

[19] 中华人民共和国国家知识产权局

[51] Int. Cl<sup>7</sup>

G06F 17/00

G11B 20/00

G11B 27/11



# [12] 发明专利申请公开说明书

[21] 申请号 200380103022.0

[43] 公开日 2005 年 12 月 21 日

[11] 公开号 CN 1711531A

[22] 申请日 2003.10.31

[21] 申请号 200380103022.0

[30] 优先权

[32] 2002.11.12 [33] EP [31] 02079720.5

[86] 国际申请 PCT/IB2003/004894 2003.10.31

[87] 国际公布 WO2004/044820 英 2004.5.27

[85] 进入国家阶段日期 2005.5.11

[71] 申请人 皇家飞利浦电子股份有限公司

地址 荷兰艾恩德霍芬

[72] 发明人 徐镇洙 J·A·海特斯马

A·A·C·M·卡克

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 吴立明 王勇

权利要求书 1 页 说明书 5 页 附图 2 页

[54] 发明名称 采指纹多媒体内容

[57] 摘要

公开了用于从多媒体信号，特别是音频信号抽取指纹的方法和装置，所述指纹对音频信号的速度改变不变。为此，该方法包括从多媒体信号，例如音频信号的功率谱抽取(12, 13)一组健壮感性特征。Fourier-Mellin 变换(15)将功率谱转换只有当音频重放速度改变时，才经受相变的傅里叶系数。它们的数值或相位差(16)构成速度改变-不变指纹。通过阈值处理操作(19)，用压缩的位数表示指纹。

I S S N 1 0 0 8 - 4 2 7 4

1. 一种从多媒体信号抽取指纹的方法，包括步骤：
  - 从所述多媒体信号抽取（12,13）一组健壮感性特征；
  - 5 -使所抽取的特征集经受（15）Fourier-Mellin 变换；
  - 将所变换的特征集转换（16,19）成构成所述指纹的序列。
2. 如权利要求 1 所述的方法，其特征在於,所述转换步骤包括转换（16, ABS）所述 Fourier-Mellin 变换的数值。
3. 如权利要求 1 所述的方法，其特征在於,所述转换步骤包括转换  
10 换（16,  $\Delta \phi$ ）所述 Fourier-Mellin 变换的相位的导数。
4. 如权利要求 1 所述的方法，其特征在於,所述多媒体信号是音频信号以及所述 Fourier-Mellin 变换包括应用于所述感性特征集的一维的对数映射。
5. 如权利要求 1 所述的方法，其特征在於,所述多媒体信号是图  
15 像或视频信号以及所述 Fourier-Mellin 变换包括应用于所述感性特征集的二维对数-极性映射过程。
6. 如权利要求 1 所述的方法，其特征在於,所述多媒体信号是图  
像或视频信号以及所述 Fourier-Mellin 变换包括应用于所述感性特征集的二维重对数映射过程。
- 20 7. 如权利要求 1 所述的方法，其特征在於,所述抽取步骤包括规格化所述感性特征集。
8. 一种用于从多媒体信号抽取指纹的装置，包括：
  - 用于从所述多媒体信号抽取一组健壮感性特征的装置（12,13）；
  - 用于使所抽取的特征集经受 Fourier-Mellin 变换的装置（15）；
  - 25 -用于将所变换的特征集转换成构成所述指纹的序列的装置（16,19）。

## 采指纹多媒体内容

## 发明领域

5 本发明涉及从多媒体信号抽取指纹的方法和装置。

## 发明背景

10 指纹，在文献中有时称为散列或签名，是从多媒体内容中抽取的二进制序列，能用来识别所述内容。不同于数据文件的加密散列（一旦该数据文件的单个位改变则会改变），多媒体内容（音频、图像、视频）的指纹对于诸如压缩和 D/A & A/D 转换的处理，在一定程度上是无变化的。这通常通过从该内容的感性基本特征抽取指纹来实现。

15 从多媒体信号抽取指纹的现有技术方法在国际专利申请 WO02/065782 中公开。该方法包括以下步骤：从多媒体信号抽取一组健壮的感性特征，以及将特征集转换成指纹。对音频信号，感性特征是所选取子频带中的音频内容的能量。对图像信号，感性特征是图像所划分的块的平均亮度。通过阈值处理，例如通过将每个特征样本与它的邻居比较，执行到二进制序列的转换。

20 采指纹的有吸引力的应用是内容识别。能通过从未知材料的摘录采指纹以及将其发送到存储所述信息的指纹的大型数据库，识别音乐歌曲或录像片段的艺术家和名称。

实验已经表明对于几乎所有的通用的音频处理操作，诸如 MP3 压缩和解压缩、均衡化、重新采样、噪声增加，和 D/A & A/D 转换，从音频信号抽取指纹的现有技术方法非常健壮。

25 无线电台加速音频几个百分比十分寻常。推测他们执行该操作有两个原因。首先，歌曲的持续时间会更短，因此允许他们广播更多广告片。第二，歌曲的节拍更快以及听众似乎更喜欢此。速度改变通常位于零和四个百分比之间。

30 音频材料的速度改变使得时域和频域中的不重合。现有技术指纹抽取法不受时域中的不重合的影响，因为指纹是从重叠音频帧中抽取的较小的子指纹的拼接。假定 2% 的速度改变仅使得在相应初始摘录的第 225 个子指纹的位置处抽取摘录的第 250 个子指纹。

频率域内中的不重合由移动到其他频率的声谱能量所引起。 2%加速的上述例子使得所有声频增加 2%。 在现有技术音频指纹抽取法中，这使得所选择的子频带中的能量（以及指纹）改变。 因此，在数据库中，不再能找到该指纹，除非对应于不同速度版本的多个指纹存储  
5 存储在用于每个歌曲的数据库中。

类似考虑适用于图像和视频材料以及用于指纹抽取的其他类型感性特征。

### 发明内容

10 本发明的目的在于提供用于从多媒体内容抽取指纹的改进方法和装置。 本发明具体的目的是提供用于从对于音频信号的速度改变基本上无变化的音频信号抽取指纹的方法和装置。

为此目的，根据本发明，从多媒体信号抽取指纹的方法包括以下步骤：从多媒体信号抽取一组健壮的感性特征；使所抽取的特征集经  
15 受 Fourier-Mellin 变换；以及将转换的特征集转换成构成指纹的序列。

按本发明的理解而采用的 Fourier-Mellin 变换包括对数映射和傅里叶变换。 由于移动中的速度改变，对数映射转换能谱的度量。 随后的傅里叶变换将移动转换成对所有傅里叶系数一样的相变。 傅里叶系数的数值不受速度改变的影响。 因此，由该数值或从由傅里叶系数的  
20 相位导数导出的指纹对速度改变无变化。

### 附图说明

图 1 示意性地表示根据本发明的用于从多媒体信号抽取指纹的装置，相当于抽取这种指纹的方法的对应步骤。

25 图 2 和 3 表示示例说明图 1 中所示的对数映射电路的操作的曲线图。

### 具体实施方式

30 将参考用于从音频信号抽取指纹的装置描述本发明。 图 1 示意性地表示根据本发明的这种装置。

该装置包括分帧电路 11，将音频信号划分成约 0.4 秒的重叠帧以及  $31/32$  的重叠因子。 选择重叠以便获得后续帧的子指纹间的高度相

关性。在划分成帧之前，音频信号已经局限于约 300Hz-3kHz 的频率范围和向下采样(未示出)，以便每个帧包括 2048 个样本。

傅里叶变换电路 12 计算每个帧的谱表示。在下一块 13 中，例如通过取(复数的)傅里叶系数的数值的平方，计算音频帧的功率谱。对 5 2048 个音频信号样本的每个帧，用 1024 个样本表示功率谱(正的和相应的负频率具有相同数值)。功率谱的样本构成一组健壮的感性特征。声谱基本上不受诸如 D/A & A/D 转换或 MP3 压缩的操作影响。

在计算功率谱后，可选的规格化电路 14 将局部规格化施加到功率谱上。这种规格化(包括解卷积和过滤)改进了性能，因为它获得更多决定性的和健壮的功率谱表示。局部规格化保留声谱的重要特征以及对于各种音频处理，包括诸如均衡化的音频声谱的局部修改，是健壮的。大部分有前途的方法是通过用其局部平均数规格化它来加重声谱的音调部分。

数学上，通过按照其局部平均数  $Lm(\omega)$  划分声谱  $A(\omega)$  来获得规格化声谱  $N(\omega)$  如下：

$$N(\omega) = \frac{A(\omega)}{Lm(\omega)}$$

能以各种方式计算局部平均数，例如：

$$Lm(\omega) = \frac{1}{2\delta} \int_{\omega-\delta}^{\omega+\delta} A(\tau) d\tau \quad (\text{算术平均}), \text{ 或者}$$

$$Lm(\omega) = \exp \left[ \frac{1}{2\delta} \int_{\omega-\delta}^{\omega+\delta} \log A(\tau) d\tau \right] \quad (\text{几何平均}) \text{ 等等。}$$

规格化声谱对均衡化保持不变。此外，音调信息直接与人的听觉有关以及在大多数音频处理后得以保留。音调信息的重要性被广泛地接受并已经用于音频识别和声频压缩的位分配中。尽管局部规格化具有许多优点，如果在  $\omega-\delta$  和  $\omega+\delta$  间没有音调分量，在压缩之后的规格化不一致。为减轻该影响，将随时间的积分和总能量项添加到  $Lm(\omega)$ 。25 然后，给出修改的局部平均值  $Lm'(\omega)$  如下：

$$Lm'(\omega) = \frac{1}{2\delta} \int_{t-\Delta}^t \int_{\omega-\delta}^{\omega+\delta} A(\tau) d\tau + \alpha \int_{t-\Delta}^t \int_{-\infty}^{\infty} A(\tau) d\tau$$

其中， $\Delta$  和  $\alpha$  是实验上确定的常数。对时间的积分使规格化更一致，以及在规格化后，总能量项限制了小的非音调分量的增加。

本发明的应用在于将 Fourier-Mellin 变换 15 应用于功率谱以便实现速度改变的弹性。 Fourier-mellin 变换包括对数映射过程 151 和傅里叶变换(或傅里叶逆变换) 152。

图 2 和 3 示出示例说明对数映射操作的曲线图。 在图 2 中, 参考标记 21 表示在正以正常速度重放音频信号情况下, 由傅里叶变换 12 提供的音频帧的功率谱的样本。 为简洁起见, 示出范围 300-3,000Hz 中的平滑功率谱。 实际上, 声谱通常显示出锯齿形的轮廓。 在图 2 中的参考标记 22 表示在正以增加的速度重放音频信号情况下, 相同音频帧的功率谱。 正如在图中所看到的那样, 速度改变引起功率谱的缩放。

图 3 示出由对数映射电路 151 计算的相应功率谱。 功率谱现在表示在所选数目的连续的对数间隔的子频带中的音频帧的能量。 参考标记 31 表示用于正以正常速度重放的音频信号的对数映射功率谱。 参考标记 32 表示用于正以增加的速度重放的音频信号的对数映射功率谱。

能以多个方式执行对数映射的过程。 在图 3 中所示的所述实施例中, 内插输入功率谱和以对数间隔的间距进行重新采样。 在另一个实施例中(未示出), 累积输入功率谱的对数间隔的(和以大小排列的)子频带内的样本以便提供对数映射功率谱的各个样本。

选择表示对数映射功率谱的样本的数量以便以足够的精度执行随后的操作。 在实际的实施例中, 由 512 个样本表示对数映射功率谱。 对图 3 的观察将可以理解, 对数映射操作将由于速度改变的功率谱的缩放(21→22)转化成移位(31→32)。 只要音频信号的重放速度不在帧周期(实际上是合理假定)内改变, 该移位对所有系数相同。

随后的傅里叶变换 152 将所述移位转化成复杂的傅里叶系数的相位的改变。 相变对所有系数相同。 因此, 如果音频信号的速度改变, 通过傅里叶变换电路 152 计算的所有傅里叶系数的相位改变相同量。 换句话说, 系数的数值和它们的相位差对于速度改变不变。 在计算电路 16 中计算它们。 因为数值和相位差对于正负频率相同, 唯一值的数量为 256。

表示音频帧的对数映射功率谱的 256 数值或相位差的向量在下文中表示  $F(k, n)$ , 其中,  $k = 1..256$  以及  $n$  为音频帧数量。 实际上,

向量构成速度改变-不变的指纹。然而，值的数量较大，以及在数字指纹系统中，每个值要求多位表示。通过仅选择最低位值，能减少表示指纹的位数。通过选择电路 17 执行此操作。已经发现 32 个最低值(最高有效系数)提供对数映射功率谱的足够精确表示。

- 5 通过使选择数值或值的相位差经受阈值处理过程，能进一步减少位数。在简单实施例中，阈值处理阶段 19 对每个特征样本产生一位，例如，如果  $F(k, n)$  高于阈值，则为‘1’，以及如果低于所述阈值，则为‘0’。可替换地，如果对应特征样本  $F(k, n)$  大于其邻居，指纹位赋予值‘1’，否则它为‘0’。为此，在一维时间滤波器 18
- 10 中，首先过滤特征样本  $F(k, n)$ 。本实施例使用后者可替换方案的改进版本。在该优选实施例中，如果特征样本  $F(k, n)$  大于其邻居并且如果对于在先前帧中也是该情形，生成指令纹位“1”，否则该指纹位为“0”。在该实施例中，过滤器 18 为二维滤波器。在数学表示法中：

$$15 \quad FP(k, n) = \begin{cases} 1 & \text{如果 } F(k, n) - F(k+1, n) - (F(k, n-1) - F(k+1, n-1)) > 0 \\ 0 & \text{如果 } F(k, n) - F(k+1, n) - (F(k, n-1) - F(k+1, n-1)) \leq 0 \end{cases}$$

当使用阈值处理，正从音频帧中抽取的每个子指纹具有 32 位。

- 尽管已经参考音频指纹描述了本发明，它也能应用于其他多媒体信号，诸如图像和动态视频。尽管速度改变通常应用于音频信号，仿射变换，诸如移位、缩放和旋转通常应用于图像和视频。根据本发明的
- 20 的方法能用来改进仿射变换的健壮性。在二维信息情况下，对数映射过程 151 被变成对数极性映射以便相对于旋转和缩放(保留纵横比)使其不变。重对数映射使它对于纵横比的改变不变。沿频率轴的 Fourier-Mellin 变换的数值(现在为二维变换)及其相位的双微分具有仿射不变特性。

- 25 公开了用于从多媒体信号，特别是音频信号抽取指纹的方法和装置，所述指纹对音频信号的速度改变不变。为此目的，该方法包括从多媒体信号，例如音频信号的功率谱抽取(12,13)一组健壮感性特征。Fourier-Mellin 变换(15)将功率谱转换只有当音频重放速度改变时，才经受相变的傅里叶系数。它们的数值或相位差(16)构成速度改变-
- 30 不变指纹。通过阈值处理操作(19)，用压缩的位数表示指纹。

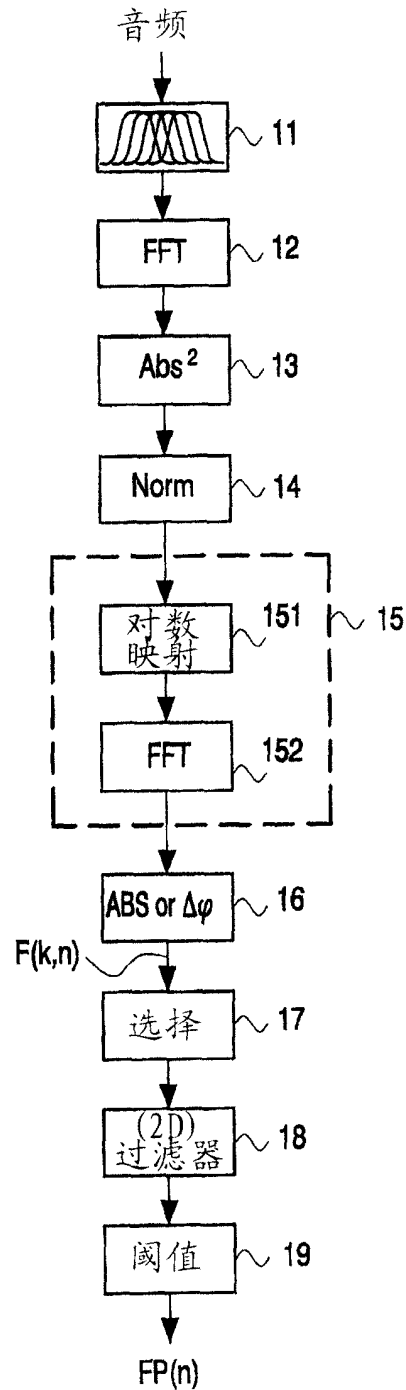


图 1



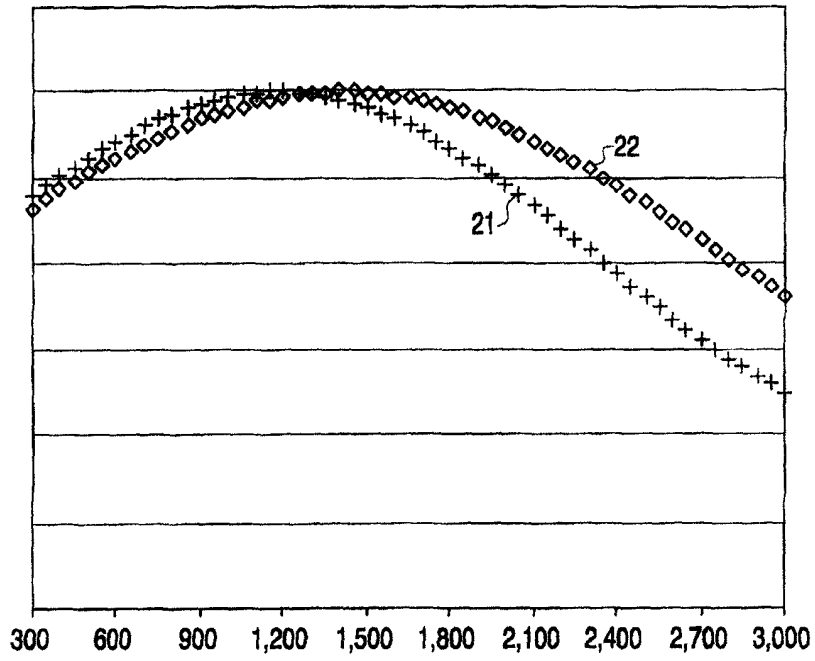


图 2

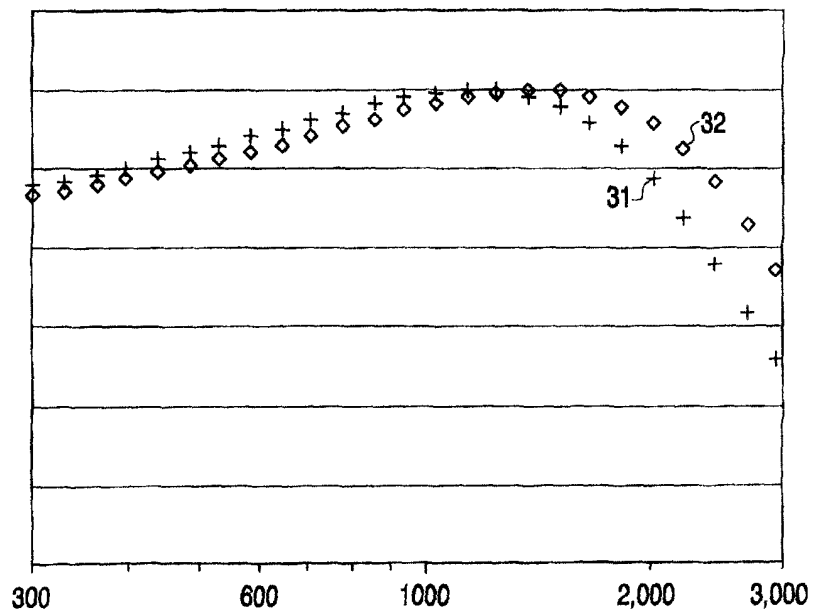


图 3