



(12) 发明专利申请

(10) 申请公布号 CN 104221018 A

(43) 申请公布日 2014. 12. 17

(21) 申请号 201380019489. 0

代理人 杜文树

(22) 申请日 2013. 04. 16

(51) Int. Cl.

(30) 优先权数据

G06F 17/30 (2006. 01)

2012-094395 2012. 04. 18 JP

G10L 25/48 (2006. 01)

G10L 25/18 (2006. 01)

(85) PCT国际申请进入国家阶段日

2014. 10. 11

(86) PCT国际申请的申请数据

PCT/JP2013/002581 2013. 04. 16

(87) PCT国际申请的公布数据

W02013/157254 EN 2013. 10. 24

(71) 申请人 索尼公司

地址 日本东京

(72) 发明人 安部素嗣 西口正之 仓田宜典

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

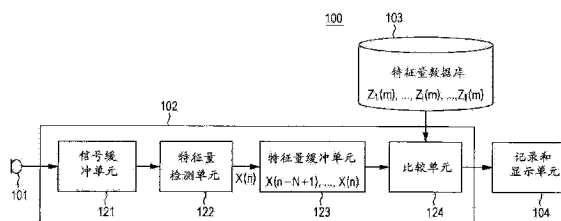
权利要求书3页 说明书21页 附图16页

(54) 发明名称

声音检测装置、声音检测方法、声音特征值检测装置、声音特征值检测方法、声音区间检测装置、声音区间检测方法及程序

(57) 摘要

提供了一种声音检测装置,包括:特征值提取单元,其从输入时间信号中提取每隔预定时间的特征值;特征值保持单元,其保持预定数目的检测目标声音项的特征值序列;以及比较单元,其每当特征值提取单元新提取出特征值时,分别将特征值提取单元提取的特征值序列与保持的预定数目的检测目标声音项的特征值序列相比较,并且获取预定数目的检测目标声音项的检测结果,其中特征值提取单元包括时频变换单元和似然分布检测单元,在频率方向和时间方向上使获得的似然分布平滑并且提取按照预定时间的特征值。



1. 一种声音检测装置,包括:

特征值提取单元,其从输入时间信号中提取每隔预定时间的特征值;

特征值保持单元,其保持预定数目的检测目标声音项的特征值序列;以及

比较单元,其每当所述特征值提取单元新提取出特征值时,分别将由所述特征值提取单元提取的特征值序列与保持的预定数目的检测目标声音项的特征值序列相比较,并且获取所述预定数目的检测目标声音项的检测结果,

其中,所述特征值提取单元包括对每个时间帧的输入时间信号执行时频变换并获取时频分布的时频变换单元,以及从所述时频分布获取音调似然分布的似然分布检测单元,所述特征值提取单元在频率方向和时间方向上使获取的似然分布平滑并且提取所述每隔预定时间的特征值。

2. 根据权利要求1所述的装置,其中,所述似然分布检测单元包括在所述时频分布的每个时间帧中检测频率方向上的峰的峰检测单元、在每个检测到的峰处拟合音调模型的拟合单元、以及基于拟合结果获取表示在每个检测到的峰处的音调成分似然性的分数的评分单元。

3. 根据权利要求1所述的装置,其中,所述特征值提取单元还包括在所述频率方向和/或所述时间方向上使经平滑的似然分布稀疏的稀疏单元。

4. 根据权利要求1所述的装置,其中,所述特征值提取单元还包括将经平滑的似然分布量化的量化单元。

5. 根据权利要求1所述的装置,其中,所述比较单元基于所述保持的检测目标声音项的特征值序列和由所述特征值提取单元针对所述预定数目的检测目标声音项中的每一个提取的特征值序列之间的对应特征值之间的关联来获取相似性,并且基于所获取的相似性获取所述检测目标声音项的检测结果。

6. 根据权利要求1所述的装置,还包括:

记录控制单元,其将所述预定数目的检测目标声音项的检测结果与时间信息一起记录在记录介质上。

7. 一种声音检测方法,包括:

从输入时间信号中提取每隔预定时间的特征值;以及

每当在所述特征值的提取中新提取出特征值时,分别将由特征值提取单元提取的特征值序列与所保持的预定数目的检测目标声音项的特征值序列相比较,并且获取所述预定数目的检测目标声音项的检测结果,

其中,在所述特征值的提取中,对每个时间帧的输入时间信号执行时频变换,获取时频分布,从所述时频分布获取音调似然分布,在频率方向和时间方向上使该似然分布平滑,并且提取所述每隔预定时间的特征值。

8. 一种程序,其使计算机执行以下操作:

从输入时间信号中提取每隔预定时间的特征值;以及

每当在所述特征值的提取中新提取出特征值时,分别将由特征值提取单元提取的特征值序列与所保持的预定数目的检测目标声音项的特征值序列相比较,并且获取所述预定数目的检测目标声音项的检测结果,

其中,在所述特征值的提取中,对每个时间帧的输入时间信号执行时频变换,获取时频

分布,从所述时频分布获取音调似然分布,在频率方向和时间方向上使该似然分布平滑,并且提取所述每隔预定时间的特征值。

9. 一种声音特征值提取装置,包括:

时频变换单元,其对每个时间帧的输入时间信号执行时频变换并且获取时频分布;

似然分布检测单元,其从所述时频分布获取音调似然分布;以及

特征值提取单元,其在频率方向和时间方向上使所述似然分布平滑并且提取每隔预定时间的特征值。

10. 根据权利要求9所述的装置,其中,所述似然分布检测单元包括在所述时频分布的每个时间帧中检测所述频率方向上的峰的峰检测单元、在每个检测到的峰处拟合音调模型的拟合单元、以及基于拟合结果获取表示在每个检测到的峰处的音调成分似然性的分数的评分单元。

11. 根据权利要求9所述的装置,还包括:

稀疏单元,其在所述频率方向和/或所述时间方向上使经平滑的似然分布稀疏。

12. 根据权利要求9所述的装置,还包括:

量化单元,其将经平滑的似然分布量化。

13. 根据权利要求9所述的装置,还包括:

声音区间检测单元,其基于所述输入时间信号检测声音区间,

其中,所述似然分布检测单元在检测到的声音区间的范围内从所述时频分布获取音调似然分布。

14. 根据权利要求13所述的装置,其中,所述声音区间检测单元包括对每个时间帧的输入时间信号执行时频变换并且获取时频分布的时频变换单元,基于该时频分布提取每个时间帧的幅度、音调成分强度和频谱大概轮廓的特征值的特征值提取单元,基于提取的特征值获取表示每个时间帧的声音区间似然性的分数的评分单元,在所述时间方向上使获取的每个时间帧的分数平滑的时间平滑单元,以及确定每个时间帧的经平滑分数的阈值并且获取声音区间信息的阈值确定单元。

15. 一种声音特征值提取方法,包括:

通过对每个时间帧的输入时间信号执行时频变换来获取时频分布;

从所述时频分布获取音调似然分布;以及

在频率方向和时间方向上使所述似然分布平滑。

16. 一种声音区间检测装置,包括:

时频变换单元,其通过对每个时间帧的输入时间信号执行时频变换来获取时频分布;

特征值提取单元,其基于该时频分布提取每个时间帧的幅度、音调成分强度和频谱大概轮廓的特征值;以及

评分单元,其基于提取的特征值获取表示每个时间帧的声音区间似然性的分数。

17. 根据权利要求16所述的装置,还包括:

时间平滑单元,其在时间方向上使获取的每个时间帧的分数平滑;以及

阈值确定单元,其确定每个时间帧的经平滑分数的阈值并且获取声音区间信息。

18. 一种声音区间检测方法,包括:

通过对每个时间帧的输入时间信号执行时频变换来获取时频分布;

基于该时频分布提取每个时间帧的幅度、音调成分强度和频谱大概轮廓的特征值 ; 以及
基于提取的特征值获取表示每个时间帧的声音区间似然性的分数。

声音检测装置、声音检测方法、声音特征值检测装置、声音特征值检测方法、声音区间检测装置、声音区间检测方法及程序

技术领域

[0001] 本技术涉及声音检测装置、声音检测方法、声音特征值检测装置、声音特征值检测方法、声音区间检测装置、声音区间检测方法及程序。

背景技术

[0002] 近年来,家用电器(家庭用途的电气设备)根据运行状态生成诸如控制声、通知声、操作声和警报声之类的各种声音(在下文中被称作“运行状态声”)。如果可以通过安装在家中某处的麦克风等观察这种运行状态声并且检测何时以及哪个家用电器执行哪种操作,则可以实现各种应用功能,诸如对作为所谓生活记录的自主行动历史的自动采集、针对具有听力困难的人将通知声可视化以及针对独自居住的老年人的行动监视。

[0003] 运行状态声可能是简单的蜂鸣声、嘟嘟声、音乐、语音声等,并且持续时间长度在短持续时间长度的情况下大约为 300ms 并且在长持续时间长度的情况下大约为几十秒。这种运行状态声被诸如安装在每个家用电器上的压电式蜂鸣器或者薄型扬声器之类的来自其的声音不足够令人满意的再现设备再现,并且被使得在周围传播。

[0004] 例如,PTL1 公开了一种技术,其中音乐作曲的部分碎片数据被转换为时频分布,特征值被提取并且随后被与已经登记的音乐作曲的特征值相比较,并且音乐作品的名称被识别出。

[0005] 引用列表

[0006] 专利文献

[0007] PTL 1:日本专利第 4788810 号

发明内容

[0008] 技术问题

[0009] 还可考虑将与 PTL1 中公开的技术相同的技术应用于对前述运行状态声音的检测。然而,关于家用电器生成的运行状态声,存在妨碍这种检测的以下事实:

[0010] (1) 有必要认出短至几百毫秒的运行状态声。

[0011] (2) 由于再现设备的不佳质量,声音变得失真,或者发生共鸣并且在一些情况下频率特性极度失真。

[0012] (3) 由于在周围的传播,幅度和相频特性较之实际家用电器生成的声音进一步失真。

[0013] 例如,图 17A 示出了在靠近家用电器的位置记录的运行状态声的波形示例。另一方面,图 17B 示出了在远离家用电器的位置记录的运行状态声的波形示例,并且波形是失真的。

[0014] (4) 诸如来自电视的输出声和交谈声之类的相对大的噪声和非恒定噪声在一些情

况下由于周围的传播而被叠加。例如,图 17C 示出了在靠近作为噪声源的电视的位置记录的运行状态声的波形示例,并且运行状态声被淹没在噪声中。

[0015] (5) 因为来自每个家用电器的声音级别和距麦克风的距离取决于每个家用电器,因此所记录的声音的音量发生变化。

[0016] 希望满意地检测诸如从家用电器生成的运行状态声之类的检测目标声音。

[0017] 问题的解决方案

[0018] 本技术的一个实施例涉及一种声音检测装置,包括:特征值提取单元,其从输入时间信号中提取每隔预定时间的特征值;特征值保持单元,其保持预定数目的检测目标声音项的特征值序列;以及比较单元,其每当特征值提取单元新提取出特征值时分别将特征值提取单元提取的特征值序列与保持的预定数目的检测目标声音项的特征值序列相比较并且获取预定数目的检测目标声音项的检测结果,其中特征值提取单元包括对每个时间帧的输入时间信号执行时频变换并且获取时频分布的时频变换单元、从该时频分布获取音调似然分布的似然分布检测单元以及在频率方向和时间方向上使似然分布平滑的平滑单元,并且该特征值提取单元从经平滑的似然分布中提取每隔预定时间的特征值。

[0019] 根据本技术,特征值提取单元从输入时间信号中提取按照预定时间的特征值。在这种情况下,特征值提取单元对每个时间帧的输入信号执行时频变换、获取时频分布、从该时频分布获取音调似然分布、在频率方向和时间方向上使似然分布平滑并且从经平滑的似然分布中提取按照预定时间的特征值。

[0020] 例如,似然分布检测单元可包括在时频分布的每个时间帧中检测频率方向上的峰的峰检测单元、在每个检测到的峰处拟合音调模型的拟合单元、以及基于拟合结果获取表示每个检测到的峰处的音调成分似然性的分数的评分单元。

[0021] 特征值保持单元保持预定数目的检测目标声音项的特征值序列。检测目标声音可包括人或动物的语音声等以及从家用电器生成的运行状态声(控制声、通知声、操作声、警报声等)。每当特征值提取单元新提取出特征值时,比较单元分别将特征值提取单元提取的特征值序列与保持的预定数目的检测目标声音的特征值序列相比较并且获取预定数目的检测目标声音项的检测结果。

[0022] 例如,比较单元可基于保持的检测目标声音项的特征值序列和由特征值提取单元针对预定数目的检测目标声音项中的每一个提取的特征值序列之间的对应特征值之间的关联来获取相似性,并且基于所获取的相似性获取检测目标声音项的检测结果。

[0023] 根据本技术,音调似然被从输入时间信号的时频分布获得,每隔预定时间的特征值被从在频率方向和时间方向上已被平滑的似然分布中提取并使用,并且可以在不依赖于麦克风的安装位置的情况下精确地检测检测目标声音(从家用电器生成的运行状态声等)。

[0024] 根据本技术,例如,特征值提取单元还可包括在频率方向和/或时间方向上使经平滑的似然分布稀疏的稀疏单元。根据本技术,例如,特征值提取单元还可包括使经平滑的似然分布量化的量化单元。在这种情况下,可以减少特征值序列的数据量并因而减少比较计算的负担。

[0025] 根据本技术,例如,该装置还可包括记录控制单元,其将预定数目的检测目标声音项的检测结果与时间信息一起记录在记录介质上。在这种情况下,例如,可以获得诸如家用

电器的操作历史之类的在家的用户动作历史。

[0026] 本技术的另一概念涉及一种声音特征值提取装置,包括:时频变换单元,其对每个时间帧的输入时间信号执行时频变换并且获取时频分布;似然分布检测单元,其从该时频分布获取音调似然分布;以及特征值提取单元,其在频率方向和时间方向上使似然分布平滑并且提取每隔预定时间的特征值。

[0027] 根据本技术,时频变换单元对每个时间帧的输入时间信号执行时频变换并且获取时频分布。似然分布检测单元从该时频分布获取音调似然分布。例如,似然分布检测单元可包括在时频分布的每个时间帧中检测频率方向上的峰的峰检测单元、在每个检测到的峰处拟合音调模型的拟合单元、以及基于拟合结果获取表示每个检测到的峰处的音调成分似然性的分数的评分单元。此外,特征值提取单元在频率方向和时间方向上使似然分布平滑并且提取按照预定时间的特征值。

[0028] 如上所述,根据本技术,音调似然性被从输入时间信号的时频分布获得,每隔预定时间的特征值被从在频率方向和时间方向上已被平滑的似然分布中提取,并且可以令人满意地提取输入时间信号中包括的声音的特征值。

[0029] 根据本技术,例如,特征值提取单元还可包括稀疏单元,其在频率方向和/或时间方向上使经平滑的似然分布稀疏。根据本技术,例如,特征值提取单元还可包括量化单元,其使经平滑的似然分布量化。在这样做时,可以减少提取的特征值的数据量。

[0030] 根据本技术,例如,该装置还可包括:声音区间检测单元,其基于输入时间信号检测声音区间,并且似然分布检测单元可在检测到的声音区间的范围内从时频分布获取音调似然分布。在这样做时,可以提取对应于声音区间的特征值。

[0031] 在这种情况下,声音区间检测单元可包括对每个时间帧的输入时间信号执行时频变换并且获取时频分布的时频变换单元,基于该时频分布提取每个时间帧的幅度、音调成分强度和频谱大概轮廓的特征值的特征值提取单元,基于提取的特征值获取表示每个时间帧的声音区间似然性的分数的评分单元,在所述时间方向上使获取的每个时间帧的分数平滑的时间平滑单元,以及确定每个时间帧的经平滑分数的阈值并且获取声音区间信息的阈值确定单元。

[0032] 此外,本技术的另一实施例涉及一种声音区间检测装置,包括:时频变换单元,其通过对每个时间帧的输入时间信号执行时频变换来获取时频分布;特征值提取单元,其基于该时频分布提取每个时间帧的幅度、音调成分强度和频谱大概轮廓的特征值;以及评分单元,其基于提取的特征值获取表示每个时间帧的声音区间似然性的分数。

[0033] 根据本技术,时频变换单元对每个时间帧的输入时间信号执行时频变换并且获取时频分布。特征值提取单元基于该时频分布提取每个时间帧的幅度、音调成分强度和频谱大概轮廓的特征值。此外,评分单元基于提取的特征值获取表示每个时间帧的声音区间似然性的分数。根据本技术,例如,该装置还可包括:时间平滑单元,其在时间方向上使获取的每个时间帧的分数平滑;以及阈值确定单元,其确定每个时间帧的经平滑分数的阈值并且获取声音区间信息。

[0034] 如上所述,根据本技术,每个时间帧的幅度、音调成分强度和频谱大概轮廓的特征值被从输入时间信号的时频分布中提取,表示每个时间帧的声音区间似然性的分数被从特征值获得,并且可以精确地获取声音区间信息。

[0035] 本发明的有利效果

[0036] 根据本技术,可以满意地检测诸如由家用电器生成的运行状态声等检测目标声音。

附图说明

[0037] [图 1] 图 1 是示出根据一个实施例的声音检测装置的配置示例的框图。

[0038] [图 2] 图 2 是示出特征值登记装置的配置示例的框图。

[0039] [图 3] 图 3 是示出声音区间和存在于该声音区间之前和之后的噪声区间的一个示例的示意图。

[0040] [图 4] 图 4 是示出构成特征值登记装置的声音区间检测单元的配置示例的框图。

[0041] [图 5A] 图 5A 是例示出音调强度特征值计算单元的示意图。

[0042] [图 5B] 图 5B 是例示出音调强度特征值计算单元的示意图。

[0043] [图 5C] 图 5C 是例示出音调强度特征值计算单元的示意图。

[0044] [图 5D] 图 5D 是例示出音调强度特征值计算单元的示意图。

[0045] [图 6] 图 6 是示出音调强度特征值计算单元中包括的、用于获取音调特性似然性的分数 $S(n, k)$ 的分布的音调似然性分布检测单元的配置示例的框图。

[0046] [图 7A] 图 7A 是示意性地例示出二次多项式函数在噪声特性的频谱峰附近不拟合良好,而该二次多项式函数在音调特性的频谱峰附近拟合良好的特性的示意图。

[0047] [图 7B] 图 7B 是示意性地例示出二次多项式函数在噪声特性的频谱峰附近不拟合良好,而该二次多项式函数在音调特性的频谱峰附近拟合良好的特性的示意图。

[0048] [图 8A] 图 8A 是示意性地示出音调特性的峰在时间方向上的变化的示意图。

[0049] [图 8B] 图 8B 是示意性地示出在频谱图上的小区域伽马 (Γ) 中拟合的示意图。

[0050] [图 9] 图 9 是示出音调似然分布检测单元检测音调似然分布的处理过程的示例的流程图。

[0051] [图 10] 图 10 是示出音调成分检测结果的一个示例的示意图。

[0052] [图 11] 图 11 是示出语音声的频谱图的一个示例的示意图。

[0053] [图 12] 图 12 是示出特征值提取单元的一个配置示例的框图。

[0054] [图 13] 图 13 是示出声音检测单元的一个配置示例的框图。

[0055] [图 14] 图 14 是例示出声音检测单元中的每个部件的操作的示意图。

[0056] [图 15] 图 15 是示出通过软件执行声音检测处理的计算装置的配置示例的框图。

[0057] [图 16] 图 16 是示出由 CPU 执行的检测目标声音检测处理的过程示例的流程图。

[0058] [图 17A] 图 17A 是例示出由实际家用电器生成的声音的记录状态的示意图。

[0059] [图 17B] 图 17B 是例示出由实际家用电器生成的声音的记录状态的示意图。

[0060] [图 17C] 图 17C 是例示出由实际家用电器生成的声音的记录状态的示意图。

具体实施方式

[0061] 在下文中,将给出对用于实现本技术的实施例(在下文中被称作“实施例”)的描述。此外,将按照以下次序给出描述。

[0062] 1. 实施例

[0063] 2. 修改示例

[0064] <1. 实施例>

[0065] “声音检测装置”

[0066] 图 1 示出了根据一个实施例的声音检测装置 100 的配置示例。声音检测装置 100 包括麦克风 101、声音检测单元 102、特征值数据库 103 和记录和显示单元 104。

[0067] 声音检测装置 100 执行用于检测由家用电器生成的运行状态声（控制声、通知声、操作声、警报声等）的声音检测处理并且记录和显示检测结果。就是说，在声音检测处理中，每隔预定时间的特征值被从通过由麦克风 101 采集声音而获得的时间信号 $f(t)$ 中提取，并且该特征值被与特征值数据库中登记的预定数目的检测目标声音项相比较。然后，如果特征值与预定检测目标声音的特征值序列大体符合的比较结果在声音检测处理中被获得，则该预定检测目标声音的时间和名称被记录和显示。

[0068] 麦克风 101 采集房间中的声音并且输出时间信号 $f(t)$ 。房间中的声音也包括由家用电器 1 至 N 生成的运行状态声（控制声、通知声、操作声、警报声等）。声音检测单元 102 获取从麦克风 101 输出的时间信号 $f(t)$ 作为输入并且从该时间信号中提取每隔预定时间的特征值。在这点上，声音检测单元 102 构成特征值提取单元。

[0069] 在构成特征值保持单元的特征值数据库 103 中，包括预定数目的检测目标声音项的特征值序列被与检测目标声音名称相关联地登记和保持。在该实施例中，预定数目的检测目标声音项例如意味着由家用电器 1 至 N 生成的运行状态声的全部或者一部分。声音检测单元 102 每当新特征值被提取时将所提取的特征值序列与特征值数据库 103 中保持的预定数目的检测目标声音项的特征值序列相比较并且获取预定数目的检测目标声音的检测结果。在这点上，声音检测单元 102 构成比较单元。

[0070] 记录和显示单元 104 将声音检测单元 102 的检测目标声音检测结果与时间一起记录在记录介质中并且将检测结果显示在显示器上。例如，当声音检测单元 102 的检测目标声音检测结果表明来自家用电器 1 的通知声 A 已被检测到时，记录和显示单元 104 将来自家用电器 1 的通知声 A 被产生的事实及其时间记录在记录介质上并且显示在显示器上。

[0071] 图 1 所示的声音检测装置 100 的操作将被描述。麦克风 101 采集房间中的声音。来自麦克风 101 的时间信号输出被提供给声音检测单元 102。声音检测单元 102 从时间信号中提取每隔预定时间的特征值。然后，声音检测单元 102 每当新特征值被提取时将所提取的特征值序列与特征值数据库 103 中保持的预定数目的检测目标声音项的特征值序列相比较并且获取预定数目的检测目标声音项的检测结果。检测结果被提供给记录和显示单元 104。记录和显示单元 104 将检测结果与时间一起记录在记录介质上并且显示在显示器上。

[0072] “特征值登记装置”

[0073] 图 2 示出了将检测目标声音的特征值序列登记在特征值数据库 103 中的特征值登记装置 200 的配置示例。特征值登记装置 200 包括麦克风 201、声音区间检测单元 202、特征值提取单元 203 和特征值登记单元 204。

[0074] 特征值登记装置 200 执行声音登记处理（声音区间检测处理和声音特征提取处理）并且将检测目标声音（家用电器生成的运行状态声）的特征值序列登记在特征值数据库 103 中。一般而言，噪声区间存在于由麦克风 201 记录的、将被登记的检测目标声音之前

和之后。因此,实际将被登记的包括有意义声音(检测目标声音)的声音区间在声音区间检测处理中被检测到。图3示出了声音区间和存在于该声音区间之前和之后的噪声区间的示例。在声音特征提取处理中,对检测检测目标声音有用的特征值被从由麦克风201获取的声音区间的时间信号 $f(t)$ 中提取并且与检测目标声音名称一起被登记在特征值数据库103中。

[0075] 麦克风201采集将被登记为检测目标声音的家用电器的运行状态声。声音区间检测单元202获取从麦克风201输出的时间信号 $f(t)$ 作为输入并且从时间信号 $f(t)$ 中检测声音区间,即由家用电器生成的运行状态声的区间。特征值提取单元203获取从麦克风201输出的时间信号 $f(t)$ 作为输入并且从时间信号 $f(t)$ 中提取每隔预定时间的特征值。

[0076] 特征值提取单元203对每个时间帧的输入时间信号 $f(t)$ 执行时频变换,获取时频分布,从时频分布获取音调似然分布,在频率方向和时间方向上使似然分布平滑,并且提取每隔预定时间的特征值。在这种情况下,特征值提取单元203在基于从声音区间检测单元202提供的声音区间信息的声音区间范围内提取特征值并且获取与家用电器生成的操作情况声的区间相对应的特征值序列。

[0077] 特征值登记单元204将与已经由特征值提取单元203获取的、作为检测目标声音的、由家用电器生成的运行状态声相对应的特征值序列与特征值数据库103中的检测目标声音名称(关于运行状态声的信息)相关联并且登记。在附图中示出的示例中,其中包括I个检测目标声音项 $Z1(m)$ 、 $Z2(m)$, ..., $Zi(m)$, ..., $ZI(m)$ 的特征值序列被登记在特征值数据库103中的状态被例示出。

[0078] “声音区间检测单元”

[0079] 图4示出声音区间检测单元202的配置示例。声音区间检测单元202的输入是由记录将被登记的检测目标声音(由家用电器生成的运行状态声)的麦克风201获取的时间信号 $f(t)$,并且噪声区间如图3所示也被包括在检测目标信号之前和之后。此外,来自声音区间检测单元202的输出是指示包括实际将被登记的有意义声音(检测目标声音)的声音区间的声音区间信息。

[0080] 声音区间检测单元202包括时频变换单元221、幅度特征值计算单元222、音调强度特征值计算单元223、频谱大概轮廓特征值计算单元224、分数计算单元225、时间平滑单元226和阈值确定单元227。

[0081] 时频变换单元221对输入时间信号 $f(t)$ 执行时频变换并且获得时频信号 $F(n, k)$ 。这里, t 表示离散时间, n 表示时间帧的数目,并且 k 表示离散频率。时频变换单元221通过短时间傅立叶变换对输入时间信号 $f(t)$ 执行时频变换并且获得如在以下式(1)中示出的时频信号 $F(n, k)$ 。

[0082] [数学1]

$$[0083] \quad F(n, k) = \log \left| \sum_{t=0}^{M-1} W(t) f(t - nR) e^{j2\pi kn} \right| \quad \dots (1)$$

[0084] 这里, $W(t)$ 表示窗口函数, M 表示窗口函数的大小,并且 R 表示帧时间间隔(=跳大小(hop size))。时频信号 $F(n, k)$ 表示时间帧 n 中的、位于频率 k 的频率成分的对数幅度值并且是所谓的频谱图(时频分布)。

[0085] 幅度特征值计算单元 222 根据时频信号 $F(n, k)$ 计算幅度特征值 $x_0(n)$ 和 $x_1(n)$ 。具体而言,幅度特征值计算单元 222 获取由以下式 (2) 表示的、预定频率范围 (具有下限 K_L 和上限 K_H) 的目标帧 n 附近的时间区间 (具有目标帧 n 之前和之后的长度 L) 的平均幅度 $A_{ave}(n)$ 。

[0086] [数学 2]

$$A_{ave}(n) = \frac{1}{2L+1} \sum_{n=-L}^L \sum_{k=K_L}^{K_H} \exp(F(n, k)) \quad \cdots (2)$$

[0088] 此外,幅度特征值计算单元 222 获取由以下式 (3) 表示的、预定频率范围 (具有下限 K_L 和上限 K_H) 的目标帧 n 中的绝对幅度 $A_{abs}(n)$ 。

[0089] [数学 3]

$$A_{abs}(n) = \sum_{k=K_L}^{K_H} \exp(F(n, k)) \quad \cdots (3)$$

[0091] 另外,幅度特征值计算单元 222 获取由以下式 (4) 表示的、预定频率范围 (具有下限 K_L 和上限 K_H) 的目标帧 n 中的相对幅度 $A_{rel}(n)$ 。

[0092] [数学 4]

$$A_{rel}(n) = \frac{A_{abs}(n)}{A_{ave}(n)} \quad \cdots (4)$$

[0094] 此外,幅度特征值计算单元 222 如以下式 (5) 所示将绝对幅度 $A_{abs}(n)$ 看作幅度特征值 $x_0(n)$ 并且将相对幅度 $A_{rel}(n)$ 看作幅度特征值 $x_1(n)$ 。

[0095] [数学 5]

$$x_0(n) = A_{abs}(n), x_1(n) = A_{rel}(n) \quad \cdots (5)$$

[0097] 音调强度特征值计算单元 223 根据时频信号 $F(n, k)$ 计算音调强度特征值 $x_2(n)$ 。音调强度特征值计算单元 223 首先将时频信号 $F(n, k)$ 的分布 (见图 5A) 变换为音调特性似然性的分数 $S(n, k)$ 的分布 (见图 5B)。每个分数 $S(n, k)$ 是从 0 到 1 的分数,其表示时频成分多么可能是每个频率 k 处的 $F(n, k)$ 的相应时间 n 中的“音调成分”。具体而言,分数 $S(n, k)$ 在 $F(n, k)$ 构成频率方向上的音调特性的峰的位置处接近 1 并且在其他位置处接近 0。

[0098] 图 6 示出了音调强度特征值计算单元 223 中包括的、用于获取音调特性似然性的分数 $S(n, k)$ 的分布的音调似然性分布检测单元 230 的配置示例。音调似然性分布检测单元 230 包括峰检测单元 231、拟合单元 232、特征值提取单元 233 和评分单元 234。

[0099] 峰检测单元 231 在频谱图 (时频信号 $F(n, k)$ 的分布) 的每个时间帧中检测频率方向上的峰。就是说,峰检测单元 231 检测某一位置是否对应于频谱图的所有频率的所有帧中的频率方向上的峰 (最大值)。

[0100] 例如通过检查以下式 (6) 是否得到满足来执行关于 $F(n, k)$ 是否对应于峰的检测。尽管使用三个点的方法被例示为峰检测方法,但是使用五个点的方法也是适用的。

[0101] $F(n, k-1) < F(n, k)$ 并且 $F(n, k) > F(n, k+1) \dots (6)$

[0102] 拟合单元 232 在已经由峰检测单元 231 检测到的每个峰附近的区域中拟合音调模型, 如下所述。首先, 拟合单元 232 执行变换为包括目标峰作为原点的坐标的坐标变换并且如以下式 (7) 所示设置邻近的时频区域。这里, 德尔塔 (Δ)_N 表示时间方向上的临近区域 (例如为三个点), 并且德尔塔 k 表示频率方向上的临近区域 (例如为两个点)。

[0103] [数学 6]

[0104] $\Gamma = [-\Delta_N \leq n \leq \Delta_N] \times [-\Delta_k \leq k \leq \Delta_k] \dots (7)$

[0105] 接下来, 拟合单元 232 使例如如以下式 (8) 所示的二次多项式函数的音调模型拟合临近区域中的时频信号。在这种情况下, 拟合单元 232 例如在峰附近的时频分布与音调模型之间基于均方误差最小准则执行拟合。

[0106] [数学 7]

[0107] $Y(k, n) = ak^2 + bk + cnk + dn^2 + en + g \dots (8)$

[0108] 就是说, 拟合单元 232 通过获取使时频信号的临近区域中的、如以下式 (9) 所示的均方误差和如以下式 (10) 所示的多项式函数最小化的系数来执行拟合。

[0109] [数学 8]

[0110] $J(a, b, c, d, e, g) = \sum_{\Gamma} (Y(k, n) - F(k, n))^2 \dots (9)$

[0111] $(\hat{a}, \hat{b}, \hat{c}, \hat{d}, \hat{e}, \hat{g}) = \arg \min J(a, b, c, d, e, g) \dots (10)$

[0112] 二次多项式函数具有二次多项式函数在音调特性的频谱峰附近拟合良好 (误差小) 并且在噪声特性的频谱峰附近拟合不佳 (误差大) 的特性。图 7A 和图 7B 是示意性地示出该状态的示图。图 7A 示意性地例示出通过前述式 (1) 获得的、第 n 个帧中的音调特性的峰附近的频谱。

[0113] 图 7B 示出了一种状态, 其中由以下式 (11) 示出的二次函数 $f_0(k)$ 被应用于图 7A 中的频谱。这里, a 表示峰曲率, k_0 表示实际峰的频率, 并且 g_0 表示实际峰的位置处的对数幅度值。二次函数在音调特性成分的频谱峰周围拟合良好然而二次函数倾向于在噪声特性的峰附近极大偏离。

[0114] [数学 9]

[0115] $f_0(k) = a(k - k_0)^2 + g_0 \dots (11)$

[0116] 图 8A 示意性地示出音调特性的峰在时间方向上的变化。音调特性的峰的幅度和频率在其大概轮廓被保持的同时在先前和后续的时间帧中改变。尽管实际获得的频谱是离散点, 但是频谱为了便利而被表示为曲线。点划线示出了前一帧, 实线示出了当前帧, 并且虚线示出了下一帧。

[0117] 在许多情况下, 音调特性成分在时间上在一定程度上是连续的并且可被表示为尽管频率和时间稍微改变但是形状大体相同的二次函数的移动。变化 $Y(k, n)$ 由以下式 (12) 表示。因为频谱被表示为对数幅度, 因此幅度的变化对应于频谱在垂直方向上的位移。这是幅度变化项 $f_1(n)$ 被添加的原因。这里, 贝塔 (β) 是频率的变化率, 并且 $f_1(n)$ 是表示峰位置处的幅度变化的时间函数。

[0118] [数学 10]

[0119] $Y(k, n) = f_0(k - \beta n) + f_1(n) \dots (12)$

[0120] 如果 $f_1(n)$ 被时间方向上的二次函数近似, 则变化 $Y(k, n)$ 可由以下式 (13) 表示。因为 a, k_0, β, d_1, e_1 和 g_0 是恒定的, 因此式 (13) 通过适当变换变量等价于前述式 (8)。

[0121] [数学 11]

[0122]

$$\begin{aligned} Y(k, n) &= a(k - k_0 - \beta n)^2 + g_0 + d_1 n^2 + e_1 n \\ &= ak^2 - 2ak_0k - 2ak_0\beta n + a\beta^2 n^2 + d_1 n^2 + 2ak_0\beta n + e_1 n + ak_0^2 + g_0 \\ &\dots (13) \end{aligned}$$

[0123] 图 8B 示意性地示出在频谱图上的小区域伽马 (Γ) 中进行拟合的示意图。因为类似的形状在音调特性的峰周围随着时间过去而逐渐改变, 因此式 (8) 倾向于适用良好。然而, 关于噪声特性的峰的附近, 峰的形状和频率发生变化, 并且式 (8) 因而适用不良, 就是说, 即使最优地将式 (8) 拟合, 也会发生大误差。

[0124] 前述式 (10) 示出了针对关于所有系数 a, b, c, d, e 和 g 进行拟合的计算。然而, 可在一些系数被预先固定为常数之后执行拟合。此外, 可利用二维或者更多维的多项式函数来执行拟合。

[0125] 返回图 6, 特征值提取单元 233 基于拟合单元 232 在每个峰处获得的拟合结果 (见前述式 (10)) 如以下式 (14) 所示提取特征值 (x_0, x_1, x_2, x_3, x_4 和 x_5)。每个特征值是表示每个峰处的频率成分的特性的特征值, 并且特征值本身可被用于分析语音声或者音乐声。

[0126] [数学 12]

[0127]

$$\left. \begin{aligned} (\text{峰的曲率}) \quad & x_0 = \hat{a} \\ (\text{峰的频率}) \quad & x_1 = -\frac{b}{2\hat{a}} \\ (\text{峰的对数幅度值}) \quad & x_2 = \hat{g} \\ (\text{频率的变化率}) \quad & x_3 = -\frac{\hat{c}}{2\hat{a}} \\ (\text{幅度的变化率}) \quad & x_4 = \hat{e} \\ (\text{拟合中的归一化误差}) \quad & x_5 = \frac{J(\hat{a}, \hat{b}, \hat{c}, \hat{d}, \hat{e}, \hat{g})}{\sum_{\Gamma} (F(n, k) - \hat{g})^2} \end{aligned} \right\} \dots (14)$$

[0128] 评分单元 234 通过使用特征值提取单元 233 针对每个峰提取的特征值来获取表示每个峰的音调成分似然性的分数 $S(n, k)$, 以使每个峰的音调成分似然性量化。评分单元

234 通过使用特征值 (x_0 、 x_1 、 x_2 、 x_3 、 x_4 和 x_5) 当中的一个或多个特征值来如以下式 (15) 所示获取分数 $S(n, k)$ 。在这种情况下, 至少拟合的归一化误差 x_5 或者峰在频率方向上的曲率 x_0 被使用。

[0129] [数学 13]

$$[0130] \quad S(n, k) = \text{Sigm} \left(\sum_{i=0}^5 w_i H_i(x_i) + w_6 \right) \quad \dots (15)$$

[0131] 这里, $\text{Sigm}(x)$ 是 S 型函数, w_i 是预定负载系数, 并且 $H_i(X_i)$ 是第 i 个特征值 x_i 的预定非线性函数。可以使用例如如以下式 (16) 所示的函数作为非线性函数 $H_i(X_i)$ 。这里, u_i 和 v_i 是预定负载系数。适当的常数可被预先确定为 w_i 、 u_i 和 v_i , 其例如可使用多个数据项通过最速下降学习来自动选择。

[0132] [数学 14]

$$[0133] \quad H_i(x_i) = \text{Sigm}(u_i x_i + v_i) \dots (16)$$

[0134] 评分单元 234 如上所述通过式 (15) 获取表示每个峰的音调成分似然性的分数 $S(n, k)$ 。此外, 评分单元 234 将除峰之外位置 (n, k) 处的分数 $S(n, k)$ 设置为 0。评分单元 234 在时频信号 $f(n, k)$ 的每个时间和每个频率处获取作为从 0 到 1 的值的音调成分似然性的分数 $S(n, k)$ 。

[0135] 图 9 中的流程图示出了音调似然分布检测单元 230 执行的音调似然分布检测的处理过程的示例。音调似然性分布检测单元 230 在步骤 ST1 中开始处理然后转移到步骤 ST2 中的处理。在步骤 ST2 中, 音调似然性分布检测单元 230 将帧 (时间帧) 的数目 n 设置为 0。

[0136] 接下来, 音调似然性分布检测单元 230 在步骤 ST3 中确定“ $n < N$ ”是否得到满足。此外, 频谱图 (时频分布) 的帧当前是从 0 到“ $N - 1$ ”。如果“ $n < N$ ”未得到满足, 则音调似然性分布检测单元 230 确定所有帧的处理已经完成, 并且在步骤 ST4 中完成处理。

[0137] 如果“ $n < N$ ”得到满足, 则音调似然性分布检测单元 230 在步骤 ST5 中将离散频率 k 设置为 0。然后, 音调似然性分布检测单元 230 在步骤 ST6 中确定“ $k < K$ ”是否得到满足。此外, 频谱图 (时频分布) 的离散频率 k 当前是从 0 到“ $k - 1$ ”。如果“ $k < K$ ”未得到满足, 则音调似然性分布检测单元 230 确定所有离散频率的处理已经完成, 在步骤 ST7 中使 n 增加, 然后返回到步骤 ST3, 并且转移到对下一帧的处理。

[0138] 如果在步骤 ST6 中“ $k < K$ ”得到满足, 则音调似然性分布检测单元 230 在步骤 ST8 中判定 $F(n, k)$ 是否对应于峰。如果 $F(n, k)$ 不对应于峰, 则音调似然性分布检测单元 230 在步骤 ST9 中将分数 $S(n, k)$ 设置为 0, 在步骤 ST10 中使 k 增加, 然后返回到步骤 ST6, 并且转移到对下一离散频率的处理。

[0139] 如果在步骤 ST8 中 $F(n, k)$ 对应于峰, 则音调似然性分布检测单元 230 转移到步骤 ST11 中的处理。在步骤 ST11 中, 音调似然性分布检测单元 230 在峰附近的区域中拟合音调模型。然后, 音调似然性分布检测单元 230 在步骤 ST12 中基于拟合结果提取各种特征值 (x_0 、 x_1 、 x_2 、 x_3 、 x_4 和 x_5)。

[0140] 接下来, 在步骤 ST13 中, 音调似然性分布检测单元 230 通过使用在步骤 ST12 中提

取的特征值来获取表示峰的音调成分似然性的、作为从 0 到 1 的值的分数 $S(n, k)$ 。音调似然性分布检测单元 230 在步骤 ST14 中的处理之后在步骤 ST10 中使 k 增加, 并且转移到对下一离散频率的处理。

[0141] 图 10 示出了根据如在图 11 中示出的时频分布 (频谱图) $F(n, k)$ 的、由在图 6 中示出的音调似然性分布检测单元 230 获得的音调成分似然性的分数 $S(n, k)$ 的一个分布示例为。分数 $S(n, k)$ 的更大值由更深的黑色示出, 并且可观察到当音调特性成分 (构成图 11 中的黑粗水平线的成分) 的峰被大概检测到时噪声特性的峰未被大概检测到。

[0142] 返回到图 4, 音调强度特征值计算单元 223 接下来创建音调成分提取滤波器 $H(n, k)$ (见图 5C), 其仅提取分数 $S(n, k)$ 大于预定阈值 S_{thsd} (见图 5B) 的位置附近的频率位置处的成分。以下式 (17) 表示音调成分提取滤波器 $H(n, k)$ 。

[0143] [数学 15]

[0144]

$$H(n, k) = \begin{cases} 1 & k_T - \Delta k < k < k_T + \Delta k \\ 0 & \text{否则} \end{cases} \quad \dots (17)$$

[0145] 然而, k_T 表示音调成分被检测到之处的频率, 并且德尔塔 k 表示预定频率宽度。这里, 当如上所述为了获得时频信号 $F(n, k)$ 的短时间傅立叶变换 (见式 (1)) 中的窗口函数 $W(t)$ 的大小是 M 时德尔塔 k 优选是 $2/M$ 。

[0146] 音调强度特征值计算单元 223 接下来将原始时频信号 $F(n, k)$ 乘以音调成分提取滤波器 $H(n, k)$ 并且获得通过如图 5D 所示仅使音调成分留下而获得的频谱 (音调成分频谱) $F_T(n, k)$ 。以下式 (18) 表示音调成分频谱 $F_T(n, k)$ 。

[0147] [数学 16]

[0148] $F_T(n, k) = H(n, k)F(n, k) \dots (18)$

[0149] 音调强度特征值计算单元 223 最终在预定频率区域 (具有下限 K_L 和上限 K_H) 求和并且获取由以下式 (19) 表示的、目标帧 n 中的音调成分强度 $A_{tone}(n)$ 。

[0150] [数学 17]

$$A_{tone}(n) = \sum_{k=K_L}^{K_H} \exp(F_T(n, k)) \quad \dots (19)$$

[0152] 然后, 音调强度特征值计算单元 223 如以下式 (20) 所示将音调成分强度 $A_{tone}(n)$ 看作音调强度特征值 $x_2(n)$ 。

[0153] [数学 18]

[0154] $x_2(n) = A_{tone}(n) \dots (20)$

[0155] 频谱大概轮廓特征值计算单元 224 如以下式 (21) 所示获取频谱大概轮廓特征值 $x_3(n)$ 、 $x_4(n)$ 、 $x_5(n)$ 和 $x_6(n)$ 。这里, L 表示特征值的维数, 并且“ $L = 7$ ”的情况在这里被示出。

[0156] [数学 19]

[0157]

$$\begin{aligned}
 x_3(n) &= \sum_{k=0}^{N/2-1} F(k, n) \cos(2\pi k / N) \\
 x_4(n) &= \sum_{k=0}^{N/2-1} F(k, n) \cos(4\pi k / N) \\
 x_5(n) &= \sum_{k=0}^{N/2-1} F(k, n) \cos(6\pi k / N) \\
 x_6(n) &= \sum_{k=0}^{N/2-1} F(k, n) \cos(8\pi k / N)
 \end{aligned}
 \quad \left. \begin{array}{l} \\ \\ \\ \end{array} \right\} \dots (21)$$

[0158] 频谱大概轮廓特征值是通过离散余弦变换形成对数频谱而获得的低维倒频谱(cepstrum)。对四维或者更少维的系数的以上描述被给出,并且更高维的系数也可被使用。另外,诸如所谓的MFCC(美尔频率倒谱系数(Mel-Frequency Cepstral Coefficient))之类的通过扭曲频率轴并对其执行离散余弦变换而获得的系数也可被使用。

[0159] 前述幅度特征值 $x_0(n)$ 和 $x_1(n)$ 、音调强度特征值 $x_2(n)$ 和频谱大概轮廓特征值 $x_3(n)$ 、 $x_4(n)$ 、 $x_5(n)$ 和 $x_6(n)$ 构成帧 n 中的 L 维(在该情况下为七维)特征值向量 $x(n)$ 。此外,“声音的音量、声音的音高和声音的音调”是三个声音因素,其是指示声音的特性的基本属性。因为特征值向量 $x(n)$ 由幅度(与声音的音量有关)、音调成分强度(与声音的音高有关)和频谱大概轮廓(与声音的音调有关)构成,因此特征值向量 $x(n)$ 构成与所有三种声音因素有关的特征值。

[0160] 分数计算单元 225 综合特征值向量 $x(n)$ 的因素并且通过从 0 到 1 的分数 $S(n)$ 来表示帧 n 是否为包括实际将被登记的有意义声音(检测目标声音)的声音区间。这例如可以通过以下式(22)来获得。这里, $\text{sigm}()$ 是 S 型函数, u_i 、 v_i 和 w_i ($i = 0, \dots, L-1$) 是基于经验从样本数据中选择的常数。

[0161] [数学 20]

$$[0162] \quad S(n) = \text{Sigm}\left(\sum_{i=0}^{L-1} w_i \xi_i(x_i(n)) + w_L\right)$$

$$[0163] \quad \xi_i(x_i) = \text{Sigm}(u_i x_i(n) + v_i) \quad \dots (22)$$

[0164] 时间平滑单元 226 在时间方向上使已经由分数计算单元 225 获得的分数 $S(n)$ 平滑。在平滑处理中,移动平均数可被简单获得,或者诸如中值滤波器之类的用于获取中间值的滤波器可被使用。以下式(23)示出了其中通过求平均处理获得经平滑分数 $S_a(n)$ 的示例。这里,德尔塔 n 表示滤波器的大小,其是基于经验确定的常数。

[0165] [数学 21]

$$[0166] \quad S_a(n) = \frac{1}{2\Delta n + 1} \sum_{\tau=n-\Delta n}^{n+\Delta n} S(n) \quad \cdots (23)$$

[0167] 阈值确定单元 227 将已经由时间平滑单元 226 获得的每个帧 n 中的经平滑分数 $S_a(n)$ 与阈值相比较, 将包括等于或者大于阈值的经平滑分数 $S_a(n)$ 的帧区间确定为声音区间, 并且输出指示该帧区间的声音区间信息。

[0168] 对图 4 中示出的声音区间检测单元 202 的操作的描述将被给出。通过由麦克风 201 采集将被登记的检测目标声音 (由家用电器生成的运行状态声) 而获得的时间信号 $f(t)$ 被提供给时频变换单元 221。时频变换单元 221 对输入时间信号 $f(t)$ 执行时频变换并且获得时频信号 $F(n, k)$ 。时频信号 $F(n, k)$ 被提供给幅度特征值计算单元 222、音调强度特征值计算单元 223 和频谱大概轮廓特征值计算单元 224。

[0169] 幅度特征值计算单元 222 根据时频信号 $F(n, k)$ 计算幅度特征值 $x_0(n)$ 和 $x_1(n)$ (见式 (5))。此外, 音调强度特征值计算单元 223 根据时频信号 $F(n, k)$ 计算音调强度特征值 $x_2(n)$ (见式 (20))。另外, 频谱大概轮廓特征值计算单元 224 计算频谱大概轮廓特征值 $x_3(n)$ 、 $x_4(n)$ 、 $x_5(n)$ 和 $x_6(n)$ (见式 (21))。

[0170] 幅度特征值 $x_0(n)$ 和 $x_1(n)$ 、音调强度特征值 $x_2(n)$ 以及频谱大概轮廓特征值 $x_3(n)$ 、 $x_4(n)$ 、 $x_5(n)$ 和 $x_6(n)$ 被作为帧 n 中的 L 维 (在该情况下为七维) 特征值向量 $x(n)$ 提供给分数计算单元 225。分数计算单元 225 综合特征值向量 $x(n)$ 的因素并且计算从 0 到 1 的分数 $S(n)$, 分数 $S(n)$ 表示帧 n 是否为包括实际将被登记的有意义声音 (检测目标声音) 的声音区间 (见式 (22))。分数 $S(n)$ 被提供给时间平滑单元 226。

[0171] 时间平滑单元 226 在时间方向上使分数 $S(n)$ 平滑 (见式 (23)), 并且经平滑的分数 $S_a(n)$ 被提供给阈值确定单元 227。阈值确定单元 227 将每个帧 n 中的经平滑分数 $S_a(n)$ 与阈值相比较, 将包括等于或者大于阈值的经平滑分数 S_a 的帧区间确定为声音区间, 并且输出指示该帧区间的声音区间信息。

[0172] 在图 4 中示出的声音区间检测单元 202 从输入时间信号 $f(t)$ 的时频分布 $F(n, k)$ 中提取每个时间帧中的幅度、音调成分强度和频谱大概轮廓的特征值并且从这些特征值获取表示每个时间帧的声音区间似然性的分数 $S(n)$ 。因此, 即使检测到的将被登记的声音是在噪声环境下记录的也可以精确地获得指示检测到的声音的区间的声音区间信息。

[0173] “特征值提取单元”

[0174] 图 12 示出了特征值提取单元 203 的一个配置示例。特征值提取单元 203 将通过由麦克风 201 记录将被登记的检测目标声音 (由家用电器生成的运行状态声) 而获得的时间信号 $f(t)$ 获取作为输入, 并且时间信号 $f(t)$ 如图 3 所示还包括检测目标声音之前和之后的噪声区间。此外, 特征值提取单元 203 输出在将被登记的检测目标声音的区间中按照每隔预定时间提取的特征值序列。

[0175] 特征值提取单元 203 包括时频变换单元 241、音调似然分布检测单元 242、时频平滑单元 243 以及稀疏和量化单元 244。时频变换单元 241 以与前面提到的声音区间检测单元 202 的时频变换单元 221 相同的方式对输入时间信号 $f(t)$ 执行时频变换并且获得时频

信号 $F(n, k)$ 。此外,特征值提取单元 203 可使用声音区间检测单元 202 的时频变换单元 221 获得的时频信号 $F(n, k)$,并且在这种情况下,提供时频变换单元 241 不是必需的。

[0176] 音调似然分布检测单元 242 基于来自声音区间检测单元 202 的声音区间信息来检测声音区间中的音调似然分布。就是说,音调似然分布检测单元 242 首先以与前面提到的声音区间检测单元 202 的音调强度特征值计算单元 223 相同的方式将时频信号 $F(n, k)$ 的分布(见图 5A)变换为音调特性似然性的分数 $S(n, k)$ 的分布(见图 5B)。

[0177] 音调似然分布检测单元 242 接下来如以下式 (24) 所示通过使用声音区间信息来获取包括将被登记的有意义声音(检测目标声音)的声音区间中的音调似然分布 $Y(n, k)$ 。

[0178] [数学 22]

[0179]

$$Y(n, k) = \begin{cases} S(n, k) & \text{当 } n \text{ 在声音部分内部时} \\ 0 & \text{当 } n \text{ 在声音部分外部时} \end{cases} \quad \dots (24)$$

[0180] 时频平滑单元 243 在时间方向和频率方向上使已经由音调似然分布检测单元 242 获得的声音区间中的音调似然分布 $Y(n, k)$ 平滑并且如以下式 (25) 所示获取经平滑的音调似然分布 $Y_a(n, k)$ 。

[0181] [数学 23]

$$Y_a(n, k) = \sum_{\tau=-\Delta_n}^{\Delta_n} \sum_{\lambda=-\Delta_k}^{\Delta_k} Y(n-\tau, k-\lambda) H(\tau, \lambda) \quad \dots (25)$$

[0183] 这里,德尔塔 k 表示平滑滤波器在频率方向的一侧上的大小,德尔塔 n 表示其在时间方向的一侧上的大小,并且 $H(n, k)$ 表示平滑滤波器的二次脉冲响应。此外,为了简化而对在频率方向上没有失真的滤波器的情况给出上面的描述。然而,可使用诸如美尔(Mel)频率之类的使频率轴失真的滤波器来执行平滑。

[0184] 稀疏和量化单元 344 使由时频平滑单元 243 获得的经平滑的音调似然分布 $Y_a(n, k)$ 稀疏,进一步量化音调似然分布 $Y_a(n, k)$,并且如以下式 (26) 所示创建将被登记的有意义声音(检测目标声音)的特征值 $Z(m, l)$ 。

[0185] [数学 24]

$$z(m, l) = \text{Quant}[Y_a(mT, lK)] \quad (0 \leq m \leq M-1, 0 \leq l \leq L-1)$$

[0187] ... (26)

[0188] 这里, T 表示时间方向上的离散化步长, K 表示频率方向上的离散化步长, m 表示经稀疏的离散时间,并且 l 表示经稀疏的离散频率。此外, M 表示时间方向上的帧的数目(对应于将被登记的有意义声音(检测目标声音)的时间长度), L 表示频率方向上的维数,并且 $\text{Quant}[]$ 表示量化的函数。

[0189] 前述特征值 $z(m, l)$ 通过如以下式 (27) 所示的频率方向上的集合向量符号可被表示为 $Z(m)$ 。

[0190] [数学 25]

[0191] $Z(m) = [z(m, 0), \dots, z(m, L-1)] (0 \leq m \leq M-1) \dots (27)$

[0192] 在这种情况下, 前述特征值 $Z(m, 1)$ 由已经在时间方向上每隔 T 提取的 M 个向量 $Z(0), \dots, Z(M-1), Z(M)$ 构成。因此, 稀疏和量化单元 244 可以获得在包括将被登记的检测目标声音的区间中按照每隔预定时间提取的特征值 (向量) 的序列 $Z(m)$ 。

[0193] 此外, 还可以考虑将已经由时频平滑单元 243 获得的经平滑音调似然分布 $Y_a(n, k)$ 原样用作来自特征值提取单元 203 的输出, 即特征值序列。然而, 因为音调似然分布 $Y_a(n, k)$ 已经被平滑, 因此没有必要准备所有的时间和频率数据。通过在时间方向和频率方向上进行稀疏可以减少信息量。此外, 通过量化可以将 8 位或者 16 位的数据变换为 2 位或者 3 位的数据。因为稀疏和量化被如上所述执行, 因此可以减少特征值 (向量) 序列 $Z(m)$ 上的信息量并因而可以减小稍后将描述的声音检测装置 100 的匹配计算的处理负担。

[0194] 将给出对图 12 中示出的特征值提取单元 203 的操作的描述。通过由麦克风 201 采集将被登记的检测目标声音 (家用电器生成的运行状态声) 而获得的时间信号 $f(t)$ 被提供给时频变换单元 241。时频变换单元 241 对输入时间信号 $f(t)$ 执行时频变换并且获得时频信号 $F(n, k)$ 。时频信号 $F(n, k)$ 被提供给音调似然分布检测单元 242。此外, 由声音区间检测单元 202 获得的声音区间信息也被提供给音调似然分布检测单元 242。

[0195] 音调似然分布检测单元 242 将时频信号 $F(n, k)$ 的分布变换为音调特性似然性的分数 $S(n, k)$ 的分布, 并且通过使用声音区间信息进一步获取包括将被登记的有意义声音 (检测目标声音) 的声音区间中的音调似然分布 $Y(n, k)$ (见式 (24))。音调似然分布 $Y(n, k)$ 被提供给时频平滑单元 243。

[0196] 时频平滑单元 243 在时间方向和频率方向上使音调似然分布 $Y(n, k)$ 平滑并且获得经平滑的音调似然分布 $Y_a(n, k)$ (见式 (25))。音调似然分布 $Y_a(n, k)$ 被提供给稀疏和量化单元 244。稀疏和量化单元 244 使音调似然分布 $Y_a(n, k)$ 稀疏, 进一步量化经稀疏的音调似然分布 $Y_a(n, k)$, 并且获得将被登记的有意义声音 (检测目标声音) 的特征值 $z(m, 1)$, 即特征值序列 $Z(m)$ (见式 (26) 和 (27))。

[0197] 返回到图 2, 特征值登记单元 204 将已经由特征值登记单元 204 创建的、将被登记的检测目标声音的特征值序列 $Z(m)$ 与特征值数据库 103 中的检测目标声音名称 (关于操作情况声的信息) 相关联并且登记。

[0198] 将给出对图 2 中示出的特征值登记装置 200 的操作的描述。麦克风 201 采集将被登记为检测目标声音的家用电器的运行状态声。从麦克风 201 输出的时间信号 $f(t)$ 被提供给声音区间检测单元 202 和特征值提取单元 203。声音区间检测单元 202 检测从输入时间信号 $f(t)$ 中检测声音区间, 即包括由家用电器生成的运行状态声的区间, 并且输出声音区间信息。声音区间信息被提供给特征值提取单元 203。

[0199] 特征值提取单元 203 对每个时间帧的输入时间信号 $f(t)$ 执行时频变换, 获取时频信号 $F(n, k)$ 的分布, 并且进一步从时频分布获取音调似然分布, 即分数 $S(n, k)$ 的分布。然后, 特征值提取单元 203 基于声音区间信息从分数 $S(n, k)$ 的分布获取声音区间的音调似然分布 $Y(n, k)$, 在时间方向和频率方向上使音调似然分布 $Y(n, k)$ 平滑, 并且进一步对其执行稀疏和量化处理以创建特征值序列 $Z(m)$ 。

[0200] 已经由特征值提取单元 203 创建的将被登记的检测目标声音 (家用电器的运行状态声) 的特征值序列 $Z(m)$ 被提供给特征值登记单元 204。特征值登记单元 204 将特征值序

列 $Z(m)$ 与特征值数据库 103 中的检测目标声音名称（关于运行状态声的信息）相关联并且登记。在以下描述中，假定 I 个检测目标声音项被登记，其特征值序列将被表示为 $Z1(m)$ 、 $Z2(m)$ 、 $\dots$ 、 $Zi(m)$ 、 $\dots$ 、 $ZI(m)$ ，并且特征值序列中的时间帧的数目（时间方向上排列的向量的数目）将被表示为 $M1$ 、 $M2$ 、 $\dots$ 、 Mi 、 $\dots$ 、 MI 。

[0201] “声音检测单元”

[0202] 图 13 示出了声音检测单元 102 的一个配置示例。声音检测单元 102 包括信号缓冲单元 121、特征值提取单元 122、特征值缓冲单元 123 和比较单元 124。信号缓冲单元 121 缓冲通过由麦克风 101 采集声音而获得的时间信号 $f(t)$ 的预定数目的信号样本。预定数目意味着特征值提取单元 122 可用来新计算与一个帧相对应的特征值序列的样本的数目。

[0203] 特征值提取单元 122 基于已经由信号缓冲单元 121 缓冲的时间信号 $f(t)$ 的信号样本提取每隔预定时间的特征值。尽管未被详细描述，但是特征值提取单元 203 被以与前面提到的特征值登记装置 200 的特征值提取单元 203（见图 12）相同的方式配置。

[0204] 然而，特征值提取单元 122 中的音调似然检测单元 242 获取所有区间中的音调似然分布 $Y(n, k)$ 。就是说，音调似然分布检测单元 242 原样输出已经从时频信号 $F(n, k)$ 的分布获得的分数 $S(n, k)$ 的分布。然后，稀疏和量化单元 244 针对输入时间信号 $f(t)$ 的所有区间输出按照 T 新提取的特征值（向量） $X(n)$ （时间方向上的离散步长）。这里， n 表示当前正被提取的特征值的帧的数目（对应于当前离散时间）。

[0205] 特征值缓冲单元 123 如图 14 所示保存从特征值提取单元 122 输出的最新 N 个特征值（向量） $X(n)$ 。这里， N 至少是等于或者大于特征值数据库 103 中登记的特征值序列 $Z1(m)$ 、 $Z2(m)$ 、 $\dots$ 、 $Zi(m)$ 、 $\dots$ 、 $ZI(m)$ 当中的最长特征值序列的帧的数目（在时间方向上对齐的向量的数目）的数目。

[0206] 比较单元 124 每当特征值提取单元 122 提取新特征值 $X(n)$ 时顺序地将信号缓冲单元 123 中保存的特征值序列与特征值数据库 103 中登记的 I 个检测目标声音项的特征值序列相比较，并且获取 I 个检测目标声音项的检测结果。这里，如果 i 表示检测目标声音数目的数目，则每个检测目标声音项的长度（帧数目 Mi ）各不相同。

[0207] 如图 14 所示，比较单元 124 将特征值缓冲单元 123 中的最新帧 n 与检测目标声音的特征值序列的最新帧 $Zi(Mi-1)$ 相结合并且通过使用具有特征值缓冲单元 123 中保存的 N 个特征值之中的检测目标声音的特征值序列的长度的帧来计算相似性。例如如以下式 (28) 所示可通过特征值之间的关联来计算相似性 $\text{Sim}(n, i)$ 。这里， $\text{Sim}(n, i)$ 意味着与第 n 个帧中的第 i 个检测目标声音的特征值序列的相似性。比较单元 124 确定“第 i 个检测目标声音是在时间 n 生成的”并且当相似性大于预定阈值时输出确定结果。

[0208] [数学 26]

[0209]

$$\text{Sim}(n, i) = \frac{\sum_{m=0}^{M_i-1} X(n - M_i - 1 + m) Z_i(m)}{\sqrt{\sum_{m=0}^{M_i-1} X^2(n - M_i - 1 + m) \sum_{m=0}^{M_i-1} Z_i^2(m)}} \quad \dots (28)$$

[0210] 将给出对在图 13 中示出的声音检测单元 102 的操作的描述。通过由麦克风 101 采集声音而获得的时间信号 $f(t)$ 被提供给信号缓冲单元 121, 并且预定数目的信号样本被缓冲。特征值提取单元 122 基于由信号缓冲单元 121 缓冲的时间信号 $f(t)$ 的信号样本提取每隔预定时间的特征值。然后, 特征值提取单元 122 顺序地输出新提取的、每隔 T (时间方向上的离散步长) 的特征值 (向量) $X(n)$ 。

[0211] 已经由特征值提取单元 122 提取的特征值 $X(n)$ 被提供给特征值缓冲单元 123, 并且最新的 N 个特征值 $X(n)$ 被保存在其中。每当新的特征值 $X(n)$ 被特征值提取单元 122 提取时, 比较单元 124 顺序地将特征值缓冲单元 123 中保存的特征值序列与特征值数据库 103 中登记的 I 个检测目标声音项的特征值序列相比较, 并且获取 I 个检测目标声音项的检测结果。

[0212] 在这种情况下, 比较单元 124 将特征值缓冲单元 123 中的最新帧 n 与检测目标声音的特征值序列的最新帧 $Z_i(M_i-1)$ 相结合并且通过使用具有检测目标声音的特征值序列的长度的帧来计算相似性 (见图 14)。然后, 比较单元 124 确定“第 i 个检测目标声音是在时间 n 生成的”并且当相似性大于预定阈值时输出确定结果。

[0213] 此外, 图 1 中示出的声音检测装置 100 可被配置为硬件或者软件。例如, 可以使图 15 中示出的计算机装置 300 包括图 1 中示出的声音检测装置 100 的部分或者所有功能并且执行与上述者相同的检测检测目标声音的处理。

[0214] 计算机装置 300 包括 CPU (中央处理单元) 301、ROM (只读存储器) 302、RAM (随机访问存储器) 303、数据输入和输出单元 (数据 I/O) 304 和 HDD (硬盘驱动器) 305。ROM 302 存储 CPU 301 的处理程序等。RAM 303 起 CPU 301 的工作区域的作用。CPU 301 按需读取 ROM 302 上存储的处理程序, 将所读取的处理程序传送到 RAM 303 并且在 RAM 303 中开发所读取的处理程序, 读取所开发的处理程序, 并且执行音调成分检测处理。

[0215] 输入时间信号 $f(t)$ 被经由数据 I/O 304 输入到计算机装置 300 并且在 HDD 305 中累积。CPU 301 如上所述对 HDD 305 中累积的输入时间信号 $f(t)$ 执行检测检测目标声音的处理。然后, 检测结果被经由数据 I/O 304 输出到外部。此外, I 个检测目标声音项的特征值序列被预先登记并保存在 HDD 305 中。

[0216] 图 16 中的流程图示出了由 CPU 301 检测检测目标声音的处理过程的一个示例。在步骤 ST21 中, CPU 301 开始处理然后转移到步骤 ST22 中的处理。在步骤 ST22 中, CPU 301 例如将输入时间信号 $f(t)$ 输入到 HDD 305 中配置的信号缓冲单元。然后, CPU 301 在步骤 ST23 中判定是否已累积可用来计算对应于一个帧的特征值序列的数目的样本。

[0217] 如果已累积对应于一个帧的数目的样本, 则 CPU 301 在步骤 ST24 中执行提取特征值 $X(n)$ 的处理。CPU 301 在步骤 ST25 中例如将所提取的特征值 $X(n)$ 输入到 HDD 305 中配

置的特征值缓冲单元。然后, CPU 301 在步骤 ST26 中将检测目标声音的数目 i 设置为零。

[0218] 接下来, CPU 301 在步骤 ST27 中判定“ $i < I$ ”是否得到满足。如果“ $i < I$ ”得到满足, 则 CPU 301 在步骤 ST28 中计算信号缓冲单元中保存的特征值序列与 HDD 305 中登记的第 i 个检测目标声音的特征值序列 $Z_i(m)$ 之间的相似性。然后, CPU 301 在步骤 ST29 中确定“相似性 $>$ 阈值”是否得到满足。

[0219] 如果“相似性 $>$ 阈值”得到满足, 则 CPU 301 在步骤 ST30 中输出指示符合的结果。就是说, “第 i 个检测目标声音在时间 n 被生成”的判定结果被输出为检测输出。此后, CPU 301 在步骤 ST31 中使 i 增加并且返回步骤 ST27 中的处理。此外, 如果“相似性 $>$ 阈值”在步骤 ST29 中未得到满足, 则 CPU 301 紧接着在步骤 ST31 中使 i 增加并且返回到步骤 ST27 中的处理。如果“ $i > I$ ”在步骤 ST27 中未得到满足, 则 CPU 301 判定对当前帧的处理已经完成, 返回到步骤 ST22 中的处理, 并且转移到对下一帧的处理。

[0220] 接下来, CPU 301 在步骤 ST3 中将帧(时间帧)的数目 n 设置为 0。然后, CPU 301 在步骤 ST4 中判定“ $n < N$ ”是否得到满足。此外, 假定频谱图(时频分布)的帧存在于从 0 到“ $N-1$ ”。如果“ $n < N$ ”未得到满足, 则 CPU 301 判定所有帧的处理已经完成并且随后在步骤 ST5 中完成处理。

[0221] 如果“ $n < N$ ”得到满足, 则 CPU 301 在步骤 ST6 中将离散频率 k 设置为 0。然后, CPU 301 在步骤 ST7 中判定“ $k < K$ ”是否得到满足。此外, 假定频谱图(时频分布)的离散频率 k 存在于从 0 到“ $k-1$ ”。如果“ $k < K$ ”未得到满足, 则 CPU 301 判定对所有离散频率的处理已经完成, 在步骤 ST8 中使 n 增加, 然后返回到步骤 ST4, 并且转移到对下一帧的处理。

[0222] 如果“ $k < K$ ”在步骤 ST7 中得到满足, 则 CPU 301 在步骤 ST9 中判定 $F(n, k)$ 是否对应于峰。如果 $F(n, k)$ 不对应于峰, 则 CPU301 在步骤 ST10 中将分数 $S(n, k)$ 设置为 0, 在步骤 ST11 中使 k 增加, 然后返回到步骤 ST7, 并且转移到对下一离散频率的处理。

[0223] 如果 $F(n, k)$ 在步骤 ST9 中对应于峰, 则 CPU 301 转移到步骤 ST12 中的处理。在步骤 ST12 中, CPU 301 在峰附近的区域中拟合音调模型。然后, CPU 301 在步骤 ST13 中基于拟合结果提取各种特征值(x_0 、 x_1 、 x_2 、 x_3 、 x_4 和 x_5)。

[0224] 接下来, 在步骤 ST14 中, CPU 301 通过使用在步骤 ST13 中提取的特征值来获取分数 $S(n, k)$, 分数 $S(n, k)$ 表示具有从 0 到 1 的值的峰的音调成分似然性。CPU 301 在步骤 ST14 中的处理之后在步骤 ST11 中使 k 增加, 然后返回到步骤 ST7, 并且转移到对下一离散频率的处理。

[0225] 如上所述, 图 1 中示出的声音检测装置 100 根据通过由麦克风 101 采集声音而获得的输入时间信号 $f(t)$ 的时频分布获取音调似然分布并且从在频率方向和时间方向上已被平滑的似然分布中提取并使用每隔预定时间的特征值。相应地, 可以在不依赖于麦克风 101 的安装位置的情况下精确地检测检测目标声音(从家用电器生成的运行状态声等)。

[0226] 此外, 图 1 中示出的声音检测装置 100 将已经由声音检测单元 102 获得的检测目标声音的检测结果与时间一起记录在记录介质上并显示在显示器上。因此, 可以在家自动地记录家用电器等的运行状态并且获得自身行动历史(所谓的生活记录)。此外, 可以为具有听觉困难的人自动地将声音通知可视化。

[0227] <2. 修改示例>

[0228] 上面的实施例示出了一个示例, 其中从在家的家用电器生成的运行状态声(控制

声、通知声、操作声、警报声等)被检测到。然而,本技术可被应用于与在制造厂中装配的产品的声音功能有关的自动检测中的用途以及家庭用途。此外,事实上本技术不仅可被应用于对运行状态声的检测而且可被应用于对特定人或者特定动物的语音声或者其他环境声的检测。

[0229] 尽管以上描述是对其中基于短时间傅立叶变换执行时频变换的实施例给出的,但是也可考虑通过使用诸如小波变换之类的另一变换方法来使输入时间信号经受时频变换。此外,尽管以上描述是对其中基于每个检测到的峰附近的时频分布与音调模型之间的均方误差最小准则来执行拟合的实施例给出的,但是也可考虑基于四次方误差最小准则、最小熵准则等来执行拟合。

[0230] 此外,本技术可被配置如下。

[0231] (1) 一种声音检测装置,包括:特征值提取单元,其从输入时间信号中提取每隔预定时间的特征值;特征值保持单元,其保持预定数目的检测目标声音项的特征值序列;以及比较单元,其每当特征值提取单元新提取出特征值时分别将特征值提取单元提取的特征值序列与保持的预定数目的检测目标声音项的特征值序列相比较并且获取预定数目的检测目标声音项的检测结果,其中特征值提取单元包括对每个时间帧的输入时间信号执行时频变换并获取时频分布的时频变换单元以及从该时频分布获取音调似然分布的似然分布检测单元,该特征值提取单元在频率方向和时间方向上使获取的似然分布平滑并且提取每隔预定时间的特征值。

[0232] (2) 根据(1)所述的装置,其中似然分布检测单元包括在时频分布的每个时间帧中检测频率方向上的峰的峰检测单元、在每个检测到的峰处拟合音调模型的拟合单元、以及基于拟合结果获取表示每个检测到的峰处的音调成分似然性的分数的评分单元。

[0233] (3) 根据(1)或(2)所述的装置,其中特征值提取单元还包括在频率方向和/或时间方向上使经平滑的似然分布稀疏的稀疏单元。

[0234] (4) 根据(1)或(2)所述的装置,其中特征值提取单元还包括将经平滑的似然分布量化的量化单元。

[0235] (5) 根据(1)至(4)中任一个所述的装置,其中比较单元基于保持的检测目标声音项的特征值序列和由特征值提取单元针对预定数目的检测目标声音项中的每一个提取的特征值序列之间的对应特征值之间的关联来获取相似性,并且基于所获取的相似性获取检测目标声音项的检测结果。

[0236] (6) 根据(1)至(5)中任一个所述的装置,还包括:

[0237] 记录控制单元,其将预定数目的检测目标声音项的检测结果与时间信息一起记录在记录介质上。

[0238] (7) 一种声音检测方法,包括:从输入时间信号中提取每隔预定时间的特征值;以及每当在特征值的提取中新提取出特征值时分别将由特征值提取单元提取的特征值序列与所保持的预定数目的检测目标声音项的特征值序列相比较,并且获取预定数目的检测目标声音项的检测结果,其中在特征值的提取中,对每个时间帧的输入时间信号执行时频变换,获取时频分布,从时频分布获取音调似然分布,在频率方向和时间方向上使该似然分布平滑,并且提取每隔预定时间的特征值。

[0239] (8) 一种程序,其使计算机执行以下操作:从输入时间信号中提取每隔预定时间

的特征值；以及每当在特征值的提取中新提取出特征值时分别将由特征值提取单元提取的特征值序列与所保持的预定数目的检测目标声音项的特征值序列相比较，并且获取预定数目的检测目标声音项的检测结果，其中在特征值的提取中，对每个时间帧的输入时间信号执行时频变换，获取时频分布，从时频分布获取音调似然分布，在频率方向和时间方向上使该似然分布平滑，并且提取每隔预定时间的特征值。

[0240] (9) 一种声音特征值提取装置，包括：时频变换单元，其对每个时间帧的输入时间信号执行时频变换并且获取时频分布；似然分布检测单元，其从该时频分布获取音调似然分布；以及特征值提取单元，其在频率方向和时间方向上使似然分布平滑并且提取每隔预定时间的特征值。

[0241] (10) 根据 (9) 所述的装置，其中似然分布检测单元包括在时频分布的每个时间帧中检测频率方向上的峰的峰检测单元、在每个检测到的峰处拟合音调模型的拟合单元、以及基于拟合结果获取表示每个检测到的峰处的音调成分似然性的分数的评分单元。

[0242] (11) 根据 (9) 或 (10) 所述的装置，还包括：稀疏单元，其在频率方向和 / 或时间方向上使经平滑的似然分布稀疏。

[0243] (12) 根据 (9) 或 (10) 所述的装置，还包括：量化单元，其将经平滑的似然分布量化。

[0244] (13) 根据 (9) 至 (12) 中任一个所述的装置，还包括：声音区间检测单元，其基于输入时间信号检测声音区间，其中似然分布检测单元在检测到的声音区间的范围内从时频分布获取音调似然分布。

[0245] (14) 根据 (13) 所述的装置，其中声音区间检测单元包括对每个时间帧的输入时间信号执行时频变换并且获取时频分布的时频变换单元，基于该时频分布提取每个时间帧的幅度、音调成分强度和频谱大概轮廓的特征值的特征值提取单元，基于提取的特征值获取表示每个时间帧的声音区间似然性的分数的评分单元，在所述时间方向上使获取的每个时间帧的分数平滑的时间平滑单元，以及确定每个时间帧的经平滑分数的阈值并且获取声音区间信息的阈值确定单元。

[0246] (15) 一种声音特征值提取方法，包括：通过对每个时间帧的输入时间信号执行时频变换来获取时频分布；从该时频分布获取音调似然分布；以及在频率方向和时间方向上使该似然分布平滑。

[0247] (16) 一种声音区间检测装置，包括：时频变换单元，其通过对每个时间帧的输入时间信号执行时频变换来获取时频分布；特征值提取单元，其基于该时频分布提取每个时间帧的幅度、音调成分强度和频谱大概轮廓的特征值；以及评分单元，其基于提取的特征值获取表示每个时间帧的声音区间似然性的分数。

[0248] (17) 根据 (16) 所述的装置，还包括：时间平滑单元，其在时间方向上使获取的每个时间帧的分数平滑；以及阈值确定单元，其确定每个时间帧的经平滑分数的阈值并且获取声音区间信息。

[0249] (18) 一种声音区间检测方法，包括：通过对每个时间帧的输入时间信号执行时频变换来获取时频分布；基于该时频分布提取每个时间帧的幅度、音调成分强度和频谱大概轮廓的特征值；以及基于提取的特征值获取表示每个时间帧的声音区间似然性的分数。

[0250] 本公开包含与在 2012 年 4 月 18 日在日本专利局提交的日本优先权专利申请 JP

2012-094395 中公开的主题有关的主题,该申请的全部内容通过引用而结合于此。

[0251] 本领域技术人员应当明白取决于设计要求和其他因素可发生各种修改、组合、子组合和变更,如果它们在所附权利要求或其等价物的范围内的话。

[0252] 标号列表

[0253] 100 :声音检测装置

[0254] 101 :麦克风

[0255] 102 :声音检测单元

[0256] 103 :特征值数据库

[0257] 104 :记录和显示单元

[0258] 121 :信号缓冲单元

[0259] 122 :特征值提取单元

[0260] 123 :特征值缓冲单元

[0261] 124 :比较单元

[0262] 200 :特征值登记装置

[0263] 201 :麦克风

[0264] 202 :声音区间检测单元

[0265] 203 :特征值提取单元

[0266] 204 :特征值登记单元

[0267] 221 :时频变换单元

[0268] 222 :幅度特征值计算单元

[0269] 223 :音调强度特征值计算单元

[0270] 224 :频谱大概轮廓特征值计算单元

[0271] 225 :分数计算单元

[0272] 226 :时间平滑单元

[0273] 227 :阈值确定单元

[0274] 230 :音调似然分布检测单元

[0275] 231 :峰检测单元

[0276] 232 :拟合单元

[0277] 233 :特征值提取单元

[0278] 234 :评分单元

[0279] 241 :时频变换单元

[0280] 242 :音调似然分布检测单元

[0281] 243 :时频变换单元

[0282] 244 :稀疏和量化单元

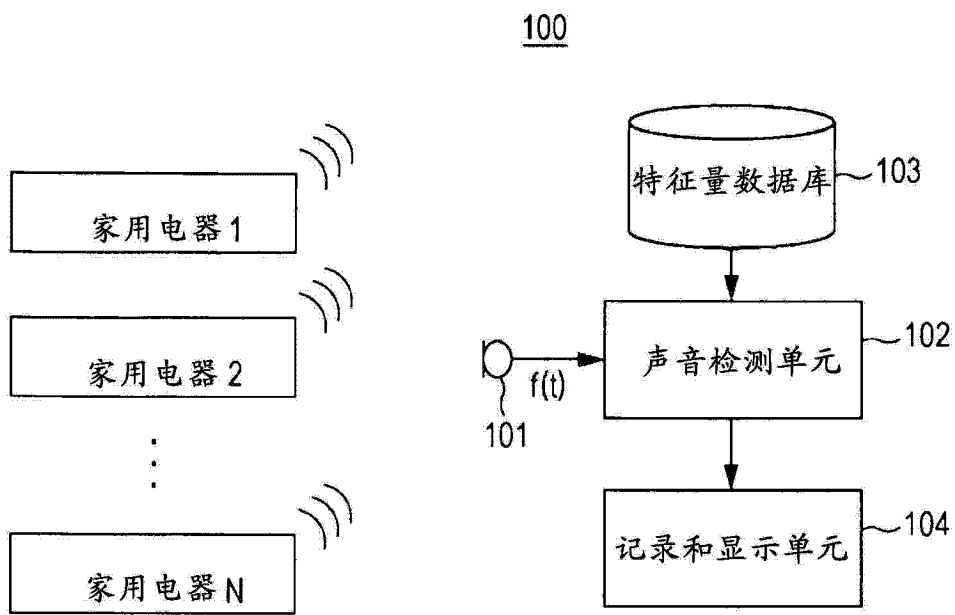


图 1

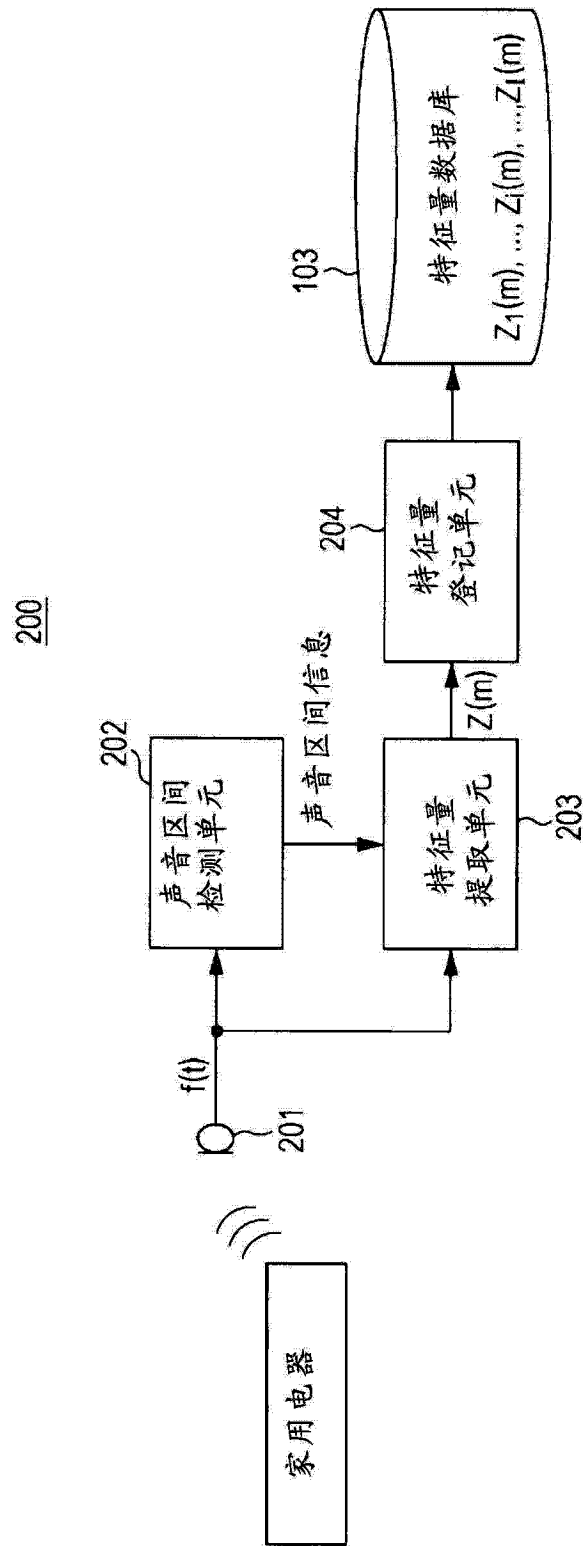


图 2

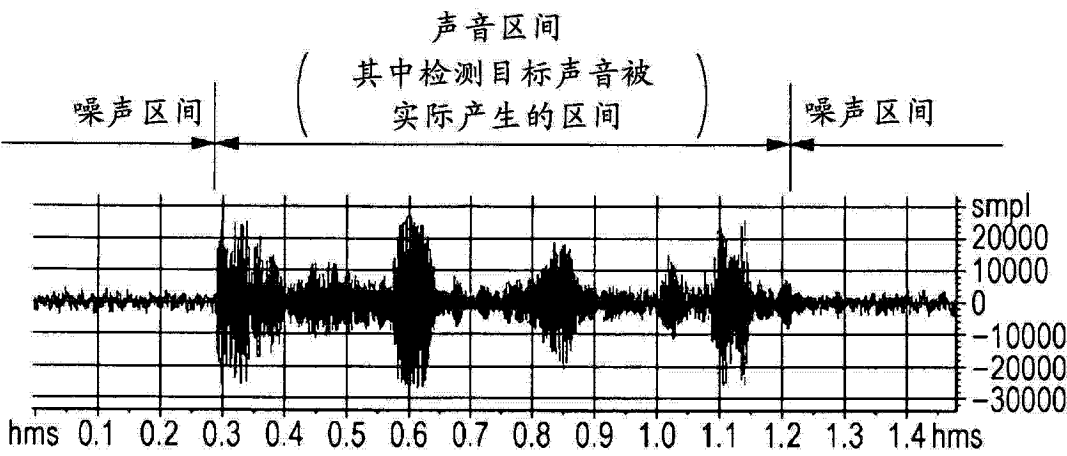


图 3

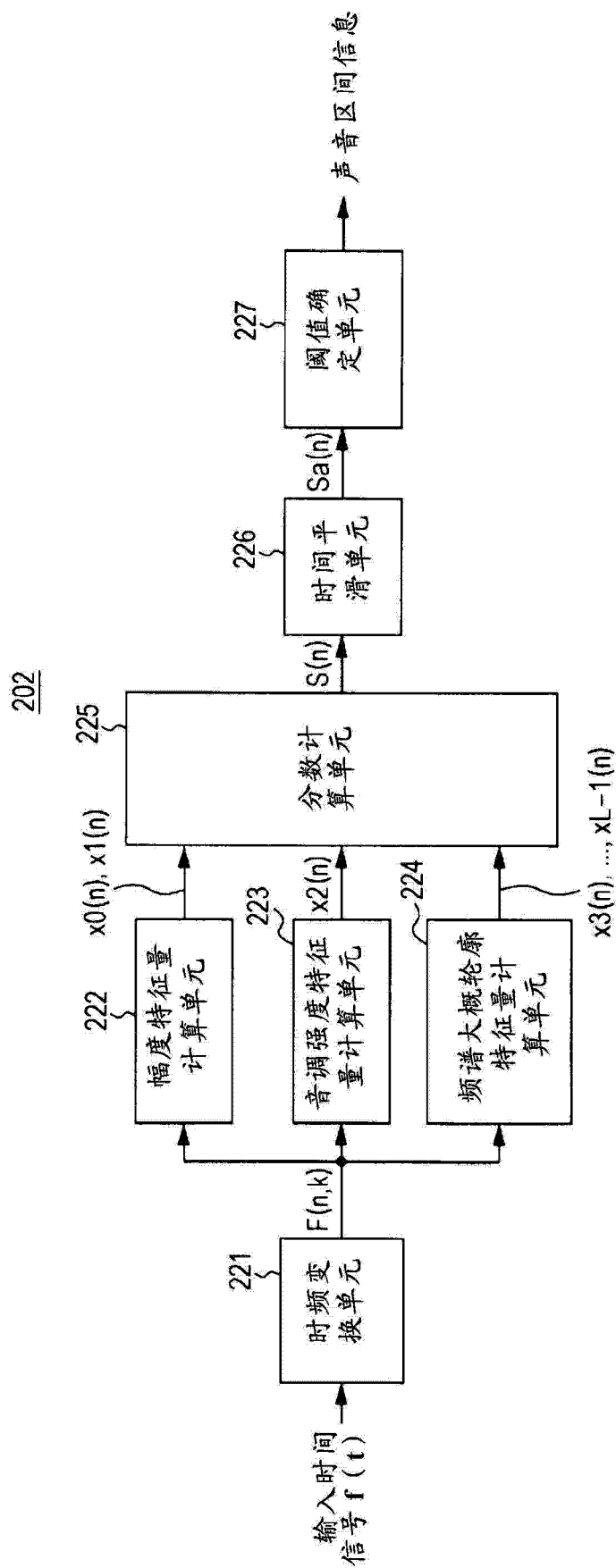


图 4

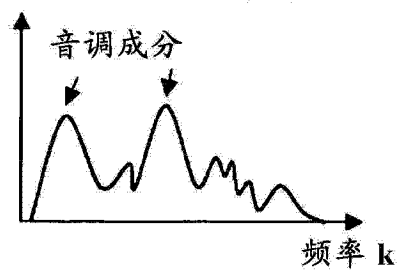
对数幅度频谱 $F(n,k)$ 

图 5A

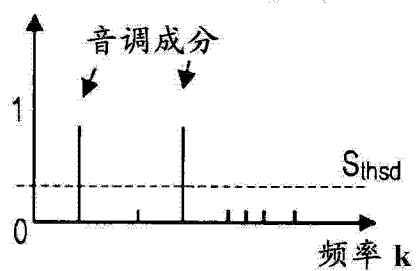
音调似然水平 $S(n,k)$ 

图 5B

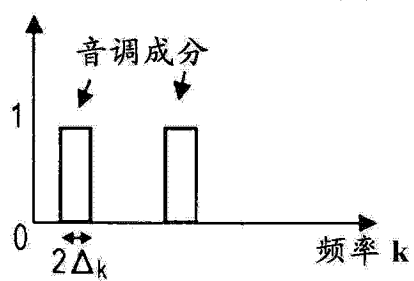
音调成分提取滤波器 $H(k)$ 

图 5C

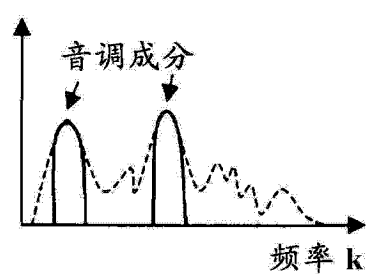
音调成分频谱 $F_T(n, k)$ 

图 5D

230

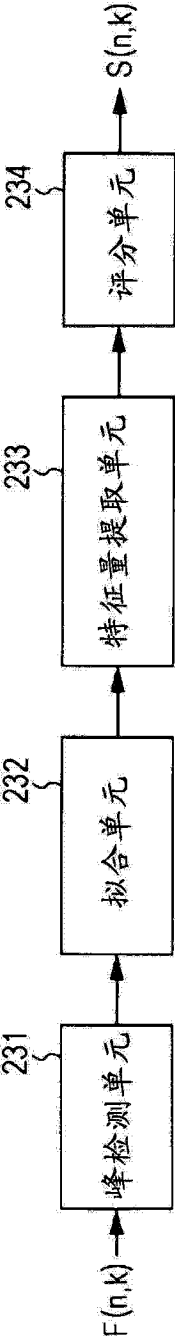


图 6

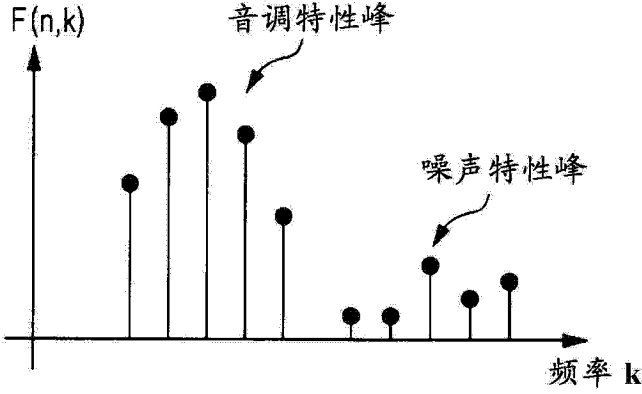


图 7A

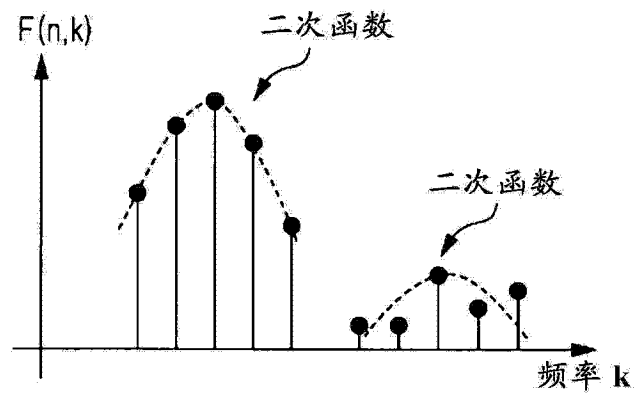


图 7B

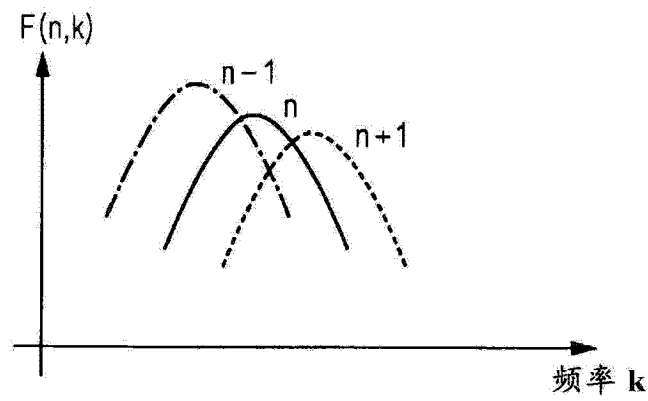


图 8A

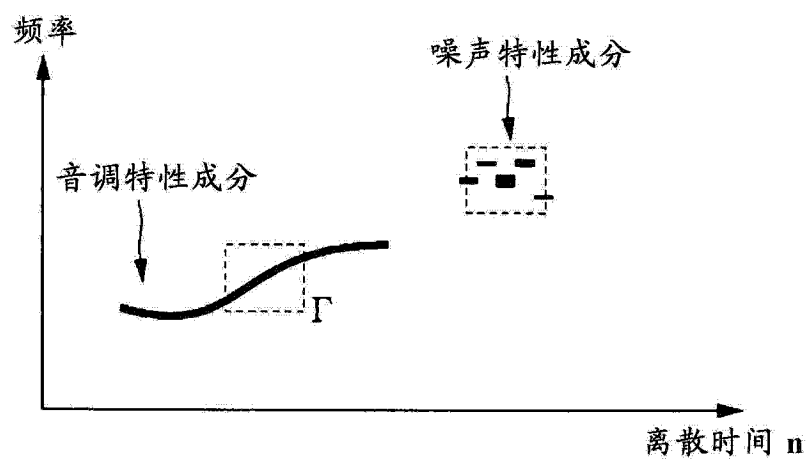


图 8B

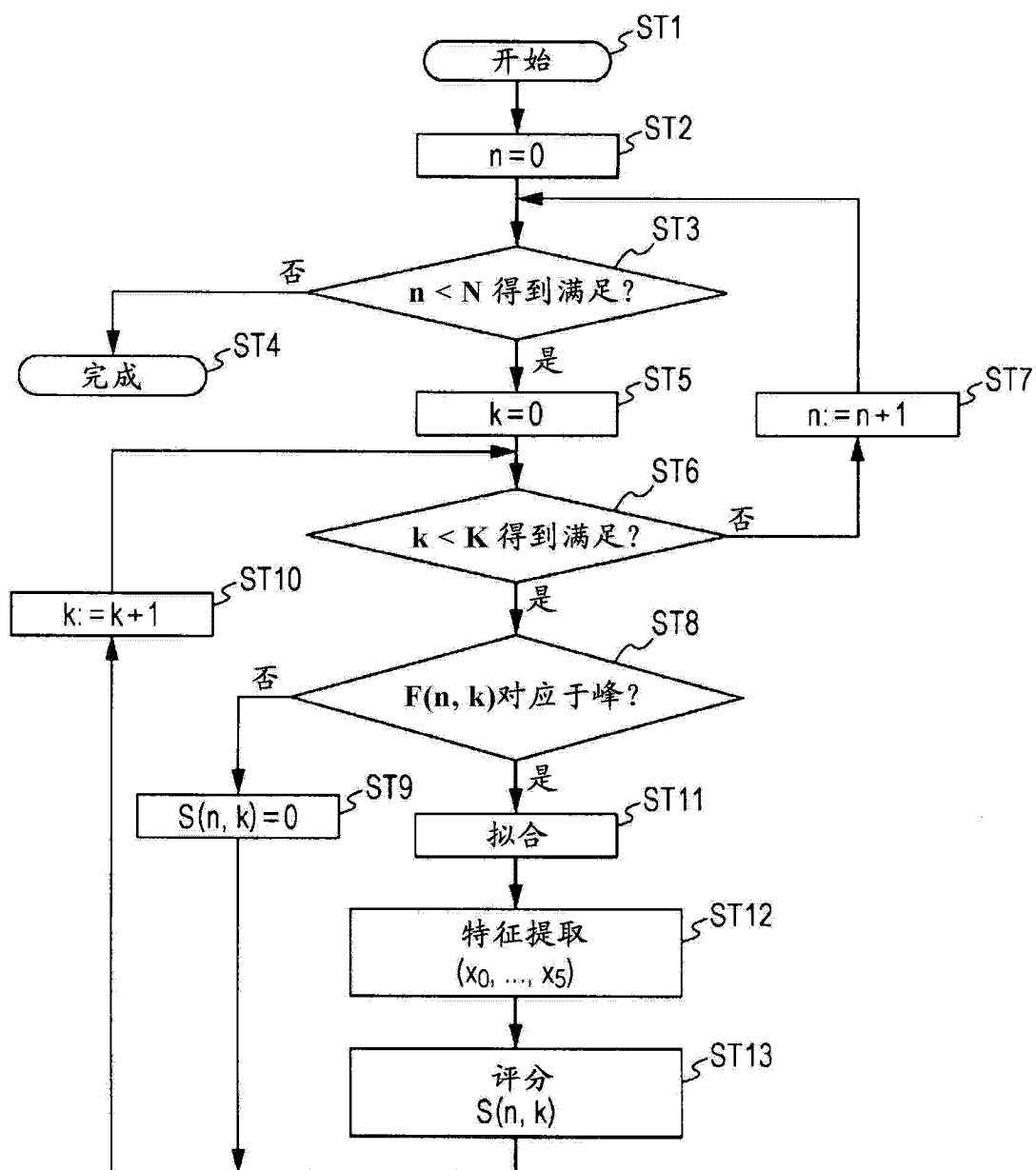


图 9

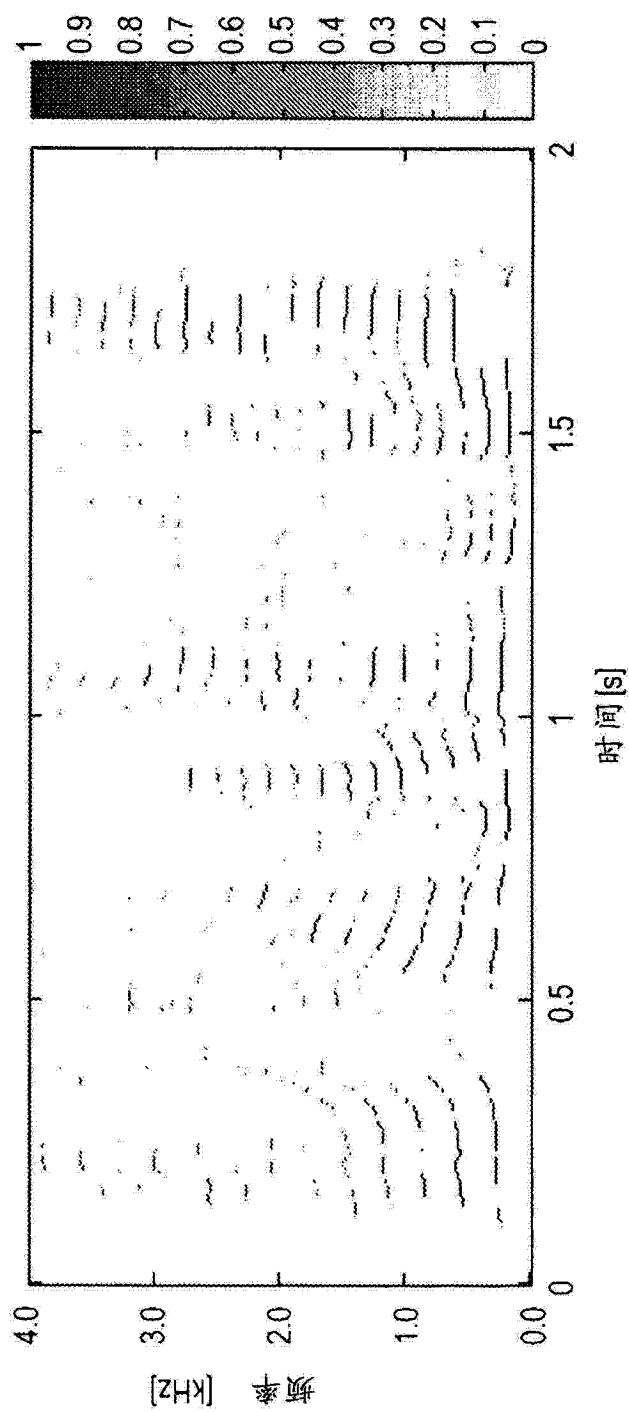


图 10

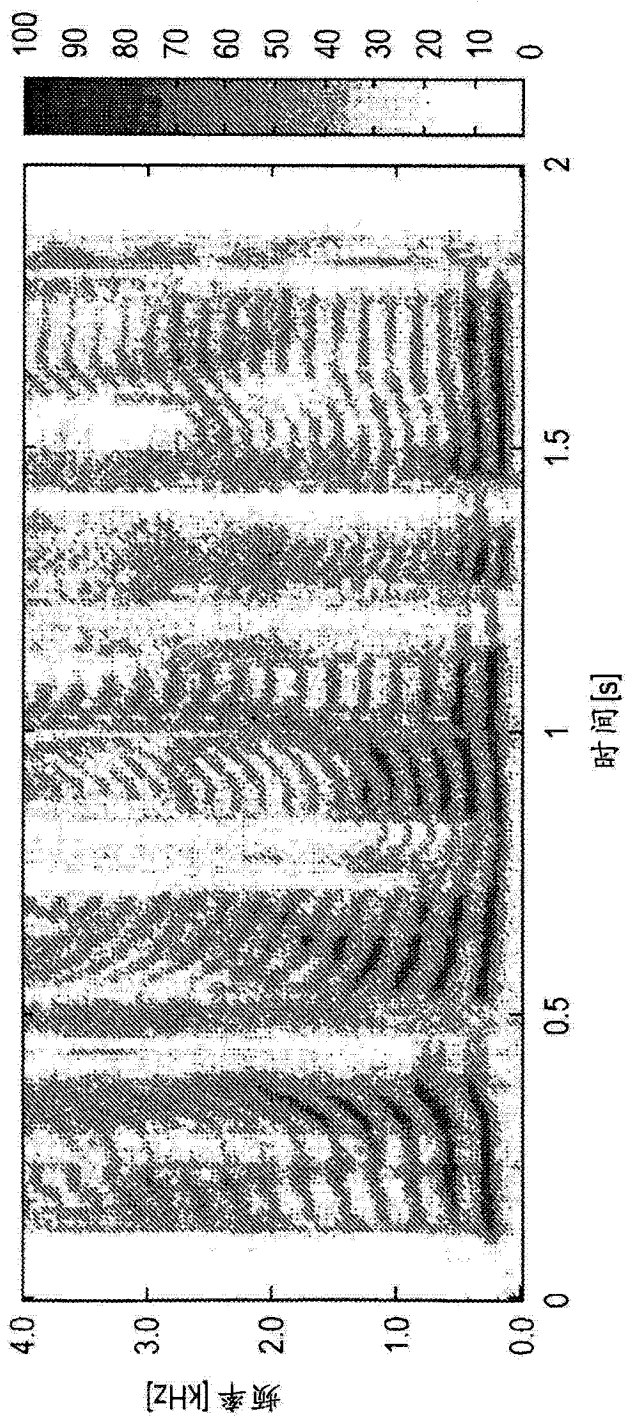


图 11

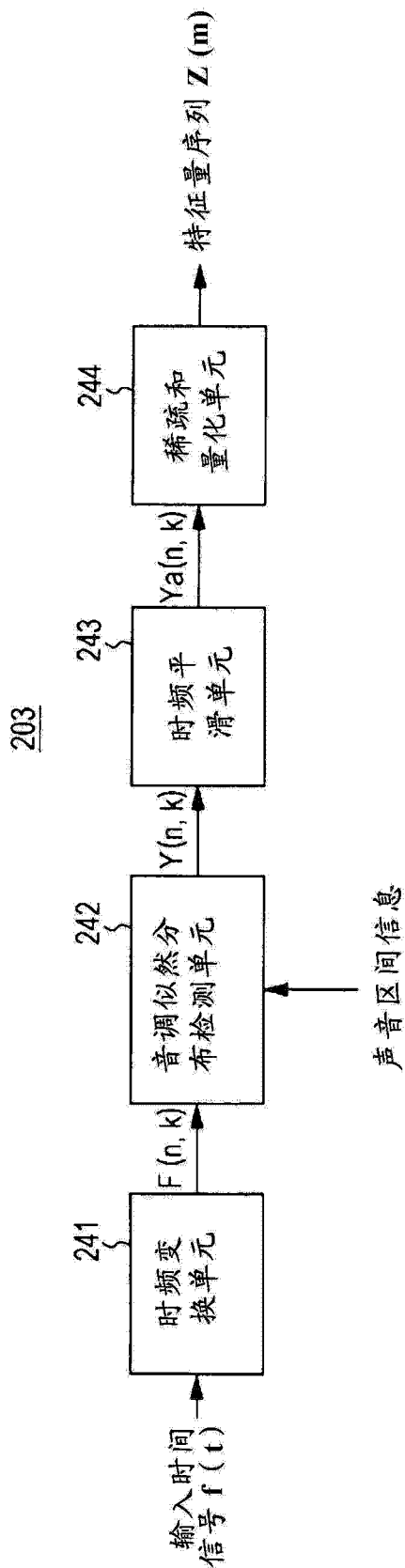


图 12

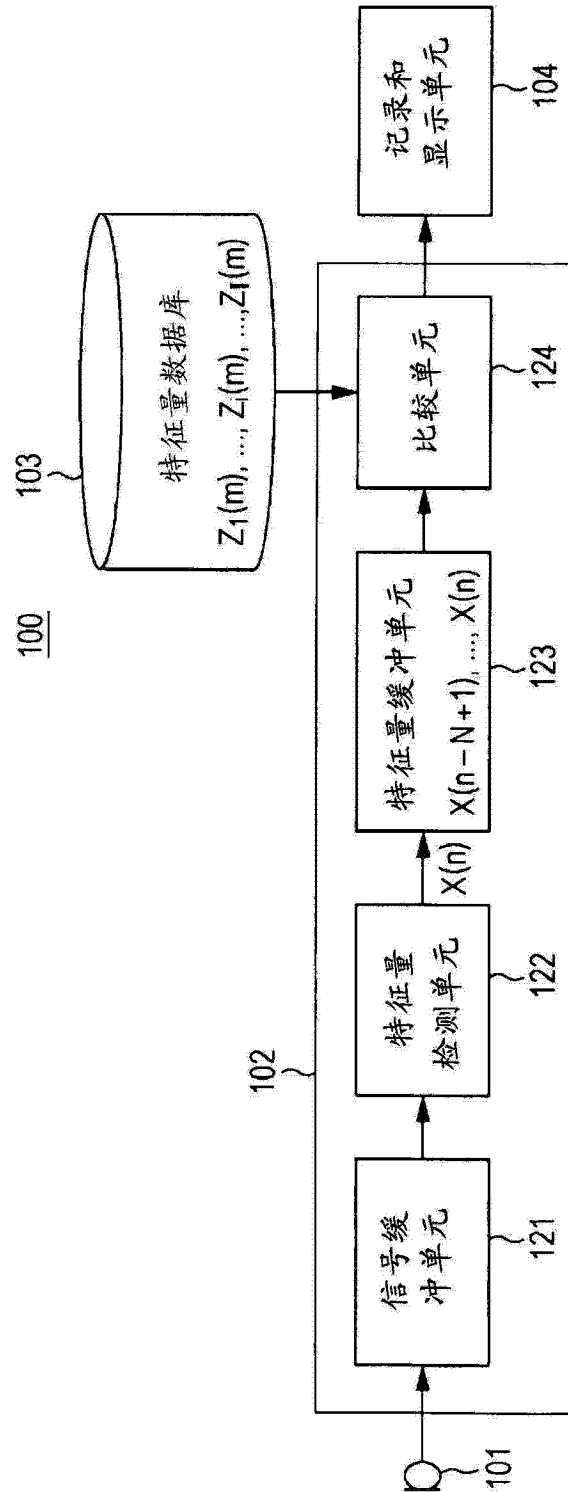


图 13

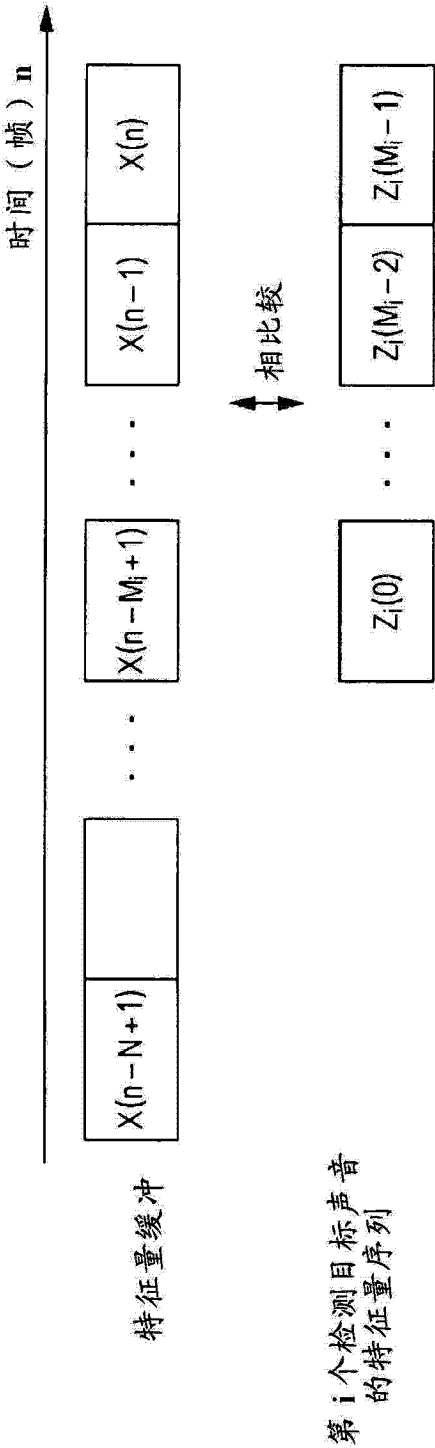


图 14

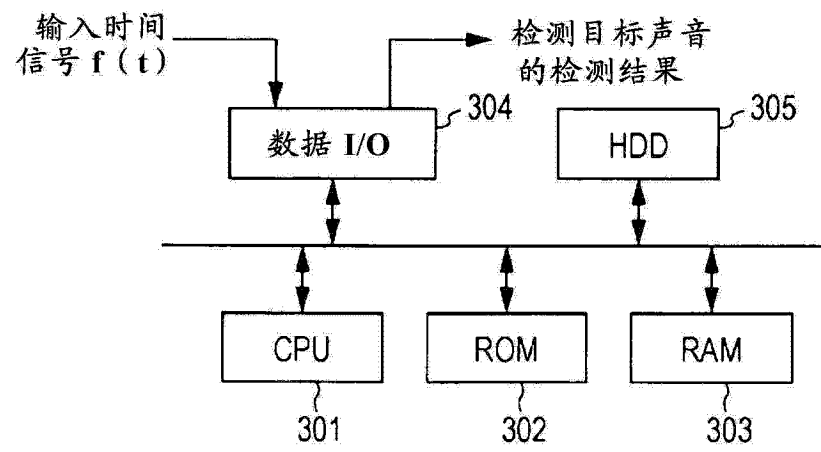


图 15

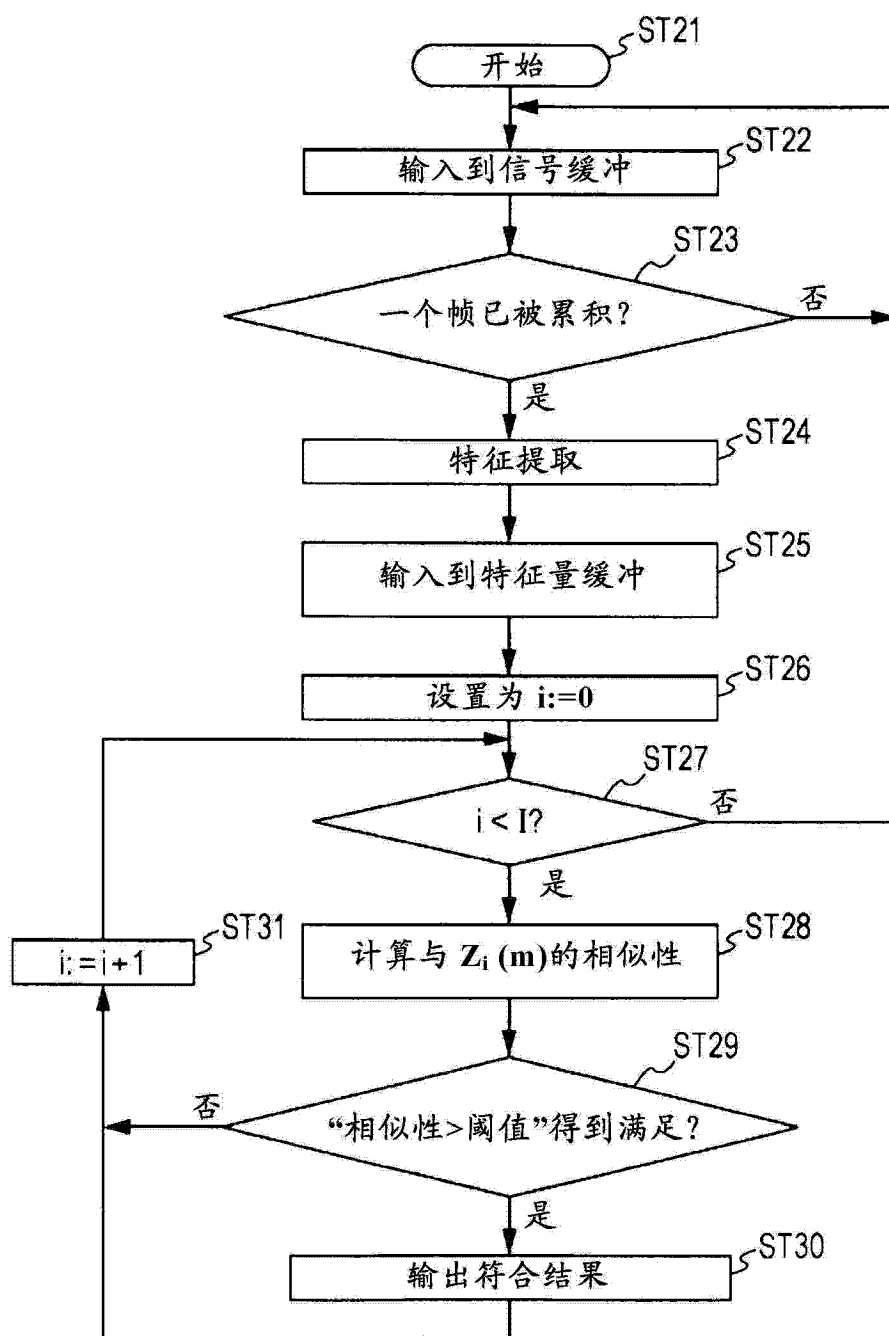


图 16

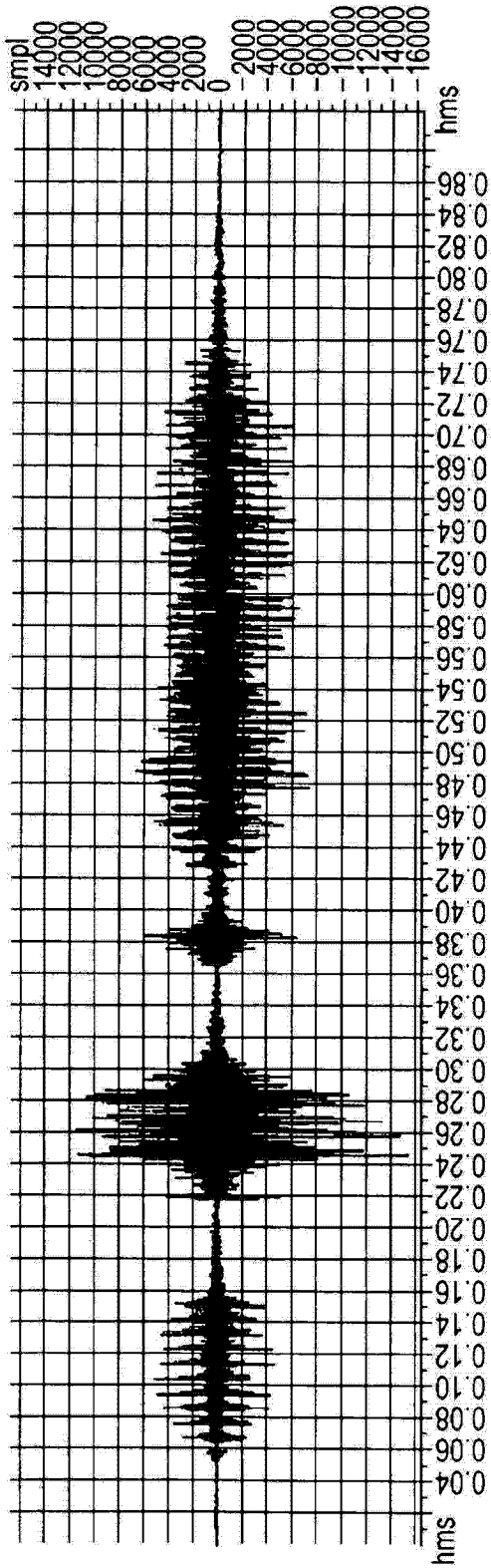


图 17A

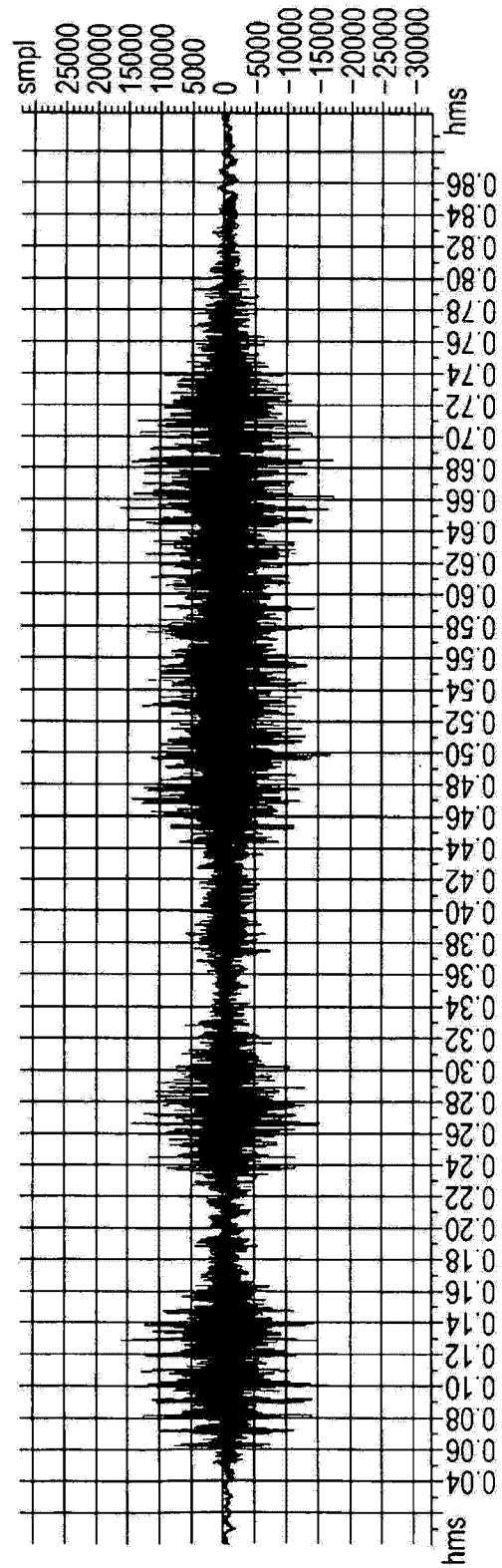


图 17B

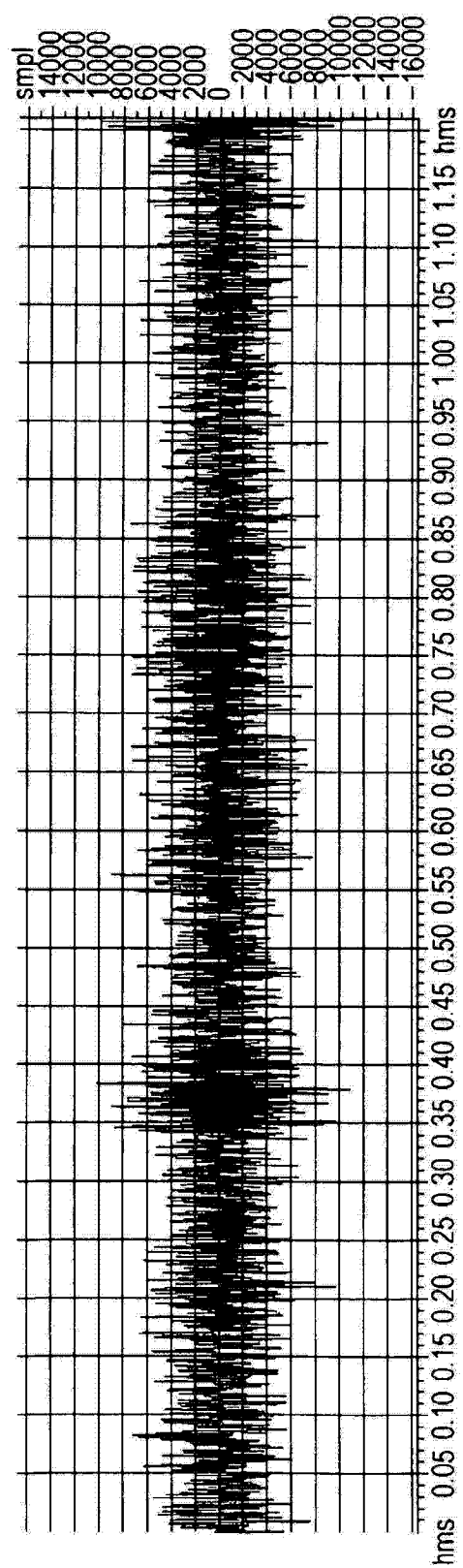


图 17C