



(12) 发明专利申请

(10) 申请公布号 CN 104766603 A

(43) 申请公布日 2015. 07. 08

(21) 申请号 201410005753. 1

(22) 申请日 2014. 01. 06

(71) 申请人 安徽科大讯飞信息科技股份有限公司

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 王影 江源 凌震华 胡国平 胡郁

(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 13/02(2013. 01)

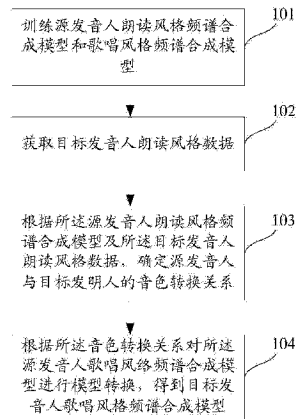
权利要求书2页 说明书8页 附图3页

(54) 发明名称

构建个性化歌唱风格频谱合成模型的方法及装置

(57) 摘要

本发明公开了一种构建个性化歌唱风格频谱合成模型的方法及装置,该方法包括:训练源发音人朗读风格频谱合成模型和歌唱风格频谱合成模型;获取目标发音人朗读风格数据;根据所述源发音人朗读风格频谱合成模型及所述目标发音人朗读风格数据,确定源发音人与目标发明人的音色转换关系;根据所述音色转换关系对所述源发音人歌唱风格频谱合成模型进行模型转换,得到目标发音人歌唱风格频谱合成模型。本发明通过目标发音人少量朗读风格数据即可得到其歌唱风格合成模型,大大降低了一般用户构建歌唱合成模型的难度,进而有效地提高了个性化歌唱合成效果。



1. 一种构建个性化歌唱风格频谱合成模型的方法,其特征在于,包括:

训练源发音人朗读风格频谱合成模型和歌唱风格频谱合成模型;

获取目标发音人朗读风格数据;

根据所述源发音人朗读风格频谱合成模型及所述目标发音人朗读风格数据,确定源发音人与目标发明人的音色转换关系;

根据所述音色转换关系对所述源发音人歌唱风格频谱合成模型进行模型转换,得到目标发音人歌唱风格频谱合成模型。

2. 根据权利要求1所述的方法,其特征在于,所述训练源发音人的朗读风格频谱合成模型和歌唱风格频谱合成模型包括:

获取源发音人的朗读风格语音数据和歌唱风格语音数据,并对所述语音数据进行标注;

提取所述语音数据对应的频谱参数;

利用源发音人的朗读风格语音数据对应的频谱参数和标注信息训练源发音人的朗读风格频谱合成模型;

利用源发音人的歌唱风格语音数据对应的声学参数和标注信息训练源发音人的歌唱风格频谱合成模型。

3. 根据权利要求2所述的方法,其特征在于,所述朗读风格语音数据标注的信息包括音素序列、音段切分、韵律词和韵律短语;所述歌唱风格语音数据的标注信息包括音素序列、音段切分、韵律词、韵律短语、以及与音乐相关的音高和时值信息。

4. 根据权利要求1所述的方法,其特征在于,所述确定所述源发音人的朗读风格频谱合成模型与所述目标发音人的朗读风格数据的转换关系包括:

从所述源发音人朗读风格频谱合成模型中选择执行自适应的模型;

从所述目标发音人的朗读风格数据中选择所述自适应模型所对应的训练数据;

根据所述训练数据对选择执行自适应的模型进行自适应,得到频谱转换矩阵。

5. 根据权利要求1至4任一项所述的方法,其特征在于,所述根据所述音色转换关系对所述源发音人歌唱风格频谱合成模型进行模型转换,得到目标发音人歌唱风格频谱合成模型包括:

根据以下公式获得目标发音人歌唱风格频谱合成模型:

$$\hat{\mu} = \Phi\mu + B$$

$$\hat{\Sigma}^{-1} = AH^{-1}A^T$$

$$\Sigma^{-1} = A A^T$$

其中, μ 和 Σ 为源发音人歌唱风格频谱合成模型的均值和方差, $\hat{\mu}$ 和 $\hat{\Sigma}$ 为转换后的频谱合成模型的均值和方差,即为目标发音人歌唱风格频谱合成模型的均值和方差, $[\Phi, B]$ 和 H 为均值和方差的全局转换矩阵, A 表示所述方差的分解矩阵。

6. 一种构建个性化歌唱风格频谱合成模型的装置,其特征在于,包括:

模型训练模块,用于训练源发音人朗读风格频谱合成模型和歌唱风格频谱合成模型;

数据获取模块,用于获取目标发音人朗读风格数据;

转换关系确定模块,用于根据所述源发音人朗读风格频谱合成模型及所述目标发音人朗读风格数据,确定源发音人与目标发明人的音色转换关系;

转换模块,用于根据所述音色转换关系对所述源发音人歌唱风格频谱合成模型进行模型转换,得到目标发音人歌唱风格频谱合成模型。

7. 根据权利要求 6 所述的装置,其特征在于,所述模型训练模块包括:

数据获取单元,用于获取源发音人的朗读风格语音数据和歌唱风格语音数据,并对所述语音数据进行标注;

参数提取单元,用于提取所述语音数据对应的频谱参数;

第一训练单元,用于利用源发音人的朗读风格语音数据对应的频谱参数和标注信息训练源发音人的朗读风格频谱合成模型;

第二训练单元,用于利用源发音人的歌唱风格语音数据对应的声学参数和标注信息训练源发音人的歌唱风格频谱合成模型。

8. 根据权利要求 7 所述的装置,其特征在于,所述朗读风格语音数据标注的信息包括音素序列、音段切分、韵律词和韵律短语;所述歌唱风格语音数据的标注信息包括音素序列、音段切分、韵律词、韵律短语、以及与音乐相关的音高和时值信息。

9. 根据权利要求 6 所述的装置,其特征在于,所述转换关系确定模块包括:

自适应模型选择单元,用于从所述源发音人朗读风格频谱合成模型中选择执行自适应的模型;

训练数据选择单元,用于从所述目标发音人的朗读风格数据中选择所述自适应模型所对应的训练数据;

自适应单元,用于根据所述训练数据对选择执行自适应的模型进行自适应,得到频谱转换矩阵。

10. 根据权利要求 6 至 9 任一项所述的装置,其特征在于,

所述转换模块,具体用于根据以下公式获得目标发音人歌唱风格频谱合成模型:

$$\hat{\mu} = \Phi\mu + B$$

$$\hat{\Sigma}^{-1} = AH^{-1}A^T$$

$$\Sigma^{-1} = A A^T$$

其中, μ 和 Σ 为源发音人歌唱风格频谱合成模型的均值和方差, $\hat{\mu}$ 和 $\hat{\Sigma}$ 为转换后的频谱合成模型的均值和方差,即为目标发音人歌唱风格频谱合成模型的均值和方差, $[\Phi, B]$ 和 H 为均值和方差的全局转换矩阵, A 表示所述方差的分解矩阵。

构建个性化歌唱风格频谱合成模型的方法及装置

技术领域

[0001] 本发明涉及语音合成技术领域，具体涉及一种构建个性化歌唱风格频谱合成模型的方法及装置。

背景技术

[0002] 近年来，歌唱合成系统得到了广泛的研究和应用，其中基于隐马尔可夫模型 (Hidden Markov Model, HMM) 的参数合成系统以及基于波形拼接合成系统达到了实用标准。相比于基于波形拼接合成系统，基于 HMM 的参数合成方法对训练数据量要求较少且系统规模较小，因而具有较好的实用价值。

[0003] 基于参数合成的歌唱合成系统首先分别确定各基本合成单元 (如音节，音素等) 的时长参数序列、基频参数序列以及频谱参数序列，然后再根据合成参数序列利用参数合成器得到连续歌唱信号。

[0004] 传统歌唱合成系统的构建通常是由固定发音人预先录制歌曲语音，并据此训练得到时长、基频及频谱的合成模型，并通过语音合成器构建歌唱合成系统。基于固定发音人数据构建的歌唱合成系统合成音色较为单一，而很多情况下用户则希望用自己的音色或其指定的音色来合成歌曲，即希望合成歌曲具有个性化的音色特点。在个性化歌唱合成模型构建时，由于普通用户很难收集到大量的歌唱数据且即使获得了少量的歌曲数据，由于受演唱水平的影响、以及演唱者在演唱时或多或少会加入一些演绎成分，也很难获得准确的标注信息 (如下文所述的音素序列、韵律词、韵律短语、音高和时值等信息)，对此，传统的基于语音数据训练得到的合成模型往往不够正确，直接影响了合成效果。

[0005] 针对个性化歌唱合成的应用需求，以及存在的训练歌唱数据不足及标注不够准确的问题，有研究人员提出直接采用个性化朗读风格的频谱模型作为歌唱频谱合成模型。由于语音数据最重要的控制参数基频和频谱具有较强的相关性，如果将一种风格语音的基频与另一种风格语音的频谱共同合成为语音，则会由于基频与频谱不搭配而导致生成的语音不自然。显然，朗读风格语音与歌唱风格语音存在较大的差别，其中歌唱风格语音基频范围跨度很大，而朗读风格语音基频范围跨度相对较小，如果使用朗读风格合成模型生成频谱，而根据歌曲简谱得到基频，则会引入较明显的基频和频谱搭配性问题，使得合成的语音听起来自然度较低。显然合理解决基频参数和频谱参数匹配的问题，获得歌唱风格的个性化频谱合成模型是提高歌唱合成效果的关键。

发明内容

[0006] 本发明提供一种构建个性化歌唱风格频谱合成模型的方法及装置，以解决歌唱风格的个性化频谱合成模型构建中存在的训练歌唱数据不足及标注不够准确的问题，满足个性化歌唱合成的应用需求。

[0007] 为此，本发明提供如下技术方案：

[0008] 一种构建个性化歌唱风格频谱合成模型的方法，包括：

- [0009] 训练源发音人朗读风格频谱合成模型和歌唱风格频谱合成模型；
- [0010] 获取目标发音人朗读风格数据；
- [0011] 根据所述源发音人朗读风格频谱合成模型及所述目标发音人朗读风格数据，确定源发音人与目标发明人的音色转换关系；
- [0012] 根据所述音色转换关系对所述源发音人歌唱风格频谱合成模型进行模型转换，得到目标发音人歌唱风格频谱合成模型。
- [0013] 优选地，所述训练源发音人的朗读风格频谱合成模型和歌唱风格频谱合成模型包括：
- [0014] 获取源发音人的朗读风格语音数据和歌唱风格语音数据，并对所述语音数据进行标注；
- [0015] 提取所述语音数据对应的频谱参数；
- [0016] 利用源发音人的朗读风格语音数据对应的频谱参数和标注信息训练源发音人的朗读风格频谱合成模型；
- [0017] 利用源发音人的歌唱风格语音数据对应的声学参数和标注信息训练源发音人的歌唱风格频谱合成模型。
- [0018] 优选地，所述朗读风格语音数据标注的信息包括音素序列、音段切分、韵律词和韵律短语；所述歌唱风格语音数据的标注信息包括音素序列、音段切分、韵律词、韵律短语、以及与音乐相关的音高和时值信息。
- [0019] 优选地，所述确定所述源发音人的朗读风格频谱合成模型与所述目标发音人的朗读风格数据的转换关系包括：
- [0020] 从所述源发音人朗读风格频谱合成模型中选择执行自适应的模型；
- [0021] 从所述目标发音人的朗读风格数据中选择所述自适应模型所对应的训练数据；
- [0022] 根据所述训练数据对选择执行自适应的模型进行自适应，得到频谱转换矩阵。
- [0023] 优选地，所述根据所述音色转换关系对所述源发音人歌唱风格频谱合成模型进行模型转换，得到目标发音人歌唱风格频谱合成模型包括：
- [0024] 根据以下公式获得目标发音人歌唱风格频谱合成模型：
- [0025] $\hat{\mu} = \Phi\mu + B$
- [0026] $\hat{\Sigma}^{-1} = AH^{-1}A^T$
- [0027] $\Sigma^{-1} = A A^T$
- [0028] 其中， μ 和 Σ 为源发音人歌唱风格频谱合成模型的均值和方差， $\hat{\mu}$ 和 $\hat{\Sigma}$ 为转换后的频谱合成模型的均值和方差，即为目标发音人歌唱风格频谱合成模型的均值和方差， $[\Phi, B]$ 和 H 为均值和方差的全局转换矩阵， A 表示所述方差的分解矩阵。
- [0029] 一种构建个性化歌唱风格频谱合成模型的装置，包括：
- [0030] 模型训练模块，用于训练源发音人朗读风格频谱合成模型和歌唱风格频谱合成模型；
- [0031] 数据获取模块，用于获取目标发音人朗读风格数据；
- [0032] 转换关系确定模块，用于根据所述源发音人朗读风格频谱合成模型及所述目标发

音人朗读风格数据,确定源发音人与目标发明人的音色转换关系;

[0033] 转换模块,用于根据所述音色转换关系对所述源发音人歌唱风格频谱合成模型进行模型转换,得到目标发音人歌唱风格频谱合成模型。

[0034] 优选地,所述模型训练模块包括:

[0035] 数据获取单元,用于获取源发音人的朗读风格语音数据和歌唱风格语音数据,并对所述语音数据进行标注;

[0036] 参数提取单元,用于提取所述语音数据对应的频谱参数;

[0037] 第一训练单元,用于利用源发音人的朗读风格语音数据对应的频谱参数和标注信息训练源发音人的朗读风格频谱合成模型;

[0038] 第二训练单元,用于利用源发音人的歌唱风格语音数据对应的声学参数和标注信息训练源发音人的歌唱风格频谱合成模型。

[0039] 优选地,所述朗读风格语音数据标注的信息包括音素序列、音段切分、韵律词和韵律短语;所述歌唱风格语音数据的标注信息包括音素序列、音段切分、韵律词、韵律短语、以及与音乐相关的音高和时值信息。

[0040] 优选地,所述转换关系确定模块包括:

[0041] 自适应模型选择单元,用于从所述源发音人朗读风格频谱合成模型中选择执行自适应的模型;

[0042] 训练数据选择单元,用于从所述目标发音人的朗读风格数据中选择所述自适应模型所对应的训练数据;

[0043] 自适应单元,用于根据所述训练数据对选择执行自适应的模型进行自适应,得到频谱转换矩阵。

[0044] 优选地,所述转换模块,具体用于根据以下公式获得目标发音人歌唱风格频谱合成模型:

$$[0045] \quad \hat{\mu} = \Phi\mu + B$$

$$[0046] \quad \hat{\Sigma}^{-1} = AH^{-1}A^T$$

$$[0047] \quad \Sigma^{-1} = A A^T$$

[0048] 其中, μ 和 Σ 为源发音人歌唱风格频谱合成模型的均值和方差, $\hat{\mu}$ 和 $\hat{\Sigma}$ 为转换后的频谱合成模型的均值和方差,即为目标发音人歌唱风格频谱合成模型的均值和方差, $[\Phi, B]$ 和 H 为均值和方差的全局转换矩阵, A 表示所述方差的分解矩阵。

[0049] 本发明实施例提供的构建个性化歌唱风格频谱合成模型的方法及装置,借助于训练得到的源发音人朗读风格频谱合成模型和歌唱风格频谱合成模型,对于目标发音人,获取其少量朗读风格数据,训练源发音人朗读风格频谱合成模型与目标说话人朗读风格数据的全局转换矩阵,得到源发音人与目标发明人的音色转换关系,进而利用该音色转换关系对源发音人歌唱风格频谱合成模型进行模型转换,从而通过目标发音人少量朗读风格数据即可得到其歌唱风格合成模型,大大降低了一般用户构建歌唱合成模型的难度,进而有效地提高了个性化歌唱合成效果。

附图说明

[0050] 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。

[0051] 图 1 是本发明实施例构建个性化歌唱风格频谱合成模型的方法的流程图;

[0052] 图 2 是本发明实施例中二叉决策树的示意图;

[0053] 图 3 是本发明实施例构建个性化歌唱风格频谱合成模型的装置的结构示意图;

[0054] 图 4 是利用本发明实施例的频谱合成模型实现个性化歌唱合成的流程图。

具体实施方式

[0055] 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。

[0056] 考虑到基频与频谱不搭配会导致生成的语音不自然的问题,而且基于歌唱合成模型生成的频谱参数相比于基于朗读风格合成系统生成的频谱参数在歌唱语音合成上将更加自然,本发明实施例提供了一种构建个性化歌唱风格频谱合成模型的方法及系统,针对一般用户的歌唱数据,包括歌唱语音数据和歌唱标注数据较难获得的问题,通过源发音人的朗读风格频谱合成模型以及目标发音人的朗读风格数据,确定源发音人和目标发音人的音色转换关系;随后根据该转换关系对源发音人的歌唱风格频谱合成模型进行转换,获得目标发音人的歌唱风格频谱合成模型。从而大大降低了对一般用户构建个性化歌唱频谱合成模型的难度,解决了个性化歌唱合成系统中基频参数和频谱参数匹配的问题,进而提高了个性化歌唱合成效果。

[0057] 如图 1 所示,是本发明实施例构建个性化歌唱风格频谱合成模型的方法的流程图,包括以下步骤:

[0058] 步骤 101,训练源发音人朗读风格频谱合成模型和歌唱风格频谱合成模型。

[0059] 所述频谱合成模型包括频谱参数模型及对应的决策树。其中,朗读风格频谱合成模型和歌唱风格频谱合成模型是各自在对应的训练数据上训练得到的,其具体实现如下:

[0060] (1) 采集训练数据,获取源发音人朗读风格语音数据和歌唱风格语音数据,并进行相应标注。

[0061] 所述朗读风格语音数据标注的信息包括音素序列、音段切分、韵律词和韵律短语等信息。

[0062] 歌唱风格语音数据标注信息不仅包括音素序列、音段切分、韵律词和韵律短语等信息,还包括音乐相关的音高和时值信息等。

[0063] (2) 提取所述语音数据对应的频谱参数。

[0064] (3) 训练频谱合成模型,包括使用源发音人朗读风格语音数据对应的声学参数和相应的标注信息训练朗读风格频谱合成模型、以及使用源发音人歌唱风格语音数据对应的声学参数和相应的标注信息训练歌唱风格频谱合成模型。

[0065] 在进行模型训练时,可以采用二叉树节点分裂的方式,获取二叉决策树文件。具体地,首先设置仅包含根节点的二叉决策树并设置根节点为新的考察节点。随后依次挑选系统预设问题集中所有问题对当前考察节点的数据进行分裂尝试并选择具有最大样本分裂

度的问题为优选问题。接着根据分裂后样本聚集度和 / 或分裂后子节点中样本数据规模判断是否分裂。若分裂后样本聚集度下降幅度小于预设的阈值或子节点中样本数据量小于预设门限,则设置当前考察节点为叶节点不再分裂;否则按照最优问题对其分裂,并获取子节点和对应的样本数据。所述样本聚集度即样本分散程度。一般来说分散度越大,则说明该节点分裂的可能性越大,否则分裂的可能性越小。具体地,可以采用样本方差来衡量节点聚集度,即计算节点包含的所有样本数据距离该类中心的距离(或距离平方)的平均值。通过该方法获得二叉树文件,并确定各叶节点对应的训练数据。

[0066] 利用各叶节点的训练数据进行频谱参数训练,即可得到各叶节点所对应的频谱合成模型。

[0067] 在采用基于 HMM 模型的参数合成模型时,可以对 HMM 模型各状态分别建模,图 2 展示了合成单元的各状态(2,3,4,5,6 状态)分别对应的合成参数模型。

[0068] 步骤 102,获取目标发音人朗读风格数据。

[0069] 具体地,可以采取现场录制的方式或直接使用目标发音人现有的语音数据,提取语音数据对应频谱参数等;对所述语音数据进行相关标注,通过自动识别方法或人工标注方法得到对应文本信息;然后通过对文本进行音素序列、韵律词和韵律短语等信息的自动分析,得到对应各基本合成单元的目标发音人朗读风格数据。

[0070] 步骤 103,根据所述源发音人朗读风格频谱合成模型及所述目标发音人朗读风格数据,确定源发音人与目标发音人的音色转换关系。

[0071] 在本发明实施例中,使用目标发音人的训练数据,即步骤 102 中获取的目标发音人的朗读风格数据,对源发音人朗读风格的频谱合成模型进行全局自适应训练,得到全局转换矩阵,确定源发音人与目标发音人的音色转换关系。

[0072] 自适应训练的具体过程如下:

[0073] (1) 确定执行自适应的模型。

[0074] 可以选择执行自适应的模型为源发音人朗读风格频谱合成模型中的所有模型,也可以选择为 HMM 模型的每个状态。考虑到在对源发音人的朗读风格合成模型进行自适应时,自适应数据往往比较少,因此可以优选对所有模型集合进行基于 MLLR(语音识别最大似然线性回归)的全局自适应。

[0075] (2) 从所述目标发音人的朗读风格数据中选择所述自适应模型所对应的训练数据。

[0076] (3) 根据所述训练数据对所述模型进行自适应,获取对应的频谱转换矩阵。

[0077] 步骤 104,根据所述音色转换关系对所述源发音人歌唱风格频谱合成模型进行模型转换,得到目标发音人歌唱风格频谱合成模型。

[0078] 可以将模型转换关系表示为: $y = Ax + b$;

[0079] 其中, x 为源发音人歌唱风格频谱合成模型, y 为转换后歌唱风格频谱合成模型,即目标发音人歌唱风格频谱合成模型, A 和 b 为通过全局自适应训练得到的转换参数。

[0080] 具体到本发明实施例,在频谱合成模型高斯拓扑设定下,其模型更新为:

$$[0081] \quad \hat{\mu} = \Phi\mu + B$$

$$[0082] \quad \hat{\Sigma}^{-1} = AH^{-1}A^T$$

[0083] $\Sigma^{-1}=A A^T$

[0084] 其中, μ 和 Σ 为源发音人歌唱风格频谱合成模型的均值和方差, $\hat{\mu}$ 和 $\hat{\Sigma}$ 为转换后的频谱合成模型的均值和方差, 即为目标发音人歌唱风格频谱合成模型的均值和方差, $[\Phi, B]$ 和 H 为均值和方差的全局转换矩阵, A 表示所述方差的分解矩阵。

[0085] 本发明实施例构建个性化歌唱风格频谱合成模型的方法, 由于只需获取目标发音人少量朗读风格语音数据, 即可得到其歌唱风格频谱合成模型, 大大降低了一般用户构建歌唱频谱合成模型的难度。

[0086] 相应地, 本发明实施例还提供一种构建个性化歌唱风格频谱合成模型的装置, 如图 3 所述, 是该装置的一种结构示意图。

[0087] 在该实施例中, 所述装置包括:

[0088] 模型训练模块 301, 用于训练源发音人朗读风格频谱合成模型和歌唱风格频谱合成模型;

[0089] 数据获取模块 302, 用于获取目标发音人朗读风格数据;

[0090] 转换关系确定模块 303, 用于根据所述源发音人朗读风格频谱合成模型及所述目标发音人朗读风格数据, 确定源发音人与目标发明人的音色转换关系;

[0091] 转换模块 304, 用于根据所述音色转换关系对所述源发音人歌唱风格频谱合成模型进行模型转换, 得到目标发音人歌唱风格频谱合成模型。

[0092] 所述频谱合成模型包括频谱参数模型及对应的决策树。其中, 朗读风格频谱合成模型和歌唱风格频谱合成模型是各自在对应的训练数据上训练得到的, 模型训练模块 301 的一种具体实现结构包括:

[0093] 数据获取单元, 用于获取源发音人的朗读风格语音数据和歌唱风格语音数据, 并对所述语音数据进行标注。其中, 朗读风格语音数据标注的信息包括音素序列、音段切分、韵律词和韵律短语等信息。歌唱风格语音数据标注信息不仅包括音素序列、音段切分、韵律词和韵律短语等信息, 还包括音乐相关的音高和时值信息等。

[0094] 参数提取单元, 用于提取所述语音数据对应的频谱参数。

[0095] 第一训练单元, 用于利用源发音人的朗读风格语音数据对应的频谱参数和标注信息训练源发音人的朗读风格频谱合成模型。

[0096] 第二训练单元, 用于利用源发音人的歌唱风格语音数据对应的声学参数和标注信息训练源发音人的歌唱风格频谱合成模型。

[0097] 需要说明的是, 上述第一训练单元和第二训练单元在进行模型训练时, 可以采用二叉树节点分裂的方式, 获取二叉决策树文件。利用二叉决策树中各叶节点的训练数据进行频谱参数训练, 即可得到各叶节点所对应的频谱合成模型, 具体过程可参照前面本发明方法实施例中的描述, 在此不再赘述。另外, 在实际应用中, 上述第一训练单元和第二训练单元可以采用同一物理实体, 也可以采用不同的物理实体。

[0098] 数据获取模块 302 具体可以采取现场录制的方式或直接获取目标发音人现有的语音数据, 提取语音数据对应频谱参数等; 对所述语音数据进行相关标注, 通过自动识别方法或人工标注方法得到对应文本信息; 然后通过对文本进行音素序列、韵律词和韵律短语等信息的自动分析, 得到对应各基本合成单元的目标发音人朗读风格数据。

[0099] 转换关系确定模块 303 具体可以使用目标发音人的朗读风格数据,对源发音人朗读风格的频谱合成模型进行全局自适应训练,得到全局转换矩阵,确定源发音人与目标发音人的音色转换关系。

[0100] 转换关系确定模块 303 可以包括:自适应模型选择单元、训练数据选择单元和自适应单元。其中:

[0101] 所述自适应模型选择单元,用于从所述源发音人朗读风格频谱合成模型中选择执行自适应的模型。具体可以选择执行自适应的模型为源发音人朗读风格频谱合成模型中的所有模型,也可以选择为 HMM 模型的每个状态。考虑到在对源发音人的朗读风格合成模型进行自适应时,自适应数据往往比较少,因此可以优选对所有模型集合进行基于 MLLR 的全局自适应。

[0102] 所述训练数据选择单元,用于从所述目标发音人的朗读风格数据中选择所述自适应模型所对应的训练数据。

[0103] 所述自适应单元,用于根据所述训练数据对选择执行自适应的模型进行自适应,得到频谱转换矩阵。

[0104] 上述转换模块 304 在频谱合成模型高斯拓扑设定下,其模型更新的具体方式可参照前面本发明方法实施例中的描述,在此不再赘述。

[0105] 需要说明的是,本发明实施例提供的构建个性化歌唱风格频谱合成模型的装置同样适用于个性化歌唱风格时长合成模型和基频合成模型的构建,使合成参数更加匹配,合成效果也可以得到进一步的提升。

[0106] 本发明实施例构建个性化歌唱风格频谱合成模型的方法及系统,由于只需获取目标发音人少量朗读风格语音数据,即可得到其歌唱风格频谱合成模型,大大降低了一般用户构建歌唱频谱合成模型的难度。将利用本发明实施例的装置构建的歌唱风格频谱合成模型应用于个性化歌唱合成系统中,可以有效提高个性化歌唱合成的效果。

[0107] 如图 4 所示,是利用本发明实施例的频谱合成模型实现个性化歌唱合成的流程图,包括以下步骤:

[0108] 步骤 401,预先训练目标发音人歌唱风格频谱合成模型、目标发音人歌唱时长合成模型、以及目标发音人歌唱基频合成模型。

[0109] 目标发音人歌唱风格频谱合成模型的具体训练过程可参照图 1 及其相应的文字描述,在此不再赘述。

[0110] 训练目标发音人歌唱时长合成模型的方式具体可以采用传统方法实现,即利用源发音人歌唱风格时长合成模型,也可以根据简谱标注获得,当然也可以采用与本发明目标发音人歌唱风格频谱合成模型构建相类似的方法,对源发音人歌唱时长合成模型优化得到,对此本发明实施例不做限定。

[0111] 需要说明的是,在实际应用中,训练目标发音人歌唱基频合成模型的方式可以采用传统方法,即利用源发音人歌唱基频合成模型,也可以采用与本发明目标发音人歌唱风格频谱合成模型构建相类似的方法,对源发音人歌唱基频合成模型优化得到。当然,也可以直接使用简谱中的基频而省略目标发音人歌唱时长合成模型的训练过程。

[0112] 步骤 402,接收待合成歌唱简谱。

[0113] 步骤 403,根据所述简谱标注音节时长及歌唱时长合成模型,确定简谱歌词中各基

本合成单元的时长合成参数。

[0114] 需要注意的是,各基本合成单元的时长参数获取方法可以有多种,对此本发明实施例不做限定。

[0115] 步骤 404,根据所述各基本合成单元的时长合成参数及目标发音人歌唱风格频谱合成模型确定各基本合成单元的频谱合成参数。

[0116] 步骤 405,根据所述各基本合成单元的时长合成参数及目标发音人的歌唱基频合成模型确定各基本合成单元的基频合成参数。

[0117] 基频合成参数的确定可以采用传统任意方法实现,对此本发明实施例不做限定。比如,可以直接根据简谱标注的基调及音高值确定各基本合成单元的基频合成参数,也可以根据所述基频合成参数模型获取各基本合成单元的基频合成参数。

[0118] 步骤 406,根据所述频谱合成参数、基频合成参数以及时长合成参数合成歌唱语音。

[0119] 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。

[0120] 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。

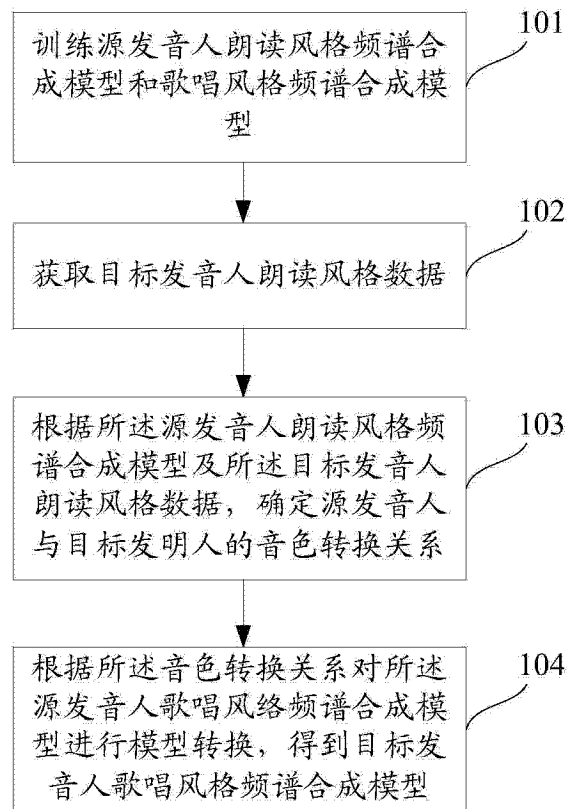


图 1

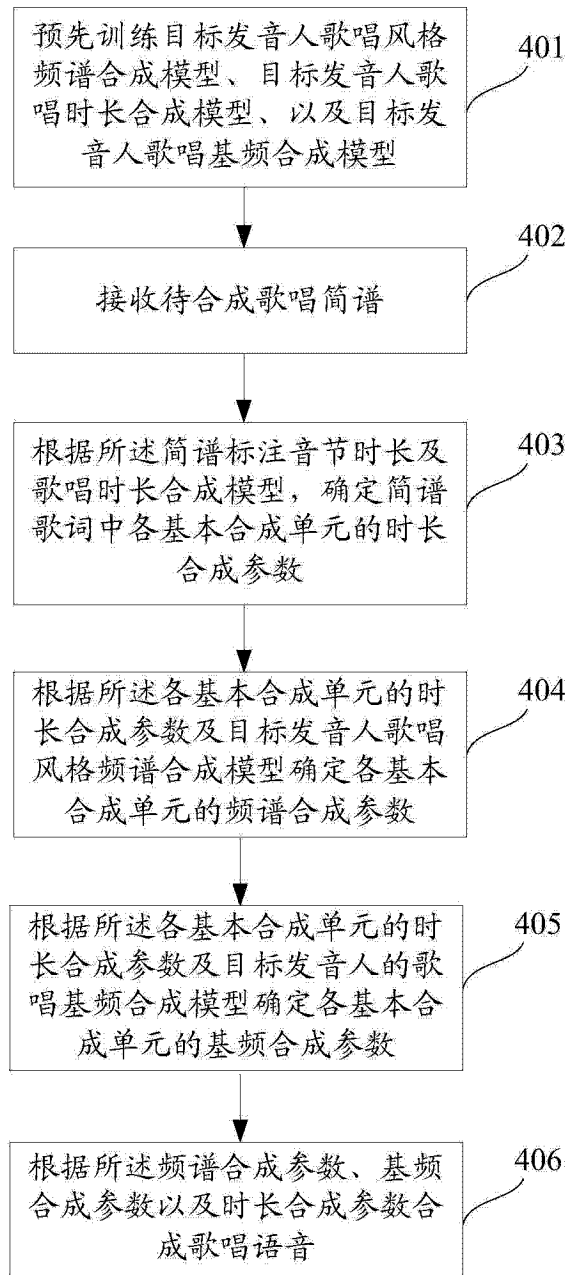


图 4