



(12) 发明专利

(10) 授权公告号 CN 110930977 B

(45) 授权公告日 2022. 07. 08

(21) 申请号 201911103193.2

G10L 13/08 (2013.01)

(22) 申请日 2019.11.12

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 110930977 A

CN 107705783 A, 2018.02.16

CN 102005205 A, 2011.04.06

CN 106205602 A, 2016.12.07

(43) 申请公布日 2020.03.27

CN 105208194 A, 2015.12.30

CN 107705783 A, 2018.02.16

(73) 专利权人 北京搜狗科技发展有限公司
地址 100084 北京市海淀区中关村东路1号
院9号楼搜狐网络大厦9层01房间

审查员 何元

(72) 发明人 孟凡博 刘恺 陈伟

(74) 专利代理机构 北京润泽恒知识产权代理有
限公司 11319
专利代理师 郑傲日

(51) Int. Cl.

G10L 13/04 (2013.01)

G10L 13/047 (2013.01)

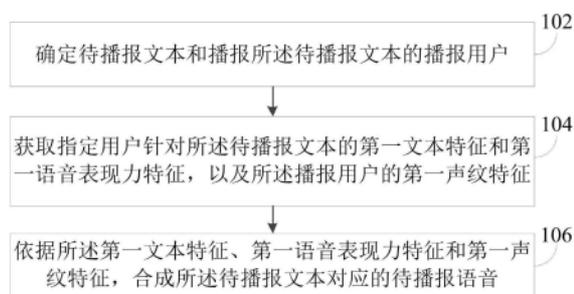
权利要求书5页 说明书17页 附图4页

(54) 发明名称

一种数据处理方法、装置和电子设备

(57) 摘要

本发明实施例提供了一种数据处理方法、装置和电子设备,其中,所述方法包括:确定待播报文本和播报所述待播报文本的播报用户;获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,以及所述播报用户的第一声纹特征;依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音;相对于训练模型提取播报用户的声学特征(如包括如语音表现力特征、声纹特征、文本特征等)而言,用于确定播报用户的声纹特征的数据更少,从而本发明实施例能够提高语音合成效率。



1. 一种数据处理方法,其特征在于,包括:

确定待播报文本和播报所述待播报文本的播报用户;

获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,以及所述播报用户的第一声纹特征;

依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音;

所述依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,包括:

将所述第一文本特征和第一语音表现力特征输入至训练后的预设模型中,得到第二声学特征,其中,所述第二声学特征包括第二声学语音表现力特征、第二声学文本特征和第二声纹特征;

采用所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征进行合成,得到所述待播报文本对应的待播报语音;

所述语音表现力特征用于表征用户朗读文本时的韵律。

2. 根据权利要求1所述的方法,其特征在于,所述获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,包括:

获取所述指定用户针对所述待播报文本的第一语音数据;

依据所述第一语音数据,提取所述指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征。

3. 根据权利要求2所述的方法,其特征在于,所述依据所述第一语音数据中,提取指定用户针对所述待播报文本所述第一文本特征,包括:

依据所述第一语音数据分别对所述待播报文本中各文字进行注音,得到所述待播报文本对应的注音信息序列;

以及依据所述第一语音数据,依次确定所述待播报文本中相邻两个文字之间的语音停顿信息;

依据所述注音信息序列和所述相邻两个文字之间的语音停顿信息,确定所述待播报文本对应的第一文本特征。

4. 根据权利要求1所述的方法,其特征在于,所述获取所述播报用户的第一声纹特征,包括:

获取所述播报用户针对参考文本的第二语音数据;

依据所述第二语音数据,提取所述播报用户针对所述参考文本的第二语音表现力特征、参考声学特征和第二文本特征;

依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,以使所述训练后的预设模型具有所述播报用户的第一声纹特征。

5. 根据权利要求4所述的方法,其特征在于,所述参考声学特征包括:参考声学语音表现力特征、参考声学文本特征和参考声纹特征;

所述依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,包括:

将所述第二语音表现力特征和第二文本特征输入至预设模型中,得到第一声学特征,

其中,所述第一声学特征包括第一声学语音表现力特征、第一声学文本特征和第一声纹特征;

将所述第一声学特征与所述参考声学特征进行比对,对所述预设模型的权重进行调整。

6. 根据权利要求1所述的方法,其特征在于,所述播报用户包括多个,一个播报用户对应该播报文本中的一个角色;

所述依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,包括:

依据各角色对应该播报文本的第一文本特征、第一语音表现力特征和对应播报用户的第一声纹特征,合成各角色对应该播报文本的待播报语音;

采用所述各角色对应该播报文本的待播报语音,生成所述待播报文本对应的待播报语音。

7. 根据权利要求1-6任一所述的方法,其特征在于,所述第一声纹特征包括以下至少一种:播报用户的音色特征、播报用户的口音特征和播报用户的方言特征。

8. 根据权利要求1所述的方法,其特征在于,所述的方法还包括:

对所述待播报语音进行播报。

9. 一种数据处理装置,其特征在于,包括:

确定模块,用于确定待播报文本和播报所述待播报文本的播报用户;

获取模块,用于获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,以及所述播报用户的第一声纹特征;

合成模块,用于依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音;

所述合成模块,包括:

第一语音合成子模块,用于将所述第一文本特征和第一语音表现力特征输入至训练后的预设模型中,得到第二声学特征,其中,所述第二声学特征包括第二声学语音表现力特征、第二声学文本特征和第二声纹特征;采用所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征进行合成,得到所述待播报文本对应的待播报语音;

所述语音表现力特征用于表征用户朗读文本时的韵律。

10. 根据权利要求9所述的装置,其特征在于,所述获取模块,包括:

第一语音获取子模块,用于获取所述指定用户针对所述待播报文本的第一语音数据;

第一特征提取子模块,用于依据所述第一语音数据,提取所述指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征。

11. 根据权利要求10所述的装置,其特征在于,

所述第一特征提取子模块,用于依据所述第一语音数据分别对所述待播报文本中各文字进行注音,得到所述待播报文本对应的注音信息序列;以及依据所述第一语音数据,依次确定所述待播报文本中相邻两个文字之间的语音停顿信息;依据所述注音信息序列和所述相邻两个文字之间的语音停顿信息,确定所述待播报文本对应的第一文本特征。

12. 根据权利要求9所述的装置,其特征在于,所述获取模块,包括:

第二语音获取子模块,用于获取所述播报用户针对参考文本的第二语音数据;

第二特征提取子模块,用于依据所述第二语音数据,提取所述播报用户针对所述参考文本的第二语音表现力特征、参考声学特征和第二文本特征;

模型训练子模块,用于依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,以使所述训练后的预设模型具有所述播报用户的第一声纹特征。

13. 根据权利要求12所述的装置,其特征在于,所述参考声学特征包括:参考声学语音表现力特征、参考声学文本特征和参考声纹特征;

所述模型训练子模块,用于将所述第二语音表现力特征和第二文本特征输入至预设模型中,得到第一声学特征,其中,所述第一声学特征包括第一声学语音表现力特征、第一声学文本特征和第一声纹特征;将所述第一声学特征与所述参考声学特征进行比对,对所述预设模型的权重进行调整。

14. 根据权利要求9所述的装置,其特征在于,所述播报用户包括多个,一个播报用户对应待播报文本中的一个角色;

所述合成模块,包括:

第二语音合成子模块,用于依据各角色对应待播报文本的第一文本特征、第一语音表现力特征和对应播报用户的第一声纹特征,合成各角色对应待播报文本的待播报语音;采用所述各角色对应待播报文本的待播报语音,生成所述待播报文本对应的待播报语音。

15. 根据权利要求9-14任一所述的装置,其特征在于,所述第一声纹特征包括以下至少一种:播报用户的音色特征、播报用户的口音特征和播报用户的方言特征。

16. 根据权利要求9所述的装置,其特征在于,所述的装置还包括:

播报模块,用于对所述待播报语音进行播报。

17. 一种可读存储介质,其特征在于,当所述存储介质中的指令由电子设备的处理器执行时,使得电子设备能够执行如方法权利要求1-8任一所述的数据处理方法。

18. 一种电子设备,其特征在于,包括有存储器,以及一个或者一个以上的程序,其中一个或者一个以上程序存储于存储器中,且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行以下操作的指令:

确定待播报文本和播报所述待播报文本的播报用户;

获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,以及所述播报用户的第一声纹特征;

依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音;

所述依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,包括:

将所述第一文本特征和第一语音表现力特征输入至训练后的预设模型中,得到第二声学特征,其中,所述第二声学特征包括第二声学语音表现力特征、第二声学文本特征和第二声纹特征;

采用所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征进行合成,得到所述待播报文本对应的待播报语音;

所述语音表现力特征用于表征用户朗读文本时的韵律。

19. 根据权利要求18所述的电子设备,其特征在于,所述获取指定用户针对所述待播报

文本的第一文本特征和第一语音表现力特征,包括:

获取所述指定用户针对所述待播报文本的第一语音数据;

依据所述第一语音数据,提取所述指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征。

20.根据权利要求19所述的电子设备,其特征在于,所述依据所述第一语音数据中,提取指定用户针对所述待播报文本所述第一文本特征,包括:

依据所述第一语音数据分别对所述待播报文本中各文字进行注音,得到所述待播报文本对应的注音信息序列;

以及依据所述第一语音数据,依次确定所述待播报文本中相邻两个文字之间的语音停顿信息;

依据所述注音信息序列和所述相邻两个文字之间的语音停顿信息,确定所述待播报文本对应的第一文本特征。

21.根据权利要求18所述的电子设备,其特征在于,所述获取所述播报用户的第一声纹特征,包括:

获取所述播报用户针对参考文本的第二语音数据;

依据所述第二语音数据,提取所述播报用户针对所述参考文本的第二语音表现力特征、参考声学特征和第二文本特征;

依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,以使所述训练后的预设模型具有所述播报用户的第一声纹特征。

22.根据权利要求21所述的电子设备,其特征在于,所述参考声学特征包括:参考声学语音表现力特征、参考声学文本特征和参考声纹特征;

所述依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,包括:

将所述第二语音表现力特征和第二文本特征输入至预设模型中,得到第一声学特征,其中,所述第一声学特征包括第一声学语音表现力特征、第一声学文本特征和第一声纹特征;

将所述第一声学特征与所述参考声学特征进行比对,对所述预设模型的权重进行调整。

23.根据权利要求18所述的电子设备,其特征在于,所述播报用户包括多个,一个播报用户对应该待播报文本中的一个角色;

所述依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,包括:

依据各角色对应该待播报文本的第一文本特征、第一语音表现力特征和对应播报用户的第一声纹特征,合成各角色对应该待播报文本的待播报语音;

采用所述各角色对应该待播报文本的待播报语音,生成所述待播报文本对应的待播报语音。

24.根据权利要求18-23任一所述的电子设备,其特征在于,所述第一声纹特征包括以下至少一种:播报用户的音色特征、播报用户的口音特征和播报用户的方言特征。

25.根据权利要求18所述的电子设备,其特征在于,还包含用于进行以下操作的指令:

对所述待播报语音进行播报。

一种数据处理方法、装置和电子设备

技术领域

[0001] 本发明涉及数据处理技术领域,特别是涉及一种数据处理方法、装置和电子设备。

背景技术

[0002] 随着终端技术的不断发展,终端中提供的功能也多种多样,如拨打电话、查看/发送短消息、观看视频、语音播报等等。

[0003] 通常,终端提供了多种声音进行语音播报,如男声、女生、童声等,以提高语音播报的趣味性。其中,为了更好的满足用户需求,终端还可以为用户定制语音,然后采用定制的语音进行语音播报。

[0004] 但是,在为用户定制语音的过程中,需要采集该用户的大量语音数据,然后采用这些语音数据训练模型,才能够使得训练后的模型能够提取该用户的声学特征(如语音表现力特征、声纹特征等);进而导致基于训练后的模型提取的声学特征合成语音的效率低。

发明内容

[0005] 本发明实施例提供一种数据处理方法,以提高语音合成的效率。

[0006] 相应的,本发明实施例还提供了一种数据处理装置和一种电子设备,用以保证上述方法的实现及应用。

[0007] 为了解决上述问题,本发明实施例公开了一种数据处理方法,具体包括:确定待播报文本和播报所述待播报文本的播报用户;获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,以及所述播报用户的第一声纹特征;依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音。

[0008] 可选地,所述获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,包括:获取所述指定用户针对所述待播报文本的第一语音数据;依据所述第一语音数据,提取所述指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征。

[0009] 可选地,所述依据所述第一语音数据中,提取指定用户针对所述待播报文本所述第一文本特征,包括:依据所述第一语音数据分别对所述待播报文本中各文字进行注音,得到所述待播报文本对应的注音信息序列;以及依据所述第一语音数据,依次确定所述待播报文本中相邻两个文字之间的语音停顿信息;依据所述注音信息序列和所述相邻两个文字之间的语音停顿信息,确定所述待播报文本对应的第一文本特征。

[0010] 可选地,所述获取所述播报用户的第一声纹特征,包括:获取所述播报用户针对参考文本的第二语音数据;依据所述第二语音数据,提取所述播报用户针对所述参考文本的第二语音表现力特征、参考声学特征和第二文本特征;依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,以使所述训练后的预设模型具有所述播报用户的第一声纹特征。

[0011] 可选地,所述参考声学特征包括:参考声学语音表现力特征、参考声学文本特征和参考声纹特征;所述依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预

设模型,包括:将所述第二语音表现力特征和第二文本特征输入至预设模型中,得到第一声学特征,其中,所述第一声学特征包括第一声学语音表现力特征、第一声学文本特征和第一声纹特征;将所述第一声学特征与所述参考声学特征进行比对,对所述预设模型的权重进行调整。

[0012] 可选地,所述依据所述第一文本特征、第一表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,包括:将所述第一文本特征和第一语音表现力特征输入至训练后的预设模型中,得到第二声学特征,其中,所述第二声学特征包括第二声学语音表现力特征、第二声学文本特征和第二声纹特征;采用所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征进行合成,得到所述待播报文本对应的待播报语音。

[0013] 可选地,所述播报用户包括多个,一个播报用户对应待播报文本中的一个角色;所述依据所述第一文本特征、第一表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,包括:依据各角色对应待播报文本的第一文本特征、第一表现力特征和对应播报用户的第一声纹特征,合成各角色对应待播报文本的待播报语音;采用所述各角色对应待播报文本的待播报语音,生成所述待播报文本对应的待播报语音。

[0014] 可选地,所述第一声纹特征包括以下至少一种:播报用户的音色特征、播报用户的口音特征和播报用户的方言特征。

[0015] 可选地,所述的方法还包括:对所述待播报语音进行播报。

[0016] 本发明实施例还公开了一种数据处理装置,具体包括:确定模块,用于确定待播报文本和播报所述待播报文本的播报用户;获取模块,用于获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,以及所述播报用户的第一声纹特征;合成模块,用于依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音。

[0017] 可选地,所述获取模块,包括:第一语音获取子模块,用于获取所述指定用户针对所述待播报文本的第一语音数据;第一特征提取子模块,用于依据所述第一语音数据,提取所述指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征。

[0018] 可选地,所述第一特征提取子模块,用于依据所述第一语音数据分别对所述待播报文本中各文字进行注音,得到所述待播报文本对应的注音信息序列;以及依据所述第一语音数据,依次确定所述待播报文本中相邻两个文字之间的语音停顿信息;依据所述注音信息序列和所述相邻两个文字之间的语音停顿信息,确定所述待播报文本对应的第一文本特征。

[0019] 可选地,所述获取模块,包括:第二语音获取子模块,用于获取所述播报用户针对参考文本的第二语音数据;第二特征提取子模块,用于依据所述第二语音数据,提取所述播报用户针对所述参考文本的第二语音表现力特征、参考声学特征和第二文本特征;模型训练子模块,用于依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,以使所述训练后的预设模型具有所述播报用户的第一声纹特征。

[0020] 可选地,所述参考声学特征包括:参考声学语音表现力特征、参考声学文本特征和参考声纹特征;所述模型训练子模块,用于将所述第二语音表现力特征和第二文本特征输入至预设模型中,得到第一声学特征,其中,所述第一声学特征包括第一声学语音表现力特征、第一声学文本特征和第一声纹特征;将所述第一声学特征与所述参考声学特征进行比

对,对所述预设模型的权重进行调整。

[0021] 可选地,所述合成模块,包括:第一语音合成子模块,用于将所述第一文本特征和第一语音表现力特征输入至训练后的预设模型中,得到第二声学特征,其中,所述第二声学特征包括第二声学语音表现力特征、第二声学文本特征和第二声纹特征;采用所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征进行合成,得到所述待播报文本对应的待播报语音。

[0022] 可选地,所述播报用户包括多个,一个播报用户对应该待播报文本中的一个角色;所述合成模块,包括:第二语音合成子模块,用于依据各角色对应待播报文本的第一文本特征、第一表现力特征和对应播报用户的第一声纹特征,合成各角色对应待播报文本的待播报语音;采用所述各角色对应待播报文本的待播报语音,生成所述待播报文本对应的待播报语音。

[0023] 可选地,所述第一声纹特征包括以下至少一种:播报用户的音色特征、播报用户的口音特征和播报用户的方言特征。

[0024] 可选地,所述的装置还包括:播报模块,用于对所述待播报语音进行播报。

[0025] 本发明实施例还公开了一种可读存储介质,当所述存储介质中的指令由电子设备的处理器执行时,使得电子设备能够执行如本发明实施例任一所述的数据处理方法。

[0026] 本发明实施例还公开了一种电子设备,包括有存储器,以及一个或者一个以上的程序,其中一个或者一个以上程序存储于存储器中,且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行以下操作的指令:确定待播报文本和播报所述待播报文本的播报用户;获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,以及所述播报用户的第一声纹特征;依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音。

[0027] 可选地,所述获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,包括:获取所述指定用户针对所述待播报文本的第一语音数据;依据所述第一语音数据,提取所述指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征。

[0028] 可选地,所述依据所述第一语音数据中,提取指定用户针对所述待播报文本所述第一文本特征,包括:依据所述第一语音数据分别对所述待播报文本中各文字进行注音,得到所述待播报文本对应的注音信息序列;以及依据所述第一语音数据,依次确定所述待播报文本中相邻两个文字之间的语音停顿信息;依据所述注音信息序列和所述相邻两个文字之间的语音停顿信息,确定所述待播报文本对应的第一文本特征。

[0029] 可选地,所述获取所述播报用户的第一声纹特征,包括:获取所述播报用户针对参考文本的第二语音数据;依据所述第二语音数据,提取所述播报用户针对所述参考文本的第二语音表现力特征、参考声学特征和第二文本特征;依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,以使所述训练后的预设模型具有所述播报用户的第一声纹特征。

[0030] 可选地,所述参考声学特征包括:参考声学语音表现力特征、参考声学文本特征和参考声纹特征;所述依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,包括:将所述第二语音表现力特征和第二文本特征输入至预设模型中,得到第一声学特征,其中,所述第一声学特征包括第一声学语音表现力特征、第一声学文本特征和第一

声纹特征;将所述第一声学特征与所述参考声学特征进行比对,对所述预设模型的权重进行调整。

[0031] 可选地,所述依据所述第一文本特征、第一表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,包括:将所述第一文本特征和第一语音表现力特征输入至训练后的预设模型中,得到第二声学特征,其中,所述第二声学特征包括第二声学语音表现力特征、第二声学文本特征和第二声纹特征;采用所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征进行合成,得到所述待播报文本对应的待播报语音。

[0032] 可选地,所述播报用户包括多个,一个播报用户对应待播报文本中的一个角色;所述依据所述第一文本特征、第一表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,包括:依据各角色对应待播报文本的第一文本特征、第一表现力特征和对应播报用户的第一声纹特征,合成各角色对应待播报文本的待播报语音;采用所述各角色对应待播报文本的待播报语音,生成所述待播报文本对应的待播报语音。

[0033] 可选地,其特征在于,所述第一声纹特征包括以下至少一种:播报用户的音色特征、播报用户的口音特征和播报用户的方言特征。

[0034] 可选地,还包含用于进行以下操作的指令:对所述待播报语音进行播报。

[0035] 本发明实施例包括以下优点:

[0036] 本发明实施例中,在确定待播报文本和播报所述待播报文本的播报用户后,可以获取依据指定用户针对所述待播报文本的第一语音表现力特征和第一文本特征,以及所述播报用户的第一声纹特征;然后通过韵律迁移,将指定用户朗读所述待播报文本时的第一语音表现力特征和第一文本特征,作为播报用户播报所述待播报文本时的语音表现力特征和文本特征,再依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成待播报语音使得合成的待播报语音即具有语音表现力又具有播报用户声纹特征;相对于训练模型提取播报用户的声学特征(如包括如语音表现力特征、声纹特征、文本特征等)而言,用于确定播报用户的声纹特征的数据更少,从而本发明实施例能够提高语音合成效率。

附图说明

[0037] 图1是本发明的一种数据处理方法实施例的步骤流程图;

[0038] 图2是本发明的一种数据处理方法可选实施例的步骤流程图;

[0039] 图3是本发明的一种数据处理装置实施例的结构框图;

[0040] 图4是本发明的一种数据处理装置可选实施例的结构框图;

[0041] 图5根据一示例性实施例示出的一种用于数据处理的电子设备的结构框图;

[0042] 图6是本发明根据另一示例性实施例示出的一种用于数据处理的电子设备的结构示意图。

具体实施方式

[0043] 为使本发明的上述目的、特征和优点能够更加明显易懂,下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0044] 本发明实施例的核心构思之一是,通过韵律迁移的方式,合成既具有语音表现力又具有播报用户声纹特征的语音数据,实现提高语音合成的效率。

[0045] 参照图1,示出了本发明的一种数据处理方法实施例的步骤流程图,具体可以包括如下步骤:

[0046] 步骤102、确定待播报文本和播报所述待播报文本的播报用户。

[0047] 步骤104、获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,以及所述播报用户的第一声纹特征。

[0048] 步骤106、依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音。

[0049] 本发明实施例中,当终端用户需要进行播报时,可以从终端提供的可用于播报的多段文本中选取一段文本,以及选取播报这段文本的用户,以使终端播报具有其选取的用户特色的语音。待用户执行选取操作后,终端可以将终端用户选取的文本确定为待播报文本,以及将终端用户选取的用户确定为播报所述待播报文本的播报用户。其中,所述终端用户与播报用户可以是同一用户,也可以是不同用户,本发明实施例对此不作限制。所述可用于播报的文本可以包括各种类型的文本,如小说、儿童故事等,本发明实施例对此不作限制。

[0050] 待确定待播报文本和对应的播报用户后,可以获取指定用户朗读所述待播报文本时的第一语音表现力特征和第一文本特征;以及播报用户的第一声纹特征。其中,所述指定用户可以是指具有高语音表现力的用户,如播音员,本发明实施例对此不作限制。其中,语音表现力特征和文本特征均可以用于表征用户朗读文本时的韵律如情感、风格等等;所述语音表现力特征可以包括多种,如语调特征、情感特征、风格特征等;所述文本特征可以包括用于描述文本被朗读时与文本相关的特征,可以包括多种特征,如被朗读时文本的拼音、被朗读时文本中相邻两个文字之间的停顿时间等;本发明实施例对此不作限制。所述声纹特征可以包括用于表征用户声音特性的特征如音色特征、口音特征等;本发明实施例对此不作限制。

[0051] 然后可以进行韵律迁移,将指定用户朗读所述待播报文本时的第一语音表现力特征和第一文本特征,作为播报用户播报所述待播报文本时的语音表现力特征和文本特征;再基于所述依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音;进而能够合成既具有语音表现力又具有播报用户声纹特征的待播报语音。相对于训练模型提取播报用户的声学特征(如包括语音表现力特征、声纹特征、文本特征等)而言,用于确定播报用户的声纹特征的数据更少,进而能够提高语音合成的效率。

[0052] 综上,本发明实施例中,在确定待播报文本和播报所述待播报文本的播报用户后,可以获取依据指定用户针对所述待播报文本的第一语音表现力特征和第一文本特征,以及所述播报用户的第一声纹特征;然后通过韵律迁移,将指定用户朗读所述待播报文本时的第一语音表现力特征和第一文本特征,作为播报用户播报所述待播报文本时的语音表现力特征和文本特征,再依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成待播报语音使得合成的待播报语音即具有语音表现力又具有播报用户声纹特征;相对于训练模型提取播报用户的声学特征(如包括如语音表现力特征、声纹特征、文本特征等)而言,用于确定播报用户的声纹特征的数据更少,从而本发明实施例能够提高语音合成效率。

[0053] 参照图2,示出了本发明的一种数据处理方法可选实施例的步骤流程图,具体可以

包括如下步骤：

[0054] 步骤202、确定待播报文本和播报所述待播报文本的播报用户。

[0055] 本发明实施例中，终端可以提供多个用于语音播报的预制用户，终端用户可以从多个预制用户中选取播报用户。若终端用户本次之前自定义了用户，则终端用户还可以从已自定义的用户中选取播报用户。当然，当终端用户确定预制用户和已自定义的用户均无法满足需求时，可以从触发终端提供的播报用户自定义功能，重新自定义播报用户；本发明实施例对此不作限制。

[0056] 步骤204、获取所述指定用户针对所述待播报文本的第一语音数据。

[0057] 本发明实施例中，可以预先由指定用户针对终端数据库中的每一段可用于播报的文本进行朗读；在指定用户朗读过程中，终端可以录制每一段可用于播报的文本对应的语音数据，后续可以将指定用户针对每一段可用于播报的文本的语音数据称为第一语音数据。然后可以建立每段可用于播报的文本与对应第一语音数据之间的第一映射关系，并存储第一语音数据和第一映射关系。进而在确定待播报文本后，可以基于所述待播报文本查找第一映射关系，查找所述指定用户针对所述待播报文本的第一语音数据。

[0058] 步骤206、依据所述第一语音数据，提取所述指定用户针对所述待播报文本的第一文本特征。

[0059] 本发明实施例的一种实现中，可以参照如下子步骤获取所述待播报文本对应的第一文本特征：

[0060] 子步骤22、依据所述第一语音数据分别对所述待播报文本中各文字进行注音，得到所述待播报文本对应的注音信息序列。

[0061] 子步骤24、依据所述第一语音数据，依次确定所述待播报文本中相邻两个文字之间的语音停顿信息。

[0062] 子步骤26、依据所述注音信息序列和所述相邻两个文字之间的语音停顿信息，确定所述待播报文本对应的第一文本特征。

[0063] 本发明实施例中，所述待播报文本可以包括多个文字，在获取所述待播报文本的第一语音数据后，可以依据所述第一语音数据，可以分别针对所述待播报文本中的每一个文字进行注音，得到每个文字对应的注音信息。然后依据所述待播报文本中各文字的顺序，将各文字对应的注音信息进行拼接，得到所述待播报文本的注音信息序列。其中，本发明实施例对所述待播报文本的语种不作限制；当所述待播报文本的语种不同时，所述注音信息序列不同。例如当所述待播报文本为汉语时，注音信息为拼音，对应的注音信息序列可以是拼音序列；当所述待播报文本为英语时，注音信息为音标，对应的注音信息序列可以是音标序列；等等。当然，当所述待播报文本为汉语时，所述每个文字对应的注音信息除了包括拼音外，还可以包括对应的声调；本发明对此不作限制。

[0064] 以及可以从所述待播报文本的第一语音数据中，提取所述待播报文本中每一对相邻的两个文字之间的语音停顿时间，然后依据该对相邻的两个文字之间的语音停顿时间，确定该对相邻的两个文字之间的语音停顿信息。其中，所述语音停顿时间可以是指指定用户朗读所述待播报文本过程中连续发出该对相邻的两个文字中第一个字的语音和发出第二个字的语音的时间差。本发明的一个示例中，可以直接将语音停顿时间确定为语音停顿信息；例如，待播报文本为：“八仙过海。相传，有一个叫做蓬莱阁的地方……。”；指定用户连

续发出“八”和“仙”这两个文字对应语音的时间差为5ms;则可以确定这“八”和“仙”两个文字之间的语音停顿时间为5ms。指定用户连续发出“仙”和“海”这两个文字的语音对应语音时间差为10ms;则可以确定“仙”和“海”这两个文字之间的语音停顿时间为10ms。指定用户连续发出“海”和“相”这两个文字对应语音的语音时间差为20ms;则可以确定“海”和“相”这两个文字之间的语音停顿时间为20ms。以此类推。然后可以确定“八”和“仙”的语音停顿信息为5ms,“仙”和“海”的语音停顿信息为10ms,“海”和“相”的语音停顿信息为20ms。当然也可以预先建立语音停顿等级与语音停顿时间的关系,例如语音停顿等级1-语音停顿时间10ms,语音停顿等级2-语音停顿时间15ms,语音停顿等级3-语音停顿时间20ms等等。然后依据语音停顿时间,确定语音停顿等级;将所述语音停顿等级确定为语音停顿信息。例如,上述示例中,“仙”和“海”的语音停顿信息为语音停顿等级1,“海”和“相”的语音停顿信息为语音停顿等级3。

[0065] 然后依据所述注音信息序列和所述相邻的两个文字之间的语音停顿信息,确定所述待播报文本对应的第一文本特征。例如,对所述待播报文本中每一对相邻的两个文字,在这两个文字的前一个文字的注音信息和后一个文字的注音信息之间,添加这两个文字之间的语音停顿信息,进而可以得到所述待播报文本对应的第一文本特征;本发明实施例对此不作限制。

[0066] 本发明的一个可选实施例中,若语音停顿信息为语音停顿时间,则当待播报文本中每一对相邻的两个文字之间语音停顿时间,小于语音停顿时间阈值时,可以无需在这两个文字的前一个文字的注音信息和后一个文字的注音信息之间,添加这两个文字之间的语音停顿时间。其中,所述语音停顿时间阈值可以按照需求设置,本发明实施例对此不作限制。

[0067] 本发明一个可选实施例中,也可以预先确定终端数据库中每一段可用于播报的文本对应的第一文本特征;并建立可用于播报的文本与对应的第一文本特征对应的第二映射关系;然后将所述第一文本特征和第二映射关系存储在数据库中。进而可以基于所述待播报文本查找所述第二映射关系,确定与所述待播报文本匹配的第一文本特征。其中,预先确定终端数据中每一段可用于播报的文本对应的第一文本特征的方式,与上述子步骤22-26类似,在此不再赘述。

[0068] 步骤208、依据所述第一语音数据,提取所述指定用户针对所述待播报文本的第一语音表现力特征。

[0069] 本发明实施例中,在基于所述待播报文本查找第一映射关系,查找所述指定用户针对所述待播报文本的第一语音数据后,可以从所述查找的第一语音数据中,提取对应的第一语音表现力特征;其中,可以采用现有的模型如深度学习模型进行语音表现力特征的提取,本发明实施例对此不作限制。所述语音表现力特征可以包括多种,如语速特征、语调特征、每个文字的发音时长特征、情感特征等等,本发明实施例对此不作限制。

[0070] 本发明的另一个可选实施例中,可以在每次录制指定用户朗读终端数据库中每一段可用于播报的文本的第一语音数据后,从该第一语音数据中,提取对应的第一语音表现力特征;然后可以建立每段可用于播报的文本与对应第一语音表现力特征之间的第三映射关系,并存储第一语音表现力特征和第三映射关系。进而可以基于所述待播报文本查找第三映射关系,确定与所述待播报数据匹配的第一语音表现力特征。当然,还可以包括其他获

取得播报文本对应的第一语音表现力特征的方式,本发明实施例对此不作限制。

[0071] 此外,为提高后续语音播报的趣味性,所述指定用户针对数据库中的每一段可用于播报的文本,可以进行高语音表现力的朗读;进而可以从具有高语音表现力的第一语音数据中,提取出具有高语音表现力的第一语音表现力特征,提高后续基于第一语音表现力特征合成的待播报语音的表现力。

[0072] 步骤210、获取所述播报用户针对参考文本的第二语音数据。

[0073] 步骤212、依据所述第二语音数据,提取所述播报用户针对所述参考文本的第二语音表现力特征、参考声学特征和第二文本特征。

[0074] 步骤214、依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,以使所述训练后的预设模型具有所述播报用户的第一声纹特征。

[0075] 其中,在自定义用户的过程中,可以由自定义的用户针对参考文本进行朗读;在自定义的用户朗读过程中,终端可以录制参考文本对应的语音数据;并将自定义的用户针对参考文本的语音数据称为第二语音数据。其中,所述参考文本可以是指符合预设条件的文本,所述预设条件可以按照需求设置,例如文字数量大于预设字数,所述预设字数可以按照需求设置;又例如朗读时长大于预设时长,所述预设时长可以按照需求设置;本发明实施例对此不作限制。所述参考文本与待播报文本可以相同也可以不同,本发明实施例对此不作限制。

[0076] 本发明的一种实施例中,可以建立自定义的用户与对应第二语音数据之间的第四映射关系,并存储第二语音数据和第四映射关系。在确定播报用户为自定义的用户(可以包括本次自定义的用户,也可以包括本次之前自定义的用户)后,可以基于播报用户查找第四映射关系,确定对应的第二语音数据。然后依据所述播报用户对应的第二语音数据,确定所述播报用户对应的第一声纹特征。

[0077] 本发明实施例中,可以获取通用的可输出声学特征的预设模型;其中,所述声学特征包括声学语音表现力特征,声学文本特征和声纹特征。所述声学语音表现力特征是语音表现力特征对应于声学维度的特征;声学语音表现力特征与语音表现力特征的区别在于,声学语音表现力特征可直接用于合成语音。所述声学文本特征是所述文本特征对应于声学维度的特征;声学文本特征与文本特征的区别在于,声学文本特征可直接用于合成语音。由于后续合成待播报语音是依据指定用户的第一文本特征和第一语音表现力特征,因此只需训练预设模型准确的学习到播报用户的声纹特征即可,进而本发明实施例采用远小于现有技术中训练数据对预设模型进行训练,就能够实现较为准确的预测出播报用户的声纹特征;从而提高语音合成效率。

[0078] 为了使得所述预设模型输出更符合播报用户的声学特征,本发明实施例中,可以从第二语音数据中,提取所述播报用户针对所述参考文本的参考声学特征、第二语音表现力特征和第二文本特征。其中,提取播报用户的第二语音表现力特征和上述提取指定用户的第一语音表现力特征采用的模型可以是相同的模型。此外,也可以采用现有的其他模型从所述第二语音数据中提取所述播报用户的参考声学特征;以及提取第二文本特征,这与上述步骤206类似,在此不再赘述。然后采用所述播报用户的第二语音表现力特征、参考声学特征和第二文本特征,对所述预设模型进行训练,使得训练后的预设模型输出的第一声学特征趋近于参考声学特征。其中,所述第一声学特征可以是指预设模型对播报用户朗读

参考文本时声学特征的预测结果;所述第一声学特征包括第一声学语音表现力特征、第一声学文本特征和第一声纹特征;进而使得训练后的模型学习到播报用户的声纹特征。

[0079] 其中,所述第一声学语音表现力特征是所述第二语音表现力特征对应于声学维度的特征;所述第一声学文本特征是所述第二文本特征对应于声学维度的特征;所述第一声纹特征可以是指,所述预设模型预测所述播报用户播报参考文本时的声纹特征。其中,所述第一声纹特征包括以下至少一种:播报用户的音色特征、播报用户的口音特征和播报用户的方言特征;当然还可以包括其他可表征播报用户声音的特性的特征,本发明实施例对此不作限制。

[0080] 其中,采用所述播报用户的第一语音表现力特征、参考声学特征和第二文本特征,训练预设模型;可以包括如下子步骤:

[0081] 子步骤62、将所述第二语音表现力特征和第二文本特征输入至预设模型中,得到第一声学特征,其中,所述第一声学特征包括第一声学语音表现力特征、第一声学文本特征和第一声纹特征。

[0082] 子步骤64、将所述第一声学特征与所述参考声学特征进行比对,对所述预设模型的权重进行调整。

[0083] 本发明实施例中,可以将参考文本中每一句文本对应的第二语音表现力特征、第二文本特征和参考声学特征,作为一组训练数据,对预设模型进行训练。以下以采用一组训练数据对训练预设模型为例进行说明:可以将该组训练数据中所述第二语音表现力特征和第二文本特征输入至预设模型中,得到所述预设模型输出的第一声学特征。然后将所述第一声学特征与参考声学特征进行比对,依据对比结果对所述预设模型的权重进行调整。

[0084] 当然,本发明的一个可选实施例中,可以在每次录制自定义的用户朗读参考文本的第二语音数据后,依据所述第二语音数据,提取所述播报用户针对所述参考文本的第二语音表现力特征、参考声学特征和第二文本特征。然后依据所述播报用户的第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型。再建立自定义的用户与训练后的预设模型之间的第五映射关系,并存储各自定义的用户训练后的模型之间的第五映射关系。进而可以基于所述播报用户查找第五映射关系,确定与所述播报用户匹配的第一声纹特征。当然,还可以包括其他获取播报用户对应的第一声纹特征的方式,本发明实施例对此不作限制。

[0085] 其中,本发明实施例不限制步骤206与步骤208的执行顺序,也不限制步骤204-步骤208和步骤210-步骤214的执行顺序。

[0086] 步骤216、将所述第一文本特征和第一语音表现力特征输入至训练后的预设模型中,得到第二声学特征,其中,所述第二声学特征包括第二声学语音表现力特征、第二声学文本特征和第二声纹特征。

[0087] 步骤218、采用所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征进行合成,得到所述待播报文本对应的待播报语音。

[0088] 然后将第一文本特征和第一语音表现力特征,输入至训练后的预设模型中,由所述训练后的预设模型基于所述第一文本特征和第一语音表现力特征,对播报用户播报待播报文本时的声学特征进行预测,得到播报用户对应的第二声学特征。其中,所述第二声学特征可以是指预设模型对播报用户朗读待播报文本时声学特征的预测结果;所述第二声

学特征包括第二声学语音表现力特征、第二声学文本特征和第二声纹特征。所述第二声学语音表现力特征是所述第一语音表现力特征对应于声学维度的特征；所述第二声学文本特征是所述第一文本特征对应于声学维度的特征；所述第二声纹特征可以是指，所述预设模型预测所述播报用户播报待播报文本时的声纹特征。所述第二声纹特征也可以包括以下至少一种：播报用户的音色特征、播报用户的口音特征和播报用户的方言特征；当然还可以包括其他可表征播报用户声音的特性的特征，本发明实施例对此不作限制

[0089] 再采用所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征进行合成，得到所述待播报文本对应的待播报语音。其中，所述待播报文本中，每一句话均存在对应的所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征；然后依次采用每一句话对应的所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征进行合成，得到待播放文本中每一句话对应的待播报语音；最终合成所述待播报文本对应的待播报语音。

[0090] 本发明的一个可选实施例中，终端还存储了预制用户针对每段可播报文本的第三语音数据；当所述播报用户为预制用户时，依据所述播报用户和待播报文本，可以查找匹配的第三语音数据。其中，所述第三语音数据即为待播报数据。

[0091] 本发明的一个可选实施例中，若所述第一声纹特征为音色特征，则在确定待播报文本和播报用户后，可以直接获取指定用户针对所述待播报文本的第一语音数据，然后播报用户的音色特征对所述第一语音数据进行变声处理，得到所述待播报文本对应的待播报语音。

[0092] 步骤220、对所述待播报语音进行播报。

[0093] 综上，本发明实施例中，在确定待播报文本和播报所述待播报文本的播报用户后，可以获取依据指定用户针对所述待播报文本的第一语音表现力特征和第一文本特征，以及所述播报用户的第一声纹特征；然后通过韵律迁移，将指定用户朗读所述待播报文本时的第一语音表现力特征和第一文本特征，作为播报用户播报所述待播报文本时的语音表现力特征和文本特征，再依据所述第一文本特征、第一语音表现力特征和第一声纹特征，合成待播报语音使得合成的待播报语音即具有语音表现力又具有播报用户声纹特征；相对于训练模型提取播报用户的声学特征（如包括如语音表现力特征、声纹特征、文本特征等）而言，用于确定播报用户的声纹特征的数据更少，从而本发明实施例能够提高语音合成效率。此外，本发明实施例中，所述指定用户可以是指具有高语音表现力的用户，进而使得合成的待播报语音具有更好语音表现力，能够提高用户的体验。

[0094] 其次，本发明实施例在合成待播报语音时，采用了待播报文本的第一文本特征，能够减少待播报文本读错的概率，增加了待播报语音的稳定性。

[0095] 再次，本发明实施例中，可以采用所述播报用户的第二语音表现力特征、参考声学特征和第二文本特征训练预设模型，使得所述训练后的预设模型具有所述播报用户的第一声纹特征；后续可以将所述第一文本特征和第一语音表现力特征输入至训练后的预设模型中，得到第二声学特征；所述第二声学特征包括第二声学语音表现力特征、第二声学文本特征和第二声纹特征；进而再采用所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征，合成所述待播报文本对应的待播报语音；能够得到待播报语音与播报用户实际发出的语音更相似，从而提高了终端用户的体验。

[0096] 此外,本发明实施例中,所述第一声纹特征包括以下至少一种:播报用户的音色特征、播报用户的口音特征和播报用户的方言特征;进而本发明实施例可以合成的待播报语音中不仅具有语音表现力,还具有播报用户音色、口音和方言至少一种声音特性;使得合成的待播报语音与播报用户实际发出的语音的音色、口音更相似,还可以生成播报用户对应方言的待播报语音,进一步提高了终端用户的体验。

[0097] 本发明的一个可选实施例中,当所述待播报文本包括多个角色的对话文本时,可以为每一个角色设置一个对应的播放用户;然后可以分角色合成所述待播报文本对应的待播报语音,进而可以提高待播报语音的趣味性,增强终端用户的用户体验。此时,所述播报用户对应可以包括多个,所述依据所述第一文本特征、第一表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,可以包括:依据各角色对应待播报文本的第一文本特征、第一表现力特征和对应播报用户的第一声纹特征,合成各角色对应待播报文本的待播报语音;采用所述各角色对应待播报文本的待播报语音,生成所述待播报文本对应的待播报语音。其中,可以针对所述待播报文本中每个角色对应的待播报文本,依据该角色对应待播报文本的第一文本特征和第一表现力特征,以及该角色对应播报用户的第一声纹特征,合成该角色对应待播报文本的待播报语音。其中,各角色对应待播报文本可以包括多段,可以确定所述待播报文本中各角色对应的各段待播报文本的播报顺序。再按照所述播报顺序,将各角色对应各段待播报文本的待播报语音数据进行拼接,合成所述待播报文本对应的待播报语音。其中,当所述待播报文本中包括旁白和对话时,所述旁白对应的播报用户和对话中其中一个角色对应的播报用户可以是同一个用户,也可以和对话中各角色对应的播报用户是不同的用户,具体可以按照需求设置,本发明实施例对此不作限制。

[0098] 作为本发明的一个示例,待播报文本为《超级奶爸》的故事文本,其中包括旁白,以及海马爸爸、海马妈妈和鳄鱼三个角色,若终端用户为宝宝,则可以将宝宝的爸爸确定为海马爸爸对应的播报用户,将宝宝的妈妈确定为海马妈妈对应的播报用户,将宝宝的爷爷确定为鳄鱼对应的播报用户,将宝宝的奶奶确定为旁白对应的播报用户。然后根据海马爸爸对应待播报文本的第一文本特征和第一语音表现力特征,以及宝宝的爸爸对应的第一声纹特征,合成海马爸爸对应待播报文本的待播报语音。根据海马妈妈对应待播报文本的第一文本特征和第一语音表现力特征,以及宝宝的妈妈对应的第一声纹特征,合成海马妈妈对应待播报文本的待播报语音。根据鳄鱼对应待播报文本的第一文本特征和第一语音表现力特征,以及宝宝的爷爷对应的第一声纹特征,合成鳄鱼对应待播报文本的待播报语音。根据旁白的第一文本特征和第一语音表现力特征,以及宝宝的奶奶对应的第一声纹特征,合成旁白对应的待播报语音。再将海马爸爸对应待播报文本的待播报语音、海马妈妈对应待播报文本的待播报语音、鳄鱼对应待播报文本的待播报语音和旁白对应的待播报语音按照对应的播报顺序进行拼接,得到故事《超级奶爸》对应的待播报语音。

[0099] 需要说明的是,对于方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本发明实施例并不受所描述的动作顺序的限制,因为依据本发明实施例,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作并不一定是本发明实施例所必须的。

[0100] 参照图3,示出了本发明的一种数据处理装置实施例的结构框图,具体可以包括如

下模块：

[0101] 确定模块302,用于确定待播报文本和播报所述待播报文本的播报用户；

[0102] 获取模块304,用于获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,以及所述播报用户的第一声纹特征；

[0103] 合成模块306,用于依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音

[0104] 参照图4,示出了本发明的一种数据处理装置可选实施例的结构框图。

[0105] 本发明一个可选的实施例中,所述获取模块304,包括：

[0106] 第一语音获取子模块3042,用于获取所述指定用户针对所述待播报文本的第一语音数据；

[0107] 第一特征提取子模块3044,用于依据所述第一语音数据,提取所述指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征。

[0108] 本发明一个可选的实施例中,所述第一特征提取子模块3044,用于依据所述第一语音数据分别对所述待播报文本中各文字进行注音,得到所述待播报文本对应的注音信息序列；以及依据所述第一语音数据,依次确定所述待播报文本中相邻两个文字之间的语音停顿信息；依据所述注音信息序列和所述相邻两个文字之间的语音停顿信息,确定所述待播报文本对应的第一文本特征。

[0109] 本发明一个可选的实施例中,所述获取模块304,包括：

[0110] 第二语音获取子模块3046,用于获取所述播报用户针对参考文本的第二语音数据；

[0111] 第二特征提取子模块3048,用于依据所述第二语音数据,提取所述播报用户针对所述参考文本的第二语音表现力特征、参考声学特征和第二文本特征；

[0112] 模型训练子模块30410,用于依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,以使所述训练后的预设模型具有所述播报用户的第一声纹特征。

[0113] 本发明一个可选的实施例中,所述参考声学特征包括：参考声学语音表现力特征、参考声学文本特征和参考声纹特征；所述模型训练子模块30410,用于将所述第二语音表现力特征和第二文本特征输入至预设模型中,得到第一声学特征,其中,所述第一声学特征包括第一声学语音表现力特征、第一声学文本特征和第一声纹特征；将所述第一声学特征与所述参考声学特征进行比对,对所述预设模型的权重进行调整。

[0114] 本发明一个可选的实施例中,所述合成模块306,包括：

[0115] 第一语音合成子模块3062,用于将所述第一文本特征和第一语音表现力特征输入至训练后的预设模型中,得到第二声学特征,其中,所述第二声学特征包括第二声学语音表现力特征、第二声学文本特征和第二声纹特征；采用所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征进行合成,得到所述待播报文本对应的待播报语音。

[0116] 本发明一个可选的实施例中,所述播报用户包括多个,一个播报用户对应该待播报文本中的一个角色；所述合成模块306,包括：

[0117] 第二语音合成子模块3064,用于依据各角色对应待播报文本的第一文本特征、第一表现力特征和对应播报用户的第一声纹特征,合成各角色对应待播报文本的待播报语音；采用所述各角色对应待播报文本的待播报语音,生成所述待播报文本对应的待播报语

音。

[0118] 本发明一个可选的实施例中,所述第一声纹特征包括以下至少一种:播报用户的音色特征、播报用户的口音特征和播报用户的方言特征。

[0119] 本发明一个可选的实施例中,所述的装置还包括:

[0120] 播报模块308,用于对所述待播报语音进行播报。

[0121] 综上,本发明实施例中,在确定待播报文本和播报所述待播报文本的播报用户后,可以获取依据指定用户针对所述待播报文本的第一语音表现力特征和第一文本特征,以及所述播报用户的第一声纹特征;然后通过韵律迁移,将指定用户朗读所述待播报文本时的第一语音表现力特征和第一文本特征,作为播报用户播报所述待播报文本时的语音表现力特征和文本特征,再依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成待播报语音使得合成的待播报语音即具有语音表现力又具有播报用户声纹特征;相对于训练模型提取播报用户的声学特征(如包括如语音表现力特征、声纹特征、文本特征等)而言,用于确定播报用户的声纹特征的数据更少,从而本发明实施例能够提高语音合成效率。

[0122] 对于装置实施例而言,由于其与方法实施例基本相似,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0123] 图5是根据一示例性实施例示出的一种用于数据处理的电子设备500的结构框图。例如,电子设备500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。

[0124] 参照图5,电子设备500可以包括以下一个或多个组件:处理组件502,存储器504,电力组件506,多媒体组件508,音频组件510,输入/输出(I/O)的接口512,传感器组件514,以及通信组件516。

[0125] 处理组件502通常控制电子设备500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。处理元件502可以包括一个或多个处理器520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502可以包括一个或多个模块,便于处理组件502和其他组件之间的交互。例如,处理部件502可以包括多媒体模块,以方便多媒体组件508和处理组件502之间的交互。

[0126] 存储器504被配置为存储各种类型的数据以支持在设备500的操作。这些数据的示例包括用于在电子设备500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504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。

[0127] 电力组件506为电子设备500的各种组件提供电力。电力组件506可以包括电源管理系统,一个或多个电源,及其他与为电子设备500生成、管理和分配电力相关联的组件。

[0128] 多媒体组件508包括在所述电子设备500和用户之间的提供一个输出接口的屏幕。在一些实施例中,屏幕可以包括液晶显示器(LCD)和触摸面板(TP)。如果屏幕包括触摸面板,屏幕可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中,

多媒体组件508包括一个前置摄像头和/或后置摄像头。当电子设备500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。

[0129] 音频组件510被配置为输出和/或输入音频信号。例如,音频组件510包括一个麦克风(MIC),当电子设备500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504或经由通信组件516发送。在一些实施例中,音频组件510还包括一个扬声器,用于输出音频信号。

[0130] I/O接口512为处理组件502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。

[0131] 传感器组件514包括一个或多个传感器,用于为电子设备500提供各个方面的状态评估。例如,传感器组件514可以检测到设备500的打开/关闭状态,组件的相对定位,例如所述组件为电子设备500的显示器和小键盘,传感器组件514还可以检测电子设备500或电子设备500一个组件的位置改变,用户与电子设备500接触的存在或不存在,电子设备500方位或加速/减速和电子设备500的温度变化。传感器组件514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514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514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。

[0132] 通信组件516被配置为便于电子设备500和其他设备之间有线或无线方式的通信。电子设备500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,通信部件514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,所述通信部件514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。

[0133] 在示例性实施例中,电子设备500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。

[0134] 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504,上述指令可由电子设备500的处理器520执行以完成上述方法。例如,所述非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0135] 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电子设备的处理器执行时,使得电子设备能够执行一种数据处理方法,所述方法包括:确定待播报文本和播报所述待播报文本的播报用户;获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,以及所述播报用户的第一声纹特征;依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音。

[0136] 可选地,所述获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,包括:获取所述指定用户针对所述待播报文本的第一语音数据;依据所述第一语音数据,提取所述指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征。

[0137] 可选地,所述依据所述第一语音数据中,提取指定用户针对所述待播报文本所述第一文本特征,包括:依据所述第一语音数据分别对所述待播报文本中各文字进行注音,得到所述待播报文本对应的注音信息序列;以及依据所述第一语音数据,依次确定所述待播报文本中相邻两个文字之间的语音停顿信息;依据所述注音信息序列和所述相邻两个文字之间的语音停顿信息,确定所述待播报文本对应的第一文本特征。

[0138] 可选地,所述获取所述播报用户的第一声纹特征,包括:获取所述播报用户针对参考文本的第二语音数据;依据所述第二语音数据,提取所述播报用户针对所述参考文本的第二语音表现力特征、参考声学特征和第二文本特征;依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,以使所述训练后的预设模型具有所述播报用户的第一声纹特征。

[0139] 可选地,所述参考声学特征包括:参考声学语音表现力特征、参考声学文本特征和参考声纹特征;所述依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,包括:将所述第二语音表现力特征和第二文本特征输入至预设模型中,得到第一声学特征,其中,所述第一声学特征包括第一声学语音表现力特征、第一声学文本特征和第一声纹特征;将所述第一声学特征与所述参考声学特征进行比对,对所述预设模型的权重进行调整。

[0140] 可选地,所述依据所述第一文本特征、第一表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,包括:将所述第一文本特征和第一语音表现力特征输入至训练后的预设模型中,得到第二声学特征,其中,所述第二声学特征包括第二声学语音表现力特征、第二声学文本特征和第二声纹特征;采用所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征进行合成,得到所述待播报文本对应的待播报语音。

[0141] 可选地,所述播报用户包括多个,一个播报用户对应待播报文本中的一个角色;所述依据所述第一文本特征、第一表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,包括:依据各角色对应待播报文本的第一文本特征、第一表现力特征和对应播报用户的第一声纹特征,合成各角色对应待播报文本的待播报语音;采用所述各角色对应待播报文本的待播报语音,生成所述待播报文本对应的待播报语音。

[0142] 可选地,所述第一声纹特征包括以下至少一种:播报用户的音色特征、播报用户的口音特征和播报用户的方言特征。

[0143] 可选地,所述的方法还包括:对所述待播报语音进行播报。

[0144] 图6是本发明根据另一示例性实施例示出的一种用于数据处理的电子设备600的结构示意图。该电子设备600可以是服务器,该服务器可因配置或性能不同而产生比较大的差异,可以包括一个或一个以上中央处理器(central processing units,CPU)622(例如,一个或一个以上处理器)和存储器632,一个或一个以上存储应用程序642或数据644的存储介质630(例如一个或一个以上海量存储设备)。其中,存储器632和存储介质630可以是短暂存储或持久存储。存储在存储介质630的程序可以包括一个或一个以上模块(图示没标出),每个模块可以包括对服务器中的一系列指令操作。更进一步地,中央处理器622可以设置为与存储介质630通信,在服务器上执行存储介质630中的一系列指令操作。

[0145] 服务器还可以包括一个或一个以上电源626,一个或一个以上有线或无线网络接口650,一个或一个以上输入输出接口658,一个或一个以上键盘656,和/或,一个或一个以

上操作系统641,例如Windows Server™,Mac OS X™,Unix™,Linux™,FreeBSD™等等。

[0146] 一种电子设备,包括有存储器,以及一个或者一个以上的程序,其中一个或者一个以上程序存储于存储器中,且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行以下操作的指令:确定待播报文本和播报所述待播报文本的播报用户;获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,以及所述播报用户的第一声纹特征;依据所述第一文本特征、第一语音表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音。

[0147] 可选地,所述获取指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征,包括:获取所述指定用户针对所述待播报文本的第一语音数据;依据所述第一语音数据,提取所述指定用户针对所述待播报文本的第一文本特征和第一语音表现力特征。

[0148] 可选地,所述依据所述第一语音数据中,提取指定用户针对所述待播报文本所述第一文本特征,包括:依据所述第一语音数据分别对所述待播报文本中各文字进行注音,得到所述待播报文本对应的注音信息序列;以及依据所述第一语音数据,依次确定所述待播报文本中相邻两个文字之间的语音停顿信息;依据所述注音信息序列和所述相邻两个文字之间的语音停顿信息,确定所述待播报文本对应的第一文本特征。

[0149] 可选地,所述获取所述播报用户的第一声纹特征,包括:获取所述播报用户针对参考文本的第二语音数据;依据所述第二语音数据,提取所述播报用户针对所述参考文本的第二语音表现力特征、参考声学特征和第二文本特征;依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,以使所述训练后的预设模型具有所述播报用户的第一声纹特征。

[0150] 可选地,所述参考声学特征包括:参考声学语音表现力特征、参考声学文本特征和参考声纹特征;所述依据所述第二语音表现力特征、参考声学特征和第二文本特征,训练预设模型,包括:将所述第二语音表现力特征和第二文本特征输入至预设模型中,得到第一声学特征,其中,所述第一声学特征包括第一声学语音表现力特征、第一声学文本特征和第一声纹特征;将所述第一声学特征与所述参考声学特征进行比对,对所述预设模型的权重进行调整。

[0151] 可选地,所述依据所述第一文本特征、第一表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,包括:将所述第一文本特征和第一语音表现力特征输入至训练后的预设模型中,得到第二声学特征,其中,所述第二声学特征包括第二声学语音表现力特征、第二声学文本特征和第二声纹特征;采用所述第二声学语音表现力特征、第二声学文本特征和第二声纹特征进行合成,得到所述待播报文本对应的待播报语音。

[0152] 可选地,所述播报用户包括多个,一个播报用户对应该待播报文本中的一个角色;所述依据所述第一文本特征、第一表现力特征和第一声纹特征,合成所述待播报文本对应的待播报语音,包括:依据各角色对应该待播报文本的第一文本特征、第一表现力特征和对应该播报用户的第一声纹特征,合成各角色对应该待播报文本的待播报语音;采用所述各角色对应该待播报文本的待播报语音,生成所述待播报文本对应的待播报语音。

[0153] 可选地,其特征在于,所述第一声纹特征包括以下至少一种:播报用户的音色特征、播报用户的口音特征和播报用户的方言特征。

[0154] 可选地,还包含用于进行以下操作的指令:对所述待播报语音进行播报。

[0155] 本说明书中的各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似的部分互相参见即可。

[0156] 本发明实施例是参照根据本发明实施例的方法、终端设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理终端设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理终端设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0157] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理终端设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0158] 这些计算机程序指令也可装载到计算机或其他可编程数据处理终端设备上,使得在计算机或其他可编程终端设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程终端设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0159] 尽管已描述了本发明实施例的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例做出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明实施例范围的所有变更和修改。

[0160] 最后,还需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者终端设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者终端设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者终端设备中还存在另外的相同要素。

[0161] 以上对本发明所提供的一种数据处理方法、一种数据处理装置和一种电子设备,进行了详细介绍,本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本发明的限制。

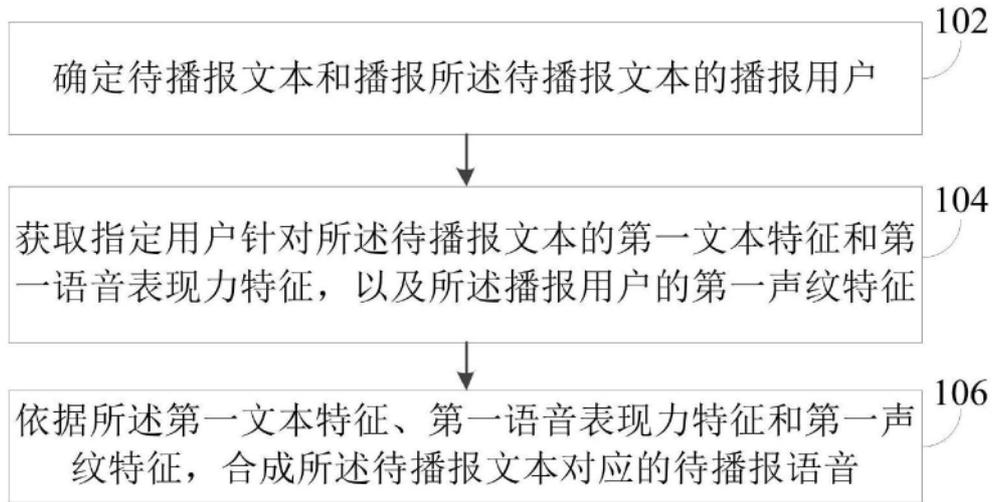


图1

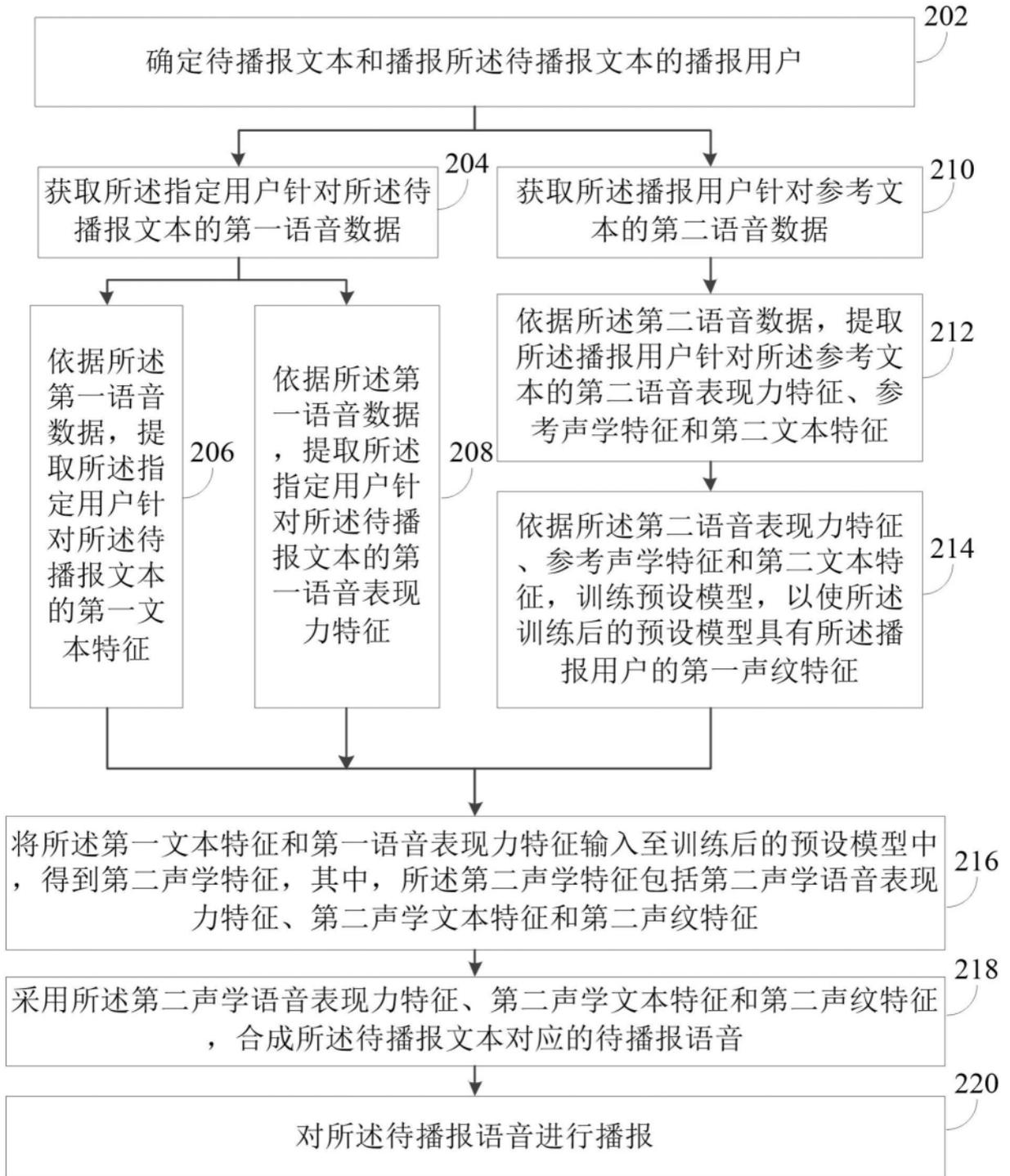


图2



图3



图4

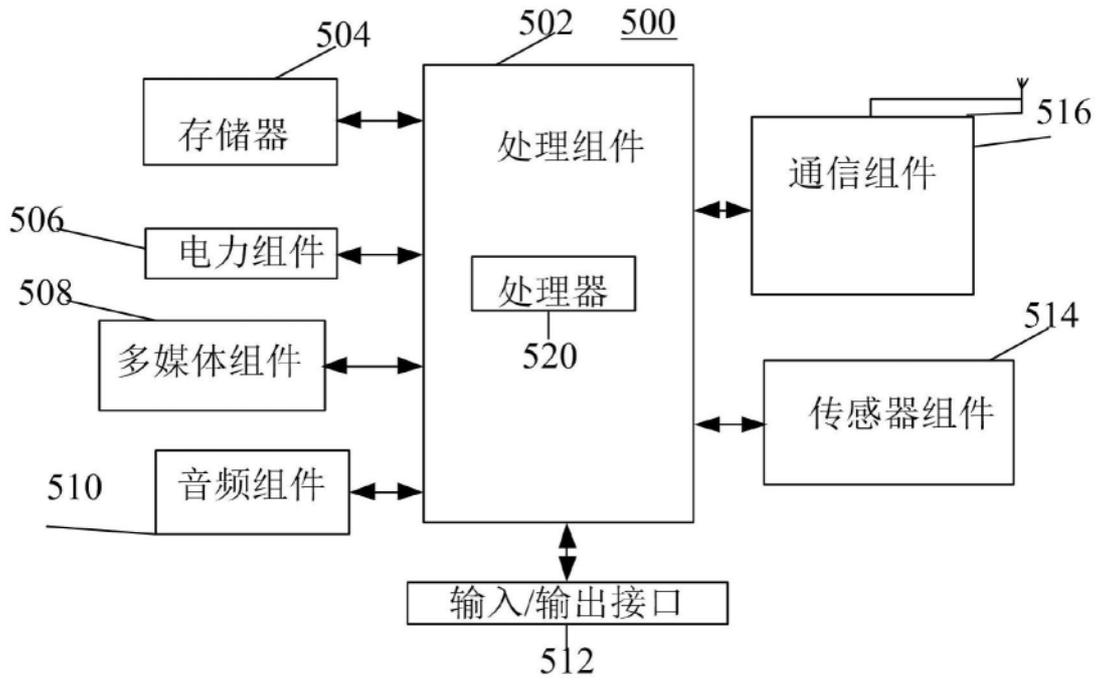


图5

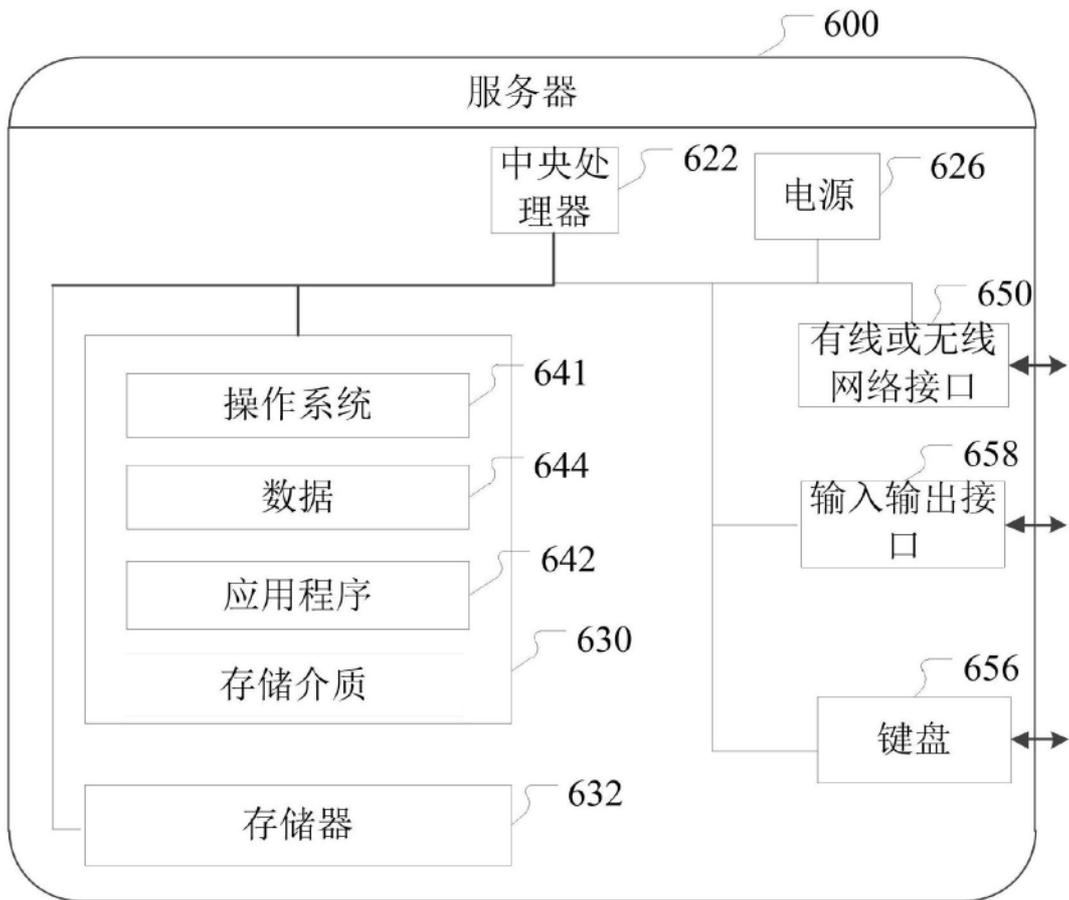


图6