(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 111929645 A
(43)申请公布日 2020.11.13

(21)申请号 202011008660.6

(22)申请日 2020.09.23

(71)申请人 深圳市友杰智新科技有限公司
地址 518000 广东省深圳市南山区招商街道蛇口南海大道1079号花园城数码大厦A座402

(72)发明人 陈俊彬 王广新 太荣鹏

(74)专利代理机构 深圳市明日今典知识产权代理事务所(普通合伙) 44343
代理人 王杰辉 赫坤鹏

(51)Int.Cl.
G01S 5/20(2006.01)

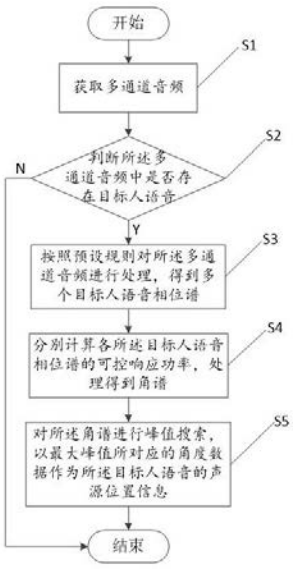
权利要求书2页 说明书18页 附图2页

(54)发明名称

特定人声的声源定位方法、装置和计算机设备

(57)摘要

本申请提供了一种特定人声的声源定位方法、装置和计算机设备，系统首先判断获取的多通道音频中是否存在目标人语音，若多通道音频中存在目标人语音，则按照预设规则对多通道音频进行处理，得到多个目标人语音相位谱。系统分别计算各目标人语音相位谱的可控响应功率，处理得到角谱。最后对角谱进行峰值搜索，以最大峰值所对应的角度数据作为目标人语音的声源位置信息。本申请在识别到目标人语音后，根据目标人语音相位谱进行相应的计算得到声源位置，在计算过程中并不涉及音频的功率谱，因而可以减少无关信息的干扰，从而准确定位特定人声的声源位置。



CN 111929645 A

1.一种特定人声的声源定位方法,其特征在于,包括:

获取多通道音频;

判断所述多通道音频中是否存在目标人语音;

若所述多通道音频中存在目标人语音,则按照预设规则对所述多通道音频进行处理,得到多个目标人语音相位谱;

分别计算各所述目标人语音相位谱的可控响应功率,处理得到角谱;

对所述角谱进行峰值搜索,以最大峰值所对应的角度数据作为所述目标人语音的声源位置信息。

2.根据权利要求1所述的特定人声的声源定位方法,其特征在于,所述判断所述多通道音频中是否存在目标人语音的步骤,包括:

从所述多通道音频中选取任一通道音频进行特征提取,得到各帧音频分别对应的MFCC;

将各所述MFCC缓存为一组,输入第一神经网络进行处理,得到识别概率,所述第一神经网络用于识别输入音频中存在目标人语音的概率;

判断所述识别概率是否大于概率阈值;

若所述识别概率大于概率阈值,则判定所述多通道音频中存在目标人语音;

若所述识别概率均小于概率阈值,则判定所述多通道音频中不存在目标人语音。

3.根据权利要求2所述的特定人声的声源定位方法,其特征在于,所述判断所述识别概率是否大于概率阈值的步骤,包括:

判断所述识别概率是否为异常概率;

若所述识别概率为异常概率,则根据所述异常概率、所述异常概率的前一识别概率、所述异常概率的后一识别概率进行均值计算,得到修正概率,所述前一识别概率为所述异常概率前一组MFCC所对应的识别概率,所述后一识别概率为所述异常概率后一组MFCC所对应的识别概率;

判断所述修正概率是否大于概率阈值;

若所述修正概率大于概率阈值,则判定所述识别概率大于概率阈值;

若所述修正概率小于概率阈值,则判定所述识别概率小于概率阈值。

4.根据权利要求1所述的特定人声的声源定位方法,其特征在于,所述按照预设规则对所述多通道音频进行处理,得到多个目标人语音相位谱的步骤,包括:

对所述多通道音频进行快速傅里叶变换,得到多个频域信号;

将各所述频域信号输入第二神经网络进行处理,得到各所述目标人语音相位谱,所述第二神经网络用于分离出输入音频信号中目标人语音的相位谱。

5.根据权利要求1所述的特定人声的声源定位方法,其特征在于,所述分别计算各所述目标人语音相位谱的可控响应功率,处理得到角谱的步骤,包括:

以麦克风阵列的几何中心为原点构建空间直角坐标系;

按照预设角度范围,在所述空间直角坐标系上选取若干个方向向量;

根据各所述方向向量和各所述目标人语音相位谱,计算得到各所述方向向量各自对应的所述可控响应功率;

将各所述方向向量转化为角度形式,得到各所述方向向量分别对应的水平角和俯仰

角；

根据各所述可控响应功率与各所述方向向量分别对应的水平角和俯仰角之间的对应关系，生成所述角谱。

6.根据权利要求5所述的特定人声的声源定位方法，其特征在于，所述根据各所述方向向量和各所述目标人语音相位谱，计算得到各所述方向向量各自对应的所述可控响应功率的步骤，包括：

将各所述目标人语音相位谱转化为复数形式的目标人语音相位谱，并计算各所述方向向量所在方向分别到达所述麦克风阵列中两个麦克风的时间差；

将所述复数形式的目标人语音相位谱、所述时间差和所述方向向量代入第一公式中，计算得到所述麦克风阵列中两个麦克风所接收的音频帧数据之间的广义互相关函数，其中，所述第一公式为：

$R_{ac}[\tau_{ac}(d_h)]=\frac{1}{K}\sum_{k=1}^{K}Y_a(k)Y_c^*(k)e^{j\Omega\tau_{ac}(d_h)}$，$\tau_{ac}(d_h)$为所述时间差，$Y_a(k)$、$Y_c(k)$均为复数形式的目标人语音相位谱，$d_h$为所述方向向量，$R_{ac}[\tau_{ac}(d_h)]$为所述广义互相关函数；

将所述广义互相关函数代入第二公式中，计算得到各所述可控响应功率，其中，所述第二公式为：

$F(d_h)=\sum_{a=1}^{C-1}\sum_{c=a+1}^{C}R_{ac}[\tau_{ac}(d_h)]$，$F(d_h)$为所述可控响应功率。

7.根据权利要求6所述的特定人声的声源定位方法，其特征在于，所述计算各所述方向向量所在方向分别到达所述麦克风阵列中两个麦克风的时间差的步骤，包括：

将所述方向向量代入第三公式中，计算得到对应的所述时间差，其中，所述第三公式为：$\tau_{ac}(d_h)=\frac{\|d_h-r_c\|-\|d_h-r_a\|}{v}$，$r_a$为第a个麦克风在所述空间直角坐标系中的坐标向量，$r_c$为第c个麦克风在所述空间直角坐标系中的坐标向量，v为音速。

8.一种特定人声的声源定位装置，其特征在于，包括：

获取模块，用于获取多通道音频；

判断模块，用于判断所述多通道音频中是否存在目标人语音；

处理模块，用于若所述多通道音频中存在目标人语音，则按照预设规则对所述多通道音频进行处理，得到多个目标人语音相位谱；

计算模块，用于分别计算各所述目标人语音相位谱的可控响应功率，处理得到角谱；

搜索模块，用于对所述角谱进行峰值搜索，以最大峰值所对应的角度数据作为所述目标人语音的声源位置信息。

9.一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机程序，其特征在于，所述处理器执行所述计算机程序时实现权利要求1至7中任一项所述方法的步骤。

10.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现权利要求1至7中任一项所述的方法的步骤。

特定人声的声源定位方法、装置和计算机设备

技术领域

[0001] 本申请涉及声源定位技术领域，特别涉及一种特定人声的声源定位方法、装置和计算机设备。

背景技术

[0002] 在特定场合，录像装置、拾音装置等需要采集某个特定人物的音视频信息，比如在大讲堂、公开课等场景，摄像头和拾音模块需要聚焦在讲课老师的方向上；在节目舞台上，摄像头和拾音模块需要聚焦在主持人的方向上。而在实际场景中，由于现场环境嘈杂，可能存在多个说话人，并且特定人物的位置并不是固定不变的(可能因互动而到处移动)。传统的声源定位算法无法区分特定人物语音以及干扰语音(比如其他人的语音)的区别，因而无法准确实现对特定人声的声源定位。

发明内容

[0003] 本申请的主要目的为提供一种特定人声的声源定位方法、装置和计算机设备，旨在解决现有声源定位算法无法准确实现对特定人声的声源定位的弊端。

[0004] 为实现上述目的，本申请提供了一种特定人声的声源定位方法，包括：

获取多通道音频；

判断所述多通道音频中是否存在目标人语音；

若所述多通道音频中存在目标人语音，则按照预设规则对所述多通道音频进行处理，得到多个目标人语音相位谱；

分别计算各所述目标人语音相位谱的可控响应功率，处理得到角谱；

对所述角谱进行峰值搜索，以最大峰值所对应的角度数据作为所述目标人语音的声源位置信息。

[0005] 进一步的，所述判断所述多通道音频中是否存在目标人语音的步骤，包括：

从所述多通道音频中选取任一通道音频进行特征提取，得到各帧音频分别对应的MFCC；

将各所述MFCC缓存为一组，输入第一神经网络进行处理，得到识别概率，所述第一神经网络用于识别输入音频中存在目标人语音的概率；

判断所述识别概率是否大于概率阈值；

若所述识别概率大于概率阈值，则判定所述多通道音频中存在目标人语音；

若所述识别概率小于概率阈值，则判定所述多通道音频中不存在目标人语音。

[0006] 进一步的，所述判断所述识别概率是否大于概率阈值的步骤，包括：

判断所述识别概率是否为异常概率；

若所述识别概率为异常概率，则根据所述异常概率、所述异常概率的前一识别概率、所述异常概率的后一识别概率进行均值计算，得到修正概率，所述前一识别概率为所述异常概率前一组MFCC所对应的识别概率，所述后一识别概率为所述异常概率后一组MFCC所对应

的识别概率；

判断所述修正概率是否大于概率阈值；

若所述修正概率大于概率阈值，则判定所述识别概率大于概率阈值；

若所述修正概率小于概率阈值，则判定所述识别概率小于概率阈值。

[0007] 进一步的，所述按照预设规则对所述多通道音频进行处理，得到多个目标人语音相位谱的步骤，包括：

对所述多通道音频进行快速傅里叶变换，得到多个频域信号；

将各所述频域信号输入第二神经网络进行处理，得到各所述目标人语音相位谱，所述第二神经网络用于分离出输入音频信号中目标人语音的相位谱。

[0008] 进一步的，所述分别计算各所述目标人语音相位谱的可控响应功率，处理得到角谱的步骤，包括：

以麦克风阵列的几何中心为原点构建空间直角坐标系；

按照预设角度范围，在所述空间直角坐标系上选取若干个方向向量；

根据各所述方向向量和各所述目标人语音相位谱，计算得到各所述方向向量各自对应的所述可控响应功率；

将各所述方向向量转化为角度形式，得到各所述方向向量分别对应的水平角和俯仰角；

根据各所述可控响应功率与各所述方向向量分别对应的水平角和俯仰角之间的对应关系，生成所述角谱。

[0009] 进一步的，所述根据各所述方向向量和各所述目标人语音相位谱，计算得到各所述方向向量各自对应的所述可控响应功率的步骤，包括：

将各所述目标人语音相位谱转化为复数形式的目标人语音相位谱，并计算各所述方向向量所在方向分别到达所述麦克风阵列中两个麦克风的时间差；

将所述复数形式的目标人语音相位谱、所述时间差和所述方向向量代入第一公式中，计算得到所述麦克风阵列中两个麦克风所接收的音频帧数据之间的广义互相关函数，其中，所述第一公式为：

$$R_{ac}[\tau_{ac}(d_h)]=\frac{1}{K}\sum_{k=1}^{K}Y_a(k)Y_c^*(k)e^{j\Omega\tau_{ac}(d_h)}$$，$\tau_{ac}(d_h)$为所述时间差，$Y_a(k)$、$Y_c(k)$均为复数形式的目标人语音相位谱，d_h为所述方向向量，$R_{ac}[\tau_{ac}(d_h)]$为所述广义互相关函数；

将所述广义互相关函数代入第二公式中，计算得到各所述可控响应功率，其中，所述第二公式为：

$$F(d_h)=\sum_{a=1}^{C-1}\sum_{c=a+1}^{C}R_{ac}[\tau_{ac}(d_h)]$$，$F(d_h)$为所述可控响应功率。

[0010] 优选的，所述计算各所述方向向量所在方向分别到达所述麦克风阵列中两个麦克风的时间差的步骤，包括：

将所述方向向量代入第三公式中，计算得到对应的所述时间差，其中，所述第三公式为：$\tau_{ac}(d_h)=\frac{\|d_h-r_c\|-\|d_h-r_a\|}{v}$，$r_a$为第a个麦克风在所述空间直角坐标系中的坐标向

量，r_c为第c个麦克风在所述空间直角坐标系中的坐标向量，v为音速。

[0011] 本申请还提供了一种特定人声的声源定位装置，包括：

获取模块，用于获取多通道音频；

判断模块，用于判断所述多通道音频中是否存在目标人语音；

处理模块，用于若所述多通道音频中存在目标人语音，则按照预设规则对所述多通道音频进行处理，得到多个目标人语音相位谱；

计算模块，用于分别计算各所述目标人语音相位谱的可控响应功率，处理得到角谱；

搜索模块，用于对所述角谱进行峰值搜索，以最大峰值所对应的角度数据作为所述目标人语音的声源位置信息。

[0012] 进一步的，所述判断模块，包括：

提取子模块，用于从所述多通道音频中选取任一通道音频进行特征提取，得到各帧音频分别对应的MFCC；

第一处理子模块，用于将各所述MFCC缓存为一组，输入第一神经网络进行处理，得到识别概率，所述第一神经网络用于识别输入音频中存在目标人语音的概率；

判断子模块，用于判断所述识别概率是否大于概率阈值；

第一判定子模块，用于若所述识别概率大于概率阈值，则判定所述多通道音频中存在目标人语音；

第二判定子模块，用于若所述识别概率小于概率阈值，则判定所述多通道音频中不存在目标人语音。

[0013] 进一步的，所述判断子模块，包括：

第一判断单元，用于判断所述识别概率是否为异常概率；

第一计算单元，用于若所述识别概率为异常概率，则根据所述异常概率、所述异常概率的前一识别概率、所述异常概率的后一识别概率进行均值计算，得到修正概率，所述前一识别概率为所述异常概率前一组MFCC所对应的识别概率，所述后一识别概率为所述异常概率后一组MFCC所对应的识别概率；

第二判断单元，用于判断所述修正概率是否大于概率阈值；

第一判定单元，用于若所述修正概率大于概率阈值，则判定所述识别概率大于概率阈值；

第二判定单元，用于若所述修正概率小于概率阈值，则判定所述识别概率小于概率阈值。

[0014] 进一步的，所述处理模块，包括：

变换子模块，用于对所述多通道音频进行快速傅里叶变换，得到多个频域信号；

第二处理子模块，用于将各所述频域信号输入第二神经网络进行处理，得到各所述目标人语音相位谱，所述第二神经网络用于分离出输入音频信号中目标人语音的相位谱。

[0015] 进一步的，所述计算模块，包括：

构建子模块，用于以麦克风阵列的几何中心为原点构建空间直角坐标系；

选取子模块，用于按照预设角度范围，在所述空间直角坐标系上选取若干个方向向量；

计算子模块，用于根据各所述方向向量和各所述目标人语音相位谱，计算得到各所述方向向量各自对应的所述可控响应功率；

转化子模块，用于将各所述方向向量转化为角度形式，得到各所述方向向量分别对应的水平角和俯仰角；

生成子模块，用于根据各所述可控响应功率与各所述方向向量分别对应的水平角和俯仰角之间的对应关系，生成所述角谱。

[0016] 进一步的，所述计算子模块，包括：

第二计算单元，用于将各所述目标人语音相位谱转化为复数形式的目标人语音相位谱，并计算各所述方向向量所在方向分别到达所述麦克风阵列中两个麦克风的时间差；

第三计算单元，用于将所述复数形式的目标人语音相位谱、所述时间差和所述方向向量代入第一公式中，计算得到所述麦克风阵列中两个麦克风所接收的音频帧数据之间的广义互相关函数，其中，所述第一公式为：

$R_{ac}[\tau_{ac}(d_h)]=\frac{1}{K}\sum_{k=1}^{K}Y_a(k)Y_c^*(k)e^{j\Omega\tau_{ac}(d_h)}$，$\tau_{ac}(d_h)$为所述时间差，$Y_a(k)$、$Y_c(k)$均为复数形式的目标人语音相位谱，$d_h$为所述方向向量，$R_{ac}[\tau_{ac}(d_h)]$为所述广义互相关函数；

第四计算单元，用于将所述广义互相关函数代入第二公式中，计算得到各所述可控响应功率，其中，所述第二公式为：

$F(d_h)=\sum_{a=1}^{C-1}\sum_{c=a+1}^{C}R_{ac}[\tau_{ac}(d_h)]$，$F(d_h)$为所述可控响应功率。

[0017] 优选的，所述第二计算单元，包括：

计算子单元，用于将所述方向向量代入第三公式中，计算得到对应的所述时间差，其中，所述第三公式为：$\tau_{ac}(d_h)=\frac{\|d_h-r_c\|-\|d_h-r_a\|}{v}$，$r_a$为第a个麦克风在所述空间直角坐标系中的坐标向量，$r_c$为第c个麦克风在所述空间直角坐标系中的坐标向量，v为音速。

[0018] 本申请还提供一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机程序，所述处理器执行所述计算机程序时实现上述任一项所述方法的步骤。

[0019] 本申请还提供一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现上述任一项所述的方法的步骤。

[0020] 本申请中提供的一种特定人声的声源定位方法、装置和计算机设备，系统首先判断获取的多通道音频中是否存在目标人语音，若多通道音频中存在目标人语音，则按照预设规则对多通道音频进行处理，得到多个目标人语音相位谱。系统分别计算各目标人语音相位谱的可控响应功率，处理得到角谱。最后对角谱进行峰值搜索，以最大峰值所对应的角度数据作为目标人语音的声源位置信息。本申请在识别到目标人语音后，根据目标人语音相位谱进行相应的计算得到声源位置，在计算过程中并不涉及音频的功率谱，因而可以减少无关信息的干扰，从而准确定位特定人声的声源位置。

附图说明

[0021] 图1是本申请一实施例中特定人声的声源定位方法步骤示意图；

图2是本申请一实施例中特定人声的声源定位装置整体结构框图；

图3是本申请一实施例的计算机设备的结构示意框图。

[0022] 本申请目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

具体实施方式

[0023] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

[0024] 参照图1，本申请一实施例中提供了一种特定人声的声源定位方法，包括：

S1:获取多通道音频；

S2:判断所述多通道音频中是否存在目标人语音；

S3:若所述多通道音频中存在目标人语音，则按照预设规则对所述多通道音频进行处理，得到多个目标人语音相位谱；

S4:分别计算各所述目标人语音相位谱的可控响应功率，处理得到角谱；

S5:对所述角谱进行峰值搜索，以最大峰值所对应的角度数据作为所述目标人语音的声源位置信息。

[0025] 本实施例中，系统通过麦克风阵列采集C个通道的音频数据，C个通道的音频经过缓存器，缓存一定长度的音频数据(比如时长为10S的音频数据)，得到多通道音频。系统从多通道音频中选取任一通道音频实时进行特征提取，得到音频中各帧音频分别对应的MFCC(梅尔倒谱系数)。系统将各个MFCC进行组合后分别输入第一神经网络进行处理，得到各个MFCC组合分别对应的识别概率。其中，第一神经网络为深度学习网络(可以选择为若干层的LSTM+DNN+softmax，或者可以用GRU来替代LSTM，或者直接用多层的DNN来实现)，预先使用包含有目标人语音的样本进行训练(第一神经网络的训练过程与现有技术相同，在此不做详述)，从而使得第一神经网络可以判断输入音频中是否存在目标人语音。系统对各个识别概率进行平滑处理，去除异常数据，从而得到修正概率。系统将修正概率与概率阈值进行比对，如果修正概率大于概率阈值，则判定多通道音频中存在目标人语音。系统对多通道音频进行快速傅里叶变换，得到多个频域信号。然后，再将各个频域信号输入第二神经网络进行处理，得到各目标人语音相位谱。其中，第二神经网络用于分离出输入音频信号的相位谱，其训练过程为：对样本语音进行随机加混响、加噪声、加干扰语音等操作，继而对带噪混合语音进行FFT变换。将FFT变换后的结果送入第二神经网络，这里可以选择为CRN(Convolutional Recurrent Network，即卷积递归神经网络)+sigmod，CRN由若干层的CNN层加若干层LSTM层，再加若干层反CNN层组成。其输出为预测纯净目标语音的相位谱$\varphi'(k),k=1,2,...,K$，K是FFT的长度。这里选用MSE作为loss函数，通过与纯净目标语音的相位谱求MSE即，$loss_2=\frac{1}{B}\sum_{b=1}^{B}\frac{1}{K}\sum_{k=1}^{K}(\varphi(k)-\varphi'(k))^2$。最后通过Adam优化器，来调节网络参数，直至收敛。在通过第二神经网络得到目标人语音相位谱后，系统以麦克风阵列的几何中心为原点构建空间直角坐标系，然后按照预设角度范围(预设角度范围由用户根据麦克风阵列的部署位置进行相应设定)，以空间直角坐标系的原点为圆心的单位圆上，选取若干个方向向量。系统根据各个方向向量和各个目标人语音相位谱，计算得到各个方向向量各自对

应的可控响应功率。系统将各方向向量转化为角度形式，得到各方向向量分别对应的水平角和俯仰角。再根据各可控响应功率与各方向向量分别对应的水平角和俯仰角之间的对应关系，生成角谱(Angle spectrum)。系统对角谱进行峰值搜索，以最大峰值所对应的角度数据(水平角和俯仰角)作为目标人语音的声源位置信息，实现对特定人声，即目标人语音的声源定位。其中，峰值搜索的计算公式为$(\theta,\phi)=\arg\max_{\theta_h,\phi_h}\left(F_{peak}(\theta_h,\phi_h)\right)$，$\theta_h$为水平角，$\phi_h$为俯仰角。

[0026] 进一步的，所述判断所述多通道音频中是否存在目标人语音的步骤，包括：

S201:从所述多通道音频中选取任一通道音频进行特征提取，得到各帧音频分别对应的MFCC；

S202:将各所述MFCC缓存为一组，输入第一神经网络进行处理，得到识别概率，所述第一神经网络用于识别输入音频中存在目标人语音的概率；

S203:判断所述识别概率是否大于概率阈值；

S204:若所述识别概率大于概率阈值，则判定所述多通道音频中存在目标人语音；

S205:若所述识别概率小于概率阈值，则判定所述多通道音频中不存在目标人语音。

[0027] 本实施例中，系统从多通道音频中选取任一通道音频进行特征提取，得到各帧音频分别对应的MFCC。MFCC的提取过程与现有技术相同，依次为：预加重、分帧、加窗、快速傅里叶变换、三角带通滤波器、计算每个滤波器组输出的对数能量、动态差分参数的提取，在此不做详述。系统将各帧MFCC进行组合，比如需要20帧MFCC才能检测特定人声，[1,2,3,4,5,...,20]为一组，然后新进来的一帧MFCC (21)，则组成[2,3,4,5,6,...,21]为新的一组，当前组MFCC输入第一神经网络，从而得到当前组MFCC对应的识别概率。为了保证数据的准确度，需要对当前的识别概率进行平滑处理。具体地，系统判断当前组的识别概率中是否为异常概率，比如设置概率阈值为0.6，此时第一神经网络输出的三组MFCC分别对应的识别概率分别为0.3、0.9、0.4，而由于0.9相比于前后两个识别概率：0.3、0.4，两者之间(0.9与0.3之间、0.9与0.4之间)的差值过大，有可能是因为识别概率0.9所对应的音频帧数据异常，因此将0.9识别为异常概率，需要对识别概率0.9进行平滑处理。系统综合异常概率的前后两组所对应的识别概率，求其均值：(0.3+0.9+0.3)/3=0.5<0.6，0.5即为修正概率。系统判断修正概率是否大于概率阈值，如果修正概率大于概率阈值，则系统判定多通道音频中存在目标人语音。如果修正概率小于概率阈值，则系统判定多通道音频中不存在目标人语音。

[0028] 进一步的，所述判断所述识别概率是否大于概率阈值的步骤，包括：

S2031:判断所述识别概率是否为异常概率；

S2032:若所述识别概率为异常概率，则根据所述异常概率、所述异常概率的前一识别概率、所述异常概率的后一识别概率进行均值计算，得到修正概率，所述前一识别概率为所述异常概率前一组MFCC所对应的识别概率，所述后一识别概率为所述异常概率后一组MFCC所对应的识别概率；

S2033:判断所述修正概率是否大于概率阈值；

S2034:若所述修正概率大于概率阈值，则判定所述识别概率大于概率阈值；

S2035:若所述修正概率小于概率阈值，则判定所述识别概率小于概率阈值。

[0029] 本实施例中，为了避免异常数据对判断多通道音频中是否存在目标人语音的准确

度，需要对当前输出的识别概率进行平滑处理。具体地，系统在判断识别概率的过程中，根据当前的识别概率与前一识别概率和后一识别概率之前的差值大小(该差值可以由开发人员定义，也可以根据前一识别概率和后一识别概率进行设定，比如差值不能大于前一识别概率和/或后一识别概率)，来判断当前的识别概率是否为异常概率。比如，第一神经网络输出的相邻三组MFCC各自对应的识别概率分别为0.3、0.9、0.4，而由于0.9相比于前后两个识别概率：0.3、0.4，两者之间(0.9与0.3之间、0.9与0.4之间)的差值过大(差值已经大于0.3、0.4)，有可能是因为识别概率0.9所对应的音频帧数据异常，因此将0.9识别为异常概率。系统根据异常概率的前一识别概率、异常概率的后一识别概率进行均值计算(如果不存在前一识别概率或后一识别概率，则不存在的前一识别概率或后一识别概率的值取0)，得到修正概率。系统判断修正概率是否大于概率阈值，若修正概率大于概率阈值，则判定识别概率大于概率阈值。若修正概率小于概率阈值，则判定识别概率小于概率阈值。在后续对识别概率的平滑处理中，不会引入修正概率，仍采用第一神经网络输出的各个识别概率进行相应的平滑处理。

[0030] 进一步的，所述按照预设规则对所述多通道音频进行处理，得到多个目标人语音相位谱的步骤，包括：

S301:对所述多通道音频进行快速傅里叶变换，得到多个频域信号；

S302:将各所述频域信号输入第二神经网络进行处理，得到各所述目标人语音相位谱，所述第二神经网络用于分离出输入音频信号中目标人语音的相位谱。

[0031] 本实施例中，系统对多通道音频的时域信号进行快速傅里叶变换，得到各个通道的音频时域信号所对应的频域信号。系统将各个频域信号作为输入，送入第二神经网络中进行相应处理，预测得到各个频域信号分别对应的目标人语音相位谱。其中，第二神经网络用于分离出输入音频信号的相位谱，其具体训练过程为：对样本语音进行随机加混响、加噪声、加干扰语音等操作，继而对带噪混合语音进行FFT变换。将FFT变换后的结果送入第二神经网络，这里可以选择为CRN(Convolutional Recurrent Network，即卷积递归神经网络)+sigmod，CRN由若干层的CNN层加若干层LSTM层，再加若干层反CNN层组成。其输出为预测纯净目标语音的相位谱$\varphi'(k), k=1,2,...,K$，K是FFT的长度。这里选用MSE作为loss函数，通过与纯净目标语音的相位谱求MSE即，$loss_2=\frac{1}{B}\sum_{b=1}^{B}\frac{1}{K}\sum_{k=1}^{K}(\varphi(k)-\varphi'(k))^2$。最后通过Adam优化器，来调节网络参数，直至收敛。训练后的第二神经网络可以从输入的频域信号提取得到对应的相位谱。

[0032] 进一步的，所述分别计算各所述目标人语音相位谱的可控响应功率，处理得到角谱的步骤，包括：

S401:以麦克风阵列的几何中心为原点构建空间直角坐标系；

S402:按照预设角度范围，在所述空间直角坐标系上选取若干个方向向量；

S403:根据各所述方向向量和各所述目标人语音相位谱，计算得到各所述方向向量各自对应的所述可控响应功率；

S404:将各所述方向向量转化为角度形式，得到各所述方向向量分别对应的水平角和俯仰角；

S405:根据各所述可控响应功率与各所述方向向量分别对应的水平角和俯仰角之间的对应关系,生成所述角谱。

[0033] 本实施例中,系统以麦克风阵列的几何中心为原点构建空间直角坐标系,然后以空间直角坐标系的原点为圆心的单位球上,在对应预设角度范围的区域均匀选取若干个点。其中,预设角度范围由用户根据麦克风阵列的部署位置进行相应的设置。以坐标原点为方向向量的起点,以各个点作为方向向量的终点,从而得到若干个方向向量。系统首先将各个目标人语音相位谱转化为复数形式的目标人语音相位谱,并计算各个方向向量所在方向分别到达所述麦克风阵列中相邻两个麦克风的时间差。然后,将复数形式的目标人语音相位谱、时间差和方向向量代入第一公式中,计算得到麦克风阵列中相邻两个麦克风所接收的音频帧数据之间的广义互相关函数,其中,第一公式为:

$R_{ac}[\tau_{ac}(d_h)]=\frac{1}{K}\sum_{k=1}^{K}Y_a(k)Y_c^*(k)e^{j\Omega\tau_{ac}(d_h)}$,$\tau_{ac}(d_h)$表示方向向量$d_h$到达第a个和第c个麦克风的到达时间差,$Y_a(k)$、$Y_c(k)$均为复数形式的目标人语音相位谱,$d_h$为方向向量,$R_{ac}[\tau_{ac}(d_h)]$为广义互相关函数。系统将广义互相关函数代入第二公式中,计算得到各个方向向量分别对应的可控响应功率,其中,第二公式为:$F(d_h)=\sum_{a=1}^{C-1}\sum_{c=a+1}^{C}R_{ac}[\tau_{ac}(d_h)]$,$F(d_h)$为可控响应功率。系统将各个方向向量转化为角度形式,即$d_h=(\theta_h,\phi_h)$,其中θ_h为水平角,ϕ_h为俯仰角。系统将θ_h放入到水平角集合中,元素个数为E1;将ϕ_h放入到俯仰角集合中,元素个数为E2(E1、E2对应方向向量的个数)。系统将$F(d_h)$按照水平角和俯仰角的对应关系(或者说$F(d_h)$与方向向量d_h之间的对应关系),生成角谱。

[0034] 进一步的,所述根据各所述方向向量和各所述目标人语音相位谱,计算得到各所述方向向量各自对应的所述可控响应功率的步骤,包括:

S4031:将各所述目标人语音相位谱转化为复数形式的目标人语音相位谱,并计算各所述方向向量所在方向分别到达所述麦克风阵列中两个麦克风的时间差;

S4032:将所述复数形式的目标人语音相位谱、所述时间差和所述方向向量代入第一公式中,计算得到所述麦克风阵列中两个麦克风所接收的音频帧数据之间的广义互相关函数,其中,所述第一公式为:

$R_{ac}[\tau_{ac}(d_h)]=\frac{1}{K}\sum_{k=1}^{K}Y_a(k)Y_c^*(k)e^{j\Omega\tau_{ac}(d_h)}$,$\tau_{ac}(d_h)$为所述时间差,$Y_a(k)$、$Y_c(k)$均为复数形式的目标人语音相位谱,$d_h$为所述方向向量,$R_{ac}[\tau_{ac}(d_h)]$为所述广义互相关函数;

S4033:将所述广义互相关函数代入第二公式中,计算得到各所述可控响应功率,其中,所述第二公式为:

$F(d_h)=\sum_{a=1}^{C-1}\sum_{c=a+1}^{C}R_{ac}[\tau_{ac}(d_h)]$,$F(d_h)$为所述可控响应功率。

[0035] 优选的,所述计算各所述方向向量所在方向分别到达所述麦克风阵列中两个麦克

风的时间差的步骤，包括：

S40311：将所述方向向量代入第三公式中，计算得到对应的所述时间差，其中，所述第三公式为：$\tau_{ac}(d_h)=\frac{\|d_h-r_c\|-\|d_h-r_a\|}{v}$，$r_a$ 为第a个麦克风在所述空间直角坐标系中的坐标向量，r_c 为第c个麦克风在所述空间直角坐标系中的坐标向量，v为音速。

[0036] 本实施例中，系统根据第四公式将各个目标人语音相位谱转化为复数形式的目标人语音相位谱，其中，第四公式为：$Y_a(k)=\cos(\varphi_a(k))+j\sin(\varphi_a(k))$，$\varphi_a$ 为目标人语音相位谱，$Y_a(k)$ 为复数形式的目标人语音相位谱。并且，系统将各个方向向量代入第三公式中，计算得到各个方向向量所在方向分别到达麦克风阵列中两个麦克风的时间差。其中，第三公式为：$\tau_{ac}(d_h)=\frac{\|d_h-r_c\|-\|d_h-r_a\|}{v}$，$r_a$ 为第a个麦克风在所述空间直角坐标系中的坐标向量，r_c 为第c个麦克风在所述空间直角坐标系中的坐标向量，v为音速。系统将复数形式的目标人语音相位谱、时间差和方向向量代入第一公式中，计算得到麦克风阵列中两个麦克风所接收的音频帧数据之间的广义互相关函数。其中，第一公式为：$R_{ac}[\tau_{ac}(d_h)]=\frac{1}{K}\sum_{k=1}^{K}Y_a(k)Y_c^*(k)e^{j\Omega\tau_{ac}(d_h)}$，$\tau_{ac}(d_h)$ 为时间差，$Y_a(k)$、$Y_c(k)$ 均为复数形式的目标人语音相位谱，d_h 为方向向量，$R_{ac}[\tau_{ac}(d_h)]$ 即为广义互相关函数。系统将广义互相关函数代入第二公式中，计算得到各个方向向量分别对应的可控响应功率。其中，第二公式为：$F(d_h)=\sum_{a=1}^{C-1}\sum_{c=a+1}^{C}R_{ac}[\tau_{ac}(d_h)]$，$F(d_h)$ 为所述可控响应功率。

[0037] 本实施例提供的一种特定人声的声源定位方法，系统首先判断获取的多通道音频中是否存在目标人语音，若多通道音频中存在目标人语音，则按照预设规则对多通道音频进行处理，得到多个目标人语音相位谱。系统分别计算各目标人语音相位谱的可控响应功率，处理得到角谱。最后对角谱进行峰值搜索，以最大峰值所对应的角度数据作为目标人语音的声源位置信息。本申请在识别到目标人语音后，根据目标人语音相位谱进行相应的计算得到声源位置，在计算过程中并不涉及音频的功率谱，因而可以减少无关信息的干扰，从而准确定位特定人声的声源位置。

[0038] 参照图2，本申请一实施例中还提供了一种特定人声的声源定位装置，包括：

获取模块，用于获取多通道音频；

判断模块，用于判断所述多通道音频中是否存在目标人语音；

处理模块，用于若所述多通道音频中存在目标人语音，则按照预设规则对所述多通道音频进行处理，得到多个目标人语音相位谱；

计算模块，用于分别计算各所述目标人语音相位谱的可控响应功率，处理得到角谱；

搜索模块，用于对所述角谱进行峰值搜索，以最大峰值所对应的角度数据作为所述目标人语音的声源位置信息。

[0039] 本实施例中，系统通过麦克风阵列采集C个通道的音频数据，C个通道的音频经过

缓存器，缓存一定长度的音频数据(比如时长为10S的音频数据)，得到多通道音频。系统从多通道音频中选取任一通道音频实时进行特征提取，得到音频中各帧音频分别对应的MFCC(梅尔倒谱系数)。系统将各个MFCC进行组合后分别输入第一神经网络进行处理，得到各个MFCC组合分别对应的识别概率。其中，第一神经网络为深度学习网络(可以选择为若干层的LSTM+DNN+softmax，或者可以用GRU来替代LSTM，或者直接用多层的DNN来实现)，预先使用包含有目标人语音的样本进行训练(第一神经网络的训练过程与现有技术相同，在此不做详述)，从而使得第一神经网络可以判断输入音频中是否存在目标人语音。系统对各个识别概率进行平滑处理，去除异常数据，从而得到修正概率。系统将修正概率与概率阈值进行比对，如果修正概率大于概率阈值，则判定多通道音频中存在目标人语音。系统对多通道音频进行快速傅里叶变换，得到多个频域信号。然后，再将各个频域信号输入第二神经网络进行处理，得到各目标人语音相位谱。其中，第二神经网络用于分离出输入音频信号的相位谱，其训练过程为：对样本语音进行随机加混响、加噪声、加干扰语音等操作，继而对带噪混合语音进行FFT变换。将FFT变换后的结果送入第二神经网络，这里可以选择为CRN(Convolutional Recurrent Network，即卷积递归神经网络)+sigmod，CRN由若干层的CNN层加若干层LSTM层，再加若干层反CNN层组成。其输出为预测纯净目标语音的相位谱$\varphi'(k), k=1,2,...,K$，K是FFT的长度。这里选用MSE作为loss函数，通过与纯净目标语音的相位谱求MSE即，$loss_2=\frac{1}{B}\sum_{b=1}^{B}\frac{1}{K}\sum_{k=1}^{K}(\varphi(k)-\varphi'(k))^2$。最后通过Adam优化器，来调节网络参数，直至收敛。在通过第二神经网络得到目标人语音相位谱后，系统以麦克风阵列的几何中心为原点构建空间直角坐标系，然后按照预设角度范围(预设角度范围由用户根据麦克风阵列的部署位置进行相应设定)，以空间直角坐标系的原点为圆心的单位圆上，选取若干个方向向量。系统根据各个方向向量和各个目标人语音相位谱，计算得到各个方向向量各自对应的可控响应功率。系统将各方向向量转化为角度形式，得到各方向向量分别对应的水平角和俯仰角。再根据各可控响应功率与各方向向量分别对应的水平角和俯仰角之间的对应关系，生成角谱(Angle spectrum)。系统对角谱进行峰值搜索，以最大峰值所对应的角度数据(水平角和俯仰角)作为目标人语音的声源位置信息，实现对特定人声，即目标人语音的声源定位。其中，峰值搜索的计算公式为$(\theta,\phi)=\arg\max_{\theta_h,\phi_h}(F_{peak}(\theta_h,\phi_h))$，$\theta_h$为水平角，$\phi_h$为俯仰角。

[0040] 进一步的，所述判断模块，包括：

提取子模块，用于从所述多通道音频中选取任一通道音频进行特征提取，得到各帧音频分别对应的MFCC；

第一处理子模块，用于将各所述MFCC缓存为一组，输入第一神经网络进行处理，得到识别概率，所述第一神经网络用于识别输入音频中存在目标人语音的概率；

判断子模块，用于判断所述识别概率是否大于概率阈值；

第一判定子模块，用于若所述识别概率大于概率阈值，则判定所述多通道音频中存在目标人语音；

第二判定子模块，用于若所述识别概率小于概率阈值，则判定所述多通道音频中不存在目标人语音。

[0041] 本实施例中，系统从多通道音频中选取任一通道音频进行特征提取，得到各帧音

频分别对应的MFCC。MFCC的提取过程与现有技术相同，依次为：预加重、分帧、加窗、快速傅里叶变换、三角带通滤波器、计算每个滤波器组输出的对数能量、动态差分参数的提取，在此不做详述。系统将各帧MFCC进行组合，比如需要20帧MFCC才能检测特定人声，[1,2,3,4,5,...,20]为一组，然后新进来的一帧MFCC (21)，则组成[2,3,4,5,6,...,21]为新的一组，当前组MFCC输入第一神经网络，从而得到当前组MFCC对应的识别概率。为了保证数据的准确度，需要对当前的识别概率进行平滑处理。具体地，系统判断当前组的识别概率中是否为异常概率，比如设置概率阈值为0.6，此时第一神经网络输出的三组MFCC分别对应的识别概率分别为0.3、0.9、0.4，而由于0.9相比于前后两个识别概率：0.3、0.4，两者之间(0.9与0.3之间、0.9与0.4之间)的差值过大，有可能是因为识别概率0.9所对应的音频帧数据异常，因此将0.9识别为异常概率，需要对识别概率0.9进行平滑处理。系统综合异常概率的前后两组所对应的识别概率，求其均值：(0.3+0.9+0.3)/3=0.5<0.6，0.5即为修正概率。系统判断修正概率是否大于概率阈值，如果修正概率大于概率阈值，则系统判定多通道音频中存在目标人语音。如果修正概率小于概率阈值，则系统判定多通道音频中不存在目标人语音。

[0042] 进一步的，所述判断子模块，包括：

第一判断单元，用于判断所述识别概率是否为异常概率；

第一计算单元，用于若所述识别概率为异常概率，则根据所述异常概率、所述异常概率的前一识别概率、所述异常概率的后一识别概率进行均值计算，得到修正概率，所述前一识别概率为所述异常概率前一组MFCC所对应的识别概率，所述后一识别概率为所述异常概率后一组MFCC所对应的识别概率；

第二判断单元，用于判断所述修正概率是否大于概率阈值；

第一判定单元，用于若所述修正概率大于概率阈值，则判定所述识别概率大于概率阈值；

第二判定单元，用于若所述修正概率小于概率阈值，则判定所述识别概率小于概率阈值。

[0043] 本实施例中，为了避免异常数据对判断多通道音频中是否存在目标人语音的准确度，需要对当前输出的识别概率进行平滑处理。具体地，系统在判断识别概率的过程中，根据当前的识别概率与前一识别概率和后一识别概率之前的差值大小(该差值可以由开发人员定义，也可以根据前一识别概率和后一识别概率进行设定，比如差值不能大于前一识别概率和/或后一识别概率)，来判断当前的识别概率是否为异常概率。比如，第一神经网络输出的相邻三组MFCC各自对应的识别概率分别为0.3、0.9、0.4，而由于0.9相比于前后两个识别概率：0.3、0.4，两者之间(0.9与0.3之间、0.9与0.4之间)的差值过大(差值已经大于0.3、0.4)，有可能是因为识别概率0.9所对应的音频帧数据异常，因此将0.9识别为异常概率。系统根据异常概率的前一识别概率、异常概率的后一识别概率进行均值计算(如果不存在前一识别概率或后一识别概率，则不存在的前一识别概率或后一识别概率的值取0)，得到修正概率。系统判断修正概率是否大于概率阈值，若修正概率大于概率阈值，则判定识别概率大于概率阈值。若修正概率小于概率阈值，则判定识别概率小于概率阈值。在后续对识别概率的平滑处理中，不会引入修正概率，仍采用第一神经网络输出的各个识别概率进行相应的平滑处理。

[0044] 进一步的，所述处理模块，包括：

变换子模块，用于对所述多通道音频进行快速傅里叶变换，得到多个频域信号；

第二处理子模块，用于将各所述频域信号输入第二神经网络进行处理，得到各所述目标人语音相位谱，所述第二神经网络用于分离出输入音频信号中目标人语音的相位谱。

[0045] 本实施例中，系统对多通道音频的时域信号进行快速傅里叶变换，得到各个通道的音频时域信号所对应的频域信号。系统将各个频域信号作为输入，送入第二神经网络中进行相应处理，预测得到各个频域信号分别对应的目标人语音相位谱。其中，第二神经网络用于分离出输入音频信号的相位谱，其具体训练过程为：对样本语音进行随机加混响、加噪声、加干扰语音等操作，继而对带噪混合语音进行FFT变换。将FFT变换后的结果送入第二神经网络，这里可以选择为CRN(Convolutional Recurrent Network，即卷积递归神经网络)+sigmod，CRN由若干层的CNN层加若干层LSTM层，再加若干层反CNN层组成。其输出为预测纯净目标语音的相位谱$\varphi'(k), k=1,2,...,K$，K是FFT的长度。这里选用MSE作为loss函数，通过与纯净目标语音的相位谱求MSE即，$loss_2=\frac{1}{B}\sum_{b=1}^{B}\frac{1}{K}\sum_{k=1}^{K}(\varphi(k)-\varphi'(k))^2$。最后通过Adam优化器，来调节网络参数，直至收敛。训练后的第二神经网络可以从输入的频域信号提取得到对应的相位谱。

[0046] 进一步的，所述计算模块，包括：

构建子模块，用于以麦克风阵列的几何中心为原点构建空间直角坐标系；

选取子模块，用于按照预设角度范围，在所述空间直角坐标系上选取若干个方向向量；

计算子模块，用于根据各所述方向向量和各所述目标人语音相位谱，计算得到各所述方向向量各自对应的所述可控响应功率；

转化子模块，用于将各所述方向向量转化为角度形式，得到各所述方向向量分别对应的水平角和俯仰角；

生成子模块，用于根据各所述可控响应功率与各所述方向向量分别对应的水平角和俯仰角之间的对应关系，生成所述角谱。

[0047] 本实施例中，系统以麦克风阵列的几何中心为原点构建空间直角坐标系，然后以空间直角坐标系的原点为圆心的单位球上，在对应预设角度范围的区域均匀选取若干个点。其中，预设角度范围由用户根据麦克风阵列的部署位置进行相应的设置。以坐标原点为方向向量的起点，以各个点作为方向向量的终点，从而得到若干个方向向量。系统首先将各个目标人语音相位谱转化为复数形式的目标人语音相位谱，并计算各个方向向量所在方向分别到达所述麦克风阵列中相邻两个麦克风的时间差。然后，将复数形式的目标人语音相位谱、时间差和方向向量代入第一公式中，计算得到麦克风阵列中相邻两个麦克风所接收的音频帧数据之间的广义互相关函数，其中，第一公式为：

$R_{ac}[\tau_{ac}(d_h)]=\frac{1}{K}\sum_{k=1}^{K}Y_a(k)Y_c^*(k)e^{j\Omega\tau_{ac}(d_h)}$，$\tau_{ac}(d_h)$为表示方向向量$d_h$到达第a个和第c个麦克风的到达时间差，$Y_a(k)$、$Y_c(k)$均为复数形式的目标人语音相位谱，$d_h$为方向向量，$R_{ac}[\tau_{ac}(d_h)]$为广义互相关函数。系统将广义互相关函数代入第二公式中，计算得到各个方

向向量分别对应的可控响应功率，其中，第二公式为：$F(d_h)=\sum_{a=1}^{C-1}\sum_{c=a+1}^{C}R_{ac}[\tau_{ac}(d_h)]$，$F(d_h)$为可控响应功率。系统将各个方向向量转化为角度形式，即$d_h=(\theta_h,\phi_h)$，其中ϕ_h为水平角，ϕ_h为俯仰角。系统将θ_h放入到水平角集合中，元素个数为E1；将ϕ_h放入到俯仰角集合中，元素个数为E2（E1、E2对应方向向量的个数）。系统将$F(d_h)$按照水平角和俯仰角的对应关系（或者说$F(d_h)$与方向向量d_h之间的对应关系），生成角谱。

[0048] 进一步的，所述计算子模块，包括：

第二计算单元，用于将各所述目标人语音相位谱转化为复数形式的目标人语音相位谱，并计算各所述方向向量所在方向分别到达所述麦克风阵列中两个麦克风的时间差；

第三计算单元，用于将所述复数形式的目标人语音相位谱、所述时间差和所述方向向量代入第一公式中，计算得到所述麦克风阵列中两个麦克风所接收的音频帧数据之间的广义互相关函数，其中，所述第一公式为：

$R_{ac}[\tau_{ac}(d_h)]=\frac{1}{K}\sum_{k=1}^{K}Y_a(k)Y_c^*(k)e^{j\Omega\tau_{ac}(d_h)}$，$\tau_{ac}(d_h)$为所述时间差，$Y_a(k)$、$Y_c(k)$均为复数形式的目标人语音相位谱，$d_h$为所述方向向量，$R_{ac}[\tau_{ac}(d_h)]$为所述广义互相关函数；

第四计算单元，用于将所述广义互相关函数代入第二公式中，计算得到各所述可控响应功率，其中，所述第二公式为：

$F(d_h)=\sum_{a=1}^{C-1}\sum_{c=a+1}^{C}R_{ac}[\tau_{ac}(d_h)]$，$F(d_h)$为所述可控响应功率。

[0049] 优选的，所述第二计算单元，包括：

计算子单元，用于将所述方向向量代入第三公式中，计算得到对应的所述时间差，其中，所述第三公式为：$\tau_{ac}(d_h)=\frac{\|d_h-r_c\|-\|d_h-r_a\|}{v}$，$r_a$为第a个麦克风在所述空间直角坐标系中的坐标向量，$r_c$为第c个麦克风在所述空间直角坐标系中的坐标向量，v为音速。

[0050] 本实施例中，系统根据第四公式将各个目标人语音相位谱转化为复数形式的目标人语音相位谱，其中，第四公式为：$Y_a(k)=\cos(\varphi_a(k))+j\sin(\varphi_a(k))$，$\varphi_a$为目标人语音相位谱，$Y_a(k)$为复数形式的目标人语音相位谱。并且，系统将各个方向向量代入第三公式中，计算得到各个方向向量所在方向分别到达麦克风阵列中相邻两个麦克风的时间差。其中，第三公式为：$\tau_{ac}(d_h)=\frac{\|d_h-r_c\|-\|d_h-r_a\|}{v}$，$r_a$为第a个麦克风在所述空间直角坐标系中的坐标向量，$r_c$为第c个麦克风在所述空间直角坐标系中的坐标向量，v为音速。系统将复数形式的目标人语音相位谱、时间差和方向向量代入第一公式中，计算得到麦克风阵列中相邻两个麦克风所接收的音频帧数据之间的广义互相关函数。其中，第一公式为：$R_{ac}[\tau_{ac}(d_h)]=\frac{1}{K}\sum_{k=1}^{K}Y_a(k)Y_c^*(k)e^{j\Omega\tau_{ac}(d_h)}$，$\tau_{ac}(d_h)$为时间差，$Y_a(k)$、$Y_c(k)$均为复数形式

的目标人语音相位谱，d_h为方向向量，$R_{ac}[\tau_{ac}(d_h)]$即为广义互相关函数。系统将广义互相关函数代入第二公式中，计算得到各个方向向量分别对应的可控响应功率。其中，第二公式为：$F(d_h)=\sum_{a=1}^{C-1}\sum_{c=a+1}^{C}R_{ac}[\tau_{ac}(d_h)]$，$F(d_h)$为所述可控响应功率。

[0051] 本实施例提供的一种特定人声的声源定位装置，系统首先判断获取的多通道音频中是否存在目标人语音，若多通道音频中存在目标人语音，则按照预设规则对多通道音频进行处理，得到多个目标人语音相位谱。系统分别计算各目标人语音相位谱的可控响应功率，处理得到角谱。最后对角谱进行峰值搜索，以最大峰值所对应的角度数据作为目标人语音的声源位置信息。本申请在识别到目标人语音后，根据目标人语音相位谱进行相应的计算得到声源位置，在计算过程中并不涉及音频的功率谱，因而可以减少无关信息的干扰，从而准确定位特定人声的声源位置。

[0052] 参照图3，本申请实施例中还提供一种计算机设备，该计算机设备可以是服务器，其内部结构可以如图3所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设计的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储第一公式等数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种特定人声的声源定位方法。

[0053] 上述处理器执行上述特定人声的声源定位方法的步骤：

S1:获取多通道音频；

S2:判断所述多通道音频中是否存在目标人语音；

S3:若所述多通道音频中存在目标人语音，则按照预设规则对所述多通道音频进行处理，得到多个目标人语音相位谱；

S4:分别计算各所述目标人语音相位谱的可控响应功率，处理得到角谱；

S5:对所述角谱进行峰值搜索，以最大峰值所对应的角度数据作为所述目标人语音的声源位置信息。

[0054] 进一步的，所述判断所述多通道音频中是否存在目标人语音的步骤，包括：

S201:从所述多通道音频中选取任一通道音频进行特征提取，得到各帧音频分别对应的MFCC；

S202:将各所述MFCC缓存为一组，输入第一神经网络进行处理，得到识别概率，所述第一神经网络用于识别输入音频中存在目标人语音的概率；

S203:判断所述识别概率是否大于概率阈值；

S204:若所述识别概率大于概率阈值，则判定所述多通道音频中存在目标人语音；

S205:若所述识别概率小于概率阈值，则判定所述多通道音频中不存在目标人语音。

[0055] 进一步的，所述判断各所述识别概率是否大于概率阈值的步骤，包括：

S2031:判断所述识别概率是否为异常概率；

S2032:若所述识别概率为异常概率，则根据所述异常概率、所述异常概率的前一识别

概率、所述异常概率的后一识别概率进行均值计算，得到修正概率，所述前一识别概率为所述异常概率前一组MFCC所对应的识别概率，所述后一识别概率为所述异常概率后一组MFCC所对应的识别概率；

S2033:判断所述修正概率是否大于概率阈值；

S2034:若所述修正概率大于概率阈值，则判定所述识别概率大于概率阈值；

S2035:若所述修正概率小于概率阈值，则判定所述识别概率小于概率阈值。

[0056] 进一步的，所述按照预设规则对所述多通道音频进行处理，得到多个目标人语音相位谱的步骤，包括：

S301:对所述多通道音频进行快速傅里叶变换，得到多个频域信号；

S302:将各所述频域信号输入第二神经网络进行处理，得到各所述目标人语音相位谱，所述第二神经网络用于分离出输入音频信号中目标人语音的相位谱。

[0057] 进一步的，所述分别计算各所述目标人语音相位谱的可控响应功率，处理得到角谱的步骤，包括：

S401:以麦克风阵列的几何中心为原点构建空间直角坐标系；

S402:按照预设角度范围，在所述空间直角坐标系上选取若干个方向向量；

S403:根据各所述方向向量和各所述目标人语音相位谱，计算得到各所述方向向量各自对应的所述可控响应功率；

S404:将各所述方向向量转化为角度形式，得到各所述方向向量分别对应的水平角和俯仰角；

S405:根据各所述可控响应功率与各所述方向向量分别对应的水平角和俯仰角之间的对应关系，生成所述角谱。

[0058] 进一步的，所述根据各所述方向向量和各所述目标人语音相位谱，计算得到各所述方向向量各自对应的所述可控响应功率的步骤，包括：

S4031:将各所述目标人语音相位谱转化为复数形式的目标人语音相位谱，并计算各所述方向向量所在方向分别到达所述麦克风阵列中两个麦克风的时间差；

S4032:将所述复数形式的目标人语音相位谱、所述时间差和所述方向向量代入第一公式中，计算得到所述麦克风阵列中两个麦克风所接收的音频帧数据之间的广义互相关函数，其中，所述第一公式为：

$$R_{ac}[\tau_{ac}(d_h)]=\frac{1}{K}\sum_{k=1}^{K}Y_a(k)Y_c^*(k)e^{j\Omega\tau_{ac}(d_h)}$$，$\tau_{ac}(d_h)$为所述时间差，$Y_a(k)$、$Y_c(k)$均为复数形式的目标人语音相位谱，d_h为所述方向向量，$R_{ac}[\tau_{ac}(d_h)]$为所述广义互相关函数；

S4033:将所述广义互相关函数代入第二公式中，计算得到各所述可控响应功率，其中，所述第二公式为：

$$F(d_h)=\sum_{a=1}^{C-1}\sum_{c=a+1}^{C}R_{ac}[\tau_{ac}(d_h)]$$，$F(d_h)$为所述可控响应功率。

[0059] 优选的，所述计算各所述方向向量所在方向分别到达所述麦克风阵列中两个麦克

风的时间差的步骤，包括：

S40311:将所述方向向量代入第三公式中，计算得到对应的所述时间差，其中，所述第三公式为：$\tau_{ac}(d_h)=\frac{\|d_h-r_c\|-\|d_h-r_a\|}{v}$，$r_a$为第a个麦克风在所述空间直角坐标系中的坐标向量，$r_c$为第c个麦克风在所述空间直角坐标系中的坐标向量，v为音速。

[0060] 本申请一实施例还提供一种计算机可读存储介质，其上存储有计算机程序，计算机程序被处理器执行时实现一种特定人声的声源定位方法，其中，所述声源定位方法具体为：

S1:获取多通道音频；

S2:判断所述多通道音频中是否存在目标人语音；

S3:若所述多通道音频中存在目标人语音，则按照预设规则对所述多通道音频进行处理，得到多个目标人语音相位谱；

S4:分别计算各所述目标人语音相位谱的可控响应功率，处理得到角谱；

S5:对所述角谱进行峰值搜索，以最大峰值所对应的角度数据作为所述目标人语音的声源位置信息。

[0061] 进一步的，所述判断所述多通道音频中是否存在目标人语音的步骤，包括：

S201:从所述多通道音频中选取任一通道音频进行特征提取，得到各帧音频分别对应的MFCC；

S202:将各所述MFCC缓存为一组，输入第一神经网络进行处理，得到识别概率，所述第一神经网络用于识别输入音频中存在目标人语音的概率；

S203:判断所述识别概率是否大于概率阈值；

S204:若所述识别概率大于概率阈值，则判定所述多通道音频中存在目标人语音；

S205:若所述识别概率小于概率阈值，则判定所述多通道音频中不存在目标人语音。

[0062] 进一步的，所述判断各所述识别概率是否大于概率阈值的步骤，包括：

S2031:判断所述识别概率是否为异常概率；

S2032:若所述识别概率为异常概率，则根据所述异常概率、所述异常概率的前一识别概率、所述异常概率的后一识别概率进行均值计算，得到修正概率，所述前一识别概率为所述异常概率前一组MFCC所对应的识别概率，所述后一识别概率为所述异常概率后一组MFCC所对应的识别概率；

S2033:判断所述修正概率是否大于概率阈值；

S2034:若所述修正概率大于概率阈值，则判定所述识别概率大于概率阈值；

S2035:若所述修正概率小于概率阈值，则判定所述识别概率小于概率阈值。

[0063] 进一步的，所述按照预设规则对所述多通道音频进行处理，得到多个目标人语音相位谱的步骤，包括：

S301:对所述多通道音频进行快速傅里叶变换，得到多个频域信号；

S302:将各所述频域信号输入第二神经网络进行处理，得到各所述目标人语音相位谱，所述第二神经网络用于分离出输入音频信号中目标人语音的相位谱。

[0064] 进一步的，所述分别计算各所述目标人语音相位谱的可控响应功率，处理得到角

谱的步骤，包括：

S401:以麦克风阵列的几何中心为原点构建空间直角坐标系；

S402:按照预设角度范围，在所述空间直角坐标系上选取若干个方向向量；

S403:根据各所述方向向量和各所述目标人语音相位谱，计算得到各所述方向向量各自对应的所述可控响应功率；

S404:将各所述方向向量转化为角度形式，得到各所述方向向量分别对应的水平角和俯仰角；

S405:根据各所述可控响应功率与各所述方向向量分别对应的水平角和俯仰角之间的对应关系，生成所述角谱。

[0065] 进一步的，所述根据各所述方向向量和各所述目标人语音相位谱，计算得到各所述方向向量各自对应的所述可控响应功率的步骤，包括：

S4031:将各所述目标人语音相位谱转化为复数形式的目标人语音相位谱，并计算各所述方向向量所在方向分别到达所述麦克风阵列中两个麦克风的时间差；

S4032:将所述复数形式的目标人语音相位谱、所述时间差和所述方向向量代入第一公式中，计算得到所述麦克风阵列中两个麦克风所接收的音频帧数据之间的广义互相关函数，其中，所述第一公式为：

$$R_{ac}[\tau_{ac}(d_h)]=\frac{1}{K}\sum_{k=1}^{K}Y_a(k)Y_c^*(k)e^{j\Omega\tau_{ac}(d_h)}$$，$\tau_{ac}(d_h)$为所述时间差，$Y_a(k)$、$Y_c(k)$均为复数形式的目标人语音相位谱，d_h为所述方向向量，$R_{ac}[\tau_{ac}(d_h)]$为所述广义互相关函数；

S4033:将所述广义互相关函数代入第二公式中，计算得到各所述可控响应功率，其中，所述第二公式为：

$$F(d_h)=\sum_{a=1}^{C-1}\sum_{c=a+1}^{C}R_{ac}[\tau_{ac}(d_h)]$$，$F(d_h)$为所述可控响应功率。

[0066] 优选的，所述计算各所述方向向量所在方向分别到达所述麦克风阵列中两个麦克风的时间差的步骤，包括：

S40311:将所述方向向量代入第三公式中，计算得到对应的所述时间差，其中，所述第三公式为：$\tau_{ac}(d_h)=\frac{\|d_h-r_c\|-\|d_h-r_a\|}{v}$，$r_a$为第a个麦克风在所述空间直角坐标系中的坐标向量，$r_c$为第c个麦克风在所述空间直角坐标系中的坐标向量，v为音速。

[0067] 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，所述的计算机程序可存储与一非易失性计算机可读取存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的和实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可以包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括

随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限,RAM通过多种形式可得,诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双速据率SDRAM(SSRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0068] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其它变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、装置、物品或者方法不仅包括那些要素,而且还包括没有明确列出的其它要素,或者是还包括为这种过程、装置、物品或者方法所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、装置、物品或者方法中还存在另外的相同要素。

[0069] 以上所述仅为本申请的优选实施例,并非因此限制本申请的专利范围,凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其它相关的技术领域,均同理包括在本申请的专利保护范围内。

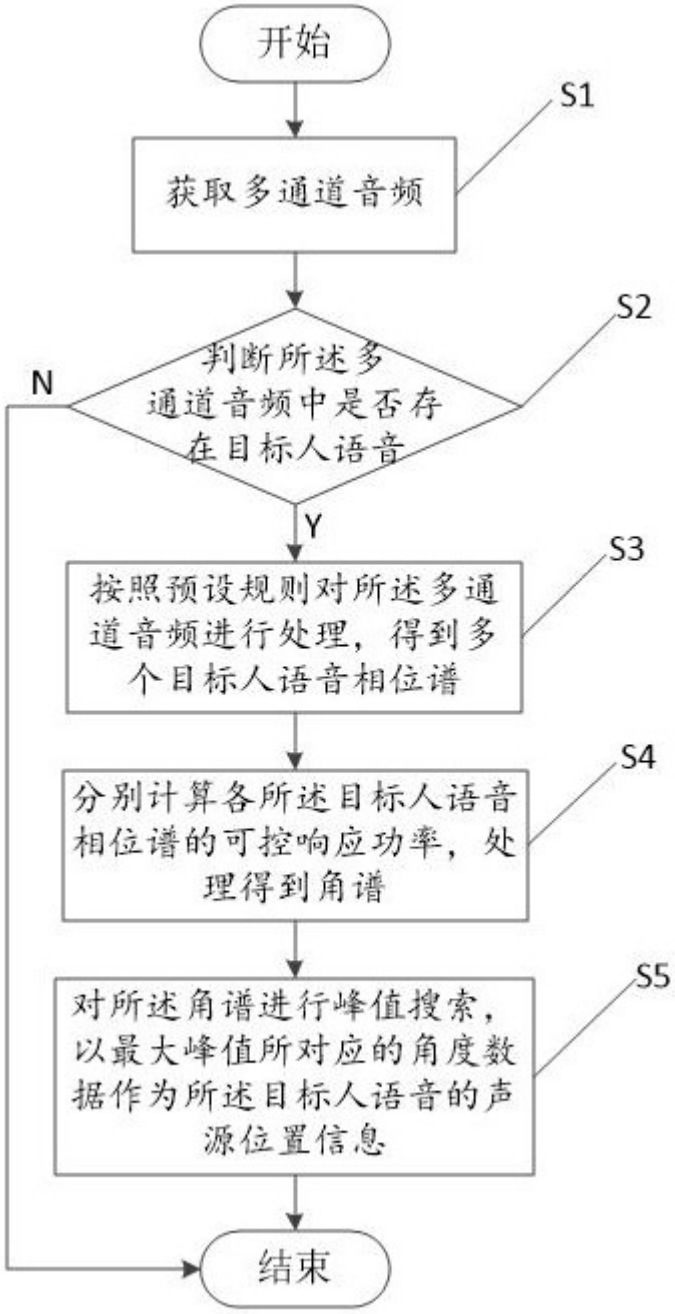

开始
S1
获取多通道音频
S2
判断所述多通道音频中是否存在目标人语音
N
Y
S3
按照预设规则对所述多通道音频进行处理，得到多个目标人语音相位谱
S4
分别计算各所述目标人语音相位谱的可控响应功率，处理得到角谱
S5
对所述角谱进行峰值搜索，以最大峰值所对应的角度数据作为所述目标人语音的声源位置信息
结束

图1

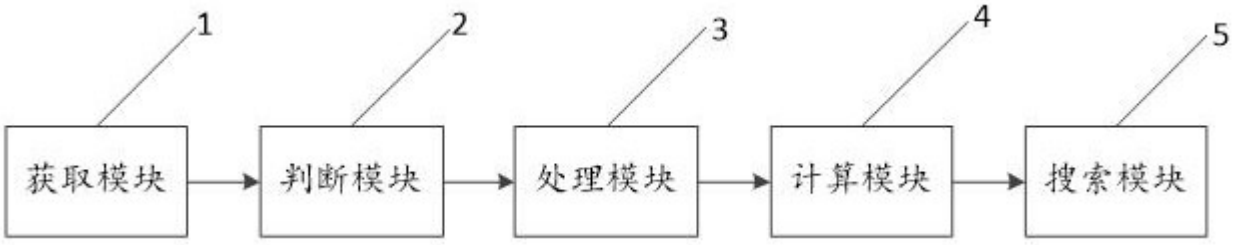

1
2
3
4
5
获取模块
判断模块
处理模块
计算模块
搜索模块

图2

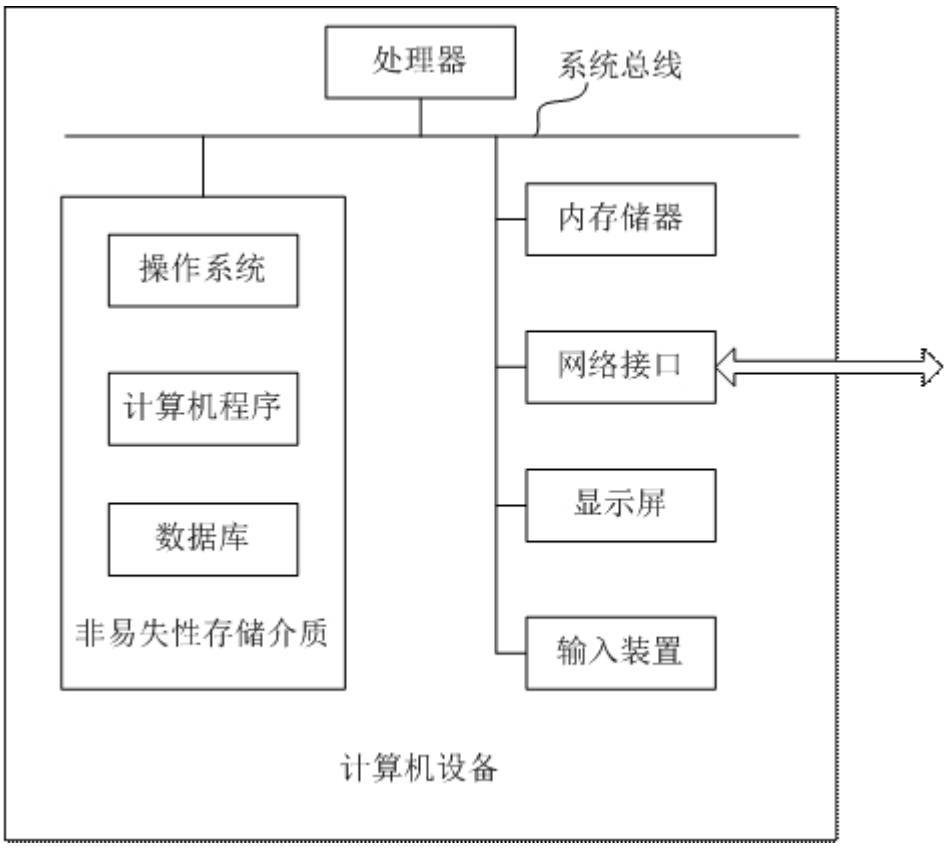


图3