



(12)发明专利

(10)授权公告号 CN 105280196 B

(45)授权公告日 2019.03.19

(21)申请号 201510810644.1

G10L 25/51(2013.01)

(22)申请日 2015.11.19

(56)对比文件

(65)同一申请的已公布的文献号
申请公布号 CN 105280196 A

US 2009/0151544 A1,2009.06.18,
CN 103824555 A,2014.05.28,
CN 102568454 A,2012.07.11,
CN 102903357 A,2013.01.30,
CN 102073636 A,2011.05.25,
CN 104282322 A,2015.01.14,
US 2009/0012637 A1,2009.01.08,

(43)申请公布日 2016.01.27

(73)专利权人 科大讯飞股份有限公司
地址 230088 安徽省合肥市高新开发区望江西路666号

审查员 李春雨

(72)发明人 李飞吾 杨溥 潘青华

(74)专利代理机构 北京维澳专利代理有限公司
11252
代理人 刘路尧 逢京喜

(51)Int.Cl.

G10L 25/18(2013.01)
G10L 25/21(2013.01)

权利要求书3页 说明书9页 附图7页

(54)发明名称

副歌检测方法及系统

(57)摘要

本发明公开了一种副歌检测方法及系统,该方法包括:接收待检测的音频数据;提取所述音频数据的检测特征,所述检测特征包括所述音频数据不同频段的频域能量包络的极值点的频率值及能量值;利用提取的检测特征进行副歌检测,得到检测结果。利用本发明,可以有效提高副歌检测的准确度。



1. 一种副歌检测方法,其特征在于,包括:
 - 接收待检测的音频数据;
 - 提取所述音频数据的检测特征,所述检测特征包括所述音频数据不同频段的频域能量包络的极值点的频率值及能量值,其中所述频率值还用于从极值点中筛选出最终参与检测的极值点;
 - 利用提取的检测特征进行副歌检测,得到检测结果。
2. 根据权利要求1所述的方法,其特征在于,所述提取所述音频数据的检测特征包括:
 - 确定所述音频数据的频域能量包络;
 - 获取所述频域能量包络的局部极值点;
 - 从所述局部极值点中选出表征所述检测特征的极值点;
 - 将选出的极值点对应的频率值及能量值作为当前帧音频数据的检测特征。
3. 根据权利要求2所述的方法,其特征在于,所述确定所述音频数据的频域能量包络包括:
 - 对所述音频数据进行频域转换;
 - 计算频域转换后的音频数据每个频点的能量值;
 - 依次连接每个频点的能量值形成所述音频数据的频域能量包络。
4. 根据权利要求2所述的方法,其特征在于,所述获取所述频域能量包络的局部极值点包括:
 - 将所述频域能量包络的整个频率区间划分为若干子区间,并获取每个子区间内的极值点;或者
 - 确定与所述频域能量包络对应的曲线函数,对所述曲线函数求导得到所述频域能量包络的局部极值点。
5. 根据权利要求2所述的方法,其特征在于,所述从所述局部极值点中选出表征所述检测特征的极值点包括:
 - 将低频至高频之间的所述局部极值点作为候选极值点;
 - 对所述候选极值点中能量高于设定值的极值点频率对应倍频邻近区域的能量值进行弱化处理;
 - 删除能量低于设定能量阈值的候选极值点;
 - 检查是否存在相近的候选极值点,所述相近的候选极值点是指频率差值小于设定值的两个候选极值点;
 - 如果有,则删除其中能量较小的候选极值点,并将剩余的候选极值点作为表征所述检测特征的极值点。
6. 根据权利要求1至5任一项所述的方法,其特征在于,所述利用提取的检测特征进行副歌检测,得到检测结果包括:
 - 依次匹配各帧音频数据,匹配过程包括:
 - 查找其它各帧音频数据检测特征中与当前帧音频数据检测特征中各极值点相匹配的极值点作为所述极值点的匹配极值点;
 - 根据当前帧音频数据检测特征中各极值点的能量值和频率值、以及所述匹配极值点的能量值和频率值,计算当前帧音频数据与各匹配极值点所在帧音频数据的匹配得分;

匹配过程完成后,依次检测各帧音频数据,检测过程包括:

将与当前帧音频数据匹配得分最低的帧作为检测起始帧;

如果所述当前帧音频数据之后连续多帧音频数据分别与所述检测起始帧后相对应的连续多帧音频数据的匹配得分均为最低,则确定当前帧音频数据及其后连续多帧音频数据、以及所述检测起始帧及其后连续多帧音频数据为副歌数据;

记录所述副歌数据的开始时间和结束时间。

7. 一种副歌检测系统,其特征在于,包括:

接收模块,用于接收待检测的音频数据;

检测特征提取模块,用于提取所述音频数据的检测特征,所述检测特征包括所述音频数据不同频段的频域能量包络的极值点的频率值及能量值,其中所述频率值还用于从极值点中筛选出最终参与检测的极值点;

检测模块,用于利用提取的检测特征进行副歌检测,得到检测结果。

8. 根据权利要求7所述的系统,其特征在于,所述检测特征提取模块包括:

包络确定单元,用于确定所述音频数据的频域能量包络;

局部极值点获取单元,用于获取所述频域能量包络的局部极值点;

选择单元,用于从所述局部极值点中选出表征所述检测特征的极值点;

特征获取单元,用于将选出的极值点对应的频率值及能量值作为当前帧音频数据的检测特征。

9. 根据权利要求8所述的系统,其特征在于,所述包络确定单元包括:

频域转换子单元,用于对所述音频数据进行频域转换;

能量值计算子单元,用于计算频域转换后的音频数据每个频点的能量值;

包络生成子单元,用于依次连接每个频点的能量值形成所述音频数据的频域能量包络。

10. 根据权利要求8所述的系统,其特征在于,

所述局部极值点获取单元,具体用于将所述频域能量包络的整个频率区间划分为若干子区间,并获取每个子区间内的极值点;或者确定与所述频域能量包络对应的曲线函数,对所述曲线函数求导得到所述频域能量包络的局部极值点。

11. 根据权利要求8所述的系统,其特征在于,所述选择单元包括:

候选极值点确定子单元,用于将低频至高频之间的所述局部极值点作为候选极值点;

弱化处理子单元,用于对所述候选极值点中能量高于设定值的极值点频率对应倍频邻近区域的能量值进行弱化处理;

筛选子单元,用于删除能量低于设定能量阈值的候选极值点,然后检查是否存在相近的候选极值点,所述相近的候选极值点是指频率差值小于设定值的两个候选极值点;如果有,则删除其中能量较小的候选极值点,并将剩余的候选极值点作为表征所述检测特征的极值点。

12. 根据权利要求8至11任一项所述的系统,其特征在于,所述检测模块包括:

匹配单元,用于依次匹配各帧音频数据,匹配过程包括:查找其它各帧音频数据检测特征中与当前帧音频数据检测特征中各极值点相匹配的极值点作为所述极值点的匹配极值点;根据当前帧音频数据检测特征中各极值点的能量值和频率值、以及所述匹配极值点的

能量值和频率值,计算当前帧音频数据与各匹配极值点所在帧音频数据的匹配得分;

检测单元,用于在匹配过程完成后,依次检测各帧音频数据,检测过程包括:将与当前帧音频数据匹配得分最低的帧作为检测起始帧;如果所述当前帧音频数据之后连续多帧音频数据分别与所述检测起始帧后相对应的连续多帧音频数据的匹配得分均为最低,则确定当前帧音频数据及其后连续多帧音频数据、以及所述检测起始帧及其后连续多帧音频数据为副歌数据,并记录所述副歌数据的开始时间和结束时间。

副歌检测方法及系统

技术领域

[0001] 本发明涉及语音信号处理领域,具体涉及一种副歌检测方法及系统。

背景技术

[0002] 流行歌曲一般来说结构简单,形式活泼,易于传唱。从作曲和编曲的角度,流行歌曲通常包含以下几个部分:主歌、副歌、桥段。主歌可以说是内容,是每首音乐的主干,副歌是指内容区别于主歌、发展与概括性比较强的且在节奏情感曲调上与主歌形成对比的段落。副歌是歌曲感情最强烈抒发的部分,乐器使用多且音量较大、饱满,概括性很强,并且对于每一首流行歌曲,其副歌都具有至少2次的旋律重复,因此它通常是每首流行歌曲中最先为听众熟知的部分。随着数字音乐行业的发展,音乐内容大量增长,越来越多的用户喜欢直接试听副歌,来选择自己喜欢的歌曲,或者将歌曲的副歌截取出来作为手机铃声或背景音。因此,如果直接将歌曲的副歌位置检测出来展现给用户,用户直接从该位置试听歌曲,无疑会给用户带来很大的便利。

[0003] 现有的副歌检测方法一般是根据歌曲歌词的相似性检测副歌,这种方法太过依赖歌词,当歌曲中的副歌歌词不相同,现有方法无法检测出副歌所在位置。另外,在实际的歌曲中,副歌前的主歌的歌词也有可能相同,如下面为《容易受伤的女人》的两段主歌及副歌的歌词,其中第一段为第一列时间点对应的歌词,第二段为第二列时间点对应的歌词:

[0004] [00:32.44] [02:14.15] 曾被破碎过的心(主歌部分)

[0005] [00:35.88] [02:16.15] 让你今天轻轻贴近

[0006] [00:39.96] [02:21.40] 多少安慰及疑问

[0007] [00:42.21] [02:25.01] 偷偷的再生

[0008] [00:46.56] [02:28.24] 情难自禁(副歌开始)

[0009] [00:48.24] [02:29.92] 我却其实属于

[0010] [00:50.06] [02:31.72] 极度容易受伤的女人

[0011] 可以看出,两段副歌前的主歌歌词完全相同。在这种情况下,容易将主歌也作为副歌检测出来,因此,现有根据歌词的相似性来检测副歌的准确度较低。

发明内容

[0012] 本发明提供一种副歌检测方法及系统,以提高副歌检测的准确度。

[0013] 为此,本发明提供如下技术方案:

[0014] 一种副歌检测方法,包括:

[0015] 接收待检测的音频数据;

[0016] 提取所述音频数据的检测特征,所述检测特征包括所述音频数据不同频段的频域能量包络的极值点的频率值及能量值;

[0017] 利用提取的检测特征进行副歌检测,得到检测结果。

[0018] 优选地,所述提取所述音频数据的检测特征包括:

- [0019] 确定所述音频数据的频域能量包络；
- [0020] 获取所述频域能量包络的局部极值点；
- [0021] 从所述局部极值点中选出表征所述检测特征的极值点；
- [0022] 将选出的极值点对应的频率值及能量值作为当前帧音频数据的检测特征。
- [0023] 优选地,所述确定所述音频数据的频域能量包络包括:
- [0024] 对所述音频数据进行频域转换;
- [0025] 计算频域转换后的音频数据每个频点的能量值;
- [0026] 依次连接每个频点的能量值形成所述音频数据的频域能量包络。
- [0027] 优选地,所述获取所述频域能量包络的局部极值点包括:
- [0028] 将所述频域能量包络的整个频率区间划分为若干子区间,并获取每个子区间内的极值点;或者
- [0029] 确定与所述频域能量包络对应的曲线函数,对所述曲线函数求导得到所述频域能量包络的局部极值点。
- [0030] 优选地,所述从所述局部极值点中选出表征所述检测特征的极值点包括:
- [0031] 将低频至高频之间的所述局部极值点作为候选极值点;
- [0032] 对所述候选极值点中能量高于设定值的极值点频率对应倍频邻近区域的能量值进行弱化处理;
- [0033] 删除能量低于设定能量阈值的候选极值点;
- [0034] 检查是否存在相近的候选极值点,所述相近的候选极值点是指频率差值小于设定值的两个候选极值点;
- [0035] 如果有,则删除其中能量较小的候选极值点,并将剩余的候选极值点作为表征所述检测特征的极值点。
- [0036] 优选地,所述利用提取的检测特征进行副歌检测,得到检测结果包括:
- [0037] 依次匹配各帧音频数据,匹配过程包括:
- [0038] 查找其它各帧音频数据检测特征中与当前帧音频数据检测特征中各极值点相匹配的极值点作为所述极值点的匹配极值点;
- [0039] 根据当前帧音频数据检测特征中各极值点的能量值和频率值、以及所述匹配极值点的能量值和频率值,计算当前帧音频数据与各匹配极值点所在帧音频数据的匹配得分;
- [0040] 匹配过程完成后,依次检测各帧音频数据,检测过程包括:
- [0041] 将与当前帧音频数据匹配得分最低的帧作为检测起始帧;
- [0042] 如果所述当前帧音频数据之后连续多帧音频数据分别与所述检测起始帧后相对应的连续多帧音频数据的匹配得分均为最低,则确定当前帧音频数据及其后连续多帧音频数据、以及所述检测起始帧及其后连续多帧音频数据为副歌数据;
- [0043] 记录所述副歌数据的开始时间和结束时间。
- [0044] 一种副歌检测系统,包括:
- [0045] 接收模块,用于接收待检测的音频数据;
- [0046] 检测特征提取模块,用于提取所述音频数据的检测特征,所述检测特征包括所述音频数据不同频段的频域能量包络的极值点的频率值及能量值;
- [0047] 检测模块,用于利用提取的检测特征进行副歌检测,得到检测结果。

- [0048] 优选地,所述检测特征提取模块包括:
- [0049] 包络确定单元,用于确定所述音频数据的频域能量包络;
- [0050] 局部极值点获取单元,用于获取所述频域能量包络的局部极值点;
- [0051] 选择单元,用于从所述局部极值点中选出表征所述检测特征的极值点;
- [0052] 特征获取单元,用于将选出的极值点对应的频率值及能量值作为当前帧音频数据的检测特征。
- [0053] 优选地,所述包络确定单元包括:
- [0054] 频域转换子单元,用于对所述音频数据进行频域转换;
- [0055] 能量值计算子单元,用于计算频域转换后的音频数据每个频点的能量值;
- [0056] 包络生成子单元,用于依次连接每个频点的能量值形成所述音频数据的频域能量包络。
- [0057] 优选地,所述局部极值点获取单元,具体用于将所述频域能量包络的整个频率区间划分为若干子区间,并获取每个子区间内的极值点;或者确定与所述频域能量包络对应的曲线函数,对所述曲线函数求导得到所述频域能量包络的局部极值点。
- [0058] 优选地,所述选择单元包括:
- [0059] 候选极值点确定子单元,用于将低频至高频之间的所述局部极值点作为候选极值点;
- [0060] 弱化处理子单元,用于对所述候选极值点中能量高于设定值的极值点频率对应倍频邻近区域的能量值进行弱化处理;
- [0061] 筛选子单元,用于删除能量低于设定能量阈值的候选极值点,然后检查是否存在相近的候选极值点,所述相近的候选极值点是指频率差值小于设定值的两个候选极值点;如果有,则删除其中能量较小的候选极值点,并将剩余的候选极值点作为表征所述检测特征的极值点。
- [0062] 优选地,所述检测模块包括:
- [0063] 匹配单元,用于依次匹配各帧音频数据,匹配过程包括:查找其它各帧音频数据检测特征中与当前帧音频数据检测特征中各极值点相匹配的极值点作为所述极值点的匹配极值点;根据当前帧音频数据检测特征中各极值点的能量值和频率值、以及所述匹配极值点的能量值和频率值,计算当前帧音频数据与各匹配极值点所在帧音频数据的匹配得分;
- [0064] 检测单元,用于在匹配过程完成后,依次检测各帧音频数据,检测过程包括:将与当前帧音频数据匹配得分最低的帧作为检测起始帧;如果所述当前帧音频数据之后连续多帧音频数据分别与所述检测起始帧后相对应的连续多帧音频数据的匹配得分均为最低,则确定当前帧音频数据及其后连续多帧音频数据、以及所述检测起始帧及其后连续多帧音频数据为副歌数据,并记录所述副歌数据的开始时间和结束时间。
- [0065] 本发明实施例提供的副歌检测方法及系统,对于待检测的音频数据,提取其检测特征,所述检测特征用该音频数据不同频段的频域能量包络的极值点来表征,从而可以全面表达该音频数据不同频段的信息,利用该检测特征进行副歌检测,可以更准确地捕捉不同频段的音乐信息,提高副歌检测的准确度。

附图说明

[0066] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例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中记载的一些实施例,对于本领域普通技术人员来讲,还可以根据这些附图获得其他的附图。

[0067] 图1是本发明实施例副歌检测方法的流程图;

[0068] 图2是本发明实施例中提取检测特征的流程图;

[0069] 图3是本发明实施例中音频数据的频域能量包络示意图;

[0070] 图4是本发明实施例中音频数据频域能量包络的局部极值点示例;

[0071] 图5是本发明实施例中候选极值点的选择示例;

[0072] 图6是本发明实施例中对倍频附近能量值进行弱化处理示意图;

[0073] 图7是本发明实施例中选出的表征检测特征的极值点示意图;

[0074] 图8是本发明实施例中不同帧音频数据进行匹配的流程图;

[0075] 图9是本发明实施例中对各帧音频数据进行检测的流程图;

[0076] 图10是本发明实施例副歌检测系统的一种结构示意图。

具体实施方式

[0077] 为了使本技术领域的人员更好地理解本发明实施例的方案,下面结合附图和实施方式对本发明实施例作进一步的详细说明。

[0078] 如图1所示,是本发明实施例副歌检测方法的流程图,包括以下步骤:

[0079] 步骤101,接收待检测的音频数据。

[0080] 所述音频数据通常是包含伴奏与人声的歌曲数据。

[0081] 步骤102,提取所述音频数据的检测特征,所述检测特征包括所述音频数据不同频段的频域能量包络的极值点的频率值及能量值。

[0082] 需要说明的是,与通常的音频数据处理中以帧为单位类似,对于本发明实施例中的音频数据,在提取检测特征之前,同样需要先对音频数据进行分帧,然后以帧为单位,提取每帧数据的检测特征。

[0083] 检测特征的提取过程需要首先计算待检测音频数据的频域能量,得到音频数据的频域能量包络,然后对所述频域能量包络进行局部极值点检测,得到频域能量包络的局部极值点,然后再从这些局部极值点中提取表征所述检测特征的极值点,将提取的这些极值点的频率值及能量值作为所述检测特征,具体提取过程将在后面详细描述。

[0084] 步骤103,利用提取的检测特征进行副歌检测,得到检测结果。

[0085] 在进行副歌检测时,可以根据不同帧中相匹配的极值点的能量值来计算每帧音频数据与各匹配极值点所在帧音频数据的匹配得分,通过匹配得分的分布情况,可以确定哪些帧音频数据为副歌数据,进而根据这些帧的开始和结束时间即可得到所述副歌数据的开始时间和结束时间,记录这些时间点,也就记录了副歌数据在整个音频数据中的位置。副歌检测包括匹配过程及检测过程,在匹配过程中,需要依次匹配各帧音频数据,也就是说,分别计算各帧音频数据与其它帧音频数据的匹配得分;匹配过程完成后,进行检测过程,在检测过程中,需要依次检测各帧音频数据的匹配得分分布情况,根据副歌在一首完整的歌曲中具有旋律重复的特点,确定副歌对应的各帧音频数据,从而可以确定副歌在所述待检测

音频数据中的位置。具体的匹配过程及检测过程将在后面详细描述。

[0086] 在实际应用中,可以根据检测到的副歌的位置,提醒用户副歌的开始和结束位置,方便用户试听。当然,也可以将检测到的副歌数据提取出来,将其作为铃声、背景音等。

[0087] 如图2所示,是本发明实施例中提取检测特征的流程图,包括以下步骤:

[0088] 步骤201,确定音频数据的频域能量包络。

[0089] 具体地,首先对所述音频数据进行频域转换,具体频域转换方法可以采用现有的一些方法,如傅立叶变换等。将所述音频数据转换到频域后,计算音频数据每个频点的能量值,然后再依次连接每个频点的能量值形成的曲线即为音频数据的频域能量包络。如图3所示,为一段采样率为8kHz音频数据的频域能量包络,其中,横坐标为音频数据的频率值,纵坐标为每个频率点音频数据的能量值。

[0090] 步骤202,获取所述频域能量包络的局部极值点。

[0091] 所述局部极值点是指在一定频率范围内的极值点。局部极值点的检测可以采用多种方式,比如:

[0092] (1) 将所述频域能量包络的整个频率区间划分为若干子区间,通过比较每个子区间内能量值的大小,获取每个子区间内的极值点,即可得到整个频率范围内的局部极值点。

[0093] (2) 确定与所述频域能量包络对应的曲线函数,比如使用多项式或样条拟合方法得到包络曲线函数,然后对所述曲线函数求导得到所述频域能量包络的局部极值点。

[0094] 如图4所示,是音频数据频域能量包络的局部极值点示例,其中,标注圆圈的点为检测到的局部极值点。

[0095] 步骤203,从所述局部极值点中选出表征所述检测特征的极值点。

[0096] 由于歌曲的极低频(0Hz-40Hz)和低频(40Hz-100Hz)部分大多为低音或重低乐器的声音,极值点较多,不具有参考性;而歌曲的高频(2560Hz-5120Hz)和极高频(5120Hz-20000Hz)部分大多为各种乐器的泛音,不具有辨识度,很难检测到极值点。因此,将低频以下的频域能量包络和高频以上的频域能量包络的极值点切除。将低频到高频之间的极值点作为候选极值点,低频和高频之间的具体候选极值点的频率范围可以根据实际应用需求进行定制,如频率范围为[187Hz,2515.625Hz],如图5为候选极值点的选择示例,其中,两条竖线之内的极值点为候选极值点,频率范围为[187Hz,2515.625Hz]。

[0097] 另外,如果能量过高极值点作为检测特征,则该极值点对应频率的倍频附近也容易产生与能量过高极值点特征相近的干扰极值点,该极值点对副歌检测没有意义,需要删除;然而,如果恰好倍频附近产生的极值点是其它乐器产生的,则该极值点具有代表性,需要选择。因此,为了防止能量过高极值点频率对应倍频附近极值点的干扰,在本发明实施例中,还可进一步将候选极值点中能量过高极值点(比如能量高于设定值的极值点)对应频率的倍频附近(比如以对应频率的倍频为中心的设定频域内)极值点的能量值进行弱化处理,具体处理时,只需要处理候选极值点倍频附近的能量值即可。

[0098] 所述弱化处理方法如使用陷波滤波器,根据当前能量超过阈值的极值点对应的频率,对该频率的所有倍频进行陷波处理,具体陷波处理及陷波程度可根据实际应用设置,如随着倍频频率值的增加,陷波程度降低,即倍频附近的能量值弱化程度降低。由图5可知,390Hz处的极值点能量值过高,超过预先设定的阈值,需要对390Hz对应倍频附近的能量值进行弱化处理,所述倍频如780Hz,1560Hz,对倍频附近的能量值进行弱化处理后,780Hz和

1560Hz附近的能量值明显降低了,如图6所示,尤其是781Hz和1578Hz处极值点的能量值降低较明显。

[0099] 进行上述弱化处理后,在选择表征检测特征的极值点时,对这些候选极值点,依次(比如频率由低至高,或频率由高至低)进行检查,从中选出用于表征检测特征的极值点。具体可以按以下原则进行:首先删除其中能量低于设定能量阈值的候选极值点;然后检查是否存在相近的候选极值点,所述相近的候选极值点是指频率差值小于设定值的两个候选极值点;如果有,则删除其中能量较小的候选极值点,并将剩余的候选极值点作为表征所述检测特征的极值点。

[0100] 步骤204,将选出的极值点对应的频率值及能量值作为当前帧音频数据的检测特征。

[0101] 如图7所示,为选出的表征检测特征的极值点示意图。其中,圆圈表示表征检测特征的极值点,三角形表示从候选极值点删除的极值点。

[0102] 通过上述过程,可以得到整个音频数据每帧的检测特征,每帧的检测特征可以由多个极值点表征。

[0103] 前面提到,在利用提取的检测特征进行副歌检测时,包括两个过程,即匹配过程及检测过程,下面分别结合附图对这两个过程进行详细说明。

[0104] 在匹配过程中,需要依次匹配各帧音频数据与其它帧音频数据,对于每帧音频数据,其匹配流程如图8所示,包括以下步骤:

[0105] 步骤801,提取当前帧音频数据检测特征中各极值点。

[0106] 步骤802,查找其它各帧音频数据检测特征中与当前帧音频数据检测特征中各极值点相匹配的极值点作为所述极值点的匹配极值点。

[0107] 具体地,依次将当前帧音频数据检测特征中各极值点作为当前极值点,查找其它各帧音频数据检测特征中的极值点频率与所述当前极值点频率差值及能量差值均小于对应的阈值的极值点,如果有满足该要求的极值点,则认为该极值点与所述当前极值点相匹配。为了描述方便,将该极值点称为当前极值点的匹配极值点。

[0108] 例如:使用 $[F_p, M_p]$ 表示当前帧音频数据的第 p 个极值点的频率值和能量值,使用 $[F_{iq}, M_{iq}]$ 表示整个音频数据中第 i 帧音频数据的检测特征中第 q 个极值点的频率值和能量值,如果 F_p 与 F_{iq} 的差值小于设定的频率差阈值,并且 M_p 与 M_{iq} 的差值小于设定的能量差阈值,则将第 i 帧音频数据的检测特征中第 q 个极值点作为当前帧音频数据第 p 个极值点的匹配极值点,依次查找到整个音频数据中除当前帧音频数据之外的所有的匹配极值点。

[0109] 步骤803,根据当前帧音频数据检测特征中各极值点的能量值和频率值、以及所述匹配极值点的能量值和频率值,计算当前帧音频数据与各匹配极值点所在帧音频数据的匹配得分。

[0110] 具体地,依次将各匹配极值点所在帧音频数据作为待匹配帧,计算当前帧音频数据检测特征中各极值点的能量值与所述待匹配帧检测特征中各匹配极值点的能量值的差值,并对这些差值加权得到当前帧音频数据与待匹配帧的匹配得分,具体计算方法如式(1)所示:

$$[0111] \quad s_i = \sum_{p=1}^{N_p} \alpha_p (|M_p - M_{pi}|) \quad (1)$$

[0112] 其中, s_i 表示当前帧音频数据相比于第*i*帧音频数据的匹配得分, M_p 表示当前帧音频数据检测特征中的第*p*个极值点的能量值, M_{pi} 表示第*i*帧待匹配音频数据检测特征中与 M_p 匹配极值点的能量值, N_p 表示当前帧音频数据检测特征中与第*i*帧待匹配音频数据检测特征中相匹配的极值点数, α_p 表示当前帧音频数据检测特征中第*p*个极值点能量差值系数, 具体计算方法如式(2)所示:

$$[0113] \quad \alpha_p = kF_p^{\frac{c}{1+N_{max}-N_p}} \quad (2)$$

[0114] 其中, N_{max} 表示匹配极值点数 N_p 的最大值, F_p 为当前帧音频数据检测特征中的第*p*个极值点的频率值, c 为常数, 取值范围为0到1之间, 具体根据实验结果取值, 如取0.4, k 为常数, 具体可以根据实验结果或实际应用取值。

[0115] 由于一般有用信息主要集中在中低频, 因此, 可以设置中低频极值点的能量差值系数取值大于高频极值点的能量差值系数取值, 如当前帧音频数据检测特征中有3个极值点, 根据频率由小到大的顺序, 能量差值系数取值依次为0.2、0.15、0.1, 从而使检测特征中中低频极值点相比高频极值点起到更大的作用。所述匹配得分越小, 当前帧音频数据与所比较的第*i*帧待匹配音频数据越相似。

[0116] 需要说明的是, 在匹配过程中, 在当前帧与各待匹配帧匹配完成后进行下一帧匹配时, 如果下一帧已和当前帧或当前帧之前的各帧进行过匹配, 则下一帧只需与未匹配的待匹配帧进行匹配即可, 无需再与已匹配的待匹配帧进行匹配。

[0117] 匹配过程完成后, 执行检测过程, 在检测过程中, 需要依次检测各帧音频数据与其它帧音频数据的匹配得分的分布情况, 以确定副歌数据所在帧, 检测流程如图9所示, 包括以下步骤:

[0118] 步骤901, 将当前帧作为待检测帧。

[0119] 步骤902, 将与待检测帧音频数据匹配得分最低的帧作为检测起始帧。

[0120] 步骤903, 判断待检测帧音频数据之后连续多帧音频数据分别与所述检测起始帧后相对应的连续多帧音频数据的匹配得分是否均为最低。如果是, 则执行步骤904; 否则, 执行步骤906。

[0121] 步骤904, 确定待检测帧音频数据及其后连续多帧音频数据、以及所述检测起始帧及其后连续多帧音频数据为副歌数据。

[0122] 步骤905, 记录所述副歌数据的开始时间和结束时间。

[0123] 步骤906, 判断待检测帧是否为最后一帧。如果是, 则执行步骤908; 否则, 执行步骤907。

[0124] 步骤907, 将下一帧作为待检测帧, 然后返回步骤902。

[0125] 步骤908, 检测过程结束。

[0126] 所述匹配得分均为最低是指: 当前帧音频数据后第一帧与所述检测起始帧后第一帧的匹配得分在所述当前帧音频数据后第一帧与其它各帧的匹配得分中是最低的; 当前帧音频数据后第二帧与所述检测起始帧后第二帧的匹配得分在所述当前帧音频数据后第二帧与其它各帧的匹配得分中也是最低的; 依此类推。如果有连续多帧满足上述条件, 如连续40帧, 则这连续多帧连同其前一帧音频数据为副歌数据, 具体连续多少帧满足上述条件, 可以根据实际应用场景确定。

[0127] 本发明实施例提供的副歌检测方法,对于待检测的音频数据,提取其检测特征,所述检测特征用该音频数据不同频段的频域能量包络的极值点来表征,从而可以全面表达该音频数据不同频段的信息,利用该检测特征进行副歌检测,可以更准确地捕捉不同频段的音乐信息,提高副歌检测的准确度。

[0128] 相应地,本发明实施例还提供一种副歌检测系统,如图10所示,是该系统的一种结构示意图。

[0129] 在该实施例中,所述系统包括:

[0130] 接收模块11,用于接收待检测的音频数据;

[0131] 检测特征提取模块12,用于提取所述音频数据的检测特征,所述检测特征包括所述音频数据不同频段的频域能量包络的极值点的频率值及能量值;

[0132] 检测模块13,用于利用提取的检测特征进行副歌检测,得到检测结果。

[0133] 所述检测特征提取模块12可以根据音频数据的频域能量包络来提取所述音频数据的检测特征,该模块的一种具体结构可以包括以下各单元:

[0134] 包络确定单元,用于确定所述音频数据的频域能量包络;

[0135] 局部极值点获取单元,用于获取所述频域能量包络的局部极值点;

[0136] 选择单元,用于从所述局部极值点中选出表征所述检测特征的极值点;

[0137] 特征获取单元,用于将选出的极值点对应的频率值及能量值作为当前帧音频数据的检测特征。

[0138] 其中,所述包络确定单元可以包括以下各子单元:

[0139] 频域转换子单元,用于对所述音频数据进行频域转换,具体频域转换方法可以采用现有的一些方法,如傅立叶变换等;

[0140] 能量值计算子单元,用于计算频域转换后的音频数据每个频点的能量值,能量值的计算可以采用现有的一些方法,在此不再详述;

[0141] 包络生成子单元,用于依次连接每个频点的能量值形成所述音频数据的频域能量包络。

[0142] 所述局部极值点获取单元具体可以通过将所述频域能量包络的整个频率区间划分为若干子区间,获取每个子区间内的极值点;或者可以通过确定与所述频域能量包络对应的曲线函数,对所述曲线函数求导得到所述频域能量包络的局部极值点。

[0143] 为了进一步减少干扰,使从所述局部极值点中选出的表征所述检测特征的极值点更具代表性,所述选择单元的一种优选结构可以包括以下各子单元:

[0144] 候选极值点确定子单元,用于将低频至高频之间的所述局部极值点作为候选极值点;

[0145] 弱化处理子单元,用于对所述候选极值点中能量高于设定值的极值点频率对应倍频邻近区域的能量值进行弱化处理;

[0146] 筛选子单元,用于删除能量低于设定能量阈值的候选极值点,然后检查是否存在相近的候选极值点,所述相近的候选极值点是指频率差值小于设定值的两个候选极值点;如果有,则删除其中能量较小的候选极值点,并将剩余的候选极值点作为表征所述检测特征的极值点。

[0147] 上述检测模块13具体可以根据不同帧中相匹配的极值点的能量值来计算每帧音

频数据与各匹配极值点所在帧音频数据的匹配得分,通过匹配得分的分布情况,可以确定哪些帧音频数据为副歌数据,进而根据这些帧的开始和结束时间即可得到所述副歌数据的开始时间和结束时间。相应地,检测模块13的一种具体结构可以包括:匹配单元和检测单元,其中:

[0148] 所述匹配单元用于依次匹配各帧音频数据,匹配过程包括:查找其它各帧音频数据检测特征中与当前帧音频数据检测特征中各极值点相匹配的极值点作为所述极值点的匹配极值点;根据当前帧音频数据检测特征中各极值点的能量值和频率值、以及所述匹配极值点的能量值和频率值,计算当前帧音频数据与各匹配极值点所在帧音频数据的匹配得分;

[0149] 所述检测单元用于在匹配过程完成后,依次检测各帧音频数据,检测过程包括:将与当前帧音频数据匹配得分最低的帧作为检测起始帧;如果所述当前帧音频数据之后连续多帧音频数据分别与所述检测起始帧后相对应的连续多帧音频数据的匹配得分均为最低,则确定当前帧音频数据及其后连续多帧音频数据、以及所述检测起始帧及其后连续多帧音频数据为副歌数据,并记录所述副歌数据的开始时间和结束时间。

[0150] 本发明实施例提供的副歌检测系统,对于待检测的音频数据,提取其检测特征,所述检测特征用该音频数据不同频段的频域能量包络的极值点来表征,从而可以全面表达该音频数据不同频段的信息,利用该检测特征进行副歌检测,可以更准确地捕捉不同频段的音乐信息,提高副歌检测的准确度。

[0151] 本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于系统实施例而言,由于其基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。以上所描述的系统实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。

[0152] 以上对本发明实施例进行了详细介绍,本文中应用了具体实施方式对本发明进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及系统;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本发明的限制。

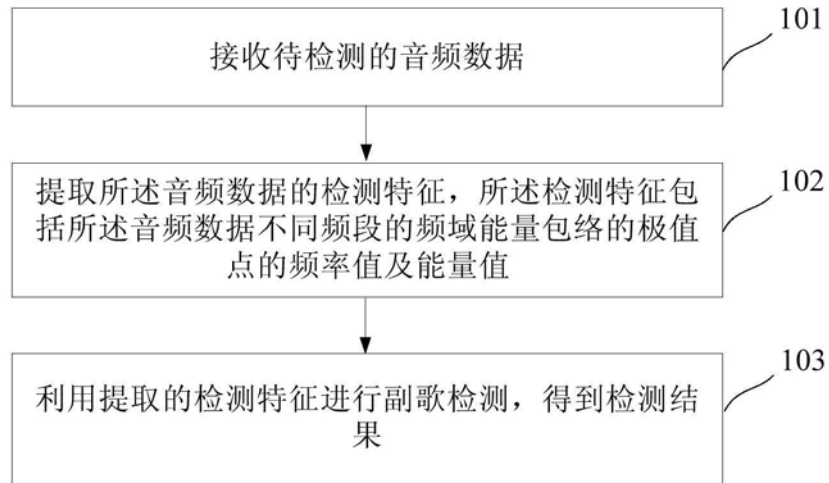


图1

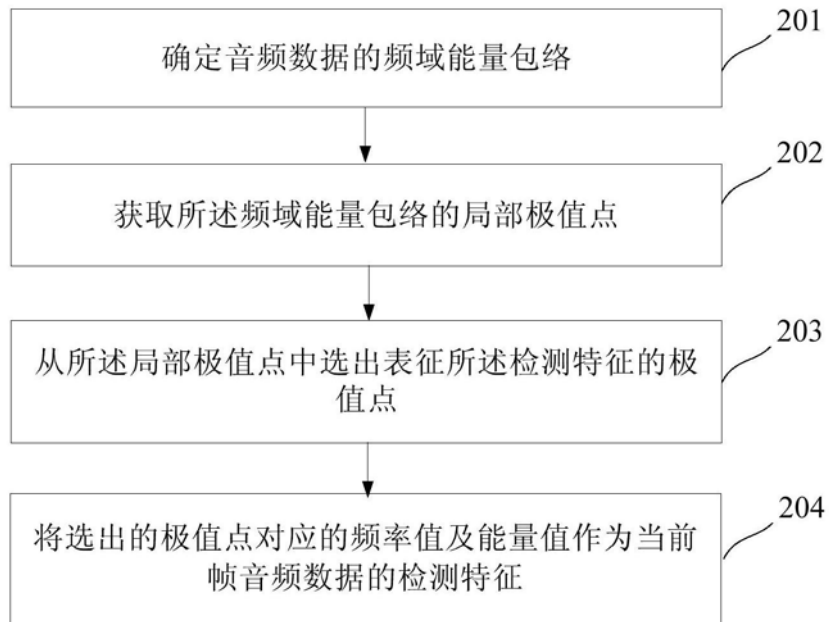


图2

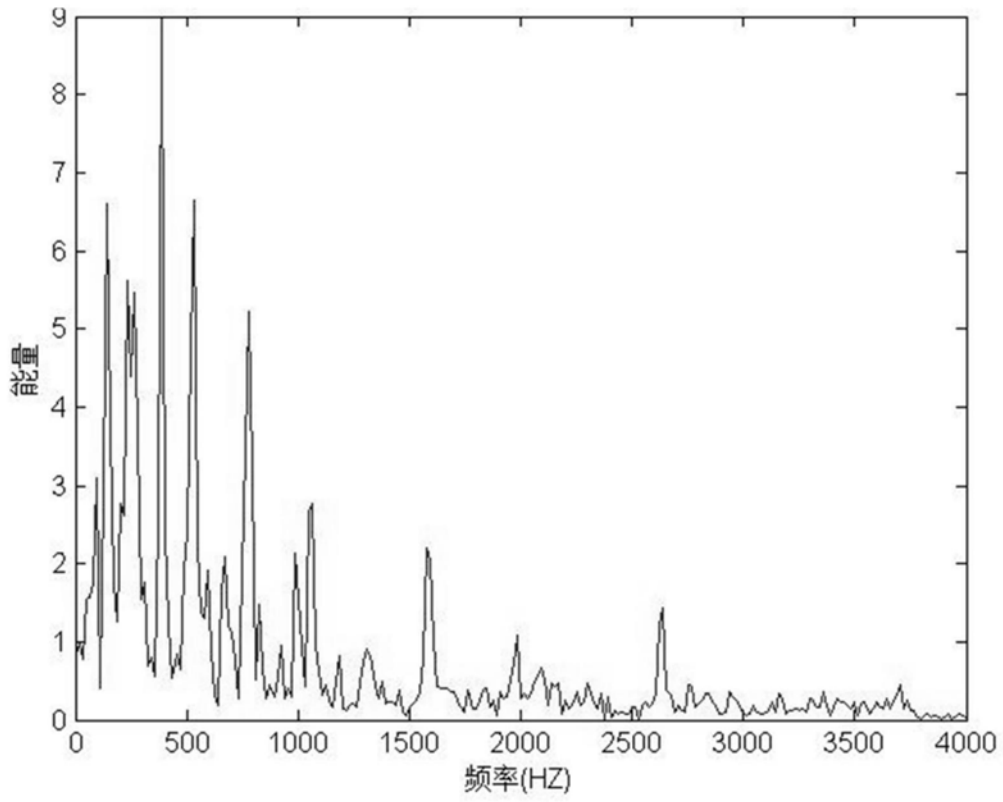


图3

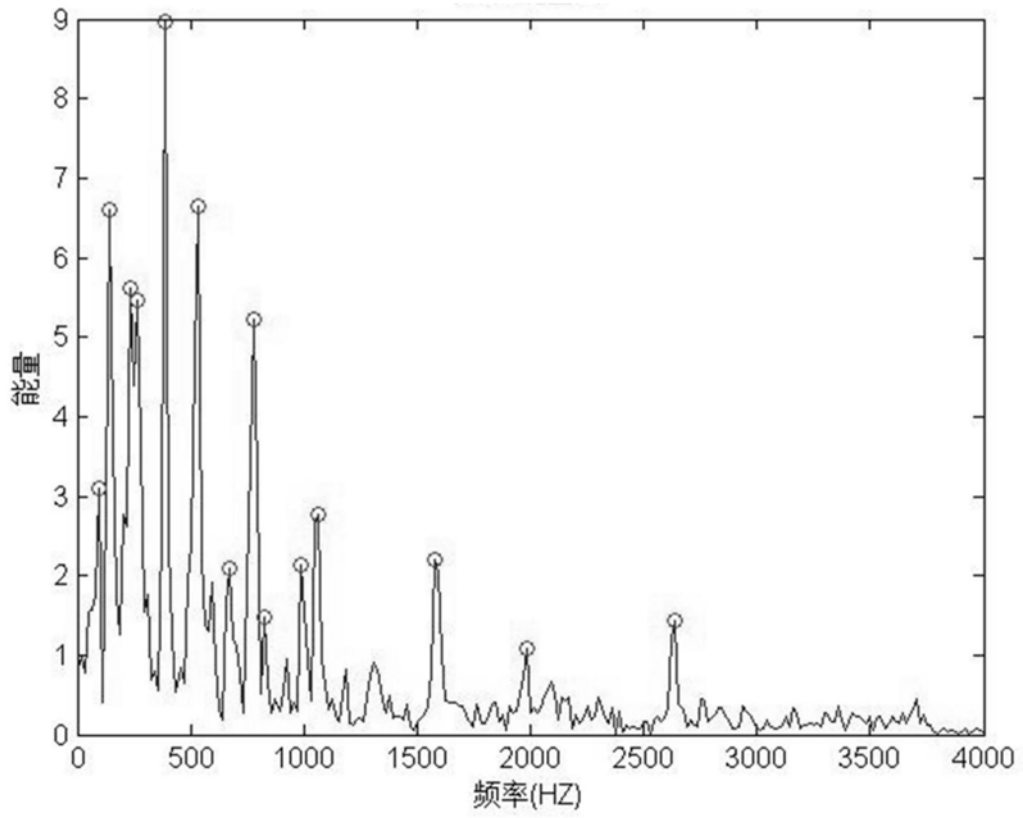


图4

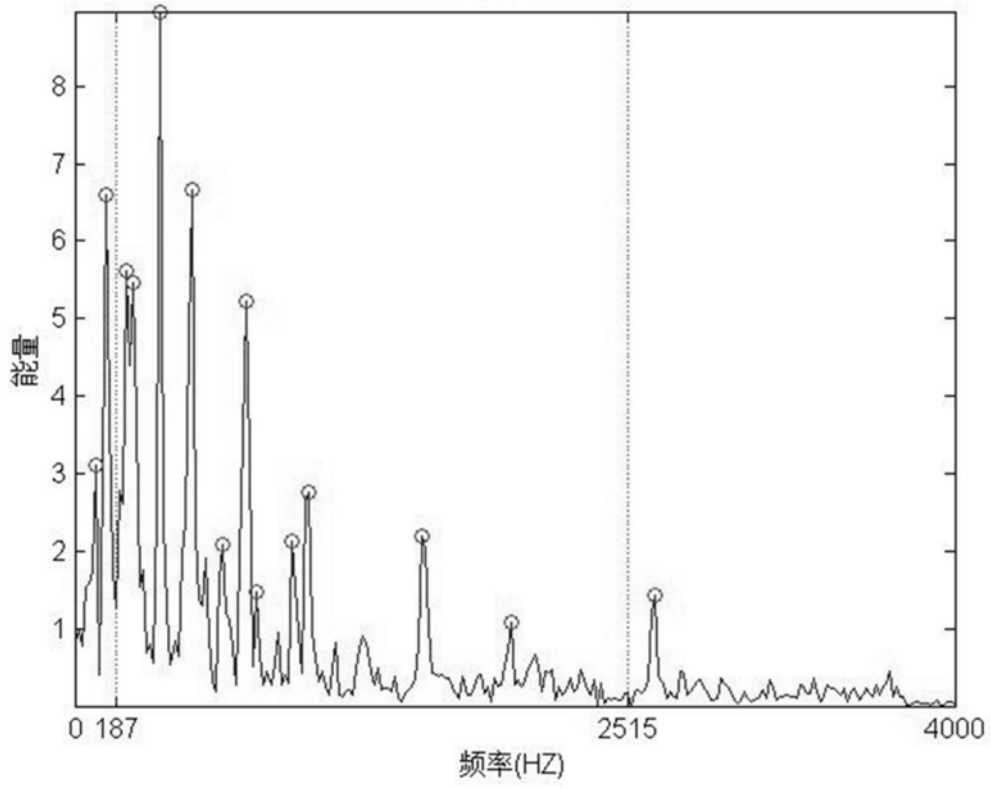


图5

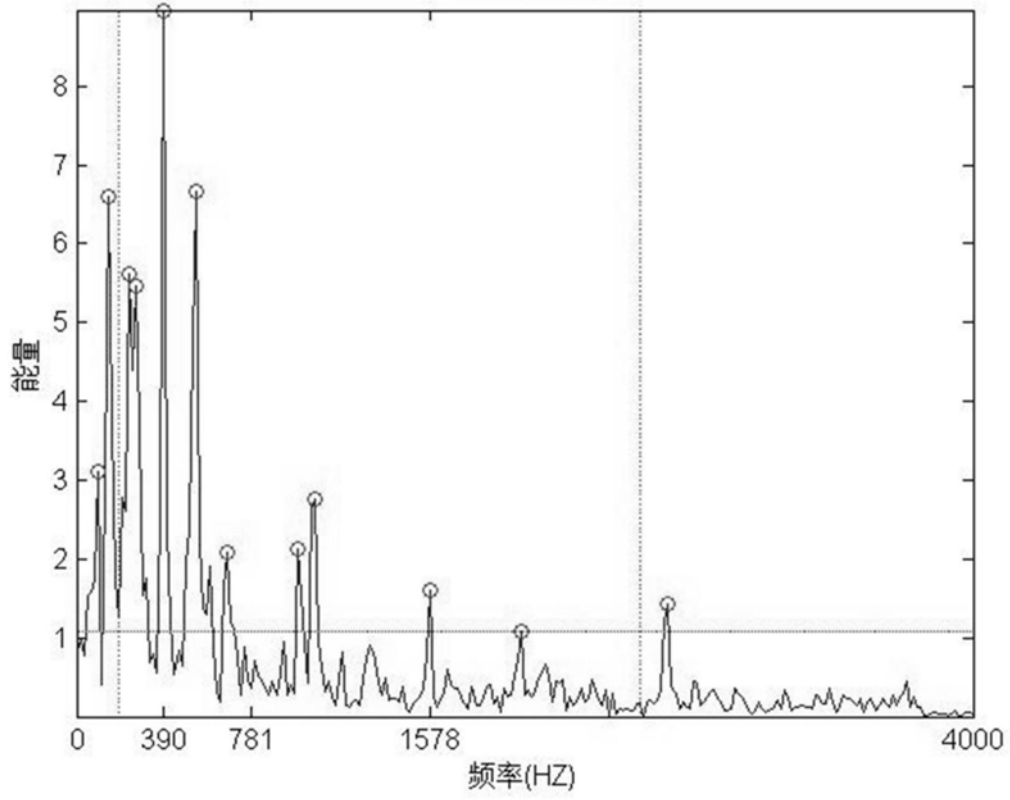


图6

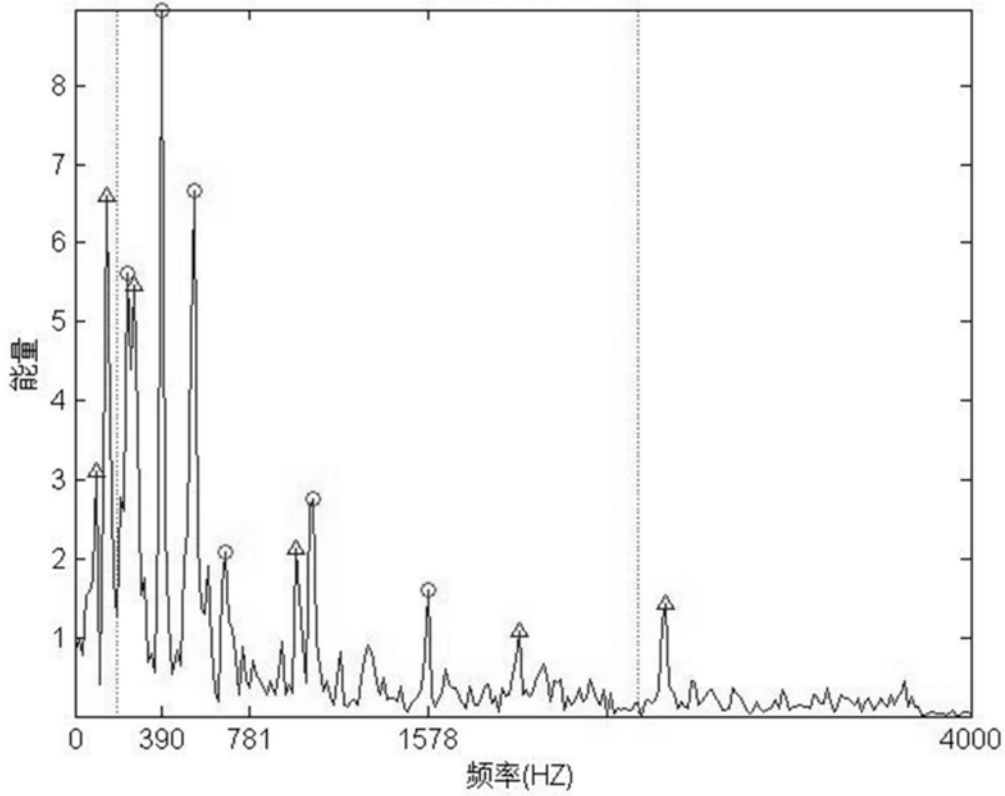


图7

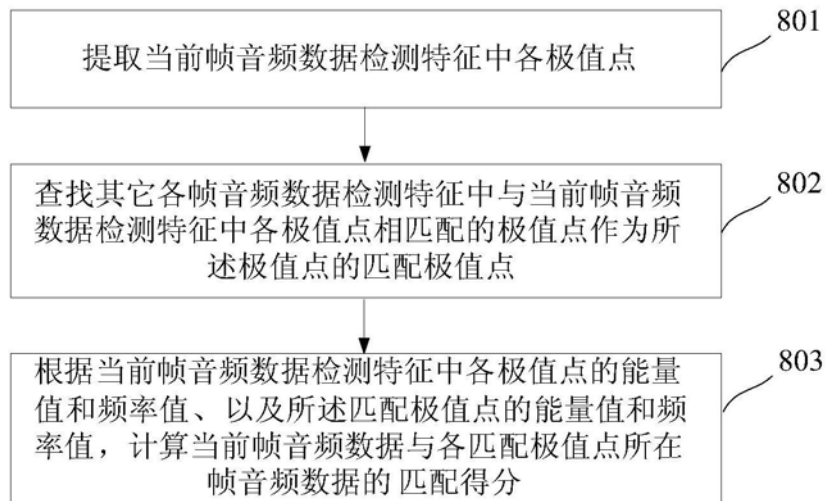


图8

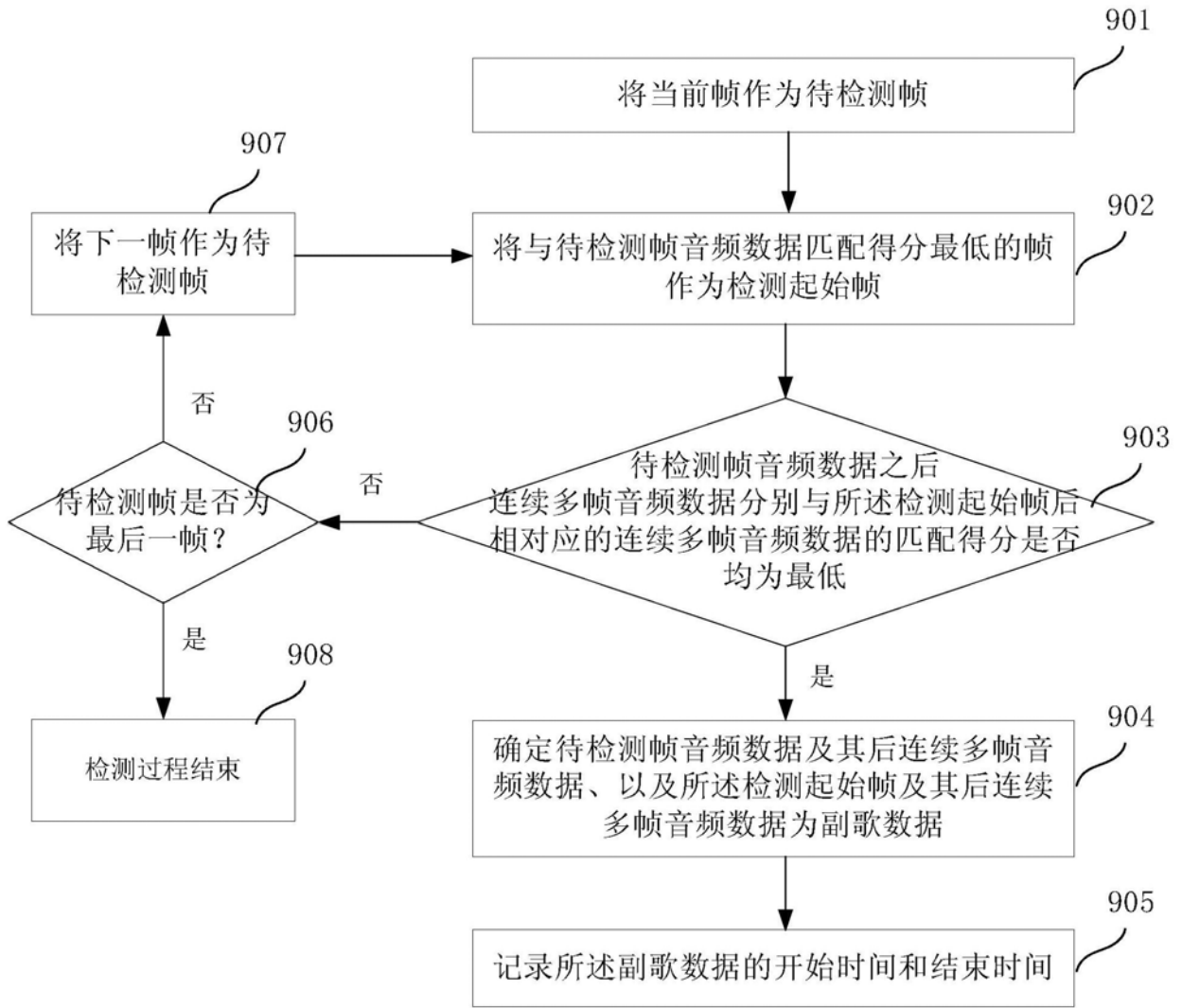


图9

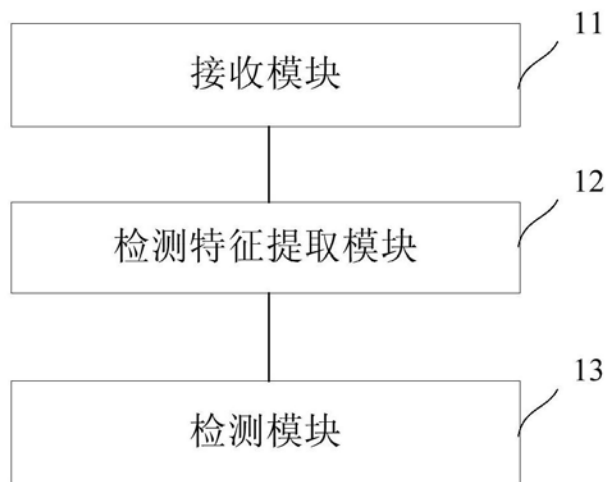


图10