



(12)发明专利申请

(10)申请公布号 CN 107293289 A

(43)申请公布日 2017. 10. 24

(21)申请号 201710442945.2

(22)申请日 2017.06.13

(71)申请人 南京医科大学

地址 211166 江苏省南京市江宁区龙眠大道101号

(72)发明人 王伟 王翰林 胡克 魏天远
张璐瑶 高珊 符凡 刘政 朱纯

(74)专利代理机构 南京苏高专利商标事务所
(普通合伙) 32204

代理人 肖明芳

(51)Int.Cl.

G10L 15/06(2013.01)

G10L 15/18(2013.01)

G10L 15/22(2006.01)

G10L 21/0208(2013.01)

权利要求书3页 说明书6页 附图2页

(54)发明名称

一种基于深度卷积生成对抗网络的语音生成方法

(57)摘要

本发明公开了一种基于深度卷积生成对抗网络的语音生成方法,其步骤为:(1)采集语音信号样本;(2)对语音信号样本进行预处理;(3)将语音信号样本输入深度卷积生成对抗网络;(4)对输入的语音信号进行训练;(5)生成接近真实语音信号的语音信号。本发明使用tensorflow作为学习框架,利用深度卷积生成对抗网络算法对大量语音信号进行训练,借助深度卷积生成对抗网络里判别网络D与生成网络G的动态博弈过程,最终生成接近原始学习内容的自然语音信号。本发明基于深度卷积生成对抗网络下生成的语音,可以解决在人机面对面交流过程中智能设备过度依赖固定的语音库来发声,模式单调缺乏变化且不够自然的问题。



1. 一种基于深度卷积生成对抗网络的语音生成方法,其特征在于,包括如下步骤:

步骤1,采集语音信号样本:随机采集M个具有相同内容的语音信号,作为语音训练样本和真实语音样本;

步骤2,对步骤1中采集到的M个语音信号进行预处理;

步骤3,将预处理后的M个语音训练样本数据和M个真实语音样本数据输入到深度卷积生成对抗网络;

步骤4,采用深度卷积生成对抗网络对输入的M个语音训练样本数据和M个真实语音样本数据进行训练;

步骤5,生成接近真实语音内容的语音信号:利用深度卷积对抗生成网络对自回归生成模型得到的波形进行训练,最终生成全新的且接近真实语音样本的语音信号。

2. 根据权利要求1所述的方法,其特征在于:步骤1中所述随机采集M个具有相同内容的语音信号,其存储格式为wav格式。

3. 根据权利要求1所述的方法,其特征在于,步骤2包括如下步骤:

步骤2-1,对采集的语音信号样本做除杂的处理;

步骤2-2,对采集的语音信号样本进行滤波处理。

4. 根据权利要求3所述的方法,其特征在于:步骤2-1包括:利用Audacity软件对采集的语音信号样本进行剪辑,过滤掉采集的语音信号样本中超出软件编辑范围的语音部分以及非语音信号的部分。

5. 根据权利要求4所述的方法,其特征在于,步骤2-2包括如下步骤:

步骤2-2-1,计算n时刻的误差信号 $\varepsilon(n)$:

$$\varepsilon(n) = d(n) - \hat{W}^H(n)X(n),$$

其中, $d(n)$ 表示n时刻主信道输入的带噪语音信号,即为自适应滤波器的期望信号,主信道表示语音输入的通道, $W(n)$ 表示n时刻对应的权重系数矢量, $X(n)$ 表示在n时刻的语音矢量, H 表示共轭转置;

步骤2-2-2,计算在n时刻的相关系数 $\gamma(n)$:

$$\gamma(n) = \frac{X^H(n)X(n-1)}{X^H(n-1)X(n-1)},$$

步骤2-2-3,计算n时刻的语音差异矢量 $U(n)$:

$$U(n) = X(n) - \gamma(n)X(n-1);$$

步骤2-2-4,通过如下公式计算迭代后每个时刻对应的权重系数矢量:

$$\hat{W}(n+1) = \hat{W}(n) + \frac{\mu}{\|U(n)\|^2 + \delta} U(n)\varepsilon(n),$$

其中, μ 表示自适应常数, $\delta > 0$, δ 表示实数;

步骤2-2-5,计算最小方差 $\min_w E\{\varepsilon^2\}$:

$$\min_w E\{\varepsilon^2\} = E\{v^2\} + \min_w E\{(n_0 - y)^2\},$$

其中, n_0 表示主信道输入的语音信号, v 表示主信道输入的噪声, y 表示自适应滤波器的输出, ε 表示误差信号, w 表示特定时刻自适应滤波器的权重矢量,使得方差 $E\{\varepsilon^2\}$ 最小;

步骤2-2-6,将步骤2-2-4中得到的每个时刻的对应的权重系数矢量分别带入步骤2-2-5的公式中,计算最小方差 $\min_w E\{\varepsilon^2\}$,若该时刻求得的不是最小方差,则通过自适应滤波器调整其下一时刻权重矢量 w 进行滤波处理,求出最小方差,得到滤波后的语音信号。

6. 根据权利要求5所述的方法,其特征在于,步骤4包括如下步骤:

步骤4-1,深度卷积生成对抗网络包括两个网络,一个是生成网络 G ,用于接收一个随机的噪声;一个是判别网络 D ,用于判别生成的数据是不是真实的;

步骤4-2,计算生成网络的损失函数:

$$(1-b) \log (1-D(G(z))),$$

其中, z 表示生成网络 G 接收的一个随机的噪声, b 表示在真实语音样本数据输入参数 c 下得到的输出参数, $G(z)$ 表示生成网络 G 的输出, \log 表示以10为底的对数操作, $D(G(z))$ 表示判别网络 D 判断生成网络 G 生成的语音数据为真实的概率;

步骤4-3,计算判别网络的损失函数:

$$-((1-b) \log (1-D(G(z))) + b \log D(c)),$$

其中, c 表示真实语音样本数据作为输入参数, $D(c)$ 表示判别网络 D 的输出,即输入真实语音样本数据参数 c 为真实的概率;

步骤4-4,计算优化函数 $\min \max V(D, G)$:

$$\min \max V(D, G) = E_{x \sim p_{data}(x)} [\log(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))],$$

其中, $p_z(z)$ 表示随机噪声的概率密度, x 表示真实语音样本数据, $p_{data}(x)$ 表示参数数据的概率密度;

步骤4-5,当判别网络 D 无法判定生成网络 G 所生成的语音数据是否真实时, $D(G(z)) = 0.5$,得到训练好的语音信号数据,否则,执行步骤4-6;

步骤4-6,利用随机梯度下降法计算判别网络 D 的梯度函数和生成网络 G 的梯度函数:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(c^{(i)}) + \log(1 - D(G(z^{(i)})))] ,$$

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^{(i)}))) ,$$

其中, θ_d 表示在判别网络 D 方向上的梯度变化量, θ_g 表示在生成网络 G 方向上的梯度变化量, $i \in [1, m]$, m 表示生成网络 G 接收随机噪声的个数, \in 表示属于符号,将训练好的判别网络 D 和生成网络 G 带入步骤4-4中。

7. 根据权利要求6所述的方法,其特征在于,步骤5包括如下步骤:

步骤5-1,计算自回归生成模型:

$$p(s|h) = \prod_{t=1}^T p(s_t | s_1, \dots, s_{t-1}, h),$$

其中, h 表示步骤4中得到训练好的语音样本数据作为输入的参数, $p(s|h)$ 表示最有可行的波形输出向量, $s(t)$ 表示在时间点 t 的输出, $t \in [1, T]$, T 表示总时间;

步骤5-2,采用步骤4中深度卷积生成对抗网络,对步骤5-1中生成的波形进行学习,再

经过自回归生成模型不断的反馈及输出,使得每一步当前输出的结果,只与之前的输出结果相关,当满足步骤4-5中 $D(G(z))=0.5$ 时,最终生成全新的且接近真实语音样本的语音信号。

一种基于深度卷积生成对抗网络的语音生成方法

技术领域

[0001] 本发明属于语音生成技术领域,尤其涉及一种基于深度卷积生成对抗网络的语音生成方法。

背景技术

[0002] 人机交互技术的研究是计算机技术研究领域的重要组成部分。使智能设备具有“说话”的功能,这在真正的“面对面人机交流”中扮演着很重要的角色。借助于语音生成系统,智能设备已经可以清晰、自然地说话,普通用户很容易听懂并接受。语音模仿作为人机语音交流的重要环节,一方面,需要在前期建立大量的语音库,另一方面,需要对大量语音信号的特征提取及训练最终生成接近原始学习内容的自然语音信号。

[0003] 生成式对抗网络,是一种近年来大热的深度学习模型,其主要就是可以用tensorflow作为学习框架,训练一个生成器G,从随机噪声或者潜在变量中生成逼真的样本,同时训练一个鉴别器D来鉴别真实数据和生成数据,两者同时训练,利用G和D构成动态“博弈过程”,直到达到一个纳什均衡,生成器生成的数据与真实样本无差别,鉴别器也无法正确的区分生成数据和真实数据。通过基于生成式对抗网络形成的语音信号,从而可以克服目前智能设备在人机对话时只能根据固定的语音库来发声及模式单调缺乏变化,不够自然等缺点。

发明内容

[0004] 本发明提出了一种基于深度卷积生成对抗网络的语音生成方法,目的在于克服目前智能设备在人机对话时只能根据固定的语音库来发声及模式单调缺乏变化的不足。

[0005] 本发明实现的具体步骤如下:

[0006] 步骤1,采集语音信号样本:随机采集M个(M一般取1000)具有相同内容的语音信号,作为语音训练样本和真实语音样本;

[0007] 步骤2,语音信号的预处理:对步骤1中采集到的M个语音信号进行预处理;

[0008] 步骤3,将语音数据样本输入深度卷积生成对抗网络:将步骤2中预处理后的M个语音训练样本数据和M个真实语音样本数据输入到深度卷积生成对抗网络;

[0009] 步骤4,对输入的语音数据进行训练:采用深度卷积生成对抗网络对输入的M个语音训练样本数据和M个真实语音样本数据进行训练;

[0010] 步骤5,生成接近真实语音内容的语音信号:利用深度卷积对抗生成网络对自回归生成模型得到的波形进行训练,最终生成全新的且接近真实语音样本的语音信号。

[0011] 步骤1中所述随机采集M个具有相同内容的语音信号,其存储格式为wav格式。

[0012] 步骤2包括如下步骤:

[0013] 步骤2-1,对采集的语音信号样本做除杂的处理;

[0014] 步骤2-2,对采集的语音信号样本进行滤波处理。

[0015] 步骤2-1包括:利用Audacity软件对采集的语音信号样本进行剪辑,过滤掉原始采

集的波形中超出软件编辑范围的语音部分以及非语音信号的部分。

[0016] 步骤2-2包括如下步骤:

[0017] 步骤2-2-1,计算n时刻的误差信号 $\varepsilon(n)$:

$$[0018] \quad \varepsilon(n) = d(n) - \hat{W}^H(n)X(n),$$

[0019] 其中, $d(n)$ 表示n时刻主信道输入的带噪语音信号,即为自适应滤波器的期望信号,主信道表示语音输入的通道, $W(n)$ 表示n时刻对应的权重系数矢量, $X(n)$ 表示在n时刻的语音矢量, H 表示共轭转置;

[0020] 步骤2-2-2,计算在n时刻的相关系数 $\gamma(n)$:

$$[0021] \quad \gamma(n) = \frac{X^H(n)X(n-1)}{X^H(n-1)X(n-1)},$$

[0022] 步骤2-2-3,计算n时刻的语音差异矢量 $U(n)$:

$$[0023] \quad U(n) = X(n) - \gamma(n)X(n-1);$$

[0024] 步骤2-2-4,通过如下公式计算迭代后每个时刻对应的权重系数矢量:

$$[0025] \quad \hat{W}(n+1) = \hat{W}(n) + \frac{\mu}{\|U(n)\|^2 + \delta} U(n)\varepsilon(n),$$

[0026] 其中, μ 表示自适应常数, $\delta > 0$, δ 表示实数;

[0027] 步骤2-2-5,计算最小方差 $\min_w E\{\varepsilon^2\}$:

$$[0028] \quad \min_w E\{\varepsilon^2\} = E\{v^2\} + \min_w E\{(n_0 - y)^2\},$$

[0029] 其中, n_0 表示主信道输入的语音信号, v 表示主信道输入的噪声, y 表示自适应滤波器的输出, ε 表示误差信号, w 表示某特定时刻自适应滤波器的权重矢量,使得方差 $E\{\varepsilon^2\}$ 最小;

[0030] 步骤2-2-6,将步骤2-2-4中得到的每个时刻的对应的权重系数矢量分别带入步骤2-2-5的公式中,计算最小方差 $\min_w E\{\varepsilon^2\}$,若该时刻求得的不是最小方差,则通过自适应滤波器调整其下一时刻权重矢量 w 进行滤波处理,求出最小方差,得到滤波后的语音信号。

[0031] 步骤4包括如下步骤:

[0032] 步骤4-1,深度卷积生成对抗网络包括两个网络,一个是生成网络 G ,用于接收一个随机的噪声;一个是判别网络 D ,用于判别生成的数据是不是真实的;

[0033] 步骤4-2,计算生成网络的损失函数:

$$[0034] \quad (1-b) \log(1-D(G(z))),$$

[0035] 其中, z 表示生成网络 G 接收的一个随机的噪声, b 表示在真实语音样本数据输入参数 c 下得到的输出参数, $G(z)$ 表示生成网络 G 的输出, \log 表示以10为底对的对数操作, $D(G(z))$ 表示判别网络 D 判断生成网络 G 生成的语音数据为真实的概率;

[0036] 步骤4-3,计算判别网络的损失函数:

$$[0037] \quad -((1-b) \log(1-D(G(z))) + b \log D(c)),$$

[0038] 其中, c 表示真实语音样本数据作为输入参数, $D(c)$ 表示判别网络 D 的输出, 即输入真实语音样本数据参数 c 为真实的概率;

[0039] 步骤4-4, 计算优化函数 $\min \max V(D, G)$:

$$[0040] \quad \min \max V(D, G) = E_{x \sim p_{data}(x)} [\log(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

[0041] 其中, $p_z(z)$ 表示随机噪声的概率密度, x 表示真实语音样本数据, $p_{data}(x)$ 表示参数数据的概率密度;

[0042] 步骤4-5, 当判别网络 D 无法判定生成网络 G 所生成的语音数据是否真实时, $D(G(z)) = 0.5$, 得到训练好的语音信号数据, 否则, 执行步骤4-6;

[0043] 步骤4-6, 利用随机梯度下降法计算判别网络 D 的梯度函数和生成网络 G 的梯度函数:

$$[0044] \quad \nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(c^{(i)}) + \log(1 - D(G(z^{(i)})))]$$

$$[0045] \quad \nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^{(i)})))$$

[0046] 其中, θ_d 表示在判别网络 D 方向上的梯度变化量, θ_g 表示在生成网络 G 方向上的梯度变化量, $i \in [1, m]$, m 表示生成网络 G 接收随机噪声的个数, \in 表示属于符号, 将训练好的判别网络 D 和生成网络 G 带入步骤4-4中。

[0047] 步骤5包括如下步骤:

[0048] 步骤5-1, 计算自回归生成模型:

$$[0049] \quad p(s|h) = \prod_{t=1}^T p(s_t | s_1, \dots, s_{t-1}, h)$$

[0050] 其中, h 表示步骤4中得到训练好的语音样本数据作为输入的参数, $p(s|h)$ 表示最有可行的波形输出向量, $s(t)$ 表示在时间点 t 的输出, $t \in [1, T]$, T 表示总时间, \in 表示属于符号;

[0051] 步骤5-2, 采用步骤4中深度卷积生成对抗网络, 对步骤5-1中生成波形进行学习, 再经过自回归生成模型不断的反馈及输出, 使得每一步当前输出的结果, 只与之前的输出结果相关, 当满足步骤4-5中 $D(G(z)) = 0.5$ 时, 最终生成全新的且接近真实语音样本的语音信号。

[0052] 有益效果: 与现有技术相比本发明具有以下优点:

[0053] 第一, 本发明主要优势在于智能设备不需要借助一个死板的语音库来实现人机交流, 而是可以通过机器训练, 让智能设备可以自主生成全新的语音, 从而使智能设备可以清晰、自然地说话, 普通用户也很容易听懂并接受。

[0054] 第二, 本发明是基于深度卷积生成对抗网络的语音生成技术, 利用判别网络和生成网络所构成的动态“博弈过程”, 最终得到语音训练数据, 再利用自回归生成模型, 将生成的语音波形转化为音频信号, 从而生成接近真实语音内容的自然语音信号。本发明可用于根据样本训练后生成的语音完成人与智能设备之间的自然语音交流。

附图说明

[0055] 下面结合附图和具体实施方式对本发明做更进一步的具体说明, 本发明的上述或

其他方面的优点将会变得更加清楚。

[0056] 图1是本发明流程图。

[0057] 图2是自适应噪声滤波流程图。

[0058] 图3是实施例生成(“你好”)语音波形图

具体实施方式

[0059] 下面结合附图及实施例对本发明做进一步说明。

[0060] 参照图1,对本发明做进一步的详细描述:

[0061] 步骤1,语音信号样本采集:

[0062] 随机采集M个具有相同内容的语音信号分别作为语音训练样本和真实语音样本;

[0063] 步骤2,语音信号的预处理:

[0064] (2a) 将采集的语音信号做除杂的处理,保留完整的语音信号;

[0065] 利用Audacity软件对语音信号进行剪辑,过滤掉原始采集的波形中超出软件编辑范围的语音部分以及非语音信号的部分;

[0066] (2b) 对完整的语音信号进行滤波处理,组成语音训练样本库和真实语音样本库,参照图2,具体步骤如下:

[0067] 第1步,计算n时刻的误差信号 $\varepsilon(n)$:

$$[0068] \quad \varepsilon(n) = d(n) - \hat{W}^H(n)X(n),$$

[0069] 其中, $d(n)$ 表示n时刻主信道输入的带噪语音信号,即为自适应滤波器的期望信号,主信道表示语音输入的通道, \hat{W} 表示n时刻对应的权重系数矢量, $X(n)$ 表示在n时刻的语音矢量, H 表示共轭转置;

[0070] 第2步,计算在n时刻的相关系数 $\gamma(n)$:

$$[0071] \quad \gamma(n) = \frac{X^H(n)X(n-1)}{X^H(n-1)X(n-1)};$$

[0072] 第3步,计算n时刻的语音差异矢量 $U(n)$:

$$[0073] \quad U(n) = X(n) - \gamma(n)X(n-1);$$

[0074] 第4步,计算迭代后每个时刻的 $W(n)$:

$$[0075] \quad \hat{W}(n+1) = \hat{W}(n) + \frac{\mu}{\|U(n)\|^2 + \delta} U(n)\varepsilon(n),$$

[0076] 其中, μ 表示自适应常数, $\delta > 0$, δ 表示实数, $W(n)$ 表示权重系数矢量;

[0077] 第5步,计算最小方差 $\min_w E\{\varepsilon^2\}$;

$$[0078] \quad \min_w E\{\varepsilon^2\} = E\{v^2\} + \min_w E\{(n_0 - y)^2\},$$

[0079] 其中, n_0 表示主信道输入的语音信号, v 表示主信道输入的噪声, y 表示自适应滤波器的输出, ε 表示误差信号, w 表示某时刻自适应滤波器的权重矢量,使得方差 $E\{\varepsilon^2\}$ 最小;

[0080] 第6步,将第4步中得到的每个时刻的 $W(n)$ 分别带入第5步中,计算最小方差

$\min_w E\{\varepsilon^2\}$, 若该时刻求得的不是最小方差, 则通过自适应滤波器调整其下一时刻权重矢量 w 进行滤波处理, 求出最小方差, 得到滤波后的语音信号;

[0081] 步骤3, 将语音数据样本输入深度卷积生成对抗网络;

[0082] 将步骤2中预处理所得的M个语音训练样本数据和M个真实语音样本数据输入到深度卷积生成对抗网络;

[0083] 步骤4, 对输入的语音数据进行训练;

[0084] 采用深度卷积生成对抗网络对输入的M个语音训练样本数据和M个真实语音样本数据进行训练, 具体步骤如下:

[0085] 第1步, 深度卷积生成对抗网络包括两个网络, 一个是生成网络G, 用于接收一个随机的噪声; 一个是判别网络D, 用于判别生成的数据是不是真实的;

[0086] 第2步, 计算生成网络的损失函数:

[0087] $(1-b) \log(1-D(G(z)))$

[0088] 其中, z 表示生成网络G接收的一个随机的噪声, b 表示第3步中在真实语音样本数据输入参数 c 下得到的输出参数, $G(z)$ 表示生成网络G的输出, \log 表示以10为底的对数操作, $D(G(z))$ 表示判别网络D判断生成网络G生成的语音数据为真实的概率;

[0089] 第3步, 计算判别网络的损失函数:

[0090] $-(1-b) \log(1-D(G(z))) + b \log D(c)$

[0091] 其中, c 表示真实语音样本数据作为输入参数, $D(c)$ 表示判别网络D的输出, 即输入真实语音样本数据参数 c 为真实的概率;

[0092] 第4步, 计算优化函数 $\min \max V(D, G)$:

[0093] $\min \max V(D, G) = E_{x \sim p_{data}(x)} [\log(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$,

[0094] 其中, $p_z(z)$ 表示随机噪声的概率密度, x 表示真实语音样本数据, $p_{data}(x)$ 表示参数数据的概率密度;

[0095] 第5步, 当判别网络D无法判定生成网络G所生成的语音数据是否真实时, $D(G(z)) = 0.5$, 得到训练好的语音数据, 否则, 执行第6步;

[0096] 第6步, 利用随机梯度下降法计算D和G网络的梯度函数:

[0097] $\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(c^{(i)}) + \log(1 - D(G(z^{(i)})))]$

[0098] $\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^{(i)})))$

[0099] 其中, θ_d 表示在判别网络D方向上的梯度变化量, θ_g 表示在生成网络G方向上的梯度变化量, $i \in [1, m]$, m 表示生成网络G接收随机噪声的个数, \in 表示属于符号, 将训练好的D和G带入第4步中;

[0100] 步骤5, 生成接近真实语音内容的语音信号;

[0101] 利用深度卷积对抗生成网络对自回归生成模型得到的波形进行训练, 最终生成全新的且接近真实语音样本的语音信号, 具体步骤如下:

[0102] 第1步, 计算自回归生成模型:

[0103] $p(s|h) = \prod_{t=1}^T p(s_t | s_1, \dots, s_{t-1}, h)$

[0104] 其中, h 表示步骤4第5步中得到训练好的语音数据作为输入的参数, $p(x|h)$ 表示最有可行的波形输出向量, $s(t)$ 表示在时间点 t 的输出, $t \in [1, T]$, T 表示总时间, \in 表示属于符号;

[0105] 第2步, 采用步骤4中深度卷积生成对抗网络, 对第一步中生成的波形进行学习, 再经过自回归生成模型不断的反馈及输出, 使得每一步当前输出的结果, 只与之前的输出结果相关, 当满足步骤4第5步中 $D(G(z)) = 0.5$ 时, 最终生成全新的且接近真实语音样本的语音信号。

[0106] 实施例

[0107] 一种基于深度卷积生成对抗网络的语音生成方法, 其主要方法的流程图如图1所示, 具体包括以下步骤:

[0108] (1) 采集语音信号样本: 随机采集1000个具有相同内容 (“你好”) 语音信号分别作为语音训练样本和真实语音样本;

[0109] (2) 语音信号的预处理: 对步骤1中采集到的1000个 (“你好”) 语音信号进行预处理。首先, 利用Audacity软件对1000个 (“你好”) 语音信号进行剪辑, 过滤掉原始采集的波形中超出软件编辑范围的语音部分以及非语音信号的部分; 其次, 进行语音滤波处理: 通过滤波算法计算 n 个时刻的误差信号、相关系数、语音差异矢量, 由迭代方法计算出每个时刻的权重系数矢量, 通过自适应滤波器调整其下一时刻权重矢量求出最小方差, 从而进行滤波处理, 便可得到滤波后的语音信号;

[0110] (3) 将语音数据样本输入深度卷积生成对抗网络: 将 (2) 中预处理都所得的1000个 (“你好”) 语音训练样本数据和真实语音样本数据输入到深度卷积生成对抗网络;

[0111] (4) 对输入的语音数据进行训练: 采用深度卷积生成对抗网络对输入的1000个 (“你好”) 语音训练样本数据和真实语音样本数据进行训练, 在该网络下, 分别计算生成网络 G 和判别网络 D 损失函数, 再利用随机梯度下降算法来训练 D 和 G , 最终求得最优函数, 得到训练好的 (“你好”) 语音数据;

[0112] (5) 生成接近真实语音内容的语音信号: 将训练好的 (“你好”) 语音数据带入自回归生成模型得到波形, 再利用深度卷积对抗生成网络对自回归生成模型得到的波形进行训练, 得到训练后的语音波形图, 如图3所示, 最终可生成全新的且接近真实 (“你好”) 语音样本的语音信号。

[0113] 本发明提供了一种基于深度卷积生成对抗网络的语音生成方法, 具体实现该技术方案的方法和途径很多, 以上所述仅是本发明的优选实施方式, 应当指出, 对于本技术领域的普通技术人员来说, 在不脱离本发明原理的前提下, 还可以做出若干改进和润饰, 这些改进和润饰也应视为本发明的保护范围。本实施例中未明确的各组成部分均可用现有技术加以实现。

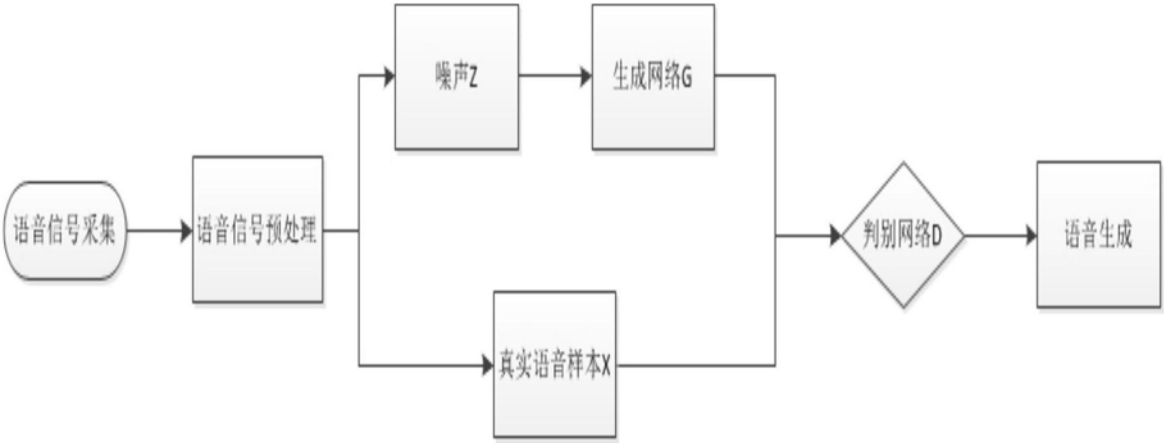


图1

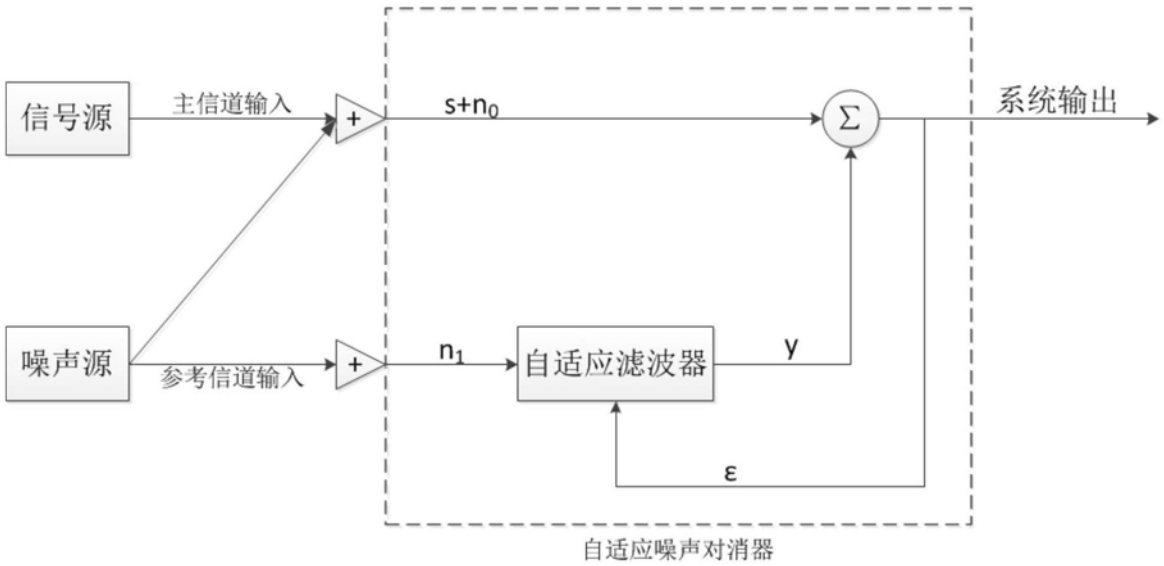


图2

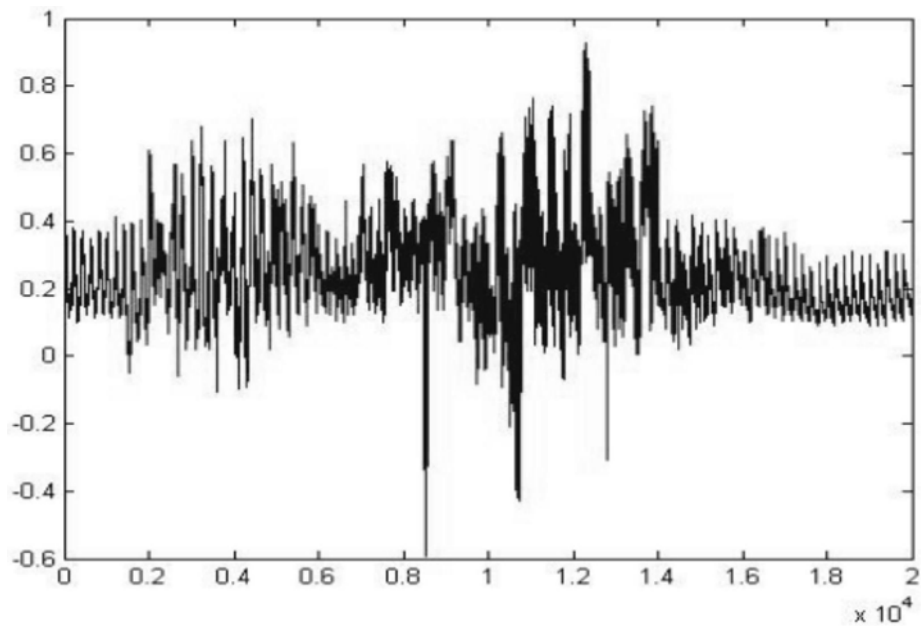


图3