



(12)发明专利申请

(10)申请公布号 CN 110033791 A

(43)申请公布日 2019. 07. 19

(21)申请号 201910232063.2

(22)申请日 2019.03.26

(71)申请人 北京雷石天地电子技术有限公司  
地址 100101 北京市朝阳区天畅园8号楼3层8-307

(72)发明人 宋冬梅 武剑 王宏

(74)专利代理机构 北京卓岚智财知识产权代理  
事务所(特殊普通合伙)  
11624

代理人 王新月

(51) Int. Cl.  
G10L 25/90(2013.01)

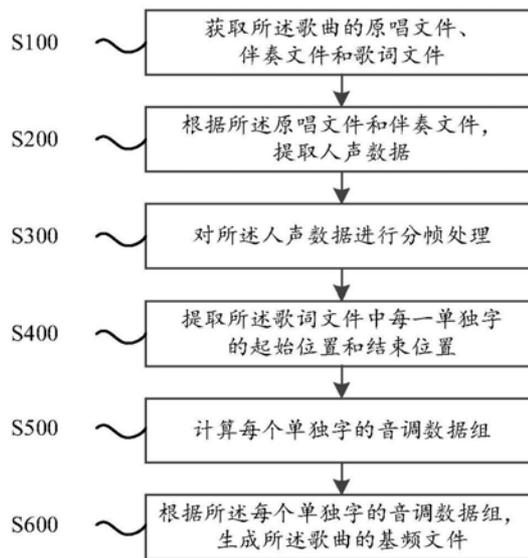
权利要求书2页 说明书9页 附图3页

(54)发明名称

一种歌曲基频提取方法及装置

(57)摘要

本发明公开了一种歌曲基频提取方法,具体包括:获取所述歌曲的原唱文件、伴奏文件和歌词文件;根据所述原唱文件和伴奏文件,提取人声数据;对所述人声数据进行分帧处理;提取所述歌词文件中每一单独字的起始位置和结束位置;计算每个单独字的音调数据组;根据所述每个单独字的音调数据组,生成所述歌曲的基频文件。本发明还公开了一种歌曲基频提取装置,能够实现上述歌曲基频提取方法。本发明能够极大地降低音调提取地错误率。



1. 一种歌曲基频提取方法,其特征在于,包括以下步骤:  
获取所述歌曲的原唱文件、伴奏文件和歌词文件;  
根据所述原唱文件和伴奏文件,提取人声数据;  
对所述人声数据进行分帧处理;  
提取所述歌词文件中每一单独字的起始位置和结束位置;  
计算每个单独字的音调数据组;  
根据所述每个单独字的音调数据组,生成所述歌曲的基频文件。
2. 如权利要求1所述的歌曲基频提取方法,其特征在于,所述提取人声数据之后,对所述人声数据进行分帧处理之前,还包括步骤:  
对所述人声数据进行升采样处理;  
所述升采样处理包括:  
对所述人声数据进行第一采样率采样,得到采样数据;  
在相邻的两个所述采样数据之间插值;  
对所有所述采样数据进行带通滤波。
3. 如权利要求1所述的歌曲基频提取方法,其特征在于,所述提取所述歌词文件中每一单独字的起始位置和结束位置包括:  
根据所述歌词文件,获取每一单独句的起始位置和结束位置;  
在所述每一单独句的起始位置之前和结束位置之后附加预设长度的余量帧;  
在所述余量帧的范围内计算所述每一单独句包括的全部所述分帧的音量;  
根据所述每一单独句范围内每一分帧的音量、所述歌词文件中每一单独字的时间长度和所述每一单独句的起始位置重新定位所述歌曲的起始位置和每一所述单独句的起始位置;  
根据所述重新定位的所述歌曲的起始位置和每一所述单独句的起始位置计算每一单独字的起始位置和结束位置。
4. 如权利要求1所述的歌曲基频提取方法,其特征在于,所述计算每个单独字的音调数据组包括:  
计算所述每一单独字的起始位置对应分帧和结束位置对应分帧的第一音调;  
计算所述每一单独字的起始位置与结束位置之间预设长度数据对应分帧的第二音调;  
根据所述第一音调和第二音调计算每个单独字的音调数据组。
5. 如权利要求4所述的歌曲基频提取方法,其特征在于,所述计算所述每一单独字的起始位置与结束位置之间预设长度数据对应分帧的第二音调包括:  
删除所述单独字范围内所有的清音帧,及被所述清音帧打断产生的不连续数据帧,保留没有被所述清音帧打断的连续数据帧;  
删除所述连续数据帧两端的音量小于预设阈值的分帧;  
判断剩余的所述连续数据帧是否满足预设长度数据对应的分帧数目,若不满足,则用0补充不足部分的数据;  
选取预设长度数据对应分帧计算第二音调。
6. 一种歌曲基频调整装置,其特征在于,包括文件获取模块、人声提取模块、音频分帧模块、位置提取模块、音调计算模块和基频生成模块,其中:

所述文件获取模块用于获取所述歌曲的原唱文件、伴奏文件和歌词文件；  
所述人声提取模块用于根据所述原唱文件和伴奏文件，提取人声数据；  
所述音频分帧模块用于对所述人声数据进行分帧处理；  
所述位置提取模块用于提取所述歌词文件中每一单独字的起始位置和结束位置；  
所述音调计算模块用于计算每个单独字的音调数据组；  
所述基频生成模块用于根据所述每个单独字的音调数据组，生成所述歌曲的基频文件。

7. 如权利要求6所述的歌曲基频调整装置，其特征在于，还包括升采样模块，所述升采样模块用于对所述人声提取模块提取的人声数据进行升采样处理；

所述升采样模块包括：

采样单元，用于对所述人声数据进行第一采样率采样，得到采样数据；

插值单元，用于在相邻的两个所述采样数据之间插值；

滤波单元，用于对所有所述采样数据进行带通滤波。

8. 如权利要求6所述的歌曲基频调整装置，其特征在于，所述位置提取模块包括：

句位置定位单元，用于根据所述歌词文件，获取每一单独句的起始位置和结束位置；

余量附加单元，用于在所述每一单独句的起始位置之前和结束位置之后附加预设长度的余量帧；

音量计算单元，用于在所述余量帧的范围内计算所述每一单独句包括的全部所述分帧的音量；

重定位单元，用于根据所述每一单独句范围内每一分帧的音量、所述歌词文件中每一单独字的时间长度和所述每一单独句的起始位置重新定位所述歌曲的起始位置和每一所述单独句的起始位置；

字位置定位单元，用于根据所述重新定位的所述歌曲的起始位置和每一所述单独句的起始位置计算每一单独字的起始位置和结束位置。

9. 如权利要求6所述的歌曲基频调整装置，其特征在于，所述音调计算模块包括：

分帧音调计算单元，用于计算所述每一单独字的起始位置对应分帧和结束位置对应分帧的第一音调；

整字音调计算单元，用于计算所述每一单独字的起始位置与结束位置之间预设长度数据对应分帧的第二音调；

字音调计算单元，用于根据所述第一音调和第二音调计算每个单独字的音调数据组。

10. 如权利要求9所述的歌曲基频调整装置，其特征在于，所述整字音调计算单元包括：

删除子单元，用于删除所述单独字范围内所有的清音帧，及被所述清音帧打断产生的不连续数据帧，保留没有被所述清音帧打断的连续数据帧；以及

删除所述连续数据帧两端的音量小于预设阈值的分帧；

判断子单元，用于判断剩余的所述连续数据帧是否满足预设长度数据对应的分帧数目；

数据补充子单元，用于当剩余的所述连续数据帧不满足预设长度数据对应的分帧数目时，用0补充不足部分的数据；

音调计算子单元，用于选取预设长度数据对应分帧计算第二音调。

## 一种歌曲基频提取方法及装置

### 技术领域

[0001] 本发明涉及音频信号处理领域,特别涉及一种歌曲基频提取方法及装置。

### 背景技术

[0002] 随着人们娱乐生活的丰富,在闲暇时间选择到KTV唱歌的民众也越来越多。不过,由于歌曲难度的差异,在KTV中对所选择的歌曲的音准进行调节以满足不同人群的需求也成为了大部分KTV中必备的功能。这也就要求必须对原歌曲的音调提取非常准确,但是,在歌曲库中的歌曲数量十分庞大,又有源源不断的新曲入库,现有的歌曲音调提取方法难以大批量快速处理,效率低、准确率差,难以满足当下的需求。

### 发明内容

[0003] 本发明提供了一种一种歌曲基频提取方法及装置,可以较大程度提高歌曲音调提取的效率和准确度。具体实现方式如下。

[0004] 一种歌曲基频提取方法,包括以下步骤:

[0005] 获取所述歌曲的原唱文件、伴奏文件和歌词文件;

[0006] 根据所述原唱文件和伴奏文件,提取人声数据;

[0007] 对所述人声数据进行分帧处理;

[0008] 提取所述歌词文件中每一单独字的起始位置和结束位置;

[0009] 计算每个单独字的音调数据组;

[0010] 根据所述每个单独字的音调数据组,生成所述歌曲的基频文件。

[0011] 优选地,所述提取人声数据之后,对所述人声数据进行分帧处理之前,还包括步骤:

[0012] 对所述人声数据进行升采样处理;

[0013] 所述升采样处理包括:

[0014] 对所述人声数据进行第一采样率采样,得到采样数据;

[0015] 在相邻的两个所述采样数据之间插值;

[0016] 对所有所述采样数据进行带通滤波。

[0017] 优选地,所述提取所述歌词文件中每一单独字的起始位置和结束位置包括:

[0018] 根据所述歌词文件,获取每一单独句的起始位置和结束位置;

[0019] 在所述每一单独句的起始位置之前和结束位置之后附加预设长度的余量帧;

[0020] 在所述余量帧的范围内计算所述每一单独句包括的全部所述分帧的音量;

[0021] 根据所述每一单独句范围内每一分帧的音量、所述歌词文件中每一单独字的时间长度和所述每一单独句的起始位置重新定位所述歌曲的起始位置和每一所述单独句的起始位置;

[0022] 根据所述重新定位的所述歌曲的起始位置和每一所述单独句的起始位置计算每一单独字的起始位置和结束位置。

- [0023] 优选地,所述计算每个单独字的音调数据组包括:
- [0024] 计算所述每一单独字的起始位置对应分帧和结束位置对应分帧的第一音调;
- [0025] 计算所述每一单独字的起始位置与结束位置之间预设长度数据对应分帧的第二音调;
- [0026] 根据所述第一音调和第二音调计算每个单独字的音调数据组。
- [0027] 优选地,所述计算所述每一单独字的起始位置与结束位置之间预设长度数据对应分帧的第二音调包括:
- [0028] 删除所述单独字范围内所有的清音帧,及被所述清音帧打断产生的不连续数据帧,保留没有被所述清音帧打断的连续数据帧;
- [0029] 删除所述连续数据帧两端的音量小于预设阈值的分帧;
- [0030] 判断剩余的所述连续数据帧是否满足预设长度数据对应的分帧数目,若不满足,则用0补充不足部分的数据;
- [0031] 选取预设长度数据对应分帧计算第二音调。
- [0032] 本发明还提供了一种歌曲基频调整装置,包括文件获取模块、人声提取模块、音频分帧模块、位置提取模块、音调计算模块和基频生成模块,其中:
- [0033] 所述文件获取模块用于获取所述歌曲的原唱文件、伴奏文件和歌词文件;
- [0034] 所述人声提取模块用于根据所述原唱文件和伴奏文件,提取人声数据;
- [0035] 所述音频分帧模块用于对所述人声数据进行分帧处理;
- [0036] 所述位置提取模块用于提取所述歌词文件中每一单独字的起始位置和结束位置;
- [0037] 所述音调计算模块用于计算每个单独字的音调数据组;
- [0038] 所述基频生成模块用于根据所述每个单独字的音调数据组,生成所述歌曲的基频文件。
- [0039] 优选地,所述歌曲基频调整装置还包括升采样模块,所述升采样模块用于对所述人声提取模块提取的人声数据进行升采样处理;
- [0040] 所述升采样模块包括:
- [0041] 采样单元,用于对所述人声数据进行第一采样率采样,得到采样数据;
- [0042] 插值单元,用于在相邻的两个所述采样数据之间插值;
- [0043] 滤波单元,用于对所有所述采样数据进行带通滤波。
- [0044] 优选地,所述位置提取模块包括:
- [0045] 句位置定位单元,用于根据所述歌词文件,获取每一单独句的起始位置和结束位置;
- [0046] 余量附加单元,用于在所述每一单独句的起始位置之前和结束位置之后附加预设长度的余量帧;
- [0047] 音量计算单元,用于在所述余量帧的范围内计算所述每一单独句包括的全部所述分帧的音量;
- [0048] 重定位单元,用于根据所述每一单独句范围内每一分帧的音量、所述歌词文件中每一单独字的时间长度和所述每一单独句的起始位置重新定位所述歌曲的起始位置和每一所述单独句的起始位置;
- [0049] 字位置定位单元,用于根据所述重新定位的所述歌曲的起始位置和每一所述单独

句的起始位置计算每一单独字的起始位置和结束位置。

[0050] 优选地,所述音调计算模块包括:

[0051] 分帧音调计算单元,用于计算所述每一单独字的起始位置对应分帧和结束位置对应分帧的第一音调;

[0052] 整字音调计算单元,用于计算所述每一单独字的起始位置与结束位置之间预设长度数据对应分帧的第二音调;

[0053] 字音调计算单元,用于根据所述第一音调和第二音调计算每个单独字的音调数据组。

[0054] 优选地,所述整字音调计算单元包括:

[0055] 删除子单元,用于删除所述单独字范围内所有的清音帧,及被所述清音帧打断产生的不连续数据帧,保留没有被所述清音帧打断的连续数据帧;以及

[0056] 删除所述连续数据帧两端的音量小于预设阈值的分帧;

[0057] 判断子单元,用于判断剩余的所述连续数据帧是否满足预设长度数据对应的分帧数目;

[0058] 数据补充子单元,用于当剩余的所述连续数据帧不满足预设长度数据对应的分帧数目时,用0补充不足部分的数据;

[0059] 音调计算子单元,用于选取预设长度数据对应分帧计算第二音调。

[0060] 本发明通过上述方案,能够根据歌曲库中的原唱、伴奏、歌词等资源文件,快速批量提取歌曲的基频数据,并能够极大地降低歌曲音调提取的错误率,提高音调选区的精度,实现了大批量歌曲的快速准确处理。

## 附图说明

[0061] 图1为本发明第一实施例中歌曲基频提取方法流程示意图;

[0062] 图2为本发明第一实施例中升采样方法流程示意图;

[0063] 图3为本发明第一实施例中单独字起始位置定位流程示意图;

[0064] 图4为本发明第一实施例中字音调计算方法流程示意图;

[0065] 图5为本发明第二实施例中歌曲基频提取装置结构示意图。

## 具体实施方式

[0066] 本发明。

[0067] 下面结合附图和具体实施方式对本发明进行详细地阐述。

[0068] 实施例一

[0069] 在本发明的一项实施例中,提供了一种歌曲基频提取方法,如图1所示,包括以下步骤:

[0070] 步骤S100:获取所述歌曲的原唱文件、伴奏文件和歌词文件;

[0071] 步骤S200:根据所述原唱文件和伴奏文件,提取人声数据;

[0072] 步骤S300:对所述人声数据进行分帧处理;

[0073] 步骤S400:提取所述歌词文件中每一单独字的起始位置和结束位置;

[0074] 步骤S500:计算每个单独字的音调数据组;

[0075] 步骤S600:根据所述每个单独字的音调数据组,生成所述歌曲的基频文件。

[0076] 本发明通过采用分帧方式提取音调与整字提取音调相结合的方式,首先提取人声数据,之后对人声数据分帧以获取每个单独字的起始位置,进而计算每个单独字的音调数据组,最后合成整首歌曲的基频文件。这一方案能够极大的减少音调提取错误率。

[0077] 优选地,在步骤S100之前,可以对待处理的音频数据进行归一化处理,将数据归一化到特定的数值范围内,例如-1~+1之间。对数据进行归一化处理可以采用如下公式进行:

$$[0078] \quad nval = \frac{val}{2^{bitnum-1}}$$

[0079] 其中,nval表示归一化后的音频数据值,一般采用浮点数据的表示形式,val表示待处理的原音频数据,一般以定点数据表示,bitnum表示定点数据val的位宽。

[0080] 优选地,在步骤S200中,提取人声数据的方法可以通过对原唱文件和伴奏文件分帧后在频域相减所得。在一个可选的实施方式中,步骤S200还包括步骤:

[0081] 步骤S210:根据所述歌词文件获取人声的起始时间;

[0082] 步骤S220:根据所述起始时间对所述原唱文件和伴奏文件进行音量对准;

[0083] 步骤S230:按照预设的窗口长度对音量对准后的所述原唱文件和伴奏文件进行分帧;

[0084] 步骤S240:计算所述原唱文件和伴奏文件每一分帧的幅度和相位;

[0085] 步骤S250:计算每一分帧的人声幅度和人声相位,所述人声幅度为所述原唱文件的幅度与所述伴奏文件的幅度之差,所述人声相位为所述原唱文件的相位;

[0086] 步骤S260:根据所述人声幅度和人声相位,计算所述人声数据。

[0087] 在本实施例一项可选的实施方式中,在步骤S200与步骤S300之间,如图2所示,还包括步骤S700:对所述人声数据进行升采样处理。步骤S700具体包括:

[0088] 步骤S710:对所述人声数据进行第一采样率采样,得到采样数据;

[0089] 步骤S720:在相邻的两个所述采样数据之间插值;

[0090] 步骤S730:对所有所述采样数据进行带通滤波。

[0091] 升采样处理的目的是提高后续音调提取算法的提取精度。通常情况下,歌曲的第一采样率一般为44.1kHz或48kHz,优选地,可以对采样率提升一倍,即采用88.2kHz或者96kHz。因此,采用步骤S720中插值的方法进行升采样率。优选地,这里可以在相邻的两个采样数据之间插值0,或者插入相邻数据的中值。插值完成后,再进行步骤S730,对采样数据进行带通滤波。

[0092] 通常情况下,人在歌唱时的音调起伏会比一般讲话时更加丰富,在低音区的部分泛音可以达到14kHz,流行歌曲中人声的基频范围在70~650Hz左右,考虑到纯人声在从歌曲中提取时会掺有少量的音乐信号,会对音调的提取产生一定影响,优选地,在进行步骤S730的带通滤波时,带通滤波器的滤波范围可以选择在50~800Hz,仅保留人声的第2~5次谐波用于音调的提取,但本发明对带通滤波器的滤波窗口选择并不加以限制,任何能够实现本发明目的合适的频率范围均可。这样处理对音调提取效果几乎不产生影响,而且可以更好的滤除人声中残余的音乐信号和低频噪声,减少音乐、噪声信号对音调提取的影响。

[0093] 优选地,所述带通滤波器除了作为升采样滤波器之外,还可以进行人声预过滤,二者在同一滤波器中完成,节省运算时间,提高运算效率。

[0094] 优选地,在步骤S300中,对人声数据进行分帧时,可以采用交叠滑动窗分帧。窗口的大小优选可以在10~40ms之间。如果数据经过了升采样,例如在96kHz的采样频率下,为了便于后续音调提取的运算,可以将窗的大小取值为2048个数据点,即21.3ms,窗口交叠范围为50%,即1024个数据点。

[0095] 经过分帧后,每一帧数据的时长FrameTime可以通过下述公式计算:

$$[0096] \quad FrameTime = \frac{1000 \times WL}{2 \times Fs}$$

[0097] 其中,WL为窗大小,Fs为采样率,分帧后的数据以帧序号FrameNum表示其时间位置。

[0098] 由于歌词文件标注的是时间点,而在进行后续处理时,往往采用的是每一帧数据,因此需要将歌词文件中的时间标注转换为帧序号:

$$[0099] \quad FrameNum = ROUND \left( \frac{Time}{FrameTime} \right)$$

[0100] 其中,FrameNum为帧序号,Time为歌词文件时间坐标,单位可以取毫秒。

[0101] 在本实施例一项可选的实施方式中,如图3所示,步骤S400还包括:

[0102] 步骤S410:根据所述歌词文件,获取每一单独句的起始位置和结束位置;

[0103] 步骤S420:在所述每一单独句的起始位置之前和结束位置之后附加预设长度的余量帧;

[0104] 步骤S430:在所述余量帧的范围内计算所述每一单独句包括的全部所述分帧的音量;

[0105] 步骤S440:根据所述每一单独句范围内每一分帧的音量、所述歌词文件中每一单独字的时间长度和所述每一单独句的起始位置重新定位所述歌曲的起始位置和每一所述单独句的起始位置;

[0106] 步骤S450:根据所述重新定位的所述歌曲的起始位置和每一所述单独句的起始位置计算每一单独字的起始位置和结束位置。

[0107] 由于歌词文件大部分为人工标注,有些仅标注一句起始位置,有些标注了一句的起始结束时间,即使每个字的都标注了时间也会存在不准确的情况,歌词整体与伴奏有时候也存在延时,造成音调提取准确率下降。所以,步骤S400对歌曲起始位置、句起始位置、每个字的起始结束位置都做了对齐处理,以提高音调检测准确度。

[0108] 在步骤S300进行分帧过后,可以确定歌词中每一单独句起始结束位置的帧序号SentenceStartIndex和SentenceStopIndex,在步骤S420中,在起始位置帧之前和结束位置帧之后附加一定的余量,优选地,余量可以为2000ms。假设余量对应OffsetNum个数据帧,每一单独句的音量计算处理范围则为[SentenceStartIndex-OffsetNum,SentenceStopIndex+OffsetNum]。

[0109] 步骤S430中,每帧音量仅需要计算每句起始结束位置附加余量对应帧的音量,句间余量交叠部分不需要重复计算。计算音量时,需要计算帧内音频数据的平均电平和所有人声数据的平均电平。人声的平均电平只需要计算每句起始结束位置对应帧的平均电平,即有歌词的时间段范围内。

[0110] 由电平计算音量的公式为:

[0111]  $UnitSI(n) = UnitdB - zerosdB$

[0112]  $AvgSI = AvgdB - zerosdB$

[0113] 其中,  $UnitSI(n)$  表示第  $n$  帧的第一音量,  $AvgSI$  表示人声平均音强,  $UnitdB$  表示平均电平值,  $AvgdB$  表示整首歌曲人声的平均电平,  $zerosdB$  表示相对静音对应的电平值, 即相对静音对应的电平值。这里以  $1dB$  表示一个强度等级, 可表示强度范围在  $0 \sim 127$ , 使得直观易于显示。

[0114] 计算完音量之后, 即可通过步骤 S440 和 S450 进行歌声对齐, 重新定位每一单独字的起始位置和结束位置。

[0115] 其中, 优选地, 对歌曲起始位置的调整可以通过如下步骤:

[0116] 步骤 S4411: 由人声的起始帧序号  $STARTFRAME$  为基准, 根据情况向前后各偏移预设的时间范围, 优选地可以在  $300 \sim 800ms$  之间选择, 例如  $500ms$ , 对应的帧序号范围为  $(STARTFRAME - OFFSET) \sim (STARTFRAME + OFFSET)$ 。  $OFFSET$  为偏移的时间范围所对应的帧个数;

[0117] 步骤 S4412: 由  $(STARTFRAME - OFFSET)$  开始向后搜索, 找到第一个音强大于  $(AvgSI - 10)$  的帧序号  $STARTFRAMEINDEX$ ;

[0118] 步骤 S4413: 计算时间调整量, 调整量  $ADJUST = STARTFRAMEINDEX - STARTFRAME$ ;

[0119] 步骤 S4414: 将歌词文件所有时间参数都换算为帧序号并与  $ADJUST$  相加, 实现歌声起始位置调整。

[0120] 优选地, 对每一单独句起始位置的调整可以通过如下步骤:

[0121] 步骤 S4421: 句子的调整范围为从第二句到最后一句的起始位置。设每句的起始帧序号为  $SentenceStartIndex$ , 对应的音量为  $SentenceStartI$ , 句起始帧前一帧和后一帧的音量为  $SentenceStartI\_A$  和  $SentenceStartI\_B$ 。确定最大调整偏移帧的数目  $X$ , 优选地, 可以选择 5 帧数据, 即在前后各 100 毫秒左右范围内调整句起始位置;

[0122] 步骤 S4422: 当  $SentenceStartI\_A > SentenceStartI > SentenceStartI\_B$  时, 向后查找  $X$  个帧的帧音量, 到找第一个音量谷点, 该点为调整后的句起始位置;

[0123] 步骤 S4423: 当  $SentenceStartI\_A < SentenceStartI < SentenceStartI\_B$  时, 向前依次查找  $X$  个帧的帧音量, 到找第一个音强谷点, 该点为调整后的句起始位置;

[0124] 步骤 S4424: 当  $SentenceStartI\_A < SentenceStartI > SentenceStartI\_B$  时, 向前后各查找第一个音量谷点的位置和对应的音量。取音量最低的谷点位置  $MINPOS$  方向为调整的方向, 调整的值取  $X$  和  $|SentenceStartI - MINPOS|$  间的最小值。若两个第一谷点音量相同, 则取距离  $SentenceStartI$  最近的谷点位置  $NEARPOS$  方向为调整方向, 调整值取  $X$  与  $|SentenceStartI - NEARPOS|$  间的最小值。

[0125] 完成步骤 S440 中的歌曲起始位置和每一单独句起始位置校正之后, 即可进行步骤 S450 中的字位置校正。字位置校正以字的起始结束时间为基, 若没有精确的起始结束时间可以以句的起始时间和每个字的时长来算出精略的起始结束时间。优选地, 进行字位置校正可以通过如下步骤:

[0126] 步骤 S451: 计算当前句起始位置与下一句起始位置间分组距离  $div$ , 再减去每个字占用的分组个数, 得到句间空白帧数  $blanknum$ ;

[0127] 步骤S452:由第一个字开始,如果字起始时间 $WordStartTime(n) < (AvgdB-20)$ ,则认为需要向后搜索字的起始位置,如果字结束时间 $WordEndTime(n) \geq (AvgdB-20)$ ,即认为当前起始结束位置就是校正后的字的起始结束位置 $WordStartIndexM(n)$ ,

[0128] 步骤S453:在向后查找时,当前字结束位置和剩余未计算的字的起始结束位置也相后累加,每次需要重新计算句间的空白帧数 $blanknum$ ,如果空白帧数为0,剩余未计算字的位置即最终校正后位置。

[0129] 同时,每句最后一个字的结束位置亦即该句的结束位置。

[0130] 在本实施例一项可选的实施方式中,如图4所示,步骤S500还包括:

[0131] 步骤S510:计算所述每一单独字的起始位置对应分帧和结束位置对应分帧的第一音调;

[0132] 步骤S520:计算所述每一单独字的起始位置与结束位置之间预设长度数据对应分帧的第二音调;

[0133] 步骤S530:根据所述第一音调和第二音调计算每个单独字的音调数据组。

[0134] 在以分帧方法识别字的音调时,一般先对目标数据进行短时傅里叶变换STFT,对运算结果求取对数幅度谱,再进行一次快速傅立叶逆变换IFFT,得到倒频谱数据,在此中求取基频数据,计算十二平均律,得到该字的音调。在运算时,可通过去除字的时间范围内分帧的清音帧,再计算音量最强的N个帧的中值来求取字的音调。但这种方式存在一定的误差,如在人声与伴奏分离不理想的情况下,特别是人声中混有鼓时,容易将鼓声与人声混合帧做为音量最大帧计算音调造成错误;另外在发音较短的字的起始结束端点时、或两个字间连接处时,往往由于音量较小帧或下一个字的帧不得不参与当前字的音调判定造成误判。因此,本发明采用提取对字内数据做整体提取音调与分帧提取音调结合的方式。

[0135] 字内整体提取音调的算法基本与分帧的处理方式相同,区别在于字内待计算音调数据的选取规则,如步骤S510~S530所示。优选地,其中,步骤S520还包括:

[0136] 步骤S521:删除所述单独字范围内所有的清音帧,及被所述清音帧打断产生的不连续数据帧,保留没有被所述清音帧打断的连续数据帧;

[0137] 步骤S522:删除所述连续数据帧两端的音量小于预设阈值的分帧;

[0138] 步骤S523:判断剩余的所述连续数据帧是否满足预设长度数据对应的分帧数目,若不满足,则进行步骤S524:用0补充不足部分的数据;

[0139] 步骤S525:选取预设长度数据对应分帧计算第二音调。

[0140] 音调的计算方法与分帧计算方法相同。

[0141] 优选地,在步骤S530中,将字的分帧音调与整字音调结合,计算字的最终音调。确定方式为:

[0142] 步骤S531:选取音量最大的的N帧数据 $SelPitch(m)$ ,计算选取的分帧的中值 $MedianPitch$ , $MedianPitch = Median(SelPitch(m))$ , $m = 1 \sim N$ ;

[0143] 步骤S532:将整字音调记为 $WordPitch$ ,若 $|MedianPitch - WordPitch| \leq 1$ ,则最终音调 $WPitch = ROUND((MedianPitch + WordPitch) * 0.5)$ ,其中 $ROUND()$ 表示对浮点数四舍五入操作;

[0144] 步骤S533:若 $1 < |MedianPitch - WordPitch| < 6$ , $WPitch = ROUND(WordPitch)$ ;

[0145] 步骤S534:若 $|MedianPitch - WordPitch| > 6$ ,那么最后得到的最终音调取值 $WPitch$

=ROUND (MIN (MedianPitch+WordPitch)),其中MIN()表示取两个数中最小值操作。

[0146] 最后,在步骤S600中,根据每个字的音调和音量对歌曲的整体音调进行修正。

[0147] 实施例二

[0148] 在本发明一项可选的实施例中,本发明还提供了一种歌曲基频调整装置,如图5所示,包括文件获取模块10、人声提取模块20、音频分帧模块30、位置提取模块40、音调计算模块50和基频生成模块60,其中:

[0149] 所述文件获取模块10用于获取所述歌曲的原唱文件、伴奏文件和歌词文件;

[0150] 所述人声提取模块20用于根据所述原唱文件和伴奏文件,提取人声数据;

[0151] 所述音频分帧模块30用于对所述人声数据进行分帧处理;

[0152] 所述位置提取模块40用于提取所述歌词文件中每一单独字的起始位置和结束位置;

[0153] 所述音调计算模块50用于计算每个单独字的音调数据组;

[0154] 所述基频生成模块60用于根据所述每个单独字的音调数据组,生成所述歌曲的基频文件。

[0155] 在一项可选的实施方式中,还包括升采样模块70,所述升采样模块70用于对所述人声提取模块提取的人声数据进行升采样处理;

[0156] 所述升采样模块70包括:

[0157] 采样单元71,用于对所述人声数据进行第一采样率采样,得到采样数据;

[0158] 插值单元72,用于在相邻的两个所述采样数据之间插值;

[0159] 滤波单元73,用于对所有所述采样数据进行带通滤波。

[0160] 在一项可选的实施方式中,所述位置提取模块40包括:

[0161] 句位置定位单元41,用于根据所述歌词文件,获取每一单独句的起始位置和结束位置;

[0162] 余量附加单元42,用于在所述每一单独句的起始位置之前和结束位置之后附加预设长度的余量帧;

[0163] 音量计算单元43,用于在所述余量帧的范围内计算所述每一单独句包括的全部所述分帧的音量;

[0164] 重定位单元44,用于根据所述每一单独句范围内每一分帧的音量、所述歌词文件中每一单独字的时间长度和所述每一单独句的起始位置重新定位所述歌曲的起始位置和每一所述单独句的起始位置;

[0165] 字位置定位单元45,用于根据所述重新定位的所述歌曲的起始位置和每一所述单独句的起始位置计算每一单独字的起始位置和结束位置。

[0166] 在一项可选的实施方式中,所述音调计算模块50包括:

[0167] 分帧音调计算单元51,用于计算所述每一单独字的起始位置对应分帧和结束位置对应分帧的第一音调;

[0168] 整字音调计算单元52,用于计算所述每一单独字的起始位置与结束位置之间预设长度数据对应分帧的第二音调;

[0169] 字音调计算单元53,用于根据所述第一音调和第二音调计算每个单独字的音调数据组。

[0170] 在一项可选的实施方式中,所述整字音调计算单元52包括:

[0171] 删除子单元521,用于删除所述单独字范围内所有的清音帧,及被所述清音帧打断产生的不连续数据帧,保留没有被所述清音帧打断的连续数据帧;以及

[0172] 删除所述连续数据帧两端的音量小于预设阈值的分帧;

[0173] 判断子单元522,用于判断剩余的所述连续数据帧是否满足预设长度数据对应的分帧数目;

[0174] 数据补充子单元533,用于当剩余的所述连续数据帧不满足预设长度数据对应的分帧数目时,用0补充不足部分的数据;

[0175] 音调计算子单元534,用于选取预设长度数据对应分帧计算第二音调。

[0176] 以上本发明各实施例提供的方案,能够根据歌曲库中的原唱、伴奏、歌词等资源文件,快速批量提取歌曲的基频数据,并能够较大程度提高歌曲音调提取准确度,实现了大批量歌曲的快速准确处理。

[0177] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,则可对这些实施例作出另外的变更和修改。

[0178] 本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样,倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。

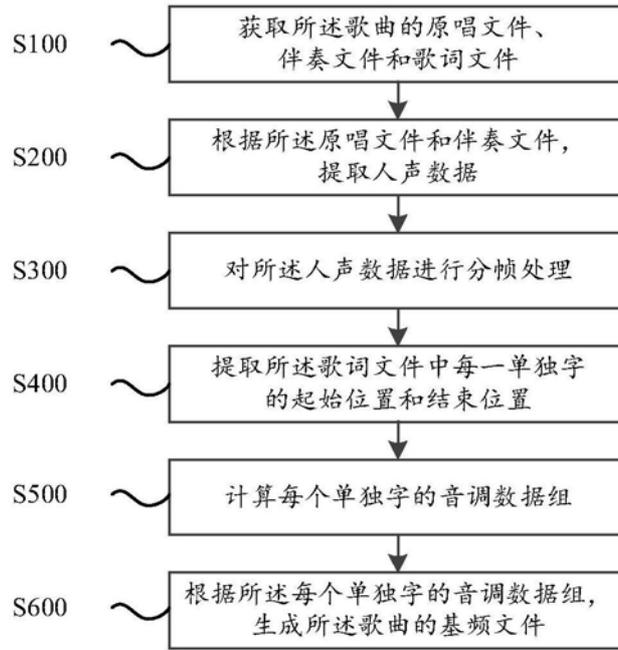


图1

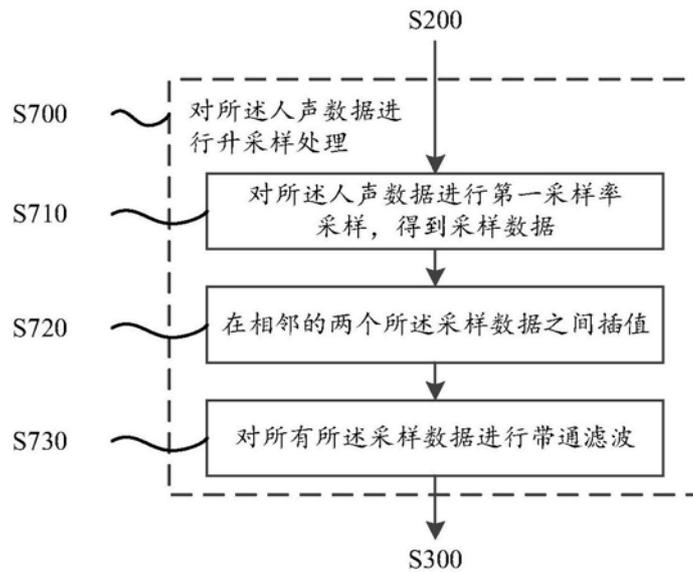


图2

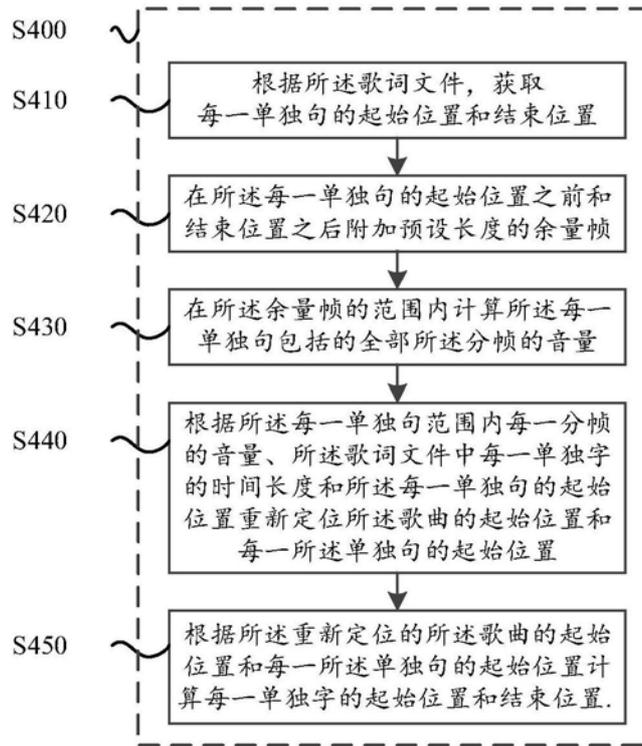


图3

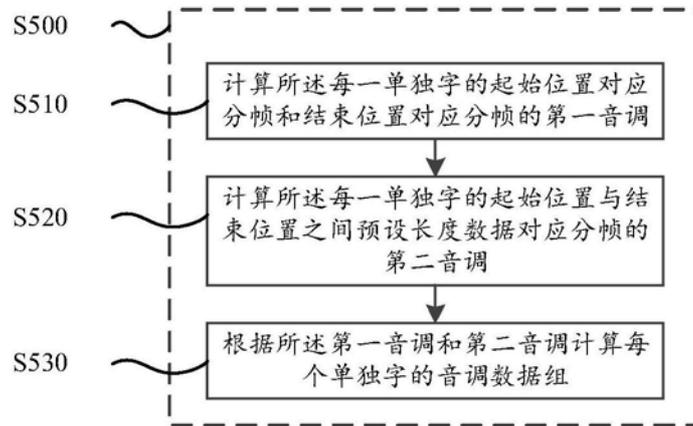


图4

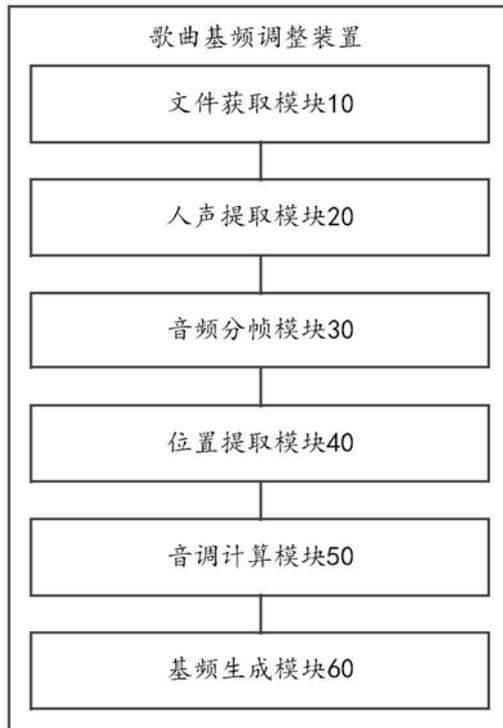


图5