



(12) 发明专利

(10) 授权公告号 CN 110148422 B

(45) 授权公告日 2021.04.16

(21) 申请号 201910503237.4

G10L 25/18 (2013.01)

(22) 申请日 2019.06.11

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 110148422 A

CN 107219512 A, 2017.09.29

CN 107919133 A, 2018.04.17

CN 101667425 A, 2010.03.10

(43) 申请公布日 2019.08.20

CN 107102295 A, 2017.08.29

CN 105989852 A, 2016.10.05

(73) 专利权人 南京地平线集成电路有限公司
地址 210000 江苏省南京市栖霞区经济技
术开发区兴智路6号兴智科技园A栋20
层

US 2018/0182412 A1, 2018.06.28

US 2018/0366135 A1, 2018.12.20

(72) 发明人 朱长宝 李锦超

陈田田. 基于稀疏分量分析的欠定盲声源分
离技术研究. 《中国优秀硕士学位论文全文数据
库 信息科技辑》. 2018, (第02期),

(74) 专利代理机构 北京思源智汇知识产权代理
有限公司 11657

Zhenghua Wu et al. Blind Source

Separation Based On Compressed Sensing.

《2011 6th International ICST Conference

on Communications and Networking in China

(CHINACOM)》. 2011,

(51) Int. Cl.

G10L 21/0272 (2013.01)

G10L 21/0308 (2013.01)

G10L 25/51 (2013.01)

G10L 25/30 (2013.01)

审查员 庞秋婵

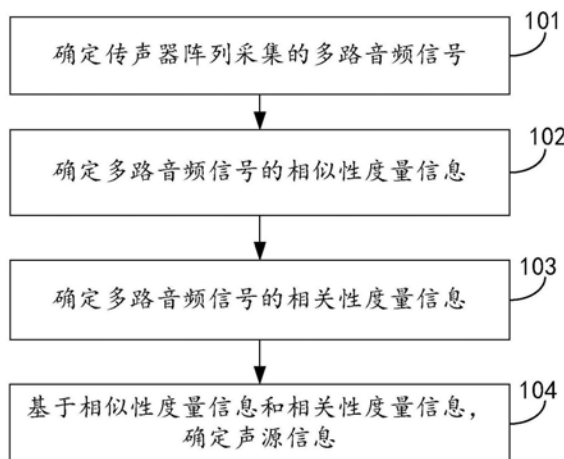
权利要求书2页 说明书17页 附图8页

(54) 发明名称

基于传声器阵列确定声源信息的方法、装置
及电子设备

(57) 摘要

本公开实施例公开了一种基于传声器阵列
确定声源信息的方法, 其中, 方法包括: 确定传声
器阵列采集的多路音频信号; 确定所述多路音频
信号的相似性度量信息; 确定所述多路音频信号
的相关性度量信息; 基于所述相似性度量信息和
所述相关性度量信息, 确定声源信息。还公开了
一种基于传声器阵列确定声源信息的装置, 其
中, 装置包括: 音频信号确定模块、相似性度量
信息确定模块、相关性度量信息确定模块和声源
信息确定模块。本公开实施例可以通过确定并基
于传声器阵列采集的多路音频信号的相似性度
量信息和相关性度量信息, 确定声源信息, 可
以通过传声器阵列得到声源信息, 方便后续进
行声源数量估计, 可以保证声源数量估计的准
确性。



1. 一种基于传声器阵列确定声源信息的方法,包括:
 - 确定传声器阵列采集的多路音频信号;
 - 确定所述多路音频信号的相似性度量信息;
 - 确定所述多路音频信号的相关性度量信息;
 - 基于所述相似性度量信息和所述相关性度量信息,确定声源信息;
 - 所述基于所述相似性度量信息和所述相关性度量信息,确定声源信息,包括:基于所述相似性度量信息和所述相关性度量信息,确定声源与所述传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵。
2. 根据权利要求1所述的方法,其中,确定所述多路音频信号的相似性度量信息,包括:
 - 确定所述多路音频信号中每路所述音频信号的盲源分离信号;
 - 确定所述盲源分离信号的频谱;
 - 基于所述盲源分离信号的频谱,确定所述盲源分离信号的相似性特征数据,其中,所述相似性特征数据为:梅尔频谱包络特征,或,频谱余弦距离;
 - 基于所述相似性特征数据,利用预设第一神经网络模型,确定所述多路音频信号的相似性度量信息。
3. 根据权利要求2所述的方法,其中,基于所述盲源分离信号的频谱,确定所述盲源分离信号的相似性特征数据,包括:
 - 确定所述盲源分离信号的频谱的盲源归一化频谱;
 - 基于所述盲源归一化频谱,确定所述盲源分离信号的相似性特征数据。
4. 根据权利要求1所述的方法,其中,确定所述多路音频信号的相关性度量信息,包括:
 - 确定所述多路音频信号中每路音频信号的频谱;
 - 基于所述每路音频信号的频谱,确定所述多路音频信号的相关性特征数据,其中,所述相关性特征数据为:广义互谱函数,或,自相关矩阵本征值;
 - 基于所述相关性特征数据,利用预设第二神经网络模型,确定所述多路音频信号的相关性度量信息。
5. 根据权利要求4所述的方法,其中,基于所述多路音频信号中每路音频信号的频谱,确定所述多路音频信号的相关性特征数据,包括:
 - 确定所述多路音频信号中每路音频信号的频谱的音频归一化频谱;
 - 基于所述音频归一化频谱,确定所述多路音频信号的相关性特征数据。
6. 根据权利要求1-5任一所述的方法,其中,基于所述相似性度量信息和所述相关性度量信息,确定声源与所述传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵,包括:
 - 基于所述相似性度量信息,确定所述相对传递函数矩阵为正定矩阵时对应的第一相似概率值、所述相对传递函数矩阵为超定矩阵时对应的第二相似概率值和所述相对传递函数矩阵为欠定矩阵时对应的第三相似概率值;
 - 基于所述相关性度量信息,确定所述相对传递函数矩阵为正定矩阵时对应的第一相关概率值、所述相对传递函数矩阵为超定矩阵时对应的第二相关概率值和所述相对传递函数矩阵为欠定矩阵时对应的第三相关概率值;
 - 根据所述第一相似概率值、所述第二相似概率值和所述第三相似概率值,以及所述第

一相关概率值、所述第二相关概率值和所述第三相关概率值,确定所述相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵。

7. 根据权利要求1-5任一所述的方法,其中,还包括:

当所述相对传递函数矩阵为正定矩阵时,基于所述传声器阵列中包括的传声器的个数确定声源数量;

当所述相对传递函数矩阵为超定矩阵时,确定所述多路音频信号的子空间的正交基个数或确定所述声源的空间谱的谱峰值个数,基于所述正交基个数或所述谱峰值个数确定声源数量,其中,所述多路音频信号的子空间为所述多路音频信号的自相关矩阵的正交量形成的空间,所述空间谱为基于所述声源的方位角设计的函数;

当所述相对传递函数矩阵为欠定矩阵时,确定所述声源的方位直方图的峰值个数,基于所述峰值个数确定声源数量。

8. 根据权利要求1所述的方法,其中,基于所述相似性度量信息和所述相关性度量信息,确定声源与所述传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵,包括:

基于所述相似性度量信息和所述相关性度量信息,利用预设第三神经网络模型,确定声源与所述传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵。

9. 根据权利要求8所述的方法,其中,还包括:

当所述相对传递函数矩阵为正定矩阵时,基于所述传声器阵列中包括的传声器的个数确定声源数量;

当所述相对传递函数矩阵为超定矩阵时,确定所述多路音频信号的子空间的正交基个数或确定所述声源的空间谱的谱峰值个数,基于所述正交基个数或所述谱峰值个数确定声源数量,其中,所述多路音频信号的子空间为所述多路音频信号的自相关矩阵的正交量形成的空间,所述空间谱为基于所述声源的方位角设计的函数;

当所述相对传递函数矩阵为欠定矩阵时,确定所述声源的方位直方图的峰值个数,基于所述峰值个数确定声源数量。

10. 一种基于传声器阵列确定声源信息的装置,包括:

音频信号确定模块,用于确定传声器阵列采集的多路音频信号;

相似性度量信息确定模块,用于确定所述多路音频信号的相似性度量信息;

相关性度量信息确定模块,用于确定所述多路音频信号的相关性度量信息;

声源信息确定模块,用于基于所述相似性度量信息和所述相关性度量信息,确定声源信息;所述声源信息确定模块具体用于基于所述相似性度量信息和所述相关性度量信息,确定声源与所述传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵。

11. 一种计算机可读存储介质,所述存储介质存储有计算机程序,所述计算机程序用于执行上述权利要求1-9任一所述的基于传声器阵列确定声源信息的方法。

12. 一种电子设备,所述电子设备包括:

处理器;

用于存储所述处理器可执行指令的存储器;

所述处理器,用于执行上述权利要求1-9任一所述的基于传声器阵列确定声源信息的方法。

基于传声器阵列确定声源信息的方法、装置及电子设备

技术领域

[0001] 本公开涉及语音信号处理技术领域,具体涉及一种基于传声器阵列确定声源信息的方法、装置及电子设备。

背景技术

[0002] 声源数量的估计对于声源的分离、识别、定位、追踪和说话者辨别等方面均起着重要的作用。目前,常利用传声器阵列进行声源数量的估计。利用传声器阵列进行声源数量的估计时,准确性难以保证。

发明内容

[0003] 为了解决上述技术问题,提出了本公开。本公开的实施例提供了一种基于传声器阵列确定声源信息的方法、装置及电子设备。

[0004] 根据本公开实施例的一个方面,提供了一种基于传声器阵列确定声源信息的方法,包括:确定传声器阵列采集的多路音频信号;确定所述多路音频信号的相似性度量信息;确定所述多路音频信号的相关性度量信息;基于所述相似性度量信息和所述相关性度量信息,确定声源信息。

[0005] 根据本公开实施例的一个方面,提供了一种基于传声器阵列确定声源信息的装置,包括:音频信号确定模块,用于确定传声器阵列采集的多路音频信号;相似性度量信息确定模块,用于确定所述多路音频信号的相似性度量信息;相关性度量信息确定模块,用于确定所述多路音频信号的相关性度量信息;声源信息确定模块,用于基于所述相似性度量信息和所述相关性度量信息,确定声源信息。

[0006] 根据本公开实施例的一个方面,提供了一种计算机可读存储介质,所述存储介质存储有计算机程序,所述计算机程序用于执行上述任一所述的基于传声器阵列确定声源信息的方法。

[0007] 根据本公开实施例的一个方面,提供了一种电子设备,所述电子设备包括:处理器;用于存储所述处理器可执行指令的存储器;所述处理器,用于执行上述任一所述的基于传声器阵列确定声源信息的方法。

[0008] 基于本公开上述实施例提供的基于传声器阵列确定声源信息的方法,通过确定传声器阵列采集的多路音频信号的相似性度量信息和相关性度量信息,基于所述相似性度量信息和所述相关性度量信息,确定声源信息,可以通过传声器阵列得到声源信息,方便后续进行声源数量估计,可以保证声源数量估计的准确性。

[0009] 下面通过附图和实施例,对本公开的技术方案做进一步的详细描述。

附图说明

[0010] 通过结合附图对本公开实施例进行更详细的描述,本公开的上述以及其他目的、特征和优势将变得更加明显。附图用来提供对本公开实施例的进一步理解,并且构成说明

书的一部分,与本公开实施例一起用于解释本公开,并不构成对本公开的限制。在附图中,相同的参考标号通常代表相同部件或步骤。

[0011] 图1是本公开一示例性实施例提供的基于传声器阵列确定声源信息的方法的流程示意图。

[0012] 图2是本公开一示例性实施例提供的确定所述多路音频信号的相似性度量信息的流程示意图。

[0013] 图3是本公开一示例性实施例提供的卷积循环神经网络的结构示意图。

[0014] 图4是本公开一示例性实施例提供的基于所述盲源分离信号的频谱,确定所述盲源分离信号的相似性特征数据的流程示意图。

[0015] 图5是本公开一示例性实施例提供的确定所述多路音频信号的相关性度量信息的流程示意图。

[0016] 图6是本公开一示例性实施例提供的基于所述多路音频信号中每路音频信号的频谱,确定所述多路音频信号的相关性特征数据的流程示意图。

[0017] 图7是本公开一示例性实施例提供的基于所述相似性度量信息和所述相关性度量信息,确定声源与所述传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵的流程示意图。

[0018] 图8是本公开另一示例性实施例提供的基于传声器阵列确定声源信息的方法的流程示意图。

[0019] 图9是本公开又一示例性实施例提供的基于传声器阵列确定声源信息的方法的流程示意图。

[0020] 图10是本公开一示例性实施例提供的基于传声器阵列确定声源信息的装置的结构示意图。

[0021] 图11是本公开一示例性实施例提供的相似性度量信息确定模块的结构示意图。

[0022] 图12是本公开一示例性实施例提供的相似性特征数据确定单元的结构示意图。

[0023] 图13是本公开另一示例性实施例提供的相关性度量信息确定模块的结构示意图。

[0024] 图14是本公开另一示例性实施例提供的相关性特征数据确定单元的结构示意图。

[0025] 图15是本公开另一示例性实施例提供的声源信息确定模块的结构示意图。

[0026] 图16是本公开另一示例性实施例提供的基于传声器阵列确定声源信息的装置的结构示意图。

[0027] 图17是本公开又一示例性实施例提供的基于传声器阵列确定声源信息的装置的结构示意图。

[0028] 图18是本公开一示例性实施例提供的电子设备的结构图。

具体实施方式

[0029] 下面,将参考附图详细地描述根据本公开的示例实施例。显然,所描述的实施例仅仅是本公开的一部分实施例,而不是本公开的全部实施例,应理解,本公开不受这里描述的示例实施例的限制。

[0030] 应注意到:除非另外具体说明,否则在这些实施例中阐述的部件和步骤的相对布置、数字表达式和数值不限制本公开的范围。

[0031] 本领域技术人员可以理解,本公开实施例中的“第一”、“第二”等术语仅用于区别不同步骤、设备或模块等,既不代表任何特定技术含义,也不表示它们之间的必然逻辑顺序。

[0032] 还应理解,在本公开实施例中,“多个”可以指两个或两个以上,“至少一个”可以指一个、两个或两个以上。

[0033] 还应理解,对于本公开实施例中提及的任一部件、数据或结构,在没有明确限定或者在前后文给出相反启示的情况下,一般可以理解为一个或多个。

[0034] 另外,本公开中术语“和/或”,仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。另外,本公开中字符“/”,一般表示前后关联对象是一种“或”的关系。

[0035] 还应理解,本公开对各个实施例的描述着重强调各个实施例之间的不同之处,其相同或相似之处可以相互参考,为了简洁,不再一一赘述。

[0036] 同时,应当明白,为了便于描述,附图中所示出的各个部分的尺寸并不是按照实际的比例关系绘制的。

[0037] 以下对至少一个示例性实施例的描述实际上仅仅是说明性的,决不作为对本公开及其应用或使用的任何限制。

[0038] 对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论,但在适当情况下,所述技术、方法和设备应当被视为说明书的一部分。

[0039] 应注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步讨论。

[0040] 本公开实施例可以应用于终端设备、计算机系统、服务器等电子设备,其可与众多其它通用或专用计算系统环境或配置一起操作。适于与终端设备、计算机系统、服务器等电子设备一起使用的众所周知的终端设备、计算系统、环境和/或配置的例子包括但不限于:个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人电脑、小型计算机系统、大型计算机系统和包括上述任何系统的分布式云计算技术环境,等等。

[0041] 终端设备、计算机系统、服务器等电子设备可以在由计算机系统执行的计算机系统可执行指令(诸如程序模块)的一般语境下描述。通常,程序模块可以包括例程、程序、目标程序、组件、逻辑、数据结构等等,它们执行特定的任务或者实现特定的抽象数据类型。计算机系统/服务器可以在分布式云计算环境中实施,分布式云计算环境中,任务是由通过通信网络链接的远程处理设备执行的。在分布式云计算环境中,程序模块可以位于包括存储设备的本地或远程计算系统存储介质上。

[0042] 申请概述

[0043] 声源数量的估计一直是音频处理的难点,目前,常利用传声器阵列进行声源数量的估计。利用传声器阵列进行声源数量估计时,声源数量估计的准确性,与声源数量相对于传声器阵列是正定、欠定、或超定有关。但目前,没有如何判定声源数量相对于传声器阵列是正定、欠定、或超定的方法,利用传声器阵列进行声源数量估计时,无法保证准确性。

[0044] 示例性方法

[0045] 图1是本公开一示例性实施例提供的基于传声器阵列确定声源信息的方法的流程

示意图。本实施例可应用在电子设备上,如图1所示,包括如下步骤:

[0046] 步骤101,确定传声器阵列采集的多路音频信号。

[0047] 其中,传声器阵列主要由一定数量的声学传感器组成,用来对声场的空间特性进行采样并处理的系统。例如:传声器阵列由2个声学传感器组成,采集得到2路音频信号,分别为 $x_1(t)$ 、 $x_2(t)$,其中,t表示时间。

[0048] 步骤102,确定多路音频信号的相似性度量信息。

[0049] 其中,相似性度量信息用于表示多路音频信号的相似性的程度。

[0050] 步骤103,确定多路音频信号的相关性度量信息。

[0051] 其中,相关性度量信息用于表示多路音频信号的相关性的程度。

[0052] 步骤104,基于相似性度量信息和相关性度量信息,确定声源信息。

[0053] 其中,声源信息是声源与传声器阵列的相关信息。

[0054] 本公开实施例所提及的基于传声器阵列确定声源信息的方法,通过确定传声器阵列采集的多路音频信号的相似性度量信息和相关性度量信息,基于相似性度量信息和相关性度量信息,确定声源信息,可以通过传声器阵列得到声源信息,方便后续进行声源数量估计,可以保证声源数量估计的准确性。

[0055] 图2是本公开一示例性实施例提供的确定多路音频信号的相似性度量信息的流程图示意图。在本公开图1所示实施例的基础上延伸出本公开图2所示实施例,下面着重叙述图2所示实施例与图1所示实施例的不同之处,相同之处不再赘述。

[0056] 如图2所示,在本公开实施例提供的基于传声器阵列确定声源信息的方法中,确定多路音频信号的相似性度量信息(即步骤102),包括:

[0057] 步骤1021,确定多路音频信号中每路音频信号的盲源分离信号。

[0058] 在一实施例中,可以利用独立成分分析等算法对传声器阵列采集的多路音频信号中每路音频信号进行盲源分离得到盲源分离信号,独立成分分析等算法的过程是:确定每路音频信号对应的解混矩阵,确定每路音频信号对应的解混矩阵与每路音频信号的乘积,得到每路音频信号的盲源分离信号。

[0059] 例如:假设有2路音频信号 $x_1(t)$ 、 $x_2(t)$,得到对应的2路盲源分离信号分别为 $s_1(t)$ 、 $s_2(t)$ 。

[0060] 步骤1022,确定盲源分离信号的频谱。

[0061] 在一实施例中,可以对盲源分离信号进行短时傅里叶变换得到盲源分离信号的频谱,例如,对2路盲源分离信号,分别进行短时傅里叶变换,得到2路盲源分离信号的频谱为 $S_1(t, f)$ 、 $S_2(t, f)$,其中,t表示时间,f表示频率。需要说明的是,本公开其它地方出现的t、f与此处含义相同,不再一一赘述。

[0062] 步骤1023,基于盲源分离信号的频谱,确定盲源分离信号的相似性特征数据。

[0063] 其中,相似性特征数据为:梅尔频谱包络特征,或,频谱余弦距离。

[0064] 在一实施例中,通过梅尔低通滤波器对盲源分离信号的频谱进行低通滤波,得到梅尔频谱包络特征,例如:2路盲源分离信号的频谱 $S_1(t, f)$ 、 $S_2(t, f)$,对应的梅尔频谱包络特征分别为 $Mel_1(t, f)$ 、 $Mel_2(t, f)$ 。

[0065] 在一实施例中,以时间为纬度,计算每2路盲源分离信号的频谱之间的余弦距离,即两向量的点乘除以二者的二范数。如2路盲源分离信号的频谱 $S_1(t, f)$ 、 $S_2(t, f)$ 的余弦距

离 $\cos(f)$ 为:

$$[0066] \quad \cos(f) = \frac{S_1(t, f) \cdot S_2(t, f)}{\|S_1(t, f)\|_2 \|S_2(t, f)\|_2}$$

[0067] 其中, $\cos(f)$ 表示余弦距离, “ \cdot ”表示点乘运算, $\|*\|_2$ 表示二范数, t, f 分别表示时间和频率。

[0068] 步骤1024, 基于相似性特征数据, 利用预设第一神经网络模型, 确定多路音频信号的相似性度量信息。

[0069] 其中, 预设第一神经网络模型可以采用分类问题中常用的各种神经网络, 对此不做具体限定。在一实施例中, 预设第一神经网络模型采用卷积循环神经网络实现, 参见图3, 卷积循环神经网络主要包括4个卷积层和1个长短时记忆层。将相似性特征数据输入卷积循环神经网络, 通过卷积循环神经网络提取相似性特征数据的静态表示、动态表示和时序表示, 得到多路各音频信号的相似性度量信息。其中, 卷积循环神经网络各个层的作用为:

[0070] 二维卷积层(convolutional layer, conv) 1~4, 均使用3*3卷积核, 用于提取相似性特征数据的静态表示、动态表示;

[0071] 最大池化层(maxpooling layer, mp) 1~3, 在尽量不丢失相似性特征数据的情况下进行降采样, 提高模型泛化能力((generalization ability) 指机器学习算法对新鲜样本的适应能力);

[0072] 重构层(reshape layer), 将上层的二维输出时频信息按时间拼接成一维向量, 以便输入时序层;

[0073] 长短时记忆层(long short-term memory layer, lstm), 用于提取相似性特征数据的时序特征;

[0074] 归一化指数函数层(softmax layer), 根据K个分类标签, 输出K*1维矢量, 其各输出值为模型判断为该对应分类标签下的概率值, 从而得到多路音频信号的相似性度量信息, 其中, K为自然数。

[0075] 其中, 各层旁边标注的数字表示这一层结构的单元数, 例如: 二维卷积层1旁边标注的数字64、最大池化层3旁边标注的数字40等。

[0076] 本公开实施例所提及的基于传声器阵列确定声源信息的方法, 通过确定多路音频信号中每路音频信号的盲源分离信号, 确定盲源分离信号的频谱, 基于盲源分离信号的频谱, 确定盲源分离信号的相似性特征数据, 基于相似性特征数据, 利用预设第一神经网络模型, 确定多路音频信号的相似性度量信息, 其中, 相似性特征数据为: 梅尔频谱包络特征, 或频谱余弦距离, 利用梅尔频谱包络特征或频谱余弦距离确定相似性度量信息作为相似性度量, 梅尔频谱包络特征或频谱余弦距离利用了语音具有辨识性的部分, 不需要假设每个频点最多只有一个主要的声源(即不需要假设频谱是稀疏的), 可以大大减少噪声和混响对结果的影响, 方便用于实际场景中, 由于神经网络模型可以提取相似性特征数据的静态表示、动态表示和时序表示, 因此大大提高了神经网络模型的泛化性能。

[0077] 图4是本公开一示例性实施例提供的基于盲源分离信号的频谱, 确定盲源分离信号的相似性特征数据的流程示意图。在本公开图2所示实施例的基础上延伸出本公开图4所示实施例, 下面着重叙述图4所示实施例与图2所示实施例的不同之处, 相同之处不再赘述。

[0078] 如图4所示, 在本公开实施例提供的基于盲源分离信号的频谱, 确定盲源分离信号

的相似性特征数据(即步骤1023),包括:

[0079] 步骤10231,确定盲源分离信号的频谱的盲源归一化频谱。

[0080] 具体地,确定盲源分离信号的频谱的盲源归一化频谱,可以采用自适应能量归一化、取对数后归一化或批归一化等方法。在一实施例中,通过对盲源分离信号的频谱进行自适应能量归一化,得到盲源归一化频谱,例如:确定盲源分离信号的频谱的幅度为 $E(t, f)$,也即盲源分离信号的频谱的能量为 $E(t, f)$,通过低通滤波器对 $E(t, f)$ 进行平滑滤波,得到平滑能量 $M(t, f)$,引入偏置因子 ϵ 、降噪因子 δ 、增益因子 α 、压缩因子 r ,得到盲源归一化频谱 $PCEN(t, f)$ 如下:

$$[0081] \quad PCEN(t, f) = \left(\frac{E(t, f)}{(\epsilon + M(t, f))^\alpha} + \delta \right)^r - \delta^r$$

[0082] 其中, ϵ 表示偏置因子、 δ 表示降噪因子、 α 表示增益因子、 r 表示压缩因子、 t 表示时间、 f 表示频率。

[0083] 需要说明的是, ϵ 、 δ 、 α 、 r 各参数的值可预先根据大数据训练结果设置,在一优选参数中,根据实验结果,设置参数为 $\epsilon = 10^{-6}$ 、 $\delta = 2$ 、 $\alpha = 0.98$ 、 $r = 0.5$ 。

[0084] 注意到, $E(t, f) / (\epsilon + M(t, f))^\alpha$ 可以表示为 $\exp(\log(E(t, f)) - \alpha \log(\epsilon + M(t, f)))$,因此整个自适应能量归一化过程可以整合为输入为 $E(t, f)$ 、 $M(t, f)$,参数为 ϵ 、 δ 、 α 、 r ,输出为 $PCEN(t, f)$ 的神经网络层。

[0085] 因为经过平滑滤波后, $E(t, f)$ 中能量高的部分会变低、能量低的部分会变高,能量 $E(t, f)$ 除以其平滑能量 $M(t, f)$ 的增益,对于能量较强的直达声部分,可以得到更大的因子,对于能量较低的混响声和噪声部分,可以得到更小的因子,也就是可以放大语音部分的直达声部分,抑制混响和噪声; $\left(\frac{E(t, f)}{(\epsilon + M(t, f))^\alpha} + \delta \right)^r - \delta^r$ 相当于对整个能量提升至某一平面进行压缩然后降回,能够减少能量中可能存在的偏置。由于可以放大语音部分的直达声部分,抑制混响和噪声,去除低频偏置,因此可以大大增强神经网络模型的鲁棒性。

[0086] 需要说明的是,为了便于描述,2路盲源分离信号的频谱 $S_1(t, f)$ 、 $S_2(t, f)$ 的盲源归一化频谱分别记作: $NS_1(t, f)$ 、 $NS_2(t, f)$ 。

[0087] 步骤10232,基于盲源归一化频谱,确定盲源分离信号的相似性特征数据。

[0088] 其中,相似性特征数据为:梅尔频谱包络特征,或,频谱余弦距离。

[0089] 在一实施例中,通过梅尔低通滤波器对盲源归一化频谱进行低通滤波,得到梅尔频谱包络特征,例如:2路盲源归一化频谱 $NS_1(t, f)$ 、 $NS_2(t, f)$,对应的梅尔频谱包络特征分别为 $Mel_{N1}(t, f)$ 、 $Mel_{N2}(t, f)$ 。

[0090] 在一实施例中,以时间为纬度,计算每2路盲源归一化频谱之间的余弦距离,即两向量的点乘除以二者的二范数。如2路盲源归一化频谱 $NS_1(t, f)$ 、 $NS_2(t, f)$ 的余弦距离 $\cos_{N1}(f)$ 为:

$$[0091] \quad \cos_{N1}(f) = \frac{NS_1(t, f) \cdot NS_2(t, f)}{\|NS_1(t, f)\|_2 \|NS_2(t, f)\|_2}$$

[0092] 其中, $\cos_{N1}(f)$ 表示余弦距离, \cdot 表示点乘运算, $\|*\|_2$ 表示二范数, t 、 f 分别表示时间和频率。

[0093] 本公开实施例所提及的基于传声器阵列确定声源信息的方法,通过确定盲源分离信号的频谱的盲源归一化频谱,基于盲源归一化频谱,确定盲源分离信号的相似性特征数据,由于归一化可以对不同能量大小的输入(分布不平衡的输入)进行调整,便于训练,且采用自适应能量归一化,可以放大语音部分的直达声部分、保留大部分主要的声学信息,可以降低能量较低的无用信息部分、抑制混响和噪声、去除低频偏置,进而增强神经网络模型的鲁棒性。

[0094] 图5是本公开一示例性实施例提供的确定多路音频信号的相关性度量信息的流程示意图。在本公开图1所示实施例的基础上延伸出本公开图5所示实施例,下面着重叙述图5所示实施例与图1所示实施例的不同之处,相同之处不再赘述。

[0095] 如图5所示,在本公开实施例提供的确定多路音频信号的相关性度量信息(即步骤103),包括:

[0096] 步骤1031,确定多路音频信号中每路音频信号的频谱。

[0097] 在一实施例中,可以对多路音频信号中每路音频信号进行短时傅里叶变换得到每路音频信号的频谱,例如,假设有2路音频信号 $x_1(t)$ 、 $x_2(t)$,对2路音频信号 $x_1(t)$ 、 $x_2(t)$,分别进行短时傅里叶变换,得到2路音频信号 $x_1(t)$ 、 $x_2(t)$ 的频谱为 $X_1(t, f)$ 、 $X_2(t, f)$,其中, t 表示时间, f 表示频率。

[0098] 步骤1032,基于多路音频信号中每路音频信号的频谱,确定多路音频信号的相关性特征数据。

[0099] 其中,相关性特征数据为:广义互谱函数,或自相关矩阵本征值。

[0100] 在一实施例中,可以利用音频信号的互功率谱进行相位变换加权,得到广义互谱函数。如2路音频信号的频谱为 $X_1(t, f)$ 、 $X_2(t, f)$,2路音频信号的频谱为 $X_1(t, f)$ 、 $X_2(t, f)$ 的广义互谱函数GCC(t, f)为:

$$[0101] \quad \text{GCC}(t, f) = \frac{X_1(t, f) \cdot X_2^*(t, f)}{\|X_1(t, f)\|_2 \|X_2(t, f)\|_2}$$

[0102] 其中,*表示共轭运算, $\|\cdot\|_2$ 表示二范数, t 表示时间, f 表示频率。

[0103] 需要说明的是,当音频信号大于2路时,对多路音频信号两两计算一次广义互谱函数。

[0104] 在一实施例中,可以通过对信号自相关矩阵进行本征值分解,得到自相关矩阵本征值。如2路音频信号的频谱为 $X_1(t, f)$ 、 $X_2(t, f)$, $X_1(t, f)$ 、 $X_2(t, f)$ 的自相关矩阵本征值 $\lambda_1(t, f)$ 、 $\lambda_2(t, f)$ 的计算步骤为:

[0105] 1) 计算每个时频点(n, m)的自功率谱 $P_{11}(n, m)$ 、 $P_{22}(n, m)$,以及互功率谱 $P_{12}(n, m)$ 、 $P_{21}(n, m)$:

$$[0106] \quad P_{11}(n, m) = X_1(n, m) X_1^*(n, m), \quad P_{22}(n, m) = X_2(n, m) X_2^*(n, m);$$

$$[0107] \quad P_{12}(n, m) = X_1(n, m) X_2^*(n, m), \quad P_{21}(n, m) = X_2(n, m) X_1^*(n, m)。$$

[0108] 2) 计算自相关矩阵 $R(n, m) = \begin{bmatrix} P_{11}(n, m) & P_{12}(n, m) \\ P_{21}(n, m) & P_{22}(n, m) \end{bmatrix}$ 的两个特征值 $\lambda_1(n, m)$

和 $\lambda_2(n, m)$,其中 $\forall(n, m) \in (t, f)$,从而得到自相关矩阵本征值 $\lambda_1(t, f)$ 、 $\lambda_2(t, f)$ 。

[0109] 需要说明的是,对于包含M个传声器的传声器阵列,计算每个时频点(n,m)的自相关矩阵为 $R(n, m) = E[\vec{x}(n, m)\vec{x}(n, m)^H]$,其中 $E[\]$ 为计算期望,

$\vec{x}(n, m) = [X_1(n, m), X_2(n, m), \dots, X_M(n, m)]^T$,T表示转置。由此可得M个自相关矩阵本征值 $\lambda_1(t, f), \dots, \lambda_M(t, f)$ 。

[0110] 步骤1033,基于相关性特征数据,利用预设第二神经网络模型,确定多路音频信号的相关性度量信息。

[0111] 其中,预设第二神经网络模型可以采用分类问题中常用的各种神经网络,对此不做具体限定。步骤1033与步骤1024类似,详见步骤1024的描述,此处不再赘述。

[0112] 本公开实施例所提及的基于传声器阵列确定声源信息的方法,通过确定多路音频信号中每路音频信号的频谱,基于多路音频信号中每路音频信号的频谱,确定多路音频信号的相关性特征数据,基于相关性特征数据,利用预设第二神经网络模型,确定多路音频信号的相关性度量信息,其中,相关性特征数据为:广义互谱函数,或自相关矩阵本征值,广义互谱函数或自相关矩阵本征值具有空间相关性,可以提供声源方位信息,因此可以结合声源方位信息确定多路音频信号的相关性度量信息,提高准确性,由于神经网络模型,可以提取相关性特征数据的静态表示、动态表示和时序表示,提高了神经网络模型的泛化性能。

[0113] 图6是本公开一示例性实施例提供的基于多路音频信号中每路音频信号的频谱,确定多路音频信号的相关性特征数据的流程示意图。在本公开图5所示实施例的基础上延伸出本公开图6所示实施例,下面着重叙述图6所示实施例与图5所示实施例的不同之处,相同之处不再赘述。

[0114] 如图6所示,在本公开实施例提供的基于传声器阵列确定声源信息的方法中,基于多路音频信号中每路音频信号的频谱,确定多路音频信号的相关性特征数据(即步骤1032),包括:

[0115] 步骤10321,确定多路音频信号中每路音频信号的频谱的音频归一化频谱。

[0116] 具体地,确定多路音频信号中每路音频信号的频谱的音频归一化频谱,可以采用自适应能量归一化、取对数后归一化或批归一化等方法。具体过程与步骤10231类似,详见步骤10231,此处不再赘述。

[0117] 为了便于描述,2路音频信号的频谱 $X_1(t, f)$ 、 $X_2(t, f)$ 的音频归一化频谱分别记作: $NX_1(t, f)$ 、 $NX_2(t, f)$ 。

[0118] 步骤10322,基于音频归一化频谱,确定多路音频信号的相关性特征数据。

[0119] 其中,相关性特征数据为:广义互谱函数,或,自相关矩阵本征值。

[0120] 在一实施例中,可以利用音频归一化频谱的互功率谱进行相位变换加权,得到广义互谱函数。如2路音频归一化频谱为 $NX_1(t, f)$ 、 $NX_2(t, f)$, $NX_1(t, f)$ 、 $NX_2(t, f)$ 的广义互谱函数 $GCC_N(t, f)$ 为:

$$[0121] \quad GCC_N(t, f) = \frac{NX_1(t, f) \cdot NX_2^*(t, f)}{\|NX_1(t, f)\|_2 \|NX_2(t, f)\|_2}$$

[0122] 其中, \cdot 表示点乘运算,*表示共轭运算,t表示时间,f表示频率。

[0123] 需要说明的是,当音频信号大于2路时,对多路音频信号两两计算一次广义互谱函数。

[0124] 在一实施例中,可以通过对信号自相关矩阵进行本征值分解,得到自相关矩阵本征值。如2路音频归一化频谱为 $NX_1(t, f)$ 、 $NX_2(t, f)$, $NX_1(t, f)$ 、 $NX_2(t, f)$ 的自相关矩阵本征值 $N\lambda_1(t, f)$ 、 $N\lambda_2(t, f)$ 的计算步骤为:

[0125] 1) 计算每个时频点 (n, m) 的自功率谱 $NP_{11}(n, m)$ 、 $NP_{22}(n, m)$,以及互功率谱 $NP_{12}(n, m)$ 、 $NP_{21}(n, m)$:

$$[0126] \quad NP_{11}(n, m) = NX_1(n, m) NX_1^*(n, m);$$

$$[0127] \quad NP_{22}(n, m) = NX_2(n, m) NX_2^*(n, m);$$

$$[0128] \quad NP_{12}(n, m) = NX_1(n, m) NX_2^*(n, m);$$

$$[0129] \quad NP_{21}(n, m) = NX_2(n, m) NX_1^*(n, m)。$$

[0130] 2) 计算自相关矩阵 $NR(n, m) = \begin{bmatrix} NP_{11}(n, m) & NP_{12}(n, m) \\ NP_{21}(n, m) & NP_{22}(n, m) \end{bmatrix}$ 的两个特征值 $N\lambda_1$

(n, m) 和 $N\lambda_2(n, m)$,其中 $\forall(n, m) \in (t, f)$,从而得到自相关矩阵本征值 $N\lambda_1(t, f)$ 、 $N\lambda_2(t, f)$ 。

[0131] 需要说明的是,对于包含M个传声器的传声器阵列,计算每个时频点 (n, m) 的互相关矩阵为 $NR(n, m) = E[\overline{nx}(n, m)\overline{nx}(n, m)^H]$,其中 $E[\]$ 为计算期望,

$\overline{nx}(n, m) = [NX_1(n, m), NX_2(n, m), \dots, NX_M(n, m)]^T$,T表示转置。由此可得M个自相关矩阵本征值 $N\lambda_1(t, f), \dots, N\lambda_M(t, f)$ 。

[0132] 本公开实施例所提及的基于传声器阵列确定声源信息的方法,通过确定多路音频信号中每路音频信号的频谱的音频归一化频谱,基于音频归一化频谱,确定多路音频信号的相关性特征数据,由于归一化可以对不同能量大小的输入(分布不平衡的输入)进行调整,便于训练,且采用自适应能量归一化,可以放大语音部分的直达声部分、保留大部分主要的声学信息,可以降低能量较低的无用信息部分、抑制混响和噪声、去除低频偏置,可以大大增强神经网络模型的鲁棒性。

[0133] 本公开一示例性实施例提供了基于相似性度量信息和相关性度量信息,确定声源信息。在本公开图1-6所示实施例的基础上延伸出本公开所示实施例,下面着重叙述本公开所示实施例与图1-6所示实施例的不同之处,相同之处不再赘述。

[0134] 在本公开实施例提供的基于相似性度量信息和相关性度量信息,确定声源信息(即步骤104),包括:

[0135] 基于相似性度量信息和相关性度量信息,确定声源与传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵。

[0136] 具体地,声音是由物体的振动产生的,把正在发声的物体叫声源。确定声源与传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵后,方便后续采用相应的方法计算声源数量、进行语音分离、定位等。

[0137] 本公开实施例所提及的基于传声器阵列确定声源信息的方法,可以基于相似性度量信息和相关性度量信息,确定声源与传声器阵列之间的相对传递函数矩阵为正定矩阵、

超定矩阵、或欠定矩阵,方便后续采用相应的方法计算声源数量、进行语音分离、声源定位等,可以提高声源数量计数、语音分离、声源定位的准确性。

[0138] 图7是本公开一示例性实施例提供的基于相似性度量信息和相关性度量信息,确定声源与传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵的流程示意图。在本公开上述所示实施例的基础上延伸出本公开图7所示实施例,下面着重叙述图7所示实施例与上述所示实施例的不同之处,相同之处不再赘述。

[0139] 如图7所示,在本公开实施例提供的基于相似性度量信息和相关性度量信息,确定声源与传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵,包括:

[0140] 步骤1041,基于相似性度量信息,确定相对传递函数矩阵为正定矩阵时对应的第一相似概率值、相对传递函数矩阵为超定矩阵时对应的第二相似概率值和相对传递函数矩阵为欠定矩阵时对应的第三相似概率值。

[0141] 例如:基于相似性度量信息,确定相对传递函数矩阵为正定矩阵时对应的第一相似概率值为80%、相对传递函数矩阵为超定矩阵时对应的第二相似概率值为15%和相对传递函数矩阵为欠定矩阵时对应的第三相似概率值为5%。

[0142] 步骤1042,基于相关性度量信息,确定相对传递函数矩阵为正定矩阵时对应的第一相关概率值、相对传递函数矩阵为超定矩阵时对应的第二相关概率值和相对传递函数矩阵为欠定矩阵时对应的第三相关概率值。

[0143] 例如:基于相关性度量信息,确定相对传递函数矩阵为正定矩阵时对应的第一相关概率值为70%、相对传递函数矩阵为超定矩阵时对应的第二相关概率值为18%和相对传递函数矩阵为欠定矩阵时对应的第三相关概率值为12%。

[0144] 步骤1043,根据第一相似概率值、第二相似概率值和第三相似概率值,以及第一相关概率值、第二相关概率值和第三相关概率值,确定相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵。

[0145] 具体地,根据第一相似概率值、第二相似概率值和第三相似概率值,以及第一相关概率值、第二相关概率值和第三相关概率值,确定相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵时,可以根据实际应用情况,选择任一可行的方式,对此不做具体限定。在一实施例中,可以计算两两相似概率值的差的绝对值;当所有相似概率值的差的绝对值大于预设概率差阈值时,选取最大相似概率值对应的结果为最终判定结果;当任一相似概率值的差的绝对值小于等于预设概率差阈值时,选取最大相关概率值对应的结果为最终判定结果,其中,当最大相关概率值存在二个以上时,选取任一最大相关概率值对应的结果为最终判定结果并报错。例如:根据步骤1041、步骤1042的例子,确定相对传递函数矩阵为正定矩阵时对应的第一相似概率值为80%、相对传递函数矩阵为超定矩阵时对应的第二相似概率值为15%和相对传递函数矩阵为欠定矩阵时对应的第三相似概率值为5%,两两相似概率值的差的绝对值分别为:第一相似概率值与第二相似概率值的差的绝对值为65%、第一相似概率值与第三相似概率值的差的绝对值为75%,第二相似概率值与第三相似概率值的差的绝对值为10%,预设概率差阈值为1%,所有相似概率值的差的绝对值大于预设概率差阈值时,选取最大相似概率值80%对应的结果相对传递函数矩阵为正定矩阵为最终判定结果。

[0146] 本公开实施例所提及的基于传声器阵列确定声源信息的方法,确定并根据第一相

似概率值、第二相似概率值和第三相似概率值,以及第一相关概率值、第二相关概率值和第三相关概率值,确定相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵,实现方式简便快捷,进而提高了确定声源信息的实现速度。

[0147] 图8是本公开另一示例性实施例提供的基于传声器阵列确定声源信息的方法的流程示意图。在本公开图7上一所示实施例的基础上延伸出本公开图8所示实施例,下面着重叙述图8所示实施例与图7上一所示实施例的不同之处,相同之处不再赘述。

[0148] 如图8所示,在本公开实施例提供的基于传声器阵列确定声源信息的方法中,还包括:

[0149] 步骤105,当相对传递函数矩阵为正定矩阵时,基于传声器阵列中包括的传声器的个数确定声源数量。

[0150] 具体地,当相对传递函数矩阵为正定矩阵时,可以确定声源数量的值为传声器阵列中包括的传声器的个数。例如:当相对传递函数矩阵为正定矩阵时,传声器阵列中包括的传声器的个数为5,则确定声源数量的值为5。

[0151] 步骤106,当相对传递函数矩阵为超定矩阵时,确定多路音频信号的子空间的正交基个数或确定声源的空间谱的谱峰值个数,基于正交基个数或谱峰值个数确定声源数量。

[0152] 其中,多路音频信号的子空间为多路音频信号的自相关矩阵的正交量形成的空间,空间谱为基于声源的方位角设计的函数。

[0153] 具体地,当相对传递函数矩阵为超定矩阵时,确定多路音频信号的子空间的正交基个数或确定声源的空间谱的谱峰值个数,确定声源数量的值为正交基个数或谱峰值个数。例如:当相对传递函数矩阵为超定矩阵时,确定多路音频信号的子空间的正交基个数为3或确定声源的空间谱的谱峰值个数为3,从而确定声源数量的值为3。

[0154] 步骤107,当相对传递函数矩阵为欠定矩阵时,确定声源的方位直方图的峰值个数,基于峰值个数确定声源数量。

[0155] 具体地,当相对传递函数矩阵为欠定矩阵时,确定声源的方位直方图的峰值个数,确定声源数量的值为峰值个数。例如:当相对传递函数矩阵为欠定矩阵时,确定声源的方位直方图的峰值个数为7,则确定声源数量的值为7。

[0156] 本公开实施例所提及的基于传声器阵列确定声源信息的方法,当相对传递函数矩阵为正定矩阵时,基于传声器阵列中包括的传声器的个数确定声源数量,当相对传递函数矩阵为超定矩阵时,确定多路音频信号的子空间的正交基个数或确定声源的空间谱的谱峰值个数,基于正交基个数或谱峰值个数确定声源数量,当相对传递函数矩阵为欠定矩阵时,确定声源的方位直方图的峰值个数,基于峰值个数确定声源数量,可以根据相对传递函数矩阵的适定情况,选择相应的方法确定声源数量,可以提高声源数量计数的准确性、可以实现复杂声场景(例如声源为终端设备,终端设备可以发出很多种声音,容易被误判断为多个声源)中声源数量的计数。

[0157] 本公开另一示例性实施例提供了基于相似性度量信息和相关性度量信息,确定声源信息。在本公开图1所示实施例的基础上延伸出本公开所示实施例,下面着重叙述本公开所示实施例与图1所示实施例的不同之处,相同之处不再赘述。

[0158] 在本公开实施例提供的基于传声器阵列确定声源信息的方法中,基于相似性度量信息和相关性度量信息,确定声源信息,包括:

[0159] 基于相似性度量信息和相关性度量信息,利用预设第三神经网络模型,确定声源与传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵。

[0160] 其中,预设第三神经网络模型可以采用分类问题中常用的各种神经网络,对此不做具体限定。具体实现与步骤1024类似,详见步骤1024,此处不再赘述。

[0161] 本公开实施例所提及的基于传声器阵列确定声源信息的方法,基于相似性度量信息和相关性度量信息,利用预设第三神经网络模型,确定声源与传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵,通过神经网络模型即可确定,实现简单快捷,且提高了神经网络模型的泛化性能。

[0162] 图9是本公开又一示例性实施例提供的基于传声器阵列确定声源信息的方法的流程图示意图。在本公开上述所示实施例的基础上延伸出本公开图9所示实施例,下面着重叙述图9所示实施例与上述所示实施例的不同之处,相同之处不再赘述。

[0163] 如图9所示,在本公开实施例提供的基于传声器阵列确定声源信息的方法中,还包括:

[0164] 步骤108,当相对传递函数矩阵为正定矩阵时,基于传声器阵列中包括的传声器的个数确定声源数量;

[0165] 具体与步骤105类似,详见步骤105,此处不再赘述。

[0166] 步骤109,当相对传递函数矩阵为超定矩阵时,确定多路音频信号的子空间的正交基个数或确定声源的空间谱的谱峰值个数,基于正交基个数或谱峰值个数确定声源数量,其中,多路音频信号的子空间为多路音频信号的自相关矩阵的正交量形成的空间,空间谱为基于声源的方位角设计的函数。

[0167] 具体与步骤106类似,详见步骤106,此处不再赘述。

[0168] 步骤110,当相对传递函数矩阵为欠定矩阵时,确定声源的方位直方图的峰值个数,基于峰值个数确定声源数量。

[0169] 具体与步骤107类似,详见步骤107,此处不再赘述。

[0170] 本公开实施例所提及的基于传声器阵列确定声源信息的方法,当相对传递函数矩阵为正定矩阵时,基于传声器阵列中包括的传声器的个数确定声源数量,当相对传递函数矩阵为超定矩阵时,确定多路音频信号的子空间的正交基个数或确定声源的空间谱的谱峰值个数,基于正交基个数或谱峰值个数确定声源数量,当相对传递函数矩阵为欠定矩阵时,确定声源的方位直方图的峰值个数,基于峰值个数确定声源数量,可以根据相对传递函数矩阵的适定情况,选择相应的方法确定声源数量,可以提高声源数量计数的准确性、可以实现复杂声场景(例如声源为终端设备,终端设备可以发出很多种声音,容易被误判断为多个声源)中声源数量的计数。

[0171] 本公开实施例提供的任一种基于传声器阵列确定声源信息的方法可以由任意适当的具有数据处理能力的设备执行,包括但不限于:终端设备和服务器等。或者,本公开实施例提供的任一种基于传声器阵列确定声源信息的方法可以由处理器执行,如处理器通过调用存储器存储的相应指令来执行本公开实施例提及的任一种基于传声器阵列确定声源信息的方法。下文不再赘述。

[0172] 示例性装置

[0173] 图10是本申请一示例性实施例提供的基于传声器阵列确定声源信息的装置的结

构示意图。如图10所示,本申请实施例提供的基于传声器阵列确定声源信息的装置,包括:

- [0174] 音频信号确定模块201,用于确定传声器阵列采集的多路音频信号;
- [0175] 相似性度量信息确定模块202,用于确定多路音频信号的相似性度量信息;
- [0176] 相关性度量信息确定模块203,用于确定多路音频信号的相关性度量信息;
- [0177] 声源信息确定模块204,用于基于相似性度量信息和相关性度量信息,确定声源信息。

[0178] 图11是本公开一示例性实施例提供的相似性度量信息确定模块的结构示意图。在本公开图10所示实施例的基础上延伸出本公开图11所示实施例,下面着重叙述图11所示实施例与图10所示实施例的不同之处,相同之处不再赘述。

[0179] 如图11所示,在本公开实施例提供的基于传声器阵列确定声源信息的装置中,相似性度量信息确定模块202,包括:

- [0180] 盲源分离信号确定单元2021,用于确定多路音频信号中每路音频信号的盲源分离信号;
- [0181] 第一频谱确定单元2022,用于确定盲源分离信号的频谱;
- [0182] 相似性特征数据确定单元2023,用于基于盲源分离信号的频谱,确定盲源分离信号的相似性特征数据,其中,相似性特征数据为:梅尔频谱包络特征,或,频谱余弦距离;
- [0183] 相似性度量信息确定单元2024,用于基于相似性特征数据,利用预设第一神经网络模型,确定多路音频信号的相似性度量信息。

[0184] 图12是本公开一示例性实施例提供的相似性特征数据确定单元的结构示意图。在本公开图11所示实施例的基础上延伸出本公开图12所示实施例,下面着重叙述图12所示实施例与图11所示实施例的不同之处,相同之处不再赘述。

[0185] 如图12所示,在本公开实施例提供的相似性特征数据确定单元2023,包括:

- [0186] 盲源归一化频谱确定子单元20231,用于确定盲源分离信号的频谱的盲源归一化频谱;
- [0187] 相似性特征数据确定子单元20232,用于基于盲源归一化频谱,确定盲源分离信号的相似性特征数据。

[0188] 图13是本公开一示例性实施例提供的相关性度量信息确定模块的结构示意图。在本公开图10所示实施例的基础上延伸出本公开图13所示实施例,下面着重叙述图13所示实施例与图10所示实施例的不同之处,相同之处不再赘述。

[0189] 如图13所示,本公开实施例提供的相关性度量信息确定模块203,包括:

- [0190] 第二频谱确定单元2031,用于确定多路音频信号中每路音频信号的频谱;
- [0191] 相关性特征数据确定单元2032,用于基于每路音频信号的频谱,确定多路音频信号的相关性特征数据,其中,相关性特征数据为:广义互谱函数,或,自相关矩阵本征值;
- [0192] 相关性度量信息确定单元2033,用于基于相关性特征数据,利用预设第二神经网络模型,确定多路音频信号的相关性度量信息。

[0193] 图14是本公开一示例性实施例提供的相关性特征数据确定单元的结构示意图。在本公开图13所示实施例的基础上延伸出本公开图14所示实施例,下面着重叙述图14所示实施例与图13所示实施例的不同之处,相同之处不再赘述。

[0194] 如图14所示,在本公开实施例提供的基于传声器阵列确定声源信息的装置中,相

关性特征数据确定单元2032,包括:

[0195] 音频归一化频谱确定子单元20321,用于确定多路音频信号中每路音频信号的频谱的音频归一化频谱;

[0196] 相关性特征数据确定子单元20322,用于基于音频归一化频谱,确定多路音频信号的相关性特征数据。

[0197] 本公开一示例性实施例提供了声源信息确定模块的结构示意图。在本公开图10所示实施例的基础上延伸出本公开所示实施例,下面着重叙述本公开所示实施例与图10所示实施例的不同之处,相同之处不再赘述。

[0198] 本公开实施例提供的声源信息确定模块204,具体用于基于相似性度量信息和相关性度量信息,确定声源与传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵。

[0199] 图15是本公开一示例性实施例提供的声源信息确定模块的结构示意图。在本公开上述所示实施例的基础上延伸出本公开图15所示实施例,下面着重叙述图15所示实施例与上述所示实施例的不同之处,相同之处不再赘述。

[0200] 如图15所示,本公开实施例提供的声源信息确定模块204,包括:

[0201] 相似概率值确定单元2041,用于基于相似性度量信息,确定相对传递函数矩阵为正定矩阵时对应的第一相似概率值、相对传递函数矩阵为超定矩阵时对应的第二相似概率值和相对传递函数矩阵为欠定矩阵时对应的第三相似概率值;

[0202] 相关概率值确定单元2042,用于基于相关性度量信息,确定相对传递函数矩阵为正定矩阵时对应的第一相关概率值、相对传递函数矩阵为超定矩阵时对应的第二相关概率值和相对传递函数矩阵为欠定矩阵时对应的第三相关概率值;

[0203] 矩阵确定单元2043,用于根据第一相似概率值、第二相似概率值和第三相似概率值,以及第一相关概率值、第二相关概率值和第三相关概率值,确定相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵。

[0204] 图16是本公开另一示例性实施例提供的基于传声器阵列确定声源信息的装置的结构示意图。在本公开图15上一所示实施例的基础上延伸出本公开图16所示实施例,下面着重叙述图16所示实施例与图15上一所示实施例的不同之处,相同之处不再赘述。

[0205] 如图16所示,在本公开实施例提供的基于传声器阵列确定声源信息的装置中,还包括:

[0206] 第一声源数量确定模块205,用于当相对传递函数矩阵为正定矩阵时,基于传声器阵列中包括的传声器的个数确定声源数量;

[0207] 第二声源数量确定模块206,用于当相对传递函数矩阵为超定矩阵时,确定多路音频信号的子空间的正交基个数或确定声源的空间谱的谱峰值个数,基于正交基个数或谱峰值个数确定声源数量,其中,多路音频信号的子空间为多路音频信号的自相关矩阵的正交量形成的空间,空间谱为基于声源的方位角设计的函数;

[0208] 第三声源数量确定模块207,用于当相对传递函数矩阵为欠定矩阵时,确定声源的方位直方图的峰值个数,基于峰值个数确定声源数量。

[0209] 本公开一示例性实施例提供了声源信息确定模块的结构示意图。在本公开图10所示实施例的基础上延伸出本公开所示实施例,下面着重叙述本公开所示实施例与图10所示

实施例的不同之处,相同之处不再赘述。

[0210] 本公开实施例提供的声源信息确定模块204,具体用于基于相似性度量信息和相关性度量信息,利用预设第三神经网络模型,确定声源与传声器阵列之间的相对传递函数矩阵为正定矩阵、超定矩阵、或欠定矩阵。

[0211] 图17是本公开又一示例性实施例提供的基于传声器阵列确定声源信息的装置的结构示意图。在本公开上述所示实施例的基础上延伸出本公开图17所示实施例,下面着重叙述图17所示实施例与上述所示实施例的不同之处,相同之处不再赘述。

[0212] 如图17所示,在本公开实施例提供的基于传声器阵列确定声源信息的装置中,还包括:

[0213] 第四声源数量确定模块208,用于当相对传递函数矩阵为正定矩阵时,基于传声器阵列中包括的传声器的个数确定声源数量;

[0214] 第五声源数量确定模块209,用于当相对传递函数矩阵为超定矩阵时,确定多路音频信号的子空间的正交基个数或确定声源的空间谱的谱峰值个数,基于正交基个数或谱峰值个数确定声源数量,其中,多路音频信号的子空间为多路音频信号的自相关矩阵的正交量形成的空间,空间谱为基于声源的方位角设计的函数;

[0215] 第六声源数量确定模块210,用于当相对传递函数矩阵为欠定矩阵时,确定声源的方位直方图的峰值个数,基于峰值个数确定声源数量。

[0216] 应当理解,图10至图17提供的基于传声器阵列确定声源信息的装置中的音频信号确定模块201、相似性度量信息确定模块202、相关性度量信息确定模块203、声源信息确定模块204、第一声源数量确定模块205、第二声源数量确定模块206、第三声源数量确定模块207、第四声源数量确定模块208、第五声源数量确定模块209和第六声源数量确定模块210,以及相似性度量信息确定模块202中包括的盲源分离信号确定单元2021、第一频谱确定单元2022、相似性特征数据确定单元2023、相似性度量信息确定单元2024,相似性特征数据确定单元2023中包括的盲源归一化频谱确定子单元20231、相似性特征数据确定子单元20232,相关性度量信息确定模块203中包括的第二频谱确定单元2031、相关性特征数据确定单元2032、相关性度量信息确定单元2033,相关性特征数据确定单元2032中包括的音频归一化频谱确定子单元20321、相关性特征数据确定子单元20322,声源信息确定模块204中包括的相似概率值确定单元2041、相关概率值确定单元2042、矩阵确定单元2043的操作和功能可以参考上述图1至图9提供的基于传声器阵列确定声源信息的方法,为了避免重复,在此不再赘述。

[0217] 示例性电子设备

[0218] 图18图示了根据本公开实施例的电子设备的框图。

[0219] 如图18所示,电子设备11包括一个或多个处理器111和存储器112。

[0220] 处理器111可以是中央处理单元(CPU)或者具有数据处理能力和/或指令执行能力的其他形式的处理单元,并且可以控制电子设备10中的其他组件以执行期望的功能。

[0221] 存储器112可以包括一个或多个计算机程序产品,计算机程序产品可以包括各种形式的计算机可读存储介质,例如易失性存储器和/或非易失性存储器。易失性存储器例如可以包括随机存取存储器(RAM)和/或高速缓冲存储器(cache)等。非易失性存储器例如可以包括只读存储器(ROM)、硬盘、闪存等。在计算机可读存储介质上可以存储一个或多个计

计算机程序指令,处理器111可以运行程序指令,以实现上文的本公开的各个实施例的基于传声器阵列确定声源信息的方法以及/或者其他期望的功能。在计算机可读存储介质中还可以存储诸如输入信号、信号分量、噪声分量等各种内容。

[0222] 在一个示例中,电子设备11还可以包括:输入装置113和输出装置114,这些组件通过总线系统和/或其他形式的连接机构(未示出)互连。

[0223] 例如,该输入装置113可以是上述的麦克风或传声器阵列,用于捕捉声源的输入信号。在该电子设备是单机设备时,该输入装置113可以是通信网络连接器,用于从网络处理器接收所采集的输入信号。

[0224] 此外,该输入设备113还可以包括例如键盘、鼠标等等。

[0225] 该输出装置114可以向外部输出各种信息,包括确定出的距离信息、方向信息等。该输出设备114可以包括例如显示器、扬声器、打印机、以及通信网络及其所连接的远程输出设备等等。

[0226] 当然,为了简化,图18中仅示出了该电子设备11中与本公开有关的组件中的一些,省略了诸如总线、输入/输出接口等等的组件。除此之外,根据具体应用情况,电子设备11还可以包括任何其他适当的组件。

[0227] 示例性计算机程序产品和计算机可读存储介质

[0228] 除了上述方法和设备以外,本公开的实施例还可以是计算机程序产品,其包括计算机程序指令,计算机程序指令在被处理器运行时使得处理器执行本说明书上述“示例性方法”部分中描述的根据本公开各种实施例的基于传声器阵列确定声源信息的方法中的步骤。

[0229] 计算机程序产品可以以一种或多种程序设计语言的任意组合来编写用于执行本公开实施例操作的程序代码,程序设计语言包括面向对象的程序设计语言,诸如Java、C++等,还包括常规的过程式程序设计语言,诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。

[0230] 此外,本公开的实施例还可以是计算机可读存储介质,其上存储有计算机程序指令,计算机程序指令在被处理器运行时使得处理器执行本说明书上述“示例性方法”部分中描述的根据本公开各种实施例的基于传声器阵列确定声源信息的方法中的步骤。

[0231] 计算机可读存储介质可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以包括但不限于电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。

[0232] 以上结合具体实施例描述了本公开的基本原理,但是,需要指出的是,在本公开中提及的优点、优势、效果等仅是示例而非限制,不能认为这些优点、优势、效果等是本公开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公开为必须采用上述具体的细节来实现。

[0233] 本说明书中各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其它实施例的不同之处,各个实施例之间相同或相似的部分相互参见即可。对于系统实施例而言,由于其与方法实施例基本对应,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0234] 本公开中涉及的器件、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。

[0235] 可能以许多方式来实现本公开的方法和装置。例如,可通过软件、硬件、固件或者软件、硬件、固件的任何组合来实现本公开的方法和装置。用于方法的步骤的上述顺序仅是为了进行说明,本公开的方法的步骤不限于以上具体描述的顺序,除非以其它方式特别说明。此外,在一些实施例中,还可将本公开实施为记录在记录介质中的程序,这些程序包括用于实现根据本公开的方法的机器可读指令。因而,本公开还覆盖存储用于执行根据本公开的方法的程序的记录介质。

[0236] 还需要指出的是,在本公开的装置、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公开的等效方案。

[0237] 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公开。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公开的范围。因此,本公开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。

[0238] 为了例示和描述的目的已经给出了以上描述。此外,此描述不意图将本公开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。

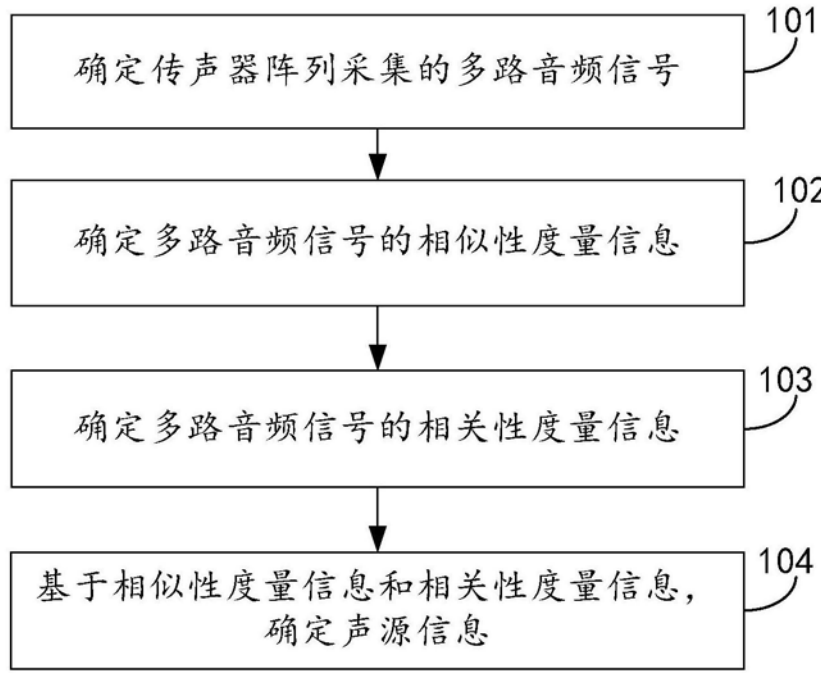


图1

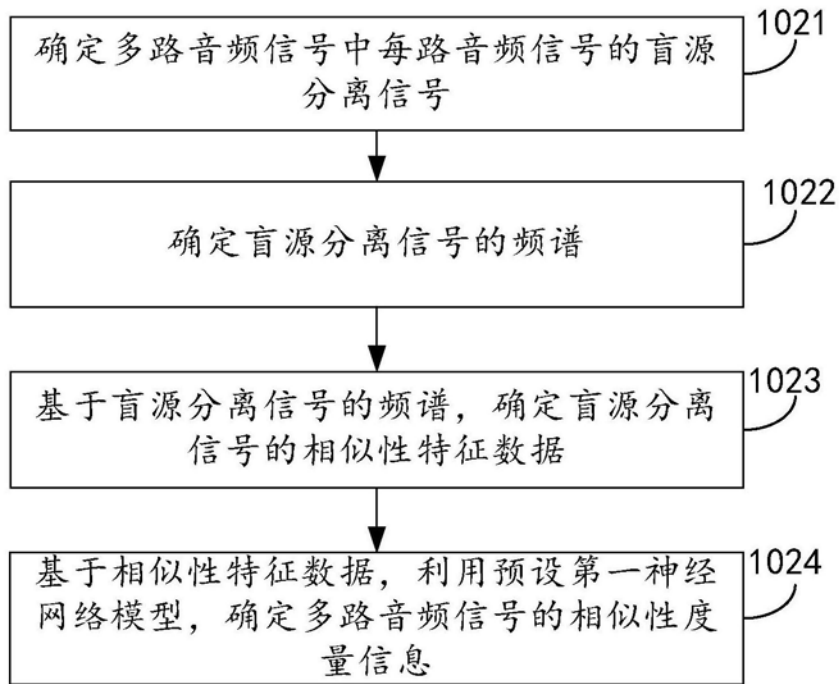


图2

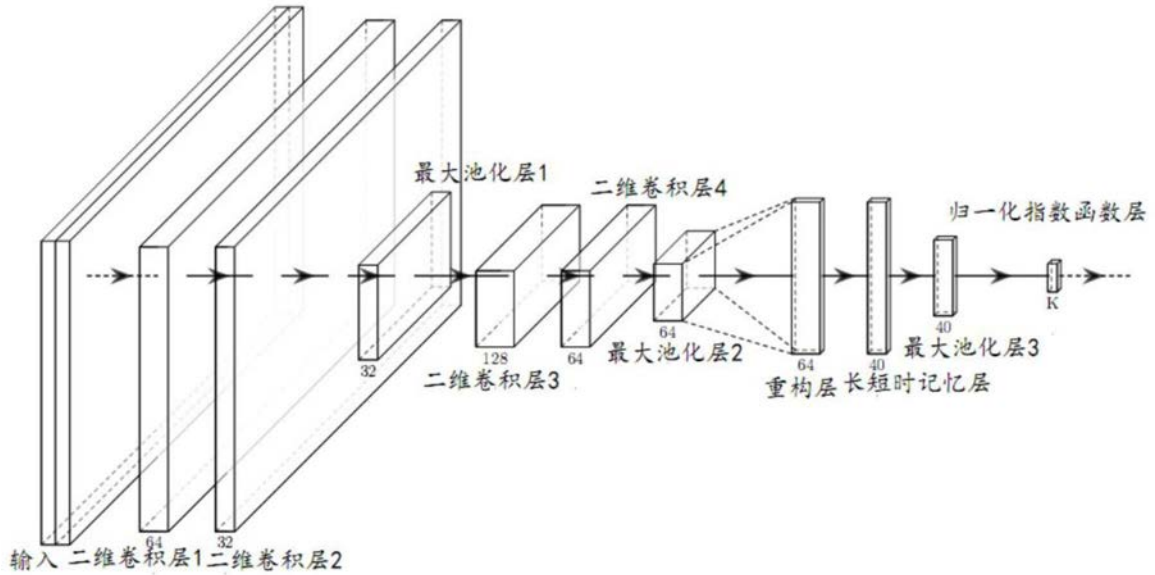


图3

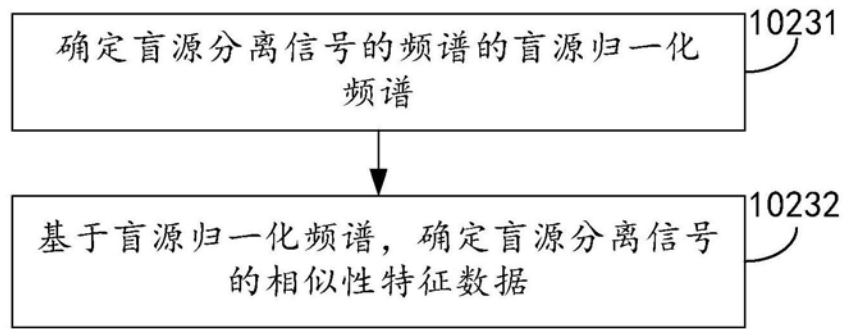


图4

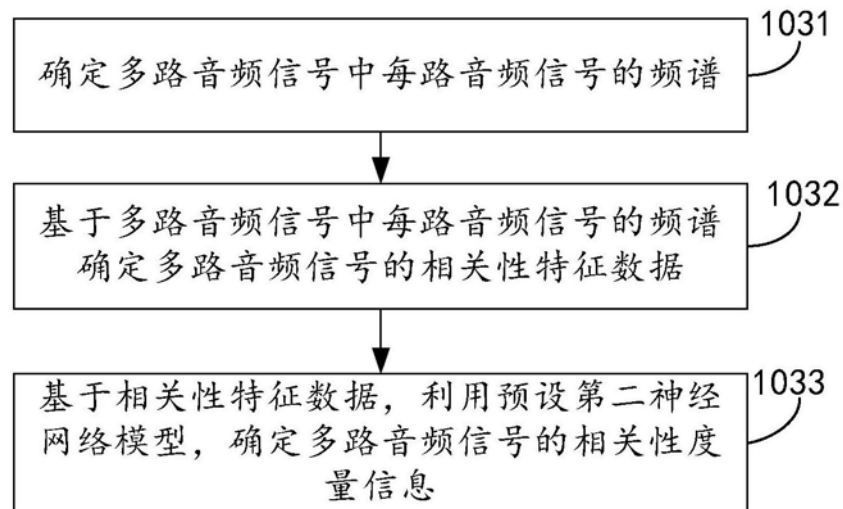


图5

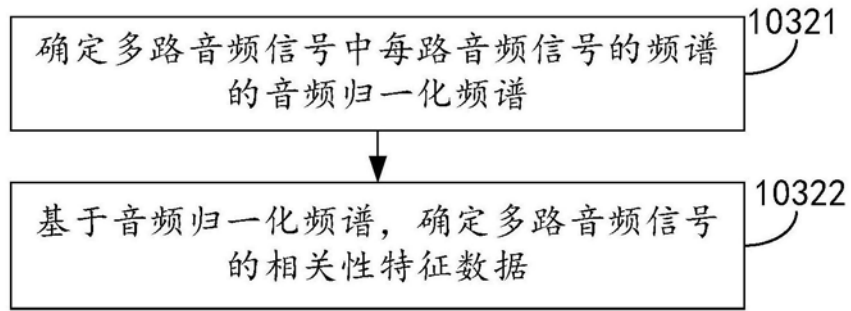


图6

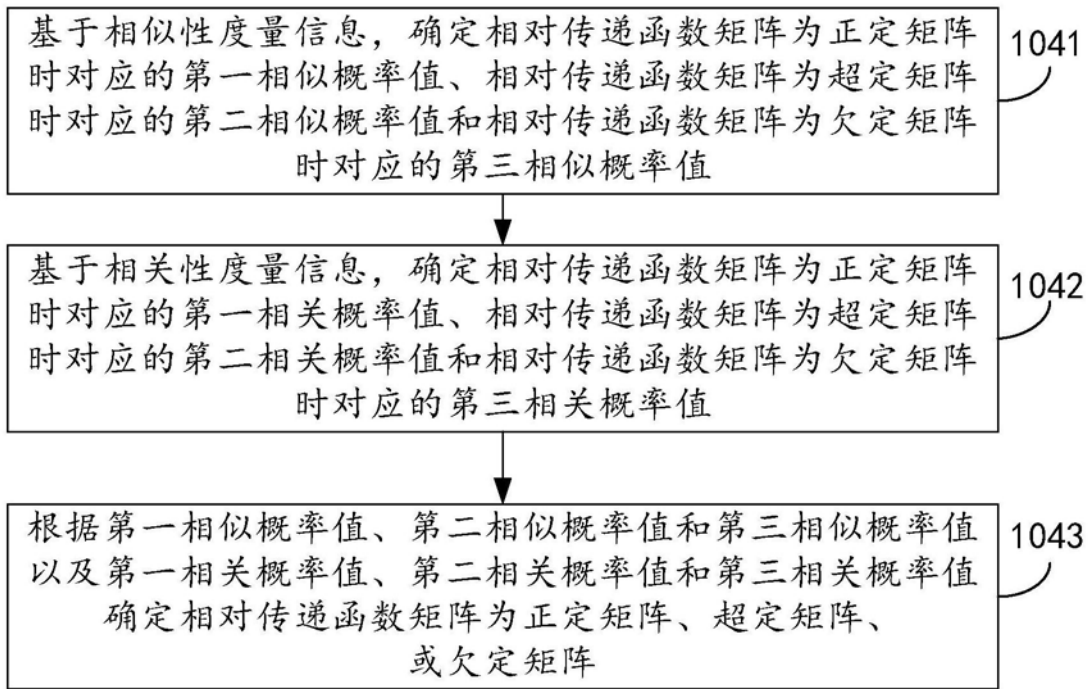


图7

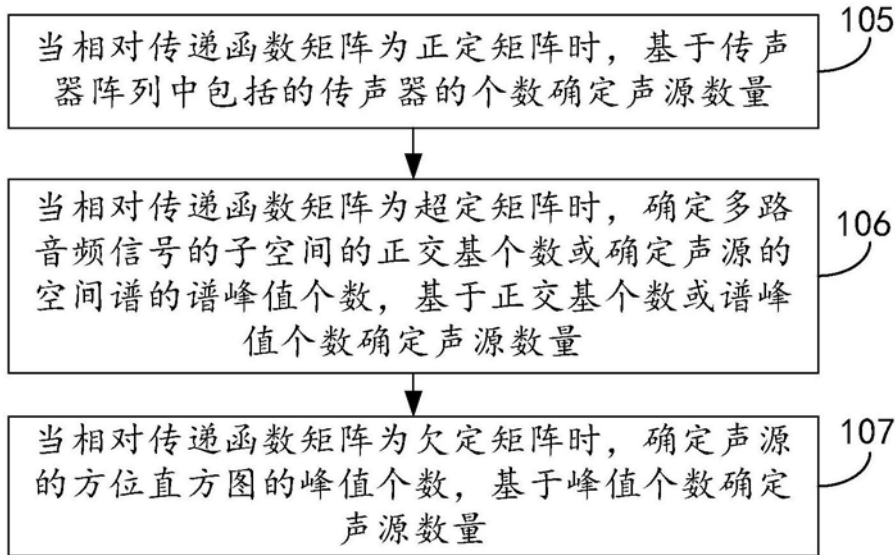


图8

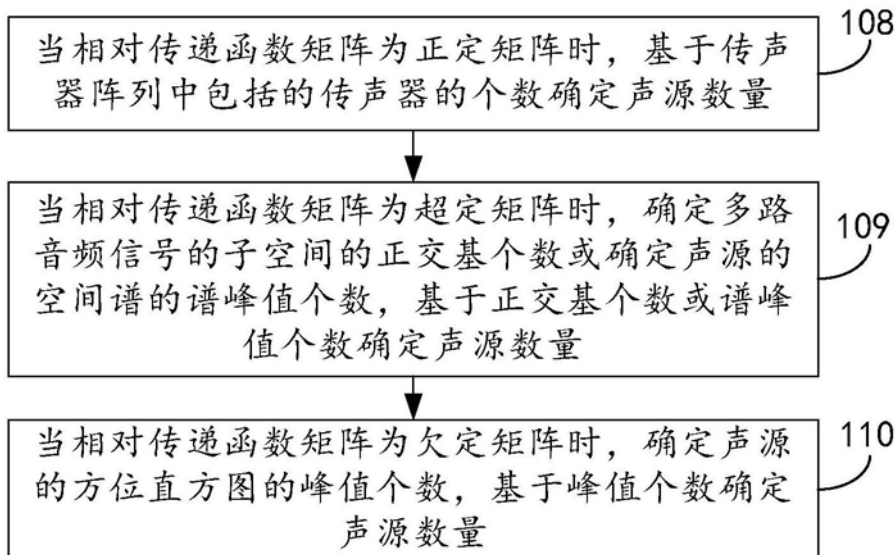


图9

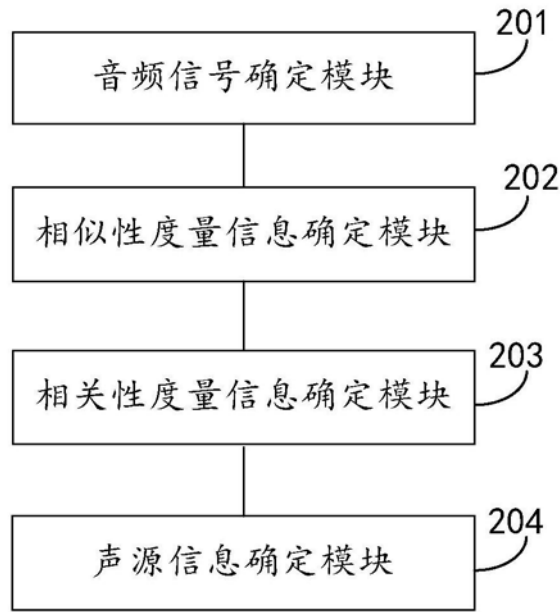


图10

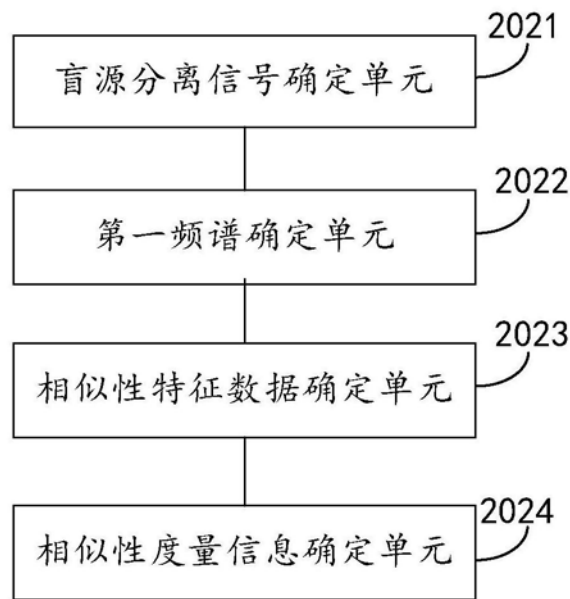


图11

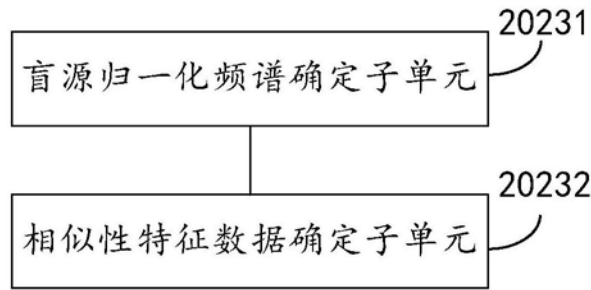


图12

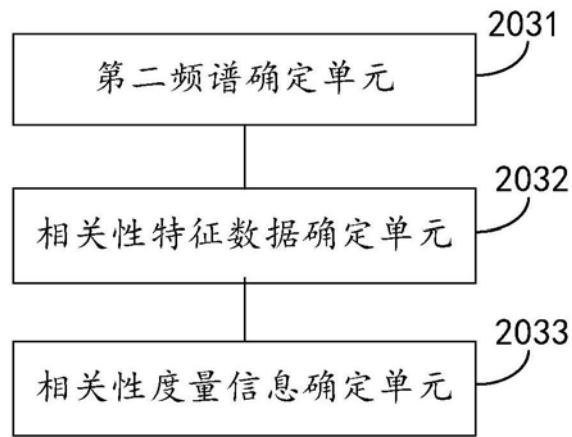


图13

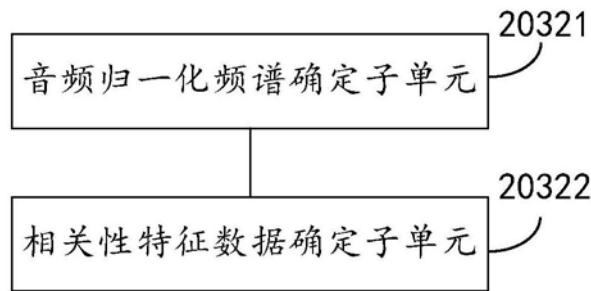


图14

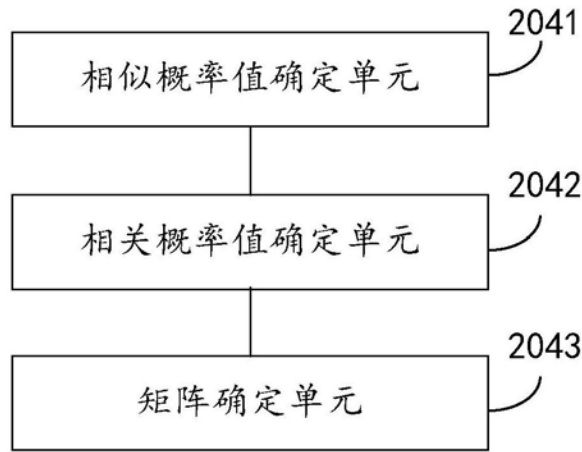


图15

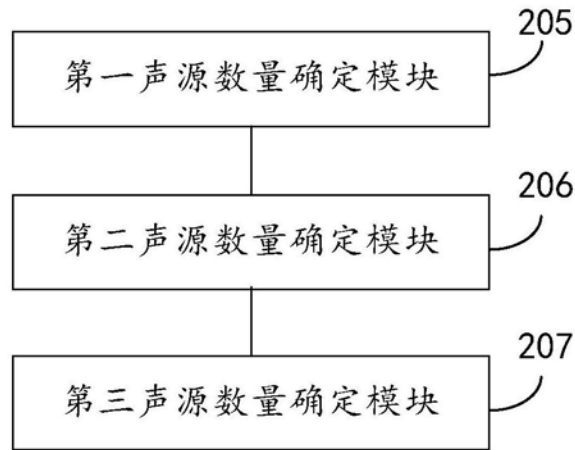


图16

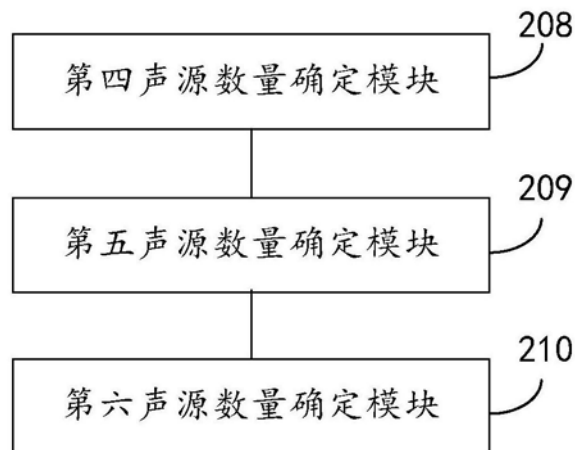


图17

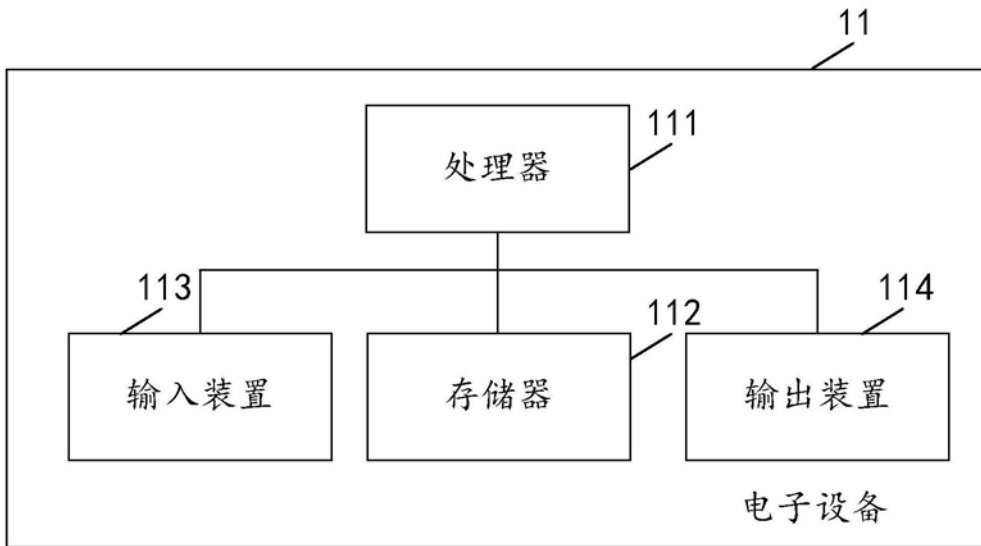


图18