



(12) 发明专利申请

(10) 申请公布号 CN 113555026 A

(43) 申请公布日 2021. 10. 26

(21) 申请号 202110835128.X

(22) 申请日 2021.07.23

(71) 申请人 平安科技(深圳)有限公司

地址 518000 广东省深圳市福田区福田街  
道福安社区益田路5033号平安金融中  
心23楼

(72) 发明人 孙奥兰 王健宗

(74) 专利代理机构 深圳市沃德知识产权代理事  
务所(普通合伙) 44347

代理人 高杰 于志光

(51) Int. Cl.

G10L 21/007 (2013.01)

G10L 25/51 (2013.01)

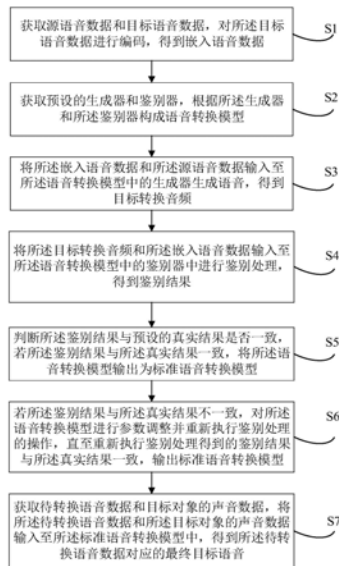
权利要求书3页 说明书16页 附图2页

(54) 发明名称

语音转换方法、装置、电子设备及介质

(57) 摘要

本发明涉及语音语义技术,揭露了一种语音转换方法,包括:对目标语音数据进行编码,得到嵌入语音数据,将嵌入语音数据和源语音数据输入至语音转换模型中的生成器生成语音,得到目标转换音频,将目标转换音频和嵌入语音数据输入至语音转换模型中的鉴别器中进行鉴别,得到鉴别结果,判断鉴别结果与真实结果是否一致,根据判断结果输出标准语音转换模型,将待转换语音数据和目标对象的声音数据输入至标准语音转换模型中,得到待转换语音数据对应的最终目标语音。此外,本发明还涉及区块链技术,所述鉴别结果可存储于区块链的节点。本发明还提出一种语音转换装置、电子设备以及计算机可读存储介质。本发明可以解决语音转换的效率较低的问题。



1. 一种语音转换方法,其特征在于,所述方法包括:

获取源语音数据和目标语音数据,对所述目标语音数据进行编码,得到嵌入语音数据;

获取预设的生成器和鉴别器,根据所述生成器和所述鉴别器构成语音转换模型;

将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生成语音,得到目标转换音频;

将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,得到鉴别结果;

判断所述鉴别结果与预设的真实结果是否一致,若所述鉴别结果与所述真实结果一致,将所述语音转换模型输出为标准语音转换模型;

若所述鉴别结果与所述真实结果不一致,对所述语音转换模型进行参数调整并重新执行鉴别处理的操作,直至重新执行鉴别处理得到的鉴别结果与所述真实结果一致,输出标准语音转换模型;

获取待转换语音数据和目标对象的声音数据,将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中,得到所述待转换语音数据对应的最终目标语音。

2. 如权利要求1所述的语音转换方法,其特征在于,所述将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生成语音,得到目标转换音频,包括:

对所述嵌入语音数据进行第一特征提取,得到第一特征数据集,对所述源语音数据进行第二特征提取,得到第二特征数据集,将所述第一特征数据集和所述第二特征数据集汇总得到总特征数据集;

利用所述生成器中的下采样层对所述总特征数据集进行下采样处理,得到下采样数据集;

将所述下采样数据集输入至所述生成器中的瓶颈层,并对经过所述瓶颈层处理的数据进行上采样处理,得到上采样数据集;

将所述上采样数据集输入至所述生成器中的动态图网络中进行转换,得到目标转换音频。

3. 如权利要求2所述的语音转换方法,其特征在于,所述对所述嵌入语音数据进行第一特征提取,得到第一特征数据集,包括:

对所述嵌入语音数据进行预加重处理、分帧处理和加窗处理以及快速傅里叶变换,得到所述嵌入语音数据的短时频谱;

将所述短时频谱输入预设的梅尔尺度滤波组中,得到梅尔频谱;

对所述梅尔频谱进行能量计算,得到对数能量;

对所述对数能量进行离散余弦变换,得到第一特征数据集。

4. 如权利要求3所述的语音转换方法,其特征在于,所述对所述对数能量进行离散余弦变换,得到第一特征数据集,包括:

利用下述公式对所述对数能量进行离散余弦变换,得到第一特征数据集:

$$C(n) = \sum_{M=0}^{N-1} T(m) \cos\left(\frac{\pi n(M-0.5)}{M}\right)$$

其中,  $C(n)$  指第一特征数据集,  $T(m)$  为对数能量,  $M$  为梅尔尺度滤波组中的滤波器个数,  $n$  为帧的个数。

5. 如权利要求1所述的语音转换方法, 其特征在于, 所述将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理, 得到鉴别结果, 包括:

分别利用所述鉴别器中的第一鉴别网络、第二鉴别网络和第三鉴别网络对所述目标转换音频和所述嵌入语音数据进行第一鉴别值、第二鉴别值和第三鉴别值;

对所述第一鉴别值、所述第二鉴别值和所述第三鉴别值进行权重归一化, 得到最终鉴别值;

若所述最终鉴别值大于或者等于预设的鉴别阈值, 得到所述目标转换音频为标准转换音频的鉴别结果;

若所述最终鉴别值小于预设的鉴别阈值, 得到所述目标转换音频为非标准转换音频的鉴别结果。

6. 如权利要求1所述的语音转换方法, 其特征在于, 所述根据所述生成器和所述鉴别器构成语音转换模型, 包括:

分别初始化所述生成器和所述鉴别器的参数;

将所述源语音数据输入至初始化后的生成器中得到生成语音数据, 判断所述生成语音数据与所述目标语音数据是否一致;

若所述生成语音数据与所述目标语音数据不一致, 则对所述生成器中的各个模块进行顺序调整, 并对调整模块顺序后的生成器重新执行语音生成处理;

若所述生成语音数据与所述目标语音数据一致, 根据预设的连接顺序将初始化后的所述生成器和所述鉴别器进行连接, 得到所述语音转换模型。

7. 如权利要求1所述的语音转换方法, 其特征在于, 所述对所述目标语音数据进行编码, 得到嵌入语音数据, 包括:

根据预设的字典获取所述目标语音数据对应的标识号;

对所述标识号和所述目标语音数据进行向量化处理, 得到嵌入语音数据。

8. 一种语音转换装置, 其特征在于, 所述装置包括:

数据编码模块, 用于获取源语音数据和目标语音数据, 对所述目标语音数据进行编码, 得到嵌入语音数据;

模型构建模块, 用于获取预设的生成器和鉴别器, 根据所述生成器和所述鉴别器构成语音转换模型;

模型训练模块, 用于将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生成语音, 得到目标转换音频, 将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理, 得到鉴别结果, 判断所述鉴别结果与预设的真实结果是否一致, 若所述鉴别结果与所述真实结果一致, 将所述语音转换模型输出为标准语音转换模型, 若所述鉴别结果与所述真实结果不一致, 对所述语音转换模型进行参数调整并重新执行鉴别处理的操作, 直至重新执行鉴别处理得到的鉴别结果与所述真实结果一致, 输出标准语音转换模型;

最终目标语音生成模块, 用于获取待转换语音数据和目标对象的声音数据, 将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中, 得到所述待转

换语音数据对应的最终目标语音。

9. 一种电子设备,其特征在于,所述电子设备包括:

至少一个处理器;以及,

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行如权利要求1至7中任意一项所述的语音转换方法。

10. 一种计算机可读存储介质,存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至7中任意一项所述的语音转换方法。

## 语音转换方法、装置、电子设备及介质

### 技术领域

[0001] 本发明涉及语音语义技术领域,尤其涉及一种语音转换方法、装置、电子设备及计算机可读存储介质。

### 背景技术

[0002] 随着多媒体通信技术的不断发展,作为人机通信重要方式之一的语音合成技术以其方便、快捷的优点收到了研究者的广泛关注。语音转换属于语音合成的一般技术领域,同时也是人工智能的重要方面之一,语音转换的研究内容为如何在不改变语言内容的情况下将一个人的声音转换为另一个人的声音。

[0003] 现有的语音转换方法是利用多阶段模型进行转换处理,即将语音转换过程拆分成音色的频谱转换和音频生成两个部分,进行语音转换的效率较低。

### 发明内容

[0004] 本发明提供一种语音转换方法、装置及计算机可读存储介质,其主要目的在于解决语音转换的效率较低的问题。

[0005] 为实现上述目的,本发明提供的一种语音转换方法,包括:

[0006] 获取源语音数据和目标语音数据,对所述目标语音数据进行编码,得到嵌入语音数据;

[0007] 获取预设的生成器和鉴别器,根据所述生成器和所述鉴别器构成语音转换模型;

[0008] 将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生成语音,得到目标转换音频;

[0009] 将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,得到鉴别结果;

[0010] 判断所述鉴别结果与预设的真实结果是否一致,若所述鉴别结果与所述真实结果一致,将所述语音转换模型输出为标准语音转换模型;

[0011] 若所述鉴别结果与所述真实结果不一致,对所述语音转换模型进行参数调整并重新执行鉴别处理的操作,直至重新执行鉴别处理得到的鉴别结果与所述真实结果一致,输出标准语音转换模型;

[0012] 获取待转换语音数据和目标对象的声音数据,将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中,得到所述待转换语音数据对应的最终目标语音。

[0013] 可选地,所述将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生成语音,得到目标转换音频,包括:

[0014] 对所述嵌入语音数据进行第一特征提取,得到第一特征数据集,对所述源语音数据进行第二特征提取,得到第二特征数据集,将所述第一特征数据集和所述第二特征数据集汇总得到总特征数据集;

[0015] 利用所述生成器中的下采样层对所述总特征数据集进行下采样处理,得到下采样数据集;

[0016] 将所述下采样数据集输入至所述生成器中的瓶颈层,并对经过所述瓶颈层处理的数据进行上采样处理,得到上采样数据集;

[0017] 将所述上采样数据集输入至所述生成器中的动态图网络中进行转换,得到目标转换音频。

[0018] 可选地,所述对所述嵌入语音数据进行第一特征提取,得到第一特征数据集,包括:

[0019] 对所述嵌入语音数据进行预加重处理、分帧处理和加窗处理以及快速傅里叶变换,得到所述嵌入语音数据的短时频谱;

[0020] 将所述短时频谱输入预设的梅尔尺度滤波组中,得到梅尔频谱;

[0021] 对所述梅尔频谱进行能量计算,得到对数能量;

[0022] 对所述对数能量进行离散余弦变换,得到第一特征数据集。

[0023] 可选地,所述对所述对数能量进行离散余弦变换,得到第一特征数据集,包括:

[0024] 利用下述公式对所述对数能量进行离散余弦变换,得到第一特征数据集:

$$[0025] \quad C(n) = \sum_{M=0}^{N-1} T(m) \cos\left(\frac{\pi n(M-0.5)}{M}\right)$$

[0026] 其中,C(n)指第一特征数据集,T(m)为对数能量,M为梅尔尺度滤波组中的滤波器个数,n为帧的个数。

[0027] 可选地,所述将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,得到鉴别结果,包括:

[0028] 分别利用所述鉴别器中的第一鉴别网络、第二鉴别网络和第三鉴别网络对所述目标转换音频和所述嵌入语音数据进行第一鉴别值、第二鉴别值和第三鉴别值;

[0029] 对所述第一鉴别值、所述第二鉴别值和所述第三鉴别值进行权重归一化,得到最终鉴别值;

[0030] 若所述最终鉴别值大于或者等于预设的鉴别阈值,得到所述目标转换音频为标准转换音频的鉴别结果;

[0031] 若所述最终鉴别值小于预设的鉴别阈值,得到所述目标转换音频为非标准转换音频的鉴别结果。

[0032] 可选地,所述根据所述生成器和所述鉴别器构成语音转换模型,包括:

[0033] 分别初始化所述生成器和所述鉴别器的参数;

[0034] 将所述源语音数据输入至初始化后的生成器中得到生成语音数据,判断所述生成语音数据与所述目标语音数据是否一致;

[0035] 若所述生成语音数据与所述目标语音数据不一致,则对所述生成器中的各个模块进行顺序调整,并对调整模块顺序后的生成器重新执行语音生成处理;

[0036] 若所述生成语音数据与所述目标语音数据一致,根据预设的连接顺序将初始化后的所述生成器和所述鉴别器进行连接,得到所述语音转换模型。

[0037] 可选地,所述对所述目标语音数据进行编码,得到嵌入语音数据,包括:

[0038] 根据预设的字典获取所述目标语音数据对应的标识号;

[0039] 对所述标识号和所述目标语音数据进行向量化处理,得到嵌入语音数据。

[0040] 为了解决上述问题,本发明还提供一种语音转换装置,所述装置包括:

[0041] 数据编码模块,用于获取源语音数据和目标语音数据,对所述目标语音数据进行编码,得到嵌入语音数据;

[0042] 模型构建模块,用于获取预设的生成器和鉴别器,根据所述生成器和所述鉴别器构成语音转换模型;

[0043] 模型训练模块,用于将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生成语音,得到目标转换音频,将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,得到鉴别结果,判断所述鉴别结果与预设的真实结果是否一致,若所述鉴别结果与所述真实结果一致,将所述语音转换模型输出为标准语音转换模型,若所述鉴别结果与所述真实结果不一致,对所述语音转换模型进行参数调整并重新执行鉴别处理的操作,直至重新执行鉴别处理得到的鉴别结果与所述真实结果一致,输出标准语音转换模型;

[0044] 最终目标语音生成模块,用于获取待转换语音数据和目标对象的声音数据,将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中,得到所述待转换语音数据对应的最终目标语音。

[0045] 为了解决上述问题,本发明还提供一种电子设备,所述电子设备包括:

[0046] 存储器,存储至少一个指令;及

[0047] 处理器,执行所述存储器中存储的指令以实现上述所述的语音转换方法。

[0048] 为了解决上述问题,本发明还提供一种计算机可读存储介质,所述计算机可读存储介质中存储有至少一个指令,所述至少一个指令被电子设备中的处理器执行以实现上述所述的语音转换方法。

[0049] 本发明实施例中,通过对目标语音数据进行编码,得到嵌入语音数据,所述嵌入语音数据包含了标识了身份信息的标识号特征和目标语音数据特征,故编码处理可以使得所述嵌入语音数据所包含的信息更加全面和丰富,根据生成器和鉴别器构成语音转换模型,所述生成器用于生成数据样本,所述鉴别器用于鉴别所述数据样本的真实性,进而去调整生成器的参数,因此所述生成器和所述鉴别器可以达到一种博弈平衡状态,进而保证所述语音转换模型输出数据的准确度,将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器进行生成转换处理,保证生成器生成的目标转换音频更具有真实性,将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,所述鉴别器可以学习音频不同频率范围的特征,判断所述鉴别结果与预设的真实结果是否一致进而输出标准语音转换模型,保证语音转换模型输出的准确性。所述语音转换模型可以将所述生成器和所述鉴别器集成在一起,并将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中,得到所述待转换语音数据对应的最终目标语音。因此本发明提出的语音转换方法、装置、电子设备及计算机可读存储介质,可以解决语音转换的效率较低的问题。

## 附图说明

[0050] 图1为本发明一实施例提供的语音转换方法的流程示意图;

- [0051] 图2为本发明一实施例提供的语音转换装置的功能模块图；
- [0052] 图3为本发明一实施例提供的实现所述语音转换方法的电子设备的结构示意图。
- [0053] 本发明目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

### 具体实施方式

- [0054] 应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。
- [0055] 本申请实施例提供一种语音转换方法。所述语音转换方法的执行主体包括但不限于服务端、终端等能够被配置为执行本申请实施例提供的该方法的电子设备中的至少一种。换言之，所述语音转换方法可以由安装在终端设备或服务端设备的软件或硬件来执行，所述软件可以是区块链平台。所述服务端包括但不限于：单台服务器、服务器集群、云端服务器或云端服务器集群等。
- [0056] 参照图1所示，为本发明一实施例提供的语音转换方法的流程示意图。
- [0057] 在本实施例中，所述语音转换方法包括：
- [0058] S1、获取源语音数据和目标语音数据，对所述目标语音数据进行编码，得到嵌入语音数据。
- [0059] 本发明实施例中，所述源语音数据为进行语音转换之前的音频数据，所述目标语音数据为语音转换的目标音频数据。例如，所述语音转换的目标是在不改变语言内容的情况下，调整音色，将A音频转换为B音频，则所述A音频为所述源语音数据，所述B音频为所述目标语音数据。
- [0060] 具体地，所述对所述目标语音数据进行编码，得到嵌入语音数据，包括：
- [0061] 根据预设的字典获取所述目标语音数据对应的标识号；
- [0062] 对所述标识号和所述目标语音数据进行向量化处理，得到嵌入语音数据。
- [0063] 详细地，所述字典中包含音频数据和标识号的一一对应关系，其中，所述音频数据为不同目标人群对应的语音数据，所述标识号是指目标人群的身份标识号，可以根据所述字典查找到所述音频数据对应的标识号，将所述标识号和所述目标语音数据输入至预获取的编码器中进行向量化处理，得到嵌入语音数据。
- [0064] 例如，所述字典中包含{源语音数据:1, 目标语音数据:2}，即所述源语音数据对应的标识号为1，所述目标语音数据对应的标识号为2，根据所述字典可以获取所述目标语音数据对应的标识号2，将所述标识号2和所述目标语音数据一起输入至所述编码器中，得到嵌入语音数据。
- [0065] 其中，通过对所述目标语音数据进行编码，可以包含标识了身份信息标识号特征和目标语音数据特征的嵌入语音数据，丰富了所述嵌入语音数据所包含的身份信息和语音信息。
- [0066] S2、获取预设的生成器和鉴别器，根据所述生成器和所述鉴别器构成语音转换模型。
- [0067] 本发明实施例中，所述生成器用于转换后的音频数据，所述鉴别器用于鉴别输入音频是真实音频或者虚假音频，其中，本方案中采用的所述生成器为StarGAN-VC2生成器，采用的所述鉴别器为MelGAN鉴别器。
- [0068] 具体地，所述根据所述生成器和所述鉴别器构成语音转换模型，包括：



- [0069] 分别初始化所述生成器和所述鉴别器的参数；
- [0070] 将所述源语音数据输入至初始化后的生成器中得到生成语音数据，判断所述生成语音数据与所述目标语音数据是否一致；
- [0071] 若所述生成语音数据与所述目标语音数据不一致，则对所述生成器中的各个模块进行顺序调整，并对调整模块顺序后的生成器重新执行语音生成处理；
- [0072] 若所述生成语音数据与所述目标语音数据一致，根据预设的连接顺序将初始化后的所述生成器和所述鉴别器进行连接，得到所述语音转换模型。
- [0073] 进一步地，所述预获取的生成器中包含多个模块，例如下采样层和上采样层等，预获取的生成器中的多个模块有固定的连接顺序，将所述源语音数据输入至初始化后的生成器中得到生成语音数据后，需要判断所述生成语音数据与所述目标语音数据是否一致，当所述生成语音数据与所述目标语音数据不一致时，可以对所述生成器中的多个模块进行顺序的调整，并将所述源语音数据重新输入至调整模块顺序后的生成器中，直至输出的语音数据与目标语音数据一致，将所述生成器与所述初始化后的鉴别器进行连接，得到所述语音转换模型。
- [0074] 其中，在本发明实施例中，以所述生成器在先，所述鉴别器在后的连接顺序构建得到所述语音转换模型。
- [0075] 详细地，利用所述生成器和所述鉴别器构成语音转换模型可以产生更好的输出样本，所述生成器用于生成转换后的数据，所述鉴别器用于判别数据的真假，所述生成器和所述鉴别器可以达到一种博弈平衡状态，进而保证所述语音转换模型输出数据的准确度。
- [0076] S3、将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生成语音，得到目标转换音频。
- [0077] 本发明实施例中，所述语音转换模型中的生成器为StarGAN-VC2生成器。
- [0078] 具体地，所述将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生成语音，得到目标转换音频，包括：
- [0079] 对所述嵌入语音数据进行第一特征提取，得到第一特征数据集，对所述源语音数据进行第二特征提取，得到第二特征数据集，将所述第一特征数据集和所述第二特征数据集汇总得到总特征数据集；
- [0080] 利用所述生成器中的下采样层对所述总特征数据集进行下采样处理，得到下采样数据集；
- [0081] 将所述下采样数据集输入至所述生成器中的瓶颈层，并对经过所述瓶颈层处理的数据进行上采样处理，得到上采样数据集；
- [0082] 将所述上采样数据集输入至所述生成器中的动态图网络中进行转换，得到目标转换音频。
- [0083] 详细地，所述生成器包含下采样层、瓶颈层和上采样层和动态图网络。
- [0084] 其中，所述生成器中的动态图网络结构可以对输入的上采样数据集进行矩阵运算，进而得到目标转换音频。
- [0085] 进一步地，所述对所述嵌入语音数据进行第一特征提取，得到第一特征数据集，包括：
- [0086] 对所述嵌入语音数据进行预加重处理、分帧处理和加窗处理以及快速傅里叶变

换,得到所述嵌入语音数据的短时频谱;

[0087] 将所述短时频谱输入预设的梅尔尺度滤波组中,得到梅尔频谱;

[0088] 对所述梅尔频谱进行能量计算,得到对数能量;

[0089] 对所述对数能量进行离散余弦变换,得到第一特征数据集。

[0090] 其中,对所述源语音数据进行第二特征提取的方法与对所述嵌入语音数据进行第一特征提取的方法一致,此处不再赘述。

[0091] 详细地,通过一个预设的高通滤波器对所述嵌入语音数据进行预加重处理,其中,所述预加重处理可以增强所述嵌入语音数据中语音信号的高频部分。利用预设的采样点将经过预加重处理的嵌入语音数据切分成多帧的数据,得到分帧数据集。

[0092] 在本申请一可选实施例中,所述加窗处理是根据预设的窗口函数对所述分帧数据集中的每一帧进行加窗处理,得到加窗信号。

[0093] 详细地,所述预设的窗口函数为:

[0094]  $S'(n) = S(n) \times W(n)$

[0095]  $W(n) = 0.5[1 - \cos\left(\frac{2\pi n}{N+1}\right)] \quad 1 \ll n \ll N$

[0096] 其中, $S'(n)$ 为加窗信号, $S(n)$ 为分帧数据集, $W(n)$ 为窗口函数, $N$ 为帧的大小, $n$ 为帧的个数。

[0097] 优选地,本申请实施例中,所述预设的窗口函数可选择三角窗,所述 $W(n)$ 为三角窗的函数表达式。

[0098] 本申请实施例对所述分帧数据集进行加窗处理可以增加帧左端和帧右端的连续性,减少频谱泄漏。

[0099] 在本申请一可选实施例中,所述对所述对数能量进行离散余弦变换,得到第一特征数据集,包括:

[0100] 利用下述公式对所述对数能量进行离散余弦变换,得到第一特征数据集:

[0101]  $C(n) = \sum_{M=0}^{N-1} T(m) \cos\left(\frac{\pi n(M-0.5)}{M}\right)$

[0102] 其中, $C(n)$ 指第一特征数据集, $T(m)$ 为对数能量, $M$ 为梅尔尺度滤波组中的滤波器个数, $n$ 为帧的个数。

[0103] 为了得到合适大小的声音特征,将所述短时频谱输入预设的梅尔尺度滤波组,变换为梅尔频谱。梅尔频谱可以让人耳对频率的感知度变为线性。对所述梅尔频谱进行倒谱分析,得到特征数据集,其中,所述倒谱分析包括对所述梅尔频谱取对数在做能量转换。

[0104] S4、将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,得到鉴别结果。

[0105] 本发明实施例中,所述鉴别器可以为Me1GAN鉴别器,其中,所述鉴别器由三层鉴别网络构成。

[0106] 具体地,所述将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,得到鉴别结果,包括:

[0107] 分别利用所述鉴别器中的第一鉴别网络、第二鉴别网络和第三鉴别网络对所述目

标转换音频和所述嵌入语音数据进行第一鉴别值、第二鉴别值和第三鉴别值；

[0108] 对所述第一鉴别值、所述第二鉴别值和所述第三鉴别值进行权重归一化，得到最终鉴别值；

[0109] 若所述最终鉴别值大于或者等于预设的鉴别阈值，得到所述目标转换音频为标准转换音频的鉴别结果；

[0110] 若所述最终鉴别值小于预设的鉴别阈值，得到所述目标转换音频为非标准转换音频的鉴别结果。

[0111] 详细地，所述鉴别器中的第一鉴别网络、第二鉴别网络和第三鉴别网络是多尺度的鉴别网络，为了实现所述鉴别器可以学习音频不同频率范围的特征，故采用了多个不同尺度的鉴别网络。

[0112] 进一步地，所述对所述第一鉴别值、所述第二鉴别值和所述第三鉴别值进行权重归一化，得到最终鉴别值，包括：

[0113] 利用如下归一化公式对所述第一鉴别值、所述第二鉴别值和所述第三鉴别值进行权重归一化：

[0114]  $D=0.1*a+0.2*b+0.3*c$

[0115] 其中，D为最终鉴别值，a为所述第一鉴别值，b为所述第二鉴别值，c为所述第三鉴别值。

[0116] 将所述最终鉴别值与预设的鉴别阈值进行比较判断，并在所述最终鉴别值大于或者等于预设的鉴别阈值，得到所述目标转换音频为标准转换音频的鉴别结果，在所述最终鉴别值小于预设的鉴别阈值，得到所述目标转换音频为非标准转换音频的鉴别结果。

[0117] S5、判断所述鉴别结果与预设的真实结果是否一致，若所述鉴别结果与所述真实结果一致，将所述语音转换模型输出为标准语音转换模型。

[0118] 本发明实施例中，判断所述鉴别结果与预设的真实结果是否一致，根据判断的结果进而对模型进行不同的处理，若所述鉴别结果与所述真实结果一致，说明此时鉴别器鉴别的正确，故将所述语音转换模型输出为标准语音转换模型。

[0119] 详细地，在本方案中，所述鉴别结果存在两种情况，分别为所述目标转换音频为标准转换音频的情况和所述目标转换音频为非标准转换音频的情况，所述预设的真实结果可以为所述目标转换音频为标准转换音频，因此，可以判断所述鉴别结果与预设的真实结果是否一致。

[0120] S6、若所述鉴别结果与所述真实结果不一致，对所述语音转换模型进行参数调整并重新执行鉴别处理的操作，直至重新执行鉴别处理得到的鉴别结果与所述真实结果一致，输出标准语音转换模型。

[0121] 本发明实施例中，当所述鉴别结果与所述真实结果不一致时，对所述语音转换模型进行参数调整，其中，主要是对所述语音数据模型中的鉴别器的模型参数进行调整，所述模型参数可以为模型权重参数或者模型梯度参数，利用调整参数后的语音转换模型重新执行鉴别处理，得到新的鉴别结果并与所述真实结果进行比对，直至重新执行鉴别处理得到的鉴别结果与所述真实结果一致，输出标准语音转换模型。

[0122] S7、获取待转换语音数据和目标对象的声音数据，将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中，得到所述待转换语音数据对应的最

终目标语音。

[0123] 本发明实施例中,可以获取标识号并根据标识号去获取目标对象的声音数据。所述标识号为目标对象的身份标识,在本方案中,想要将所述待转换语音数据转换为与目标对象的声音数据的音色相同,但同时保持所述待转换语音数据的语音内容不变的最终目标语音。

[0124] 具体地,将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中,所述标准语音转换模型会输出语音内容不改变,但是音色变为目标对象的声音数据的最终目标语音。

[0125] 例如,所述待转换语音数据为F,若在保证所述待转换语音数据F中的语音内容不变的前提下将所述待转换语音数据的音色转为最终目标语音G的音色,则需要获取所述目标对象的声音数据G对应的标识号g,并将所述待转换语音数据F和所述目标对象的声音数据G输入至所述标准语音转换模型中,得到语音内容与所述待转换语音数据F相同但是音色与所述目标对象的声音数据G一致的最终目标语音。

[0126] 本发明实施例中,通过对目标语音数据进行编码,得到嵌入语音数据,所述嵌入语音数据包含了标识了身份信息标识号特征和目标语音数据特征,故编码处理可以使得所述嵌入语音数据所包含的信息更加全面和丰富,根据生成器和鉴别器构成语音转换模型,所述生成器用于生成数据样本,所述鉴别器用于鉴别所述数据样本的真实性,进而去调整生成器的参数,因此所述生成器和所述鉴别器可以达到一种博弈平衡状态,进而保证所述语音转换模型输出数据的准确度,将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器进行生成转换处理,保证生成器生成的目标转换音频更具有真实性,将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,所述鉴别器可以学习音频不同频率范围的特征,判断所述鉴别结果与预设的真实结果是否一致进而输出标准语音转换模型,保证语音转换模型输出的准确性。所述语音转换模型可以将所述生成器和所述鉴别器集成在一起,并将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中,得到所述待转换语音数据对应的最终目标语音。因此本发明提出的语音转换方法可以解决语音转换的效率较低的问题。

[0127] 如图2所示,是本发明一实施例提供的语音转换装置的功能模块图。

[0128] 本发明所述语音转换装置100可以安装于电子设备中。根据实现的功能,所述语音转换装置100可以包括数据编码模块101、模型构建模块102、模型训练模块103及最终目标语音生成模块104。本发明所述模块也可以称之为单元,是指一种能够被电子设备处理器所执行,并且能够完成固定功能的一系列计算机程序段,其存储在电子设备的存储器中。

[0129] 在本实施例中,关于各模块/单元的功能如下:

[0130] 所述数据编码模块101,用于获取源语音数据和目标语音数据,对所述目标语音数据进行编码,得到嵌入语音数据;

[0131] 所述模型构建模块102,用于获取预设的生成器和鉴别器,根据所述生成器和所述鉴别器构成语音转换模型;

[0132] 所述模型训练模块103,用于将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生成语音,得到目标转换音频,将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,得到鉴别结果,判断所述

鉴别结果与预设的真实结果是否一致,若所述鉴别结果与所述真实结果一致,将所述语音转换模型输出为标准语音转换模型,若所述鉴别结果与所述真实结果不一致,对所述语音转换模型进行参数调整并重新执行鉴别处理的操作,直至重新执行鉴别处理得到的鉴别结果与所述真实结果一致,输出标准语音转换模型;

[0133] 所述最终目标语音生成模块104,用于获取待转换语音数据和目标对象的声音数据,将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中,得到所述待转换语音数据对应的最终目标语音。

[0134] 详细地,所述语音转换装置100各模块的具体实施方式如下:

[0135] 步骤一、获取源语音数据和目标语音数据,对所述目标语音数据进行编码,得到嵌入语音数据。

[0136] 本发明实施例中,所述源语音数据为进行语音转换之前的音频数据,所述目标语音数据为语音转换的目标音频数据。例如,所述语音转换的目标是在不改变语言内容的情况下,调整音色,将A音频转换为B音频,则所述A音频为所述源语音数据,所述B音频为所述目标语音数据。

[0137] 具体地,所述对所述目标语音数据进行编码,得到嵌入语音数据,包括:

[0138] 根据预设的字典获取所述目标语音数据对应的标识号;

[0139] 对所述标识号和所述目标语音数据进行向量化处理,得到嵌入语音数据。

[0140] 详细地,所述字典中包含音频数据和标识号的一一对应关系,其中,所述音频数据为不同目标人群对应的语音数据,所述标识号是指目标人群的身份标识号,可以根据所述字典查找到所述音频数据对应的标识号,将所述标识号和所述目标语音数据输入至预获取的编码器中进行向量化处理,得到嵌入语音数据。

[0141] 例如,所述字典中包含{源语音数据:1,目标语音数据:2},即所述源语音数据对应的标识号为1,所述目标语音数据对应的标识号为2,根据所述字典可以获取所述目标语音数据对应的标识号2,将所述标识号2和所述目标语音数据一起输入至所述编码器中,得到嵌入语音数据。

[0142] 其中,通过对所述目标语音数据进行编码,可以包含标识了身份信息标识号特征和目标语音数据特征的嵌入语音数据,丰富了所述嵌入语音数据所包含的身份信息和语音信息。

[0143] 步骤二、获取预设的生成器和鉴别器,根据所述生成器和所述鉴别器构成语音转换模型。

[0144] 本发明实施例中,所述生成器用于转换后的音频数据,所述鉴别器用于鉴别输入音频是真实音频或者虚假音频,其中,本方案中采用的所述生成器为StarGAN-VC2生成器,采用的所述鉴别器为MelGAN鉴别器。

[0145] 具体地,所述根据所述生成器和所述鉴别器构成语音转换模型,包括:

[0146] 分别初始化所述生成器和所述鉴别器的参数;

[0147] 将所述源语音数据输入至初始化后的生成器中得到生成语音数据,判断所述生成语音数据与所述目标语音数据是否一致;

[0148] 若所述生成语音数据与所述目标语音数据不一致,则对所述生成器中的各个模块进行顺序调整,并对调整模块顺序后的生成器重新执行语音生成处理;

[0149] 若所述生成语音数据与所述目标语音数据一致,根据预设的连接顺序将初始化后的所述生成器和所述鉴别器进行连接,得到所述语音转换模型。

[0150] 进一步地,所述预获取的生成器中包含多个模块,例如下采样层和上采样层等,预获取的生成器中的多个模块有固定的连接顺序,将所述源语音数据输入至初始化后的生成器中得到生成语音数据后,需要判断所述生成语音数据与所述目标语音数据是否一致,当所述生成语音数据与所述目标语音数据不一致时,可以对所述生成器中的多个模块进行顺序的调整,并将所述源语音数据重新输入至调整模块顺序后的生成器中,直至输出的语音数据与目标语音数据一致,将所述生成器与所述初始化后的鉴别器进行连接,得到所述语音转换模型。

[0151] 其中,在本发明实施例中,以所述生成器在先,所述鉴别器在后的连接顺序构建得到所述语音转换模型。

[0152] 详细地,利用所述生成器和所述鉴别器构成语音转换模型可以产生更好的输出样本,所述生成器用于生成转换后的数据,所述鉴别器用于判别数据的真假,所述生成器和所述鉴别器可以达到一种博弈平衡状态,进而保证所述语音转换模型输出数据的准确度。

[0153] 步骤三、将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生成语音,得到目标转换音频。

[0154] 本发明实施例中,所述语音转换模型中的生成器为StarGAN-VC2生成器。

[0155] 具体地,所述将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生成语音,得到目标转换音频,包括:

[0156] 对所述嵌入语音数据进行第一特征提取,得到第一特征数据集,对所述源语音数据进行第二特征提取,得到第二特征数据集,将所述第一特征数据集和所述第二特征数据集汇总得到总特征数据集;

[0157] 利用所述生成器中的下采样层对所述总特征数据集进行下采样处理,得到下采样数据集;

[0158] 将所述下采样数据集输入至所述生成器中的瓶颈层,并对经过所述瓶颈层处理的数据进行上采样处理,得到上采样数据集;

[0159] 将所述上采样数据集输入至所述生成器中的动态图网络中进行转换,得到目标转换音频。

[0160] 详细地,所述生成器包含下采样层、瓶颈层和上采样层和动态图网络。

[0161] 其中,所述生成器中的动态图网络结构可以对输入的上采样数据集进行矩阵运算,进而得到目标转换音频。

[0162] 进一步地,所述对所述嵌入语音数据进行第一特征提取,得到第一特征数据集,包括:

[0163] 对所述嵌入语音数据进行预加重处理、分帧处理和加窗处理以及快速傅里叶变换,得到所述嵌入语音数据的短时频谱;

[0164] 将所述短时频谱输入预设的梅尔尺度滤波组中,得到梅尔频谱;

[0165] 对所述梅尔频谱进行能量计算,得到对数能量;

[0166] 对所述对数能量进行离散余弦变换,得到第一特征数据集。

[0167] 其中,对所述源语音数据进行第二特征提取的方法与对所述嵌入语音数据进行第

一特征提取的方法一致,此处不再赘述。

[0168] 详细地,通过一个预设的高通滤波器对所述嵌入语音数据进行预加重处理,其中,所述预加重处理可以增强所述嵌入语音数据中语音信号的高频部分。利用预设的采样点将经过预加重处理的嵌入语音数据切分成多帧的数据,得到分帧数据集。

[0169] 在本申请一可选实施例中,所述加窗处理是根据预设的窗口函数对所述分帧数据集中的每一帧进行加窗处理,得到加窗信号。

[0170] 详细地,所述预设的窗口函数为:

$$[0171] \quad S'(n) = S(n) \times W(n)$$

$$[0172] \quad W(n) = 0.5 \left[ 1 - \cos \left( \frac{2\pi n}{N+1} \right) \right] \quad 1 \ll n \ll N$$

[0173] 其中, $S'(n)$ 为加窗信号, $S(n)$ 为分帧数据集, $W(n)$ 为窗口函数, $N$ 为帧的大小, $n$ 为帧的个数。

[0174] 优选地,本申请实施例中,所述预设的窗口函数可选择三角窗,所述 $W(n)$ 为三角窗的函数表达式。

[0175] 本申请实施例对所述分帧数据集进行加窗处理可以增加帧左端和帧右端的连续性,减少频谱泄漏。

[0176] 在本申请一可选实施例中,所述对所述对数能量进行离散余弦变换,得到第一特征数据集,包括:

[0177] 利用下述公式对所述对数能量进行离散余弦变换,得到第一特征数据集:

$$[0178] \quad C(n) = \sum_{m=0}^{N-1} T(m) \cos \left( \frac{\pi n (M - 0.5)}{M} \right)$$

[0179] 其中, $C(n)$ 指第一特征数据集, $T(m)$ 为对数能量, $M$ 为梅尔尺度滤波组中的滤波器个数, $n$ 为帧的个数。

[0180] 为了得到合适大小的声音特征,将所述短时频谱输入预设的梅尔尺度滤波组,变换为梅尔频谱。梅尔频谱可以让人耳对频率的感知度变为线性。对所述梅尔频谱进行倒谱分析,得到特征数据集,其中,所述倒谱分析包括对所述梅尔频谱取对数在做能量转换。

[0181] 步骤四、将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,得到鉴别结果。

[0182] 本发明实施例中,所述鉴别器可以为Me1GAN鉴别器,其中,所述鉴别器由三层鉴别网络构成。

[0183] 具体地,所述将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,得到鉴别结果,包括:

[0184] 分别利用所述鉴别器中的第一鉴别网络、第二鉴别网络和第三鉴别网络对所述目标转换音频和所述嵌入语音数据进行第一鉴别值、第二鉴别值和第三鉴别值;

[0185] 对所述第一鉴别值、所述第二鉴别值和所述第三鉴别值进行权重归一化,得到最终鉴别值;

[0186] 若所述最终鉴别值大于或者等于预设的鉴别阈值,得到所述目标转换音频为标准转换音频的鉴别结果;

[0187] 若所述最终鉴别值小于预设的鉴别阈值,得到所述目标转换音频为非标准转换音频的鉴别结果。

[0188] 详细地,所述鉴别器中的第一鉴别网络、第二鉴别网络和第三鉴别网络是多尺度的鉴别网络,为了实现所述鉴别器可以学习音频不同频率范围的特征,故采用了多个不同尺度的鉴别网络。

[0189] 进一步地,所述对所述第一鉴别值、所述第二鉴别值和所述第三鉴别值进行权重归一化,得到最终鉴别值,包括:

[0190] 利用如下归一化公式对所述第一鉴别值、所述第二鉴别值和所述第三鉴别值进行权重归一化:

$$[0191] \quad D=0.1*a+0.2*b+0.3*c$$

[0192] 其中,D为最终鉴别值,a为所述第一鉴别值,b为所述第二鉴别值,c为所述第三鉴别值。

[0193] 将所述最终鉴别值与预设的鉴别阈值进行比较判断,并在所述最终鉴别值大于或者等于预设的鉴别阈值,得到所述目标转换音频为标准转换音频的鉴别结果,在所述最终鉴别值小于预设的鉴别阈值,得到所述目标转换音频为非标准转换音频的鉴别结果。

[0194] 步骤五、判断所述鉴别结果与预设的真实结果是否一致,若所述鉴别结果与所述真实结果一致,将所述语音转换模型输出为标准语音转换模型。

[0195] 本发明实施例中,判断所述鉴别结果与预设的真实结果是否一致,根据判断的结果进而对模型进行不同的处理,若所述鉴别结果与所述真实结果一致,说明此时鉴别器鉴别的正确,故将所述语音转换模型输出为标准语音转换模型。

[0196] 详细地,在本方案中,所述鉴别结果存在两种情况,分别为所述目标转换音频为标准转换音频的情况和所述目标转换音频为非标准转换音频的情况,所述预设的真实结果可以为所述目标转换音频为标准转换音频,因此,可以判断所述鉴别结果与预设的真实结果是否一致。

[0197] 步骤六、若所述鉴别结果与所述真实结果不一致,对所述语音转换模型进行参数调整并重新执行鉴别处理的操作,直至重新执行鉴别处理得到的鉴别结果与所述真实结果一致,输出标准语音转换模型。

[0198] 本发明实施例中,当所述鉴别结果与所述真实结果不一致时,对所述语音转换模型进行参数调整,其中,主要是对所述语音数据模型中的鉴别器的模型参数进行调整,所述模型参数可以为模型权重参数或者模型梯度参数,利用调整参数后的语音转换模型重新执行鉴别处理,得到新的鉴别结果并与所述真实结果进行比对,直至重新执行鉴别处理得到的鉴别结果与所述真实结果一致,输出标准语音转换模型。

[0199] 步骤七、获取待转换语音数据和目标对象的声音数据,将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中,得到所述待转换语音数据对应的最终目标语音。

[0200] 本发明实施例中,可以获取标识号并根据标识号去获取目标对象的声音数据。所述标识号为目标对象的身份标识,在本方案中,想要将所述待转换语音数据转换为与目标对象的声音数据的音色相同,但同时保持所述待转换语音数据的语音内容不变的最终目标语音。



[0201] 具体地,将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中,所述标准语音转换模型会输出语音内容不改变,但是音色变为目标对象的声音数据的最终目标语音。

[0202] 例如,所述待转换语音数据为F,若在保证所述待转换语音数据F中的语音内容不变的前提下将所述待转换语音数据的音色转为最终目标语音G的音色,则需要获取所述目标对象的声音数据G对应的标识号g,并将所述待转换语音数据F和所述目标对象的声音数据G输入至所述标准语音转换模型中,得到语音内容与所述待转换语音数据F相同但是音色与所述目标对象的声音数据G一致的最终目标语音。

[0203] 本发明实施例中,通过对目标语音数据进行编码,得到嵌入语音数据,所述嵌入语音数据包含了标识了身份信息的标识号特征和目标语音数据特征,故编码处理可以使得所述嵌入语音数据所包含的信息更加全面和丰富,根据生成器和鉴别器构成语音转换模型,所述生成器用于生成数据样本,所述鉴别器用于鉴别所述数据样本的真实性,进而去调整生成器的参数,因此所述生成器和所述鉴别器可以达到一种博弈平衡状态,进而保证所述语音转换模型输出数据的准确度,将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器进行生成转换处理,保证生成器生成的目标转换音频更具有真实性,将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,所述鉴别器可以学习音频不同频率范围的特征,判断所述鉴别结果与预设的真实结果是否一致进而输出标准语音转换模型,保证语音转换模型输出的准确性。所述语音转换模型可以将所述生成器和所述鉴别器集成在一起,并将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中,得到所述待转换语音数据对应的最终目标语音。因此本发明提出的语音转换装置可以解决语音转换的效率较低的问题。

[0204] 如图3所示,是本发明一实施例提供的实现语音转换方法的电子设备的结构示意图。

[0205] 所述电子设备可以包括处理器10、存储器11、通信接口12和总线13,还可以包括存储在所述存储器11中并可在所述处理器10上运行的计算机程序,如语音转换程序。

[0206] 其中,所述存储器11至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、移动硬盘、多媒体卡、卡型存储器(例如:SD或DX存储器等)、磁性存储器、磁盘、光盘等。所述存储器11在一些实施例中可以是电子设备的内部存储单元,例如该电子设备的移动硬盘。所述存储器11在另一些实施例中也可以是电子设备的外部存储设备,例如电子设备上配备的插接式移动硬盘、智能存储卡(Smart Media Card, SMC)、安全数字(Secure Digital, SD)卡、闪存卡(Flash Card)等。进一步地,所述存储器11还可以既包括电子设备的内部存储单元也包括外部存储设备。所述存储器11不仅可以用于存储安装于电子设备的应用软件及各类数据,例如语音转换程序的代码等,还可以用于暂时地存储已经输出或者将要输出的数据。

[0207] 所述处理器10在一些实施例中可以由集成电路组成,例如可以由单个封装的集成电路所组成,也可以是由多个相同功能或不同功能封装的集成电路所组成,包括一个或者多个中央处理器(Central Processing unit, CPU)、微处理器、数字处理芯片、图形处理器及各种控制芯片的组合等。所述处理器10是所述电子设备的控制核心(Control Unit),利用各种接口和线路连接整个电子设备的各个部件,通过运行或执行存储在所述存储器11内

的程序或者模块(例如语音转换程序等),以及调用存储在所述存储器11内的数据,以执行电子设备的各种功能和处理数据。

[0208] 所述通信接口12用于上述电子设备与其他设备之间的通信,包括网络接口和用户接口。可选地,所述网络接口可以包括有线接口和/或无线接口(如WI-FI接口、蓝牙接口等),通常用于在该电子设备与其他电子设备之间建立通信连接。所述用户接口可以是显示器(Display)、输入单元(比如键盘(Keyboard)),可选地,用户接口还可以是标准的有线接口、无线接口。可选地,在一些实施例中,显示器可以是LED显示器、液晶显示器、触控式液晶显示器以及OLED(Organic Light-Emitting Diode,有机发光二极管)触摸器等。其中,显示器也可以适当的称为显示屏或显示单元,用于显示在电子设备中处理的信息以及用于显示可视化的用户界面。

[0209] 所述总线13可以是外设部件互连标准(peripheral component interconnect,简称PCI)总线或扩展工业标准结构(extended industry standard architecture,简称EISA)总线等。该总线13可以分为地址总线、数据总线、控制总线等。所述总线13被设置为实现所述存储器11以及至少一个处理器10等之间的连接通信。

[0210] 图3仅示出了具有部件的电子设备,本领域技术人员可以理解的是,图3示出的结构并不构成对所述电子设备的限定,可以包括比图示更少或者更多的部件,或者组合某些部件,或者不同的部件布置。

[0211] 例如,尽管未示出,所述电子设备还可以包括给各个部件供电的电源(比如电池),优选地,电源可以通过电源管理装置与所述至少一个处理器10逻辑相连,从而通过电源管理装置实现充电管理、放电管理、以及功耗管理等功能。电源还可以包括一个或一个以上的直流或交流电源、再充电装置、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。所述电子设备还可以包括多种传感器、蓝牙模块、Wi-Fi模块等,在此不再赘述。

[0212] 进一步地,所述电子设备还可以包括网络接口,可选地,所述网络接口可以包括有线接口和/或无线接口(如WI-FI接口、蓝牙接口等),通常用于在该电子设备与其他电子设备之间建立通信连接。

[0213] 可选地,该电子设备还可以包括用户接口,用户接口可以是显示器(Display)、输入单元(比如键盘(Keyboard)),可选地,用户接口还可以是标准的有线接口、无线接口。可选地,在一些实施例中,显示器可以是LED显示器、液晶显示器、触控式液晶显示器以及OLED(Organic Light-Emitting Diode,有机发光二极管)触摸器等。其中,显示器也可以适当的称为显示屏或显示单元,用于显示在电子设备中处理的信息以及用于显示可视化的用户界面。

[0214] 应该了解,所述实施例仅为说明之用,在专利申请范围上并不受此结构的限制。

[0215] 所述电子设备中的所述存储器11存储的语音转换程序是多个指令的组合,在所述处理器10中运行时,可以实现:

[0216] 获取源语音数据和目标语音数据,对所述目标语音数据进行编码,得到嵌入语音数据;

[0217] 获取预设的生成器和鉴别器,根据所述生成器和所述鉴别器构成语音转换模型;

[0218] 将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生

成语音,得到目标转换音频;

[0219] 将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,得到鉴别结果;

[0220] 判断所述鉴别结果与预设的真实结果是否一致,若所述鉴别结果与所述真实结果一致,将所述语音转换模型输出为标准语音转换模型;

[0221] 若所述鉴别结果与所述真实结果不一致,对所述语音转换模型进行参数调整并重新执行鉴别处理的操作,直至重新执行鉴别处理得到的鉴别结果与所述真实结果一致,输出标准语音转换模型;

[0222] 获取待转换语音数据和目标对象的声音数据,将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中,得到所述待转换语音数据对应的最终目标语音。

[0223] 具体地,所述处理器10对上述指令的具体实现方法可参考图1对应实施例中相关步骤的描述,在此不赘述。

[0224] 进一步地,所述电子设备集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读存储介质中。所述计算机可读存储介质可以是易失性的,也可以是非易失性的。例如,所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)。

[0225] 本发明还提供一种计算机可读存储介质,所述可读存储介质存储有计算机程序,所述计算机程序在被电子设备的处理器所执行时,可以实现:

[0226] 获取源语音数据和目标语音数据,对所述目标语音数据进行编码,得到嵌入语音数据;

[0227] 获取预设的生成器和鉴别器,根据所述生成器和所述鉴别器构成语音转换模型;

[0228] 将所述嵌入语音数据和所述源语音数据输入至所述语音转换模型中的生成器生成语音,得到目标转换音频;

[0229] 将所述目标转换音频和所述嵌入语音数据输入至所述语音转换模型中的鉴别器中进行鉴别处理,得到鉴别结果;

[0230] 判断所述鉴别结果与预设的真实结果是否一致,若所述鉴别结果与所述真实结果一致,将所述语音转换模型输出为标准语音转换模型;

[0231] 若所述鉴别结果与所述真实结果不一致,对所述语音转换模型进行参数调整并重新执行鉴别处理的操作,直至重新执行鉴别处理得到的鉴别结果与所述真实结果一致,输出标准语音转换模型;

[0232] 获取待转换语音数据和目标对象的声音数据,将所述待转换语音数据和所述目标对象的声音数据输入至所述标准语音转换模型中,得到所述待转换语音数据对应的最终目标语音。

[0233] 在本发明所提供的几个实施例中,应该理解到,所揭露的设备,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。

[0234] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显

示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

[0235] 另外,在本发明各个实施例中的各功能模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能模块的形式实现。

[0236] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。

[0237] 因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附关联图标记视为限制所涉及的权利要求。

[0238] 本发明所指区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain),本质上是一个去中心化的数据库,是一串使用密码学方法相关联产生的数据块,每一个数据块中包含了一批次网络交易的信息,用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层等。

[0239] 此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。系统权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第二等词语用来表示名称,而并不表示任何特定的顺序。

[0240] 最后应说明的是,以上实施例仅用以说明本发明的技术方案而非限制,尽管参照较佳实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或等同替换,而不脱离本发明技术方案的精神和范围。

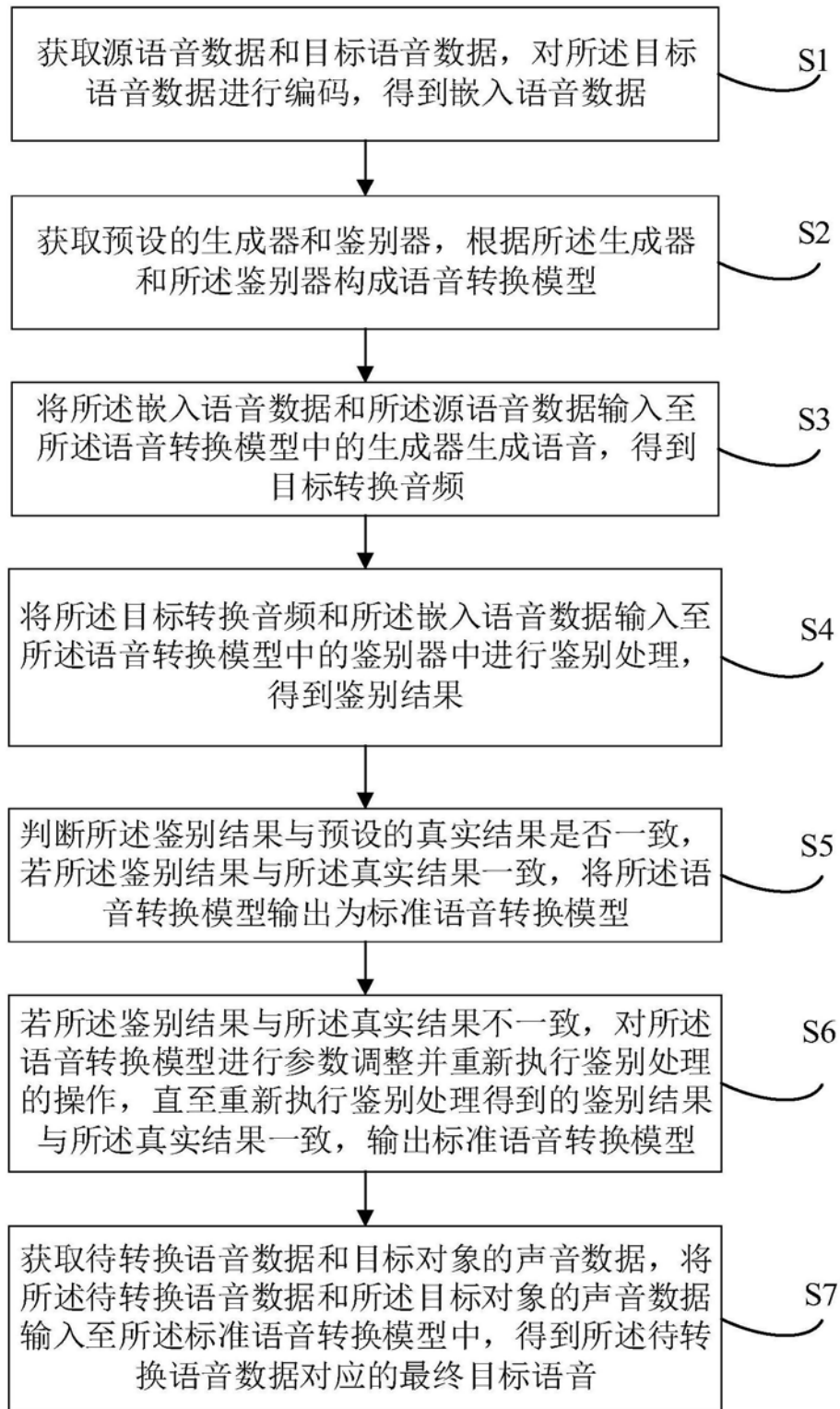


图1

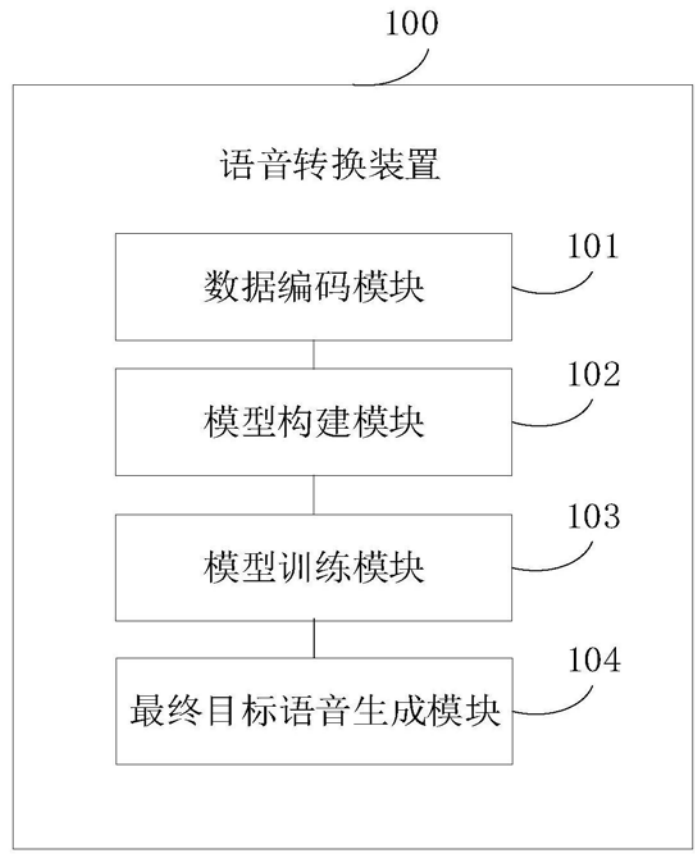


图2

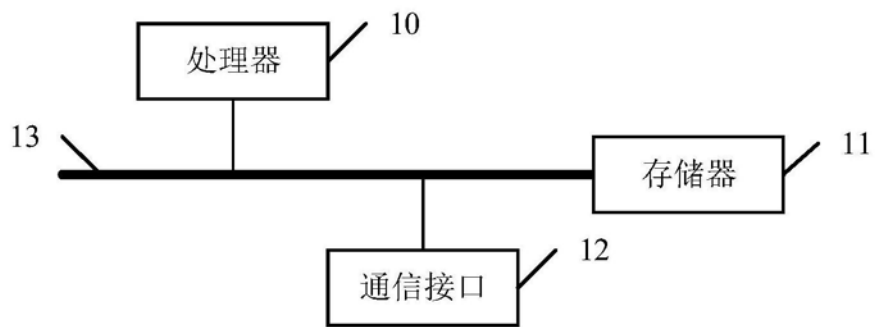


图3