



(12) 发明专利

(10) 授权公告号 CN 116110422 B

(45) 授权公告日 2023.07.04

(21) 申请号 202310389820.3

审查员 张鹤

(22) 申请日 2023.04.13

(65) 同一申请的已公布的文献号

申请公布号 CN 116110422 A

(43) 申请公布日 2023.05.12

(73) 专利权人 南京熊大巨幕智能科技有限公司

地址 211100 江苏省南京市江宁区江宁高

新园科宁路777号

(72) 发明人 叶玲

(51) Int. Cl.

G10L 21/0216 (2013.01)

G10L 25/30 (2013.01)

G06N 3/0464 (2023.01)

G06N 3/0475 (2023.01)

G06N 3/08 (2023.01)

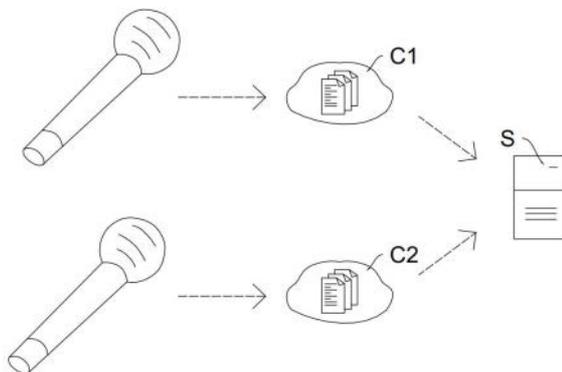
权利要求书3页 说明书11页 附图4页

(54) 发明名称

全向级联麦克风阵列降噪方法及系统

(57) 摘要

本发明公开了一种全向级联麦克风阵列降噪方法及系统,涉及智能化降噪技术领域,全向级联麦克风阵列降噪方法包括:获取由第一全向麦克风单元采集的第一目标声源信号和由第二全向麦克风单元采集的第二目标声源信号;采用基于深度学习的人工智能技术,基于所述第一目标声源信号和所述第二目标声源信号在高维特征空间中的全时序空间联立表示为上下文语义背景,将其映射到第一目标声源的高维特征空间中以进行特征分布校正,并基于对抗生成思想生成降噪后第一目标声源信号。本发明可以增大降噪后第一目标声源信号的准确性,以提高噪声消除效果。



1. 一种全向级联麦克风阵列降噪方法,其特征在于,包括:

获取由第一全向麦克风单元采集的第一目标声源信号和由第二全向麦克风单元采集的第二目标声源信号,所述第一全向麦克风单元和所述第二全向麦克风单元相互级联;

将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量;

将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵;

以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量;

对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量;以及

将所述解码特征向量通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号;

所述对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量,包括:以如下公式对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量;

其中,所述公式为:

$$V' = V \otimes (\|V\|_2^2 \odot V_{order} \otimes V^T);$$

其中, $V$ 表示所述校正第一目标声源波形特征向量, $V'$ 表示所述解码特征向量, $\|V\|_2^2$ 表示所述校正第一目标声源波形特征向量的二范数的平方, $V_{order}$ 表示所述校正第一目标声源波形特征向量的各特征值按大小次序排列的有序向量,且所述校正第一目标声源波形特征向量 $V$ 是列向量形式, $V^T$ 表示 $V$ 的转置矩阵, $\odot$ 表示向量点乘, $\otimes$ 表示矩阵乘法。

2. 根据权利要求1所述的全向级联麦克风阵列降噪方法,其特征在于,所述将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量,包括:使用所述作为过滤器的第一卷积神经网络模型各层在层的正向传递中分别对输入数据进行处理:

对所述输入数据进行卷积处理以得到卷积特征图;

对所述卷积特征图进行基于特征矩阵的均值池化处理以得到池化特征图;以及

对所述池化特征图进行非线性激活以得到激活特征图;

其中,所述作为过滤器的第一卷积神经网络模型的最后一层的输出为所述第一目标声源波形特征向量,所述作为过滤器的第一卷积神经网络模型的第一层的输入为所述第一目标声源信号的波形图。

3. 根据权利要求2所述的全向级联麦克风阵列降噪方法,其特征在于,所述将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵,包括:使用所述作为特征提取器的第二卷积神经网络模型各层在层的正向传递中分别对

输入数据进行处理：

对所述输入数据进行卷积处理以得到卷积特征图；

对所述卷积特征图进行沿通道维度的池化处理以得到池化特征图；以及

对所述池化特征图进行非线性激活以得到激活特征图；

其中，所述作为特征提取器的第二卷积神经网络模型的最后一层的输出为所述协同声音波形特征矩阵，所述作为特征提取器的第二卷积神经网络模型的第一层的输入为所述多通道声音波形图。

4. 根据权利要求3所述的全向级联麦克风阵列降噪方法，其特征在于，所述作为特征提取器的第二卷积神经网络模型为深度残差网络。

5. 根据权利要求4所述的全向级联麦克风阵列降噪方法，其特征在于，所述对抗生成网络包括鉴别器和生成器。

6. 根据权利要求5所述的全向级联麦克风阵列降噪方法，其特征在于，所述将所述解码特征向量通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号，包括：

将所述解码特征向量输入所述基于对抗生成网络的声音信号生成器以由所述对抗生成网络的生成器通过至少两次反卷积编码生成所述降噪后第一目标声源信号。

7. 一种全向级联麦克风阵列降噪系统，其特征在于，包括：

信号获取模块，用于获取由第一全向麦克风单元采集的第一目标声源信号和由第二全向麦克风单元采集的第二目标声源信号，所述第一全向麦克风单元和所述第二全向麦克风单元相互级联；

第一特征提取模块，用于将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量；

第二特征提取模块，用于将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵；

校正模块，用于以所述第一目标声源波形特征向量作为查询特征向量，计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量；

数据流形局部优化模块，用于对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量；以及

声音信号生成模块，用于将所述解码特征向量通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号；

所述数据流形局部优化模块，还用于：以如下公式对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量；

其中，所述公式为：

$$V' = V \otimes (\|V\|_2^2 \odot V_{order} \otimes V^T);$$

其中， $V$ 表示所述校正第一目标声源波形特征向量， $V'$ 表示所述解码特征向量， $\|V\|_2^2$ 表示所述校正第一目标声源波形特征向量的二范数的平方， $V_{order}$ 表示所述校正第一目标声源波形特征向量的各特征值按大小次序排列的有序向量，且所述校正第一目标声源波形

特征向量 $\mathbf{V}$ 是列向量形式， $\mathbf{V}^T$ 表示 $\mathbf{V}$ 的转置矩阵， $\odot$ 表示向量点乘， $\otimes$ 表示矩阵乘法。

8. 根据权利要求7所述的全向级联麦克风阵列降噪系统,其特征在於,所述第一特征提取模块,进一步用于:使用所述作为过滤器的第一卷积神经网络模型的各层在层的正向传递中分别对输入数据进行处理:

对所述输入数据进行卷积处理以得到卷积特征图;

对所述卷积特征图进行基于特征矩阵的均值池化处理以得到池化特征图;以及

对所述池化特征图进行非线性激活以得到激活特征图;

其中,所述作为过滤器的第一卷积神经网络模型的最后一层的输出为所述第一目标声源波形特征向量,所述作为过滤器的第一卷积神经网络模型的第一层的输入为所述第一目标声源信号的波形图。

9. 根据权利要求8所述的全向级联麦克风阵列降噪系统,其特征在於,所述第二特征提取模块,进一步用于:使用所述作为特征提取器的第二卷积神经网络模型的各层在层的正向传递中分别对输入数据进行处理:

对所述输入数据进行卷积处理以得到卷积特征图;

对所述卷积特征图进行沿通道维度的池化处理以得到池化特征图;以及

对所述池化特征图进行非线性激活以得到激活特征图;

其中,所述作为特征提取器的第二卷积神经网络模型的最后一层的输出为所述协同声音波形特征矩阵,所述作为特征提取器的第二卷积神经网络模型的第一层的输入为所述多通道声音波形图。

## 全向级联麦克风阵列降噪方法及系统

### 技术领域

[0001] 本发明涉及智能化降噪技术领域,且更为具体地,涉及一种全向级联麦克风阵列降噪方法及系统。

### 背景技术

[0002] 目前,麦克风阵列有尺寸较大的,由至少两个麦克风组成。现行的一种麦克风阵列设备包括两个全指向性麦克风构成,且两个麦克风间距离较小。

[0003] 目前,麦克风阵列算法被越来越多的应用在噪声消除中。麦克风阵列算法主要利用目标声源和噪声源在空间域上的差别,即各声源到麦克风的距离和方向不同,来进行信号分离,从而实现噪声消除。

[0004] 现有的小距离双麦克风阵列算法能够正常工作的前提条件是:两个麦克风的匹配性很好。但是在实际应用中,这一点是不能完全保证的,即使是同批次的普通麦克风,其幅频特性有时相差会在3-4db左右,且随着时间的推移和外界环境温度、湿度的变化,麦克风的幅频特性也会发生改变,因此当两个麦克风不完全匹配时,就会大大影响噪声消除效果,导致影响语音质量。

### 发明内容

[0005] 为了解决上述技术问题,提出了本发明。本发明的实施例提供了一种全向级联麦克风阵列降噪方法及系统,其获取由第一全向麦克风单元采集的第一目标声源信号和由第二全向麦克风单元采集的第二目标声源信号;采用基于深度学习的人工智能技术,基于所述第一目标声源信号和所述第二目标声源信号在高维特征空间中的全时序空间联立表示为上下文语义背景,将其映射到第一目标声源的高维特征空间中以进行特征分布校正,并基于对抗生成思想生成降噪后第一目标声源信号。这样,可以增大降噪后第一目标声源信号的准确性,以提高噪声消除效果。

[0006] 根据本发明的一个方面,提供了一种全向级联麦克风阵列降噪方法,其包括:

[0007] 获取由第一全向麦克风单元采集的第一目标声源信号和由第二全向麦克风单元采集的第二目标声源信号,所述第一全向麦克风单元和所述第二全向麦克风单元相互级联;

[0008] 将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量;

[0009] 将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵;

[0010] 以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量;

[0011] 对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征

向量;以及

[0012] 将所述解码特征向量通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号。

[0013] 在上述全向级联麦克风阵列降噪方法中,所述将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量,包括:使用所述作为过滤器的第一卷积神经网络模型的各层在层的正向传递中分别对输入数据进行处理:对所述输入数据进行卷积处理以得到卷积特征图;对所述卷积特征图进行基于特征矩阵的均值池化处理以得到池化特征图;以及,对所述池化特征图进行非线性激活以得到激活特征图;其中,所述作为过滤器的第一卷积神经网络模型的最后一层的输出为所述第一目标声源波形特征向量,所述作为过滤器的第一卷积神经网络模型的第一层的输入为所述第一目标声源信号的波形图。

[0014] 在上述全向级联麦克风阵列降噪方法中,所述将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵,包括:使用所述作为特征提取器的第二卷积神经网络模型的各层在层的正向传递中分别对输入数据进行处理:对所述输入数据进行卷积处理以得到卷积特征图;对所述卷积特征图进行沿通道维度的池化处理以得到池化特征图;以及,对所述池化特征图进行非线性激活以得到激活特征图;其中,所述作为特征提取器的第二卷积神经网络模型的最后一层的输出为所述协同声音波形特征矩阵,所述作为特征提取器的第二卷积神经网络模型的第一层的输入为所述多通道声音波形图。

[0015] 在上述全向级联麦克风阵列降噪方法中,所述作为特征提取器的第二卷积神经网络模型为深度残差网络。

[0016] 在上述全向级联麦克风阵列降噪方法中,所述对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量,包括:以如下公式对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量;其中,所述公式为:

$$[0017] \quad V' = V \otimes (\|V\|_2^2 \odot V_{order} \otimes V^T);$$

[0018] 其中, $V$ 表示所述校正第一目标声源波形特征向量, $V'$ 表示所述解码特征向量, $\|V\|_2^2$ 表示所述校正第一目标声源波形特征向量的二范数的平方, $V_{order}$ 表示所述校正第一目标声源波形特征向量的各特征值按大小次序排列的有序向量,且所述校正第一目标声源波形特征向量 $V$ 是列向量形式, $V^T$ 表示 $V$ 的转置矩阵, $\odot$ 表示向量点乘, $\otimes$ 表示矩阵乘法。

[0019] 在上述全向级联麦克风阵列降噪方法中,所述对抗生成网络包括鉴别器和生成器。

[0020] 在上述全向级联麦克风阵列降噪方法中,所述将所述解码特征向量通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号,包括:将所述解码特征向量输入所述基于对抗生成网络的声音信号生成器以由所述对抗生成网络的生成器通过至少两次反卷积编码生成所述降噪后第一目标声源信号。

[0021] 根据本发明的另一个方面,提供了一种全向级联麦克风阵列降噪系统,其包括:

[0022] 信号获取模块,用于获取由第一全向麦克风单元采集的第一目标声源信号和由第

二全向麦克风单元采集的第二目标声源信号,所述第一全向麦克风单元和所述第二全向麦克风单元相互级联;

[0023] 第一特征提取模块,用于将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量;

[0024] 第二特征提取模块,用于将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵;

[0025] 校正模块,用于以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量;

[0026] 数据流形局部优化模块,用于对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量;以及

[0027] 声音信号生成模块,用于将所述解码特征向量通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号。

[0028] 在上述全向级联麦克风阵列降噪系统中,所述第一特征提取模块,进一步用于:使用所述作为过滤器的第一卷积神经网络模型的各层在层的正向传递中分别对输入数据进行处理:对所述输入数据进行卷积处理以得到卷积特征图;对所述卷积特征图进行基于特征矩阵的均值池化处理以得到池化特征图;以及,对所述池化特征图进行非线性激活以得到激活特征图;其中,所述作为过滤器的第一卷积神经网络模型的最后一层的输出为所述第一目标声源波形特征向量,所述作为过滤器的第一卷积神经网络模型的第一层的输入为所述第一目标声源信号的波形图。

[0029] 在上述全向级联麦克风阵列降噪系统中,所述第二特征提取模块,进一步用于:使用所述作为特征提取器的第二卷积神经网络模型的各层在层的正向传递中分别对输入数据进行处理:对所述输入数据进行卷积处理以得到卷积特征图;对所述卷积特征图进行沿通道维度的池化处理以得到池化特征图;以及,对所述池化特征图进行非线性激活以得到激活特征图;其中,所述作为特征提取器的第二卷积神经网络模型的最后一层的输出为所述协同声音波形特征矩阵,所述作为特征提取器的第二卷积神经网络模型的第一层的输入为所述多通道声音波形图。

[0030] 与现有技术相比,本发明提供的全向级联麦克风阵列降噪方法及系统,其获取由第一全向麦克风单元采集的第一目标声源信号和由第二全向麦克风单元采集的第二目标声源信号;采用基于深度学习的人工智能技术,基于所述第一目标声源信号和所述第二目标声源信号在高维特征空间中的全时序空间联立表示为上下文语义背景,将其映射到第一目标声源的高维特征空间中以进行特征分布校正,并基于对抗生成思想生成降噪后第一目标声源信号。这样,可以增大降噪后第一目标声源信号的准确性,以提高噪声消除效果。

## 附图说明

[0031] 通过结合附图对本发明实施例进行更详细的描述,本发明的上述以及其他目的、特征和优势将变得更加明显。附图用来提供对本发明实施例的进一步理解,并且构成说明书的一部分,与本发明实施例一起用于解释本发明,并不构成对本发明的限制。在附图中,相同的参考标号通常代表相同部件或步骤。

- [0032] 图1为根据本发明实施例的全向级联麦克风阵列降噪方法的场景示意图。
- [0033] 图2为根据本发明实施例的全向级联麦克风阵列降噪方法的流程图。
- [0034] 图3为根据本发明实施例的全向级联麦克风阵列降噪方法的架构示意图。
- [0035] 图4为根据本发明实施例的全向级联麦克风阵列降噪系统的框图。

### 具体实施方式

[0036] 下面,将参考附图详细地描述根据本发明的示例实施例。显然,所描述的实施例仅仅是本发明的一部分实施例,而不是本发明的全部实施例,应理解,本发明不受这里描述的示例实施例的限制。

#### [0037] 场景概述

[0038] 如上所述,现有的小距离双麦克风阵列算法能够正常工作的一个前提条件是:两个麦克风的匹配性很好。但是在实际应用中,这一点是不能完全保证的,即使是同批次的普通麦克风,其幅频特性有时相差会在3-4db左右,且随着时间的推移和外界环境温度、湿度的变化,麦克风的幅频特性也会发生改变,因此当两个麦克风不完全匹配时,就会大大影响噪声消除效果,导致影响语音质量。因此,期待一种优化的用于麦克风阵列降噪方案。

[0039] 近年来,深度学习以及神经网络已经广泛应用于计算机视觉、自然语言处理、文本信号处理等领域。深度学习以及神经网络的发展,为麦克风阵列降噪提供了新的解决思路和方案。

[0040] 具体地,在本发明的技术方案中,考虑到在麦克风阵列接收声源信号时,所述麦克风阵列的第一全向麦克风单元和第二全向麦克风单元都会接收到目标声源信号,为了便于描述,定义为第一目标声源信号和第二目标声源信号。而所述第一目标声源信号和所述第二目标声源信号之间的声音特征协同可用于降噪。

[0041] 相应地,首先将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量。也就是,将所述第一目标声源信号的波形图视为一个二维图数据,并使用在图像特征提取领域具有优异性能表现的卷积神经网络模型作为特征提取器来提取所述第一目标声源信号中的高维局部隐含特征,即,所述第一目标声源信号中的有效声音特征。

[0042] 同时,将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵。也就是,在信号源域端,将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图沿着通道维度进行聚合,并同样使用在图像特征提取领域具有优异性能表现的卷积神经网络模型提取所述第一目标声源信号和所述第二目标声源信号的有效声音特征,以及,所述第一目标声源信号和所述第二目标声源信号的基于通道维度的高维语义关联特征。在本发明一个具体的示例中,所述作为特征提取器的第二卷积神经网络模型为深度残差网络。

[0043] 进而,以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量。也就是,基于所述第一目标声源信号和所述第二目标声源信号在高维特征空间中的全时序空间联立表示为上下文语义背景,并基于矩阵相乘的方式将其映射到所述第一目标声源波形特征向量的高

维特征空间中以得到所述校正第一目标声源波形特征向量。继而,以所述校正第一目标声源波形特征向量作为解码特征向量,并通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号。也就是,基于对抗生成思想来生成所述降噪后第一目标声源信号。

[0044] 在本发明的技术方案中,通过以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量,可以基于所述第一目标声源信号和所述第二目标声源信号在高维特征空间中的全时序空间联立表示为上下文语义背景,并基于矩阵相乘的方式将其映射到所述第一目标声源波形特征向量的高维特征空间中以得到所述校正第一目标声源波形特征向量以进行特征分布校正。但是,由于所述协同声音波形特征矩阵本身表达的是所述第一目标声源信号和所述第二目标声源信号的图像波形时序关联特征的高阶关联表达,将其映射到所述第一目标声源波形特征向量的低阶图像波形时序关联空间后,可能导致所述校正第一目标声源波形特征向量的特征分布的结构模糊,从而降低所述校正第一目标声源波形特征向量的表达确定性,影响所述校正第一目标声源波形特征向量通过基于对抗生成网络的声音信号生成器得到的降噪后第一目标声源信号的准确性。

[0045] 基于此,本发明对所述校正第一目标声源波形特征向量进行向量的有序希尔伯特完备化,表示为:

$$[0046] \quad V' = V \otimes (\|V\|_2^2 \odot V_{order} \otimes V^T);$$

[0047] 其中, $V$ 表示所述校正第一目标声源波形特征向量, $V'$ 表示所述解码特征向量, $\|V\|_2^2$ 表示所述校正第一目标声源波形特征向量的二范数的平方, $V_{order}$ 表示所述校正第一目标声源波形特征向量的各特征值按大小次序排列的有序向量,且所述校正第一目标声源波形特征向量 $V$ 是列向量形式, $V^T$ 表示 $V$ 的转置矩阵。

[0048] 这里,通过将有序向量映射到由向量的自内积定义的希尔伯特空间内,可以实现特征集合的数值关系在一致性空间内的有意义的度量,以此为基础通过与特征向量的相对位置嵌入构造具有正交性结构的特征空间,并基于向量查询对特征向量的高维流形进行特征空间内的结构完备化,可以避免由于模糊化结构降低特征向量的表达确定性,从而增大所述校正第一目标声源波形特征向量通过基于对抗生成网络的声音信号生成器得到的降噪后第一目标声源信号的准确性。

[0049] 基于此,本发明提供了一种全向级联麦克风阵列降噪方法,其包括:获取由第一全向麦克风单元采集的第一目标声源信号和由第二全向麦克风单元采集的第二目标声源信号,所述第一全向麦克风单元和所述第二全向麦克风单元相互级联;将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量;将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵;以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量;对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量;以及,将所述解码特征向量通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号。

[0050] 图1为根据本发明实施例的全向级联麦克风阵列降噪方法的场景示意图。如图1所

示,在该应用场景中,首先,获取由第一全向麦克风单元采集的第一目标声源信号(例如,如图1中所示意的C1)和由第二全向麦克风单元采集的第二目标声源信号(例如,如图1中所示意的C2);然后,将获取的第一目标声源信号和第二目标声源信号输入至部署有全向级联麦克风阵列降噪算法的服务器(例如,如图1中所示意的S)中,其中所述服务器能够基于全向级联麦克风阵列降噪算法对所述第一目标声源信号和所述第二目标声源信号进行处理,以生成降噪后第一目标声源信号。

[0051] 在介绍了本发明的基本原理之后,下面将参考附图来具体介绍本发明的各种非限制性实施例。

[0052] 示例性方法

[0053] 图2为根据本发明实施例的全向级联麦克风阵列降噪方法的流程图。如图2所示,根据本发明实施例的全向级联麦克风阵列降噪方法,包括步骤:S110,获取由第一全向麦克风单元采集的第一目标声源信号和由第二全向麦克风单元采集的第二目标声源信号,所述第一全向麦克风单元和所述第二全向麦克风单元相互级联;S120,将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量;S130,将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵;S140,以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量;S150,对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量;以及,S160,将所述解码特征向量通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号。

[0054] 图3为根据本发明实施例的全向级联麦克风阵列降噪方法的架构示意图。如图3所示,在该网络架构中,首先,获取由第一全向麦克风单元采集的第一目标声源信号和由第二全向麦克风单元采集的第二目标声源信号,所述第一全向麦克风单元和所述第二全向麦克风单元相互级联;然后,将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量;接着,将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵;然后,以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量;接着,对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量;以及,最后,将所述解码特征向量通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号。

[0055] 具体地,在步骤S110中,获取由第一全向麦克风单元采集的第一目标声源信号和由第二全向麦克风单元采集的第二目标声源信号,所述第一全向麦克风单元和所述第二全向麦克风单元相互级联。

[0056] 如前所述,现有的小距离双麦克风阵列算法能够正常工作的一个前提条件是:两个麦克风的匹配性很好。但是在实际应用中,这一点是不能完全保证的,即使是同批次的普通麦克风,其幅频特性有时相差会在3-4db左右,且随着时间的推移和外界环境温度、湿度的变化,麦克风的幅频特性也会发生改变,因此当两个麦克风不完全匹配时,就会大大影响

噪声消除效果,导致影响语音质量。因此,期待一种优化的用于麦克风阵列降噪方案。

[0057] 近年来,深度学习以及神经网络已经广泛应用于计算机视觉、自然语言处理、文本信号处理等领域。深度学习以及神经网络的发展,为麦克风阵列降噪提供了新的解决思路和方案。

[0058] 具体地,在本发明的技术方案中,考虑到在麦克风阵列接收声源信号时,所述麦克风阵列的第一全向麦克风单元和第二全向麦克风单元都会接收到目标声源信号,为了便于描述,定义为第一目标声源信号和第二目标声源信号。而所述第一目标声源信号和所述第二目标声源信号之间的声音特征协同可用于降噪。

[0059] 具体地,在步骤S120中,将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量。相应地,首先将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量。也就是,将所述第一目标声源信号的波形图视为一个二维图数据,并使用在图像特征提取领域具有优异性能表现的卷积神经网络模型作为特征提取器来提取所述第一目标声源信号中的高维局部隐含特征,即,所述第一目标声源信号中的有效声音特征。

[0060] 其中,所述将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量,包括:使用所述作为过滤器的第一卷积神经网络模型的各层在层的正向传递中分别对输入数据进行处理:对所述输入数据进行卷积处理以得到卷积特征图;对所述卷积特征图进行基于特征矩阵的均值池化处理以得到池化特征图;以及,对所述池化特征图进行非线性激活以得到激活特征图;其中,所述作为过滤器的第一卷积神经网络模型的最后一层的输出为所述第一目标声源波形特征向量,所述作为过滤器的第一卷积神经网络模型的第一层的输入为所述第一目标声源信号的波形图。

[0061] 具体地,在步骤S130中,将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵。同时,将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵。

[0062] 也就是,在信号源域端,将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图沿着通道维度进行聚合,并同样使用在图像特征提取领域具有优异性能表现的卷积神经网络模型提取所述第一目标声源信号和所述第二目标声源信号的有效声音特征,以及,所述第一目标声源信号和所述第二目标声源信号的基于通道维度的高维语义关联特征。在本发明一个具体的示例中,所述作为特征提取器的第二卷积神经网络模型为深度残差网络。

[0063] 其中,所述将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵,包括:使用所述作为特征提取器的第二卷积神经网络模型的各层在层的正向传递中分别对输入数据进行处理:对所述输入数据进行卷积处理以得到卷积特征图;对所述卷积特征图进行沿通道维度的池化处理以得到池化特征图;以及,对所述池化特征图进行非线性激活以得到激活特征图;其中,所述作为特征提取器的第二卷积神经网络模型的最后一层的输出为所述协同声音波形特征矩阵,所述作为特征提取器的第

二卷积神经网络模型的第一层的输入为所述多通道声音波形图。

[0064] 具体地,在步骤S140中,以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量。

[0065] 进而,以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量。也就是,基于所述第一目标声源信号和所述第二目标声源信号在高维特征空间中的全时序空间联立表示为上下文语义背景,并基于矩阵相乘的方式将其映射到所述第一目标声源波形特征向量的高维特征空间中以得到所述校正第一目标声源波形特征向量。

[0066] 具体地,在步骤S150中,对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量。在本发明的技术方案中,通过以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量,可以基于所述第一目标声源信号和所述第二目标声源信号在高维特征空间中的全时序空间联立表示为上下文语义背景,并基于矩阵相乘的方式将其映射到所述第一目标声源波形特征向量的高维特征空间中以得到所述校正第一目标声源波形特征向量以进行特征分布校正。但是,由于所述协同声音波形特征矩阵本身表达的是所述第一目标声源信号和所述第二目标声源信号的图像波形时序关联特征的高阶关联表达,将其映射到所述第一目标声源波形特征向量的低阶图像波形时序关联空间后,可能导致所述校正第一目标声源波形特征向量的特征分布的结构模糊,从而降低所述校正第一目标声源波形特征向量的表达确定性,影响所述校正第一目标声源波形特征向量通过基于对抗生成网络的语音信号生成器得到的降噪后第一目标声源信号的准确性。

[0067] 基于此,本发明对所述校正第一目标声源波形特征向量进行向量的有序希尔伯特完备化,也就是,所述对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量,包括:以如下公式对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量;其中,所述公式为:

$$[0068] \quad V' = V \otimes (\|V\|_2^2 \odot V_{order} \otimes V^T);$$

[0069] 其中, $V$ 表示所述校正第一目标声源波形特征向量, $V'$ 表示所述解码特征向量, $\|V\|_2^2$ 表示所述校正第一目标声源波形特征向量的二范数的平方, $V_{order}$ 表示所述校正第一目标声源波形特征向量的各特征值按大小次序排列的有序向量,且所述校正第一目标声源波形特征向量 $V$ 是列向量形式, $V^T$ 表示 $V$ 的转置矩阵, $\odot$ 表示向量点乘, $\otimes$ 表示矩阵乘法。

[0070] 这里,通过将有序向量映射到由向量的自内积定义的希尔伯特空间内,可以实现特征集合的数值关系在一致性空间内的有意义的度量,以此为基础通过与特征向量的相对位置嵌入构造具有正交性结构的特征空间,并基于向量查询对特征向量的高维流形进行特征空间内的结构完备化,可以避免由于模糊化结构降低特征向量的表达确定性,从而增大所述校正第一目标声源波形特征向量通过基于对抗生成网络的语音信号生成器得到的降噪后第一目标声源信号的准确性。

[0071] 具体地,在步骤S160中,将所述解码特征向量通过基于对抗生成网络的语音信号生成器以得到降噪后第一目标声源信号。继而,以所述校正第一目标声源波形特征向量作为解码特征向量,并通过基于对抗生成网络的语音信号生成器以得到降噪后第一目标声源

信号。也就是,基于对抗生成思想来生成所述降噪后第一目标声源信号。

[0072] 所述将所述解码特征向量通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号,包括:将所述解码特征向量输入所述基于对抗生成网络的声音信号生成器以由所述对抗生成网络的生成器通过至少两次反卷积编码生成所述降噪后第一目标声源信号。

[0073] 在本发明一具体示例中,所述对抗生成网络包括鉴别器和生成器。

[0074] 综上,基于本发明实施例的全向级联麦克风阵列降噪方法,其获取由第一全向麦克风单元采集的第一目标声源信号和由第二全向麦克风单元采集的第二目标声源信号;采用基于深度学习的人工智能技术,基于所述第一目标声源信号和所述第二目标声源信号在高维特征空间中的全时序空间联立表示为上下文语义背景,将其映射到第一目标声源的高维特征空间中以进行特征分布校正,并基于对抗生成思想生成降噪后第一目标声源信号。这样,可以增大降噪后第一目标声源信号的准确性,以提高噪声消除效果。

[0075] 示例性系统

[0076] 图4为根据本发明实施例的全向级联麦克风阵列降噪系统的框图。如图4所示,根据本发明实施例的全向级联麦克风阵列降噪系统100,包括:信号获取模块110,用于获取由第一全向麦克风单元采集的第一目标声源信号和由第二全向麦克风单元采集的第二目标声源信号,所述第一全向麦克风单元和所述第二全向麦克风单元相互级联;第一特征提取模块120,用于将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量;第二特征提取模块130,用于将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵;校正模块140,用于以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量;数据流形局部优化模块150,用于对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量;以及,声音信号生成模块160,用于将所述解码特征向量通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号。

[0077] 在一个示例中,在上述全向级联麦克风阵列降噪系统100中,所述第一特征提取模块,进一步用于:使用所述作为过滤器的第一卷积神经网络模型的各层在层的正向传递中分别对输入数据进行处理:对所述输入数据进行卷积处理以得到卷积特征图;对所述卷积特征图进行基于特征矩阵的均值池化处理以得到池化特征图;以及,对所述池化特征图进行非线性激活以得到激活特征图;其中,所述作为过滤器的第一卷积神经网络模型的最后一层的输出为所述第一目标声源波形特征向量,所述作为过滤器的第一卷积神经网络模型的第一层的输入为所述第一目标声源信号的波形图。

[0078] 在本发明的技术方案中,考虑到在麦克风阵列接收声源信号时,所述麦克风阵列的第一全向麦克风单元和第二全向麦克风单元都会接收到目标声源信号,为了便于描述,定义为第一目标声源信号和第二目标声源信号。而所述第一目标声源信号和所述第二目标声源信号之间的声音特征协同可用于降噪。

[0079] 相应地,首先将所述第一目标声源信号的波形图通过作为过滤器的第一卷积神经网络模型以得到第一目标声源波形特征向量。也就是,将所述第一目标声源信号的波形图

视为一个二维图数据,并使用在图像特征提取领域具有优异性能表现的卷积神经网络模型作为特征提取器来提取所述第一目标声源信号中的高维局部隐含特征,即,所述第一目标声源信号中的有效声音特征。

[0080] 在一个示例中,在上述全向级联麦克风阵列降噪系统100中,所述第二特征提取模块,进一步用于:使用所述作为特征提取器的第二卷积神经网络模型的各层在层的正向传递中分别对输入数据进行处理:对所述输入数据进行卷积处理以得到卷积特征图;对所述卷积特征图进行沿通道维度的池化处理以得到池化特征图;以及,对所述池化特征图进行非线性激活以得到激活特征图;其中,所述作为特征提取器的第二卷积神经网络模型的最后一层的输出为所述协同声音波形特征矩阵,所述作为特征提取器的第二卷积神经网络模型的第一层的输入为所述多通道声音波形图。

[0081] 同时,将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图按照通道维度聚合为多通道声音波形图后通过作为特征提取器的第二卷积神经网络模型以得到协同声音波形特征矩阵。也就是,在信号源域端,将所述第一目标声源信号的波形图和所述第二目标声源信号的波形图沿着通道维度进行聚合,并同样使用在图像特征提取领域具有优异性能表现的卷积神经网络模型提取所述第一目标声源信号和所述第二目标声源信号的有效声音特征,以及,所述第一目标声源信号和所述第二目标声源信号的基于通道维度的高维语义关联特征。

[0082] 在一个示例中,在上述全向级联麦克风阵列降噪系统100中,所述作为特征提取器的第二卷积神经网络模型为深度残差网络。

[0083] 在本发明的技术方案中,通过以所述第一目标声源波形特征向量作为查询特征向量,计算其与所述协同声音波形特征矩阵之间的乘积以得到校正第一目标声源波形特征向量,可以基于所述第一目标声源信号和所述第二目标声源信号在高维特征空间中的全时序空间联立表示为上下文语义背景,并基于矩阵相乘的方式将其映射到所述第一目标声源波形特征向量的高维特征空间中以得到所述校正第一目标声源波形特征向量以进行特征分布校正。但是,由于所述协同声音波形特征矩阵本身表达的是所述第一目标声源信号和所述第二目标声源信号的图像波形时序关联特征的高阶关联表达,将其映射到所述第一目标声源波形特征向量的低阶图像波形时序关联空间后,可能导致所述校正第一目标声源波形特征向量的特征分布的结构模糊,从而降低所述校正第一目标声源波形特征向量的表达确定性,影响所述校正第一目标声源波形特征向量通过基于对抗生成网络的声音信号生成器得到的降噪后第一目标声源信号的准确性。

[0084] 在一个示例中,在上述全向级联麦克风阵列降噪系统100中,所述数据流形局部优化模块,进一步用于:以如下公式对所述校正第一目标声源波形特征向量进行数据流形局部优化以得到解码特征向量;其中,所述公式为:

$$[0085] \quad V' = V \otimes (\|V\|_2^2 \odot V_{order} \otimes V^T);$$

[0086] 其中, $V$ 表示所述校正第一目标声源波形特征向量, $V'$ 表示所述解码特征向量, $\|V\|_2^2$ 表示所述校正第一目标声源波形特征向量的二范数的平方, $V_{order}$ 表示所述校正第一目标声源波形特征向量的各特征值按大小次序排列的有序向量,且所述校正第一目标声源波形特征向量 $V$ 是列向量形式, $V^T$ 表示 $V$ 的转置矩阵, $\odot$ 表示向量点乘, $\otimes$ 表示矩阵乘法。

[0087] 这里,通过将有序向量映射到由向量的自内积定义的希尔伯特空间内,可以实现特征集合的数值关系在一致性空间内的有意义的度量,以此为基础通过与特征向量的相对位置嵌入构造具有正交性结构的特征空间,并基于向量查询对特征向量的高维流形进行特征空间内的结构完备化,可以避免由于模糊化结构降低特征向量的表达确定性,从而增大所述校正第一目标声源波形特征向量通过基于对抗生成网络的声音信号生成器得到的降噪后第一目标声源信号的准确性。

[0088] 在一个示例中,在上述全向级联麦克风阵列降噪系统100中,所述对抗生成网络包括鉴别器和生成器。

[0089] 在一个示例中,在上述全向级联麦克风阵列降噪系统100中,所述声音信号生成模块,用于:将所述解码特征向量输入所述基于对抗生成网络的声音信号生成器以由所述对抗生成网络的生成器通过至少两次反卷积编码生成所述降噪后第一目标声源信号。

[0090] 以所述校正第一目标声源波形特征向量作为解码特征向量,并通过基于对抗生成网络的声音信号生成器以得到降噪后第一目标声源信号。也就是,基于对抗生成思想来生成所述降噪后第一目标声源信号。

[0091] 这里,本领域技术人员可以理解,上述全向级联麦克风阵列降噪系统100中的各个单元和模块的具体功能和操作已经在上面参考图1到图3的全向级联麦克风阵列降噪方法的描述中得到了详细介绍,并因此,将省略其重复描述。

[0092] 以上结合具体实施例描述了本发明的基本原理,但是,需要指出的是,在本发明中提及的优点、优势、效果等仅是示例而非限制,不能认为这些优点、优势、效果等是本发明的各个实施例必须具备的。另外,上述公开的具体细节仅是为了示例的作用和便于理解的作用,而非限制,上述细节并不限制本发明为必须采用上述具体的细节来实现。

[0093] 本发明中涉及的器件、装置、设备、系统的方框图仅作为例示性的例子并且不意图要求或暗示必须按照方框图示出的方式进行连接、布置、配置。如本领域技术人员将认识到的,可以按任意方式连接、布置、配置这些器件、装置、设备、系统。诸如“包括”、“包含”、“具有”等等的词语是开放性词汇,指“包括但不限于”,且可与其互换使用。这里所使用的词汇“或”和“和”指词汇“和/或”,且可与其互换使用,除非上下文明确指示不是如此。这里所使用的词汇“诸如”指词组“诸如但不限于”,且可与其互换使用。

[0094] 还需要指出的是,在本发明的装置、设备和方法中,各部件或各步骤是可以分解和/或重新组合的。这些分解和/或重新组合应视为本发明的等效方案。

[0095] 提供所公开的方面的以上描述以使本领域的任何技术人员能够做出或者使用本发明。对这些方面的各种修改对于本领域技术人员而言是非常显而易见的,并且在此定义的一般原理可以应用于其他方面而不脱离本发明的范围。因此,本发明不意图被限制到在此示出的方面,而是按照与在此公开的原理和新颖的特征一致的最宽范围。

[0096] 为了例示和描述的目的已经给出了以上描述。此外,此描述不意图将本发明的实施例限制到在此公开的形式。尽管以上已经讨论了多个示例方面和实施例,但是本领域技术人员将认识到其某些变型、修改、改变、添加和子组合。

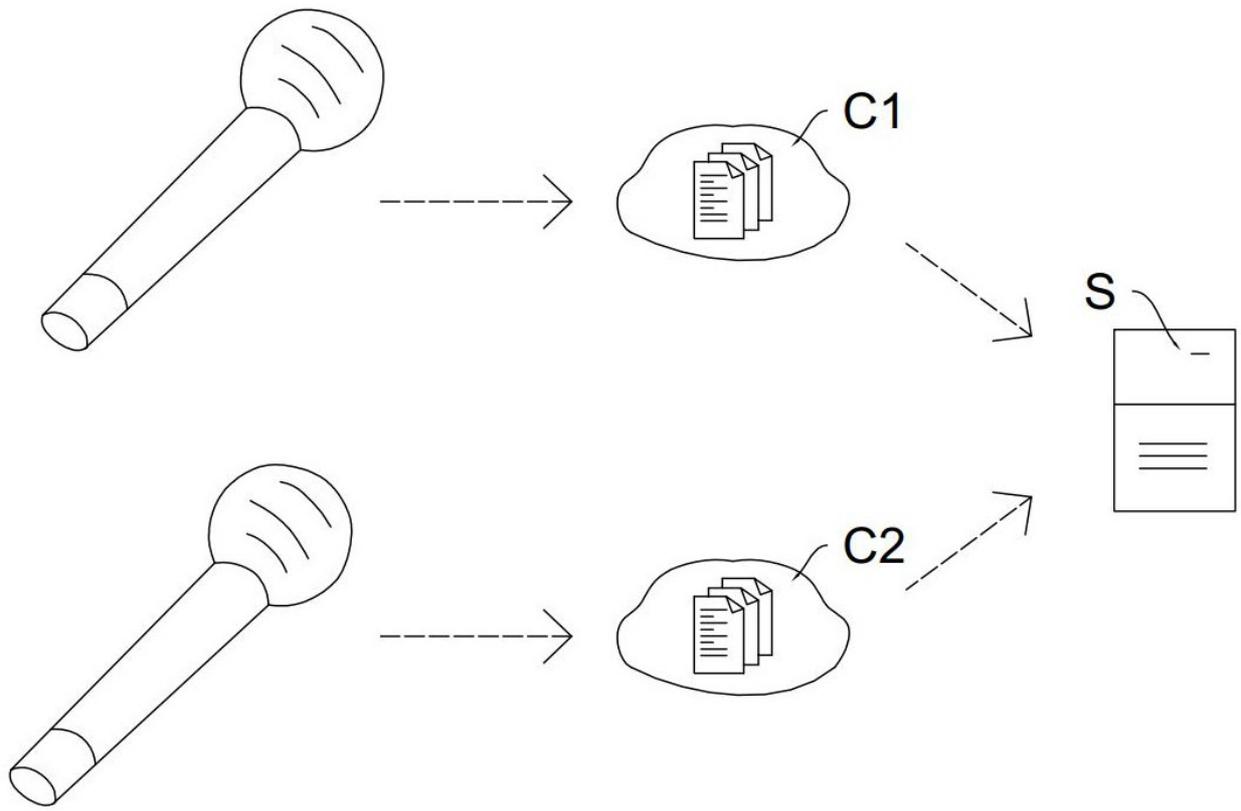


图 1

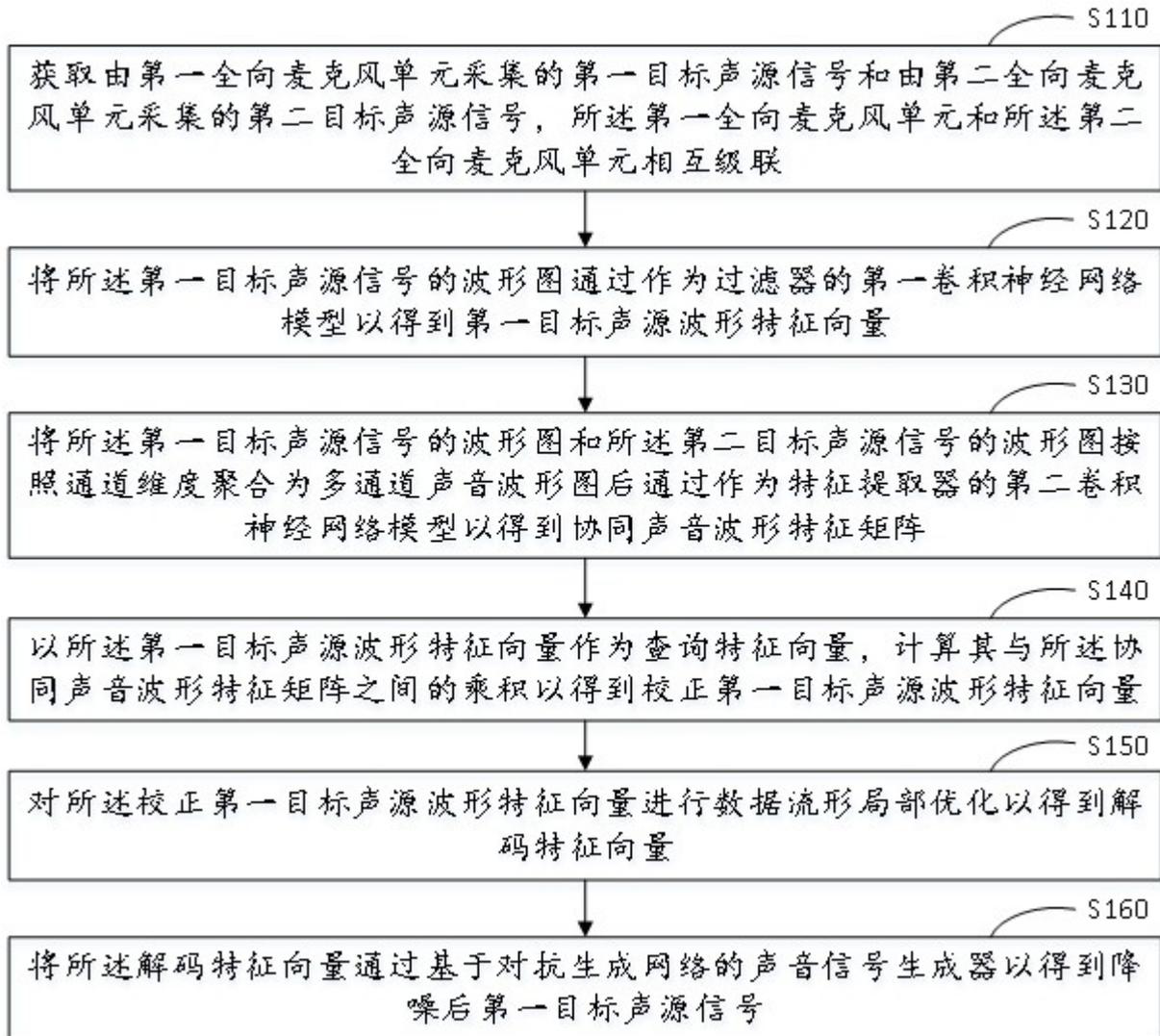


图 2

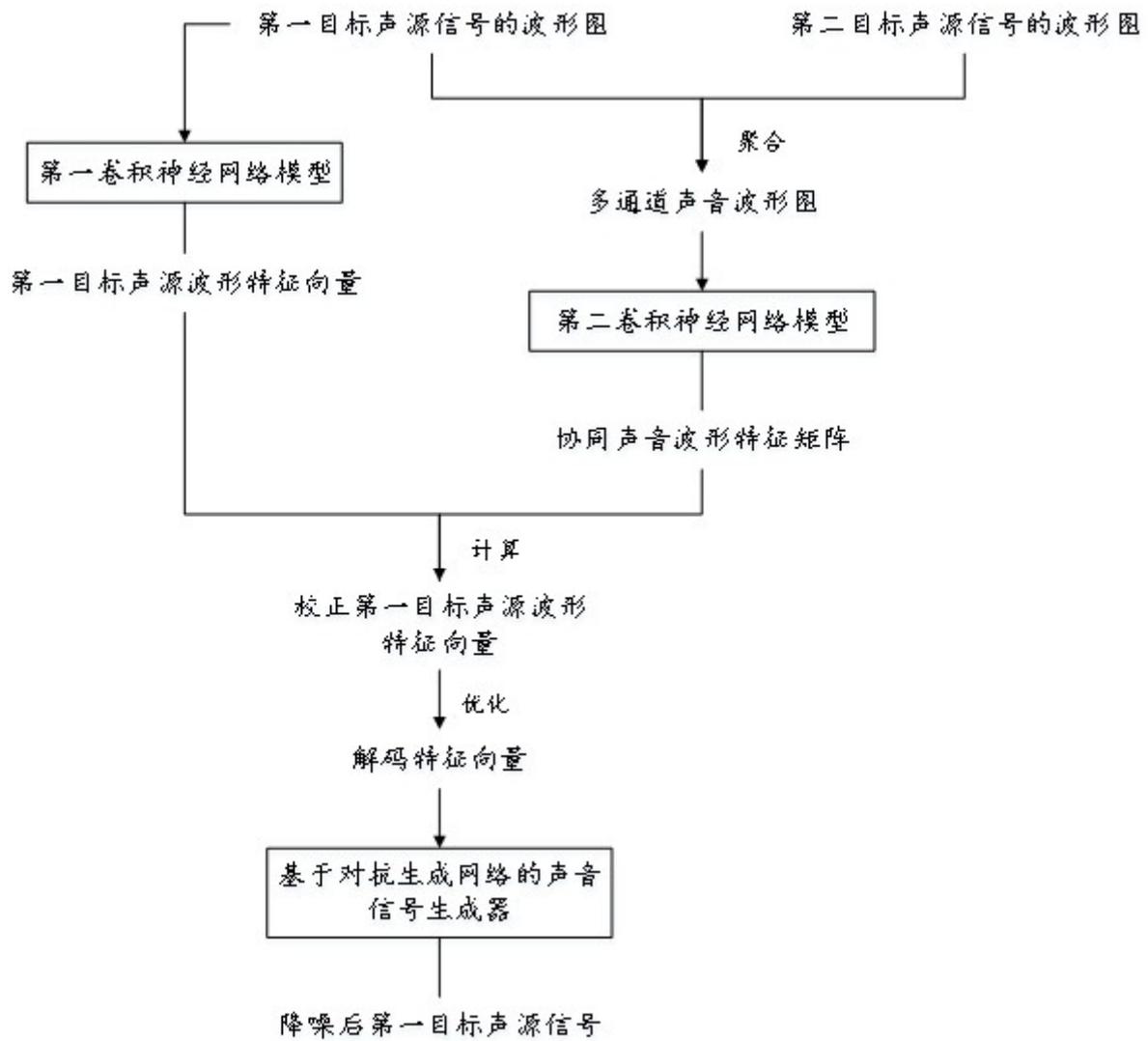


图 3

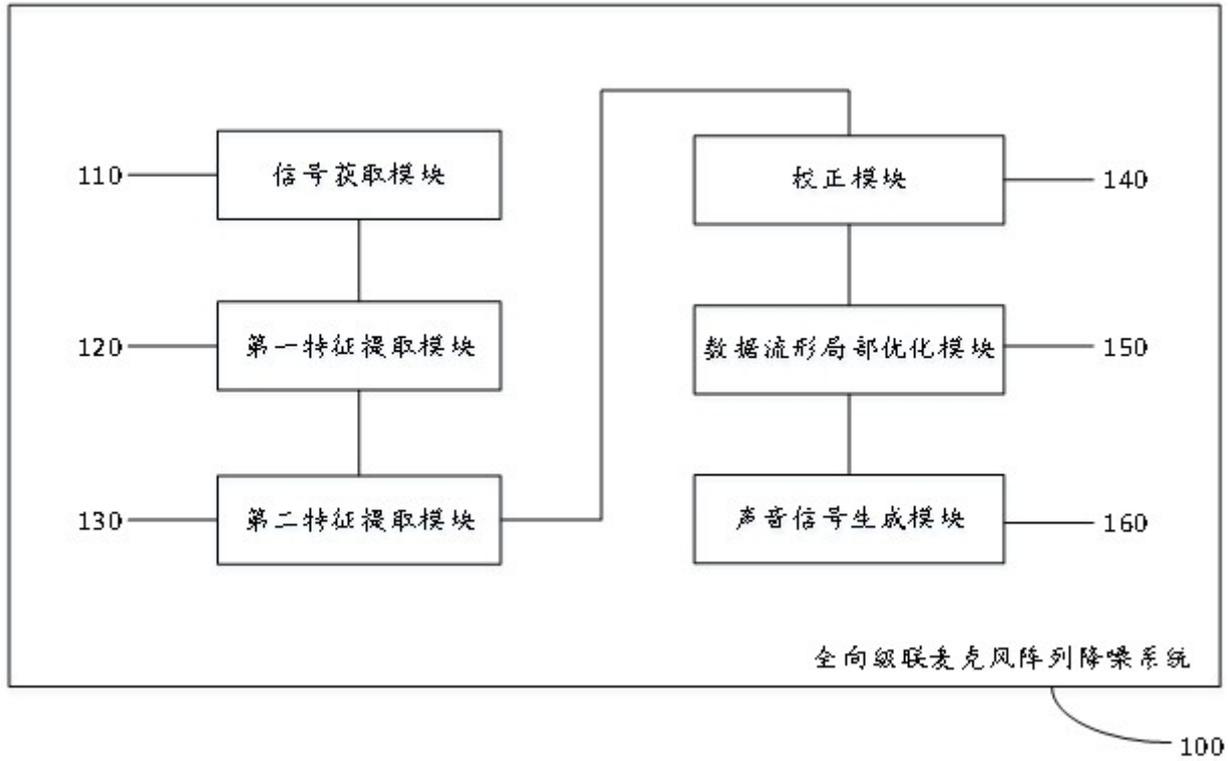


图 4