

权 利 要 求

1、一种音频检索识别方法，所述方法包括：

获取待识别音频的音频指纹，其中，所述音频指纹包括用于表示所述
5 待识别音频的内容特征的第一部分以及用于表示所述第一部分的可靠程度
的第二部分；

根据所述音频指纹对所述待识别音频进行识别，得到识别结果。

2、根据权利要求1所述的音频检索识别方法，其中，所述获取待识别
音频的音频指纹包括：

10 将所述待识别音频转换成声谱图；

确定所述声谱图中的特征点；

在所述声谱图上，为所述特征点确定一个或多个掩模，每个所述掩模
包含多个谱区域；

确定每个所述谱区域的均值能量；

15 根据所述掩模中的所述多个谱区域的均值能量确定音频指纹比特；

判断所述音频指纹比特的可靠程度以确定强弱权重比特；

根据所述音频指纹比特和所述强弱权重比特确定所述待识别音频的音频
指纹。

3、根据权利要求2所述的音频检索识别方法，其中，所述将所述待识
20 别音频转换成声谱图包括：通过短时傅里叶变换将所述待识别音频转换成
时间-频率的二维声谱图，所述声谱图中每个点的取值代表所述待识别音频
的能量。

4、根据权利要求3所述的音频检索识别方法，其中，所述将所述待识
别音频转换成声谱图还包括：对所述声谱图进行梅尔变化。

25 5、根据权利要求3所述的音频检索识别方法，其中，所述特征点为所
述声谱图中的固定点。

6、根据权利要求5所述的音频检索识别方法，其中，所述特征点为频
率值与预设的多个频率设定值相等的点。

7、根据权利要求3所述的音频检索识别方法，其中，所述特征点为所
30 述声谱图中的能量极大值点，或者，所述特征点为所述声谱图中的能量极
小值点。

8、根据权利要求2所述的音频检索识别方法，其中，所述掩模所包含
的多个所述谱区域是对称分布的。

9、根据权利要求8所述的音频检索识别方法，其中，所述掩模所包含
35 的多个所述谱区域具有相同的频率范围、和/或具有相同的时间范围、和/
或以所述特征点为中心而中心对称分布。

10、根据权利要求 2 所述的音频检索识别方法，其中，所述谱区域均值能量为所述谱区域所包含的所有点的能量值的平均值。

11、根据权利要求 2 所述的音频检索识别方法，其中，所述的根据所述掩模中的所述多个谱区域的均值能量确定音频指纹比特包括：

5 根据一个所述掩模所包含的多个所述谱区域的均值能量的差值确定一个音频指纹比特。

12、根据权利要求 11 所述的音频检索识别方法，其中，所述的判断所述音频指纹比特的可信程度以确定强弱权重比特包括：

10 判断所述差值的绝对值是否达到或超过预设的强弱比特阈值，如果达到或超过所述强弱比特阈值，则将所述音频指纹比特确定为强比特，否则降所述音频指纹比特确定为弱比特；根据所述音频指纹比特是强比特还是弱比特来确定所述强弱权重比特。

13、根据权利要求 2 所述的音频检索识别方法，所述方法还包括：

将待识别音频按时间分成多段子音频；

15 提取每段所述子音频的所述音频指纹；

将提取得到的各个所述子音频的所述音频指纹进行组合，得到所述待识别音频的音频指纹。

14、根据权利要求 2 所述的音频检索识别方法，其中，将所述待识别音频的音频指纹定义为第一音频指纹，所述第一音频指纹包含多个第一音频指纹单体以及与各个所述第一音频指纹单体相对应的第一强弱权重单体，所述第一音频指纹单体包含所述待识别音频的多个所述音频指纹比特，所述第一强弱权重单体包含与所述多个音频指纹比特相对应的多个所述强弱权重比特。

15、根据权利要求 14 所述的音频检索识别方法，其中，所述根据所述音频指纹对所述待识别音频进行识别包括：

根据每个单独的所述第一音频指纹单体对多个已知音频进行第一排名，根据所述第一排名的结果，取出前 k 个所述已知音频作为第一候选音频集合，其中 k 为正整数；

16 根据多个顺序排列的所述第一音频指纹单体对所述第一候选音频集合进行第二排名，根据所述第二排名的结果，取出前 n 个所述第一候选音频作为识别结果，其中 n 为正整数。

16、根据权利要求 15 所述的音频检索识别方法，还包括：

35 预先获取所述已知音频的音频指纹作为第二音频指纹，所述第二音频指纹包含多个第二音频指纹单体以及与所述第二音频指纹单体相对应的第二强弱权重单体；

对所述第二音频指纹进行索引，以预先得到所述已知音频的指纹索引。

17、根据权利要求 16 所述的音频检索识别方法，其中，在进行所述第一排名和/或进行所述第二排名的过程中，根据所述第一强弱权重单体和/或第二强弱权重单体，对所述第一音频指纹单体和/或所述第二音频指纹单体进行加权。

5 18、根据权利要求 16 所述的音频检索识别方法，其中，所述根据每个单独的所述第一音频指纹单体对多个已知音频进行第一排名包括：根据每个单独的所述第一音频指纹单体对多个已知音频进行词频-逆向文件频率 TF-IDF 排名。

19、根据权利要求 17 所述的音频检索识别方法，其中，所述根据每个单独的所述第一音频指纹单体对多个已知音频进行词频-逆向文件频率 TF-IDF 方式的第一排名包括：

将所述已知音频的指纹索引与所述第一音频指纹单体进行匹配，以对所述已知音频进行所述 TF-IDF 排名。

20、根据权利要求 19 所述的音频检索识别方法，其中，所述预先得到所述已知音频的指纹索引包括：

根据所述第二强弱权重单体，预先得到所述已知音频的正排指纹索引和/或倒排指纹索引。

21、根据权利要求 19 所述的音频检索识别方法，其中，所述将所述已知音频的指纹索引与所述第一音频指纹单体进行匹配包括：

20 根据所述第一强弱权重单体，将所述音频的指纹索引与所述第一音频指纹单体进行绝对匹配。

22、根据权利要求 16 所述的音频检索识别方法，其中，所述根据多个顺序排列的所述第一音频指纹单体对所述第一候选音频集合进行第二排名包括：

25 根据所述已知音频的指纹索引与所述第一音频指纹得到所述第一候选音频集合中的音频的相似度矩阵，根据所述相似度矩阵对所述第一候选音频集合中的音频进行排名。

23、根据权利要求 22 所述的音频检索识别方法，其中，所述的根据所述已知音频的指纹索引与所述第一音频指纹得到所述第一候选音频集合中的音频的相似度矩阵，根据所述相似度矩阵对所述第一候选音频集合中的音频进行排名包括：

利用所述第一强弱权重单体和/或所述第二强弱权重单体进行加权，得到加权的所述相似度矩阵，根据所述加权的相似度矩阵对所述第一候选音频集合中的音频进行排名。

35 24、根据权利要求 22 所述的音频检索识别方法，其中，所述根据所述相似度矩阵对所述第一候选音频集合中的音频进行排名包括：

根据所述相似度矩阵中的直线对所述第一候选音频集合中的音频进行排名。

25、根据权利要求 22 所述的音频检索识别方法，其中：

所述获取待识别音频的音频指纹包括，获取所述待识别音频的多种类型的第一音频指纹；

所述预先获取所述已知音频的音频指纹作为第二音频指纹包括，获取所述第一候选音频集合中的音频的多种类型的第二音频指纹；

所述的根据所述已知音频的指纹索引与所述第一音频指纹得到所述第一候选音频集合中的音频的相似度矩阵包括，根据所述多种类型的第一音频指纹和所述多种类型的第二音频指纹确定所述相似度矩阵。

26、根据权利要求 25 所述的音频检索识别方法，其中，

每种类型的所述第一音频指纹包含多个第一音频指纹单体，每种类型的所述第二音频指纹包含多个第二音频指纹单体；

所述的根据所述多种类型的第一音频指纹和所述多种类型的第二音频指纹确定所述相似度矩阵包括：

分别确定同种类型的所述第二音频指纹单体与所述第一音频指纹单体之间的单体相似度，以得到多种所述单体相似度；根据所述多种单体相似度的平均值或最小值确定所述相似度矩阵。

27、根据权利要求 16 所述的音频检索识别方法，还包括：

预先对待识别音频和已知音频按照预设的时间长度切片，得到多段待识别子音频和多段已知子音频，对所述多段待识别子音频和所述多段已知子音频分别提取音频指纹，以得到长度相同的多个第一子音频指纹和多个第二子音频指纹。

28、根据权利要求 16 所述的音频检索识别方法，还包括：

在进行所述第一排名之前，对获得的待识别音频的所述第一音频指纹和已知音频的所述第二音频指纹按照预设的长度切片，以得到长度相同的多个第一子音频指纹和多个第二子音频指纹。

29、根据权利要求 22 所述的音频检索识别方法，其中，所述多个第一音频指纹单体在所述第一音频指纹中按时间顺序排列，所述多个第二音频指纹单体在所述第二音频指纹中按时间顺序排列。

30、根据权利要求 29 所述的音频检索识别方法，还包括：

根据所述相似度矩阵确定所述待识别音频与所述识别结果中的音频的重复片段。

31、一种音频检索识别装置，所述装置包括：

音频指纹获取系统，用于获取待识别音频的音频指纹，其中，所述音频指纹包括用于表示所述待识别音频的内容特征的第一部分以及用于表示

所述第一部分的可信程度的第二部分；

检索识别系统，用于根据所述音频指纹对所述待识别音频进行识别，得到识别结果。

5 32、根据权利要求 31 所述的音频检索识别装置，所述装置还包括执行权利要求 2 到 30 中任一权利要求所述步骤的模块。

33、一种音频检索识别硬件装置，包括：

存储器，用于存储非暂时性计算机可读指令；以及

10 处理器，用于运行所述计算机可读指令，使得所述计算机可读指令被所述处理器执行时实现根据权利要求 1 到 30 中任意一项所述的音频检索识别方法。

34、一种计算机可读存储介质，用于存储非暂时性计算机可读指令，当所述非暂时性计算机可读指令由计算机执行时，使得所述计算机执行权利要求 1 到 30 中任意一项所述的音频检索识别方法。

15 35、一种终端设备，包括权利要求 31 或 32 所述的一种音频检索识别装置。