(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 110010144 A
(43)申请公布日 2019.07.12

(21)申请号 201910336274.0

(22)申请日 2019.04.24

(71)申请人 厦门亿联网络技术股份有限公司
地址 361009 福建省厦门市湖里区高新园区岭下北路1号亿联研发大楼

(72)发明人 冯万健 张联昌 刘键涛

(74)专利代理机构 广州三环专利商标代理有限公司 44202
代理人 颜希文 麦小婵

(51)Int.Cl.
G10L 21/02(2013.01)
G10L 25/30(2013.01)
G10L 25/45(2013.01)

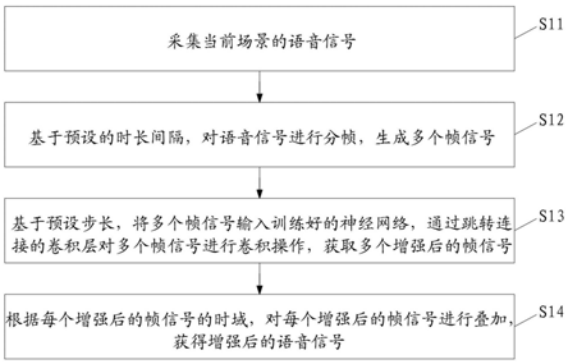
权利要求书2页 说明书6页 附图1页

(54)发明名称

语音信号增强方法及装置

(57)摘要

本申请公开了一种语音信号增强方法及装置，通过将当前场景的语音信号基于预设的时长间隔分割成多个帧信号；基于预设步长，将多个帧信号输入训练好的神经网络，通过跳转连接的卷积层对多个帧信号进行卷积操作，获取多个增强后的帧信号；根据每个增强后的帧信号的时域，对每个增强后的帧信号进行叠加，获得增强后的语音信号。与现有技术相比，本申请通过神经网络对语音信号进行自动增强，无需人工干预，使得语音增强的效果和应用场景无需受限于预设方法及方法设计者，从而降低信号失真和额外杂音的出现频率，进而提高语音信号增强效果。



CN 110010144 A

1.一种语音信号增强方法,其特征在于,包括:

采集当前场景的语音信号;

基于预设的时长间隔,对所述语音信号进行分帧,生成多个帧信号;

基于预设步长,将多个所述帧信号输入训练好的神经网络,通过跳转连接的卷积层对多个所述帧信号进行卷积操作,获取多个增强后的帧信号;

根据每个增强后的帧信号的时域,对每个增强后的帧信号进行叠加,获得增强后的所述语音信号。

2.根据权利要求1所述的语音信号增强方法,其特征在于,所述基于预设的时长间隔,对所述语音信号进行分帧,生成多个帧信号,具体为:

基于预设的时长间隔,对所述语音信号进行分帧,并将分帧后的所述语音信号加以汉宁窗后进行DFT,生成多个帧信号。

3.根据权利要求1所述的语音信号增强方法,其特征在于,所述神经网络的训练方法为:

采集多个噪声信号及不带噪声的多个清晰信号;

基于随机生成的混合系数,将多个所述噪声信号与多个所述清晰信号一一进行混合,获得多个所述带噪信号;其中,一个所述噪声信号与一个所述清晰信号混合成一个所述带噪信号;

将多个所述带噪信号依次输入所述神经网络进行信号增强,产生一一对应的多个降噪信号,并根据各所述降噪信号与各降噪信号一一对应的各所述清晰信号的最小平方误差,调整所述神经网络。

4.根据权利要求3所述的语音信号增强方法,其特征在于,所述将多个所述带噪信号依次输入所述神经网络进行信号增强,产生一一对应的多个降噪信号,并根据各所述降噪信号与各降噪信号一一对应的各所述清晰信号的最小平方误差,调整所述神经网络,具体为:

将所述带噪信号输入所述神经网络,根据所述带噪信号通过所述神经网络进行信号增强后产生的降噪信号,与对应的清晰信号的最小平方误差,调整所述神经网络,并根据下一所述带噪信号通过调整后的神经网络产生的降噪信号,与对应的清晰信号的最小平方误差,继续调整所述神经网络,直至利用不同的带噪信号获得的最小平方误差不再产生变化时,完成所述神经网络的训练。

5.根据权利要求1-4任意一项所述的语音信号增强方法,其特征在于,所述神经网络包括N个依次排序的卷积层;以第N/2层卷积层为对称轴,两两对称的卷积层之间跳转连接;其中,N为偶数。

6.一种语音信号增强装置,其特征在于,包括:

信号采集模块,用于采集当前场景的语音信号;

信号分帧模块,用于基于预设的时长间隔,对所述语音信号进行分帧,生成多个帧信号;

信号增强模块,用于基于预设步长,将多个所述帧信号输入训练好的神经网络,通过跳转连接的卷积层对多个所述帧信号进行卷积操作,获取多个增强后的帧信号;

信号输出模块,用于根据每个增强后的帧信号的时域,对每个增强后的帧信号进行叠加,获得增强后的所述语音信号。

7.根据权利要求6所述的语音信号增强装置,其特征在于,所述信号分帧模块具体用于:

基于预设的时长间隔,对所述语音信号进行分帧,并将分帧后的所述语音信号加以汉宁窗后进行DFT,生成多个帧信号。

8.根据权利要求6所述的语音信号增强装置,其特征在于,所述神经网络的训练方法为:

采集多个噪声信号及不带噪声的多个清晰信号;

基于随机生成的混合系数,将多个所述噪声信号与多个所述清晰信号一一进行混合,获得多个所述带噪信号;其中,一个所述噪声信号与一个所述清晰信号混合成一个所述带噪信号;

将多个所述带噪信号依次输入所述神经网络进行信号增强,产生一一对应的多个降噪信号,并根据各所述降噪信号与各降噪信号一一对应的各所述清晰信号的最小平方误差,调整所述神经网络。

9.根据权利要求8所述的语音信号增强装置,其特征在于,所述将多个所述带噪信号依次输入所述神经网络进行信号增强,产生一一对应的多个降噪信号,并根据各所述降噪信号与各降噪信号一一对应的各所述清晰信号的最小平方误差,调整所述神经网络,具体为:

将所述带噪信号输入所述神经网络,根据所述带噪信号通过所述神经网络进行信号增强后产生的降噪信号,与对应的清晰信号的最小平方误差,调整所述神经网络,并根据下一所述带噪信号通过调整后的神经网络产生的降噪信号,与对应的清晰信号的最小平方误差,继续调整所述神经网络,直至利用不同的带噪信号获得的最小平方误差不再产生变化时,完成所述神经网络的训练。

10.根据权利要求6-9任意一项所述的语音信号增强装置,其特征在于,所述神经网络包括N个依次排序的卷积层;以第N/2层卷积层为对称轴,两两对称的卷积层之间跳转连接。

语音信号增强方法及装置

技术领域

[0001] 本申请涉及语音信号处理技术领域，尤其涉及一种语音信号增强方法及装置。

背景技术

[0002] 语音信号增强是为了提升语音的可懂性，和提升那些被加性噪声所污染的语音，其主要应用于主要应用于通信设备，同样也有应用在听力辅助，人工耳蜗等植入设备。现有的语音信号增强方法，通常采用“谱减法”、“维纳滤波”、“统计模型方法”、“子空间法”等。但在采用现有技术进行语音信号增强时发现，由于这些语音信号增强方法在原理上属于人工预设方法，因此效果和应用场景均受限于预设方法及方法设计者，且现实中语音场景多种多样，采用现有技术进行语音增强的过程中不可避免地会出现信号失真以及出现额外杂音的情况，因此，在面对复杂的语音场景时，现有的语音增强技术的鲁棒性较差。

发明内容

[0003] 本申请实施例所要解决的技术问题在于，提供一种语音信号增强方法及装置，实现对不同场景的语音信号的增强。

[0004] 为解决上述问题，本申请实施例提供一种语音信号增强方法，至少包括：

[0005] 采集当前场景的语音信号；

[0006] 基于预设的时长间隔，对所述语音信号进行分帧，生成多个帧信号；

[0007] 基于预设步长，将多个所述帧信号输入训练好的神经网络，通过跳转连接的卷积层对多个所述帧信号进行卷积操作，获取多个增强后的帧信号；

[0008] 根据每个增强后的帧信号的时域，对每个增强后的帧信号进行叠加，获得增强后的所述语音信号。

[0009] 进一步的，所述基于预设的时长间隔，对所述语音信号进行分帧，生成多个帧信号，具体为：

[0010] 基于预设的时长间隔，对所述语音信号进行分帧，并将分帧后的所述语音信号加以汉宁窗后进行DFT，生成多个帧信号。

[0011] 进一步的，所述神经网络的训练方法为：

[0012] 采集多个噪声信号及不带噪声的多个清晰信号；

[0013] 基于随机生成的混合系数，将多个所述噪声信号与多个所述清晰信号一一进行混合，获得多个所述带噪信号；其中，一个所述噪声信号与一个所述清晰信号混合成一个所述带噪信号；

[0014] 将多个所述带噪信号依次输入所述神经网络进行信号增强，产生一一对应的多个降噪信号，并根据各所述降噪信号与各降噪信号一一对应的各所述清晰信号的最小平方误差，调整所述神经网络。

[0015] 进一步的，所述将多个所述带噪信号依次输入所述神经网络进行信号增强，产生一一对应的多个降噪信号，并根据各所述降噪信号与各降噪信号一一对应的各所述清晰信

号的最小平方误差，调整所述神经网络，具体为：

[0016] 将所述带噪信号输入所述神经网络，根据所述带噪信号通过所述神经网络进行信号增强后产生的降噪信号，与对应的清晰信号的最小平方误差，调整所述神经网络，并根据下一所述带噪信号通过调整后的神经网络产生的降噪信号，与对应的清晰信号的最小平方误差，继续调整所述神经网络，直至利用不同的带噪信号获得的最小平方误差不再产生变化时，完成所述神经网络的训练。

[0017] 进一步的，所述神经网络包括N个依次排序的卷积层；以第N/2层卷积层为对称轴，两两对称的卷积层之间跳转连接；其中，N为偶数。

[0018] 进一步的，还提供一种语音信号增强装置，包括：

[0019] 信号采集模块，用于采集当前场景的语音信号；

[0020] 信号分帧模块，用于基于预设的时长间隔，对所述语音信号进行分帧，生成多个帧信号；

[0021] 信号增强模块，用于基于预设步长，将多个所述帧信号输入训练好的神经网络，通过跳转连接的卷积层对多个所述帧信号进行卷积操作，获取多个增强后的帧信号；

[0022] 信号输出模块，用于根据每个增强后的帧信号的时域，对每个增强后的帧信号进行叠加，获得增强后的所述语音信号。

[0023] 进一步的，所述信号分帧模块具体用于：

[0024] 基于预设的时长间隔，对所述语音信号进行分帧，并将分帧后的所述语音信号加以汉宁窗后进行DFT，生成多个帧信号。

[0025] 进一步的，所述神经网络的训练方法为：

[0026] 采集多个噪声信号及不带噪声的多个清晰信号；

[0027] 基于随机生成的混合系数，将多个所述噪声信号与多个所述清晰信号一一进行混合，获得多个所述带噪信号；其中，一个所述噪声信号与一个所述清晰信号混合成一个所述带噪信号；

[0028] 将多个所述带噪信号依次输入所述神经网络进行信号增强，产生一一对应的多个降噪信号，并根据各所述降噪信号与各降噪信号一一对应的各所述清晰信号的最小平方误差，调整所述神经网络。

[0029] 进一步的，所述将多个所述带噪信号依次输入所述神经网络进行信号增强，产生一一对应的多个降噪信号，并根据各所述降噪信号与各降噪信号一一对应的各所述清晰信号的最小平方误差，调整所述神经网络，具体为：

[0030] 将所述带噪信号输入所述神经网络，根据所述带噪信号通过所述神经网络进行信号增强后产生的降噪信号，与对应的清晰信号的最小平方误差，调整所述神经网络，并根据下一所述带噪信号通过调整后的神经网络产生的降噪信号，与对应的清晰信号的最小平方误差，继续调整所述神经网络，直至利用不同的带噪信号获得的最小平方误差不再产生变化时，完成所述神经网络的训练。

[0031] 进一步的，所述神经网络包括N个依次排序的卷积层；以第N/2层卷积层为对称轴，两两对称的卷积层之间跳转连接。

[0032] 实施本申请实施例，具有如下有益效果：

[0033] 本申请实施例提供的一种语音信号增强方法及装置，通过将当前场景的语音信号

基于预设的时长间隔分割成多个帧信号;基于预设步长,将多个帧信号输入训练好的神经网络,通过跳转连接的卷积层对多个帧信号进行卷积操作,获取多个增强后的帧信号;根据每个增强后的帧信号的时域,对每个增强后的帧信号进行叠加,获得增强后的语音信号。与现有技术相比,本申请通过神经网络对语音信号进行自动增强,无需人工干预,使得语音增强的效果和应用场景无需受限于预设方法及方法设计者,从而降低信号失真和额外杂音的出现频率,进而提高语音信号增强效果。

附图说明

[0034] 图1是本申请的一个实施例提供的语音信号增强方法的流程示意图;

[0035] 图2是本申请的一个实施例提供的神经网络训练方法的流程示意图;

[0036] 图3是本申请的一个实施例提供的语音信号增强装置的结构示意图。

具体实施方式

[0037] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0038] 请参见图1。

[0039] 参见图1,是本申请的一个实施例提供的语音信号增强方法的流程示意图,如图1所示,该语音信号增强方法包括:

[0040] 步骤S11、采集当前场景的语音信号。

[0041] 由于采样频率为22.05KHz的音源已经达到了FM广播的声音品质,能够被清楚识别,若采集22.05KHz以上的音源进行语音信号增强,效果也并不显著,因此在本实施例中,采集当前场景中采样频率为16KHz的音源作为语音信号。

[0042] 步骤S12、基于预设的时长间隔,对语音信号进行分帧,生成多个帧信号。

[0043] 具体的,基于预设的时长间隔,对语音信号进行分帧,并将分帧后的语音信号加以汉宁窗后进行DFT,生成多个帧信号。

[0044] 在本实施例中,预设的时长间隔为16ms。

[0045] 步骤S13,基于预设步长,将多个帧信号输入训练好的神经网络,通过跳转连接的卷积层对多个帧信号进行卷积操作,获取多个增强后的帧信号。

[0046] 由于多个帧信号中存在信号重叠,因此在本实施例中,以50%的帧长为步长,将多个帧信号按生成顺序,每10帧输入训练好的神经网络。

[0047] 步骤S14,根据每个增强后的帧信号的时域,对每个增强后的帧信号进行叠加,获得增强后的语音信号。

[0048] 考虑到增强后的多个帧信号之间存在信号重叠,因此在本实施例中,通过重叠叠加法,将每个增强后的帧信号重构成时域信号,该时域信号即为增强后的语音信号。

[0049] 请参见图2。

[0050] 进一步的,参见图2,是本申请的一个实施例提供的神经网络训练方法的流程示意图。包括:

[0051] S21,采集多个噪声信号及不带噪声的多个清晰信号。

[0052] 在本实施例中,将采集的多个噪声信号及不带噪声的多个清晰信号整理归入数据池后,对数据池中的多个噪声信号分别标以噪声1、噪声2……噪声N1,对数据池中的多个清晰信号分别标以语音1、语音2……语音N2。

[0053] S22,基于随机生成的混合系数,将多个噪声信号与多个清晰信号一一进行混合,获得多个带噪信号。

[0054] 其中,一个噪声信号与一个清晰信号混合成一个带噪信号。

[0055] 在本实施例中,随机生成一个范围在1到N1的随机整数RND1,根据该随机整数,从数据池中获取对应数字编号的噪声信号并标记为噪声RND1,再随机生成一个范围在1到N2的随机整数RND2,根据该随机整数,从数据池中获取对应数字编号的清晰信号并标记为语音RND2,并生成范围在0到1间的随机数RND3,将噪声RND1及语音RND2按RND3的系数进行混合,生成带噪信号后,重复上述过程,从而获得多个带噪信号。

[0056] S23,将多个带噪信号依次输入神经网络进行信号增强,产生一一对应的多个降噪信号,并根据各降噪信号与各降噪信号一一对应的各清晰信号的最小平方误差,调整神经网络。

[0057] 具体的,将带噪信号输入神经网络,根据带噪信号通过神经网络进行信号增强后产生的降噪信号,与对应的清晰信号的最小平方误差,调整神经网络,并根据下一带噪信号通过调整后的神经网络产生的降噪信号,与对应的清晰信号的最小平方误差,继续调整神经网络,直至利用不同的带噪信号获得的最小平方误差不再产生变化时,完成神经网络的训练。

[0058] 在本实施例中,将带噪信号输入神经网络,获得降噪信号后,计算降噪信号与对应的清晰信号之间的最小平方误差,并根据该最小平方误差,利用Adam优化器优化神经网络的网络参数,从而调整神经网络。

[0059] 在本实施例中,神经网络的网络架构包括N个依次排序的卷积层,且以第N/2层卷积层为对称轴,两两对称的卷积层之间跳转连接。其中,N为偶数。

[0060] 具体的,神经网络由12个卷积层组成,每个卷积层后跟以一个批量标准化(BN)层,最后以线性整流单元(ReLU)激活函数进行激活。每层卷积层通道数以第六层为中心,向两侧对称排布,所对称的卷积对,2层与10层、3层与9层、4层与8层、5层与7层分别跳转连接,输入神经网络的数据通过前11层卷积后,进行最后一次卷积操作,最终获得输出与输入相同形状的数据。具体通道数及卷积核大小如下表所示:

[0061]

Layer name	Input feature	Output feature	Kernal size
Convolution1	1	8	9
Convolution2	8	12	9
Convolution3	12	16	7
Convolution4	16	20	7
Convolution5	20	24	5
Convolution6	24	28	5
Convolution7	28	24	5

Convolution8	24	20	7
Convolution9	20	16	7
Convolution10	16	12	9
Convolution11	12	8	9
FinalConvolution	8	1	129

[0062] 请参见图3。

[0063] 进一步的,参见图3,是本申请的一个实施例提供的语音信号增强装置的结构示意图。包括:

[0064] 信号采集模块101,用于采集当前场景的语音信号。

[0065] 由于采样频率为22.05KHz的音源已经达到了FM广播的声音品质,能够被清楚识别,若采集22.05KHz以上的音源进行语音信号增强,效果也并不显著,因此在本实施例中,信号采集模块101用于采集当前场景中采样频率为16KHz的音源作为语音信号。

[0066] 信号分帧模块102,用于基于预设的时长间隔,对语音信号进行分帧,生成多个帧信号。

[0067] 在本实施例中,信号分帧模块102具体用于,基于预设的时长间隔,对语音信号进行分帧,并将分帧后的语音信号加以汉宁窗后进行DFT,生成多个帧信号。

[0068] 在本实施例中,预设的时长间隔为16ms。

[0069] 信号增强模块103,用于基于预设步长,将多个帧信号输入训练好的神经网络,通过跳转连接的卷积层对多个帧信号进行卷积操作,获取多个增强后的帧信号。

[0070] 由于多个帧信号中存在信号重叠,因此在本实施例中,信号增强模块103用于以50%的帧长为步长,将多个帧信号按生成顺序,每10帧输入训练好的神经网络。

[0071] 信号输出模块104,用于根据每个增强后的帧信号的时域,对每个增强后的帧信号进行叠加,获得增强后的语音信号。

[0072] 考虑到增强后的多个帧信号之间存在信号重叠,因此在本实施例中,信号输出模块104用于通过重叠叠加法,将每个增强后的帧信号重构成时域信号,该时域信号即为增强后的语音信号。

[0073] 本申请实施例提供一种语音信号增强方法及装置,通过将当前场景的语音信号基于预设的时长间隔分割成多个帧信号;基于预设步长,将多个帧信号输入训练好的神经网络,通过跳转连接的卷积层对多个帧信号进行卷积操作,获取多个增强后的帧信号;根据每个增强后的帧信号的时域,对每个增强后的帧信号进行叠加,获得增强后的语音信号。与现有技术相比,本申请通过神经网络对语音信号进行自动增强,无需人工干预,使得语音增强的效果和应用场景无需受限于预设方法及方法设计者,从而降低信号失真和额外杂音的出现频率,进而提高语音信号增强效果。

[0074] 本申请的又一的实施例还提供了一种语音信号增强终端设备,包括处理器、存储器以及存储在所述存储器中且被配置为由所述处理器执行的计算机程序,所述处理器执行所述计算机程序时实现如上述实施例所述的语音信号增强方法。

[0075] 以上所述是本申请的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本申请原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也视为本申请的保护范围。

[0076] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的程序可存储于一计算机可读取存储介质中,该程序在执行时,可包括如上述各方法的实施例的流程。其中,所述的存储介质可为磁碟、光盘、只读存储记忆体(Read-Only Memory,ROM)或随机存储记忆体(Random Access Memory,RAM)等。

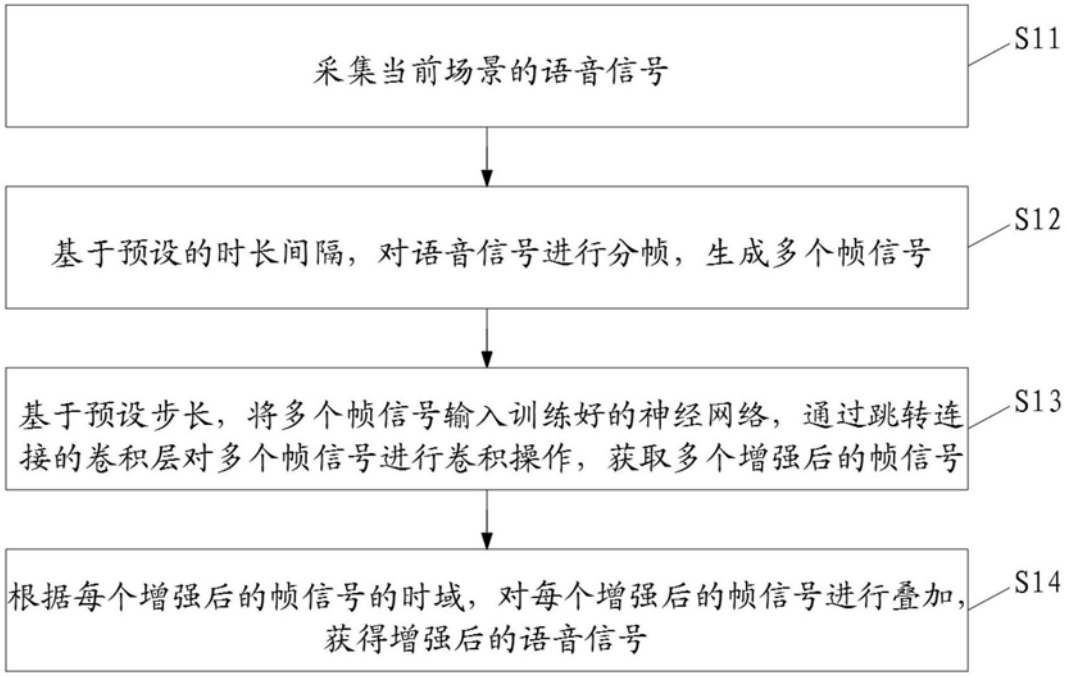

采集当前场景的语音信号
S11
基于预设的时长间隔，对语音信号进行分帧，生成多个帧信号
S12
基于预设步长，将多个帧信号输入训练好的神经网络，通过跳转连接的卷积层对多个帧信号进行卷积操作，获取多个增强后的帧信号
S13
根据每个增强后的帧信号的时域，对每个增强后的帧信号进行叠加，获得增强后的语音信号
S14

图1

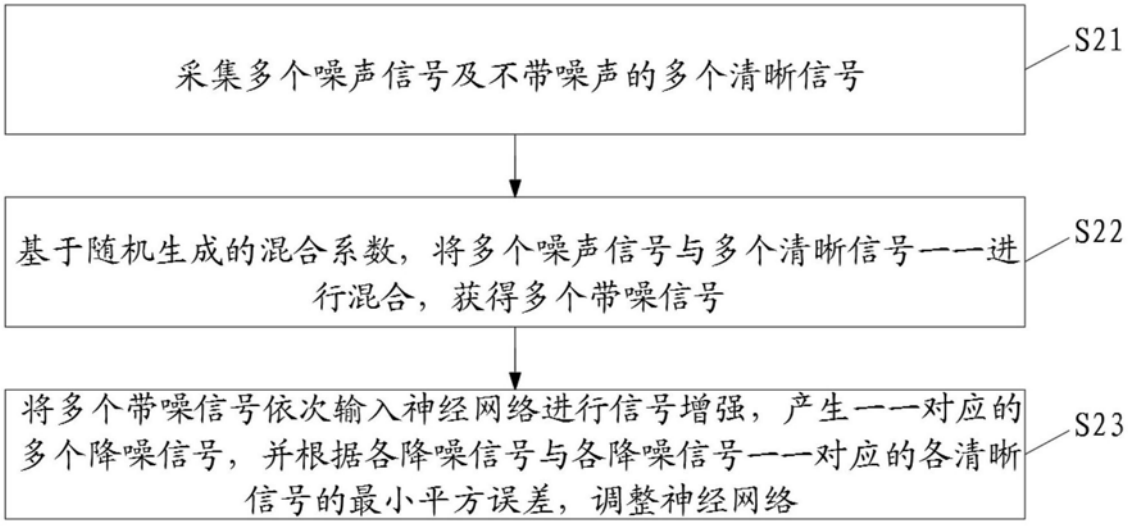

采集多个噪声信号及不带噪声的多个清晰信号
S21
基于随机生成的混合系数，将多个噪声信号与多个清晰信号一一进行混合，获得多个带噪信号
S22
将多个带噪信号依次输入神经网络进行信号增强，产生一一对应的多个降噪信号，并根据各降噪信号与各降噪信号一一对应的各清晰信号的最小平方误差，调整神经网络
S23

图2

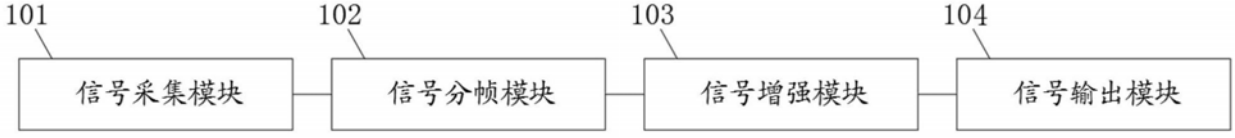

101
信号采集模块
102
信号分帧模块
103
信号增强模块
104
信号输出模块

图3