

音频信号底鼓节拍点的检测方法以及终端

5 技术领域

本发明涉及多媒体信息技术领域，具体而言，本发明涉及一种音频信号底鼓节拍点的检测方法以及终端。

背景技术

10 底鼓又称地鼓、低音鼓，是架子鼓中脚踏的低音大鼓。带有底鼓演奏的音乐中其底鼓的节拍点往往具有较强的节奏性。因此将底鼓的节拍点检测出来，以应用于用户所需要的各个场景中，具备重要的意义。通常音乐中包含了多种乐器混合演奏，因此难以直接检测出底鼓的节拍点。传统技术中，一般需要人工对每一首音乐中的底鼓的节拍点进行检测，效率较低。

15

发明内容

本发明针对现有方式的缺点，提出一种音频信号底鼓节拍点的检测方法以及终端，用以解决现有技术中存在的底鼓的节拍点检测效率低的问题，以提高底鼓的节拍点检测效率。

20

本发明的实施例根据第一个方面，提供了一种音频信号底鼓节拍点的检测方法，包括步骤：

根据输入的待检测音频信号获得若干个本征模函数；

计算若干个本征模函数对应的瞬时强度信号以及瞬时频率信号；

根据若干个本征模函数对应的瞬时强度信号和瞬时频率信号，获得底鼓的特征

25 信号；

对所述特征信号进行峰值检测，获得若干个峰值点；

根据若干个峰值点获得底鼓的节拍点。

本实施例提供的音频信号底鼓节拍点的检测方法，利用本征模函数进行底鼓特征信号的提取，通过对特征信号进行峰值检测获取到峰值点，该峰值点即为音乐中底鼓被敲击的时间点，根据峰值点即可以得到其节拍点，实现了底鼓节拍点的自动
30 化获取，效率较高。

在一个实施例中，所述对所述特征信号进行峰值检测，获得若干个峰值点，包括：

对所述特征信号进行峰值检测，获取各个极大值点；

从各个极大值点中选取满足预设条件的极大值点，将选取的极大值点判定为峰
5 值点；

其中，所述预设条件包括：两个连续的极大值点之间的特征信号中任何一个点均不是极大值点，且两个连续的极大值点之间的特征信号中最小值远小于该两个连续的极大值点。

本实施例充分结合底鼓特征信号的性状以及底鼓本身的声学特性，设计出了一套独特的预设条件以用于峰值检测，从而最大限度的保证了底鼓的检测准确度，降低误判的概率。
10

在一个实施例中，所述将选取的极大值点判定为峰值点之后，还包括：

计算由每个峰值点与其邻近点所构成的信号峰经过高斯拟合后的半高全宽；

若半高全宽小于预设门限，将对应的峰值点保留，否则将对应的峰值点剔除。

15 在一个实施例中，所述根据若干个本征模函数对应的瞬时强度信号和瞬时频率信号，获得底鼓的特征信号，包括：

将瞬时强度信号平方后与瞬时频率信号相乘，得到每个本征模函数的等效瞬时频率；

对所有本征模函数的等效瞬时频率求和，获得所述底鼓的特征信号。

20 在一个实施例中，所述获得所述底鼓的特征信号之后，所述对所述特征信号进行峰值检测，获得若干个峰值点之前，还包括：

获取所述特征信号的所有谷值点；

计算由每个谷值点与其最近的两个峰值点所构成的信号谷的半高全宽；

获取半高全宽小于预设第一阈值的信号谷，利用该信号谷邻近的特征信号通过
25 插值的方式将该信号谷去除；

将预设第一阈值更新为预设第二阈值，返回获取半高全宽小于预设第一阈值的信号谷的步骤，直至得到平滑的特征信号。

本实施例通过对特征信号进行平滑，有效的降低了模态混叠对检测结果的影响，提高了检测的准确度。

30 在一个实施例中，所述对所述特征信号进行峰值检测，获得若干个峰值点之后，所述根据若干个峰值点获得底鼓的节拍点之前，还包括：

从所述特征信号中每个峰值点所指示位置的邻近区域内寻找最大值，将寻找到的最大值作为对齐后的峰值点。

本实施例通过寻找最大值对齐峰值点的操作，进一步提高了检测的准确度。

在一个实施例中，所述对所述特征信号进行峰值检测，获得若干个峰值点之后，
5 所述根据若干个峰值点获得底鼓的节拍点之前，还包括：

从所述特征信号中每个峰值点所指示位置的邻近区域内，统计所述邻近区域内的特征信号数值超过对应峰值点对应的特征信号数值预设比例的数量，若所述数量超过预设门限值，将对应的峰值点剔除；

和/或，

10 当两个连续的峰值点之间的间隔小于预设的间隔门限时，将对应的特征信号数值低的峰值点剔除；

和/或，

当一个峰值点对应的特征信号数值均低于其他峰值点对应的特征信号数值时，计算该个峰值点相较于其邻域区域内的特征信号的凸显度，若所述凸显度小于预设
15 阈值，将该个峰值点剔除。

本实施例通过根据各个条件进一步剔除峰值点的操作，进一步提高了检测的准确度。

在一个实施例中，所述根据若干个峰值点获得底鼓的节拍点之后，还包括：

在所述底鼓的节拍点所在的位置添加预设的音视频特效。

20 本实施例通过在底鼓的节拍点上添加特定的音视频特效，能够使得相应的特效效果与音乐本身更加的贴合，从而达到较好的产品效果，且能够在产品上实时呈现。

本发明的实施例根据第二个方面，还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现前述任意一项所述的音频信号底鼓节拍点的检测方法。

25 本实施例提供的计算机可读存储介质，利用本征模函数进行底鼓特征信号的提取，通过对特征信号进行峰值检测获取到峰值点，该峰值点即为音乐中底鼓被敲击的时间点，根据峰值点即可以得到其节拍点，实现了底鼓节拍点的自动化获取，效率较高。

本发明的实施例根据第三个方面，还提供了一种终端，所述终端包括：

30 一个或多个处理器；

存储装置，用于存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现前述任意一项所述的音频信号底鼓节拍点的检测方法。

本实施例提供的终端，利用本征模函数进行底鼓特征信号的提取，通过对特征信号进行峰值检测获取到峰值点，该峰值点即为音乐中底鼓被敲击的时间点，根据
5 峰值点即可以得到其节拍点，实现了底鼓节拍点的自动化获取，效率较高。

本发明附加的方面和优点将在下面的描述中部分给出，这些将从下面的描述中变得明显，或通过本发明的实践了解到。

附图说明

10 本发明上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解，其中：

图 1 为本发明一个实施例的音频信号底鼓节拍点的检测方法的流程示意图；

图 2 为本发明一实施例的采用经验模式分解获得若干个本征模函数的流程示意图；

15 图 3 为本发明一具体实施例的音频信号底鼓节拍点的检测方法的流程示意图；

图 4 为本发明一具体实施例的终端的结构示意图。

具体实施方式

下面详细描述本发明的实施例，所述实施例的示例在附图中示出，其中自始至终
20 终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，仅用于解释本发明，而不能解释为对本发明的限制。

本技术领域技术人员可以理解，除非特意声明，这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是，本发明的说明书中使用的措辞“包括”是指存在所述特征、整数、步骤、操作、元件和/或组件，但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、组件和/或它们的组。这里使用的措辞“和/或”包括一个或更多个相关联的列出项的全部或任一单元和全部组合。

本技术领域技术人员可以理解，除非另外定义，这里使用的所有术语（包括技
30 术术语和科学术语），具有与本发明所属领域中的普通技术人员的一般理解相同的意义。还应该理解的是，诸如通用字典中定义的那些术语，应该被理解为具有与现

有技术的上下文中的意义一致的意义，并且除非像这里一样被特定定义，否则不会用理想化或过于正式的含义来解释。

本技术领域技术人员可以理解，这里所使用的“终端”既包括无线信号接收器的设备，其仅具备无发射能力的无线信号接收器的设备，又包括接收和发射硬件的设备，其具有能够在双向通信链路上，执行双向通信的接收和发射硬件的设备。这种设备可以包括：蜂窝或其他通信设备，其具有单线路显示器或多线路显示器或没有多线路显示器的蜂窝或其他通信设备；PCS（Personal Communications Service，个人通信系统），其可以组合语音、数据处理、传真和/或数据通信能力；PDA（Personal Digital Assistant，个人数字助理），其可以包括射频接收器、寻呼机、互联网/内联网访问、网络浏览器、记事本、日历和/或GPS（Global Positioning System，全球定位系统）接收器；常规膝上型和/或掌上型计算机或其他设备，其具有和/或包括射频接收器的常规膝上型和/或掌上型计算机或其他设备。这里所使用的“终端”可以是便携式、可运输、安装在交通工具（航空、海运和/或陆地）中的，或者适合于和/或配置为在本地运行，和/或以分布形式，运行在地球和/或空间的任何其他位置运行。这里所使用的“终端”还可以是通信终端、上网终端、音乐/视频播放终端，例如可以是PDA、MID（Mobile Internet Device，移动互联网设备）和/或具有音乐/视频播放功能的移动电话，也可以是智能电视、机顶盒等设备。

有必要先对本发明的技术构思进行如下的先导性说明。

本发明实施例提供的音频信号底鼓节拍点的检测方法以及终端，首先从音频信号中针对底鼓的声学特性提取出底鼓的特征信息，然后利用底鼓特征信息计算出峰值点，峰值点即为音乐中底鼓敲击事件发生的准确时间点，进而根据峰值点得到节拍点信息，以用于用户所需要的各种场景，例如音视频特效添加等。

下面结合附图对本发明的具体实施方式进行详细介绍。

如图1所示，为一实施例的音频信号底鼓节拍点的检测方法的流程示意图，该检测方法包括步骤：

S110、根据输入的待检测音频信号获得若干个本征模函数。

由于本发明要对底鼓节拍点进行检测，所以待检测音频信号一般为包括底鼓演奏的音频信号。用户可以通过选择音乐库中的音乐或者自己上传的音乐的方式输入待检测音频信号。

为了满足用户多样化的需求，例如仅需要对某一些音频信号进行检测，而对另一些音频信号不需要进行检测，可选的，首先判断是否需要对输入的音乐实行节拍

点检测，对于需要实行节拍点检测的音乐，调用本发明实施例提供的方法对该音乐进行底鼓节拍点检测，否则按照常规的操作方法执行。在具体实现时，可以设置弹窗显示是否对输入的音乐进行底鼓节拍点检测，然后根据用户触发的相应功能选项确定是否执行本发明实施例所提供的方法。

- 5 通过希尔伯特变换的方式所定义的瞬时频率值在某些情况下不具备明确的物理意义，研究表明，只有满足特定条件的信号才具备具有物理意义的瞬时频率，称这类信号为本征模函数（IMF，Intrinsic Mode Function），在此基础上创建的一套信号自适应分解得到本征模函数的方法即为经验模式分解（EMD，Empirical Mode Decomposition）。其中，瞬时频率为：对任意的时间序列，通过希尔伯特变换的方式可以唯一的得到其复解析信号，定义该复解析信号的相位随时间的变化率为瞬时频率。

S120、计算若干个本征模函数对应的瞬时强度信号以及瞬时频率信号。

对于每一个本征模函数，计算对应的瞬时强度信号和瞬时频率信号，即可以得到所有本征模函数对应的瞬时强度信号以及瞬时频率信号。

- 15 S130、根据若干个本征模函数对应的瞬时强度信号和瞬时频率信号，获得底鼓的特征信号。

特征信号用于表征底鼓所独有的不同于其它乐器或者人物声音的特征，获取所有本征模函数对应的瞬时强度信号以及瞬时频率信号后，就可以计算出底鼓的特征信号。

- 20 S140、对所述特征信号进行峰值检测，获得若干个峰值点。

峰值检测用于检测出特征信号的峰值点，每一个峰值点均代表了底鼓被敲击的一个时间点，即用户敲击底鼓的时间点。

S150、根据若干个峰值点获得底鼓的节拍点。

- 25 获得峰值点，即该音乐中所有底鼓被敲击发生的具体时间点，而后，再利用得到的时间点进行进一步的音乐节奏信息分析，得到最终的节拍点信息。其中，根据时间点进行音乐节奏信息分析从而得到节拍点信息可以根据现有技术中已有的方式实现。

- 30 上述实施例利用本征模函数进行底鼓特征信号的提取，通过对特征信号进行峰值检测获取到峰值点，即音乐中底鼓被敲击的时间点，根据底鼓被敲击的时间点即可以得到其节拍点，实现了底鼓的节拍点的自动化获取，效率较高。

输入待检测音频信号之后，根据待检测音频信号获得若干个本征模函数之前，

可选的，还包括对待检测音频信号进行预处理的步骤。预处理的方式有很多，下面结合一个具体实施例进行介绍，应当理解，本发明并不限制于下述预处理的方式，用户还可以根据需要采取其他预处理操作。

具体的，所述对待检测音频信号进行预处理，包括：

5 S1101、将待检测音频信号以设定的采样率进行重采样。重采样能够降低输入的信号量，从而大幅降低本发明方法运算所消耗的时间，使本发明方法能够在一个可接受的时间范围内给出处理结果，以便后续使用。本发明的发明人经过反复试验与分析发现，当采样率为 2kHz（千赫兹）时达到较好的效果。

10 S1102、对重采样后的待检测音频信号进行低通滤波。本发明的发明人经过反复试验与分析发现，使用滤波器为 8 阶巴特沃斯低通滤波器（截止频率 150Hz），能够有效的减少待检测音频信号所包含的不同乐器、人声演唱等干扰成分，同时最大限度的保留下底鼓的成分，从而使后续的特征提取更加准确。

经验模式分解是希尔伯特变换中的重要步骤，如图 2 所示，为采用经验模式分解获得若干个本征模函数的流程示意图，具体包括步骤：

15 S1105、对输入的待检测音频信号进行峰谷检测，分别得到峰值序列与谷值序列；

S1106、对峰值序列与谷值序列分别进行三次样条插值，得到待检测音频信号的上包络线（峰值线）与下包络线（谷值线）；

S1107、将上下包络线相加后平均，得到均值线；

20 S1108、将待检测音频信号减去均值线得到信号的无偏高频分量；

S1109、判断得到的无偏高频分量是否满足本征条件，如果满足，则将这一信号记录为一个本征模式，否则的话将得到的无偏高频分量设为输入信号重新进行 S1105- S1108 步骤，得到新的无偏高频分量；

25 可选的，本征条件的判定准则为：对无偏高频成分而言，其极值点的数量与过零点的数量相差不超过 1 个，或连续两次迭代的无偏高频成分之间的标准差小于设定的大小，或者连续迭代次数超过设定的次数。这里标准差的定义为：

$$S_d = \sum_{t=0}^T \frac{\|h_{k-1}(t) - h_k(t)\|^2}{h_k^2(t)}$$

其中 $h_k(t)$ 即为第 k 次迭代得到的无偏高频成分。

30 S1110、将输入的待检测音频信号减去得到的本征模式信号得到余量信号，判断余量信号是否满足结束判定，若满足，得到余量模式，若否，将余量信号设为的

待检测音频信号，重新进行 S1105- S1109 步骤得到下一个本征模式信号；

可选的，结束判定的判定准则为：当余量信号的所有数值的绝对值均小于某一门限值，或者其经过峰谷检测得到的峰值序列或谷值序列的数量少于设定的门限。

最终经验模式分解将把输入的待检测音频信号分解为若干个本征模式信号以及一个余量模式信号，我们称这些信号为输入的待检测音频信号的本征模函数。

经验模式分解有两个固有的问题，其一为端点效应，其二则为模态混叠。其中，模态混叠为：当两组强度相当且频率相差很小的谐波相互叠加时，通过经验模式分解无法完全的将这两个谐波分量分离出来，分解出的信号存在有模态混叠的现象。对于存在混叠的本征模函数，其瞬时频率不再具有准确的物理意义，会导致最终提取的底鼓特征存在偏差。

由于前述的低通滤波器，以及后述的特征平滑的作用，模态混叠能够得到有效的抑制。而针对端点效应所导致的误差，可选的，本发明采用周期延拓的方式进行抑制，具体过程为：

S1103、选取端点处的一段特定长度的信号，并在端点附近一定范围内寻找与之最相近的信号；

S1104、利用找到的信号的前续信号对原先的端点处进行信号延拓；

S1105、利用延拓后的信号来进行峰谷检测，以得到更加准确的峰值序列和谷值序列。

通过上述方案，可有效降低端点对经验模式分解造成的误差。

在一个实施例中，计算若干个本征模函数对应的瞬时强度信号以及瞬时频率信号包括：

S1201、对前述计算得到的所有本征模式函数 Imf_i 进行希尔伯特变换，得到对应的复解析信号 H_i ；

S1202、对每一条复解析信号 H_i ，计算瞬时强度信号 $A_i = \sqrt{R_i^2 + I_i^2}$ 和瞬时相位信号 $\Phi_i = \tan^{-1}(I_i/R_i)$ ，其中 R_i 和 I_i 分别为 H_i 的实部和虚部；

S1203、对每一条复解析信号 H_i ，计算瞬时频率信号 $\omega_i = (\Phi_i - \Phi_{i-1})/\Delta t$ 。

特别的，由于 Φ_i 的数值经过了 $[0, 2\pi]$ 的取模，因此需要对 ω_i 进行一定的调整以消除取模导致的突变，具体而言，当 ω_i 小于某一负数值时将其加上一个正偏移量，而当 ω_i 大于某一数值时则将其加上一个负偏移量。

在一个实施例中，所述根据若干个本征模函数对应的瞬时强度信号和瞬时频率信号，获得底鼓的特征信号，包括：将瞬时强度信号平方后与瞬时频率信号相乘，

得到每个本征模函数的等效瞬时频率；对所有本征模函数的等效瞬时频率求和，获得所述底鼓的特征信号。

上述实施例中，对每一条本征模函数分别计算 A_i 与 ω_i ，并最终计算特征信号 $\zeta = \sum A_i^2 \omega_i$ 。采用该种方式计算特征信号，能够最大限度的突出底鼓信号的特征。

5 通过前述的方式计算得到的特征信号在底鼓被敲击处会呈现明显的峰值特性，因而可以通过对特征信号进行峰值检测的方式获取到底鼓被敲击的准确时间点。在一个实施例中，所述对所述特征信号进行峰值检测，获得若干个峰值点，包括：对所述特征信号进行峰值检测，获取各个极大值点；从各个极大值点中选取满足预设条件的极大值点，将选取的极大值点判定为峰值点；其中，所述预设条件包括：两
10 个连续的极大值点之间的特征信号中任何一个点均不是极大值点，且两个连续的极大值点之间的特征信号中最小值远小于该两个连续的极大值点。

上述峰值检测的实施例实现的是条件检峰，条件检峰指的是当且仅当特征信号的某个极大值点满足预设条件时，才将这一极大值点判定为峰值点。其中，远小于定义为最小值与该两个连续的极大值点的比值均小于一个设定的比值门限，或其
15 最小值与该两个连续的极大值点的差值均大于一个设定的差值门限。

为了进一步提高检测结果的准确性，还需要对峰值点进行二次筛选，因此，在一个实施例中，所述将选取的极大值点判定为峰值点之后，还包括：计算由每个峰值点与其邻近点所构成的信号峰经过高斯拟合后的半高全宽；若半高全宽小于预设
20 门限，将对应的峰值点保留，否则将对应的峰值点剔除。

上述实施例中：一个峰值点的邻近点指的是该个峰值点附近的信号点，也即是与峰值点的差值小于预设阈值的信号点。对于每一个峰值点而言，该个峰值点与其附近的信号点构成一个信号峰。半高全宽为在信号的一个峰当中，前后两个信号值等于该峰值一半的点之间的距离，通常用于表征信号峰的持续时间。对任何一个峰值点而言，其与其附近的信号点所构成的信号峰经过高斯拟合后的半高全宽应小于
25 某一门限，如果不小于则将该峰值点删除。

上述实施例充分结合了经验模式分解所获得的底鼓特征信号的性状以及底鼓本身的声学特性，而设计出了一套独特的检峰判决条件（即预设条件），从而最大限度的保证了底鼓的检测准确度，降低误判的概率。

虽然前述使用的低通滤波器有效的降低了经验模式分解的模态混叠的影响，但
30 依然会有少量的干扰残留，具体表现为计算得到的特征信号时常会有轻微的上下抖动，对于底鼓强度足够的地方，这一抖动不会对结果产生太大的干扰，但对于一些

强度不足的底鼓点，以及诸如强低音贝斯处的干扰点，这一抖动均会对检测结果产生影响，使得最终的准确度下降。为解决这一问题，还需要对所述特征信号进行平滑。因此，在一个实施例中，所述获得所述底鼓的特征信号之后，所述对所述特征信号进行峰值检测，获得若干个峰值点之前，还包括：

5 S131、获取所述特征信号的所有谷值点。

谷值点为特征信号的极小值点，获取谷值点可以根据现有技术中已有的方式实现。

S132、计算由每个谷值点与其最近的两个峰值点所构成的信号谷的半高全宽。

对于每一个谷值点，该个谷值点与其最近的两个峰值点构成一个信号谷。计算
10 每一个信号谷的半高全宽。

S133、获取半高全宽小于预设第一阈值的信号谷，利用该信号谷邻近的特征信号通过插值的方式将该信号谷去除。

信号谷邻近的特征信号指的是与信号谷的距离小于预设阈值的特征信号。当某一信号谷的半高全宽小于设定的阈值时，利用该信号谷附近的特征信号通过插值的方式将这一信号谷抹去。也即是，对该信号谷附近的特征信号进行插值，将该信号
15 谷替换为插值获得的信号。

S134、将预设第一阈值更新为预设第二阈值，返回获取半高全宽小于预设第一阈值的信号谷的步骤，直至得到平滑的特征信号。

以不同的阈值重复上述步骤若干次，直至得到平滑的特征信号。后续即可以利用该平滑的特征信号进行峰值检测（即条件检峰），进一步提高了检测结果的准确性。
20

应当理解，本发明并不限制于上述平滑的方案，任何具备低通滤波特性的平滑操作，如均值滤波、高斯平滑等都应当被视为等同的过程。

由于平滑的影响，通过条件检峰得到的峰值点未必准确的对应了原始特征信号的峰值点，因此需要进行一定的时间对齐，在一个实施例中，所述对所述特征信号进行峰值检测，获得若干个峰值点之后，所述根据若干个峰值点获得底鼓的节拍点之前，还包括：从所述特征信号中每个峰值点所指示位置的邻近区域内寻找最大值，将寻找到的最大值作为对齐后的峰值点。
25

上述实施例中，对于每一个峰值点，邻近区域指的是其中的每一个点与对应峰值点的距离小于预设阈值。对于每一个峰值点，从特征信号上该位置附近一定范围内寻找最大值，并将该最大值位置作为对齐后的峰值点输出。
30

对于绝大多数带底鼓的音乐而言，通过上述步骤得到的峰值点已经具备了较高的准确度，然而，仍然有少部分音乐，特别是对一些具有较强的低频干扰源如贝斯、手敲鼓、男低音等的音乐而言，其得到的峰值点具有一些的误判点，针对这一问题，本发明使用二次筛选的方式对获取到的峰值点进行进一步的筛选。因此，在一个实施例中，所述对所述特征信号进行峰值检测，获得若干个峰值点之后，所述根据若干个峰值点获得底鼓的节拍点之前，还包括：

S141、从所述特征信号中每个峰值点所指示位置的邻近区域内，统计所述邻近区域内的特征信号数值超过对应峰值点对应的特征信号数值预设比例的数量，若所述数量超过预设门限值，将对应的峰值点剔除。

10 该步骤中，特征信号数值为 ζ 的值，以特征信号采用 X 轴和 Y 轴坐标系表示为例，X 轴用于表征位置（也即是时间点），Y 轴用于表征 ζ 的值。邻近区域指的是该区域中的每一个点与对应峰值点的距离小于预设阈值。预设比例和预设门限值均可以根据实际需要进行设置。对于每一个峰值点，计算该峰值点附近的特征信号中超过该峰值点对应的特征信号数值预设比例的点的数量，当数量超过预设的门限时，
15 将该峰值点剔除。

和/或，

S142、当两个连续的峰值点之间的间隔小于预设的间隔门限时，将对应的特征信号数值低的峰值点剔除。

上述步骤中，两个连续的峰值点指的是两个相邻的峰值点。当相邻两个峰值点
20 之间的间隔小于设定的门限时，剔除其中对应特征信号数值更低的峰值点。

和/或，

S143、当一个峰值点对应的特征信号数值均低于其他峰值点对应的特征信号数值时，计算该个峰值点相较于其邻域区域内的特征信号的凸显度，若所述凸显度小于预设阈值，将该个峰值点剔除。

25 上述步骤中，可选的，当某个峰值点对应的特征信号数值明显的低于整首音乐的其他的峰值点对应的特征信号数值时，才计算凸显度，明显的低于指的是该峰值点的特征信号数值与其他峰值点对应的特征信号数值之间的差值均大于一个设定的数值，或者该峰值点的特征信号数值与其他峰值点对应的特征信号数值之间的比值均小于一个设定的数值。当所述凸显度小于预设阈值时将该个峰值点剔除即当该
30 峰值点相较于周围特征信号没有明显凸显时将该峰值点剔除。可选的，凸显度指的是该峰值点与其左右两侧一定范围内的特征信号的（均值+方差的 1.5 倍）的比值。

利用本发明得到的节拍点信息可供产品用于对该音乐进行所需要的处理，例如，在一个实施例中，所述根据若干个峰值点获得底鼓的节拍点之后，还包括：在所述底鼓的节拍点所在的位置添加预设的音视频特效。通过在本发明给出的底鼓节拍点位置添加上一系列音视频特效，能够使得最终视频效果达到与音乐节奏与情感相统一，从而具备较好的整体呈现效果。

应当理解，在得到底鼓的节拍点之后，本发明并不限制于在该节拍点处添加音视频特效，用户还可以根据得到的节拍点进行其它操作，例如音乐游戏等。

如图3所示，为一具体实施例的音频信号底鼓节拍点的检测方法的流程示意图，该方法可以由C++代码构成的数字信号处理程序实现，可运行于任何支持C++运行环境的计算硬件之上。应当理解，本发明并不限制于由C++代码实现，用户还可以采用其它编程语言。

具体而言，该具体实施例包含6个部分，各部分之间的关系及数据处理流程如下所述：

15 S1、数据预处理

将原始音频数据以2kHz的采样率进行重采样，对重采样后的信号进行低通滤波，使用的滤波器为8阶巴特沃斯低通滤波器，截止频率150Hz。

S2、经验模式分解

对低通滤波后的音频数据进行周期延拓，利用延拓后的信号进行经验模式分解，得到若干个本征模式信号以及一个余量模式信号，称这些信号为原始音频数据的本征模函数。

S3、特征计算

对每一条本征模函数分别计算 A_i 与 ω_i ，并最终得到特征信号 $\zeta = \sum A_i^2 \omega_i$ 。

S4、特征检峰

25 特征检峰包括两个步骤：信号平滑与条件检峰。通过信号平滑得到平滑的特征信号，然后对该特征信号进行条件检峰，获得筛选后的峰值点。

S5、时间对齐

对特征检峰输出的每一个峰值点，从特征计算得到的特征信号上该位置附近一定范围内寻找最大值，并将该最大值位置作为对齐后的峰值点输出给二次筛选步骤。

30 S6、二次筛选

二次筛选包含三个过程：

1. 对于时间对齐步骤后的输出的每一个峰值点，计算该个峰值点的附近的特征信号中，超过该峰值点对应的特征信号数值特定比例的点的数量，当数量超过预设的门限时，将该峰值点剔除；

2. 对于时间对齐步骤后的输出的每一个峰值点，当连续两个峰值点之间的间隔
5 小于设定的门限时，剔除其中对应特征信号数值更低的峰值点；

3. 对于时间对齐步骤后的输出的每一个峰值点，当某个峰值点对应的特征信号数值明显的低于整首音乐的其他的峰值点对应的特征信号数值时，分析该峰值点相较于其附近一定范围内的特征信号的凸显度，当该峰值点相较于周围特征信号没有明显凸显时将该峰值点剔除。

10 通过上述 6 个部分即可以得到准确的峰值点，该峰值点即为底鼓被敲击的准确时间点，再利用得到的底鼓敲击时间点进行进一步的音乐节奏信息分析，得到最终的节拍点信息。

本发明实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，该
15 程序被处理器执行时实现前述任意一项所述的音频信号底鼓节拍点的检测方法。所述存储介质包括但不限于任何类型的盘（包括软盘、硬盘、光盘、CD-ROM、和磁盘）、ROM（Read-Only Memory，只读存储器）、RAM（Random Access Memory，随机存储器）、EPROM（Erasable Programmable Read-Only Memory，可擦写可编程只读存储器）、EEPROM（Electrically Erasable Programmable Read-Only Memory，
20 电可擦可编程只读存储器）、闪存、磁性卡片或光线卡片。也就是，存储介质包括由设备（例如，计算机）以能够读的形式存储或传输信息的任何介质。可以是只读存储器，磁盘或光盘等。

本发明实施例还提供了一种终端，所述终端包括：

25 一个或多个处理器；

存储装置，用于存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如前述任意一项所述的音频信号底鼓节拍点的检测方法。

如图 4 所示，为了便于说明，仅示出了与本发明实施例相关的部分，具体技术
30 细节未揭示的，请参照本发明实施例方法部分。该终端可以为包括手机、平板电脑、PDA（Personal Digital Assistant，个人数字助理）、POS（Point of Sales，销售终端）、

车载电脑等任意终端设备，以终端为手机为例：

图 4 示出的是与本发明实施例提供的终端相关的手机的部分结构的框图。参考图 4，手机包括：射频（Radio Frequency，RF）电路 1510、存储器 1520、输入单元 1530、显示单元 1540、传感器 1550、音频电路 1560、无线保真（wireless fidelity，Wi-Fi）模块 1570、处理器 1580、以及电源 1590 等部件。本领域技术人员可以理解，图 4 中示出的手机结构并不构成对手机的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置。

下面结合图 4 对手机的各个构成部件进行具体的介绍：

RF 电路 1510 可用于收发信息或通话过程中，信号的接收和发送，特别地，将基站的下行信息接收后，给处理器 1580 处理；另外，将设计上的数据发送给基站。通常，RF 电路 1510 包括但不限于天线、至少一个放大器、收发信机、耦合器、低噪声放大器（Low Noise Amplifier，LNA）、双工器等。此外，RF 电路 1510 还可以通过无线通信与网络和其他设备通信。上述无线通信可以使用任一通信标准或协议，包括但不限于全球移动通讯系统（Global System of Mobile communication，GSM）、通用分组无线服务（General Packet Radio Service，GPRS）、码分多址（Code Division Multiple Access，CDMA）、宽带码分多址（Wideband Code Division Multiple Access，WCDMA）、长期演进（Long Term Evolution，LTE）、电子邮件、短消息服务（Short Messaging Service，SMS）等。

存储器 1520 可用于存储软件程序以及模块，处理器 1580 通过运行存储在存储器 1520 的软件程序以及模块，从而执行手机的各种功能应用以及数据处理。存储器 1520 可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序（比如节拍点检测功能等）等；存储数据区可存储根据手机的使用所创建的数据（比如峰值点等）等。此外，存储器 1520 可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。

输入单元 1530 可用于接收输入的数字或字符信息，以及产生与手机的用户设置以及功能控制有关的键信号输入。具体地，输入单元 1530 可包括触控面板 1531 以及其他输入设备 1532。触控面板 1531，也称为触摸屏，可收集用户在其上或附近的触摸操作（比如用户使用手指、触笔等任何适合的物体或附件在触控面板 1531 上或在触控面板 1531 附近的操作），并根据预先预设的程式驱动相应的连接装置。可选的，触控面板 1531 可包括触摸检测装置和触摸控制器两个部分。其中，触摸

检测装置检测用户的触摸方位，并检测触摸操作带来的信号，将信号传送给触摸控制器；触摸控制器从触摸检测装置上接收触摸信息，并将它转换成触点坐标，再送给处理器 1580，并能接收处理器 1580 发来的命令并加以执行。此外，可以采用电阻式、电容式、红外线以及表面声波等多种类型实现触控面板 1531。除了触控面板 5 1531，输入单元 1530 还可以包括其他输入设备 1532。具体地，其他输入设备 1532 可以包括但不限于物理键盘、功能键（比如音量控制按键、开关按键等）、轨迹球、鼠标、操作杆等中的一种或多种。

显示单元 1540 可用于显示由用户输入的信息或提供给用户的信息以及手机的各种菜单。显示单元 1540 可包括显示面板 1541，可选的，可以采用液晶显示器 10 （Liquid Crystal Display，LCD）、有机发光二极管（Organic Light-Emitting Diode，OLED）等形式来配置显示面板 1541。进一步的，触控面板 1531 可覆盖显示面板 1541，当触控面板 1531 检测到在其上或附近的触摸操作后，传送给处理器 1580 以确定触摸事件的类型，随后处理器 1580 根据触摸事件的类型在显示面板 1541 上提供相应的视觉输出。虽然在图 4 中，触控面板 1531 与显示面板 1541 是作为两个独立 15 的部件来实现手机的输入和输入功能，但是在某些实施例中，可以将触控面板 1531 与显示面板 1541 集成而实现手机的输入和输出功能。

手机还可包括至少一种传感器 1550，比如光传感器、运动传感器以及其他传感器。具体地，光传感器可包括环境光传感器及接近传感器，其中，环境光传感器可根据环境光线的明暗来调节显示面板 1541 的亮度，接近传感器可在手机移动到耳 20 边时，关闭显示面板 1541 和/或背光。作为运动传感器的一种，加速计传感器可检测各个方向上（一般为三轴）加速度的大小，静止时可检测出重力的大小及方向，可用于识别手机姿态的应用（比如横竖屏切换、相关游戏、磁力计姿态校准）、振动识别相关功能（比如计步器、被敲击）等；至于手机还可配置的陀螺仪、气压计、湿度计、温度计、红外线传感器等其他传感器，在此不再赘述。

25 音频电路 1560、扬声器 1561，传声器 1562 可提供用户与手机之间的音频接口。音频电路 1560 可将接收到的音频数据转换后的电信号，传输到扬声器 1561，由扬声器 1561 转换为声纹信号输出；另一方面，传声器 1562 将收集的声纹信号转换为电信号，由音频电路 1560 接收后转换为音频数据，再将音频数据输出处理器 1580 处理后，经 RF 电路 1510 以发送给比如另一手机，或者将音频数据输出至存储器 30 1520 以便进一步处理。

Wi-Fi 属于短距离无线传输技术，手机通过 Wi-Fi 模块 1570 可以帮助用户收发

电子邮件、浏览网页和访问流式媒体等，它为用户提供了无线的宽带互联网访问。虽然图 4 示出了 Wi-Fi 模块 1570，但是可以理解的是，其并不属于手机的必须构成，完全可以根据需要在不改变发明的本质的范围内而省略。

处理器 1580 是手机的控制中心，利用各种接口和线路连接整个手机的各个部分，通过运行或执行存储在存储器 1520 内的软件程序和/或模块，以及调用存储在存储器 1520 内的数据，执行手机的各种功能和处理数据，从而对手机进行整体监控。可选的，处理器 1580 可包括一个或多个处理单元；优选的，处理器 1580 可集成应用处理器和调制解调处理器，其中，应用处理器主要处理操作系统、用户界面和应用程序等，调制解调处理器主要处理无线通信。可以理解的是，上述调制解调处理器也可以不集成到处理器 1580 中。

手机还包括给各个部件供电的电源 1590（比如电池），优选的，电源可以通过电源管理系统与处理器 1580 逻辑相连，从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。

尽管未示出，手机还可以包括摄像头、蓝牙模块等，在此不再赘述。

通过本发明实施例所提供的方案能够自动化的获取到音乐中底鼓被敲击的准确时间点，从而能提供对整首音乐的节奏与情绪流进行分析的信息，效率较高；通过这些底鼓节拍点上添加特定的音视频特效，能够使得相应的特效效果与音乐本身更加的贴合，从而达到较好的产品效果，且可以于产品上实时呈现。

应该理解的是，虽然附图的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，其可以以其他的顺序执行。而且，附图的流程图中的至少一部分步骤可以包括多个子步骤或者多个阶段，这些子步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，其执行顺序也不必然是依次进行，而是可以与其他步骤或者其他步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

以上所述仅是本发明的部分实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

- 1.一种音频信号底鼓节拍点的检测方法，其特征在于，包括步骤：
根据输入的待检测音频信号获得若干个本征模函数；
5 计算若干个本征模函数对应的瞬时强度信号以及瞬时频率信号；
根据若干个本征模函数对应的瞬时强度信号和瞬时频率信号，获得底鼓的特征信号；
对所述特征信号进行峰值检测，获得若干个峰值点；
根据若干个峰值点获得底鼓的节拍点。
- 10 2.根据权利要求1所述的音频信号底鼓节拍点的检测方法，其特征在于，所述对所述特征信号进行峰值检测，获得若干个峰值点，包括：
对所述特征信号进行峰值检测，获取各个极大值点；
从各个极大值点中选取满足预设条件的极大值点，将选取的极大值点判定为峰值点；
15 其中，所述预设条件包括：两个连续的极大值点之间的特征信号中任何一个点均不是极大值点，且两个连续的极大值点之间的特征信号中最小值远小于该两个连续的极大值点。
- 3.根据权利要求2所述的音频信号底鼓节拍点的检测方法，其特征在于，所述将选取的极大值点判定为峰值点之后，还包括：
20 计算由每个峰值点与其邻近点所构成的信号峰经过高斯拟合后的半高全宽；
若半高全宽小于预设门限，将对应的峰值点保留，否则将对应的峰值点剔除。
- 4.根据权利要求1至3任意一项所述的音频信号底鼓节拍点的检测方法，其特征在于，所述根据若干个本征模函数对应的瞬时强度信号和瞬时频率信号，获得底鼓的特征信号，包括：
25 将瞬时强度信号平方后与瞬时频率信号相乘，得到每个本征模函数的等效瞬时频率；
对所有本征模函数的等效瞬时频率求和，获得所述底鼓的特征信号。
- 5.根据权利要求1至3任意一项所述的音频信号底鼓节拍点的检测方法，其特征在于，所述获得底鼓的特征信号之后，所述对所述特征信号进行峰值检测，获得
30 若干个峰值点之前，还包括：
获取所述特征信号的所有谷值点；

计算由每个谷值点与其最近的两个峰值点所构成的信号谷的半高全宽；

获取半高全宽小于预设第一阈值的信号谷，利用该信号谷邻近的特征信号通过插值的方式将该信号谷去除；

将预设第一阈值更新为预设第二阈值，返回获取半高全宽小于预设第一阈值的信号谷的步骤，直至得到平滑的特征信号。

6.根据权利要求5所述的音频信号底鼓节拍点的检测方法，其特征在于，所述对所述特征信号进行峰值检测，获得若干个峰值点之后，所述根据若干个峰值点获得底鼓的节拍点之前，还包括：

从所述特征信号中每个峰值点所指示位置的邻近区域内寻找最大值，将寻找到的最大值作为对齐后的峰值点。

7.根据权利要求1至3任意一项所述的音频信号底鼓节拍点的检测方法，其特征在于，所述对所述特征信号进行峰值检测，获得若干个峰值点之后，所述根据若干个峰值点获得底鼓的节拍点之前，还包括：

从所述特征信号中每个峰值点所指示位置的邻近区域内，统计所述邻近区域内的特征信号数值超过对应峰值点对应的特征信号数值预设比例的数量，若所述数量超过预设门限值，将对应的峰值点剔除；

和/或，

当两个连续的峰值点之间的间隔小于预设的间隔门限时，将对应的特征信号数值低的峰值点剔除；

和/或，

当一个峰值点对应的特征信号数值均低于其他峰值点对应的特征信号数值时，计算该个峰值点相较于其邻域区域内的特征信号的凸显度，若所述凸显度小于预设阈值，将该个峰值点剔除。

8.根据权利要求1至3任意一项所述的音频信号底鼓节拍点的检测方法，其特征在于，所述根据若干个峰值点获得底鼓的节拍点之后，还包括：

在所述底鼓的节拍点所在的位置添加预设的音视频特效。

9.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现如权利要求1至8中任意一项所述的音频信号底鼓节拍点的检测方法。

10.一种终端，其特征在于，所述终端包括：

一个或多个处理器；

存储装置，用于存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求 1 至 8 中任意一项所述的音频信号底鼓节拍点的检测方法。