

音频信号的分类方法及装置

本申请要求于 2009 年 3 月 27 日提交中国专利局，申请号为 200910129157.3，发明名称为“音频信号的分类方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5

技术领域

本发明涉及通信技术领域，尤其涉及一种音频信号的分类方法及装置。

背景技术

语音编码器擅长于在中低码率下对语音类型的音频信号进行编码，而对音乐类型的音频信号编码效果则欠佳；音频编码器适用于在高码率下对语音类型和音乐类型的音频信号进行编码，但在中低码率下对语音类型的音频信号编码效果不够理想。为了使语音音频混合的音频信号在中低码率下能够取得良好的编码效果，一个适用于在中低码率下的语音音频编码器的编码过程主要包括：首先利用信号分类模块判别音频信号的类型，再根据判别出来的音频信号的类型选择对应的编码方法，对于语音类型的音频信号选择语音编码器，对于音乐类型的音频信号选择音频编码器。

在现有技术当中，对于上述判别音频信号的类型的方法主要包括：

- 1、利用窗函数将输入信号划分为一系列的重叠的帧；
- 2、利用快速傅里叶变换（FFT）计算每帧的频谱系数；
- 20 3、根据每帧的频谱系数，对于每个段计算五个方面的特征参数：谐波、噪音、拖尾、拖延及节奏；
- 4、基于上述特征参数的值，把音频信号分为六类：语音类型、音乐类型、噪音类型、短序列、待定序列、及短待定序列。

在实现上述判别音频信号类型的过程中，发明人发现现有技术中至少存

在如下问题：该方法在分类的过程中需要计算多方面的特征参数，对于音频信号
5 的分类也较复杂，由此而导致了分类复杂度较高。

发明内容

本发明的实施例提供一种音频信号的分类方法及装置，降低音频信号分类

5 复杂度，减少运算量。

为达到上述目的，本发明的实施例采用如下技术方案：

一种音频信号的分类的方法，包括：

获取待分类音频信号在至少一个子带中的音调特征参数；

根据获取的特征参数判定所述待分类音频信号的类型。

10 一种音频信号的分类的装置，包括：

音调获取模块，用于获取待分类音频信号在至少一个子带中的音调特征参
数；

分类模块，用于根据获取的特征参数判定所述待分类音频信号的类型。

本发明实施例提供的方案通过采用音频信号的音调特性对音频信号进行
15 分类的技术手段，克服了现有技术中在对音频信号分类时分类复杂的技术问
题，进而达到了降低音频信号分类复杂度，减少分类时所需要的运算量的技术
效果。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施
20 例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述
中的附图仅仅是本发明的一些实施例，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。

图 1 为本发明实施例 1 音频信号的分类方法的流程图；

图 2 为本发明实施例 2 音频信号的分类方法的流程图；

图 3 为本发明实施例 3 音频信号的分类方法的流程图;

图 4 为本发明实施例 4 音频信号的分类装置的框图;

图 5 为本发明实施例 5 音频信号的分类装置的框图;

图 6 为本发明实施例 6 音频信号的分类装置的框图。

5 具体实施方式

下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,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,都属于本发明保护的范围。

10 本发明实施例提供了一种音频信号的分类方法及装置,该方法的具体执行过程包括:获取待分类音频信号在至少一个子带中的音调特征参数;根据获取的特征参数判定所述待分类音频信号的类型。

该方法通过包括如下模块的装置实现:音调获取模块和分类模块。其中,该音调获取模块用于获取待分类音频信号在至少一个子带中的音调特征参数;
15 分类模块用于根据获取的特征参数判定所述待分类音频信号的类型。

本发明实施例提供音频信号的分类方法及装置,通过获取音调特征参数便可以判断出待分类音频信号的类型,需要计算的特征参数的方面少,分类方法简单,降低了分类过程中的运算量。

实施例 1

20 本实施例提供一种音频信号的分类方法,如图 1 所示,该方法包括如下步骤:

501,接收到当前一帧音频信号,该音频信号即为待分类音频信号。

具体为:设采样频率为 48kHz,帧长 $N=1024$ 个样本点,接收到的当前一帧音频信号为第 k 帧音频信号。

25 下述为计算当前一帧音频信号的音调特征参数的过程。

502,计算该当前一帧音频信号的功率谱密度。

具体为：对第 k 帧音频信号的时域数据进行加汉宁窗的加窗处理。

可通过如下汉宁窗的公式计算：

$$h(l) = \sqrt{\frac{8}{3}} \cdot 0.5 \cdot \left[1 - \cos\left(2\pi \cdot \frac{l}{N}\right) \right], \quad 0 \leq l \leq N-1 \quad (1)$$

5 其中，N 代表帧长，h(1) 代表第 k 帧音频信号的第 1 个样本点的汉宁窗数据。

对加窗后的第 k 帧音频信号的时域数据进行长度为 N 的 FFT 变换（因为 FFT 变换是关于 N/2 对称的，所以实际计算长度为 N/2 的 FFT 变换即可），并利用 FFT 变换系数计算该第 k 帧音频信号中第 k' 个功率谱密度。

10 该第 k 帧音频信号中第 k' 个功率谱密度可通过如下计算公式计算：

$$X(k') = 10 \cdot \log_{10} \left| \frac{1}{N} \sum_{l=0}^{N-1} \{h(l) \cdot s(l) \cdot e^{[-jk'l \cdot 2\pi/N]}\} \right|^2 = 20 \cdot \log_{10} \left| \frac{1}{N} \sum_{l=0}^{N-1} \{h(l) \cdot s(l) \cdot e^{[-jk'l \cdot 2\pi/N]}\} \right| \text{ dB} \quad (2)$$

$0 \leq k' \leq N/2, 0 \leq l \leq N-1$

15

其中 s(l) 代表第 k 帧音频信号的原始输入样本点，X(k') 代表第 k 帧音频信号中第 k' 个功率谱密度。

对计算出的功率谱密度 X(k') 进行校正，使得该功率谱密度的最大值为参考声压级(96dB)。

20 503，利用上述功率谱密度检测在频率区域的每个子带中是否有音调的存在，并统计在对应子带中存在的音调的个数，将该音调个数作为在该子带中的子带音调个数。

具体为：将频率区域划分为四个频率子带，分别用 sb_0 、 sb_1 、 sb_2 及 sb_3 表示这四个频率子带。如果功率谱密度 X(k') 与相邻的第若干个功率谱密度之间
25 满足一定的条件，该一定条件在本实施例中可以为如下公式 (3) 的所示的条件，则认为与该 X(k') 对应的子带中含有音调，并对该音调的个数进行统计，

得出在该子带中的子带音调个数 NT_{k_i} ，该 NT_{k_i} 代表第 k 帧音频信号在子带 sbi (i 代表子带的编号，并且 $i=0, 1, 2, 3$) 中的子带音调个数。

$$X(k'-1) < X(k') \leq X(k'+1) \quad \text{and} \quad X(k') - X(k'+j) \geq 7dB$$

(3)

5 其中， j 的取值规定如下：

$$j = \begin{cases} -2, +2 & \text{for } 2 \leq k' < 63 \\ -3, -2, +2, +3 & \text{for } 63 \leq k' < 127 \\ -6, \Delta, -2, +2, \Delta, +6 & \text{for } 127 \leq k' < 255 \\ -12, \Delta, -2, +2, \Delta, +12 & \text{for } 255 \leq k' < 500 \end{cases}$$

10

在本实施例中，已知功率谱密度的系数个数（即长度）为 $N/2$ ，对应于上述 j 的取值规定，对于 k' 值的取值区间的意义进一步说明如下：

sb_0 ：对应 $2 \leq k' < 63$ ，对应的功率谱密度系数为第 0 个到第 $(N/16-1)$ 个，对应的频率范围是 $[0\text{kHz}, 3\text{kHz})$ ；

15 sb_1 ：对应 $63 \leq k' < 127$ ，对应的功率谱密度系数为第 $N/16$ 个到第 $(N/8-1)$ 个，对应的频率范围是 $[3\text{kHz}, 6\text{kHz})$ ；

sb_2 ：对应 $127 \leq k' < 255$ ，对应的功率谱密度系数为第 $N/8$ 个到第 $(N/4-1)$ 个，对应的频率范围是 $[6\text{kHz}, 12\text{kHz})$ ；

20 sb_3 ：对应 $255 \leq k' < 500$ ，对应的功率谱密度系数为第 $N/4$ 个到第 $N/2$ 个，对应的频率范围是 $[12\text{kHz}, 24\text{kHz})$ 。

其中， sb_0 及 sb_1 对应低频子带部分； sb_2 对应较高频子带部分； sb_3 对应高频子带部分。

具体统计 NT_{k_i} 的过程如下：

25 对于子带 sb_0 ，使 k' 在大于等于 2 小于 63 的区间内逐一取值，对于每一个 k' 的取值，判断其是否满足公式 (3) 的条件，在遍历完整个 k' 的取值区间后，统计满足条件的 k' 的个数，该满足条件的 k' 个数，即为第 k 帧音频信号在子

带 sb_0 中的存在的子带音调个数 NT_{k_0} 。

例如：若当 $k'=3$ ， $k'=5$ ， $k'=10$ 时，公式（3）成立，则认为在子带 sb_0 中有3个子带音调，即 $NT_{k_0}=3$ 。

同样地，对于子带 sb_1 ，使 k' 在大于等于63小于127的区间内逐一取值，

- 5 对于每一个 k' 的取值，判断其是否满足公式（3）的条件，在遍历完整个 k' 的取值区间后，统计满足条件的 k' 的个数，该满足条件的 k' 个数，即为第 k 帧音频信号在子带 sb_1 中的存在的子带音调个数 NT_{k_1} 。

同样地，对于子带 sb_2 ，使 k' 在大于等于127小于255的区间内逐一取值，

- 10 对于每一个 k' 的取值，判断其是否满足公式（3）的条件，在遍历完整个 k' 的取值区间后，统计满足条件的 k' 的个数，该满足条件的 k' 个数，即为第 k 帧音频信号在子带 sb_2 中的存在的子带音调个数 NT_{k_2} 。

利用同样的方法，也可统计该第 k 帧音频信号在子带 sb_3 中的存在的子带音调个数 NT_{k_3} 。

504，计算当前一帧音频信号的总音调个数。

- 15 具体为：根据503统计出的 NT_{k_i} 计算第 k 帧音频信号在四个子带 sb_0 、 sb_1 、 sb_2 及 sb_3 中的子带音调个数之和。

该第 k 帧音频信号在四个子带 sb_0 、 sb_1 、 sb_2 及 sb_3 中的子带音调个数之和即为该第 k 帧音频信号的中的音调个数，具体可通过如下公式计算：

$$NT_{k_sum} = \sum_{i=0}^3 NT_{k_i} \quad (4)$$

20

其中， NT_{k_sum} 代表第 k 帧音频信号的总音调个数。

505，计算在规定帧数内当前一帧音频信号在对应子带中的子带音调个数均值。

- 25 具体为：设该规定帧数为 M ，在该 M 帧内包括第 k 帧音频信号和第 k 帧的前 $(M-1)$ 帧音频信号，根据 M 的值与 k 的值之间关系计算第 k 帧音频信号在这 M 帧音频信号每个子带中的子带音调个数均值。

该子带音调个数均值具体可通过如下公式 (5) 计算:

$$ave_NT_i = \begin{cases} \frac{\sum_{j=0}^k NT_{j_i}}{k+1} & \text{if } k < (M-1) \\ \frac{\sum_{j=k-M+1}^k NT_{j_i}}{M} & \text{if } k \geq (M-1) \end{cases} \quad (5)$$

其中, NT_{j_i} 代表第 j 帧音频信号在子带 i 中的子带音调个数, ave_NT_i 代表在子带 i 中的子带音调个数均值。特别地, 由公式 (5) 可知, 在计算时需根据 k 的值与 M 的值的 关系选择适当的公式进行计算。

特别地, 在本实施例中根据设计的需要, 不必对每个子带都计算子带音调个数均值, 计算在低频子带 sb_0 中的子带音调个数均值 ave_NT_0 , 及在较高频子带 sb_2 中的子带音调个数 ave_NT_2 即可。

506, 计算在规定帧数内当前一帧音频信号总的音调个数均值。

具体为: 设该规定帧数为 M , 在该 M 帧内包括第 k 帧音频信号和第 k 帧的前 $(M-1)$ 帧音频信号, 根据 M 的值与 k 的值之间的关系计算第 k 帧音频信号在这 M 帧的音频信号内平均每帧音频信号包含的总音调个数。

该总音调个数具体可如下公式 (6) 计算:

$$ave_NT_{sum} = \begin{cases} \frac{\sum_{j=0}^k NT_{j_sum}}{k+1} & \text{if } k < (M-1) \\ \frac{\sum_{j=k-M+1}^k NT_{j_sum}}{M} & \text{if } k \geq (M-1) \end{cases} \quad (6)$$

其中, NT_{j_sum} 代表第 j 帧总音调个数, ave_NT_{sum} 代表总的音调个数均值。特别地, 由公式 (6) 可知, 在计算时需根据 k 的值与 M 的值的 关系选择适当的公式进行计算。

507, 将计算出的在至少一个子带中的子带音调个数均值与总的音调个数均值之比分别作为当前一帧音频信号在对应子带中的音调特征参数。

该音调特征参数具体可通过如下公式 (7) 计算:

$$ave_NT_ratio_i = \frac{ave_NT_i}{ave_NT_{sum}}$$

(7)

5 其中, ave_NT_i 代表在子带 i 中的子带音调个数均值, ave_NT_{sum} 代表总的音调个数均值, $ave_NT_ratio_i$ 代表第 k 帧音频信号在子带 i 中的子带音调个数均值与总的音调个数均值的比值。

特别地, 在本实施例中, 利用 205 计算出来的在低频子带 sb_0 中的子带音调个数均值 ave_NT_0 及在较高频子带 sb_2 中的子带音调个数均值 ave_NT_2 , 通过公式 (7) 可分别计算出第 k 帧音频信号在子带 sb_0 中的音调特征参数 $ave_NT_ratio_0$ 和在子带 sb_2 中的音调特征参数 $ave_NT_ratio_2$, 并将该 $ave_NT_ratio_0$ 和 $ave_NT_ratio_2$ 作为第 k 帧音频信号的音调特征参数。

15 在本实施例中, 需要考虑的音调特征参数是在低频子带中和在较高频子带中的音调特征参数, 但本发明的设计方案并不仅限于在本实施例中的这一个, 根据设计的需要, 还可以计算在其它子带中的音调特征参数。

508, 根据上述过程中计算得出的音调特征参数判断当前一帧音频信号的类型。

具体为: 判断 507 中计算得出的在子带 sb_0 中的音调特征参数 $ave_NT_ratio_0$ 、在子带 sb_2 中的音调特征参数 $ave_NT_ratio_2$ 是否与第一参数和第二参数满足一定关系, 该一定关系在本实施例中可如下关系式 (12):

$$(ave_NT_ratio_0 > \alpha) \text{ and } (ave_NT_ratio_2 < \beta)$$

(12)

25 其中, $ave_NT_ratio_0$ 代表第 k 帧音频信号在低频子带中的音调特征参数, $ave_NT_ratio_2$ 代表第 k 帧音频信号在较高频子带中的音调特征参数, α 代表第一系数, β 代表第二系数。

如果满足上述关系式(12), 则判定第 k 帧音频信号为语音类型的音频信号,

否则为音乐类型的音频信号。

下述为当前一帧音频信号进行平滑处理的过程。

509, 对于已判断出音频信号的类型的前一帧音频信号, 再判断该当前一帧音频信号的前一帧音频信号的类型是否与当前音频信号的后一帧音频信号的类型相同, 如果判定为两者相同, 则执行 510, 否则执行 512。

具体为: 判断第 $(k-1)$ 帧音频信号的类型是否与第 $(k+1)$ 帧音频信号的类型相同, 如果判定的结果为第 $(k-1)$ 帧音频信号的类型与第 $(k+1)$ 帧音频信号的类型相同, 则执行 510, 否则执行 512。

510, 判断当前一帧音频信号的类型是否与当前一帧音频信号的前一帧音频信号的类型相同, 如果判定为不相同则执行 511, 否则执行 512。

具体为: 判断第 k 帧音频信号的类型是否与第 $(k-1)$ 帧音频信号的类型相同, 如果判断的结果为第 k 帧音频信号的类型与第 $(k-1)$ 帧音频信号的类型不相同, 则执行 511, 否则执行 512。

511, 将当前一帧音频信号的类型修改为前一帧音频信号的类型。

具体为: 将第 k 帧音频信号的类型修改为第 $(k-1)$ 帧音频信号的类型。

本实施例在所述当前一帧音频信号进行平滑处理的过程中, 具体判断是否需要当前一帧音频信号的类型进行平滑处理时, 采用了需了解前一帧和后一帧音频信号的技术方案, 但该方法属于了解前后帧相关信息的过程, 具体采用了解前几帧和后几帧的方法并不受本实施例所描述的限制。在该过程中具体了解至少前一帧或至少后一帧音频信号的类型方案都适用与本发明的实施例。

512, 结束流程。

在现有技术中, 对音频信号的类型进行分类时需要考虑五种特征参数, 在本实施例所提供的方法中, 通过采用计算音频信号的音调特征参数, 即可判断出大部分的音频信号的类型。与现有技术相比, 分类方法简单, 运算量低。

实施例 2

本实施例公开一种音频信号的分类方法，如图 2 所示，该方法包括：

101，接收到当前一帧的音频信号，该音频信号即为待分类音频信号。

102，获取当前一帧音频信号在至少一个子带中的音调特征参数。

5 一般将频率区域分成 4 个频率子带，在每一个子带中，当前一帧音频信号都可以获取一个对应的音调特征参数。当然，根据设计的需要，也可以选择获取当前一帧音频信号在其中的一个或两个子带中的音调特征参数。

103，获取当前一帧音频信号的频谱倾斜度特征参数。

在本实施例中，102、103 不限定执行的顺序，甚至可以同时执行。

10 104，根据在 102 中获取的至少一个音调特征参数和在 103 中获取的频谱倾斜度特征参数，判断当前一帧音频信号的类型。

本实施例提供的技术方案，通过采取根据音频信号的音调特征参数及频谱倾斜度特征参数来判断音频信号的类型的技术手段，解决了现有技术中，在对音频信号的类型进行分类时需要谐波、噪音及节奏等五种特征参数导致分类方法复杂的技术问题，进而取得了降低在对音频信号进行分类时，分类方法的复
15 杂度，减少分类时的运算量的技术效果。

实施例 3

本实施例公开一种音频信号的分类方法，如图 3 所示，该方法包括如下步骤：

201，接收到当前一帧音频信号，该音频信号即为待分类音频信号。

20 具体为：设采样频率为 48kHz，帧长 $N=1024$ 个样本点，接收到的当前一帧音频信号为第 k 帧音频信号。

下述为计算当前一帧音频信号的音调特征参数的过程。

202，计算该当前一帧音频信号的功率谱密度。

具体为：对第 k 帧音频信号的时域数据进行加汉宁窗的加窗处理。

25 可通过如下汉宁窗的公式计算：

$$h(l) = \sqrt{\frac{8}{3}} \cdot 0.5 \cdot \left[1 - \cos\left(2\pi \cdot \frac{l}{N}\right) \right], \quad 0 \leq l \leq N-1 \quad (1)$$

其中, N 代表帧长, $h(1)$ 代表第 k 帧音频信号的第 1 个样本点的汉宁窗数据。

- 5 对加窗后的第 k 帧音频信号的时域数据进行长度为 N 的 FFT 变换 (因为 FFT 变换是关于 N/2 对称的, 所以实际计算长度为 N/2 的 FFT 变换即可), 并利用 FFT 变换系数计算该第 k 帧音频信号中第 k 个的功率谱密度。

该第 k 帧音频信号中第 k' 个功率谱密度可通过如下计算公式计算:

$$10 \quad X(k') = 10 \cdot \log_{10} \left| \frac{1}{N} \sum_{l=0}^{N-1} \{h(l) \cdot s(l) \cdot e^{[-jk'l \cdot 2\pi/N]}\} \right|^2 = 20 \cdot \log_{10} \left| \frac{1}{N} \sum_{l=0}^{N-1} \{h(l) \cdot s(l) \cdot e^{[-jk'l \cdot 2\pi/N]}\} \right| \text{ dB} \quad (2)$$

$$0 \leq k' \leq N/2, 0 \leq l \leq N-1$$

其中 $s(l)$ 代表第 k 帧音频信号的原始输入样本点, $X(k')$ 代表第 k 帧音频信号中第 k' 个功率谱密度。

- 15 对计算出的功率谱密度 $X(k')$ 进行校正, 使得该功率谱密度的最大值为参考声压级(96dB)。

203, 利用上述功率谱密度检测在频率区域的每个子带中是否有音调的存在, 并统计在对应子带中存在的音调的个数, 将该音调个数作为在该子带中的子带音调个数。

- 20 具体为: 将频率区域划分为四个频率子带, 分别用 sb_0 、 sb_1 、 sb_2 及 sb_3 表示这四个频率子带。如果功率谱密度 $X(k')$ 与相邻的第若干个功率谱密度之间满足一定的条件, 该一定条件在本实施例中可以如下公式 (3) 的所示的条件, 则认为与该 $X(k')$ 对应的子带中含有音调, 并对该音调的个数进行统计, 得出在该子带中的子带音调个数 NT_{k_i} , 该 NT_{k_i} 代表第 k 帧音频信号在子带 sb_i (i 代表子带的编号, 并且 $i=0, 1, 2, 3$) 中的子带音调个数。

$$X(k'-1) < X(k') \leq X(k'+1) \quad \text{and} \quad X(k') - X(k'+j) \geq 7 \text{ dB}$$

(3)

其中, j 的取值规定如下:

$$j = \begin{cases} -2, +2 & \text{for } 2 \leq k' < 63 \\ -3, -2, +2, +3 & \text{for } 63 \leq k' < 127 \\ -6, \Lambda, -2, +2, \Lambda, +6 & \text{for } 127 \leq k' < 255 \\ -12, \Lambda, -2, +2, \Lambda, +12 & \text{for } 255 \leq k' < 500 \end{cases}$$

在本实施例中, 已知功率谱密度的系数个数 (即长度) 为 $N/2$, 对应于上述 j 的取值规定, 对于 k' 值的取值区间的意义进一步说明如下:

sb_0 : 对应 $2 \leq k' < 63$, 对应的功率谱密度系数为第 0 个到第 $(N/16-1)$ 个, 对应的频率范围是 $[0\text{kHz}, 3\text{kHz})$;

10 sb_1 : 对应 $63 \leq k' < 127$, 对应的功率谱密度系数为第 $N/16$ 个到第 $(N/8-1)$ 个, 对应的频率范围是 $[3\text{kHz}, 6\text{kHz})$;

sb_2 : 对应 $127 \leq k' < 255$, 对应的功率谱密度系数为第 $N/8$ 个到第 $(N/4-1)$ 个, 对应的频率范围是 $[6\text{kHz}, 12\text{kHz})$;

15 sb_3 : 对应 $255 \leq k' < 500$, 对应的功率谱密度系数为第 $N/4$ 个到第 $N/2$ 个, 对应的频率范围是 $[12\text{kHz}, 24\text{kHz})$ 。

其中, sb_0 及 sb_1 对应低频子带部分; sb_2 对应较高频子带部分; sb_3 对应高频子带部分。

具体统计 NT_{k_i} 的过程如下:

20 对于子带 sb_0 , 使 k' 在大于等于 2 小于 63 的区间内逐一取值, 对于每一个 k' 的取值, 判断其是否满足公式 (3) 的条件, 在遍历完整个 k' 的取值区间后, 统计满足条件的 k' 的个数, 该满足条件的 k' 个数, 即为第 k 帧音频信号在子带 sb_0 中的存在的子带音调个数 NT_{k_0} 。

例如: 若当 $k'=3$, $k'=5$, $k'=10$ 时, 公式 (3) 成立, 则认为在子带 sb_0 中有 3 个子带音调, 即 $NT_{k_0}=3$ 。

25 同样地, 对于子带 sb_1 , 使 k' 在大于等于 63 小于 127 的区间内逐一取值, 对于每一个 k' 的取值, 判断其是否满足公式 (3) 的条件, 在遍历完整个 k' 的

取值区间后,统计满足条件的 k' 的个数,该满足条件的 k' 个数,即为第 k 帧音频信号在子带 sb_1 中的存在的子带音调个数 NT_{k_1} 。

同样地,对于子带 sb_2 ,使 k' 在大于等于127小于255的区间内逐一取值,对于每一个 k' 的取值,判断其是否满足公式(3)的条件,在遍历完整个 k' 的取值区间后,统计满足条件的 k' 的个数,该满足条件的 k' 个数,即为第 k 帧音频信号在子带 sb_2 中的存在的子带音调个数 NT_{k_2} 。

利用同样的方法,也可统计该第 k 帧音频信号在子带 sb_3 中的存在的子带音调个数 NT_{k_3} 。

204,计算当前一帧音频信号的总音调个数。

10 具体为:根据203统计出的 NT_{k_i} 计算第 k 帧音频信号在四个子带 sb_0 、 sb_1 、 sb_2 及 sb_3 中的子带音调个数之和。

该第 k 帧音频信号在四个子带 sb_0 、 sb_1 、 sb_2 及 sb_3 中的子带音调个数之和即为该第 k 帧音频信号中的音调个数,具体可通过如下公式计算:

$$NT_{k_sum} = \sum_{i=0}^3 NT_{k_i} \quad (4)$$

15

其中, NT_{k_sum} 代表第 k 帧音频信号的总音调个数。

205,计算在规定帧数内当前一帧音频信号在对应子带中的子带音调个数均值。

具体为:设该规定帧数为 M ,在该 M 帧内包括第 k 帧音频信号和第 k 帧的前 $(M-1)$ 帧音频信号,根据 M 的值与 k 的值之间关系计算第 k 帧音频信号在这 M 帧音频信号每个子带中的子带音调个数均值。

该子带音调个数均值具体可通过如下公式(5)计算:

$$ave_NT_i = \begin{cases} \frac{\sum_{j=0}^k NT_{j_i}}{k+1} & \text{if } k < (M-1) \\ \frac{\sum_{j=k-M+1}^k NT_{j_i}}{M} & \text{if } k \geq (M-1) \end{cases} \quad (5)$$

25

其中, NT_{j-i} 代表第 j 帧音频信号在子带 i 中的子带音调个数, ave_NT_i 代表在子带 i 中的子带音调个数均值。特别地, 由公式 (5) 可知, 在计算时需根据 k 的值与 M 的值的 关系选择适当的公式进行计算。

5 特别地, 在本实施例中根据设计的需要, 不必对每个子带都计算子带音调个数均值, 计算在低频子带 sb_0 中的子带音调个数均值 ave_NT_0 , 及在较高频子带 sb_2 中的子带音调个数 ave_NT_2 即可。

206, 计算在规定帧数内当前一帧音频信号总的音调个数均值。

具体为: 设该规定帧数为 M , 在该 M 帧内包括第 k 帧音频信号和第 k 帧
10 的前 $(M-1)$ 帧音频信号, 根据 M 的值与 k 的值之间的关系计算第 k 帧音频信号在这 M 帧的音频信号内平均每帧音频信号包含的总音调个数。

该总音调个数具体可如下公式 (6) 计算:

$$15 \quad ave_NT_{sum} = \begin{cases} \frac{\sum_{j=0}^k NT_{j_sum}}{k+1} & \text{if } k < (M-1) \\ \frac{\sum_{j=k-M+1}^k NT_{j_sum}}{M} & \text{if } k \geq (M-1) \end{cases} \quad (6)$$

其中, NT_{j_sum} 代表第 j 帧总音调个数, ave_NT_{sum} 代表总的音调个数均值。特别地, 由公式 (6) 可知, 在计算时需根据 k 的值与 M 的值的 关系选择适当的公式进行计算。

20 207, 将计算出的在至少一个子带中的子带音调个数均值与总的音调个数均值之比分别作为当前一帧音频信号在对应子带中的音调特征参数。

该音调特征参数具体可通过如下公式 (7) 计算:

$$25 \quad (7) \quad ave_NT_ratio_i = \frac{ave_NT_i}{ave_NT_{sum}}$$

其中, ave_NT_i 代表在子带 i 中的子带音调个数均值, ave_NT_{sum} 代表总

的音调个数均值, ave_NT_ratio_i 代表第 k 帧音频信号在子带 i 中的子带音调个数均值与总的音调个数均值的比值。

特别地, 在本实施例中, 利用 205 计算出来的在低频子带 sb_0 中的子带音调个数均值 ave_NT_0 及在较高频子带 sb_2 中的子带音调个数均值 ave_NT_2 , 通过公式 (7) 可分别计算出第 k 帧音频信号在子带 sb_0 中的音调特征参数 ave_NT_ratio_0 和在子带 sb_2 中的音调特征参数 ave_NT_ratio_2 , 并将该 ave_NT_ratio_0 和 ave_NT_ratio_2 作为第 k 帧音频信号的音调特征参数。

在本实施例中, 需要考虑的音调特征参数是在低频子带中和在较高频子带中的音调特征参数, 但本发明的设计方案并不仅限于在本实施例中的这一个, 根据设计的需要, 还可以计算在其它子带中的音调特征参数。

下述为计算当前一帧音频信号的频谱倾斜度特征参数的过程。

208, 计算一帧音频信号的频谱倾斜度。

具体为: 计算第 k 帧音频信号的频谱倾斜度。

该第 k 帧音频信号的频谱倾斜度可通过如下公式 (8) 计算:

$$\text{spec_tilt}_k = \frac{r(1)}{r(0)} = \frac{\sum_{n=(k-1)N}^{kN-1} [s(n) \cdot s(n-1)]}{\sum_{n=(k-1)N}^{kN-1} [s(n) \cdot s(n)]} \quad (8)$$

其中, $s(n)$ 代表第 k 帧音频信号的第 n 个时域样本点, r 代表自相关参数, spec_tilt_k 代表该第 k 帧音频信号的频谱倾斜度。

209, 根据上述计算出的一帧频谱倾斜度, 计算当前一帧音频信号在规定帧数内的频谱倾斜度均值。

具体为: 设该规定帧数为 M , 在该 M 帧内包括第 k 帧音频信号和第 k 帧的前 $(M-1)$ 帧音频信号, 根据 M 的值与 k 的值之间的关系计算在这 M 帧的音频信号内平均每帧音频信号的频谱倾斜度, 即在这 M 帧的音频信号内的频谱倾斜度均值。

该频谱倾斜度均值具体可通过如下公式 (9) 计算:

$$ave_spec_tilt = \begin{cases} \frac{\sum_{j=0}^k spec_tilt_j}{k+1} & \text{if } k < (M-1) \\ \frac{\sum_{j=k-M+1}^k spec_tilt_j}{M} & \text{if } k \geq (M-1) \end{cases} \quad (9)$$

其中，k 代表当前一帧音频信号的帧号，M 代表规定帧数，spec_tilt_j 代表第 j 帧的音频信号的频谱倾斜度，ave_spec_tilt 为频谱倾斜度均值。特别地，由公式 (9) 可知，在计算时需根据 k 的值与 M 的值的关

210，将至少一个音频信号的频谱倾斜度与上述计算得出的频谱倾斜度均值的均方差作为该当前一帧音频信号的频谱倾斜度特征参数。

具体为：设该规定帧数为 M，在该 M 帧内包括第 k 帧音频信号和第 k 帧的前 (M-1) 帧音频信号，根据 M 的值与 k 的值之间的关系计算至少一个音频信号的频谱倾斜度与频谱倾斜度均值的均方差。该均方差即为当前一帧音频信号的频谱倾斜度特征参数。

该频谱倾斜度特征参数可通过如下公式 (10) 计算：

$$dif_spec_tilt = \begin{cases} \frac{\sum_{j=0}^k [(spec_tilt_j - ave_spec_tilt)^2]}{k+1} & \text{if } k < (M-1) \\ \frac{\sum_{j=k-M+1}^k [(spec_tilt_j - ave_spec_tilt)^2]}{M} & \text{if } k \geq (M-1) \end{cases} \quad (10)$$

其中，k 代表当前一帧音频信号的帧号，ave_spec_tilt 为频谱倾斜度均值，dif_spec_tilt 为频谱倾斜度特征参数。特别地，由公式 (10) 可知，在计算时需根据 k 的值与 M 的值的关

上述实施例中描述的计算音调特征参数的过程 (202 到 207) 和频谱倾斜度特征参数的过程 (208 到 210) 并不限定执行的顺序，甚至可以同时执行。

211，根据上述过程中计算得出的音调特征参数和频谱倾斜度特征参数判

断当前一帧音频信号的类型。

具体为：判断 207 中计算得出的在子带 sb_0 中的音调特征参数 $ave_NT_ratio_0$ 、在子带 sb_2 中的音调特征参数 $ave_NT_ratio_2$ 及在 210 中计算得出的频谱倾斜度特征参数 dif_spec_tilt 是否与第一参数、第二参数和第三参数
5 满足一定关系，该关系在本实施例中可如下关系式 (11)：

$$(ave_NT_ratio_0 > \alpha) \text{ and } (ave_NT_ratio_2 < \beta) \text{ and } (dif_spec_tilt > \gamma) \quad (11)$$

其中， $ave_NT_ratio_0$ 代表第 k 帧音频信号在低频子带中的音调特征参数， $ave_NT_ratio_2$ 代表第 k 帧音频信号在较高频子带中的音调特征参数，
10 dif_spec_tilt 代表第 k 帧音频信号的频谱倾斜度特征参数， α 代表第一系数， β 代表第二系数， γ 代表第三系数。

如果满足所述一定关系，即上述关系式(11)，则判定第 k 帧音频信号为语音类型的音频信号，否则为音乐类型的音频信号。

下述为当前一帧音频信号进行平滑处理的过程。

15 212，对于已判断出音频信号类型的当前一帧音频信号，再判断该当前一帧音频信号的前一帧音频信号的类型是否与当前音频信号的后一帧音频信号的类型相同，如果判定为两者相同，则执行 213，否则执行 215。

具体为：判断第 $(k-1)$ 帧音频信号的类型是否与第 $(k+1)$ 帧音频信号的类型相同，如果判定的结果为第 $(k-1)$ 帧音频信号的类型与第 $(k+1)$ 帧音频
20 信号的类型相同，则执行 213，否则执行 215。

213，判断当前一帧音频信号的类型是否与当前一帧音频信号的前一帧音频信号的类型相同，如果判定为不相同则执行 214，否则执行 215。

具体为：判断第 k 帧音频信号的类型是否与第 $(k-1)$ 帧音频信号的类型相同，如果判断的结果为第 k 帧音频信号的类型与第 $(k-1)$ 帧音频信号的类型
25 不相同，则执行 214，否则执行 215。

214，将当前一帧音频信号的类型修改为前一帧音频信号的类型。

具体为：将第 k 帧音频信号的类型修改为第 $(k-1)$ 帧音频信号的类型。

在本实施例描述的对当前一帧音频信号进程平滑处理的过程中，步骤 212 在判断当前一帧音频信号的类型时，即第 k 帧音频信号的类型时，需要等待第 $(k+1)$ 帧音频信号的类型判断出来后才能进行下一步骤 213，在这里似乎是引入了一帧的延时用于等待判断出第 $(k+1)$ 帧音频信号的类型，但通常编码器算法本身在对每帧音频信号进行编码时均会有一帧的延时，本实施例正好利用了这一帧的延时来实施平滑处理的过程，既可避免对当前一帧音频信号的类型误判，又不会引入额外的延时，取得了可对音频信号进行实时分类的技术效果。

在对于延时上的要求不是很严格的情况下，在本实施例的当前一帧音频信号进行平滑处理的过程中，还可以通过判断当前音频信号的前三帧的类型和后三帧的类型，或者当前音频信号的前五帧的类型和后五帧的类型等来决定是否需要当前音频信号进行平滑处理，具体需要了解的前后相关帧的个数并不受本实施例中所描述的限制。因为多了解一些前后相关信息，这样的平滑处理后的效果可能会更好。

215，流程结束。

与现有技术需要根据五种特征参数来实现对音频信号的类型进行分类相比本实施例提供的音频信号的分类方法根据两种特征参数即可实现对音频信号的类型分类，分类算法简单，复杂度低，降低了分类过程的运算量；同时，本实施例的方案还采用了对分类后的音频信号进行平滑处理的技术手段，取得了可提高对音频信号的类型识别率，使得后续编码过程中能够充分发挥语音编码器及音频编码器作用的有益效果。

实施例 4

对应与上述实施 1，本实施例具体提供一种音频信号的分类装置，如图 4 所示，该装置包括：接收模块 40，音调获取模块 41，分类模块 43，第一判断模块 44，第二判断模块 45，平滑模块 46，第一设定模块 47。

接收模块 40 用于接收当前一帧的音频信号，该当前一帧的音频信号即为待分类音频信号；音调获取模块 41 用于获取待分类音频信号在至少一个子带中的音调特征参数；分类模块 43 用于根据音调获取模块 41 获取的音调特征参数判定所述待分类音频信号的类型；第一判断模块 44 用于在分类模块 43 对待
5 分类音频信号的类型分类后，判断在所述待分类音频信号之前的至少前一帧音频信号的类型是否与在所述待分类音频信号之后对应的至少后一帧音频信号的类型相同；第二判断模块 45 用于当第一判断模块 44 判定与在所述待分类音频信号之后对应的至少后一帧音频信号的类型相同时，判断所述待分类音频信号的类型是否与所述至少前一帧音频信号的类型不同；平滑模块 46 用于当第
10 二判断模块 45 判定与所述至少前一帧音频信号的类型不同时，对所述待分类音频信号的类型进行平滑处理；第一设定模块 47 用于预先设定规定计算的帧数。

在本实施例中，若所述音调获取模块 41 获取的在至少一个子带中的音调特征参数为：在低频子带中的音调特征参数和在较高频子带中的音调特征参
15 数，则所述分类模块 43 包括：判断单元 431，分类单元 432。

判断单元 431 用于判断所述待分类音频信号是否在低频子带中的音调特征参数大于第一系数，并且在较高频子带中的音调特征参数小于第二系数；分类单元 432 用于当判断单元 431 判定所述待分类音频信号在低频子带中的音调特征参数大于第一系数，并且在较高频子带中的音调特征参数小于第二系数
20 时，判定所述待分类音频信号的类型为语音类型，否则为音乐类型。

其中，音调获取模块 41 是根据待分类音频信号在至少一个子带中的音调个数和所述待分类音频信号总的音调个数计算所述音调特征参数的。

进一步地，在本实施例中音调获取模块 41 包括：第一计算单元 411，第二计算单元 412，音调特征单元 413。

25 第一计算单元 411 用于计算待分类音频信号在至少一个子带中的子带音调个数均值；第二计算单元 412 用于计算待分类音频信号总的音调个数均值；

音调特征单元 413 用于将所述在至少一个子带中的子带音调个数均值与所述总的音调个数均值的比值分别作为所述待分类音频信号在对应子带中的音调特征参数。

其中，第一计算单元 411 计算待分类音频信号在至少一个子带中的子带音调个数均值包括：根据第一设定模块 47 设定的规定计算的帧数与待分类音频信号的帧号的关系计算在一个子带中的子带音调个数均值。

第二计算单元 412 计算待分类音频信号总的音调个数均值包括：根据第一设定模块设定的规定计算的帧数与待分类音频信号的帧号的关系计算总的音调个数均值。

10 本实施例提供的音频信号的分类装置通过采用获取音频信号的音调特征参数的技术手段，取得了可判断出大部分音频信号的类型的技术效果，降低了对音频信号的分类过程中分类方法的难度，同时也减少了运算量。

实施例 5

15 对应与上述实施 2 的音频信号的分类方法，本实施例公开一种音频信号的分类装置，如图 5 所示，该装置包括：接收模块 30，音调获取模块 31，频谱倾斜度获取模块 32，分类模块 33。

接收模块 30 用于接收当前一帧的音频信号；音调获取模块 31 用于获取待分类音频信号在至少一个子带中的音调特征参数；频谱倾斜度获取模块 32 用于获取待分类音频信号的频谱倾斜度特征参数；分类模块 33 用于根据音调获取模块 31 获取的所述音调特征参数和频谱倾斜度获取模块 32 获取的频谱倾斜度特征参数确定所述待分类音频信号的类型。

25 在现有技术中，在对音频信号进行分类时的需要参考音频信号的多方面的特征参数，使得分类复杂度高、运算量大，而本实施例提供的方案在对音频信号进行分类时，根据该音频信号的音调和频谱倾斜度两种特征参数即可分辨出该音频信号的类型，使对音频信号的分类变得简单，同时也减少了在分类过程中的运算量。

实施例 6

本实施例具体提供一种音频信号的分类装置，如图 6 所示，该装置包括：接收模块 40，音调获取模块 41，频谱倾斜度获取模块 42，分类模块 43，第一判断模块 44，第二判断模块 45，平滑模块 46，第一设定模块 47 和第二设定模块 48。

5 接收模块 40 用于接收当前一帧的音频信号，该当前一帧的音频信号即为待分类音频信号；音调获取模块 41 用于获取待分类音频信号在至少一个子带中的音调特征参数；频谱倾斜度获取模块 42 用于获取待分类音频信号的频谱倾斜度特征参数；分类模块 43 用于根据音调获取模块 41 获取的所述音调特征参数和频谱倾斜度获取模块 42 获取的频谱倾斜度特征参数判断所述待分类音频信号的类型；第一判断模块 44 用于在分类模块 43 对待分类音频信号的类型分类后，判断在所述待分类音频信号之前的至少前一帧音频信号的类型是否与在所述待分类音频信号之后对应的至少后一帧音频信号的类型相同；第二判断模块 45 用于当第一判断模块 44 判定与在所述待分类音频信号之后对应的至少后一帧音频信号的类型相同时，判断所述待分类音频信号的类型是否与所述至少前一帧音频信号的类型不同；平滑模块 46 用于当第二判断模块 45 判定与所述至少前一帧音频信号的类型不同时，对所述待分类音频信号的类型进行平滑处理；第一设定模块 47 用于预先设定计算音调特征参数时需要规定计算的帧数；第二设定模块 48 用于预先设定计算频谱倾斜度特征参数时需要规定计算的帧数。

20 其中，音调获取模块 41 是根据待分类音频信号在至少一个子带中的音调个数和所述待分类音频信号总的音调个数计算所述音调特征参数的。

在本实施例中，若所述音调获取模块 41 获取的在至少一个子带中的音调特征参数为：在低频子带中的音调特征参数和在较高频子带中的音调特征参数，则所述分类模块 43 包括：判断单元 431，分类单元 432。

25 判断单元 431 用于当所述待分类音频信号在低频子带中的音调特征参数大于第一系数，并且在较高频子带中的音调特征参数小于第二系数时，判断所

述音频信号的频谱倾斜度特征参数是否大于第三系数;分类单元 432 用于当判断单元判定所述待分类音频信号的频谱倾斜度特征参数大于第三系数时,判定所述待分类音频信号的类型为语音类型,否则为音乐类型。

进一步的,在本实施例中音调获取模块 41 包括:第一计算单元 411,第二计算单元 412,音调特征单元 413。

第一计算单元 411 用于计算待分类音频信号在至少一个子带中的子带音调个数均值;第二计算单元 412 用于计算待分类音频信号总的音调个数均值;音调特征单元 413 用于将所述在至少一个子带中的子带音调个数均值与所述的音调个数均值的比值分别作为所述待分类音频信号在对应子带中的音调特征参数。

其中,第一计算单元 411 计算待分类音频信号在至少一个子带中的子带音调个数均值根据第一设定模块 47 设定的规定计算的帧数与待分类音频信号的帧号的关系计算在一个子带中的子带音调个数均值。

第二计算单元 412 计算待分类音频信号总的音调个数均值包括:根据第一设定模块 47 设定的规定计算的帧数与待分类音频信号的帧号的关系计算总的音调个数均值。

进一步的,本实施中,频谱倾斜度获取模块 42 包括:第三计算单元 421,频谱倾斜度特征单元 422。

第三计算单元 421 用于计算待分类音频信号的频谱倾斜度均值;频谱倾斜度特征单元 422 用于将至少一个音频信号的频谱倾斜度与所述频谱倾斜度均值的均方差作为所述待分类音频信号的频谱倾斜度特征参数。

其中,第三计算单元 421 计算待分类音频信号的频谱倾斜度均值包括:根据第二设定模块 48 设定的规定计算的帧数与待分类音频信号的帧号的关系计算频谱倾斜度均值。

频谱倾斜度特征单元 422 计算至少一个音频信号的频谱倾斜度与所述频谱倾斜度均值的均方差包括:根据第二设定模块 48 设定的规定计算的帧数与

待分类音频信号的帧号的关系计算频谱倾斜度特征参数。

本实施例中的第一设定模块 47 和第二设定模块 48 可以通过一个程序或者模块实现，甚至也可以设定相同的规定计算的帧数的值。

本实施例提供的方案具有如下有益效果：分类简单，复杂度低，运算量小，
5 不为编码器引入额外的延时，可满足中低码率下的语音音频编码器在分类过程中要求的实时编码、低复杂度的需求。

本发明实施例主要运用于通信技术领域，实现对音频信号的类型进行快速、准确并实时的分类。随着网络技术的发展有可能应用到本领域的其它场景，也有可能转用到类似或者相近的技术领域上去。

10 通过以上的实施方式的描述，所属领域的技术人员可以清楚地了解到本发明可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在可读取的存储介质中，如计算机的软盘，硬盘或光盘等，包
15 括若干指令用以使得一台编码器执行本发明各个实施例所述的方法。

以上所述，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应以所述权利要求的保护范围为准。

20

权 利 要 求

1、一种音频信号的分类方法，其特征在于，包括：

获取待分类音频信号在至少一个子带中的音调特征参数；

5 根据获取的所述音调特征参数判定所述待分类音频信号的类型。

2、根据权利要求 1 所述的音频信号的分类方法，其特征在于，该方法还包括：

获取待分类音频信号的频谱倾斜度特征参数；

10 根据获取的所述频谱倾斜度特征参数确定所述判定的待分类音频信号的类型。

3、根据权利要求 1 所述的音频信号的分类方法，其特征在于，若所述在至少一个子带中的音调特征参数为：在低频子带中的音调特征参数和在较高频子带中的音调特征参数，则所述根据获取的特征参数判定所述待分类音频信号的类型包括：

15 判断所述待分类音频信号在低频子带中的音调特征参数是否大于第一系数，并且在较高频子带中的音调特征参数小于第二系数；

如果所述待分类音频信号在低频子带中的音调特征参数大于第一系数，并且在较高频子带中的音调特征参数小于第二系数，则所述待分类音频信号的类型为语音类型，否则为音乐类型。

20 4、根据权利要求 2 所述的音频信号的分类方法，其特征在于，若所述在至少一个子带中的音调特征参数为：在低频子带中的音调特征参数和在较高频子带中的音调特征参数，则所述根据获取的所述频谱倾斜度特征参数确定所述判定的待分类音频信号的类型包括：

25 当所述待分类音频信号在低频子带中的音调特征参数大于第一系数，并且在较高频子带中的音调特征参数小于第二系数时，判断所述待分类音频信号的频谱倾斜度特征参数是否大于第三系数；

如果所述待分类音频信号的频谱倾斜度特征参数大于第三系数，则所述待

分类音频信号的类型为语音类型，否则为音乐类型。

5、根据权利要求 1 所述的音频信号的分类方法，其特征在于，所述获取待分类音频信号在至少一个子带中的音调特征参数为：

5 根据待分类音频信号在至少一个子带中的音调个数和所述待分类音频信号总的音调个数计算所述音调特征参数。

6、根据权利要求 5 所述的音频信号的分类方法，其特征在于，所述根据待分类音频信号在至少一个子带中的音调个数和所述待分类音频信号总的音调个数计算所述音调特征参数包括：

10 计算待分类音频信号在至少一个子带中的子带音调个数均值；

计算待分类音频信号总的音调个数均值；

将所述在至少一个子带中的子带音调个数均值与所述总的音调个数均值的比值分别作为所述待分类音频信号在对应子带中的音调特征参数。

7、根据权利要求 6 所述的音频信号的分类方法，其特征在于，

15 预先设定规定计算的帧数；所述计算待分类音频信号在至少一个子带中的子带音调个数均值包括：

根据规定计算的帧数与待分类音频信号的帧号的关系计算在一个子带中的子带音调个数均值。

8、根据权利要求 6 所述的音频信号的分类方法，其特征在于，预先设定规定计算的帧数；所述计算待分类音频信号总的音调个数均值包括：

20 根据规定计算的帧数与待分类音频信号的帧号的关系计算总的音调个数均值。

9、根据权利要求 2 所述的音频信号的分类方法，其特征在于，所述获取所述待分类音频信号的频谱倾斜度特征参数包括：

计算待分类音频信号的频谱倾斜度均值；

25 将至少一个音频信号的频谱倾斜度与所述频谱倾斜度均值的均方差作为所述待分类音频信号的频谱倾斜度特征参数。

10、根据权利要求 9 所述的音频信号的分类方法，其特征在于，

预先设定规定计算的帧数；所述计算待分类音频信号的频谱倾斜度均值包括：根据规定计算的帧数与待分类音频信号的帧号的关系计算频谱倾斜度均值。

5 11、根据权利要求 9 所述的音频信号的分类方法，其特征在于，

预先设定规定计算的帧数；所述至少一个音频信号的频谱倾斜度与所述频谱倾斜度均值的均方差包括：根据规定计算的帧数与待分类音频信号的帧号的关系计算频谱倾斜度特征参数。

12、一种音频信号的分类装置，其特征在于，包括：

10 音调获取模块，用于获取待分类音频信号在至少一个子带中的音调特征参数；

分类模块，用于根据获取的所述音调特征参数判定所述待分类音频信号的类型。

15 13、根据权利要求 12 所述的音频信号的分类装置，其特征在于，该装置还包括：

频谱倾斜度获取模块，用于获取待分类音频信号的频谱倾斜度特征参数；

则所述分类模块还用于根据所述频谱倾斜度获取模块获取的频谱倾斜度特征参数确定所述判定的待分类音频信号的类型。

20 14、根据权利要求 12 所述的音频信号的分类装置，其特征在于，当所述音调获取模块获取的在至少一个子带中的音调特征参数为：在低频子带中的音调特征参数和在较高频子带中的音调特征参数时，所述分类模块包括：

判断单元，用于判断所述待分类音频信号是否在低频子带中的音调特征参数大于第一系数，并且在较高频子带中的音调特征参数小于第二系数：

25 分类单元，用于当判断单元判定所述待分类音频信号在低频子带中的音调特征参数大于第一系数，并且在较高频子带中的音调特征参数小于第二系数时，判定所述待分类音频信号的类型为语音类型，否则为音乐类型。

15、根据权利要求 13 所述的音频信号的分类装置，其特征在于，当所述音调获取模块获取的在至少一个子带中的音调特征参数为：在低频子带中的音调特征参数和在较高频子带中的音调特征参数时，所述分类模块包括的

判断单元还用于当所述待分类音频信号在低频子带中的音调特征参数大于第一系数，并且在高频子带中的音调特征参数小于第二系数时，判断所述音频信号的频谱倾斜度特征参数是否大于第三系数；

分类单元还用于当判断单元判定所述待分类音频信号的频谱倾斜度特征参数大于第三系数时，判定所述待分类音频信号的类型为语音类型，否则为音乐类型。

10 16、根据权利要求 12 所述的音频信号的分类装置，其特征在于，所述音调获取模块根据待分类音频信号在至少一个子带中的音调个数和所述待分类音频信号总的音调个数计算所述音调特征参数。

17、根据权利要求 12 或 16 所述的音频信号的分类装置，其特征在于，所述音调获取模块包括：

15 第一计算单元，用于计算待分类音频信号在至少一个子带中的子带音调个数均值；

第二计算单元，用于计算待分类音频信号总的音调个数均值；

20 音调特征单元，用于将所述在至少一个子带中的子带音调个数均值与所述总的音调个数均值的比值分别作为所述待分类音频信号在对应子带中的音调特征参数。

18、根据权利要求 17 所述的音频信号的分类装置，其特征在于，该装置还包括：

第一设定模块，用于预先设定规定计算的帧数；

25 所述第一计算单元计算待分类音频信号在至少一个子带中的子带音调个数均值包括：根据第一设定模块设定的规定计算的帧数与待分类音频信号的帧数的关系计算在一个子带中的子带音调个数均值。

19、根据权利要求 17 所述的音频信号的分类装置，其特征在于，该装置还包括：

第一设定模块，用于预先设定规定计算的帧数；

5 所述第二计算单元计算待分类音频信号总的音调个数均值包括：根据第一设定模块设定的规定计算的帧数与待分类音频信号的帧号的关系计算总的音调个数均值。

20、根据权利要求 12 所述的音频信号的分类装置，其特征在于，所述频谱倾斜度获取模块包括：

第三计算单元，用于计算待分类音频信号的频谱倾斜度均值；

10 频谱倾斜度特征单元，用于将至少一个音频信号的频谱倾斜度与所述频谱倾斜度均值的均方差作为所述待分类音频信号的频谱倾斜度特征参数。

21、根据权利要求 20 所述的音频信号的分类装置，其特征在于，该装置还包括：

第二设定模块，用于预先设定规定计算的帧数；

15 所述第三计算单元计算待分类音频信号的频谱倾斜度均值包括：根据第二设定模块设定的规定计算的帧数与待分类音频信号的帧号的关系计算频谱倾斜度均值。

22、根据权利要求 20 所述的音频信号的分类装置，其特征在于，该装置还包括：

20 第二设定模块，用于预先设定规定计算的帧数；

所述频谱倾斜度特征单元计算至少一个音频信号的频谱倾斜度与所述频谱倾斜度均值的均方差包括：根据第二设定模块设定的规定计算的帧数与待分类音频信号的帧号的关系计算频谱倾斜度特征参数。