



(12)发明专利申请

(10)申请公布号 CN 111508457 A

(43)申请公布日 2020.08.07

(21)申请号 202010291577.8

(22)申请日 2020.04.14

(71)申请人 上海影卓信息科技有限公司

地址 201100 上海市闵行区金都路4299号5
幢1楼1A室

(72)发明人 邓海峰 林立

(74)专利代理机构 上海段和段律师事务所

31334

代理人 李佳俊 郭国中

(51)Int.Cl.

G10H 1/40(2006.01)

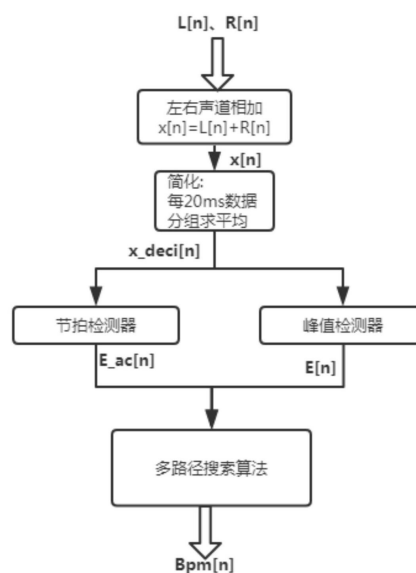
权利要求书2页 说明书5页 附图2页

(54)发明名称

音乐节拍检测方法和系统

(57)摘要

本发明提供一种音乐节拍检测方法和系统,接收音乐后,将双通道音频进行左右声道相加,得到音乐PCM类型的数据序列,对数据序列进行分组相加,求取平均值以进行简化,得到简化后音频信号;对简化后音频信号进行自相关函数计算,再进行波峰检测得到所有波峰;对简化后音频信号进行包络提取,得到音频包络信号,对音频包络信号进行一阶差分处理、半波整流处理和加权补偿处理,得到音频峰值信号,利用动态阈值波峰检测输出所有节拍点;利用所有波峰和所有节拍点,通过多路径搜索得到最优节拍路径,记录最优节拍路径对应的音频节奏和节拍点。解决音乐节拍检测和具体的节奏点精准定位问题,通过寻找可能的节拍值,便于实现的自相关函数来估算节拍值。



1. 一种音乐节拍检测方法,其特征在于,包括以下步骤:

音频分段步骤:接收音乐后,将双通道音频进行左右声道相加,得到音乐PCM类型的数据序列,对数据序列进行分组相加,求取平均值以进行简化,得到简化后音频信号;

检测节拍步骤:对简化后音频信号进行自相关函数计算,再进行波峰检测,得到所有波峰;

提取峰值步骤:对简化后音频信号进行包络提取,得到音频包络信号,对音频包络信号进行一阶差分处理、半波整流处理和加权补偿处理,得到音频峰值信号,利用动态阈值波峰检测输出所有节拍点;

多路搜索步骤:利用所有波峰和所有节拍点,通过多路径搜索得到最优节拍路径,记录最优节拍路径对应的音频节奏和节拍点。

2. 根据权利要求1所述的音乐节拍检测方法,其特征在于,所述音乐PCM类型的数据序列 $x[n]$ 表示一维离散时间信号,音乐双声道分别为 $L[n]$ 、 $R[n]$, $x[n]=L[n]+R[n]$;

每20ms将一组音频分组数据相加,求取平均
$$x_{deci}[n] = \frac{\sum_{i=0}^m x[n * \frac{m}{2} + i]}{m},$$
 m 表示20毫秒长度

的音频采样点个数, m 是一个常数, $0 \leq i < m$ 。

3. 根据权利要求2所述的音乐节拍检测方法,其特征在于,所述对简化后音频信号进行包络提取是对简化后音频信号做截止频率为20Hz的低通滤波处理,再用快充慢放的包络检波器进行处理保留包络的突起,输出为音频包络信号 $x_{env}[n]$;

设 $t[n]$ 是 $x_{deci}[n]$ 经过截止频率为20Hz的低通滤波器的输出,则快充慢放包络检波器输出为

$$x_{env}[n] < t[n], \partial = 0.6$$

$$x_{env}[n+1] = (1-\partial) * x_{env}[n] + \partial * t[n], \quad x_{env}[n] \geq t[n], \partial = 0.9。$$

4. 根据权利要求3所述的音乐节拍检测方法,其特征在于,令所述音频包络信号 $x_{env}[n]$ 进行一阶差分、半波整流、加权补偿,突出包络中的峰值点,输出处理后的突出的音频峰值信号,表述为音频峰值信号 $x_s[n]$, $x_s[n]$ 是 $x_{env}[n]$ 经过一阶差分、半波整流、加权补偿之后获取到的峰值信号;

所述一阶差分 $dx[n] = x_{env}[n] - x_{env}[n-1]$,半波整流 $hx[n] = \max(dx[n], 0)$,加权补偿 $x_s[n] = 0.5 * 31.5 * hx[n] + 0.5 * dx[n]$;

再利用基于动态阈值波峰检测器检测 $x_s[n]$,输出所有可能的节拍点 $E[n]$ 。

5. 根据权利要求2所述的音乐节拍检测方法,其特征在于,所述对简化后音频信号进行自相关函数计算是对 $x_{deci}[n]$ 调用自相关函数,输出自相关序列 $ac[n]$,对自相关序列 $ac[n]$ 调用波峰检测器,找出所有波峰 $E_{acPeak}[n]$,每个 $E_{acPeak}[n]$ 保存可能的节拍值。

6. 一种音乐节拍检测系统,其特征在于,包括以下模块:

音频分段模块:接收音乐后,将双通道音频进行左右声道相加,得到音乐PCM类型的数据序列,对数据序列进行分组相加,求取平均值以进行简化,得到简化后音频信号;

检测节拍模块:对简化后音频信号进行自相关函数计算,再进行波峰检测,得到所有波峰;

提取峰值模块:对简化后音频信号进行包络提取,得到音频包络信号,对音频包络信号

进行一阶差分处理、半波整流处理和加权补偿处理,得到音频峰值信号,利用动态阈值波峰检测输出所有节拍点;

多路搜索模块:利用所有波峰和所有节拍点,通过多路径搜索得到最优节拍路径,记录最优节拍路径对应的音频节奏和节拍点。

7. 根据权利要求6所述的音乐节拍检测系统,其特征在于,所述音乐PCM类型的数据序列 $x[n]$ 表示一维离散时间信号,音乐双声道分别为 $L[n]$ 、 $R[n]$, $x[n]=L[n]+R[n]$;

每20ms将一组音频分组数据相加,求取平均
$$x_{\text{deci}}[n] = \frac{\sum_{i=0}^m x[n * \frac{m}{2} + i]}{m},$$
 m 表示20毫秒长度

的音频采样点个数, m 是一个常数, $0 \leq i < m$ 。

8. 根据权利要求7所述的音乐节拍检测系统,其特征在于,所述对简化后音频信号进行包络提取是对简化后音频信号做截止频率为20Hz的低通滤波处理,再用快充慢放的包络检波器进行处理保留包络的突起,输出为音频包络信号 $x_{\text{env}}[n]$;

设 $t[n]$ 是 $x_{\text{deci}}[n]$ 经过截止频率为20Hz的低通滤波器的输出,则快充慢放包络检波器输出为
$$x_{\text{env}}[n] < t[n], \partial = 0.6$$

$$x_{\text{env}}[n+1] = (1-\partial) * x_{\text{env}}[n] + \partial * t[n], \quad x_{\text{env}}[n] \geq t[n], \partial = 0.9。$$

9. 根据权利要求8所述的音乐节拍检测系统,其特征在于,令所述音频包络信号 $x_{\text{env}}[n]$ 进行一阶差分、半波整流、加权补偿,突出包络中的峰值点,输出处理后的突出的音频峰值信号,表述为音频峰值信号 $x_s[n]$, $x_s[n]$ 是 $x_{\text{env}}[n]$ 经过一阶差分、半波整流、加权补偿之后获取到的峰值信号;

所述一阶差分 $dx[n] = x_{\text{env}}[n] - x_{\text{env}}[n-1]$,半波整流 $hx[n] = \max(dx[n], 0)$,加权补偿 $x_s[n] = 0.5 * 31.5 * hx[n] + 0.5 * dx[n]$;

再利用基于动态阈值波峰检测器检测 $x_s[n]$,输出所有可能的节拍点 $E[n]$ 。

10. 根据权利要求7所述的音乐节拍检测系统,其特征在于,所述对简化后音频信号进行自相关函数计算是对 $x_{\text{deci}}[n]$ 调用自相关函数,输出自相关序列 $ac[n]$,对自相关序列 $ac[n]$ 调用波峰检测器,找出所有波峰 $E_{\text{acPeak}}[n]$,每个 $E_{\text{acPeak}}[n]$ 保存可能的节拍值。

音乐节拍检测方法和系统

技术领域

[0001] 本发明涉及音频数据处理技术领域,具体地,涉及一种音乐节拍检测方法和系统。

背景技术

[0002] 乐曲的音乐节奏检测对音乐理解以及可视化至关重要,如何通过音乐节奏进行检测并合理识别出音乐节拍的准确位置,以实现任何人都能伴随着一段乐曲打拍子。随着计算机以及多媒体技术的迅速发展,目前出现了很多在智能节奏检测、节奏跟踪等方面的研究,例如,基于贝叶斯理论的单音音乐节奏提取方法,引入贝叶斯节奏模型,然后采用基于贝叶斯理论的序列蒙特卡洛方法推断音乐片段的小节和节拍的位置。或者基于内外概率算法的音乐节奏提取方法,对音乐元素进行分析,定义了一个概率上下文无关文法系统来描述相对独立的音乐节奏元素,以将音乐节奏转化为形式化的文法语句,采用内外概率算法得到文法系统中每条文法语句被使用的概率值,利用概率上下文无关文法指导计算机生成音乐节奏。以上研究基于复杂的概率及相关理论来推导计算音乐片段中的节奏数值,均存在计算复杂度较高、准确度较低的缺点。

[0003] 与本申请相关的现有技术是专利文献CN 107103917A,公开了一种音乐节奏检测方法及其系统,方法包括:获取音乐的音频数据;在所述音频数据中依序获取一音频帧作为当前音频帧,将当前音频帧与上一音频帧的频谱能量和的差值,作为当前音频帧的能量差值并保存所述能量差值;确定当前音频帧对应的能量阈值;获取当前音频帧及其之前的与其相邻的两个以上连续的音频帧的能量差值,得到三个以上的音频帧的能量差值;若所述三个以上的音频帧的能量差值中存在峰值,且所述峰值大于当前音频帧对应的能量阈值,则将所述峰值对应的音频帧标记为节奏点。

发明内容

[0004] 针对现有技术中的缺陷,本发明的目的是提供一种音乐节拍检测方法和系统。

[0005] 根据本发明提供一种音乐节拍检测方法,包括以下步骤:

[0006] 音频分段步骤:接收音乐后,将双通道音频进行左右声道相加,得到音乐PCM类型的数据序列,对数据序列进行分组相加,求取平均值以进行简化,得到简化后音频信号;

[0007] 检测节拍步骤:对简化后音频信号进行自相关函数计算,再进行波峰检测,得到所有波峰;

[0008] 提取峰值步骤:对简化后音频信号进行包络提取,得到音频包络信号,对音频包络信号进行一阶差分处理、半波整流处理和加权补偿处理,得到音频峰值信号,利用动态阈值波峰检测输出所有节拍点;

[0009] 多路搜索步骤:利用所有波峰和所有节拍点,通过多路径搜索得到最优节拍路径,记录最优节拍路径对应的音频节奏和节拍点。

[0010] 优选地,所述音乐PCM类型的数据序列 $x[n]$ 表示一维离散时间信号,音乐双声道分别为 $L[n]$ 、 $R[n]$, $x[n]=L[n]+R[n]$;

[0011] 每20ms将一组音频分组数据相加,求取平均 $x_{\text{deci}}[n] = \frac{\sum_{i=0}^m x[n * \frac{m}{2} + i]}{m}$, m表示20毫秒

长度的音频采样点个数,m是一个常数, $0 \leq i < m$ 。

[0012] 优选地,所述对简化后音频信号进行包络提取是对简化后音频信号做截止频率为20Hz的低通滤波处理,再用快充慢放的包络检波器进行处理保留包络的突起,输出为音频包络信号 $x_{\text{env}}[n]$;

[0013] 设 $t[n]$ 是 $x_{\text{deci}}[n]$ 经过截止频率为20Hz的低通滤波器的输出,则快充慢放包络检

波器输出为 $x_{\text{env}}[n] < t[n], \partial = 0.6$
 $x_{\text{env}}[n+1] = (1-\partial) * x_{\text{env}}[n] + \partial * t[n], x_{\text{env}}[n] \geq t[n], \partial = 0.9$ 。

[0014] 优选地,令所述音频包络信号 $x_{\text{env}}[n]$ 进行一阶差分、半波整流、加权补偿,突出包络中的峰值点,输出处理后的突出的音频峰值信号,表述为音频峰值信号 $x_s[n]$, $x_s[n]$ 是 $x_{\text{env}}[n]$ 经过一阶差分、半波整流、加权补偿之后获取到的峰值信号;

[0015] 所述一阶差分 $dx[n] = x_{\text{env}}[n] - x_{\text{env}}[n-1]$,半波整流 $hx[n] = \max(dx[n], 0)$,加权补偿 $x_s[n] = 0.5 * 31.5 * hx[n] + 0.5 * dx[n]$;

[0016] 再利用基于动态阈值波峰检测器检测 $x_s[n]$,输出所有可能的节拍点 $E[n]$ 。

[0017] 优选地,所述对简化后音频信号进行自相关函数计算是对 $x_{\text{deci}}[n]$ 调用自相关函数,输出自相关序列 $ac[n]$,对自相关序列 $ac[n]$ 调用波峰检测器,找出所有波峰 $E_{\text{acPeak}}[n]$,每个 $E_{\text{acPeak}}[n]$ 保存可能的节拍值。

[0018] 根据本发明提供一种音乐节拍检测系统,包括以下模块:

[0019] 音频分段模块:接收音乐后,将双通道音频进行左右声道相加,得到音乐PCM类型的数据序列,对数据序列进行分组相加,求取平均值以进行简化,得到简化后音频信号;

[0020] 检测节拍模块:对简化后音频信号进行自相关函数计算,再进行波峰检测,得到所有波峰;

[0021] 提取峰值模块:对简化后音频信号进行包络提取,得到音频包络信号,对音频包络信号进行一阶差分处理、半波整流处理和加权补偿处理,得到音频峰值信号,利用动态阈值波峰检测输出所有节拍点;

[0022] 多路搜索模块:利用所有波峰和所有节拍点,通过多路径搜索得到最优节拍路径,记录最优节拍路径对应的音频节奏和节拍点。

[0023] 与现有技术相比,本发明具有如下的有益效果:

[0024] 1、本发明解决了音乐节奏检测和具体的节奏点精准定位问题。

[0025] 2、本发明通过寻找可能的节拍值,便于实现的自相关函数来估算节拍值。

附图说明

[0026] 通过阅读参照以下附图对非限制性实施例所作的详细描述,本发明的其它特征、目的和优点将会变得更明显:

[0027] 图1为本发明的处理流程图;

[0028] 图2为本发明中节拍检测器流程图;

- [0029] 图3为本发明中峰值检测器流程图；
 [0030] 图4为本发明中包络提取器流程图；
 [0031] 图5为本发明中峰值提取器流程图。

具体实施方式

[0032] 下面结合具体实施例对本发明进行详细说明。以下实施例将有助于本领域的技术人员进一步理解本发明,但不以任何形式限制本发明。应当指出的是,对本领域的普通技术人员来说,在不脱离本发明构思的前提下,还可以做出若干变化和改进。这些都属于本发明的保护范围。

[0033] 本发明是基于时域的“波峰检测算法”以找到所有可能的节拍点;通过自相关函数,提取出音乐的bpm值,根据所有可能的节拍点和bpm值,利用多路径搜索,查找所有节拍的位置。其中,“波峰检测算法”主要是检测声音等波形的波峰和波谷,是基于动态阈值的波

峰检测算法,将音频序列 $A[n]$ 按2秒分段,分别计算每段的平均值

$$AVG[n] = \frac{\sum_{i=0}^l A[n \cdot l + i]}{l},$$

$AVG[n]$ 表示第 n 段的平均值(avg), $A[n]$ 是待处理的音频序列,同时也表示第 n 点的音频强度值, l 表示2秒长度的音频采样点个数, a_i 表示第 i 个音频序列的动态阈值,设定 $a_i = A[n \cdot l + i]$,此时 n 为固定值, $0 \leq i < l$, $AVG[n]$ 是第 n 段的动态阈值,若 $a_i = A[n \cdot l + i] > AVG[n]$,则 $n \cdot l + i$ 为可能的波峰,若 $a_i > avg$, i 为可能的波峰,若在第 i 个左右50ms附近有大于 a_i 的值的波峰,则将第 i 个从波峰候选点中删除。

[0034] 首先对接收音乐中双通道音频进行左右声道相加,得到音乐PCM类型的数据序列 $x[n]$ 表示一维离散时间信号,一般音乐都是双声道,即2个一维离散时间信号,分别为 $L[n]$, $R[n]$, $x[n] = L[n] + R[n]$,即刚开始得到的数据,其中 n 表示数据序列的序号,取值为自然数,

每20ms将一组音频分组数据相加,求取平均 $x_{deci}[n] = \frac{\sum_{i=0}^m x[n \cdot \frac{m}{2} + i]}{m}$, m 表示20毫秒长度的音

频采样点个数,对于一个44.1kHz的音频来说, $m=882$ 表示20毫秒的音频长度的采样个数, m 是一个常数, $0 \leq i < m$,每一组音频分组数据和邻近的音频分组数据有10ms的重叠,使得到的

的 $x_{deci}[n]$ 信号更平滑,输出 $x_{deci}[n] = \sum_{i=0}^m x[n \cdot m/2 + i]$, $x_{deci}[n]$ 表示对 $x[n]$ 信号简化,即简

化后的音频信号,其中 m 表示20ms的序列长度, i 表示序列长度的序号, n 表示数据序列的序

号; $\frac{\sum_{i=0}^m x[n \cdot \frac{m}{2} + i]}{m}$ 中 $\frac{m}{2}$ 即包含了10ms的数据重叠操作。

[0035] 其次,对 $x_{deci}[n]$ 做截止频率为20Hz的低通滤波处理,并用快充慢放的包络检波器进行处理最大限度的保留包络的突起,输出为音频包络信号 $x_{env}[n]$, $x_{env}[n]$ 是 $x_{deci}[n]$ 先经过截止频率为20Hz的低通滤波,去掉音频信号中的高频变化成分,再进过快充慢放的包络检波器突出并保留包络的突起,对 $x_{env}[n]$ 进行一阶差分,半波整流,加权补偿,突出包络中的峰值点,输出处理后的突出的音频峰值信号,可表述为音频峰值信号 $x_s[n]$, $x_s[n]$ 是 x_{env}

[n]经过一阶差分、半波整流、加权补偿之后获取到的峰值信号,一阶差分公式是 $dx[n] = x_{env}[n] - x_{env}[n-1]$,半波整流公式是 $hx[n] = \max(dx[n], 0)$,加权补偿公式是 $x_s[n] = 0.5 * 31.5 * hx[n] + 0.5 * dx[n]$,利用基于动态阈值波峰检测器检测 $x_s[n]$,输出所有可能的节拍点 $E[n]$;同时,对 $x_{deci}[n]$ 调用自相关函数,输出自相关序列 $ac[n]$,对自相关序列 $ac[n]$ 调用波峰检测器,找出所有波峰 $E_{acPeak}[n]$,每个 $E_{acPeak}[n]$ 保存可能的节拍值,利用节拍点 $E[n]$ 和波峰 $E_{acPeak}[n]$ 通过多路径搜索算法获取最优的节拍路径的位置, $E[n]$ 表示最终峰值信号,节拍点应该都包含在峰值信号中, $E_{acPeak}[n]$ 表示所有可能的节奏,音乐节奏应该包含在内,路劲定义 $Path = \{b, \{p1, p2, \dots\}\}$, b 表示节奏值, $p1, p2, p3, \dots$ 分别表示第一、第二、第三等节奏点,定义一个对象函数 $F(Path)$,若这个路劲的节奏点和 b 值越匹配则 $F(Path)$ 值越大,反之则越小,用 $E_{acPeak}[n]$ 包含的节奏和 $E[n]$ 的起始节拍点初始化多条路径 $\{b, E[s]\}$,并预测跟踪下一个节拍点,判断是否将此节拍点添加入路径,并更新对应的 $F(Path)$ 函数,找到 $F(Path)$ 值最大的路径,从而找到音频的节奏和对应的节拍点。

[0036] 如图1所示,对双通道音频,进行左右声道相加,得到音乐PCM类型的数据序列 $x[n]$,然后每20ms一组音频分组数据相加,求取平均,每一组音频分组数据和邻近的音频分

组数据有10ms的重叠,输出 $x_{deci}[n] = \frac{\sum_{i=0}^m x[n * \frac{m}{2} + i]}{m}$, m 表示20ms的序列长度。

[0037] 如图4所示,对 $x_{deci}[n]$ 做截止频率为20Hz的低通滤波,并用快充慢放包络检波器进行处理最大限度的保留包络的突起,输出为 $x_{env}[n]$,设 $t[n]$ 是 $x_{deci}[n]$ 经过截止频率为20Hz的低通滤波器的输出,则快充慢放包络检波器输出为

$$x_{env}[n] < t[n], \partial = 0.6$$

$$x_{env}[n+1] = (1-\partial) * x_{env}[n] + \partial * t[n], \quad x_{env}[n] \geq t[n], \partial = 0.9。$$

[0038] 如图5所示,对 $x_{env}[n]$ 进行一阶差分、半波整流、加权补偿突出包络中的峰值点,输出 $x_s[n]$ 。

[0039] 如图2图3所示,利用基于动态阈值的波峰检测器检测 $x_s[n]$,输出所有可能的节拍点 $E[i]$, $E[i]$ 保存的是可能的节拍点的位置,对 $x_{deci}[n]$ 调用自相关函数,输出自相关序列。对自相关序列 $ac[n]$ 调用波峰检测器,找出所有的波峰 $E_{acPeak}[n]$,每个 $E_{acPeak}[n]$ 保存可能的节拍值,从 $ac[n]$ 中可以估算出多个bpm的值,可以参考soundtouch实现,bmp就是节拍周期,找到了多个bmp的值,根据 $E[i]$,我们可以选出一条最适合的节奏点。利用 $E_{acPeak}[n]$ 和 $E[n]$ 通过多路径搜索算法获取最优的节拍路径的位置。

[0040] 本领域技术人员知道,除了以纯计算机可读程序代码方式实现本发明提供的系统、装置及其各个模块以外,完全可以通过将方法步骤进行逻辑编程来使得本发明提供的系统、装置及其各个模块以逻辑门、开关、专用集成电路、可编程逻辑控制器以及嵌入式微控制器等的形式来实现相同程序。所以,本发明提供的系统、装置及其各个模块可以被认为是—种硬件部件,而对其内包括的用于实现各种程序的模块也可以视为硬件部件内的结构;也可以将用于实现各种功能的模块视为既可以是实现方法的软件程序又可以是硬件部件内的结构。

[0041] 以上对本发明的具体实施例进行了描述。需要理解的是,本发明并不局限于上述

特定实施方式,本领域技术人员可以在权利要求的范围内做出各种变化或修改,这并不影响本发明的实质内容。在不冲突的情况下,本申请的实施例和实施例中的特征可以任意相互组合。

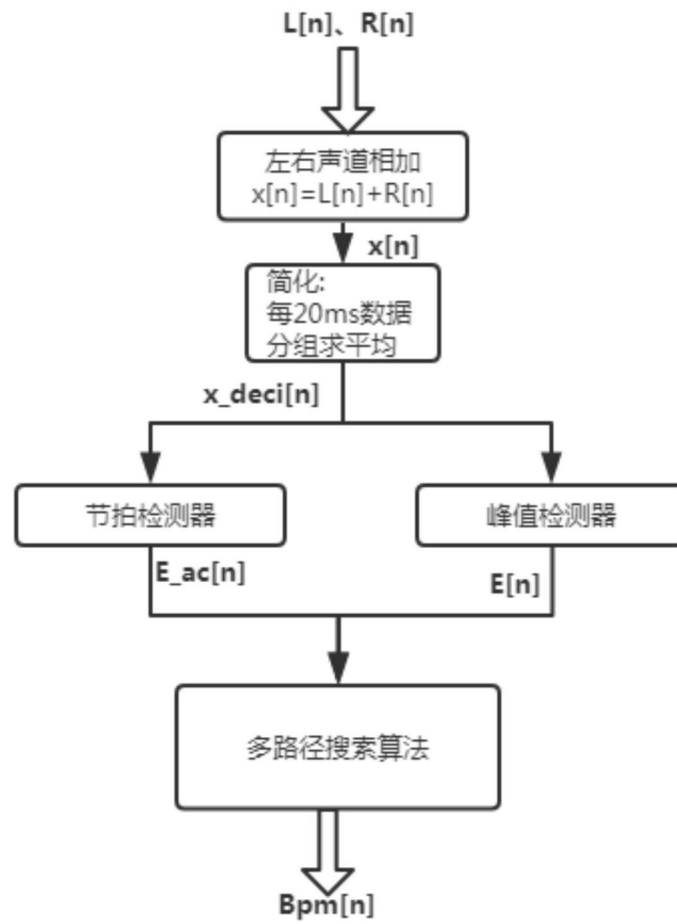


图1

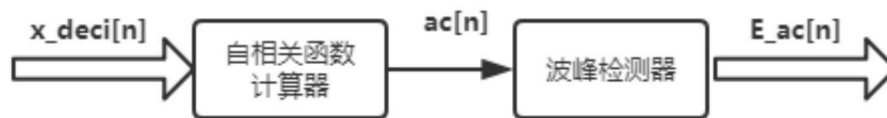


图2

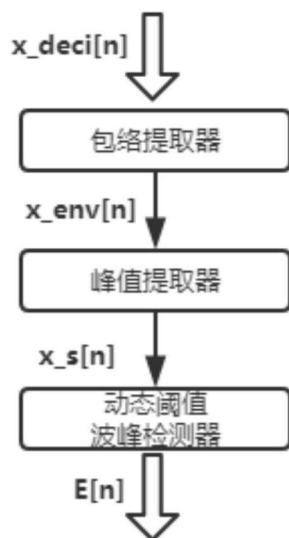


图3

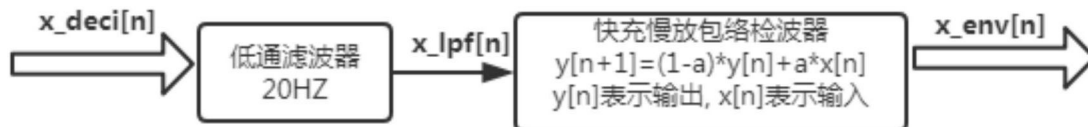


图4

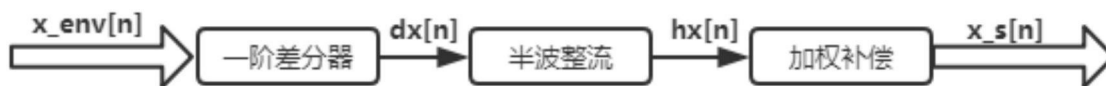


图5