(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 106294331 B
(45)授权公告日 2020.01.21

(21)申请号 201510235892.8
(22)申请日 2015.05.11
(65)同一申请的已公布的文献号
申请公布号 CN 106294331 A
(43)申请公布日 2017.01.04
(73)专利权人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼资本大厦一座四层847号邮箱
(72)发明人 易东 肖业鸣 刘荣 张伦 楚汝峰
(74)专利代理机构 北京鸿德海业知识产权代理事务所(普通合伙) 11412
代理人 倪志华
(51)Int.Cl.
G06F 16/63(2019.01)
G06F 16/683(2019.01)

(56)对比文件
CN 1647160 A,2005.07.27,
CN 103729368 A,2014.04.16,
CN 1647160 A,2005.07.27,
CN 102959543 A,2013.03.06,
CN 103093761 A,2013.05.08,

审查员 马金驹

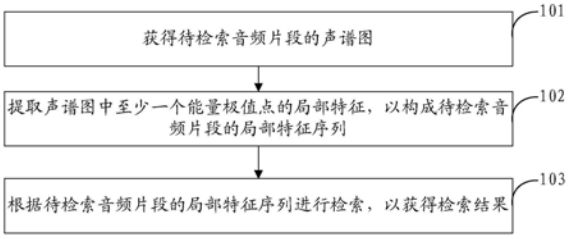
权利要求书4页 说明书10页 附图4页

(54)发明名称

音频信息检索方法及装置

(57)摘要

本申请提供一种音频信息检索方法及装置。方法包括:获得待检索音频片段的声谱图;提取声谱图中至少一个能量极值点的局部特征,以构成待检索音频片段的局部特征序列;根据待检索音频片段的局部特征序列进行检索,以获得检索结果。本申请可以降低漏匹配的概率,提高检索结果的准确度。



CN 106294331 B

1.一种音频信息检索方法,其特征在于,包括:

获得待检索音频片段的声谱图;

对所述声谱图进行极值点检测,以获得至少一个能量极值点;

确定所述至少一个能量极值点中每个能量极值点在所述声谱图上所属的图像块;

提取所述每个能量极值点所属的图像块的特征,以构成所述待检索音频片段的局部特征序列;

根据所述待检索音频片段的局部特征序列进行检索,以获得检索结果。

2.根据权利要求1所述的方法,其特征在于,所述确定所述至少一个能量极值点中每个能量极值点在所述声谱图上所属的图像块,包括:

在所述声谱图上取以所述每个能量极值点为中心的窗口区域,作为所述每个能量极值点所属的图像块。

3.根据权利要求1所述的方法,其特征在于,所述提取所述每个能量极值点所属的图像块的特征,以构成所述待检索音频片段的局部特征序列,包括:

对所述每个能量极值点所属的图像块,按照指定的编码次数,对所述图像块中像素点代表的能量值之间的大小关系进行随机编码,以获得所述图像块的特征,将所述图像块的特征作为所述待检索音频片段的局部特征序列中的一个局部特征。

4.根据权利要求3所述的方法,其特征在于,所述按照指定的编码次数,对所述图像块中像素点代表的能量值之间的大小关系进行随机编码,以获得所述图像块的特征,包括:

每次随机从所述图像块中获取两个像素点,对所述两个像素点代表的能量值之间的大小关系进行编码,以获得一个编码结果,直到编码次数达到所述指定的编码次数时,根据所有编码结果获得所述图像块的特征。

5.根据权利要求1-4任一项所述的方法,其特征在于,所述根据所述待检索音频片段的局部特征序列进行检索,以获得检索结果,包括:

将所述待检索音频片段的局部特征序列与音频特征库中每个音频文件的局部特征序列进行匹配,以获得所述待检索音频片段与相似音频文件之间的匹配点对,所述相似音频文件是指所述音频特征库中与所述待检索音频片段相似的音频文件;

根据所述待检索音频片段与所述相似音频文件之间的匹配点对,获取所述待检索音频片段与所述相似音频文件的匹配度;

获取最大匹配度对应的相似音频文件的信息作为所述检索结果。

6.根据权利要求5所述的方法,其特征在于,所述将所述待检索音频片段的局部特征序列与音频特征库中每个音频文件的局部特征序列进行匹配,以获得所述待检索音频片段与相似音频文件之间的匹配点对,包括:

根据所述待检索音频片段中每个能量极值点的频率坐标和所述音频文件中每个能量极值点的频率坐标,确定所述待检索音频片段中每个能量极值点对应于所述音频文件中的极值点子集;

根据所述待检索音频片段中每个能量极值点的局部特征和对应的极值点子集中各能量极值点的局部特征,获取所述待检索音频片段中每个能量极值点与所述对应的极值点子集的距离,所述待检索音频片段中每个能量极值点与所述对应的极值点子集的距离是指所述待检索音频片段中每个能量极值点与所述对应的极值点子集中各能量极值点的距离中

的最小距离；

若所述待检索音频片段中的能量极值点与所述对应的极值点子集的距离中存在小于预设特征阈值的距离，则将所述音频文件作为所述待检索音频片段的相似音频文件，并将所述小于预设特征阈值的距离对应的所述待检索音频片段中的能量极值点和所述音频文件中的能量极值点作为匹配点对。

7.根据权利要求6所述的方法，其特征在于，所述根据所述待检索音频片段与所述相似音频文件之间的匹配点对，获取所述待检索音频片段与所述相似音频文件的匹配度，包括：

采用随机抽样一致算法或霍夫变换算法，对所述待检索音频片段与所述相似音频文件之间的匹配点对进行处理，以获取所述待检索音频片段与所述相似音频文件的匹配度。

8.根据权利要求5所述的方法，其特征在于，还包括：

根据所述待检索音频片段与所述相似音频文件之间的匹配点对，获取所述待检索音频片段在所述相似音频文件中的时间偏移量；

获取最大匹配度对应的时间偏移量作为所述检索结果。

9.根据权利要求6所述的方法，其特征在于，还包括：构建所述音频特征库；

所述构建所述音频特征库，包括：

获得所述音频文件的声谱图；

提取所述音频文件的声谱图中至少一个能量极值点的局部特征，以构成所述音频文件的局部特征序列；

将所述音频文件的局部特征序列存储到所述音频特征库中。

10.一种音频特征库构建方法，其特征在于，包括：

获得音频文件的声谱图；

对所述声谱图进行极值点检测，以获得至少一个能量极值点；

确定所述至少一个能量极值点中每个能量极值点在所述声谱图上所属的图像块；

提取所述每个能量极值点所属的图像块的特征，以构成所述音频文件的局部特征序列；

将所述音频文件的局部特征序列存储到音频特征库中。

11.一种音频信息检索装置，其特征在于，包括：

获取模块，用于获得待检索音频片段的声谱图；

提取模块，用于对所述声谱图进行极值点检测，以获得至少一个能量极值点，确定所述至少一个能量极值点中每个能量极值点在所述声谱图上所属的图像块，提取所述每个能量极值点所属的图像块的特征，以构成所述待检索音频片段的局部特征序列；

检索模块，用于根据所述待检索音频片段的局部特征序列进行检索，以获得检索结果。

12.根据权利要求11所述的装置，其特征在于，所述提取模块具体用于：

在所述声谱图上取以所述每个能量极值点为中心的窗口区域，作为所述每个能量极值点所属的图像块。

13.根据权利要求11所述的装置，其特征在于，所述提取模块具体用于：

对所述每个能量极值点所属的图像块，按照指定的编码次数，对所述图像块中像素点代表的能量值之间的大小关系进行随机编码，以获得所述图像块的特征，将所述图像块的特征作为所述待检索音频片段的局部特征序列中的一个局部特征。

14.根据权利要求13所述的装置,其特征在于,所述提取模块具体用于:

每次随机从所述图像块中获取两个像素点,对所述两个像素点代表的能量值之间的大小关系进行编码,以获得一个编码结果,直到编码次数达到所述指定的编码次数时,根据所有编码结果获得所述图像块的特征。

15.根据权利要求11-14任一项所述的装置,其特征在于,所述检索模块具体用于:

将所述待检索音频片段的局部特征序列与音频特征库中每个音频文件的局部特征序列进行匹配,以获得所述待检索音频片段与相似音频文件之间的匹配点对,所述相似音频文件是指所述音频特征库中与所述待检索音频片段相似的音频文件;

根据所述待检索音频片段与所述相似音频文件之间的匹配点对,获取所述待检索音频片段与所述相似音频文件的匹配度;

获取最大匹配度对应的相似音频文件的信息作为所述检索结果。

16.根据权利要求15所述的装置,其特征在于,所述检索模块具体用于:

根据所述待检索音频片段中每个能量极值点的频率坐标和所述音频文件中每个能量极值点的频率坐标,确定所述待检索音频片段中每个能量极值点对应于所述音频文件中的极值点子集;

根据所述待检索音频片段中每个能量极值点的局部特征和对应的极值点子集中各能量极值点的局部特征,获取所述待检索音频片段中每个能量极值点与所述对应的极值点子集的距离,所述待检索音频片段中每个能量极值点与所述对应的极值点子集的距离是指所述待检索音频片段中每个能量极值点与所述对应的极值点子集中各能量极值点的距离中的最小距离;

若所述待检索音频片段中的能量极值点与所述对应的极值点子集的距离中存在小于预设特征阈值的距离,则将所述音频文件作为所述待检索音频片段的相似音频文件,并将所述小于预设特征阈值的距离对应的所述待检索音频片段中的能量极值点和所述音频文件中的能量极值点作为匹配点对。

17.根据权利要求16所述的装置,其特征在于,所述检索模块具体用于:

采用随机抽样一致算法或霍夫变换算法,对所述待检索音频片段与所述相似音频文件之间的匹配点对进行处理,以获取所述待检索音频片段与所述相似音频文件的匹配度。

18.根据权利要求15所述的装置,其特征在于,所述检索模块还用于:

根据所述待检索音频片段与所述相似音频文件之间的匹配点对,获取所述待检索音频片段在所述相似音频文件中的时间偏移量;

获取最大匹配度对应的时间偏移量作为所述检索结果。

19.根据权利要求16所述的装置,其特征在于,还包括:

构建模块,用于构建所述音频特征库;

所述构建模块具体用于:

获得所述音频文件的声谱图;

提取所述音频文件的声谱图中至少一个能量极值点的局部特征,以构成所述音频文件的局部特征序列;

将所述音频文件的局部特征序列存储到所述音频特征库中。

20.一种音频特征库构建装置,其特征在于,包括:

获得模块，用于获得音频文件的声谱图；

提取模块，用于对所述声谱图进行极值点检测，以获得至少一个能量极值点，确定所述至少一个能量极值点中每个能量极值点在所述声谱图上所属的图像块，提取所述每个能量极值点所属的图像块的特征，以构成所述音频文件的局部特征序列；

存储模块，用于将所述音频文件的局部特征序列存储到音频特征库中。

音频信息检索方法及装置

【技术领域】

[0001] 本申请涉及音频处理技术领域，尤其涉及一种音频信息检索方法及装置。

【背景技术】

[0002] 音乐检索研究始于上世纪90年代，2000年后开始逐步成熟且走进实际应用。已有的音乐检索方法大多基于声谱图进行分析，可分为两类：基于极值点的方法和基于纹理分析的方法。

[0003] 一种基于纹理分析的音乐检索方法，首先对音乐片段采用短时傅立叶变换以生成声谱图，然后将该声谱图分解为32个子带，并计算相邻子带的梯度极性，从而将原始信号压缩为紧致的二进制编码，在检索时采用哈希表进行加速。基于纹理分析的音乐检索方法对块状噪声不鲁棒，且运算复杂度较高，检索时间较长。于是提出一种对块状噪声更鲁棒，且检索速度较快的方法，即基于极值点的方法。

[0004] 基于极值点的方法，首先对音乐片段采用短时傅立叶变换以生成声谱图，然后检测声谱图中的极大值点，然后直接根据相邻极值点对的频率和时间差生成哈希表。在检索时，首先使用哈希表匹配待检索音乐片段和音乐库之间对应的匹配点，然后根据匹配点的时间坐标估计每首音乐的偏移量和置信度，置信度最大且超过阈值的音乐即为检索结果。但是，该方法中极值点的检测对随机噪声和椒盐噪声比较敏感，容易在频率和时间方向上产生偏移，极值点的轻微偏移会完全改变哈希值，这会导致漏匹配，影响检索结果的准确度。

【发明内容】

[0005] 本申请的多个方面提供一种音频信息检索方法及装置，用以降低漏匹配的概率，提高检索结果的准确度。

[0006] 本申请的一方面，提供一种音频信息检索方法，包括：

[0007] 获得待检索音频片段的声谱图；

[0008] 提取所述声谱图中至少一个能量极值点的局部特征，以构成所述待检索音频片段的局部特征序列；

[0009] 根据所述待检索音频片段的局部特征序列进行检索，以获得检索结果。

[0010] 本申请的另一方面，提供一种音频信息检索装置，包括：

[0011] 获取模块，用于获得待检索音频片段的声谱图；

[0012] 提取模块，用于提取所述声谱图中至少一个能量极值点的局部特征，以构成所述待检索音频片段的局部特征序列；

[0013] 检索模块，用于根据所述待检索音频片段的局部特征序列进行检索，以获得检索结果。

[0014] 本申请的又一方面，提供一种音频特征库构建方法，包括：

[0015] 获得音频文件的声谱图；

[0016] 提取所述音频文件的声谱图中至少一个能量极值点的局部特征，以构成所述音频文件的局部特征序列；

[0017] 将所述音频文件的局部特征序列存储到音频特征库中。

[0018] 本申请的又一方面，提供一种音频特征库构建装置，包括：

[0019] 获得模块，用于获得音频文件的声谱图；

[0020] 提取模块，用于提取所述音频文件的声谱图中至少一个能量极值点的局部特征，以构成所述音频文件的局部特征序列；

[0021] 存储模块，用于将所述音频文件的局部特征序列存储到音频特征库中。

[0022] 在本申请中，获得待检索音频片段的声谱图，提取声谱图中至少一个能量极值点的局部特征，构成该待检索音频片段的局部特征序列，根据待检索音频片段的局部特征序列进行检索，获得检索结果。本申请在检索过程中使用的是声谱图中能量极值点的局部特征，而不是能量极值点，相当于放宽了在时间坐标和频率坐标上的匹配范围，能够增加匹配中的点数，意味着能量极值点的局部特征要比能量极值点对随机噪声和椒盐噪声的敏感性低，即使发生轻微偏移也不会对匹配结果产生太大影响，解决了现有技术中因极值点偏移导致漏匹配的问题，有利于提高检索结果的准确度。

【附图说明】

[0023] 为了更清楚地说明本申请实施例中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[0024] 图1为本申请一实施例提供的音频信息检索方法的流程示意图；

[0025] 图2为本申请一实施例提供的音频信号的时域波形图；

[0026] 图3为图2所示音频信号的声谱图；

[0027] 图4为本申请一实施例提供的被噪声污染的音频信号的能量极值点分布图；

[0028] 图5为本申请一实施例提供的未被噪声污染的音频信号的能量极值点分布图；

[0029] 图6为本申请一实施例提供的相同能量极值点及不同能量极值点根据其局部特征计算出的汉明距离的分布示意图；

[0030] 图7为本申请一实施例提供的音频信息检索装置的结构示意图；

[0031] 图8为本申请另一实施例提供的音频信息检索装置的结构示意图；

[0032] 图9为本申请一实施例提供的音频特征库构建装置的结构示意图。

【具体实施方式】

[0033] 为使本申请实施例的目的、技术方案和优点更加清楚，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作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。

[0034] 图1为本申请一实施例提供的音频信息检索方法的流程示意图。如图1所示，该方法包括：

[0035] 101、获得待检索音频片段的声谱图。

[0036] 102、提取声谱图中至少一个能量极值点的局部特征，以构成待检索音频片段的局部特征序列。

[0037] 103、根据待检索音频片段的局部特征序列进行检索，以获得检索结果。

[0038] 本实施例提供一种音频信息检索方法，主要用于对待检索音频片段进行检索，获取与待检索音频片段相关的信息。该方法的主要原理是利用待检索音频片段的能量极值点的局部特征代替能量极值点，根据能量极值点的局部特征进行检索，利用能量极值点的局部特征比极值点对随机噪声和椒盐噪声的敏感性低的优势，解决了现有技术中因极值点偏移导致漏匹配的问题，提高了检索结果的准确度。

[0039] 下面对本实施例进行详细介绍：

[0040] 声音为周期性机械波，人耳能感知到的频率范围为20到20000赫兹(Hz)，频率越小音调越低，频率越大则音调越高。通过在时间轴上对各种频率进行组合，即可形成不同的音频信号。为区分不同的音频信号，研究者一般对信号在某段时间内进行频谱分解，然后分析每一时段的频谱特性。

[0041] 音频信号在计算机中使用一组采样和量化过的一维信号表示，采样率通常为但不限于：11025Hz，量化级别通常为但不限于：16。如图2所示，为一段长度为12秒的音频信号的时域波形，该音频信号的采样率为11025Hz，量化等级为16，图2的横轴为时间，纵轴为信号强度。对图2所示音频信号进行短时傅立叶变换(STFT)，可得到其声谱图，如图3所示。图3的横轴为时间，纵轴为频率，图3右侧的灰度条为能量与灰度值的对应关系，不同灰度值代表不同能量。将图2与图3进行比对可知，与时域波形图相比，声谱图更直观，更能反映出音频信号在各时段、各频率的能量分布。

[0042] 在了解音频信号的基础上，当用户需要进行音频信息检索时，可以获取待检索音频片段。例如，用户可以录制一段音频，如一句话、几句音乐或一段旋律等作为待检索音频片段。或者，用户可以从互联网上下载一段音频，如一首完整音乐、一段音乐片段、一首铃声或一段演讲内容等作为待检索音频片段。或者，用户还可以获取本地的音频，如一首铃声、一段演讲内容或一段音乐等作为待检索音频片段。之后，对待检索音频片段进行时频变换，如短时傅里叶变换，从而获得待检索音频片段的声谱图。

[0043] 虽然音频信号的特性大部分包含在其声谱图中，但声谱图包含的数据量较大，且易受噪声影响，因此不宜直接将声谱图用于音频检索。为了提升检索过程对噪声的鲁棒性和减小计算量，本实施例主要关注声谱图中的能量极值点。图4和图5显示了同一段音频信号的能量极值点的分布，其中，图4所示为被噪声污染的音频信号的能量极值点分布，图5所示为未被噪声污染的音频信号的能量极值点分布。从图4和图5中可以看出，同一段音频信号的能量极值点大致分布在相同的位置，基本能反映音频信号的特性，虽然局部有偏移，但整体上相对稳定，由此可见，通过声谱图中的能量极值点能够区分出相同音频信号和不同音频信号，声谱图上的能量极值点可代替声谱图用于音频检索过程。

[0044] 进一步对图4和图5进行分析可得出：因受噪声影响，图4和图5所示能量极值点中在时间坐标和频率坐标上能严格匹配上的点数低于总点数的10％，而如果采用能量极值点周围的特性，即放宽在时间坐标和频率坐标上的限制，大致能匹配上的点数可达总点数的50％。由于匹配上的点数越多，匹配结果就会越精确，基于此，为匹配上更多的点，本实施例

采用能量极值点周围的特性来代表能量极值点，使得检索过程基于能量极值点周围的特性进行。能量极值点周围的特性称为能量极值点的局部特征，该局部特征可以是各种类型的特征，例如纹理特征、边缘特征等。

[0045] 于是，在获得待检索音频片段的声谱图后，提取声谱图中至少一个能量极值点的局部特征，以构成该检索音频片段的局部特征序列，以便用于后续检索步骤。

[0046] 考虑到能量极值点的数量较多，本实施例优先采用相对简单的特征来表达能量极值点的局部特性，从而达到降低数据量和比对损耗的目的。例如，可以采用局部二值模式(Local Binary Pattern,LBP)特征、梯度方向直方图(Histogram of Oriented Gradient,HoG)特征、Haar特征等来表达能量极值点的局部特征。针对LBP特征、HoG特征及Haar特征，可以采用相应的算法来获取。由于获取LBP特征、HoG特征及Haar特征的过程属于现有技术，在此不再赘述。

[0047] 本实施例提供一种提取声谱图中至少一个能量极值点的局部特征的方法，包括：

[0048] 首先，对声谱图进行极值点检测，以获得至少一个能量极值点；例如，可以采用滤波器在待检索音频片段的声谱图上进行极大值滤波，以获得声谱图上能量极大值的位置；滤波器的大小和形状可以根据具体应用进行调整；

[0049] 接着，确定至少一个能量极值点中每个能量极值点在声谱图上所属的图像块；例如，可以在声谱图上取以每个能量极值点为中心的窗口区域，作为每个能量极值点所属的图像块，窗口区域的大小不做限定，可根据具体应用适应性设置；

[0050] 提取每个能量极值点所属的图像块的特征，以构成待检索音频片段的局部特征序列。也就是说，将能量极值点所属图像块的特征作为能量极值点的局部特征。

[0051] 进一步，提取每个能量极值点所属的图像块的特征的过程具体为：对每个能量块所属的图像块，按照指定的编码次数，对图像块中像素点代表的能量值之间的大小关系进行随机编码，以获得该图像块的特征，将该图像块的特征作为待检索音频片段的局部特征序列中的一个局部特征。

[0052] 具体的，预先指定编码次数，每次随机从图像块中获取两个像素点，对随机获取的两个像素点代表的能量之间的大小关系进行编码，以获得一个编码结果，当编码次数达到指定的编码次数时，根据所有编码结果获得图像块的特征。

[0053] 其中，上述编码可以是二进制编码。例如，若第一个获取的像素点代表的能量大于第二个获取的像素点代表的能量时，可以编码为1；若第一个获取的像素点代表的能量小于或等于第二个获取的像素点代表的能量时，可以编码为0。或者，若第一个获取的像素点代表的能量大于第二个获取的像素点代表的能量时，可以编码为0；若第一个获取的像素点代表的能量小于或等于第二个获取的像素点代表的能量时，可以编码为1。

[0054] 基于上述二进制编码，根据所有编码结果获得图像块的特征的过程具体为：按照编码先后顺序依次将所有编码结果结合起来作为一个二进制序列，该二进制序列即为图像块的特征。

[0055] 当对随机选取的两个像素代表的能量值之间的大小进行二进制编码时，上述指定编码次数与上述二进制序列的长度相同，例如，可以是33位、256位等，一次编码处理的结果作为二进制序列的一位。

[0056] 在获得待检索音频片段的局部特征序列后，根据待检索音频片段的局部特征序列

进行检索，以获得检索结果。这里的检索结果可以是各种与待检索音频片段相关的信息，检索结果包括但不限于：待检索音频片段所属的音频文件的信息。例如，检索结果还可以包括：待检索音频片段在所属音频文件中的时间偏移量。

[0057] 上述检索过程的实质是利用能量极值点的局部特征在预先建立的音乐特征库中进行匹配，根据匹配度输出与该待检索音频片段有关的信息。其中，音频特征库中存储有大量音频文件的局部特征序列。关于音频特征库的建立过程在后续实施方式中进行详述。

[0058] 上述检索过程具体包括：

[0059] 将待检索音频片段的局部特征序列与音频特征库中每个音频文件的局部特征序列进行匹配，以获得待检索音频片段与音频特征库中与该待检索音频片段相似的音频文件之间的匹配点对；为便于描述，将音频特征库中与待检索音频片段相似的音频文件称为相似音频文件；相似音频文件为一个或多个；

[0060] 根据上述待检索音频片段与相似音频文件之间的匹配点对，获取待检索音频片段与相似音频文件的匹配度；

[0061] 获取最大匹配度对应的相似音频文件的信息作为检索结果。

[0062] 可选的，除了获取待检索音频片段与相似音频文件的匹配度之外，还可以根据上述待检索音频片段与相似音频文件之间的匹配点对，获取待检索音频片段在相似音频文件中的时间偏移量；进一步，还可以获取最大匹配度对应的时间偏移量作为检索结果。

[0063] 在一可选实施方式中，待检索音频片段的局部特征序列包括待检索音频片段中每个能量极值点的局部特征；相应的，每个音频文件的局部特征序列包括该音频文件中每个能量极值点的局部特征。进一步，待检索音频片段的局部特征序列还可以包括待检索音频片段中每个能量极值点的时间坐标和频率坐标；相应的，每个音频文件的局部特征序列还可以包括该音频文件中每个能量极值点的时间坐标和频率坐标。例如，可以将待检索音频片段中每个能量极值点的信息记为$\left(f_x^k \quad t_x^k \quad b_x^k\right)$，并将音频文件中每个能量极值点的信息记为$\left(f_y^l \quad t_y^l \quad b_y^l\right)$。其中，f表示频率坐标，t表示时间坐标，b为局部特征，k为待检索音频片段中极值点的序号，l为音频文件中能量极值点的序号。

[0064] 可选的，可以根据能量极值点的局部特征之间的距离，来判断待检索音频片段是否与音频文件相似以及在相似时存在的匹配点对。所述距离可以采用但不限于：汉明距离。

[0065] 本申请发明人经过大量实验发现：局部特征之间的汉明距离可以表征能量极值点是否匹配。实验过程具体为：对于大量来自同一音频源的两个音频片段，两个音频片段的区别在于：一个是未被噪声污染的信号，一个是被噪声污染的信号，分别计算两个音频片段中所有能量极值点的局部特征之间的汉明距离，其汉明距离的分布如图6所示，其中实线表示相同能量极值点之间的汉明距离的分布，虚线表示不同能量极值点之间的汉明距离的分布。从图6中可以看出，相同能量极值点之间的汉明距离明显小于不同能量极值点之间的汉明距离，因此汉明距离可用来判断两个能量极值点是否匹配。

[0066] 基于上述，一种获得待检索音频片段与相似音频文件之间的匹配点对的实施方式包括：

[0067] 根据待检索音频片段中每个能量极值点的频率坐标和音频文件中每个能量极值点的频率坐标，确定待检索音频片段中每个能量极值点对应于该音频文件中的极值点子

集;例如,可以从$\left(f_y^l \quad t_y^l \quad b_y^l\right)$中选择频率坐标在频率范围$\left[f_x^k - T_f \quad f_x^k + T_f\right]$内的能量极值点构成极值点子集;$T_f$为频率误差阈值;

[0068] 根据待检索音频片段中每个能量极值点的局部特征和所对应的极值点子集中各能量极值点的局部特征,获取待检索音频片段中每个能量极值点与所对应极值点子集的距离,待检索音频片段中每个能量极值点与所对应的极值点子集的距离是指待检索音频片段中每个能量极值点与对应极值点子集中各能量极值点的距离中的最小距离;

[0069] 若待检索音频片段中的能量极值点与所对应的极值点子集的距离中存在小于预设特征阈值的距离,则将该音频文件作为待检索音频片段的相似音频文件,并将小于预设特征阈值的距离对应的待检索音频片段中的能量极值点和该音频文件中的能量极值点作为匹配点对。

[0070] 值得说明的是,若待检索音频片段中所有能量极值点与所对应的极值点子集的距离中不存在小于预设特征阈值的距离,说明待检索音频片段与该音频文件不相似,可以将该音频文件忽略,不用进行后续处理,以便节约检索资源。

[0071] 进一步,在获得匹配点对后,根据匹配点对,获取待检索音频片段与相似音频文件之间的相似度及时间偏移量的方式可以有多种。例如,可以用匹配点对的个数来衡量两者之间的相似度,用匹配点对的时间坐标的均值的差异作为时间偏移量。或者,可以对匹配点对的个数、以及匹配点对在时间坐标上的差异以及在频率坐标上的差异进行加权处理,获得相似度及时间偏移量等。

[0072] 考虑到匹配点对中有可能包括误匹配的点,若直接根据匹配点对计算匹配度和时间偏移量,有可能导致计算结果不鲁棒,因此,本实施例提供一种具有鲁棒效果的方法,即采用随机抽样一致算法(RANSAC)或霍夫变换(Hough Transform)算法,对待检索音频片段与相似音频文件之间的匹配点对进行处理,以获取待检索音频片段与所述相似音频文件的匹配度。值得说明的是,若需要还可以获取待检索音频片段在相似音频文件中的时间偏移量。

[0073] 其中,随机抽样一致算法是一种基于随机采样的鲁棒的模型参数估计方法。在本实施例中的应用原理是:每次随机从待检索音频片段与所述相似音频文件之间的匹配点对中选取部分匹配点对,根据随机选取的部分匹配点对的时间坐标进行模型拟合,获得待检索音频片段在相似音频文件中的候选时间偏移量,并将此次选取的匹配点对中的非噪声点对的个数作为待检索音频片段与相似音频文件的候选匹配度;经过多次模型拟合,获得多个候选时间偏移量和候选匹配度;从中选择最大候选匹配度作为待检索音频片段与相似音频文件的匹配度,将最大候选匹配度对应的候选时间偏移量作为待检索音频片段在相似音频文件中的时间偏移量。

[0074] 上述模型拟合的公式具体为:$t_y = t_x + o$,其中,o表示时间偏移量。

[0075] 由于经过多次模型拟合,每次都随机选择部分匹配点对,终会有一次选择使用的匹配点对都是非噪声点,进而给出合理的结果,可以降低噪声影响,对噪声具有鲁棒性。另外,该算法内存消耗小,尤其适用于对内存消耗有要求的场景。

[0076] 由上述可见,本申请技术方案对噪声具有更好的鲁棒性,且内存消耗更小,能检索时长更短的音频片段。

[0077] 在进行检索之前，本实施例提供的方法还包括：构建音频特征库的步骤。一种构成音频特征库的方式包括：

[0078] 获得音频文件的声谱图；例如，可以对音频文件进行时频变换，如短时傅立叶变换，以获得其声谱图；

[0079] 提取音频文件的声谱图中至少一个能量极值点的局部特征，以构成音频文件的局部特征序列；

[0080] 将音频文件的局部特征序列存储到音频特征库中。

[0081] 可选的，上述提取音频文件的声谱图中至少一个能量极值点的局部特征，以构成音频文件的局部特征序列，包括：首先对音频文件的声谱图进行极值点检测，以获得至少一个能量极值点；例如，可以采用滤波器在音频文件的声谱图上进行极大值滤波，以获得声谱图上能量极大值的位置；滤波器的大小和形状可以根据具体应用进行调整；接着，确定至少一个能量极值点中每个能量极值点在声谱图上所属的图像块；例如，可以在声谱图上取以每个能量极值点为中心的窗口区域，作为每个能量极值点所属的图像块，窗口区域的大小不做限定，可根据具体应用适应性设置；之后，提取每个能量极值点所属的图像块的特征，以构成音频文件的局部特征序列。也就是说，将能量极值点所属图像块的特征作为能量极值点的局部特征。

[0082] 进一步，提取每个能量极值点所属的图像块的特征的过程具体为：对每个能量块所属的图像块，按照指定的编码次数，对图像块中像素点代表的能量值之间的大小关系进行随机编码，以获得该图像块的特征，将该图像块的特征作为待检索音频片段的局部特征序列中的一个局部特征。

[0083] 具体的，预先指定编码次数，每次随机从图像块中获取两个像素点，对随机获取的两个像素点代表的能量之间的大小关系进行编码，以获得一个编码结果，当编码次数达到指定的编码次数时，根据所有编码结果获得图像块的特征。

[0084] 值得说明的是，上述音频文件的数量越多，该音频特征库存储到信息就越丰富。另外，随着时间的推移，可以随机对音频特征库进行更新。

[0085] 需要说明的是，对于前述的各方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请并不受所描述的动作顺序的限制，因为依据本申请，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本申请所必须的。

[0086] 在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中没有详述的部分，可以参见其他实施例的相关描述。

[0087] 图7为本申请一实施例提供的音频信息检索装置的结构示意图。如图7所示，该装置包括：获取模块71、提取模块72和检索模块73。

[0088] 获取模块71，用于获得待检索音频片段的声谱图。

[0089] 提取模块72，用于提取获取模块71获取的声谱图中至少一个能量极值点的局部特征，以构成待检索音频片段的局部特征序列。

[0090] 检索模块73，用于根据提取模块72获取的待检索音频片段的局部特征序列进行检索，以获得检索结果。

[0091] 在一可选实施方式中，提取模块72具体用于：

[0092] 对声谱图进行极值点检测，以获得至少一个能量极值点；

[0093] 确定至少一个能量极值点中每个能量极值点在声谱图上所属的图像块；

[0094] 提取每个能量极值点所属的图像块的特征，以构成待检索音频片段的局部特征序列。

[0095] 进一步，提取模块72在确定至少一个能量极值点中每个能量极值点在声谱图上所属的图像块时，具体用于：在声谱图上取以每个能量极值点为中心的窗口区域，作为每个能量极值点所属的图像块。

[0096] 进一步，提取模块72在提取每个能量极值点所属的图像块的特征，以构成待检索音频片段的局部特征序列时，具体用于：对每个能量极值点所属的图像块，按照指定的编码次数，对图像块中像素点代表的能量值之间的大小关系进行随机编码，以获得图像块的特征，将图像块的特征作为待检索音频片段的局部特征序列中的一个局部特征。

[0097] 更进一步，提取模块72具体用于：每次随机从图像块中获取两个像素点，对两个像素点代表的能量值之间的大小关系进行编码，以获得一个编码结果，直到编码次数达到指定的编码次数时，根据所有编码结果获得图像块的特征。

[0098] 在一可选实施方式中，检索模块73具体用于：

[0099] 将待检索音频片段的局部特征序列与音频特征库中每个音频文件的局部特征序列进行匹配，以获得待检索音频片段与相似音频文件之间的匹配点对，相似音频文件是指音频特征库中与待检索音频片段相似的音频文件；

[0100] 根据待检索音频片段与相似音频文件之间的匹配点对，获取待检索音频片段与相似音频文件的匹配度；

[0101] 获取最大匹配度对应的相似音频文件的信息作为检索结果。

[0102] 进一步，检索模块73在获得待检索音频片段与相似音频文件之间的匹配点对时，具体用于：

[0103] 根据待检索音频片段中每个能量极值点的频率坐标和音频文件中每个能量极值点的频率坐标，确定待检索音频片段中每个能量极值点对应于音频文件中的极值点子集；

[0104] 根据待检索音频片段中每个能量极值点的局部特征和对应的极值点子集中各能量极值点的局部特征，获取待检索音频片段中每个能量极值点与对应的极值点子集的距离，待检索音频片段中每个能量极值点与对应的极值点子集的距离是指待检索音频片段中每个能量极值点与对应的极值点子集中各能量极值点的距离中的最小距离；

[0105] 若待检索音频片段中的能量极值点与对应的极值点子集的距离中存在小于预设特征阈值的距离，则将音频文件作为待检索音频片段的相似音频文件，并将小于预设特征阈值的距离对应的待检索音频片段中的能量极值点和音频文件中的能量极值点作为匹配点对。

[0106] 进一步，检索模块73在获取待检索音频片段与相似音频文件的匹配度时，具体用于：

[0107] 采用随机抽样一致算法或霍夫变换算法，对待检索音频片段与相似音频文件之间的匹配点对进行处理，以获取待检索音频片段与相似音频文件的匹配度。

[0108] 进一步，检索模块73还用于：根据待检索音频片段与相似音频文件之间的匹配点

对，获取待检索音频片段在相似音频文件中的时间偏移量；获取最大匹配度对应的时间偏移量作为检索结果。

[0109] 在一可选实施方式中，如图8所示，该音频信息检索装置还包括：构建模块74。

[0110] 构建模块74，用于构建音频特征库。

[0111] 构建模块74具体用于采用以下方式构建音频特征库：

[0112] 获得音频文件的声谱图；

[0113] 提取音频文件的声谱图中至少一个能量极值点的局部特征，以构成音频文件的局部特征序列；

[0114] 将音频文件的局部特征序列存储到音频特征库中。

[0115] 本实施例提供的音频信息检索装置，获得待检索音频片段的声谱图，提取声谱图中至少一个能量极值点的局部特征，构成该待检索音频片段的局部特征序列，根据待检索音频片段的局部特征序列进行检索，获得检索结果。本实施例提供的音频信息检索装置在检索过程中使用的是声谱图中能量极值点的局部特征，而不是能量极值点，相当于放宽了在时间坐标和频率坐标上的匹配范围，能够增加匹配中的点数，意味着能量极值点的局部特征要比能量极值点对随机噪声和椒盐噪声的敏感性低，即使发生轻微偏移也不会对匹配结果产生太大影响，解决了现有技术中因极值点偏移导致漏匹配的问题，有利于提高检索结果的准确度。

[0116] 图9为本申请一实施例提供的音频特征库构建装置的结构示意图。如图9所示，该装置包括：获得模块91、提取模块92和存储模块93。

[0117] 获得模块91，用于获得音频文件的声谱图。

[0118] 提取模块92，用于提取获得模块91获得的音频文件的声谱图中至少一个能量极值点的局部特征，以构成音频文件的局部特征序列。

[0119] 存储模块93，用于将提取模块92所提取的音频文件的局部特征序列存储到音频特征库中。

[0120] 提取模块92具体用于：对音频文件的声谱图进行极值点检测，以获得至少一个能量极值点；例如，可以采用滤波器在音频文件的声谱图上进行极大值滤波，以获得声谱图上能量极大值的位置；滤波器的大小和形状可以根据具体应用进行调整；接着，确定至少一个能量极值点中每个能量极值点在声谱图上所属的图像块；例如，可以在声谱图上取以每个能量极值点为中心的窗口区域，作为每个能量极值点所属的图像块，窗口区域的大小不做限定，可根据具体应用适应性设置；之后，提取每个能量极值点所属的图像块的特征，以构成音频文件的局部特征序列。也就是说，将能量极值点所属图像块的特征作为能量极值点的局部特征。

[0121] 进一步，提取模块92在用于提取每个能量极值点所属的图像块的特征时，具体用于：对每个能量块所属的图像块，按照指定的编码次数，对图像块中像素点代表的能量值之间的大小关系进行随机编码，以获得该图像块的特征，将该图像块的特征作为待检索音频片段的局部特征序列中的一个局部特征。

[0122] 具体的，预先指定编码次数，每次随机从图像块中获取两个像素点，对随机获取的两个像素点代表的能量之间的大小关系进行编码，以获得一个编码结果，当编码次数达到指定的编码次数时，根据所有编码结果获得图像块的特征。

[0123] 值得说明的是，上述音频文件的数量越多，该音频特征库存储到信息就越丰富。另外，随着时间的推移，可以随机对音频特征库进行更新。

[0124] 本实施例提供的音频特征库构建装置，获得待检索音频文件的声谱图，提取声谱图中至少一个能量极值点的局部特征，构成该待检索音频文件的局部特征序列，将待检索音频文件的局部特征序列存储到音频特征库中，以构建音频特征库。本实施例提供的音频特征库中存储的是声谱图中能量极值点的局部特征，而不是能量极值点，相当于在匹配过程中放宽了在时间坐标和频率坐标上的匹配范围，能够增加匹配中的点数，意味着能量极值点的局部特征要比能量极值点对随机噪声和椒盐噪声的敏感性低，即使发生轻微偏移也不会对匹配结果产生太大影响，解决了现有技术中因极值点偏移导致漏匹配的问题，有利于提高检索结果的准确度。所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统，装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

[0125] 在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

[0126] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0127] 另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用硬件加软件功能单元的形式实现。

[0128] 上述以软件功能单元的形式实现的集成的单元，可以存储在一个计算机可读取存储介质中。上述软件功能单元存储在一个存储介质中，包括若干指令用以使得一台计算机设备(可以是个人计算机，服务器，或者网络设备等)或处理器(processor)执行本申请各个实施例所述方法的部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器(Read-Only Memory，ROM)、随机存取存储器(Random Access Memory，RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0129] 最后应说明的是：以上实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围。

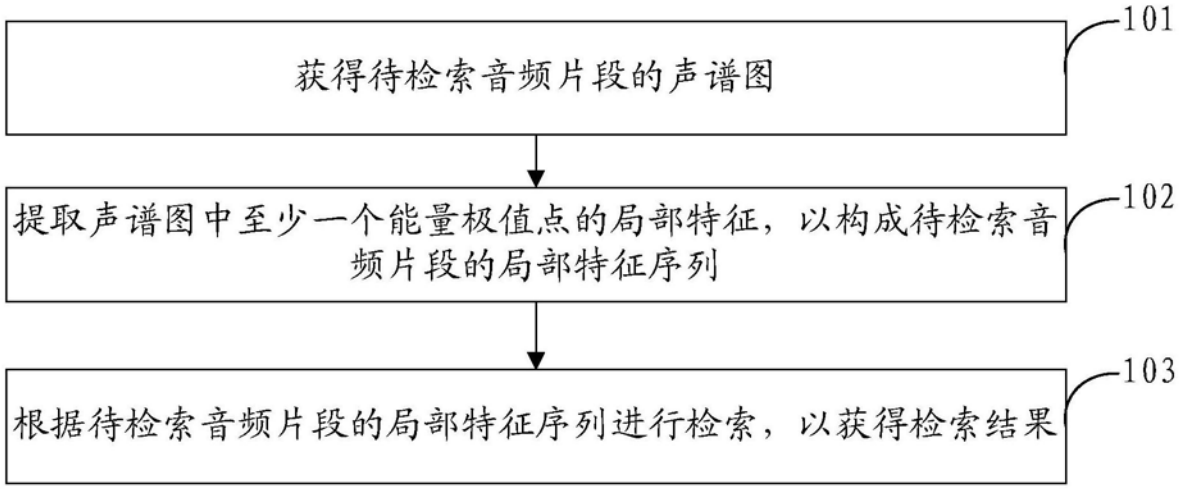

获得待检索音频片段的声谱图
101
提取声谱图中至少一个能量极值点的局部特征，以构成待检索音频片段的局部特征序列
102
根据待检索音频片段的局部特征序列进行检索，以获得检索结果
103

图1

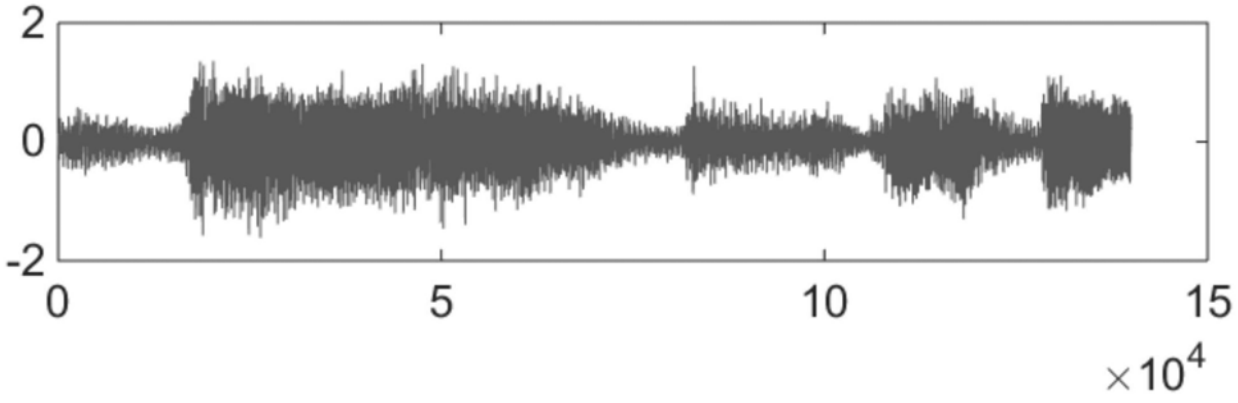

2
0
-2
0
5
10
15
×10^4

图2

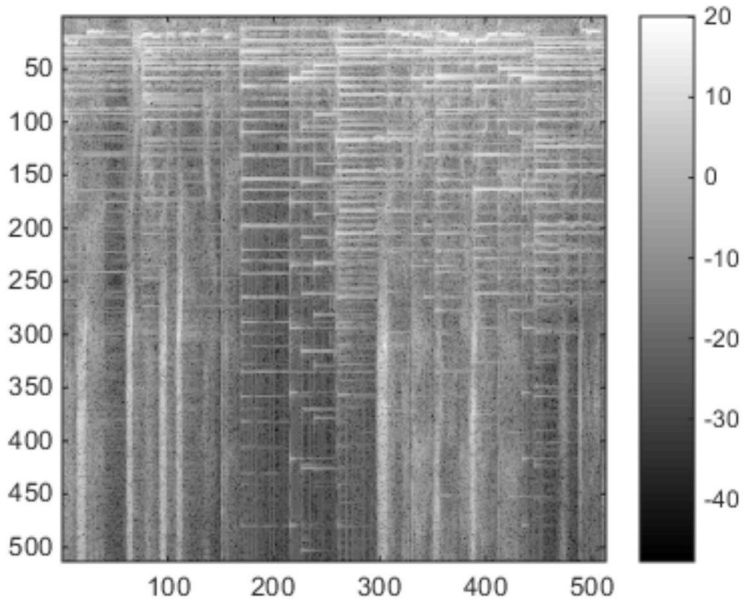

50
100
150
200
250
300
350
400
450
500
100
200
300
400
500
20
10
0
-10
-20
-30
-40

图3

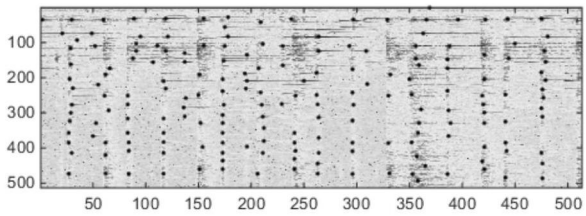

100
200
300
400
500
50
100
150
200
250
300
350
400
450
500

图4

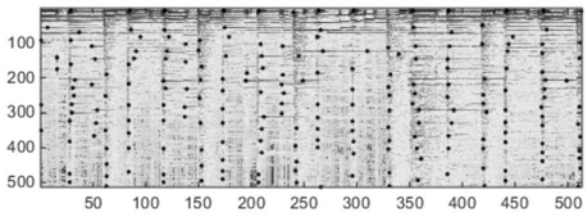

100
200
300
400
500
50
100
150
200
250
300
350
400
450
500

图5

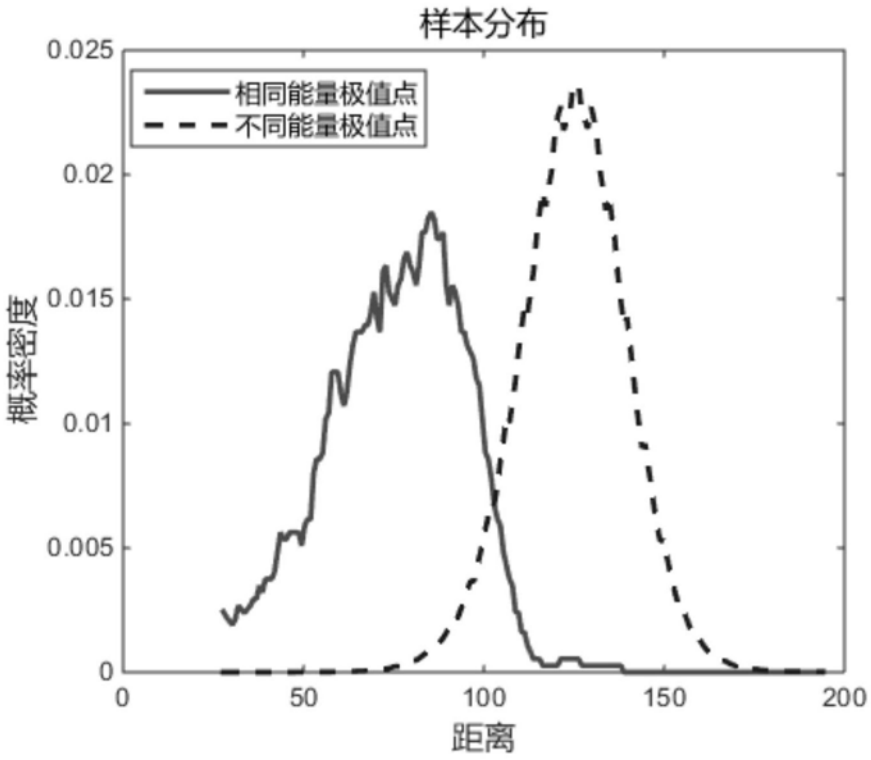

样本分布
相同能量极值点
不同能量极值点
概率密度
距离
0.025
0.02
0.015
0.01
0.005
0
0
50
100
150
200

图6

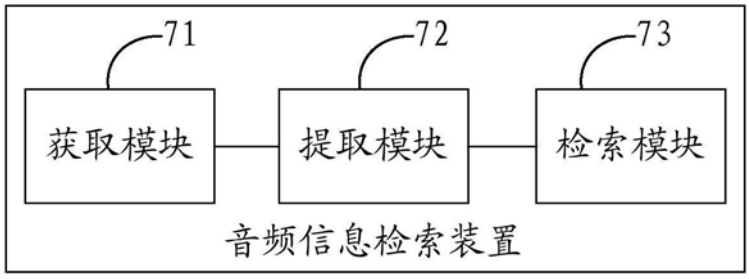

71
72
73
获取模块
提取模块
检索模块
音频信息检索装置

图7

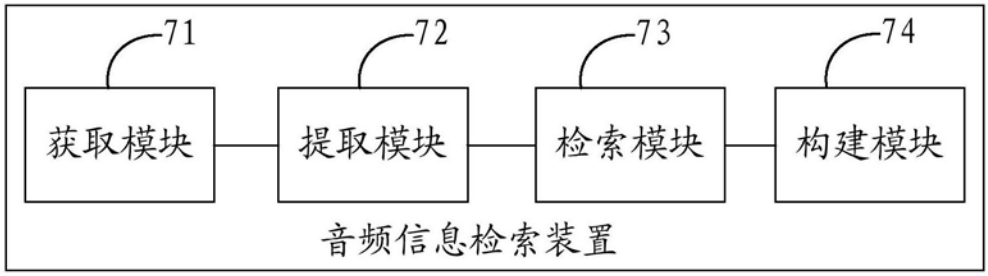

71
72
73
74
获取模块
提取模块
检索模块
构建模块
音频信息检索装置

图8

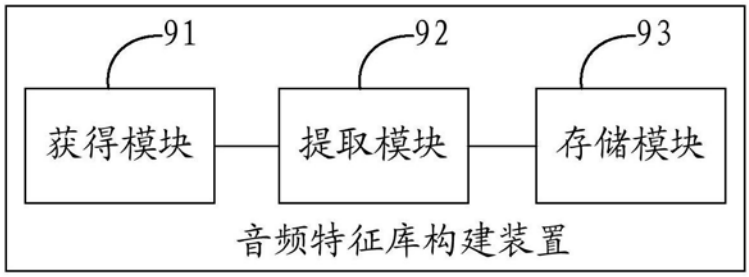


图9