。

11. 如权利要求2所述的方法，还包括：

获取所述用户的非语音输入，并且

确定后续信息流的呈现内容包括：

基于所述非语音输入，确定后续信息流的呈现内容。

12. 如权利要求2所述的方法，其中，基于所述当前信息流和所述语音输入，确定后续信息流的呈现内容包括：

基于所述信息流的当前分支和所述语音输入，更改所述剧情分支的结构。

13. 如权利要求1所述的方法，其中，呈现当前信息流包括：

呈现虚拟化身。

14. 如权利要求13所述的方法，其中，所述虚拟化身包括用户虚拟化身，并且

获取来自用户的语音输入包括：

获取所述用户对所述用户虚拟化身的语音控制，并且

确定后续信息流的呈现内容包括：

基于所述语音输入，控制所述用户虚拟化身的呈现。

15. 如权利要求14所述的方法，其中，所述虚拟化身包括用户虚拟化身之外的其他虚拟化身，并且

获取来自用户的语音输入包括：

获取所述用户与所述其他虚拟化身的语音交互，并且

确定后续信息流的呈现内容包括：

基于所述语音交互，控制所述其他虚拟化身的呈现。

16. 如权利要求15所述的方法，其中，控制所述其他虚拟化身的呈现包括如下至少一项：

获取触发剧情分支的线索；

获取触发剧情分支的交互点。

17. 如权利要求1所述的方法，其中，呈现当前信息流包括：

语音播报所述当前信息流。

18. 一种语音交互方法，包括：

语音播报剧情故事；

语音播报用于触发不同剧情分支的多个选项；

获取用户对所述多个选项中一个选项的语音选择；以及

基于所述语音选择，触发与被选选项相对应的剧情分支。

19. 如权利要求18所述的方法，还包括：

获取用户在语音播报所述多个选项的其他时段内插入的语音输入；

基于所述语音输入，呈现剧情交互点；

获取所述用户针对所述剧情交互点的语音交互；以及

基于所述语音交互，生成或触发后续剧情分支。

20. 一种语音交互系统，服务端和多个终端，其中，

- 所述终端用于：
呈现从所述服务端获取的信息流；
采集来自用户的语音输入；
向所述服务端上传所述语音输入；
获取所述服务端下发的语音输入反馈；以及
基于所述语音输入反馈，呈现后续信息流，
所述服务端用于：
下发用于呈现的当前信息流；
获取所述终端上传的所述语音输入；
基于所述语音输入，生成并下发所述语音输入反馈。
21. 如权利要求20所述的系统，其中，所述服务端用于：
确定并下发用于后续信息流的呈现内容。
22. 如权利要求21所述的系统，其中，所述服务端用于：
获取所述语音输入的文本信息、语音属性信息和环境信息；以及
基于所述文本信息、语音属性信息和环境信息，确定并下发用于后续信息流的呈现内容。
23. 如权利要求20所述的系统，其中，所述终端用于：
向用户呈现用于触发不同剧情分支的多个选项；
采集用户对所述多个选项中一个选项的语音选择；以及
基于所述语音选择，呈现后续信息流。
24. 一种语音交互终端，包括：
呈现装置，用于呈现当前信息流；
输入装置，用于获取来自用户的语音输入；
处理装置，用于基于所述当前信息流和所述语音输入，确定后续信息流的呈现内容。
25. 如权利要求24所述的终端，还包括：
联网装置，用于：
获取要呈现的信息；
上传获取的所述语音输入；以及
获取用于确定后续信息流的语音输入反馈。
26. 如权利要求24所述的终端，其中，所述呈现装置包括：
语音输出装置，用于语音播报所述当前信息流和后续信息流。
27. 如权利要求26所述的终端，其中，所述呈现装置包括：
显示装置，用于显示所述当前信息流和/或后续信息流。
28. 如权利要求26所述的终端，其中，所述输入装置包括：
操作控制装置，用于获取所述用户的操作控制输入，并且
所述处理装置，用于：
基于所述操作控制输入，确定后续信息流的呈现内容。
29. 一种语音交互方法，包括：
呈现当前信息流；

获取来自多个用户的多个语音输入;以及
基于所述当前信息流和所述多个语音输入,确定后续信息流的呈现内容。

30. 如权利要求29所述的方法,其中,获取来自多个用户的多个语音输入包括如下至少一项:

针对相继呈现的不同当前信息流,分别获取来自不同的用户的语音输入;以及
针对一个当前信息流,获取来自不同用户的多个语音输入。

31. 如权利要求29所述的方法,还包括:

判定所述多个语音输入来自不同的用户,

其中,基于所述当前信息流和所述多个语音输入,确定后续信息流的呈现内容包括如下至少一项:

针对不同的用户,生成子信息流并确定所述子信息流的呈现内容;以及
综合判定所述多个语音输入的用户身份和输入内容,确定后续信息流的呈现内容。

32. 一种语音交互方法,包括:

呈现当前信息流;

获取来自用户的多轮语音输入;以及

基于所述当前信息流和所述多轮语音输入,确定后续信息流的呈现内容。

33. 如权利要求32所述的方法,其中,获取来自用户的多轮语音输入包括:

根据预定框架,呈现本轮交互内容;

获取用户针对本轮交互内容生成的本轮语音输入;以及

基于所述预定框架和/或所述本轮语音输入,呈现下轮交互内容。

34. 如权利要求32所述的方法,其中,所述信息流包括剧情故事,并且基于所述当前信息流和所述多轮语音输入,确定后续信息流的呈现内容包括:

基于所述多轮语音输入,构建所述剧情故事的剧情框架。

35. 一种计算设备,包括:

处理器;以及

存储器,其上存储有可执行代码,当所述可执行代码被所述处理器执行时,使所述处理器执行如权利要求1-17及29-34中任一项所述的方法。

36. 一种非暂时性机器可读存储介质,其上存储有可执行代码,当所述可执行代码被电子设备的处理器执行时,使所述处理器执行如权利要求1-17及29-34中任一项所述的方法。

语音交互方法、系统和终端

技术领域

[0001] 本公开涉及一种语音处理技术,尤其涉及一种语音交互方法、系统和终端。

背景技术

[0002] 随着语音交互技术的发展,可以使用语音指令进行各类控制与内容获取的智能音箱得到普及。使用智能音箱听歌、听故事等内容类功能颇受欢迎。在带屏音箱上,还可以结合视频、图片、文字、音频等多种媒介进行内容呈现。对于有剧情的内容节目,可用用户指令触发后一次性播报完成。例如,用户可以对智能音箱说话,“XXX,我要听故事”。智能音箱就会给用户播报一个故事,直到本集播放完毕。虽然用户可以进行诸如播放、暂停、选集等的操作,但这类操控性操作无法主动影响内容的走向,使得用户缺乏沉浸与参与感。

[0003] 因此,需要一种用户能够主动影响内容走向的交互方案。

发明内容

[0004] 本公开要解决的一个技术问题是提供一种用户能够通过语音交互主动影响内容走向的方案。用户通过语音输入,能够决定当前信息流的后续走向,尤其是能够通过语音输入决定剧情类游戏的剧情分支,从而增强用户的沉浸感和参与感,提升游戏的可玩性。

[0005] 根据本公开的第一个方面,提供了一种语音交互方法,包括:呈现当前信息流;获取来自用户的语音输入;基于所述当前信息流和所述语音输入,确定后续信息流的呈现内容。所述信息流可以是包括含剧情分支的信息流,或是包括可操控虚拟化身的信息流。

[0006] 根据本公开的第二个方面,提供了一种语音交互方法,包括:语音播报剧情故事;语音播报用于触发不同剧情分支的多个选项;获取用户对所述多个选项中一个选项的语音选择;以及基于所述语音选择,触发与被选选项相对应的剧情分支。

[0007] 根据本公开的第三个方面,提供了一种语音交互系统,服务端和多个终端,其中,所述终端用于:呈现从所述服务端获取的信息流;采集来自用户的语音输入;向所述服务端上传所述语音输入;获取所述服务端下发的语音输入反馈;以及基于所述语音输入反馈,呈现后续信息流,所述服务端用于:下发用于呈现的当前信息流;获取所述终端上传的所述语音输入;基于所述语音输入,生成并下发所述语音输入反馈。

[0008] 根据本公开的第四个方面,提供了一种语音交互终端,包括:呈现装置,用于呈现当前信息流;输入装置,用于获取来自用户的语音输入;处理装置,用于基于所述当前信息流和所述语音输入,确定后续信息流的呈现内容。

[0009] 根据本公开的第五个方面,提供了一种语音交互方法,包括:呈现当前信息流;获取来自多个用户的多个语音输入;基于所述当前信息流和所述多个语音输入,确定后续信息流的呈现内容。

[0010] 根据本公开的第六个方面,提供了一种语音交互方法,包括:呈现当前信息流;获取来自用户的多轮语音输入;基于所述当前信息流和所述多轮语音输入,确定后续信息流的呈现内容。

[0011] 根据本公开的第七个方面,提供了一种计算设备,包括:处理器;以及存储器,其上存储有可执行代码,当可执行代码被处理器执行时,使处理器执行如上述第一和第二方面以及第五和第六方面所述的方法。

[0012] 根据本公开的第八个方面,提供了一种非暂时性机器可读存储介质,其上存储有可执行代码,当可执行代码被电子设备的处理器执行时,使处理器执行如上述第一和第二方面以及第五和第六方面所述的方法。

[0013] 由此,,本发明能够实现以语音交互形式影响剧情走向,播报不同的音视频内容的技术效果。具体地,语音指令可以用于触发后续的内容播报,能够将语音识别出来的文本信息、执续时间、是否打断音箱播报、指令发出时间、语料情感等多维度信息一并做为后续播报内容的决策与生成依据。

附图说明

[0014] 通过结合附图对本公开示例性实施方式进行更详细的描述,本公开的上述以及其它目的、特征和优势将变得更加明显,其中,在本公开示例性实施方式中,相同的参考标号通常代表相同部件。

[0015] 图1示出了根据本发明一个实施例的语音交互方法的示意性流程图。

[0016] 图2示出了剧情分支结构的一个例子。

[0017] 图3示出了通过选择触发剧情分支的一个例子。

[0018] 图4示出了语音播报剧情故事进行交互的一个例子。

[0019] 图5示出了可以实施本发明的语音交互系统的组成示意图。

[0020] 图6示出了根据本发明一个实施例的语音交互终端的组成示意图。

具体实施方式

[0021] 下面将参照附图更详细地描述本公开的优选实施方式。虽然附图中显示了本公开的优选实施方式,然而应该理解,可以以各种形式实现本公开而不应被这里阐述的实施方式所限制。相反,提供这些实施方式是为了使本公开更加透彻和完整,并且能够将本公开的范围完整地传达给本领域的技术人员。

[0022] 随着语音交互技术的发展,可以使用语音指令进行各类控制与内容获取的智能音箱得到普及。使用智能音箱听歌、听故事等内容类功能颇受欢迎。在带屏音箱上,还可以结合视频、图片、文字、音频等多种媒介进行内容呈现。对于有剧情的内容节目,可用户指令触发后一次性播报完成。例如,用户可以对智能音箱说话,“XXX,我要听故事”。智能音箱就会给用户播报一个故事,直到本集播放完毕。虽然用户可以进行诸如播放、暂停、选集等的操作,但这类操控性操作无法主动影响内容的走向,使得用户缺乏沉浸与参与感。

[0023] 为此,本发明提供一种用户能够通过语音交互主动影响内容走向的方案。用户通过语音输入,能够决定当前信息流的后续走向,尤其是能够通过语音输入决定剧情类游戏的剧情分支,从而增强用户的沉浸感和参与感,提升游戏的可玩性。

[0024] 图1示出了根据本发明一个实施例的语音交互方法的示意性流程图。在某些实施例中,该方法可由语音交互终端通过与终端用户的交互实现。在更多的实施例中,语音交互终端还需部分借助云端的处理和/或存储能力来实现上述方案。

[0025] 在步骤S110中,呈现当前信息流。在此,“呈现”指代通过经由终端设备的各种感知方式让用户获知。在一个实施例中,呈现的信息流可以是在显示屏中显示的信息页面,例如,在手机安装的APP内显示。作为替换或是补充,上述呈现还可以包括声音呈现。此时,扬声器或是耳机可以播放相应的场景声音,例如,音乐、语音提示或是描述、或是模拟某一真实场景(例如,下雨、刮风等)的声音。在其他实施例中,还可以通过振动等其他方式进行信息流的呈现。

[0026] 在此,“信息流”指代能够更新呈现内容的信息。例如,智能音箱可以经由音频流进行故事朗读,此时朗读的故事可以看作是“信息流”。另外,在游戏场景中,基于用户输入而变换的内容,也可以看作是信息流。

[0027] 针对呈现的信息流,用户可以进行的相应的语音反馈。为此,在步骤S120,可以获取来自用户的语音输入。在获取了用户的语音输入之后,可以在步骤S130基于当前信息流和所述语音输入,确定后续信息流的呈现内容。由此,实现基于用户的语音输入,对后续信息流呈现的主动影响,由此增加用户与终端设备的互动体验,并通过语音参与增加沉浸感。

[0028] 由前所述,在本发明中“信息流”指代能够更新呈现内容的信息。在一个优选实现中,本发明的信息流尤其可以指代包括含剧情分支的信息流。例如,可以是包含剧情分支的游戏、剧集(例如,电视剧、电影、动画片等)或是小说。“剧情分支”是根据用户不同选择,导向不同剧情的玩法设计,也是文字冒险游戏和互动小说中最经典、重要的交互元素之一,可以让用户在抉择、改变中获得巨大的成就感。

[0029] 图2示出了剧情分支结构的一个例子。该剧情可以从一个公共序幕部分20开始,用来设定故事情节的场景,向玩游戏者介绍角色等等。在A到G的分支点上需要进行判断,因为关系到故事情节取哪个通路,从而用户通过可行的故事情节网到达四个可能的端点W到Z之一。相比于需要判断的分支点A-G(其中可以部分实现为下文提及的交互点),某些通路还可以在结点H,J和K处相结合,即,不同的剧情分支还可以回到相同的剧情线上,并自然向前行进,直至端点W到Z之一。

[0030] 在此,端点W到Z可以连接后续剧情,即图2所示分支结构可以是某一游戏或是剧集的整体分支结构的一部分。在其他实施例中,端点W到Z可以对应于剧情的四个结局,例如,在较为简单的游戏中。图中的通路22可以指代常规的剧情通路,而双线24则可以指代需要特定条件触发的剧情通路。

[0031] 不同的剧情分支可以在用户或是用户控制的虚拟化身符合某个条件(例如,升到一定级数)或是完成某一任务之后而被触发,也可以是在剧情交互点通过用户的多个选项进行选择来对触发。图3示出了通过选择触发剧情分支的一个例子。如图所示,在一个航海游戏中,用户所带领的船队遇到了食物不够的问题,此时在游戏交互点,可以向用户示出不同的应对选项。用户选择不同的选项,会进行到不同的分支,例如,导致此次航行任务失败或成功的分支等。

[0032] 本发明的语音交互方案尤其适用于针对如上所述是具有剧情分支的信息流,用户可以通过语音交互来影响信息流的剧情走向,从而达到沉浸式的交互效果。例如,在以语音播报呈现当前信息流时,可以向用户呈现用于触发不同剧情分支的多个选项,例如,用语音播报选项。随后,可以获取用户对所述多个选项中一个选项的语音选择。由此可以基于所述语音选择,触发与被选选项相对应的剧情分支。例如,在侦探类语音游戏中,可以语音播报

侦探故事,并推动用户去寻找线索。例如,智能语音设备可以播报“现在你来到一个岔路口,左侧通向森林,右侧通向河边,你要走哪条路?”。用户可以直接通过语音回复“左侧去森林”来进行选择,并基于上述选择行进至对应的剧情分支。由此,在语音播报信息流的情况下,通过引入用户的语音回复,增加了用户的参与度,并且能够让用户始终沉浸在语音播报所营造的氛围之内,提升了可玩性

[0033] 在其他实施例中,在呈现选项之前,还可以包括:向用户呈现用于触发不同剧情分支的交互点提示。随后,可以获取用户对所述交互点提示的语音指令,并且基于所述语音选择,触发所述交互点相对应的剧情分支或分支选项。在此,交互点可以指代用于引发不同剧情分支的交互发生处。例如,前述图3的图像显示选项和语音选项。在交互点不是剧情发展必然经过时,可以对上述交互点进行提示。例如,在语音播报例中,可以包括“路过一个大门,是否需要进行调查”。此时,进入大门本身可以不直接引发剧情分支,而是在用户语音选择进入大门内存在引发剧情分支的选项。用户可以通过后续对门内选项的选择,或是对门内线索的获知,来进行引发后续的剧情分支。

[0034] 对于获取的用户语音输入,首先可以获取该语音输入的文本信息,并将其转换为机器能够理解的指令。例如,上例中对“左侧”和/或“森林”的理解,从而实现例如现有游戏交互中基于点选等操作进行选择的效果,以确定后续信息流的呈现内容。除了作为已有交互方式(例如,鼠标点击、手指点击触摸屏)的代替,语音输入还可以为终端提供其自有的独特信息,以帮助后续信息流的确定或是生成。为此,获取来自用户的语音输入可以包括:获取所述语音输入的文本信息;获取所述语音输入的语音属性信息,并且确定后续信息流的呈现内容可以包括:基于所述文本信息和所述语音属性信息,确定后续信息流的呈现内容。除了呈现内容之外,还可以进一步地基于所述文本信息和所述语音属性信息,确定后续信息流的呈现方式。例如,在语音侦探游戏中,可以基于用户的语音输入所包含的各类信息,决定后续要播报的剧情分支,以及该剧情分支的呈现方式,例如,以更为神秘或是紧张的语调进行播报等等。

[0035] 具体地,语音属性信息可以指代语音输入的语义文本内容之外的、与输入的语音本身相关联的信息。

[0036] 在一个实施例中,语音属性信息可以是语音输入相对应当前信息流的起始时刻。如前所述,在步骤S130中,需要基于当前信息流的行进状况,再结合语音输入的相关指令来确定后续信息流的流向。在本实施例中,不仅是用户说了什么,在什么时候说也成为了后续信息流生成的判定来源。例如,在如上基于交互点和选择的分支确定的情况下,可以基于用户是否打断音箱播报来判定用户心态,并且给出更为贴合用户当前心态的后续呈现内容或其呈现方式。另外,在某些实施例中,可以允许用户在交互点之外的其他位置进行语音输入,例如图2中A-G交互点之外的通路22或24上,上述语音输入的内容和起始时刻也可作为后续剧情分支的选择标准。

[0037] 作为替换或者补充,语音属性信息可以是语音输入的持续时间。可以从上述持续时间中判断用户的语速或是心态,上述信息也可作为后续剧情分支的选择标准。

[0038] 同样作为作为替换或者补充,语音属性信息还可以是所述语音输入的情感信息和/或语调信息。在获取了用户的语音输入之后,可以对输入的语音本身进行分析,基于其说话的激烈程度、用词等来作为用户情感信息的判断,并且在后续给出与之相应的剧情分

支(例如,难度更大的玩法等)。另外,在表达相同文本意思的时候,用户也可以采取不同的遣词造句或是语气语调,这些语气语调同样可以用来进行后续剧情分支的生成、确定或是呈现方式的确定。例如,在用户使用川普(四川普通话)进行应答时,在后续游戏的交互中,进行虚拟人物对话时,也可以使用川普进行。而在诸如日语等不同身份的人使用不同句式进行沟通的语种中,也可以基于当前用户的遣词造句,选择对应的句式进行后续的信息流呈现。

[0039] 另外,语音属性信息还可以包括所述语音输入对应的用户身份。例如,用户可以对获取到的语音进行声纹比对,以确定用户的身份,例如在前录入的年龄、性别、偏好和积分等资料,并且可以基于确定的用户身份,决定后续的信息流呈现。例如,某一战斗类游戏可以具有R-13、R-18等不同版本,在验证了用户的声纹后,可以根据用户的年龄,确定是否开放如图2所示中需要特定条件触发的剧情通路24(例如,成年用户通路)。

[0040] 进一步地,本发明的语音交互方法还可以包括:获取所述语音输入生成时的环境信息,并且基于所述环境信息,确定后续信息流的呈现内容。上述环境信息可以是用户进行语音输入时所处小环境的信息,例如,家中温度、是否有其他人等,也可以是更大尺度上的环境信息,例如所处时段(节假日、上下班高峰期、深夜)、天气状况、地理位置信息等。上述信息同样可以用来确定后续信息流的生成或选择,及其呈现方式。

[0041] 另外,除了接收语音输入之外,本发明的语音交互方法还可以获取所述用户的非语音输入,并且基于所述非语音输入,确定后续信息流的呈现内容。例如,终端可以获取用户的体感输入(例如,基于体感传感器)或是影像输入(例如,基于3D摄像头),并结合语音输入进行综合判断。在例如,在某些交互点上可以通过语音输入进行交互,在其他输入点上可以基于例如鼠标或是屏幕点击来进行交互。

[0042] 作为应用与剧情分支类信息流的替换或者补充,本发明的语音交互方法还可以用于呈现虚拟化身的当前信息流。

[0043] 所述虚拟化身可以包括用户虚拟化身。例如,在经典RPG(角色扮演)游戏中,用户控制的游戏主角。此时,用户可以通过语音来控制其虚拟化身,例如,“向左走,出城门”等,由此代替繁琐的手部鼠标点选或是键盘控制。为此,获取来自用户的语音输入可以包括:获取所述用户对所述用户虚拟化身的语音控制,并且确定后续信息流的呈现内容包括:基于所述语音输入,控制所述用户虚拟化身的呈现。

[0044] 进一步地,所述虚拟化身包括用户虚拟化身之外的其他虚拟化身。其他虚拟化身可以是单机游戏中的虚拟人物或其他生物体,也可以是网络游戏中的其他真人用户的虚拟化身或是游戏中自带的虚拟人物或其他生物体。此时,获取来自用户的语音输入可以包括:获取所述用户与所述其他虚拟化身的语音交互,并且确定后续信息流的呈现内容包括:基于所述语音交互,控制所述其他虚拟化身的呈现。控制所述其他虚拟化身的呈现可以包括获取触发剧情分支的线索;和/或获取触发剧情分支的交互点。

[0045] 具体地,用户可以直接或通过其虚拟化身,与其他化身进行语音交谈,从而相比于现有的点击操作获取更强烈的沉浸感。用户与虚拟化身的对话内容、持续时间等,可以引发例如游戏中的线索人物提供线索,或是直接触发剧情分支的交互点,从而促进游戏的进行。

[0046] 如上结合图1描述了根据本发明的语音交互方案,该方案可以适用于各种被用户以语音交互主动影响能够带来更佳交互体验的信息流,上述信息流可以使用一种或多种方

式呈现,例如,通过图像和/或声音,并且可以接收包括语音交互在内的各种输入。

[0047] 在这其中,本发明尤其适于实现为一种用于语音播报剧情故事的交互方法。为此,终端可以语音播报剧情故事;并且语音播报用于触发不同剧情分支的多个选项;获取用户对所述多个选项中一个选项的语音选择;以及基于所述语音选择,触发与被选选项相对应的剧情分支。进一步地,还可以获取用户在语音播报所述多个选项的其他时段内插入的语音输入;基于所述语音输入,呈现剧情交互点;获取所述用户针对所述剧情交互点的语音交互;以及基于所述语音交互,生成或触发后续剧情分支。

[0048] 图4示出了语音播报剧情故事进行交互的一个例子。在此剧情故事例如可以是一个侦探故事,系统可以参考使用用户语音指令,决定后续播报内容,其主要流程如下:

[0049] 游戏/节目开始,开始播报内容。随后可以开始接收用户指令。用户指令可以在特定时间点进行接收,也可以在任意时间段内开始接收,换句话说,可以是当前播报内开始接收,也可以是在播报过程中开始接收。

[0050] 接收的指令可以是用户语音类的指令,对于语音类指令,会做相应的指令识别,转化为程序可理解的结构化数据,这里接收的信息可以包含识别出来的文本信息、语音持续是时间、是否在播报完成后发出指令、识别来的用户情感等。例如,在语音播报交互点的多个选项时(例如,播报“走左侧森林还是右侧河边”时),可以根据用户语音输出的文本内容(例如,“去森林”)来实现对后续剧情分支的明确选择。再例如,可以根据声纹判断的用户年龄(例如,已满18岁)以及用户语音输入所显示的情绪状态来确定是否需要开启“恐怖探案”隐藏模式(例如,图2所示的双线剧情通路24),以增加用户对游戏的深刻体验程度。

[0051] 接收的指令也可以是其他非语音事件触发,例如用户的鼠标、按键操作、地理位置变化等等,也可是用户无任何指令输入导致输入超时事件。例如,在带屏音箱的情况下,可以在屏幕上同时显示语音播报的内容选项,用户也可以通过点击触摸屏来完成选择。再例如,在更为深度的交互类游戏中,可以根据3D摄像机捕捉的用户形体动作、用户面部表情和用户语音交互,共同确定后续的游戏走向。

[0052] 根据上述指令信息与各类已知的上下文信息(如用信积分、所在城市等)决策产出播报内容,播报内容可能是事先准备好的、待选择的,也可以根据决策结果动态生成的。随后可以继续播报最新内容,直到满足结束条件。

[0053] 由此,本发明能够实现以语音交互形式影响剧情走向,播报不同的音视频内容的技术效果。具体地,语音指令可以用于触发后续的内容播报,能够将语音识别出来的文本信息、执续时间、是否打断音箱播报、指令发出时间、语料情感等多维度信息一并做为后续播报内容的决策与生成依据。

[0054] 在具体的应用场景中,本发明的语音交互方案还可以实现为涉及多用户交互,和/或多轮交互的更为复杂的语音交互和信息流呈现方案。

[0055] 在一个实施例中,本发明可以实现为一种语音交互方法,包括:呈现当前信息流;获取来自多个用户的多个语音输入;基于所述当前信息流和所述多个语音输入,确定后续信息流的呈现内容。

[0056] 在此,呈现信息流并获取语音输入可以是每次获取一个用户输入,剧情行进,并获取下一个用户输入的分获取方式,也可以是一次性获取多个用户输入的方式。于是,获取来自多个用户的多个语音输入包括如下至少一项:针对相继呈现的不同当前信息流,分别

获取来自不同的用户的语音输入;以及针对一个当前信息流,获取来自不同用户的多个语音输入。

[0057] 进一步地,本方法还可以包括:判定所述多个语音输入来自不同的用户。于是,基于所述当前信息流和所述多个语音输入,确定后续信息流的呈现内容包括如下至少一项:针对不同的用户,生成子信息流并确定所述子信息流的呈现内容;以及综合判定所述多个语音输入的用户身份和输入内容,确定后续信息流的呈现内容。

[0058] 在涉及多人的角色扮演游戏中,例如,三玩家A、B和C参与的游戏中,本方法可以在不同的询问点分别询问不同的玩家A、B和C,也可以同时询问这三个玩家。在同时询问时,如果三个玩家具有各自不相互干扰的语音输入设备,例如,各自佩戴有麦克风,则可以同时进行回答以便系统进行获取(例如,在线上游戏的情况下);而如果三个玩家面对一个语音交互设备,例如,一个智能音箱时,则优选在三个玩家不同时发声的情况下获取三个玩家的语音输入,并且根据语音输入的内容,以及输入的相继关系等,确定后续信息流的呈现内容。在线上游戏的情况下,可以分别为玩家A、B、C呈现(例如,语音播报)其各自的子信息流。而在本地游戏的情况下,则可以在同一个信息流中进行呈现。

[0059] 作为替换或者补充,在另一个实施例中,本发明可以实现为一种语音交互方法,包括:呈现当前信息流;获取来自用户的多轮语音输入;以及基于所述当前信息流和所述多轮语音输入,确定后续信息流的呈现内容。

[0060] 用户可以在系统的引导下,进行多轮语音输入。为此,获取来自用户的多轮语音输入可以包括:根据预定框架,呈现本轮交互内容;获取用户针对本轮交互内容生成的本轮语音输入;以及基于所述预定框架和/或所述本轮语音输入,呈现下轮交互内容。例如,所述信息流可以包括剧情故事,并且可以基于所述多轮语音输入,构建所述剧情故事的剧情框架。

[0061] 例如,在剧情故事播报的场景中,系统可以让用户选择或是确定故事发生的背景,例如,19世纪的伦敦,未来22世纪的虚拟世界等,随后用户可以选择故事类型,例如侦探推理类、喜剧类,甚至可以让用户确定主人公的性格特征等,由此使得用户能够深度参与故事创作,实现与系统的共创,进一步提升参与度和趣味性。

[0062] 如前所述,本发明的语音交互方法可由语音交互终端通过与终端用户的交互实现。在更多的实施例中,语音交互终端还需部分借助云端的处理和/或存储能力来实现上述方案。

[0063] 为此,本发明还可以实现为一种语音交互系统。图5示出了可以实施本发明的语音交互系统的组成示意图。如图所示,语音交互系统可以包括服务端510以及多个终端520。服务端510可以包括多个平台,来为海量的终端520提供本发明语音交互所涉及的各类服务。如图所示,终端520可以是各类智能音箱,例如图中示出的圆柱形智能音箱、带屏智能音箱或是移动智能终端,例如手机。

[0064] 在此,终端可以用于:呈现从所述服务端获取的信息流;采集来自用户的语音输入;向所述服务端上传所述语音输入;获取所述服务端下发的语音输入反馈;以及基于所述语音输入反馈,呈现后续信息流。在一个实施例中,所述终端可以是一个物理终端,例如图中示出的智能音箱和移动终端,其可以独自实现信息流呈现(例如,播报和显示)、语音采集、网络传输和后续呈现等功能,包括可以包括部分能够本地执行的处理能力。在其他实施例中,所述终端可以包括多个物理终端,例如,智能音箱可以与本地安装的智能语音贴短距

离通信,并基于上述语音贴完成语音的采集和播报,对此本发明不做限制。

[0065] 相应地,服务端510可以用于:下发用于呈现的当前信息流;获取所述终端上传的所述语音输入;基于所述语音输入,生成并下发所述语音输入反馈。

[0066] 在某些实施例中,终端520可以根据获取的语音输入反馈,从本机已有信息中确定后续信息流呈现内容,或是直接生成后续内容。在其他实施例中,服务端510可以用于确定并下发用于后续信息流的呈现内容,即,终端直接获取服务端下发的后续内容。

[0067] 进一步地,服务端510可以用于:获取所述语音输入的文本信息、语音属性信息和环境信息;以及基于所述文本信息、语音属性信息和环境信息,确定并下发用于后续信息流的呈现内容。

[0068] 所述终端510可以用于:向用户呈现用于触发不同剧情分支的多个选项;采集用户对所述多个选项中一个选项的语音选择;以及基于所述语音选择,呈现后续信息流。

[0069] 进一步地,本发明还可以实现为一种语音交互终端,用于实现如上结合图1和图4描述的语音交互方法。图6示出了根据本发明一个实施例的语音交互终端的组成示意图。该终端可以执行如上结合图1和图4描述的语音交互方法,或者至少是可以经由该终端并在云端的参与下完成执行。该终端也可以作为图5所示系统中的终端510。

[0070] 具体地,终端600可以包括呈现装置610、输入装置620以及处理装置630。

[0071] 呈现装置610可以用于呈现当前信息流。输入装置620可以用于获取来自用户的语音输入。处理装置630则可用于进行处理,例如基于所述当前信息流和所述语音输入,确定后续信息流的呈现内容。

[0072] 进一步地,在终端600需要与服务端交互以借助云端平台的处理能力进行语音处理、后续信息流的确定(包括选择和生成)时,终端600还可以包括联网装置640,用于获取要呈现的信息;上传获取的所述语音输入;以及获取用于确定后续信息流的语音输入反馈。在必要时,联网装置640还可以获取用于呈现的信息流,例如在前的游戏数据下载,或是随着信息流播放过程中的实时下载。

[0073] 在不同的实施例中,呈现装置610可以具有不同的形态。例如在某些实施例中,呈现装置可以包括显示装置,用于视觉输出要呈现的内容,例如显示所述当前信息流和/或后续信息流。作为替换或者补充,呈现装置还可以包括:语音输出装置,用于语音播报所述当前信息流和后续信息流。

[0074] 另外,输入装置620还可以包括:操作控制装置,用于获取所述用户的操作控制输入。例如,操作控制装置可以包括键盘、鼠标、触摸屏、游戏手柄等。随后,处理装置630可以用于:基于所述操作控制输入,确定后续信息流的呈现内容。

[0075] 进一步地,所述终端600可以实现为一种包含常规计算处理能力的计算设备。处理装置620可以实现为该计算设备的处理器,并且该计算设备还可以包括存储器,用于存储计算所需的数据和指令。

[0076] 处理器620可以是一个多核的处理器,也可以包含多个处理器。在一些实施例中,处理器620可以包含一个通用的主处理器以及一个或多个特殊的协处理器,例如图形处理器(GPU)、数字信号处理器(DSP)等等。在一些实施例中,处理器620可以使用定制的电路实现,例如特定用途集成电路(ASIC, Application Specific Integrated Circuit)或者现场可编程逻辑门阵列(FPGA, Field Programmable Gate Arrays)。

[0077] 存储器可以包括各种类型的存储单元,例如系统内存、只读存储器(ROM)和永久存储装置。其中,ROM可以存储处理器620或者计算机的其他模块需要的静态数据或者指令。永久存储装置可以是可读写的存储装置。永久存储装置可以是即使计算机断电后也不会失去存储的指令和数据的非易失性存储设备。在一些实施方式中,永久性存储装置采用大容量存储装置(例如磁或光盘、闪存)作为永久存储装置。另外一些实施方式中,永久性存储装置可以是可移除的存储设备(例如软盘、光驱)。系统内存可以是可读写存储设备或者易失性可读写存储设备,例如动态随机访问内存。系统内存可以存储一些或者所有处理器在运行时需要的指令和数据。此外,存储器可以包括任意计算机可读存储媒介的组合,包括各种类型的半导体存储芯片(DRAM,SRAM,SDRAM,闪存,可编程只读存储器),磁盘和/或光盘也可以采用。在一些实施方式中,存储器可以包括可读和/或写的可移除的存储设备,例如激光唱片(CD)、只读数字多功能光盘(例如DVD-ROM,双层DVD-ROM)、只读蓝光光盘、超密度光盘、闪存卡(例如SD卡、min SD卡、Micro-SD卡等等)、磁性软盘等等。计算机可读存储媒介不包含载波和通过无线或有线传输的瞬间电子信号。

[0078] 存储器上还可以存储有可执行代码,当可执行代码被处理器620处理时,可以使处理器620执行上文述及的语音交互方法。

[0079] 上文中已经参考附图详细描述了根据本发明的语音交互方法、系统和终端。本发明通过语音交互使得用户能够主动影响信息流的呈现,尤其适用于提升用户对剧情分支类信息流的沉浸式体验。进一步地,本发明的用户还可以经由虚拟化身与游戏中的人物进行语音对话,进一步提升游戏的代入感和可玩性。

[0080] 此外,根据本发明的方法还可以实现为一种计算机程序或计算机程序产品,该计算机程序或计算机程序产品包括用于执行本发明的上述方法中限定的上述各步骤的计算机程序代码指令。

[0081] 或者,本发明还可以实施为一种非暂时性机器可读存储介质(或计算机可读存储介质、或机器可读存储介质),其上存储有可执行代码(或计算机程序、或计算机指令代码),当所述可执行代码(或计算机程序、或计算机指令代码)被电子设备(或计算设备、服务端等)的处理器执行时,使所述处理器执行根据本发明的上述方法的各个步骤。

[0082] 本领域技术人员还将明白的是,结合这里的公开所描述的各种示例性逻辑块、模块、电路和算法步骤可以被实现为电子硬件、计算机软件或两者的组合。

[0083] 附图中的流程图和框图显示了根据本发明的多个实施例的系统和方法的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分,所述模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标记的功能也可以以不同于附图中所标记的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0084] 以上已经描述了本发明的各实施例,上述说明是示例性的,并非穷尽性的,并且也不限于所披露的各实施例。在不偏离所说明的各实施例的范围和精神的情况下,对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择,旨

在最好地解释各实施例的原理、实际应用或对市场中的技术的改进,或者使本技术领域的其它普通技术人员能理解本文披露的各实施例。

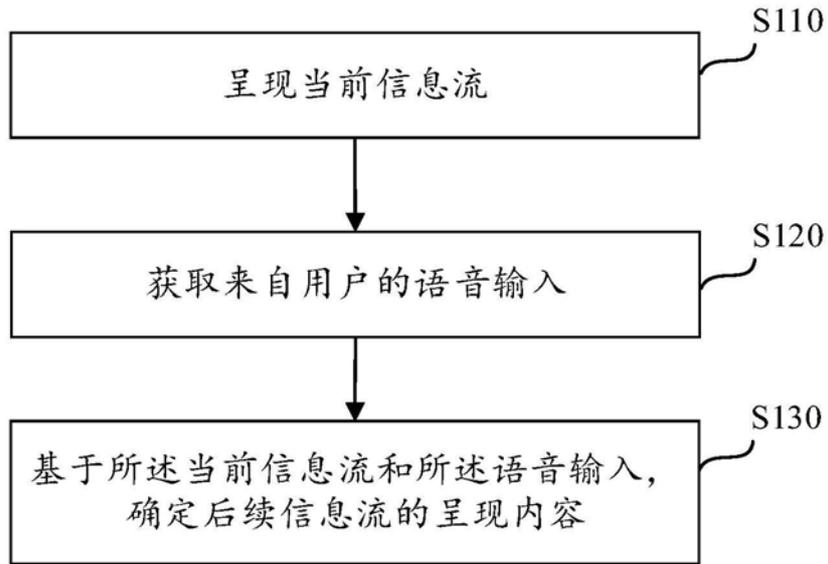


图1

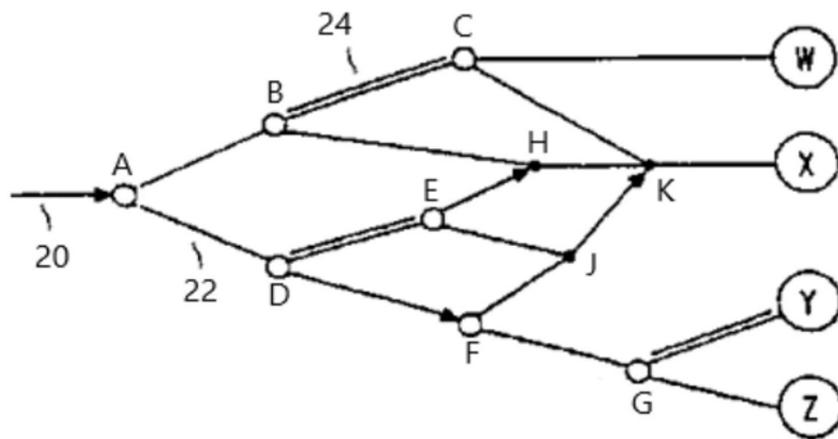


图2



图3

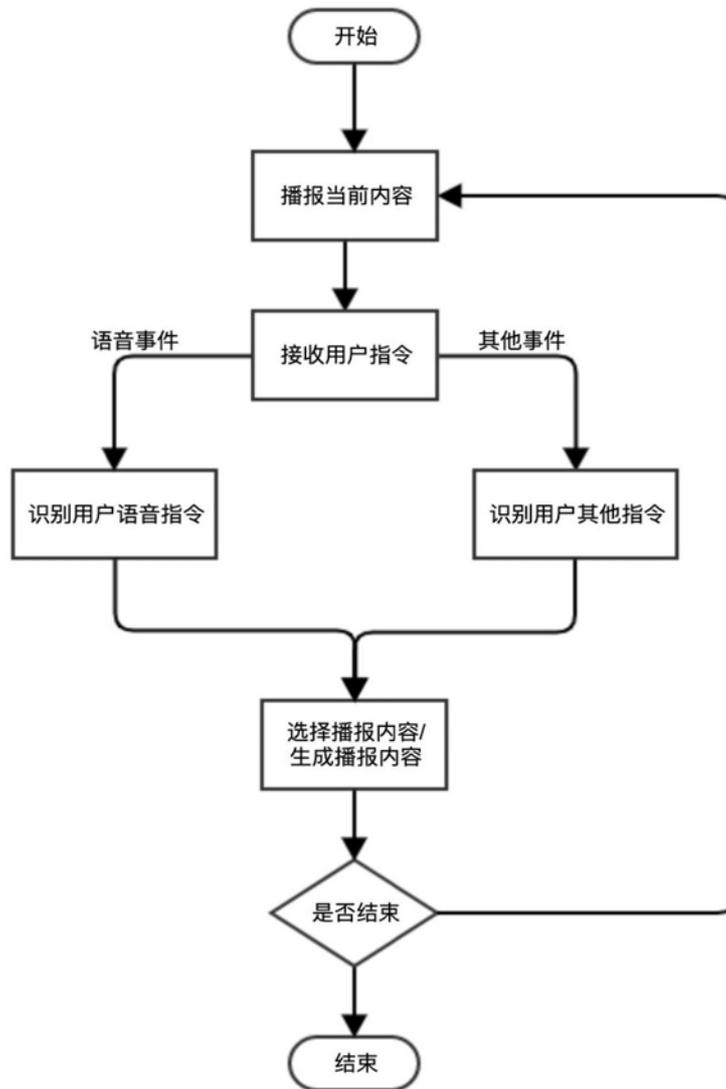


图4

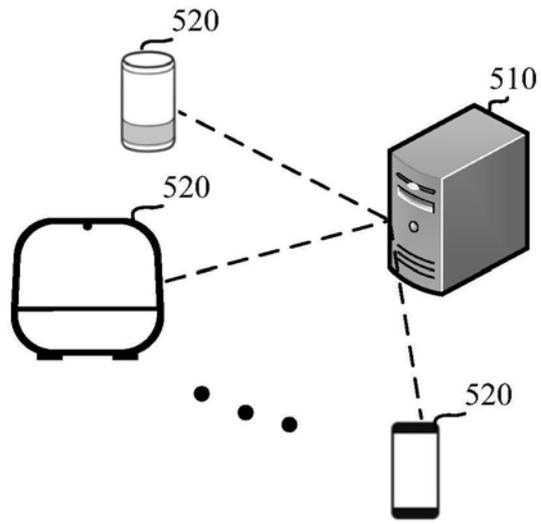


图5

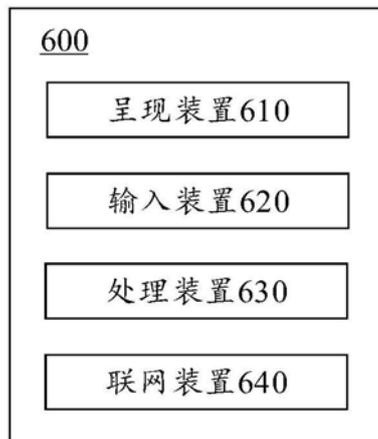


图6