(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 1815552 B
(45)授权公告日 2010.05.12

(21)申请号 200610038589.X
(22)申请日 2006.02.28
(73)专利权人 安徽中科大讯飞信息科技有限公司
地址 230088 安徽省合肥市黄山路616号
(72)发明人 凌震华 王玉平 王仁华
(74)专利代理机构 安徽合肥华信知识产权代理有限公司 34112
代理人 余成俊
(51)Int.Cl.
G10L 13/08(2006.01)
G10L 13/02(2006.01)
G10L 13/00(2006.01)
G10L 21/02(2006.01)

(56)对比文件
US 6205423 B1,2001.03.20, 全文.
CN 1667703 A,2005.09.14, 全文.
EP 0852376 A2,1998.07.08, 全文.

审查员 文娟

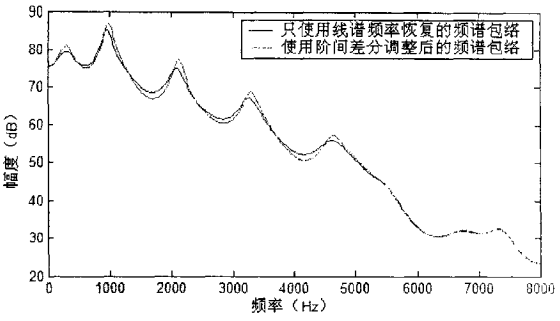
权利要求书 1 页 说明书 4 页 附图 1 页

(54)发明名称

基于线谱频率及其阶间差分参数的频谱建模与语音增强方法

(57)摘要

本发明公开了基于线谱频率及其阶间差分参数的频谱建模与语音增强方法,包括在频谱参数提取时将线谱频率阶间差分作为提取结果的一部分;在模型建模和训练时分别对线谱频率及其阶间差分参数进行独立建模和训练;在预测时分别预测线谱频率及其阶间差分参数,并利用阶间差分对线谱频率参数进行调整;最终利用调整后的线谱频率参数合成输出语音以达到通过增强和锐化合成语音的共振峰而提高合成语音音质的目的。



CN 1815552 B

1. 基于线谱频率及其阶间差分参数的频谱建模与语音增强方法，包括以下步骤：

(1)、对语音信号分帧求取线性预测系数；

(2)、线谱频率及其阶间差分参数的获得：将线性预测系数转换成对应阶数的线谱频率参数，同时，对相邻阶的线谱频率计算其差分参数；

(3)、对于各阶线谱频率及其阶间差分参数分别独立进行声学模型的训练，采用的模型为隐马尔可夫模型，在模型训练过程中，通过结合语音单元的上下文属性利用决策树对各参数对应的模型进行较为细致的分类，保证得到的声学模型均可以实现依据上下文属性输入的参数预测；

(4)、合成阶段的语音增强处理：

a、对用户输入的文本进行分析，利用分析得到的各语音单元对应的上下文属性输入训练得到的声学模型，预测合成时使用的各帧线谱频率及阶间差分参数，由于线谱频率和阶间差分参数是分别通过二个独立的声学模型预测的，所以预测得到的阶间差分参数与预测得到的线谱频率的实际阶间差分参数相比并不一致；

b、利用各帧预测得到的阶间差分参数依据下式对预测得到的线谱频率进行调整：

$$l_i' = l_{i-1} + c_{i-1} + \frac{c_{i-1}^2}{c_{i-1}^2 + c_i^2}[(l_{i+1} - l_{i-1}) - (c_i + c_{i-1})]$$

其中，l_i，i = 1,2,...,N 为预测得到的当前帧第 i 阶的线谱频率，N 为线谱频率参数的阶数；c_i，i = 1,2,...,M 为预测得到的当前帧第 i+1 阶和第 i 阶线谱频率之间的阶间差分参数，M 为阶间差分参数的阶数，M < N；l'_i，i = 1,2,...,N 为调整后当前帧的 N 阶线谱频率，对于各阶线谱频率，选择从低阶第 2 阶到高阶第 M 阶的调整顺序，或者选择从高阶第 M 阶到低阶第 2 阶的调整顺序，或者同时通过调整遍数来控制这种频谱峰值增强作用的强弱；

c、将调整后的线谱频率转换为线性预测系数，同时结合韵律预测模块生成的基音频率参数，送入线性预测滤波器，合成语音并输出。

2. 根据权利要求 1 所述的方法，其特征在于对语音信号分帧求取线性预测系数是：通过固定帧移加窗乘取的方法获得各帧语音的短时信号波形，然后求取该帧信号对应的各阶线性预测系数，求取方法为基于时域波形自相关系数的线性预测系数求取方法；或者自适应加权谱内插的方法，首先计算该帧语音对应的频谱包络，再利用全极点模型拟合求解线性预测系数。

3. 根据权利要求 1 所述的方法，其特征在于线谱频率及其阶间差分参数的获得过程中，选择保留所有的阶间差分参数，或者为了降低参数维数选择仅保留较低阶的阶间差分参数。

4. 根据权利要求 2 所述的方法，其特征在于所述加窗是指高斯窗，窗宽为基音周期长度的两倍，帧移 5 毫秒。

基于线谱频率及其阶间差分参数的频谱建模与语音增强方法

技术领域

[0001] 本发明涉及语音合成方法，具体是在基于线谱频率的语音频谱参数化与建模过程中加入对其阶间差分参数的考虑，通过对线谱频率阶间差分参数的合理利用达到对合成语音共振峰的增强的目的，提高合成语音清晰度。

背景技术

[0002] 现有的语音合成技术主要有基于波形拼接的语音合成方法和基于参数合成的语音合成方法两大类。前者通过利用包含自然声学样本的语音音库和在合成时进行单元选择的方法可以取得较高的合成语音的音质与自然度。但是由于语音音库的使用，往往在存储量上有比较大的消耗，难以实现在嵌入式平台等资源受限领域的使用。

[0003] 另一种基于参数合成的语音合成方法首先需要对语音信号进行参数化分析，一般包括表征激励信息的基音频率参数和表征声道滤波器频谱特征的频谱参数，然后对分析得到的参数进行建模，在合成时利用模型进行相关声学参数的预测，最终通过参数合成器还原语音信号。这种方法同样能够取得较好的合成语音的流畅度和自然度，并且由于在合成阶段脱离的音库的限制，消耗存储资源很小。但是由于在对参数的建模过程中，往往会引入一定的平均化处理，这样使得模型预测输出的频谱参数对应的频谱包络过于平滑，共振峰被削弱，从而造成合成语音清晰度的下降。

发明内容

[0004] 本发明的目的就是为了提供一种语音合成系统中基于线谱频率及其阶间差分参数的频谱建模与语音增强方法，以达到提高合成语音效果的目的。

[0005] 本发明的技术方案如下：

[0006] 基于线谱频率及其阶间差分参数的频谱建模与语音增强方法，其特征在于包括以下步骤：

[0007] (1)、对语音信号分帧求取线性预测系数；

[0008] (2)、线谱频率及其阶间差分参数的获得：将线性预测系数转换成对应阶数的线谱频率参数，同时，对相邻阶的线谱频率计算其差分参数；

[0009] (3)、对于各阶线谱频率及其阶间差分参数分别独立进行声学模型的训练，采用的模型为隐马尔可夫模型，在模型训练过程中，通过结合语音单元的上下文属性利用决策树对各参数对应的模型进行较为细致的分类，保证得到的声学模型均可以实现依据上下文属性输入的参数预测；

[0010] (4)、合成阶段的语音增强处理：

[0011] d、对用户输入的文本进行分析，利用分析得到的各语音单元对应的上下文属性输入训练得到的声学模型，预测合成时使用的各帧线谱频率及阶间差分参数，由于线谱频率和阶间差分参数是分别通过二个独立的声学模型预测的，所以预测得到的阶间差分参数与

预测得到的线谱频率的实际阶间差分参数相比并不一致；

[0012] e、利用各帧预测得到的阶间差分参数依据下式对预测得到的线谱频率进行调整：

[0013] $$l_i' = l_{i-1} + c_{i-1} + \frac{c_{i-1}^2}{c_{i-1}^2 + c_i^2}[(l_{i+1} - l_{i-1}) - (c_i + c_{i-1})]$$

[0014] 其中，l_i，i = 1,2,...,N 为预测得到的当前帧第 i 阶的线谱频率，N 为线谱频率参数的阶数；c_i，i = 1,2,...,M 为预测得到的当前帧第 i+1 阶和第 i 阶线谱频率之间的阶间差分参数，M 为阶间差分参数的阶数，M < N；l'_i，i = 1,2,...,N 为调整后当前帧的 N 阶线谱频率。对于各阶线谱频率，可以选择从低阶（第 2 阶）到高阶（第 M 阶）的调整顺序，也可以选择从高阶（第 M 阶）到低阶（第 2 阶）的调整顺序，同时可以通过调整遍数来控制这种频谱峰值增强作用的强弱；

[0015] f、将调整后的线谱频率转换为线性预测系数，同时结合韵律预测模块生成的基音频率参数，送入线性预测滤波器，合成语音并输出。

[0016] 对语音信号分帧求取线性预测系数是：通过固定帧移加窗乘取的方法获得各帧语音的短时信号波形，然后求取该帧信号对应的各阶线性预测系数，求取方法为基于时域波形自相关系数的线性预测系数求取方法；或者自适应加权谱内插的方法，首先计算该帧语音对应的频谱包络，再利用全极点模型拟合求解线性预测系数。

[0017] 线谱频率及其阶间差分参数的获得过程中，选择保留所有的阶间差分参数，或者为了降低参数维数选择仅保留较低阶的阶间差分参数。

[0018] 所述加窗是指高斯窗，窗宽为基音周期长度的两倍，帧移 5 毫秒，

[0019] 这里提出的在语音合成系统中基于线谱频率及其阶间差分参数的频谱建模与语音增强方法就是为了提高参数合成方法的语音清晰度，主要基于以下几点考虑：

[0020] (1) 线谱频率参数相对于线性预测系数更加稳定，相对于倒谱系数更加能够反映与频谱峰值相关的一些频谱局部特征，相对于共振峰参数在求解上更加容易与鲁棒；

[0021] (2) 线谱频率对于频谱局部特征的反映，主要是通过其相邻阶差分表现出来的，线谱频率具有 0 ～ π 的顺序排列特征，当两个线谱频率比较接近，即阶间差分较小时，会在频谱包络对应频率处形成一个峰，差分越小，峰值越尖锐，反之，频谱越平坦。

[0022] 通过观察合成语音的频谱可以发现，在使用基于线谱频率及其阶间差分参数的频谱建模与语音增强方法后，对比只使用线谱频率参数，频谱中的共振峰部分得到了有效的锐化和增强。

[0023] 通过对合成语音的实际测听表明，使用该方法后，对比只使用线谱频率参数，合成语音的清晰度得到明显提高，更容易被使用者接受。

[0024] 同时，对比其他的语音增强算法，由于该方法只是对各帧的频谱参数进行了调整，而没有引入后滤波等额外处理，所以对与整个合成系统不会增加运算量的消耗。

[0025] 术语解释

[0026] 语音合成 (Text-To-Speech)：又称为文语转化。它涉及声学、语言学、数字信号处理、多媒体等多种学科，是中文信息处理领域的一项前沿技术。语音合成技术解决的主要问题是：如何将电子化文本的文字信息转化为能够播放的声音信息。近代语音合成技术是随着计算机技术和数字信号处理技术的发展而发展起来的，目的是让计算机能够产生高清晰

度、高自然度的连续语音。

[0027] 线性预测系数(Linear Prediction Coefficient):线性预测分析从人的发声机理入手,通过对声道的短管级联模型的研究,认为系统的传递函数符合全极点数字滤波器的形式,从而当前时刻的信号可以用前若干时刻的信号的线性组合来估计,通过使实际语音的采样值和线性预测采样值之间达到均方差最小,即可得到线性预测系数。

[0028] 线谱频率(Linear Spectral Frequency):线谱频率是一种和线性预测系数等价的声道模型描述参数,具有 0 ～ π 的顺序分布特征,可以依据线性预测系数求解获得。

[0029] 自适应加权谱内插(Speech Transformation and Representation usingAdaptive Interpolation of weiGHTed spectrum,STRAIGHT):一种针对语音信号的分析合成算法,它通过对语音短时谱进行时频域的自适应内插平滑来提取精确的谱包络。

[0030] 隐马尔可夫模型(Hidden Markov Model):马尔可夫模型的概念是一个离散时域有限状态自动机,隐马尔可夫模型是指这一马尔可夫模型的内部状态外界不可见,外界只能看到各个时刻的输出值。用隐马尔可夫刻画语音信号需作出两个假设,一是内部状态的转移只与上一状态有关,另一是输出值只与当前状态(或当前的状态转移)有关,这两个假设大大降低了模型的复杂度。

附图说明

[0031] 图 1 :利用预测得到阶间差分参数对线谱频率调整后合成语音频谱的增强情况示例

[0032] 图 2 :本发明模型训练阶段流程图。

[0033] 图 3 :本发明合成阶段流程图。

具体实施方式

[0034] 本发明具体的实现方式如下 :

[0035] 1. 训练语音数据的频谱参数化分析

[0036] 1) 对语音信号分帧求取线性预测系数 :通过固定帧移加窗乘取(高斯窗,窗宽为基音周期长度的两倍,帧移 5 毫秒)的方法获得各帧语音的短时信号波形,然后求取该帧信号对应的各阶线性预测系数。求取方法可以采用基于时域波形自相关系数的线性预测系数求取方法 ;也可以采用自适应加权谱内插的方法,首先计算该帧语音对应的频谱包络,再利用全极点模型拟合求解线性预测系数。计算时,可以根据语音信号采样率的不同而对参数阶数进行不同的设定 ;

[0037] 2) 线谱频率及其阶间差分参数的获得 :将线性预测系数转换成对应阶数的线谱频率参数,同时,对相邻阶的线谱频率计算其差分值(差分参数),作为频谱参数提取结果的一部分,可以选择保留所有的阶间差分参数,也可以为了降低参数维数选择只保留较低阶的阶间差分参数,因为人耳对于语音低频区域更加敏感。本

[0038] 2. 对于各阶线谱频率及其阶间差分参数分别进行声学模型的训练,采用的模型为隐马尔可夫模型(Hidden Markov Model,HMM),在模型训练过程中,通过结合语音单元的上下文属性利用决策树对各参数对应的模型进行较为细致的分类,保证得到的声学模型可以实现依据上下文属性输入的参数预测 ;

[0039] 3. 合成阶段的语音增强处理

[0040] 1) 对用户输入的文本进行分析，利用分析得到的各语音单元对应的上下文属性输入训练得到的参数模型，预测合成时使用的各帧线谱频率及阶间差分参数，由于线谱频率和阶间差分参数是分别独立建模与预测的，所以预测得到的阶间差分参数与预测得到的线谱频率的实际阶间差分参数相比并不一致；

[0041] 2) 利用各帧预测得到的阶间差分参数依据下式对线谱频率进行调整：

[0042] $$l_i' = l_{i-1} + c_{i-1} + \frac{c_{i-1}^2}{c_{i-1}^2 + c_i^2}[(l_{i+1} - l_{i-1}) - (c_i + c_{i-1})]$$

[0043] 其中，l_i，i = 1,2,...,N 为预测得到的当前帧第 i 阶的线谱频率，N 为线谱频率参数的阶数；c_i，i = 1,2,...,M 为预测得到的当前帧第 i+1 阶和第 i 阶线谱频率之间的阶间差分参数，M 为阶间差分参数的阶数，M < N；l'_i，i = 1,2,...,N 为调整后当前帧的 N 阶线谱频率。对于各阶线谱频率，可以选择从低阶（第 2 阶）到高阶（第 M 阶）的调整顺序，也可以选择从高阶（第 M 阶）到低阶（第 2 阶）的调整顺序，同时可以通过调整遍数来控制这种频谱峰值增强作用的强弱。

[0044] 3) 将调整后的线谱频率转换为线性预测系数，同时结合韵律预测模块生成的基音频率参数，送入线性预测滤波器，合成语音并输出。

[0045] 图 1：利用预测得到阶间差分对线谱频率调整后对应合成语音频谱的变化情况，以上为一帧合成语音 /a/ 所对应的幅度谱，采样率为 16kHz，线谱频率阶数为 24，使用的阶间差分参数阶数为 16，调整方法为由低阶向高阶调整一遍。

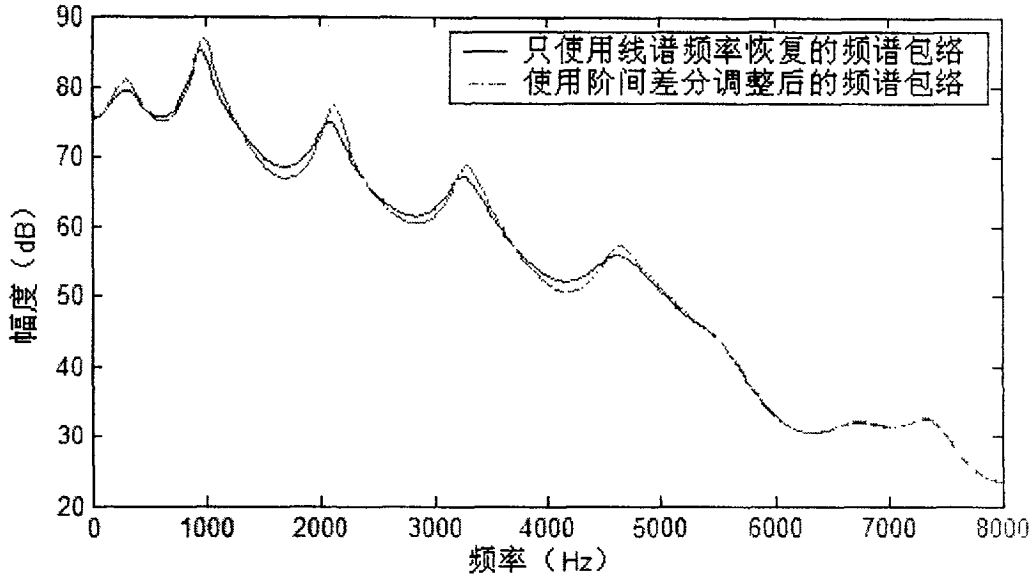

只使用线谱频率恢复的频谱包络
使用阶间差分调整后的频谱包络
幅度（dB）
频率（Hz）
90
80
70
60
50
40
30
20
0
1000
2000
3000
4000
5000
6000
7000
8000

图 1

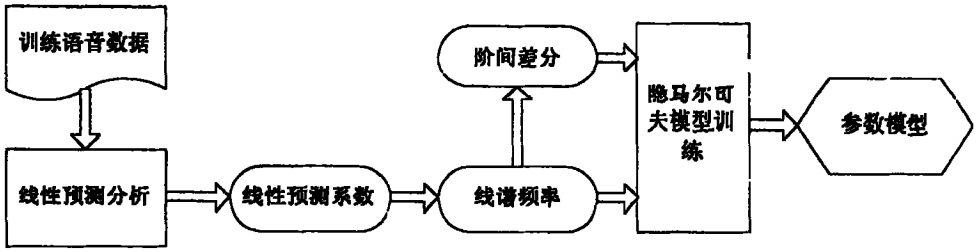

训练语音数据
线性预测分析
线性预测系数
线谱频率
阶间差分
隐马尔可夫模型训练
参数模型

图 2

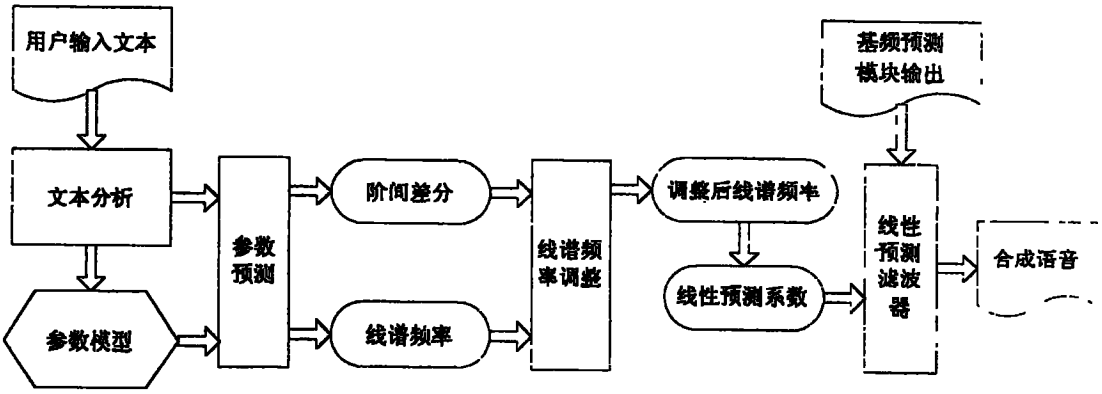

用户输入文本
文本分析
参数模型
参数预测
阶间差分
线谱频率
线谱频率调整
调整后线谱频率
线性预测系数
基频预测模块输出
线性预测滤波器
合成语音

图 3