



(12)发明专利

(10)授权公告号 CN 105023574 B

(45)授权公告日 2018.06.15

(21)申请号 201410182886.6

(22)申请日 2014.04.30

(65)同一申请的已公布的文献号

申请公布号 CN 105023574 A

(43)申请公布日 2015.11.04

(73)专利权人 科大讯飞股份有限公司

地址 230088 安徽省合肥市高新开发区望江西路666号

(72)发明人 孙见青 陈凌辉 凌震华 江源

胡国平 胡郁 刘庆峰

(74)专利代理机构 北京维澳专利代理有限公司

11252

代理人 王立民 吉海莲

(51)Int.Cl.

G10L 15/06(2013.01)

G10L 21/02(2013.01)

(56)对比文件

CN 103065619 A,2013.04.24,

CN 1835074 A,2006.09.20,

CN 102982809 A,2013.03.20,

CN 102568476 A,2012.07.11,

JP 特開2007-193139 A,2007.08.02,

US 6038533 A,2000.03.14,

CN 1835075 A,2006.09.20,

陈凌辉等.《基于话者无关模型的说话人转换方法》.《模式识别与人工智能》.2013,第26卷(第3期),

胡郁等.《基于声学统计建模的语音合成技术研究》.《中文信息学报》.2011,第25卷(第6期),

Ling-Hui Chen ET AL..《Estimation of Window Coefficients for Dynamic Feature Extraction for HMM-Based Speech Synthesis》.《INTERSPEECH 2011 12th Annual Conference of the International Speech Communication Association》.2011,

审查员 李海龙

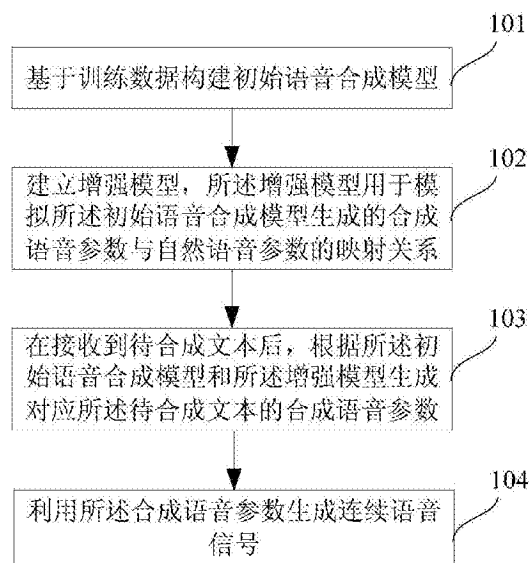
权利要求书2页 说明书9页 附图3页

(54)发明名称

一种实现合成语音增强的方法及系统

(57)摘要

本发明涉及语音合成技术领域,公开了一种实现合成语音增强的方法及系统,该方法包括:基于训练数据构建初始语音合成模型,所述训练数据包括文本数据及与所述文本数据对应的语音数据;建立增强模型,所述增强模型用于模拟所述初始语音合成模型生成的合成语音参数与自然语音参数的映射关系;在接收到待合成文本后,根据所述初始语音合成模型和所述增强模型生成对应所述待合成文本的合成语音参数;利用所述合成语音参数生成连续语音信号。利用本发明,可以有效提高合成语音的增强效果。



1. 一种实现合成语音增强的方法,其特征在于,包括:

基于训练数据构建初始语音合成模型,所述训练数据包括文本数据及与所述文本数据对应的语音数据;所述初始语音合成模型包括:时长模型、频谱模型、基频模型;

建立增强模型,所述增强模型用于模拟所述初始语音合成模型生成的合成语音参数与自然语音参数的映射关系;

在接收到待合成文本后,根据所述初始语音合成模型和所述增强模型生成对应所述待合成文本的合成语音参数,包括:根据所述增强模型对所述初始语音合成模型中的频谱模型和/或基频模型进行增强处理,得到增强的频谱模型和/或基频模型;利用所述增强的频谱模型和/或基频模型生成对应所述待合成文本的频谱参数和/或基频参数;利用所述初始语音合成模型生成对应所述待合成文本的除频谱模型和/或基频模型之外的其它语音参数;

利用所述合成语音参数生成连续语音信号。

2. 根据权利要求1所述的方法,其特征在于,所述建立增强模型包括:

根据所述初始语音合成模型生成所有训练数据的合成语音参数;

提取所有训练数据的自然语音参数;

确定增强模型的拓扑结构;

将对应所述训练数据的合成语音参数和自然语音参数的数据对作为训练集合,根据所述拓扑结构进行参数训练,得到增强模型。

3. 根据权利要求2所述的方法,其特征在于,所述增强模型为:线性函数的映射模型、或者GMM模型、或者DNN模型。

4. 根据权利要求1所述的方法,其特征在于,所述初始语音合成模型生成的合成语音参数与自然语音参数的映射关系为所述初始语音合成模型生成的合成语音参数与自然语音参数的条件分布。

5. 根据权利要求1所述的方法,其特征在于,所述根据所述增强模型对所述初始语音合成模型中的频谱模型和/或基频模型进行增强处理,得到增强的频谱模型和/或基频模型包括:

从所述初始语音合成模型中获取频谱模型和/或基频模型的模型参数;

利用所述增强模型对所述模型参数进行增强处理,得到增强后的模型参数;

将增强后的模型参数替代对应的频谱模型和/或基频模型的模型参数,得到增强的频谱模型和/或基频模型。

6. 一种实现合成语音增强的系统,其特征在于,包括:

初始模型建立模块,用于基于训练数据构建初始语音合成模型,所述训练数据包括文本数据及与所述文本数据对应的语音数据;所述初始语音合成模型包括:时长模型、频谱模型、基频模型;

增强模型建立模块,用于建立增强模型,所述增强模型用于模拟所述初始语音合成模型生成的合成语音参数与自然语音参数的映射关系;

接收模块,用于接收待合成文本;

参数生成模块,用于根据所述初始语音合成模型和所述增强模型生成对应所述待合成文本的合成语音参数;所述参数生成模块包括:模型增强单元,用于根据所述增强模型对所

述初始语音合成模型中的频谱模型和/或基频模型进行增强处理,得到增强的频谱模型和/或基频模型;增强语音参数生成单元,用于利用所述增强的频谱模型和/或基频模型生成对应所述待合成文本的频谱参数和/或基频参数;初始语音参数生成单元,用于利用所述初始语音合成模型生成对应所述待合成文本的除频谱模型和/或基频模型之外的其它语音参数;

合成模块,用于利用所述合成语音参数生成连续语音信号。

7. 根据权利要求6所述的系统,其特征在于,所述增强模型建立模块包括:

合成语音参数生成单元,用于根据所述初始语音合成模型生成所有训练数据的合成语音参数;

自然语音参数提取单元,用于提取所有训练数据的自然语音参数;

拓扑结构确定单元,用于确定增强模型的拓扑结构;

训练单元,用于将对应所述训练数据的合成语音参数和自然语音参数的数据对作为训练集合,根据所述拓扑结构进行参数训练,得到增强模型。

8. 根据权利要求6所述的系统,其特征在于,所述模型增强单元包括:

模型参数获取单元,用于从所述初始语音合成模型中获取频谱模型和/或基频模型的模型参数;

模型参数增强单元,用于利用所述增强模型对所述模型参数进行增强处理,得到增强后的模型参数;

增强模型生成单元,用于将增强后的模型参数替代对应的频谱模型和/或基频模型的模型参数,得到增强的频谱模型和/或基频模型。

一种实现合成语音增强的方法及系统

技术领域

[0001] 本发明涉及语音合成技术领域,具体涉及一种实现合成语音增强的方法及系统。

背景技术

[0002] 实现人机之间人性化、智能化的有效交互,构建高效自然的人机交流环境,已经成为当前信息技术应用和发展的迫切需求。作为语音技术中实用的一项重要技术,语音合成技术或称文语转换技术(Text-To-Speech,TTS),将文字信息转化为自然的语音信号,实现文本的实时转换,改变了传统通过录音回放实现机器开口说话的繁琐操作,节省了系统存储空间,在信息交互日益增多的当今,特别是在信息内容需要经常变动的动态查询应用领域发挥了越来越重要的作用。

[0003] 基于参数合成的语音合成系统由于具有较好的鲁棒性和推广性得到了广泛的应用,然而该方法具有较强的平滑作用,合成的语音平淡且音质容易受损,在合成自然度方面表现不够理想,实际应用中存在一定的提升空间。如何提高合成语音的自然度是合成系统实用化的重要保障。

[0004] 为此,现有技术中主要采用合成语音增强的方法来提高合成语音的自然度,其主要技术可以归纳为:基于人的听感特性等经验知识对生成频谱参数或合成语音进行后滤波处理,比如对合成语音的共振峰进行增强处理、加强生成频谱参数的动态特性,从而改善合成语音的音质。

[0005] 实际上,不同发音人的声学特性存在细节差异,且对于同一个发音人,其在发不同音时,声学特性也存在细节差异。而基于经验知识的合成语音增强方法,只能使得增强后的合成语音在总体上符合人的听感,其增强效果并不理想。

发明内容

[0006] 本发明实施例提供一种实现合成语音增强的方法及系统,以提高合成语音的增强效果。

[0007] 为此,本发明实施例提供如下技术方案:

[0008] 一种实现合成语音增强的方法,包括:

[0009] 基于训练数据构建初始语音合成模型,所述训练数据包括文本数据及与所述文本数据对应的语音数据;所述初始语音合成模型包括:时长模型、频谱模型、基频模型;

[0010] 建立增强模型,所述增强模型用于模拟所述初始语音合成模型生成的合成语音参数与自然语音参数的映射关系;

[0011] 在接收到待合成文本后,根据所述初始语音合成模型和所述增强模型生成对应所述待合成文本的合成语音参数,包括:根据所述增强模型对所述初始语音合成模型中的频谱模型和/或基频模型进行增强处理,得到增强的频谱模型和/或基频模型;利用所述增强的频谱模型和/或基频模型生成对应所述待合成文本的频谱参数和/或基频参数;利用所述初始语音合成模型生成对应所述待合成文本的除频谱模型和/或基频模型之外的其它语音

参数;

[0012] 利用所述合成语音参数生成连续语音信号。

[0013] 优选地,所述建立增强模型包括:

[0014] 根据所述初始语音合成模型生成所有训练数据的合成语音参数;

[0015] 提取所有训练数据的自然语音参数;

[0016] 确定增强模型的拓扑结构;

[0017] 将对应所述训练数据的合成语音参数和自然语音参数的数据对作为训练集合,根据所述拓扑结构进行参数训练,得到增强模型。

[0018] 优选地,所述增强模型为:线性函数的映射模型、或者GMM模型、或者DNN模型。

[0019] 优选地,所述初始语音合成模型生成的合成语音参数与自然语音参数的映射关系为所述初始语音合成模型生成的合成语音参数与自然语音参数的条件分布。

[0020] 优选地,所述根据所述增强模型对所述初始语音合成模型中的频谱模型和/或基频模型进行增强处理,得到增强的频谱模型和/或基频模型包括:

[0021] 从所述初始语音合成模型中获取频谱模型和/或基频模型的模型参数;

[0022] 利用所述增强模型对所述模型参数进行增强处理,得到增强后的模型参数;

[0023] 将增强后的模型参数替代对应的频谱模型和/或基频模型的模型参数,得到增强的频谱模型和/或基频模型。

[0024] 一种实现合成语音增强的系统,包括:

[0025] 初始模型建立模块,用于基于训练数据构建初始语音合成模型,所述训练数据包括文本数据及与所述文本数据对应的语音数据;所述初始语音合成模型包括:时长模型、频谱模型、基频模型;

[0026] 增强模型建立模块,用于建立增强模型,所述增强模型用于模拟所述初始语音合成模型生成的合成语音参数与自然语音参数的映射关系;

[0027] 接收模块,用于接收待合成文本;

[0028] 参数生成模块,用于根据所述初始语音合成模型和所述增强模型生成对应所述待合成文本的合成语音参数;所述参数生成模块包括:模型增强单元,用于根据所述增强模型对所述初始语音合成模型中的频谱模型和/或基频模型进行增强处理,得到增强的频谱模型和/或基频模型;增强语音参数生成单元,用于利用所述增强的频谱模型和/或基频模型生成对应所述待合成文本的频谱参数和/或基频参数;初始语音参数生成单元,用于利用所述初始语音合成模型生成对应所述待合成文本的除频谱模型和/或基频模型之外的其它语音参数;

[0029] 合成模块,用于利用所述合成语音参数生成连续语音信号。

[0030] 优选地,所述增强模型建立模块包括:

[0031] 合成语音参数生成单元,用于根据所述初始语音合成模型生成所有训练数据的合成语音参数;

[0032] 自然语音参数提取单元,用于提取所有训练数据的自然语音参数;

[0033] 拓扑结构确定单元,用于确定增强模型的拓扑结构;

[0034] 训练单元,用于将对应所述训练数据的合成语音参数和自然语音参数的数据对作为训练集合,根据所述拓扑结构进行参数训练,得到增强模型。

[0035] 优选地,所述模型增强单元包括:

[0036] 模型参数获取单元,用于从所述初始语音合成模型中获取频谱模型和/或基频模型的模型参数;

[0037] 模型参数增强单元,用于利用所述增强模型对所述模型参数进行增强处理,得到增强后的模型参数;

[0038] 增强模型生成单元,用于将增强后的模型参数替代对应的频谱模型和/或基频模型的模型参数,得到增强的频谱模型和/或基频模型。

[0039] 本发明实施例提供的实现合成语音增强的方法及系统,基于统计的方法构建用于模拟传统语音合成模型生成的合成语音参数与自然语音参数的映射关系的增强模型,然后利用该增强模型及传统语音合成模型生成对应待合成文本的合成语音参数,进而利用所述合成语音参数生成连续语音信号。本发明实施例的方案以自然声学参数作为指导,因此可以对不同发音人、以及同一发音人发不同音时声学参数的细节特征有较强的把握,可以抓住特定发音人的特性,使合成语音增强的效果更好。

附图说明

[0040] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明中记载的一些实施例,对于本领域普通技术人员来讲,还可以根据这些附图获得其他的附图。

[0041] 图1是本发明实施例实现合成语音增强的方法的流程图;

[0042] 图2是本发明实施例中根据初始语音合成模型和增强模型合成语音的一种流程图;

[0043] 图3是本发明实施例中根据初始语音合成模型和增强模型合成语音的另一种流程图;

[0044] 图4是本发明实施例实现合成语音增强的系统的结构示意图;

[0045] 图5是本发明实施例中参数生成模块的一种具体实现结构示意图;

[0046] 图6是本发明实施例中参数生成模块的另一种具体实现结构示意图。

具体实施方式

[0047] 为了使本技术领域的人员更好地理解本发明实施例的方案,下面结合附图和实施方式对本发明实施例作进一步的详细说明。

[0048] 由于不同发音人的声学特性存在细节差异,且对于同一个发音人,其在发不同音时,声学特性也存在细节差异。而现有的合成语音增强方法基于人的听感特性等经验知识对生成频谱参数或合成语音进行后滤波处理,没有关注发音人声学参数的细节特性,只能使得增强后的合成语音在总体上符合人的听感,不能获得理想的增强效果。为此,本发明实施例针对现有技术存在的问题,提供一种实现合成语音增强的方法及系统,基于统计的方法构建用于模拟传统语音合成模型生成的合成语音参数与自然语音参数的映射关系的增强模型,然后利用该增强模型及传统语音合成模型生成对应待合成文本的合成语音参数,进而利用所述合成语音参数生成连续语音信号。

[0049] 如图1所示,是本发明实施例实现合成语音增强的方法的流程图,包括以下步骤:

[0050] 步骤101,基于训练数据构建初始语音合成模型,所述训练数据包括文本数据及与所述文本数据对应的语音数据。

[0051] 所述初始语音合成模型可以使用传统的参数合成方法进行构建,其包括:各基本合成单元对应的二叉决策树、频谱模型、基频模型、时长模型等。比如,可以采用基于HMM的参数合成方法,对于频谱模型,采用GMM(Gaussian Mixture Mode,高斯混合模型)来模拟叶节点的频谱分布,其高斯数通常可以参考训练数据规模确定为正整数,比如选择高斯数为1。

[0052] 步骤102,建立增强模型,所述增强模型用于模拟所述初始语音合成模型生成的合成语音参数与自然语音参数的映射关系。

[0053] 由于增强模型的设置和优化对合成语音增强效果有着重要的影响,因此,在本发明实施例中,采用基于数据驱动的增强模型设置方式,以自然声学参数作为指导,真实体现不同发音人、以及同一发音人发不同音时声学参数的细节特征,进而提高合成语音增强的效果。

[0054] 增强模型的构建过程如下:

[0055] (1) 根据初始语音合成模型生成所有训练数据的合成语音参数;

[0056] (2) 提取所有训练数据的自然语音参数;

[0057] (3) 确定增强模型的拓扑结构;

[0058] (4) 将对应所述训练数据的合成语音参数和自然语音参数的数据对作为训练集合,根据所述拓扑结构进行参数训练,得到增强模型。

[0059] 需要说明的是,在实际应用中,可以分别构建针对频谱特性和/或基频特性的增强模型。比如,针对频谱特性的增强模型,具体的构建过程如下:

[0060] (1) 根据初始语音合成模型中的频谱模型生成所有训练数据的合成频谱参数。

[0061] 根据所述频谱模型、以及强制对齐结果,可确定训练数据对应的频谱模型序列。具体地,对于单个基本语音单元,根据强制对齐时长信息将选定的频谱模型进行多次拷贝,获取该基本语音单元的频谱特征系列模型。

[0062] 统计训练数据对应的频谱模型序列的似然度总和,计算如下:

[0063]

$$\log P(W|C_s, Q, \lambda) = -\frac{1}{2} C_s^T W^T U_s^{-1} W C_s + C_s^T W^T U_s^{-1} M_s + \text{const} \quad (1)$$

[0064] 其中W是计算动态参数的窗函数矩阵, C_s 为待生成的频谱参数, M_s 和 U_s 分别为频谱模型的均值和协方差矩阵。显然频谱模型的似然度总和是目标频谱特征矢量的函数。

[0065] (2) 提取所有训练数据的自然频谱参数。

[0066] (3) 确定频谱增强模型的拓扑结构。

[0067] 频谱增强模型用于模拟传统语音合成模型生成的频谱参数与自然频谱参数的映射关系,在本发明实施例中,可以采用线性函数的映射模型,也可以采用GMM模型或DNN模型等数学统计模型。一般来说,模型越精细则在数据充分的情况下其模拟效果越好。

[0068] (4) 根据所述拓扑结构对频谱增强模型进行参数训练,获取优化的频谱增强模型,即建立合成频谱参数 x_t 与自然频谱参数 y_t 的条件分布 $p(y_t|x_t)$ 。

[0069] 针对基频特性的增强模型的构建过程与上述类似,在此不再详细描述。

[0070] 步骤103,在接收到待合成文本后,根据所述初始语音合成模型和所述增强模型生成对应所述待合成文本的合成语音参数。

[0071] 基于上述增强模型,在实际应用中,可以采用多种方式对初始语音合成模型或合成语音参数进行增强,均可以得到很好的增强效果,具体实现过程将在后面详细描述。

[0072] 步骤104,利用所述合成语音参数生成连续语音信号。

[0073] 本发明实施例提供的实现合成语音增强的方法,基于统计的方法构建用于模拟传统语音合成模型生成的合成语音参数与自然语音参数的映射关系的增强模型,然后利用该增强模型及传统语音合成模型生成对应待合成文本的合成语音参数,进而利用所述合成语音参数生成连续语音信号。由于所述增强模型是以自然声学参数作为指导,因此可以对不同发音人、以及同一发音人发不同音时声学参数的细节特征有较强的把握,可以抓住特定发音人的特性,使合成语音增强的效果更好。而且,本发明实施例的方案在实际合成任务中不会增加运算量,有利于产品的实时化。

[0074] 需要说明的是,在实际应用中,根据初始语音合成模型和增强模型生成合成语音参数的方式有多种。比如,可以利用相应的增强模型对初始语音合成模型中的频谱模型和/或基频模型进行增强处理,利用增强处理后的频谱模型和/或基频模型生成对应待合成文本的频谱参数和/或基频参数,由初始语音合成模型生成其它语音合成参数,然后利用这些语音合成参数生成连续语音信号。再比如,还可以先利用初始语音合成模型生成对应待合成文本的语音合成参数(包括时长参数、频谱参数、基频参数),然后再利用相应的增强模型对其中的一些语音合成参数(包括频谱参数和/或基频参数)进行增强处理,最后利用这些增强后的语音合成参数及另外一些未增强处理的语音合成参数(主要是时长参数)生成连续语音信号。

[0075] 下面分别举例详细说明本发明实施例中根据初始语音合成模型和增强模型生成合成语音参数的过程。

[0076] 如图2所示,是本发明实施例中根据初始语音合成模型和增强模型生成合成语音参数的一种流程图,包括以下步骤:

[0077] 步骤201,利用初始语音合成模型生成对应待合成文本的时长参数和基频参数。

[0078] 步骤202,根据增强模型对初始语音合成模型中的频谱模型进行增强处理,得到增强的频谱模型。

[0079] 首先,从初始的频谱模型中获取模型参数,比如基于GMM的频谱模型的,记为 x_t ;然后利用预先训练好的增强模型,对模型参数 x_t 进行增强处理,即根据 $p(y_t|x_t)$,求得增强后的模型参数 y_t ;最后用增强后的模型参数 y_t 替换频谱模型的模型参数,得到新的频谱模型,此模型即为增强后的频谱模型。

[0080] 步骤203,利用增强的频谱模型生成对应所述待合成文本的频谱参数。

[0081] 步骤204,利用对应所述待合成文本的时长参数、基频参数、以及频谱参数生成连续语音信号。

[0082] 需要说明的是,在实际应用中,可以分别生成针对频谱特性的增强模型及针对基频特性的增强模型,因此,可以单独采用针对频谱特性的增强模型对初始语音合成模型中的频谱模型进行增强处理,或者单独采用针对基频特性的增强模型对初始语音合成模型中的基频模型进行增强处理,也可以综合采用上述两种针对不同特性的增强模型分别对初始

语音合成模型中的频谱模型和基频模型进行增强处理。相应地,利用增强后的频谱模型和/或基频模型得到对应待合成文本的基频参数和/或频谱参数,利用这些语音合成参数及通过初始语音合成模型得到的其它语音合成参数,即可生成连续语音信号。

[0083] 利用上述各语音合成参数生成连续语音信号的过程与现有技术类似,在此不再赘述。

[0084] 由此可见,本发明实施例的方法对传统语音合成模型进行增强处理,在后续的合成任务中,只需使用增强后的语音合成模型得到相应的语音合成参数即可,不会增加运算量,而且可以达到很好的增强效果。

[0085] 如图3所示,是本发明实施例中根据初始语音合成模型和增强模型合成语音的另一种流程图。

[0086] 步骤301,利用初始语音合成模型分别生成对应待合成文本的时长参数、频谱参数和基频参数。

[0087] 步骤302,利用增强模型对频谱参数进行增强处理,得到增强后的频谱参数。

[0088] 具体地,将前面公式(1)中的频谱参数 C_s 代入前面增强处理模型 $p(y_t|x_t)$ 中的 x_t ,可得到增强后的频谱参数 y_t 。

[0089] 步骤303,利用对应所述待合成文本的时长参数、基频参数、以及增强后的频谱参数生成连续语音信号。

[0090] 由上述流程可以看出,图3所示流程与图2所示流程不同的是,在该实施例中,先由初始语音合成模型生成对应待合成文本的基频参数、频谱参数、以及时长参数,然后再由相应的增强模型对其中的频谱参数进行增强处理,从而使增强后的语音合成参数更好地体现不同发音人、以及同一发音人发不同音时的声学特性上的细节差异。将这些增强处理后的语音合成参数与通过传统语音合成模型得到的其它语音合成参数结合在一起,通过合成器合成出语音。

[0091] 需要说明的是,在实际应用中,同样可以采用与上述图3类似的流程,利用相应的增强模型对基频参数进行增强处理,得到增强后的基频参数。然后利用对应待合成文本的时长参数、频谱参数、以及增强后的基频参数生成连续语音信号。或者利用针对频谱特性的增强模型对初始语音合成模型生成的频谱参数进行增强处理,同时利用针对基频特性的增强模型对初始语音合成模型生成的基频参数进行增强处理,然后利用初始语音合成模型生成的时长参数、以及增强后的基频参数和频谱参数生成连续语音信号。

[0092] 相应地,本发明实施例还提供一种实现合成语音增强的系统,如图4所示,是该系统的结构示意图。

[0093] 在该实施例中,所述系统包括:

[0094] 初始模型建立模块401,用于基于训练数据构建初始语音合成模型,所述训练数据包括文本数据及与所述文本数据对应的语音数据;

[0095] 增强模型建立模块402,用于建立增强模型,所述增强模型用于模拟所述初始语音合成模型生成的合成语音参数与自然语音参数的映射关系;

[0096] 接收模块403,用于接收待合成文本;

[0097] 参数生成模块404,用于根据所述初始语音合成模型和所述增强模型生成对应所述待合成文本的合成语音参数;

[0098] 合成模块405,用于利用所述合成语音参数生成连续语音信号。

[0099] 上述初始模型建立模块401可以使用传统的参数合成方法构建初始语音合成模型,所述初始语音合成模型包括:各基本合成单元对应的二叉决策树、频谱模型、基频模型、时长模型等。比如,可以采用基于HMM的参数合成方法,对于频谱模型,采用GMM来模拟叶节点的频谱分布,其高斯数通常可以参考训练数据规模确定为正整数,比如选择高斯数为1。

[0100] 由于增强模型的设置和优化对合成语音增强效果有着重要的影响,因此,在本发明实施例中,增强模型建立模块402采用基于数据驱动的增强模型设置方式,以自然声学参数作为指导,真实体现不同发音人、以及同一发音人发不同音时声学参数的细节特征,进而提高合成语音增强的效果。

[0101] 上述增强模型建立模块402具体可以包括以下各单元:

[0102] 合成语音参数生成单元,用于根据所述初始语音合成模型生成所有训练数据的合成语音参数;

[0103] 自然语音参数提取单元,用于提取所有训练数据的自然语音参数;

[0104] 拓扑结构确定单元,用于确定增强模型的拓扑结构;

[0105] 训练单元,用于将对应所述训练数据的合成语音参数和自然语音参数的数据对作为训练集合,根据所述拓扑结构进行参数训练,得到增强模型。

[0106] 需要说明的是,在实际应用中,上述增强模型建立模块402可以分别构建针对频谱特性和/或基频特性的增强模型。相应地,在构建针对频谱特性的增强模型时,所述合成语音参数生成单元需要根据初始语音合成模型中的频谱模型生成所有训练的合成频谱参数;自然语音参数提取单元需要提取所有训练数据的自然频谱参数。类似地,在构建针对基频特性的增强模型时,所述合成语音参数生成单元需要根据初始语音合成模型中的基频模型生成所有训练的合成基频参数;自然语音参数提取单元需要提取所有训练数据的自然基频参数。

[0107] 上述参数生成模块404基于增强模型建立模块402建立的增强模型,可以采用多种方式对初始语音合成模型或合成语音参数进行增强,均可以得到很好的增强效果,相应地,参数生成模块404可以有多种具体实现结构,将在后面详细描述。

[0108] 本发明实施例提供的实现合成语音增强的系统,基于统计的方法构建用于模拟传统语音合成模型生成的合成语音参数与自然语音参数的映射关系的增强模型,然后利用该增强模型及传统语音合成模型生成对应待合成文本的合成语音参数,进而利用所述合成语音参数生成连续语音信号。由于所述增强模型是以自然声学参数作为指导,因此可以对不同发音人、以及同一发音人发不同音时声学参数的细节特征有较强的把握,可以抓住特定发音人的特性,使合成语音增强的效果更好。

[0109] 如图5所示,是本发明实施例中参数生成模块的一种具体实现结构示意图。

[0110] 在该实施例中,所述参数生成模块包括:

[0111] 模型增强单元501,用于根据所述增强模型对所述初始语音合成模型中的频谱模型和/或基频模型进行增强处理,得到增强的频谱模型和/或基频模型;

[0112] 增强语音参数生成单元502,用于利用所述增强的频谱模型和/或基频模型生成对应所述待合成文本的频谱参数和/或基频参数;

[0113] 初始语音参数生成单元503,用于利用所述初始语音合成模型生成对应所述待合

成文本的除频谱模型和/或基频模型之外的其它语音参数。

[0114] 需要说明的是,在实际应用中,前述增强模型建立模块402可以分别生成针对频谱特性的增强模型及针对基频特性的增强模型,因此,在图5所示实施例中,模型增强单元501可以单独采用针对频谱特性的增强模型对初始语音合成模型中的频谱模型进行增强处理,或者单独采用针对基频特性的增强模型对初始语音合成模型中的基频模型进行增强处理,也可以综合采用上述两种针对不同特性的增强模型分别对初始语音合成模型中的频谱模型和基频模型进行增强处理。相应地,增强语音参数生成单元502可以利用增强后的频谱模型和/或基频模型得到对应待合成文本的基频参数和/或频谱参数,图4中的合成模块405利用这些语音合成参数及初始语音参数生成单元503得到的其它语音合成参数,即可生成连续语音信号。

[0115] 由此可见,本发明实施例实现合成语音增强的系统对传统语音合成模型进行增强处理,在后续的合成任务中,只需使用增强后的语音合成模型得到相应的语音合成参数即可,不会增加运算量,而且可以达到很好的增强效果。

[0116] 上述模型增强单元501可以通过对模型均值的增强来得到增强的频谱模型和/或基频模型,模型增强单元501的一种具体结构可以包括以下各单元:

[0117] 模型参数获取单元,用于从所述初始语音合成模型中获取频谱模型和/或基频模型的模型参数;

[0118] 模型参数增强单元,用于利用所述增强模型对所述模型参数进行增强处理,得到增强后的模型参数;

[0119] 增强模型生成单元,用于将增强后的模型参数替代对应的频谱模型和/或基频模型的模型参数,得到增强的频谱模型和/或基频模型。

[0120] 如图6所示,是本发明实施例中参数生成模块的另一种具体实现结构示意图。

[0121] 在该实施例中,所述参数生成模块包括:

[0122] 初始语音参数生成单元601,用于利用初始语音合成模型分别生成对应所述待合成文本的时长参数、频谱参数和基频参数;

[0123] 参数增强单元602,用于利用增强模型对所述频谱参数和/或基频参数进行增强处理,得到增强后的频谱参数和/或基频参数,并将所述增强后的频谱参数和/或基频参数作为合成语音时对应所述待合成文本的频谱参数和/或基频参数。

[0124] 与图5所示框图结构不同的是,在该实施例中,先由初始语音参数生成单元601利用初始语音合成模型生成对应待合成文本的基频参数、频谱参数、以及时长参数,然后再由参数增强单元602利用相应的增强模型对其中的频谱参数进行增强处理,从而使增强后的语音合成参数更好地体现不同发音人、以及同一发音人发不同音时的声学特性上的细节差异。图4中的合成模块405将这些增强处理后的语音合成参数与通过传统语音合成模型得到的其它语音合成参数结合在一起,通过合成器合成出语音。

[0125] 利用本发明实施例实现合成语音增强的系统,通过统计方式获取不同发音人、以及同一发音人发不同音时声学参数的细节特征,进而利用这些细节特征对合成语音进行增强处理,从而可以得到更好的增强效果。

[0126] 本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于系统实

施例而言,由于其基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。以上所描述的系统实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元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。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。

[0127] 以上对本发明实施例进行了详细介绍,本文中应用了具体实施方式对本发明进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及设备;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。

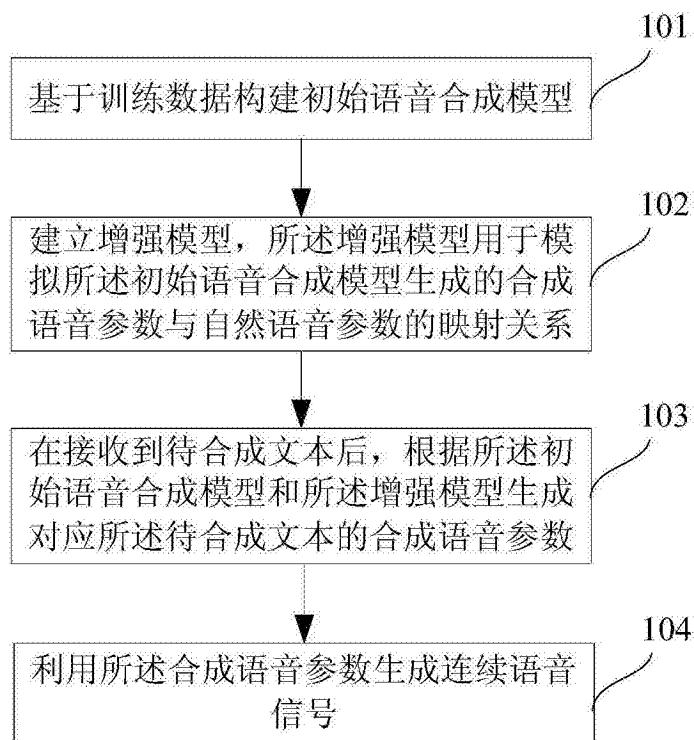


图1

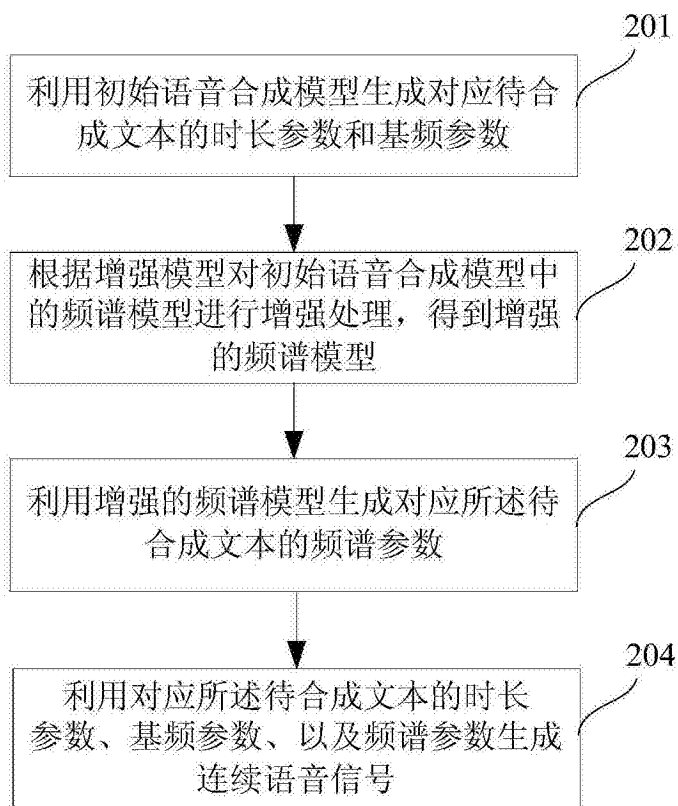


图2

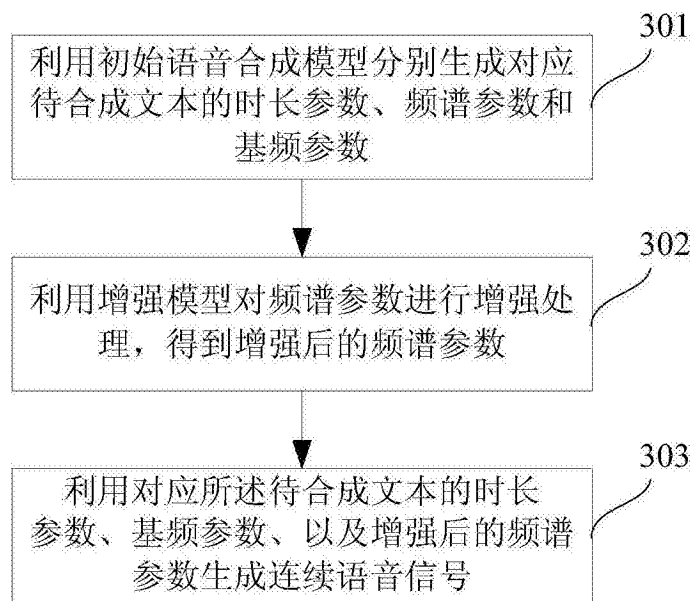


图3

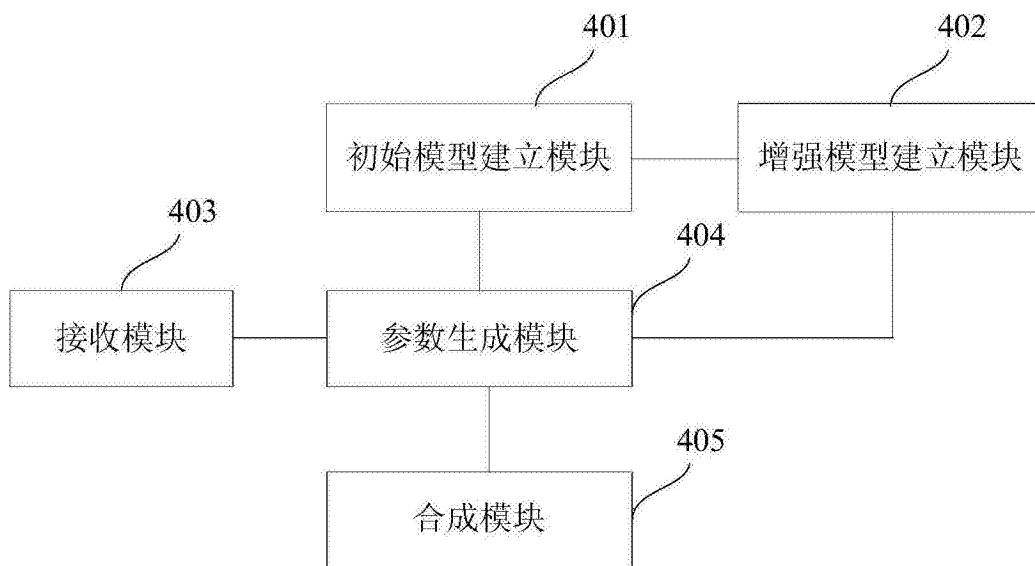


图4

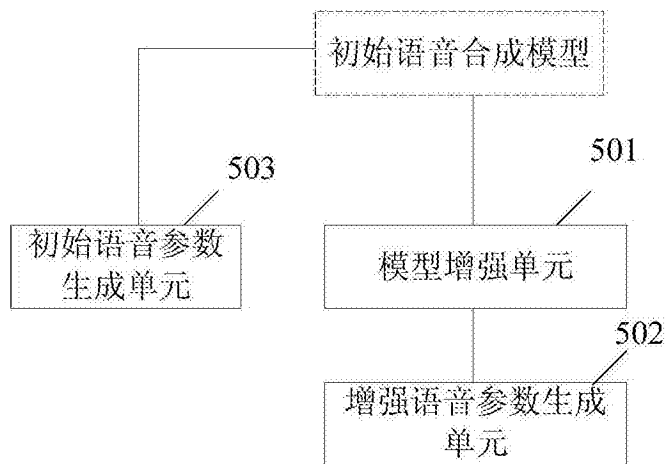


图5

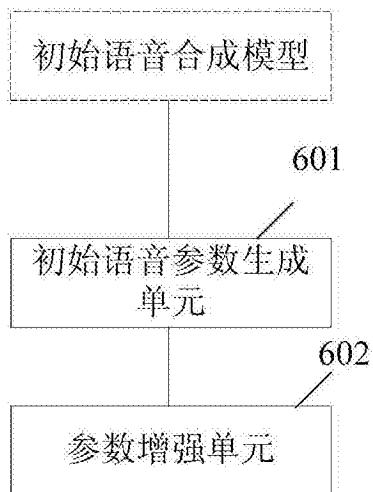


图6