一种音频指纹提取方法及装置

相关申请的交叉引用

本申请要求申请号为201810273669.6、申请日为2018年3月29日的中国专利申请的优先权，该文献的全部内容以引用方式并入本文。

技术领域

本公开涉及音频处理技术领域，特别是涉及一种音频指纹提取方法及装置。

背景技术

音频指纹（或者称为音频特征）以及音频指纹检索在如今的“多媒体信息社会”中具有广泛的应用。音频指纹检索最初被应用到听歌识曲之中，也就是输入一段音频，通过提取和比对该段音频的指纹特征，就能识别出对应的歌曲。另外，音频指纹检索也可应用到内容监控之中，比如音频消重、基于检索的语音广告监控、音频版权等。

现有的音频指纹检索方法存在准确性差的问题，在一定程度上这是由于所提取的音频指纹的准确性较差导致的。现有的音频指纹提取方法存在着对噪声的鲁棒性较差、处理复杂等问题。

发明内容

本公开的目的在于提供一种新的音频指纹提取方法及装置。

本公开的目的是采用以下的技术方案来实现的。依据本公开提出的音频指纹提取方法，包括以下步骤：将音频信号转换成声谱图；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；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信程度以确定强弱权重比特；将所述音频指纹比特以及所述强弱权重比特进行组合，得到音频指纹。

本公开的目的还可以采用以下的技术措施来进一步实现。

前述的音频指纹提取方法，其中所述将音频信号转换成声谱图包括：通过短时傅里叶变换将音频信号转换成时间-频率的二维声谱图，所述声谱图中每个点的取值代表所述音频信号的能量。

前述的音频指纹提取方法，其中所述将音频信号转换成声谱图还包括：对所述声谱图进行梅尔变化。

前述的音频指纹提取方法，其中所述将音频信号转换成声谱图还包括：

对所述声谱图进行人类听觉系统滤波。

前述的音频指纹提取方法，其中所述特征点为所述声谱图中的固定点。

前述的音频指纹提取方法，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。

前述的音频指纹提取方法，其中所述特征点为所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。

前述的音频指纹提取方法，其中所述掩模所包含的多个所述谱区域是对称分布的。

前述的音频指纹提取方法，其中所述掩模所包含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。

前述的音频指纹提取方法，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。

前述的音频指纹提取方法，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：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的取值。

前述的音频指纹提取方法，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：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，否则将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。

前述的音频指纹提取方法，其中所述的强弱比特阈值为固定值、或者为基于所述差值的值、或者为比例值。

前述的音频指纹提取方法，其还包括：将音频信号按时间分成多段音频子信号；提取所述音频子信号的所述音频指纹；将提取得到的各个所述音频子信号的所述音频指纹进行组合，得到所述音频信号的音频指纹。

本公开的目的还采用以下技术方案来实现。依据本公开提出的音频指纹库构建方法，包括：按照前述的任意一项的音频指纹提取方法提取音频信号的音频指纹；将所述音频指纹存储到音频指纹库中。

本公开的目的还采用以下技术方案来实现。依据本公开提出的音频指纹提取装置，包括：声谱图转换模块，用于将音频信号转换成声谱图；特征点确定模块，用于确定所述声谱图中的特征点；掩模确定模块，用于在所述声谱图上，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；均值能量确定模块，用于确定每个所述谱区域的均值能量；音频指纹比特确定模块，用于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；强弱权重比特确定模块，用于判断所述音频指纹比特

的可信程度以确定强弱权重比特；音频指纹确定模块，用于将所述音频指纹比特以及所述强弱权重比特进行组合，得到音频指纹。

本公开的目的还可以采用以下的技术措施来进一步实现。

前述的音频指纹提取装置，其还包括执行前述任一音频指纹提取方法步骤的模块。

本公开的目的还采用以下技术方案来实现。依据本公开提出的音频指纹库构建装置，包括：音频指纹提取模块，用于按照前述任意一项的音频指纹提取方法提取音频信号的音频指纹；音频指纹存储模块，用于将所述音频指纹存储到音频指纹库中；音频指纹库，用于存储所述音频指纹。

本公开的目的还采用以下技术方案来实现。依据本公开提出的一种音频指纹提取硬件装置，包括：存储器，用于存储非暂时性计算机可读指令；以及处理器，用于运行所述计算机可读指令，使得所述处理器执行时实现前述任意一种音频指纹提取方法。

本公开的目的还采用以下技术方案来实现。依据本公开提出的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前述任意一种音频指纹提取方法。

本公开的目的还采用以下技术方案来实现。依据本公开提出的一种终端设备，包括前述任意一种音频指纹提取装置。

上述说明仅是本公开技术方案的概述，为了能更清楚了解本公开的技术手段，而可依照说明书的内容予以实施，并且为让本公开的上述和其他目的、特征和优点能够更明显易懂，以下特举较佳实施例，并配合附图，详细说明如下。

附图说明

图1是本公开一个实施例的音频指纹提取方法的流程示意图。

图2是本公开一个实施例的音频指纹库构建方法的流程示意图。

图3是本公开一个实施例的音频指纹提取装置的结构框图。

图4是本公开一个实施例的音频指纹库构建装置的结构框图。

图5是本公开一个实施例的音频指纹提取硬件装置的硬件框图。

图6是本公开一个实施例的计算机可读存储介质的示意图。

图7是本公开一个实施例的终端设备的结构框图。

具体实施方式

为更进一步阐述本公开为达成预定发明目的所采取的技术手段及功效，以下结合附图及较佳实施例，对依据本公开提出的音频指纹提取方法及装置的具体实施方式、结构、特征及其功效，详细说明如后。

图 1 为本公开的音频指纹提取方法一个实施例的示意性流程图。请参阅图 1，本公开示例的音频指纹提取方法，主要包括以下步骤：

步骤 S11，将音频信号转换成声谱图（Spectrogram）。具体地，通过短时傅里叶变换（Fast Fourier Transformation）将音频信号转换成时间-频率声谱图。其中的声谱图是一种常用的音频信号的二维频谱图，横轴是时间 t，纵轴是频率 f，图中每个点(t,f)的具体的取值 E(t,f)代表了信号的能量。需注意，对音频信号的具体类型不做限制，可以是静态文件（static file）也可以是流音频（streaming audio）。此后，处理进到步骤 S12。

在本公开的实施例中，可利用梅尔（MEL）变换对声谱图进行预处理，利用梅尔变换能够将频谱分成多个频率区块（频率 bin），而所分成的频率区块的数目是可以配置的。另外，还可以对声谱图进行人类听觉系统滤波（Human Auditory System filtering），利用人类听觉系统滤波等非线性变换，能够使得声谱图中的频谱分布更适合人耳感知。

需要说明的是，可以通过调整短时傅里叶变换中的各个超参数以适应不同的实际情况。在本公开的实施例中，可将步骤 S11 中的各个超参数设置为：在短时傅里叶变换中，时间窗设置为 100ms，间隔设置为 50ms；在梅尔变换中，频率区块的数目设置为 32～128。

步骤 S12，确定声谱图中的特征点。

具体地，采用多种标准中的一种来确定特征点，例如，可以将特征点选为声谱图中的能量的极大值点，或者，也可以选为能量的极小值点。其中，如果声谱图中的一个点(t,f)的能量 E(t,f)能够同时满足：E(t,f)>E(t+1,f) 、 E(t,f)>E(t-1,f) 、 E(t,f)>E(t,f+1) 且 E(t,f)>E(t,f-1)，则该(t,f)点为声谱图中的能量极大值点。类似地，如果一个点(t,f)的能量 E(t,f)能够同时满足：E(t,f)<E(t+1,f)、E(t,f)<E(t-1,f)、E(t,f)<E(t,f+1) 且 E(t,f)<E(t,f-1)，则该(t,f)点为声谱图中的能量极小值点。此后，处理进到步骤 S12。

在本公开的实施例中，由于选取能量极值点作为特征点存在：能量极值点易受噪声影响；不易控制极值点的数量，可能一个声谱图中没有极值点，而另一个声谱图中有多个极值点，导致特征点不均匀；需要存储额外的时间戳以记录能量极值点在声谱图中的位置等问题。因此，也可以不选能量的极值点作为特征点，而是选取固定点作为特征点，例如可以选取频率值与预设的频率设定值相等的点（频率固定的点）。进一步地，可按照频率大小预设低频、中频、高频的多个频率设定值（低频、中频、高频的具体值是可以设置的）。通过选取频率为低频、中频、高频的多个固定点作为特征点，可以使得选取的特征点更加均匀。需要注意的是，也可以按照其他标准选取固定点，如选取与一个或多个预设能量值相等的点。

需要说明的是，可以通过调整所选取的特征点的数量以适应不同的实际情况。在本公开的实施例中，可将步骤 S12 中的超参数设置为：特征点的密度设置为每秒 20～80 个。

步骤 S13，在声谱图上，在特征点的附近，为特征点确定一个或多个掩模（mask），每个掩模包含（或者说，覆盖）多块声谱图上的区域（不妨称为谱区域）。此后，处理进到步骤 S14。

具体地，在声谱图中，每个掩模所包含的多块谱区域可以是对称分布的：

以时间轴对称（即，多个谱区域具有相同的频率范围），例如，在一个梅尔-声谱图中，可以为特征点确定一个包含 R11 和 R12 这两块谱区域的掩模，R11、R12 均位于特征点的左侧，且 R11 位于 R12 的左侧，并且 R11 与 R12 覆盖相同的频率区块；

或者以频率轴对称（即，多个谱区域具有相同的时间范围）。例如，在一个梅尔-声谱图中，可以为特征点确定一个包含 R13 和 R14 这两块谱区域的掩模，R13 位于特征点的上侧，R14 位于特征点的下侧，并且 R13 与 R14 具有相同的时间范围；

或者以特征点为中心而中心对称分布，例如，在一个梅尔-声谱图中，可以为特征点确定一个包含 R15 和 R16 这两块谱区域的掩模，R15 位于特征点的左上侧，R16 位于特征点的右下侧，并且 R15 与 R16 以特征点为中心而相互对称。

当然，一个掩模所包含的多块谱区域也可以同时满足多种对称分布情况。例如，可以为特征点确定一个包含 R21、R22、R23 和 R24 这四块谱区域的掩模，R21、R22、R23、R24 分别位于特征点的左上、右上、左下、右下，并且 R21 与 R22 具有相同的频率范围、R23 与 R24 具有相同的频率范围、R21 与 R23 具有相同的时间范围、R22 与 R24 具有相同的时间范围，而且这四块谱区域还以特征点为中心而中心对称。需要说明的是，一个掩模的四个谱区域并非一定以特征点为中心而中心对称分布，例如，可以均位于特征点的左侧，且在频率轴上分布于特征点的两侧。

需要说明的是，属于同一掩模的多块谱区域之间是可以相互交叠的。另外，不同的掩模之间也是可以相互交叠的。可选地，每个掩模可包含偶数个谱区域。

需要注意的是，掩模可以是按照固定的预设标准确定的，即每个掩模在声谱图中的位置及覆盖的区域是预先设置好的。或者，也可以不预先固定掩模的位置和范围，而是使用数据驱动的方式自动确定掩模区域：从大量掩模中选取协方差最小、最有区分度的掩模。

步骤 S14，确定每个谱区域的均值能量。具体地，对于仅包含一个点的

谱区域，该谱区域的均值能量就是这个点的能量值；当谱区域由多个点组成时，可以将该谱区域的均值能量设置为这多个点的能量值的平均值。此后，处理进到步骤S15。

步骤 S15，根据掩模中的多块谱区域的均值能量，确定音频指纹比特(bit)。此后，处理进到步骤S16。

在本公开实施例的步骤 S15 中，可根据一个掩模所包含的多个谱区域的均值能量的差值确定一个音频指纹比特。

具体地，如果一个掩模包含两个谱区域，例如前述的包含 R11 和 R12 两块谱区域的示例，可以按照下面的公式一来计算R11、R12的均值能量的差值D1：

$$D1 = E(R11) - E(R12), \quad \text{（公式一）}$$

然后判断差值 D1 的正负，如果差值D1 为正值，则得到一个取值为 1 的音频指纹比特，如果差值D1 为负值，则得到一个取值为 0 的音频指纹比特。

如果一个掩模包含四个谱区域，例如前述的包含 R21、R22、R23、R24 四块谱区域的示例，可以按照下面的公式二来计算R21、R22、R23、R24的均值能量的差值D2：

$$D2 = (E(R21) + E(R22)) - (E(R23) + E(R24)), \quad \text{（公式二）}$$

然后判断差值 D2 的正负，如果差值D2 为正值，则得到一个取值为 1 的音频指纹比特，如果差值D2 为负值，则得到一个取值为 0 的音频指纹比特。需要说明的是，并非必须通过差值D2 来确定一个包含四个谱区域的掩模的音频指纹比特，也可以利用其他形式的差值来确定音频指纹比特。例如，也可以计算这四个谱区域的均值能量的二阶差值D3：

$$D3 = (E(R23) - E(R24)) - (E(R21) - E(R22)), \quad \text{（公式三）}$$

然后判断差值D1 的正负来确定音频指纹比特。

需要说明的是，如果为特征点确定了多个掩模，则能够对应地得到多个音频指纹比特。

步骤S16，确定音频指纹比特对应的强弱权重比特，该强弱权重比特用于表示该音频指纹比特的可信程度。具体他，将可信度高的音频指纹比特定义为强比特，将可信度低的音频指纹比特定义为弱比特。判断一个音频指纹比特的可信程度，并根据该音频指纹比特是强比特还是弱比特来确定强弱权重比特的取值。此后，处理进到步骤S17。

在本公开的实施例中，如果音频指纹比特是根据一个掩模所包含的多个谱区域均值能量的差值来确定的，则步骤S16 具体包括：判断生成该音频指纹比特所使用的该差值的绝对值是否达到（或超过）预设的强弱比特阈值；如果达到强弱比特阈值，则将该音频指纹比特确定为强比特，并得到一个与该音频指纹比特对应的取值为 1 的强弱权重比特；如果未达到强弱

比特阈值，则将该音频指纹比特确定为弱比特，并得到一个与该音频指纹比特对应的取值为 0 的强弱权重比特。

作为一个具体示例，如果一个音频指纹比特是通过判断前述公式二的四个谱区域均值能量的差值 D2 的正负来确定的，则步骤 S16 包括：判断该差值 D2 的绝对值与预设的强弱比特阈值 T 的大小关系，如果|D2|≥T，则该音频指纹比特是强比特，并将该音频指纹比特对应的强弱权重比特取值设置为 1；如果|D2|＜T，则该音频指纹比特是弱比特，并将该音频指纹比特对应的强弱权重比特取值设置为 0。需要说明的是，该强弱比特阈值可以是多种类型的阈值：该强弱比特阈值可以是个预设的固定值，例如可以固定取为 1；或者，该强弱比特阈值也可以是基于均值能量的差值而得到的数值，例如可将该强弱比特阈值设置为多个掩模（或多个特征点）对应的多个差值的平均数（事实上不限于平均数，也可以是任意一个介于最大的差值与最小的差值之间的数值），并且将差值达到该平均数的音频指纹比特确定为强比特，将差值未达到该平均数的音频指纹比特确定为弱比特；再或者，该强弱比特阈值也可以是个比例值，例如可将该强弱比特阈值设置为 60%，在多个掩模（或多个特征点）对应的多个差值中，如果一个差值的绝对值位于所有差值中的前 60%，则将该音频指纹比特确定为强比特，否则将该音频指纹比特确定为弱比特。

步骤 S17，将得到的多个音频指纹比特以及多个强弱权重比特组合在一起，得到音频指纹。具体地，对音频指纹的组合方式、音频指纹的长度不做限制。例如，一个音频指纹可以包括两部分，一部分是将一个特征点的所有掩模所对应的音频指纹比特组合在一起而得到的音频指纹比特序列，然后将多个特征点对应的多个音频指纹比特序列按照特征点的时间顺序排列得到音频指纹的第一部分；另一部分是将对应的强弱权重比特组合在一起而得到的与音频指纹比特序列长度相等的强弱权重比特序列，然后将多个特征点对应的多个强弱权重比特序列按照特征点的时间顺序排列得到音频指纹的第二部分。可选地，获得的音频指纹比特序列的长度可以是 32bits。

本公开通过在提取音频指纹比特的同时，提取该指纹比特对应的强弱权重比特，能够为一段音频生成一个准确性高、鲁棒性好的音频指纹。

可选地，该音频指纹提取方法还包括：为音频指纹添加一个时间戳字段，用于表示音频起始位置与该特征点的时间差的字段，该字段可以是一个 hash 值。而如果将特征点设为固定点，则可以不必包含本步骤，即不必记录该时间戳。

可选地，该音频指纹提取方法还包括：为音频指纹添加一个音频信号标识字段，用于记录该音频指纹所对应的音频信号的 ID 标识信息，该字段

可以是一个 hash 值。

可选地，该音频指纹提取方法还包括：将音频信号按时间分成多段音频子信号；按照前述方法的步骤，对各段音频子信号提取音频指纹，得到多个音频指纹；将提取的各段音频子信号的各个特征点的音频指纹组合在一起，得到该整段音频信号的音频指纹。

作为一种可选示例，在利用本公开提取的音频指纹进行音频检索、音频识别的过程中，在计算两个音频指纹间的距离（例如汉明距离）时，针对每个音频指纹比特，利用对应的强弱权重比特进行加权，强比特的权重高，弱比特的权重低（也可将弱比特的权重设为零），以削弱或移除弱比特所占的权重，进而使音频检索对噪声更加鲁棒，有效解决噪声误码率问题。

其中的汉明距离是一种信息论领域中常用的度量，两个等长字符串之间的汉明距离是两个字符串对应位置的不同字符的个数。在实际计算汉明距离时，可以对两个字符串进行异或运算，并统计结果为 1 的个数，而这个数就是汉明距离。

图 2 为本公开的音频指纹库构建方法一个实施例的示意性流程图。请参阅图 2，本公开示例的音频指纹库构建方法，主要包括以下步骤：

步骤 S21，按照前述的本公开示例的音频指纹提取方法的步骤提取音频信号的音频指纹。此后，处理进到步骤 S22。

步骤 S22，将得到的音频信号的音频指纹存储到音频指纹库中。

需要说明的是，上述音频信号的数量越多，该音频指纹库存储到信息就越丰富。另外，随着时间的推移，可以随时对音频指纹库进行更新。

图 3 为本公开的音频指纹提取装置一个实施例的示意性结构图。请参阅图 3，本公开示例的音频指纹提取装置 100 主要包括：声谱图转换模块 101、特征点确定模块 102、掩模确定模块 103、均值能量确定模块 104、音频指纹比特确定模块 105、强弱权重比特确定模块 106 以及音频指纹确定模块 107。

其中，该声谱图转换模块 101 用于将音频信号转换成声谱图（Spectrogram）。具体地，声谱图转换模块 101 可具体用于通过短时傅里叶变换（Fast Fourier Transformation）将音频信号转换成时间-频率声谱图。

在本公开的实施例中，声谱图转换模块 101 可包括梅尔变换子模块，用于利用梅尔（MEL）变换对声谱图进行预处理，利用梅尔变换能够将频谱分成多个频率区块（bin），其中所分成的频率区块的数目是可以配置的。另外，声谱图转换模块 101 还可以包括人类听觉系统滤波子模块，用于对声谱图进行人类听觉系统滤波（Human Auditory System filtering），利用人类听觉系统滤波等非线性变换，能够使得声谱图中的频谱分布更适合人

耳感知。

该特征点确定模块 102 用于确定声谱图中的特征点。

具体地，该特征点确定模块 102 可以具体用于采用多种标准中的一种来确定特征点，例如，可以将特征点选为声谱图中的能量的极大值点，或者也可以选为能量的极小值点。

在本公开的实施例中，该特征点确定模块 102 也可以不选能量的极值点作为特征点，而是用于选取固定点作为特征点，例如可以选取频率值与预设的频率设定值相等的点（频率固定的点）。进一步地，该特征点确定模块 102 可用于按照频率大小分别选取低频、中频、高频的多个频率设定值（低频、中频、高频的具体值是可以设置的）。

该掩模确定模块 103 用于在声谱图上，在特征点的附近，为特征点确定一个或多个掩模（mask），每个掩模包含多个谱区域。具体地，在声谱图中，每个掩模所包含的多块谱区域可以是对称分布的。

该均值能量确定模块 104，用于分别确定每个谱区域的均值能量。

该音频指纹比特确定模块 105，用于根据一个掩模中的多块谱区域的均值能量，来确定一个音频指纹比特。

在本公开的实施例中，该音频指纹比特确定模块 105 可具体用于根据一个掩模所包含的多个谱区域的均值能量的差值确定一个音频指纹比特。

该强弱权重比特确定模块 106，用于判断音频指纹比特的可信程度，以确定每个音频指纹比特对应的强弱权重比特。

在本公开的实施例中，如果音频指纹比特是根据一个掩模所包含的多个谱区域均值能量的差值确定的，则该强弱权重比特确定模块 106 具体用于：判断生成该音频指纹比特所使用的该差值的绝对值是否达到（或超过）预设的强弱比特阈值；如果达到强弱比特阈值，则将该音频指纹比特确定为强比特，并得到一个取值为 1 的强弱权重比特；如果未达到强弱比特阈值，则将该音频指纹比特确定为弱比特，并得到一个取值为 0 的强弱权重比特。

该音频指纹确定模块 107，用于将得到的多个音频指纹比特以及多个强弱权重比特组合在一起，得到音频指纹。

可选地，该音频指纹提取装置 100 还包括时间戳添加模块（图中未示出），用于为音频指纹添加一个时间戳字段，用于表示音频起始位置与该特征点的时间差的字段，该字段可以是一个 hash 值。而如果将特征点设为固定点，则可以不必包含该时间戳添加模块。

可选地，该音频指纹提取装置 100 还包括音频信号标识添加模块（图中未示出），用于为音频指纹添加一个音频信号标识字段，以记录该音频指纹所对应的音频信号的 ID 标识信息。

可选地，该音频指纹提取装置 100 还包括音频分割模块（图中未示出）和音频指纹组合模块（图中未示出）。该音频分割模块用于将音频信号按时间分成多段音频子信号。利用音频指纹提取装置所包含的模块，对各段音频子信号提取音频指纹，以得到多个音频指纹。而音频指纹组合模块用于将提取的各段音频子信号的各个特征点的音频指纹组合在一起，得到该整段音频信号的音频指纹。

图 4 为本公开的音频指纹库构建装置一个实施例的示意性结构图。请参阅图 4，本公开示例的音频指纹库构建装置 200 主要包括：

音频指纹提取模块 201，包括前述的本公开示例的音频指纹提取装置 100 的声谱图转换模块 101、特征点确定模块 102、掩模确定模块 103、均值能量确定模块 104、音频指纹比特确定模块 105、强弱权重比特确定模块 106 及音频指纹确定模块 107，用于按照前述的本公开示例的音频指纹提取方法的步骤提取音频信号的音频指纹。

音频指纹存储模块 202，用于将由音频指纹提取模块 201 得到的音频信号的音频指纹存储到音频指纹库 203 中。

音频指纹数据库 203，用于存储各个音频信号的音频指纹。

图 5 是图示根据本公开的实施例的音频指纹提取硬件装置的硬件框图。如图 5 所示，根据本公开实施例的音频指纹提取硬件装置 300 包括存储器 301 和处理器 302。音频指纹提取硬件装置 300 中的各组件通过总线系统和/或其它形式的连接机构（未示出）互连。

该存储器 301 用于存储非暂时性计算机可读指令。具体地，存储器 301 可以包括一个或多个计算机程序产品，该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。该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。该非易失性存储器例如可以包括只读存储器（ROM）、硬盘、闪存等。

该处理器 302 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它形式的处理单元，并且可以控制音频指纹提取硬件装置 300 中的其它组件以执行期望的功能。在本公开的一个实施例中，该处理器 302 用于运行该存储器 301 中存储的该计算机可读指令，使得该音频指纹提取硬件装置 300 执行前述的本公开各实施例的音频指纹提取方法的全部或部分步骤。

图 6 是图示根据本公开的实施例的计算机可读存储介质的示意图。如图 6 所示，根据本公开实施例的计算机可读存储介质 400，其上存储有非暂时性计算机可读指令 401。当该非暂时性计算机可读指令 401 由处理器运行时，执行前述的本公开各实施例的音频指纹提取方法的全部或部分步骤。

图 7 是图示根据本公开实施例的终端设备的硬件结构示意图。终端设备可以以各种形式来实施，本公开中的终端设备可以包括但不限于诸如移动电话、智能电话、笔记本电脑、数字广播接收器、PDA（个人数字助理）、PAD（平板电脑）、PMP（便携式多媒体播放器）、导航装置、车载终端设备、车载显示终端、车载电子后视镜等等的移动终端设备以及诸如数字 TV、台式计算机等等的固定终端设备。

如图 7 所示，终端设备 1100 可以包括无线通信单元 1110、A/V（音频/视频）输入单元 1120、用户输入单元 1130、感测单元 1140、输出单元 1150、存储器 1160、接口单元 1170、控制器 1180 和电源单元 1190 等等。图 7 示出了具有各种组件的终端设备，但是应理解的是，并不要求实施所有示出的组件。可以替代地实施更多或更少的组件。

其中，无线通信单元 1110 允许终端设备 1100 与无线通信系统或网络之间的无线电通信。A/V 输入单元 1120 用于接收音频或视频信号。用户输入单元 1130 可以根据用户输入的命令生成键输入数据以控制终端设备的各种操作。感测单元 1140 检测终端设备 1100 的当前状态、终端设备 1100 的位置、用户对于终端设备 1100 的触摸输入的有无、终端设备 1100 的取向、终端设备 1100 的加速或减速移动和方向等等，并且生成用于控制终端设备 1100 的操作的命令或信号。接口单元 1170 用作至少一个外部装置与终端设备 1100 连接可以通过的接口。输出单元 1150 被构造为以视觉、音频和/或触觉方式提供输出信号。存储器 1160 可以存储由控制器 1180 执行的处理和控制操作的软件程序等等，或者可以暂时地存储已经输出或将要输出的数据。存储器 1160 可以包括至少一种类型的存储介质。而且，终端设备 1100 可以与通过网络连接执行存储器 1160 的存储功能的网络存储装置协作。控制器 1180 通常控制终端设备的总体操作。另外，控制器 1180 可以包括用于再现或回放多媒体数据的多媒体模块。控制器 1180 可以执行模式识别处理，以将在触摸屏上执行的手写输入或者图片绘制输入识别为字符或图像。电源单元 1190 在控制器 1180 的控制下接收外部电力或内部电力并且提供操作各元件和组件所需的适当的电力。

本公开提出的音频指纹提取方法的各种实施方式可以以使用例如计算机软件、硬件或其任何组合的计算机可读介质来实施。对于硬件实施，本公开提出的音频指纹提取方法的各种实施方式可以通过使用特定用途集成电路（ASIC）、数字信号处理器（DSP）、数字信号处理装置（DSPD）、可编程逻辑装置（PLD）、现场可编程门阵列（FPGA）、处理器、控制器、微控制器、微处理器、被设计为执行这里描述的功能的电子单元中的至少一种来实施，在一些情况下，本公开提出的音频指纹提取方法的各种实施方式可以在控制器 1180 中实施。对于软件实施，本公开提出的音频指纹提取方法

的各种实施方式可以与允许执行至少一种功能或操作的单独的软件模块来实施。软件代码可以由以任何适当的编程语言编写的软件应用程序（或程序）来实施，软件代码可以存储在存储器 1160 中并且由控制器 1180 执行。

以上，根据本公开实施例的音频指纹提取方法、装置、硬件装置、计算机可读存储介质以及终端设备，利用掩模提取音频指纹比特，并且提取对应的强弱权重比特，能够大大提高音频指纹提取的准确性和提取的效率，为音频信号生成优良程度高、鲁棒性好的音频指纹，进而使得基于本公开的音频指纹提取方法得到的音频指纹进行的音频比对、音频检索、音频消重以及音频内容监测具有更高的准确率、更高的效率和更好的鲁棒性。

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。

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。

另外，如在此使用的，在以“至少一个”开始的项的列举中使用的“或”指示分离的列举，以便例如“A、B 或 C 的至少一个”的列举意味着 A 或 B 或 C，或 AB 或 AC 或 BC，或 ABC（即 A 和 B 和 C）。此外，措辞“示例的”不意味着描述的例子是优选的或者比其他例子更好。

还需要指出的是，在本公开的系统和方法中，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。

可以不脱离由所附权利要求定义的教导的技术而进行对在此所述的技术的各种改变、替换和更改。此外，本公开的权利要求的范围不限于以上所述的处理、机器、制造、事件的组成、手段、方法和动作的具体方面。可以利用与在此所述的相应方面进行基本相同的功能或者实现基本相同的结果的当前存在的或者稍后要开发的处理、机器、制造、事件的组成、手段、方法或动作。因而，所附权利要求包括在其范围内的这样的处理、机器、制造、事件的组成、手段、方法或动作。

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。

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。

权 利 要 求

1、一种音频指纹提取方法，所述方法包括：

将音频信号转换成声谱图；

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；

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信程度以确定强弱权重比特；

将所述音频指纹比特和所述强弱权重比特进行组合，得到音频指纹。

2、根据权利要求 1 所述的音频指纹提取方法，其中，所述将所述音频信号转换成声谱图包括：通过短时傅里叶变换将所述音频信号转换成时间-频率的二维声谱图，所述声谱图中每个点的取值代表所述音频信号的能量。

3、根据权利要求 2 所述的音频指纹提取方法，其中，所述将音频信号转换成声谱图还包括：对所述声谱图进行梅尔变化。

4、根据权利要求 2 所述的音频指纹提取方法，其中，所述将音频信号转换成声谱图还包括：对所述声谱图进行人类听觉系统滤波。

5、根据权利要求 2 所述的音频指纹提取方法，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 所述的音频指纹提取方法，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 2 所述的音频指纹提取方法，其中，所述特征点为所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 1 所述的音频指纹提取方法，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 所述的音频指纹提取方法，其中，所述掩模所包含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 1 所述的音频指纹提取方法，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 1 所述的音频指纹提取方法，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：

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 所述的音频指纹提取方法，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：

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，否则将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 12 所述的音频指纹提取方法，其中，所述的强弱比特阈值为固定值、或者为基于所述差值的值、或者为比例值。

14、根据权利要求 1 所述的音频指纹提取方法，所述方法还包括：

将音频信号按时间分成多段音频子信号；

提取所述音频子信号的所述音频指纹；

将提取得到的各个所述音频子信号的所述音频指纹进行组合，得到所述音频信号的音频指纹。

15、一种音频指纹库构建方法，所述方法包括：

按照如权利要求 1 到 14 中任意一项所述的音频指纹提取方法提取音频信号的音频指纹；

将所述音频指纹存储到音频指纹库中。

16、一种音频指纹提取装置，所述装置包括：

声谱图转换模块，用于将音频信号转换成声谱图；

特征点确定模块，用于确定所述声谱图中的特征点；

掩模确定模块，用于在所述声谱图上，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；

均值能量确定模块，用于确定每个所述谱区域的均值能量；

音频指纹比特确定模块，用于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；

强弱权重比特确定模块，用于判断所述音频指纹比特的可信程度以确定强弱权重比特；

音频指纹确定模块，用于将所述音频指纹比特和所述强弱权重比特进行组合，得到音频指纹。

17、根据权利要求 16 所述的音频指纹提取装置，所述装置还包括执行权利要求 2 到 14 中任一权利要求所述步骤的模块。

18、一种音频指纹库构建装置，所述装置包括：

音频指纹提取模块，用于按照如权利要求 1 到 14 中任意一项所述的音频指纹提取方法提取音频信号的音频指纹；

音频指纹存储模块，用于将所述音频指纹存储到音频指纹库中；

音频指纹库，用于存储所述音频指纹。

19、一种音频指纹提取硬件装置，包括：

存储器，用于存储非暂时性计算机可读指令；以及

处理器，用于运行所述计算机可读指令，使得所述处理器执行时实现根据权利要求 1 到 14 中任意一项所述的音频指纹提取方法。

20、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 14 中任意一项所述的音频指纹提取方法。

21、一种终端设备，包括权利要求 16 或 17 所述的一种音频指纹提取装置。