



(12) 发明专利

(10) 授权公告号 CN 110168640 B

(45) 授权公告日 2021.08.03

(21) 申请号 201780082684.6

(22) 申请日 2017.01.23

(65) 同一申请的已公布的文献号  
申请公布号 CN 110168640 A

(43) 申请公布日 2019.08.23

(85) PCT国际申请进入国家阶段日  
2019.07.08

(86) PCT国际申请的申请数据  
PCT/EP2017/051311 2017.01.23

(87) PCT国际申请的公布数据  
W02018/133951 EN 2018.07.26

(73) 专利权人 华为技术有限公司  
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72) 发明人 肖玮 金文字

(74) 专利代理机构 广州三环专利商标代理有限公司 44202

代理人 熊永强 李稷芳

(51) Int.Cl.  
G10L 21/0232 (2006.01)  
G10L 25/84 (2006.01)

(56) 对比文件  
US 2003/0023430 A1, 2003.01.30  
CN 102314883 A, 2012.01.11  
CN 105489226 A, 2016.04.13  
CN 102804260 A, 2012.11.28  
CN 101278337 A, 2008.10.01  
CN 102792374 A, 2012.11.21  
US 2011/0125492 A1, 2011.05.26  
US 2016/0314805 A1, 2016.10.27  
EP 2228902 A1, 2010.09.15

审查员 李召卿

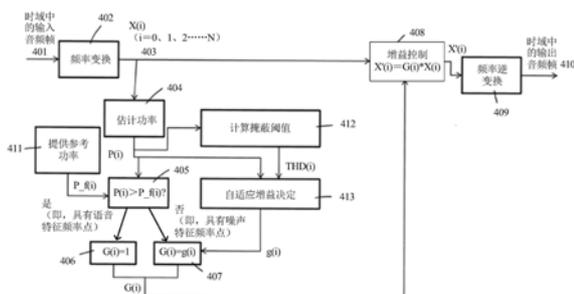
权利要求书3页 说明书9页 附图9页

(54) 发明名称

用于增强信号中需要分量的装置和方法

(57) 摘要

一种信号增强器,包括输入,用于接收具有需要分量和不需要分量的音频信号。所述信号增强器还包括感知分析器,用于将所述音频信号分离成多个频谱分量。所述感知分析器还用于:对于每个频谱分量,根据与所述频谱分量相关的功率估计,将所述频谱分量指定为属于所述需要分量或所述不需要分量。如果频谱分量被指定为属于所述不需要分量,则所述感知分析器将通过对所述频谱分量应用自适应增益来调整其功率;其中,所述自适应增益根据所述频谱分量对用户的期望可感知程度进行选择。这提高了所述需要分量的清晰度。



1. 一种信号增强器,其特征在于,包括:  
输入,用于获取具有需要分量和不需要分量的音频信号;  
感知分析器,用于:  
将所述音频信号分离成多个频谱分量;  
对于每个频谱分量,根据与所述频谱分量相关的功率估计,将所述频谱分量指定为属于所述需要分量或所述不需要分量;  
当所述频谱分量被指定为属于所述不需要分量时,通过对所述频谱分量应用自适应增益来调整其功率;其中,所述自适应增益根据所述频谱分量对用户的期望可感知程度进行选择;  
其中,所述感知分析器用于针对被指定为属于所述不需要分量的每个频谱分量,将其功率估计与功率阈值进行比较;  
当所述功率估计低于所述功率阈值时,将使与所述频谱分量相关的功率保持不变的增益选择作为所述自适应增益;  
当所述功率估计高于所述功率阈值时,将降低与所述频谱分量相关的功率的增益选择作为所述自适应增益。
2. 根据权利要求1所述的信号增强器,其特征在于,根据使用户期望的频谱分量变得可感知的功率来选择功率阈值。
3. 根据权利要求1所述的信号增强器,其特征在于,在给定与一个或多个其它频谱分量相关的功率的情况下,根据频谱分量对用户的期望可感知程度来选择所述功率阈值。
4. 根据权利要求2所述的信号增强器,其特征在于,在给定与一个或多个其它频谱分量相关的功率的情况下,根据频谱分量对用户的期望可感知程度来选择所述功率阈值。
5. 根据权利要求1至4中任一项所述的信号增强器,其特征在于,所述感知分析器用于:根据与每个频谱分量相关的组来为所述频谱分量选择所述功率阈值;其中,相同的功率阈值应用于特定组中包含的所有频谱分量的所述功率估计。
6. 根据权利要求5所述的信号增强器,其特征在于,所述感知分析器用于将每组频谱分量的所述功率阈值选择为预定阈值,所述预定阈值根据由所述组中的频谱分量表示的一个或多个频率分配给所述特定组。
7. 根据权利要求5所述的信号增强器,其特征在于,所述感知分析器用于确定一组频谱分量的所述功率阈值,所述确定是根据所述特定组中的频谱分量的所述功率估计进行的。
8. 根据权利要求7所述的信号增强器,其特征在于,所述感知分析器用于通过以下方法确定一组特定频谱分量的所述功率阈值:  
识别针对所述特定组中的频谱分量估计的最高功率;  
通过将所述最高功率减去预定量生成所述功率阈值。
9. 根据权利要求5所述的信号增强器,其特征在于,所述感知分析器用于通过比较以下项选择一组频谱分量的所述功率阈值:  
第一阈值,其根据由所述特定组中的频谱分量表示的一个或多个频率分配给所述组;  
第二阈值,其根据所述组中的频谱分量的所述功率估计确定;  
所述感知分析器用于选择所述第一和第二阈值中的较低者作为所述组的所述功率阈值。

10. 根据权利要求1至4中任一项所述的信号增强器,其特征在于,所述感知分析器用于:针对每个具有以下指定的频谱分量:(i)属于所述不需要分量;(ii)功率估计高于所述功率阈值,将所述自适应增益选择为所述功率阈值与所述频谱分量的所述功率估计之间的比率。

11. 根据前述权利要求中任一项所述的信号增强器,其特征在于,所述信号增强器包括变换单元,用于:

接收时域中的所述音频信号,并将所述音频信号转换到频域,其中,频域版本的所述音频信号通过相应的系数表示所述音频信号的每个频谱分量;

其中,所述感知分析器用于调整与频谱分量相关的所述功率,所述调整是通过将所述自适应增益应用于表示频域版本的所述音频信号中所述频谱分量的所述系数进行的。

12. 根据权利要求11所述的信号增强器,其特征在于,所述感知分析器用于产生目标音频信号,所述目标音频信号包括:

未调整的系数,其表示被指定为属于所述音频信号的所述需要分量的频谱分量;

调整的系数,其表示被指定为属于所述音频信号的所述不需要分量的频谱分量。

13. 根据权利要求12所述的信号增强器,其特征在于,所述变换单元用于接收频域中的所述目标音频信号,并将转换到时域;其中,所述信号增强器的输出用于输出时域版本的所述目标音频信号。

14. 一种信号处理方法,其特征在于,包括:

获取具有需要分量和不需要分量的音频信号;

将所述音频信号分离成多个频谱分量;

对于每个频谱分量,根据与所述频谱分量相关的功率估计,将所述频谱分量指定为属于所述需要分量或所述不需要分量;

当所述频谱分量被指定为属于所述不需要分量时,通过对所述频谱分量应用自适应增益来调整其功率;其中,所述自适应增益根据所述频谱分量对用户的期望可感知程度进行选择;

所述方法还包括:

其中,针对被指定为属于所述不需要分量的每个频谱分量,将其功率估计与功率阈值进行比较;

当所述功率估计低于所述功率阈值时,将使与所述频谱分量相关的功率保持不变的增益选择作为所述自适应增益;

当所述功率估计高于所述功率阈值时,将降低与所述频谱分量相关的功率的增益选择作为所述自适应增益。

15. 一种非瞬机器可读存储介质,其上存储有实现方法的处理器可执行指令,其特征在于,所述方法包括:

获取具有需要分量和不需要分量的音频信号;

将所述音频信号分离成多个频谱分量;

对于每个频谱分量,根据与所述频谱分量相关的功率估计,将所述频谱分量指定为属于所述需要分量或所述不需要分量;

当所述频谱分量被指定为属于所述不需要分量时,通过对所述频谱分量应用自适应增

益来调整其功率;其中,所述自适应增益根据所述频谱分量对用户的期望可感知程度进行选择;

其中,

针对被指定为属于所述不需要分量的每个频谱分量,将其功率估计与功率阈值进行比较:

当所述功率估计低于所述功率阈值时,将使与所述频谱分量相关的功率保持不变的增益选择作为所述自适应增益;

当所述功率估计高于所述功率阈值时,将降低与所述频谱分量相关的功率的增益选择作为所述自适应增益。

## 用于增强信号中需要分量的装置和方法

### 技术领域

[0001] 本发明涉及一种用于增强信号的装置和方法,其中所述信号具有需要分量和不需要分量。

### 背景技术

[0002] 增强噪声信号中的语音分量是有帮助的。例如,语音增强可以帮助改善通过电信网络等进行的语音通信的主观质量。另一个例子是自动语音识别(automatic speech recognition,简称ASR)。如果要扩展ASR的使用,则需要提高其对噪声条件的可靠性。一些商用ASR解决方案声称其提供良好性能,例如,词错误率(word error rate,简称WER)低于10%。但是,这种性能通常只有在几乎没有噪音的良好情况下才能实现。在复杂的噪声条件下,WER会高于40%。

[0003] 增强语音的一种方法是使用多个麦克风捕获音频信号,然后使用最佳滤波器对这些信号进行滤波。所述最佳滤波器通常是自适应滤波器,其受制于某些约束,例如最大限度增加信噪比(signal-to-noise ratio,简称SNR)。这种技术主要基于噪声控制,并且很少考虑听觉感知。其高噪声水平下不够稳定。太强的处理也会削弱语音分量,导致ASR性能低下。

[0004] 另一种方法主要基于对前景语音的控制,因为与噪声相比,语音成分往往具有独特的特征。这种方法使用所谓的“掩蔽效应”增加了语音和噪声之间的功率差异。根据心理声学,如果两个信号分量之间的功率差异足够大,则掩蔽者(具有较高功率)将掩蔽被掩蔽者(具有较低功率),使得被掩蔽者不可再听觉感知。得到的信号是具有更高清晰度的增强信号。

[0005] 利用掩蔽效应的一种技术是计算听觉场景分析(Computational Auditory Scene Analysis,简称CASA)。其工作原理是:检测信号中的语音分量和噪声分量并掩蔽噪声分量。CN105096961中描述了特定CASA方法的一个示例。图1示出了概览。在这种技术中,一组多个麦克风信号之一被选择作为主通道并进行处理以产生目标信号。然后,所述目标信号用于定义用于产生增强语音信号的最佳滤波器的约束。这种技术利用二进制掩码,所述二进制掩码通过将主信号频谱中低于参考功率的时间和频率点设置为0,将高于所述参考功率的频率点设置为1进行生成。这是一种简单的技术,尽管CN105096961提出了一些附加处理,但通过这种方法产生的目标信号通常具有许多频谱空洞。所述附加处理还在这种技术中引入了一些不希望的复杂性,包括需要两次时频变换及其逆变换。

### 发明内容

[0006] 本发明的目的在于提供用于增强信号中需要分量的改进概念。

[0007] 上述及其它目的通过独立权利要求的特征来实现。根据从属权利要求、说明书以及附图,进一步的实现形式是显而易见的。

[0008] 根据第一方面,提供一种信号增强器,所述信号增强器包括输入,用于接收具有需

要分量和不需要分量的音频信号。所述信号增强器还包括感知分析器,用于将所述音频信号分离成多个频谱分量。所述感知分析器还用于:对于每个频谱分量,根据与所述频谱分量相关的功率估计,将所述频谱分量指定为属于所述需要分量或所述不需要分量。如果频谱分量被指定为属于所述不需要分量,则所述感知分析器将通过与所述频谱分量应用自适应增益来调整其功率;其中,所述自适应增益根据所述频谱分量对用户的期望可感知程度进行选择。这提高了所述需要分量的清晰度。

[0009] 在所述第一方面的第一种实现方式中,所述感知分析器可以用于针对被指定为属于所述不需要分量的每个频谱分量,将其功率估计与功率阈值进行比较。所述感知分析器可以用于:当所述功率估计低于所述功率阈值时,将使与所述频谱分量相关的功率保持不变的增益选择作为所述自适应增益。所述感知分析器可以用于:当所述功率估计高于所述功率阈值时,将降低与所述频谱分量相关的功率的增益选择作为所述自适应增益。这增加了所述需要分量相对于所述不需要分量的相对功率,从而提高了所述需要分量的清晰度。

[0010] 在所述第一方面的第二种实现方式中,可以根据用户预期可感知所述频谱分量的功率来选择所述第一种实现方式中的所述功率阈值。这在实际意义上提高了所述需要分量的清晰度,因为人类用户对不同频率分量的感知不同。

[0011] 在所述第一方面的第三种实现方式中,在给定与一个或多个其它频谱分量相关的功率的情况下,可以根据频谱分量对用户的期望可感知程度来选择所述第一种或第二种实现方式中的所述功率阈值。这提高了所述增强信号中所述需要分量的可感知性。

[0012] 在所述第一方面的第四种实现方式中,根据所述第一种至第三种实现方式中任一项所述的感知分析器可以用于:根据与每个频谱分量相关的组来为所述频谱分量选择所述功率阈值;其中,相同的功率阈值应用于特定组中包含的所有光谱分量的所述功率估计。这符合心理声学原理。

[0013] 在所述第一方面的第五种实现方式中,根据所述第四种实现方式所述的感知分析器可以用于将每组频谱分量的所述功率阈值选择为预定阈值,所述预定阈值根据由所述组中的频谱分量表示的一个或多个频率分配给所述特定组。这符合心理声学原理。

[0014] 在所述第一方面的第六种实现方式中,根据所述第四种或第五种实现方式所述的感知分析器可以用于确定一组频谱分量的所述功率阈值,所述确定是根据所述特定组中的频谱分量的所述功率估计进行的。这考虑了信号中人类用户可类似感知的频谱分量的相对强度。

[0015] 在所述第一方面的第七种实现方式中,根据所述第六种实现方式所述的感知分析器可以用于确定一组特定频谱分量的所述功率阈值,所述确定是通过识别针对所述特定组中的频谱分量估计的最高功率,然后通过将所述最高功率减去预定量来生成所述功率阈值进行的。这考虑了特定频谱分量在被赋予其频谱组中的其它频谱分量的功率时的可能可感知度。

[0016] 在所述第一方面的第八种实现方式中,根据所述第四种至第七种实现方式中任一项所述的感知分析器可以用于通过比较第一阈值和第二阈值来选择一组频谱分量的所述功率阈值。所述第一阈值可以根据由特定组中的频谱分量表示的一个或多个频率分配给所述组。所述第二阈值可以根据所述组中的频谱分量的所述功率估计来确定。所述感知分析器可以用于选择所述第一和第二阈值中的较低者作为所述组的所述功率阈值。所述信号增

强器因此能够选择更合适的阈值。

[0017] 在所述第一方面的第九种实现方式中,根据所述第一种至第八种实现方式中任一项所述的感知分析器可以用于:针对每个具有以下指定的频谱分量:(i)属于所述不需要分量;(ii)功率估计高于所述功率阈值,将所述自适应增益选择为所述功率阈值与所述频谱分量的所述功率估计之间的比率。这将所述不需要分量的功率降至可接受的水平。

[0018] 在所述第一方面的第十种实现方式中,所述信号增强器,尤其是根据上述实现方式中任一项所述的信号增强器,包括变换单元。所述变换单元可以用于接收时域中的所述音频信号,并将所述音频信号转换到频域,其中,频域版本的所述音频信号通过相应的系数表示所述音频信号的每个频谱分量。所述感知分析器可以用于调整与频谱分量相关的所述功率,所述调整是通过将所述自适应增益应用于表示频域版本的所述音频信号中所述频谱分量的所述系数进行的。在频域中执行所述调整很方便,因为在频域中所述音频信号的不同部分之间的感知差异变得明显。

[0019] 在所述第一方面的第十一种实现方式中,根据所述第十种实现方式所述的感知分析器可以用于产生目标音频信号以包括:未调整的系数,其表示被指定为属于所述音频信号的所述需要分量的频谱分量;调整的系数,其表示被指定为属于所述音频信号的所述不需要分量的频谱分量。所述目标音频信号可以产生用于优化所述音频信号和其它音频信号的滤波的约束。所述目标音频信号可以在频域或时域中产生。

[0020] 在所述第一方面的第十二种实现方式中,根据所述第十一种实现方式所述的信号增强器的所述变换单元可以用于接收频域中的所述目标音频信号并将其转换到时域,其中所述变换单元用于输出所述目标音频的时域信号。这产生可以用作目标音频信号的时域信号。

[0021] 根据第二方面,提供一种方法,所述方法包括获取具有需要分量和不需要分量的音频信号。所述方法包括将所述音频信号分离成多个频谱分量。所述方法还包括:对于每个频谱分量,根据与所述频谱分量相关的功率估计,将所述频谱分量指定为属于所述需要分量或所述不需要分量。所述方法包括:当频谱分量被指定为属于所述不需要分量时,通过对所述频谱分量应用自适应增益来调整其功率;其中,所述自适应增益根据所述频谱分量对用户的期望可感知程度进行选择。

[0022] 根据第三方面,提供一种非瞬时机器可读存储介质,其上存储有处理器可执行指令,用于实现方法。所述方法包括获取具有需要分量和不需要分量的音频信号。所述方法还包括:对于每个频谱分量,根据与所述频谱分量相关的功率估计,将所述频谱分量指定为属于所述需要分量或所述不需要分量。所述方法包括:当频谱分量被指定为属于所述不需要分量时,通过对所述频谱分量应用自适应增益来调整其功率;其中,所述自适应增益根据所述频谱分量对用户的期望可感知程度进行选择。

## 附图说明

[0023] 现将参考附图通过示例的方式对本发明进行描述。在附图中:

[0024] 图1涉及一种用于增强语音信号的现有技术;

[0025] 图2示出了根据本发明实施例的信号增强器的示例;

[0026] 图3示出了根据本发明实施例的用于增强信号的过程的示例;

[0027] 图4示出了用于增强信号的过程的更详细示例；

[0028] 图5a至5c示出了根据本发明实施例的用于增强信号的过程，其中所述过程使用不同的最终功率阈值来设置所述自适应增益；

[0029] 图6示出了不同临界频带的绝对听觉阈值；

[0030] 图7示出了根据本发明的一个或多个实施例的技术如何影响需要与不需要信号分量的所述相对功率；

[0031] 图8a和图8b示出了用于从两个麦克风接收输入信号的语音处理系统的示例。

## 具体实施方式

[0032] 图2示出了信号增强器。所述信号增强器，通常在200处示出，包括输入201和感知分析器202。所述输入用于接收信号。所述信号包括需要分量和不需要分量。在许多实施例中，所述信号将是表示由麦克风捕获的声音的音频信号。所述需要分量通常是语音。所述不需要分量通常是噪声。如果麦克风处于包含语音和噪声的环境中，则其通常将捕获代表两者的音频信号。然而，所述需要分量和不需要分量不限于语音或噪声。它们可以是任何类型的信号。

[0033] 所述感知分析器202包括频率变换单元207，用于将所述输入信号分离成多个频谱分量。每个频谱分量代表特定频带或频率点中的所述输入信号的一部分。所述感知分析器还包括掩蔽单元203，用于分析每个频谱分量并将其指定为属于所述需要分量或属于所述不需要分量。所述掩蔽单元根据与所述频谱分量相关的功率估计做出此决定。所述感知分析器还包括自适应增益控制器204。当频谱分量被指定为属于所述不需要分量时，所述自适应增益控制器对所述频谱分量应用自适应增益。所述自适应增益根据所述频谱分量对用户的期望可感知程度进行选择。

[0034] 图2所示的信号增强器还包括合成器210。所述合成器是可选组件，用于将所有频谱分量组合在一起以产生增强信号，即，所述合成器将所述不需要分量和所述调整的不需要分量组合成单个输出信号。所述合成器可以用于在频域或时域中执行所述操作。例如，所述频谱分量可以先转换到时域，然后进行组合；或者它们可以先在频域中进行组合，然后转换到时域。

[0035] 图3示出了用于增强信号的方法的示例。所述方法的第一步骤是S301：获取具有需要分量和不需要分量的音频信号。然后，所述信号在步骤S302中分离成多个频谱分量。然后，每个频谱分量被指定为属于所述需要分量或属于所述不需要分量（步骤S303）。所述指定可以根据与每个频谱分量相关的功率估计来进行。在步骤S304中，如果任何分量被指定为属于所述不需要分量，则通过对所述频谱分量应用自适应增益来调整其功率。所述自适应增益根据所述频谱分量对用户的期望可感知程度进行选择。

[0036] 图2所示的结构（以及其中包括的所有框装置图）意在对应于许多功能块。这仅用于说明目的。图2不意在界定芯片上硬件的不同部分之间或软件中不同程序、过程或功能之间的严格划分。在一些实施例中，本文描述的部分或全部信号处理技术可能全部或部分地采用硬件执行。这尤其适用于结合重复操作的技术，例如傅里叶变换和阈值比较。在一些实现方式中，至少部分所述功能块可能全部或部分地由在软件控制下操作的处理器实现。任何所述软件适当地存储在非瞬态机器可读存储介质上。例如，所述处理器可以是手机、智能

手机、平板电脑或任何通用用户设备或通用计算设备的数字信号处理器 (Digital Signal Processor, 简称DSP)。

[0037] 本文所述的装置和方法可以用于在使用来自任何数量的麦克风的信号的系统中实现语音增强。在一示例中, 本文所述的技术可以并入多通道麦克风阵列语音增强系统中, 所述系统使用空间滤波来滤波多个输入并产生单通道增强输出信号。由这些技术产生的所述增强信号可以通过用作目标信号来为空间滤波提供新约束。意在用作目标信号的信号优选地通过考虑心理声学原理来产生。这可以通过使用自适应增益控制来实现, 所述自适应增益控制考虑频域中不同频率分量的估计感知阈值。

[0038] 图4示出了更详细的实施例。该图提供了下方参考图2中所示的一些功能块描述的语音增强技术的概览。音频信号的帧401被输入到时域中的系统。为简单起见, 图4和下方描述描述了单个帧的处理。此输入帧由频率变换单元207、402进行处理, 以输出一系列频率系数403。每个频率系数 $X(i)$ 表示频率点 $i$ 中频谱分量的幅度。然后, 由频谱功率估计器208使用所述系数403估计与每个所述频谱分量相关的功率。掩蔽单元203使用所述估计功率来将所述相应的频谱分量指定为属于噪声和/或属于语音。因此, 一些所述频谱分量将被指定为属于噪声, 而其它所述频谱分量将被指定为属于语音。在图4所示的示例中, 这在405中由掩蔽发生器206执行, 所述掩蔽发生器将每个频谱分量的所述估计功率与参考功率阈值411 (由参考功率估计器205生成) 进行比较。所述参考功率阈值可以设置为依赖于所述信号中所述噪声分量的预期功率的功率水平。功率高于此功率水平的频谱分量被假定为包括语音。具有更低功率的频谱分量被假定为仅包含噪声。

[0039] 然后, 被指定为属于语音的所述频谱分量与掩蔽功率阈值412 (由掩蔽功率发生器209输出) 进行比较。此附加功率阈值 $THD(i)$ 涉及用户预期可感知不同频谱分量的功率。(有各种参数可用于设置所述阈值, 下方将描述一些示例)。所述掩蔽功率阈值控制由所述自适应增益控制器204做出的自适应增益决定413。当频率点 $i$ 中的所述频谱分量被指定为噪声时由所述控制器对所述频谱分量应用的增益 $g(i)$ 会根据所述频谱分量是否满足所述掩蔽功率阈值发生变化 (在407中)。在一示例中, 如果频谱分量的功率估计低于所述掩蔽功率阈值, 则其功率保持不变; 如果其功率估计高于所述掩蔽功率阈值, 则其功率会减小。已被指定为包括语音的频谱分量将保持不变, 因此在406中为它们选择增益为1。这只是一个示例, 因为可以应用任何合适的增益。例如, 在一个示例中, 被指定为包括语音的所述频谱分量可以放大。

[0040] 为每个频谱分量选择的所述增益被应用于每个相应系数408。新系数 $X'(i)$ 形成频率逆变换409的基础, 用于在时域410中构造输出帧。

[0041] 图5a示出了与图4中概述的过程类似的过程。图5a包括如何执行所述过程中的特定步骤的具体示例。当总体考虑时, 这些具体示例还可以用于与图5a中所示的过程可能不同的过程中的等效步骤。例如, 图5a示出了从两个选项 $THD1$ 和 $THD2$ 中选择所述掩蔽功率阈值的整个过程。但是, 这并不意味着下方描述的所有示例步骤都限于从 $THD1$ 和 $THD2$ 中选择所述掩蔽功率阈值的实施例。例如, 图5a包括如何估计每个频谱分量的功率的示例。所述示例是普遍适用的, 并且可以容易地引入整个过程, 所述整个过程采用与图5a所示的方式不同的方式选择所述掩蔽功率阈值。

[0042] 在图5a中, 所述输入信号同样以帧为单位进行处理。这实现了所述信号的实时处

理。每个输入信号可以划分为多个具有固定帧长度(例如,16ms)的帧。对所有帧应用相同的处理。所述单通道输入501可以称为“Mic-1”。所述输入可以是一组麦克风信号之一,所述一组麦克风信号全部包括需要分量,例如语音,以及不需要分量,例如噪声。所述一组信号不需要是音频信号,并且可以通过麦克风捕获之外的方法产生。

[0043] 在框502中,对所述输入接收的所述音频信号执行时频变换以获得其频谱。此步骤可以通过执行短时离散傅里叶变换(Short-Time Discrete Fourier Transform,简称SDFT)算法实现。SDFT是傅里叶相关的变换,用于确定信号随时间变化时其局部部分的正弦频率和相含量。可以通过将所述音频信号分成相等长度的短分段(例如帧501),然后分别对每个短分段计算傅里叶变换来计算SDFT。其结果是所述音频信号的每个短分段的傅里叶频谱,其捕获作为时间函数的所述音频信号的变化频谱。因此,每个频谱成分均具有幅度和时间广度。

[0044] 对所述输入信号501的每个帧执行SDFT 502。如果采样率是16kHz,则帧大小可以设置为16ms。这只是一个示例,可以使用其它采样率和帧大小。还应注意,采样率和帧大小之间没有固定的关系。因此,例如,所述采样率可以为48kHz,帧大小为16ms。可以在所述输入信号上实现512点SDFT。执行所述SDFT会在频域中生成一系列复值系数 $X(i)$ (其中,系数索引 $i=0,1,2,3$ 等,可用于在时域中指定所述信号的索引或在频域中指定系数的索引)。这些系数是傅里叶系数,也可以称为频谱系数或频率系数。

[0045] 对于每个系数 $X(i)$ ,计算相应的功率 $P(i) = |X(i)|^2$ 503。这可以通过以下等式定义:

$$[0046] \quad P(i) = \text{real}(X(i))^2 + \text{imag}(X(i))^2$$

[0047] 其中, $\text{real}(*)$ 和 $\text{imag}(*)$ 是所述相应SDFT系数的实部和虚部。

[0048] 还针对每个傅里叶系数 $X(i)$  516估计参考功率 $P_f(i)$ 。图2中的所述感知分析器202可以引入参考功率估计器205用生成所述估计。任何合适的噪声估计(noise estimation,简称NE)方法都可以用于成所述估计。一种简单的方法是计算当前帧和一个或多个先前帧的每个系数的功率密度平均值。根据语音处理理论,这种简单的方法最适合于所述音频信号可能包含平稳噪声的场景。另一种选择是使用先进的噪声估计方法,所述方法往往适合于包含非平稳噪声的场景。在一些实施例中,所述参考功率估计器可以用于根据预期噪声场景选择适当的功率估计算法,例如,噪声在本质上预期是平稳的或非平稳的。

[0049] 下一步骤是通过将每个系数的所述功率 $P(i)$ 与所述相应的参考功率 $P_f(i)$  505进行比较来实现二进制掩蔽。这产生二进制掩蔽矩阵 $M(i)$ :

$$[0050] \quad M(i) = \begin{cases} 1, & \text{if } P(i) > \gamma \cdot P_f(i) \\ 0, & \text{if } P(i) \leq \gamma \cdot P_f(i) \end{cases}$$

[0051] 其中, $\gamma$ 是预定义的功率频谱密度因子。

[0052] 所述二进制掩码中的值 $M(i) = 1$ 表示相应的所述系数 $P(i)$ 具有语音特征。在这种情况下,不需要更改所述系数 $P(i)$ 。值 $M(i) = 0$ 表示相应的所述系数 $P(i)$ 具有噪声特征。在这种情况下,应使用自适应增益控制506、507调整相应的所述系数。

[0053] 在图5a所示的示例中,自适应增益控制由三个操作组成。

[0054] 在第一操作中,为每个频率索引( $i=0,1,2,\dots$ )提供绝对听觉阈值(THD1)  $Th1(i)$  ( $i=0,1,2,\dots$ ) 511。所述阈值根据每个频谱分量对用户的期望可感知程度进行设置。每个频率

可以与相应的功率阈值THD1相关联,这由心理声学原理确定。高于所述阈值,则人类听觉系统可感知所述频率的频谱分量;否则,不可感知。因此,THD1可以预定义,并且可以作为查找表提供给掩蔽功率发生器209。

[0055] 在实践中,不一定要为每个单独频谱分量定义绝对听觉阈值。相反,所述SDFT系数可以分成多个组,同一个组内的所有系数可以与相同的绝对听觉阈值相关联。换句话说,对于属于相同相关组的任何系数索引*i*、*j*来说,  $Th1(i) = Th1(j)$ 。

[0056] 所述SDFT系数集优选地分成系数彼此相邻的系数组(即所述系数代表相邻的频率点,因此具有相邻的索引)。一种简单的方法是将所述系数均匀地分成*N*组(例如, $N=30$ ),其中每组具有相同数量的SDFT系数。或者,所述组可以以某些频率集中。对于低频率组和高频率组而言,分配至特定组的系数的数量可以不同。优选方法是使用所谓的临界频带尺度,其类似于对数尺度。这符合心理声学基本原理。通常,低频率组中的系数的数量应小于高频率组中的系数的数量。图6示出了不同临界频带的所述绝对听觉阈值,其假设有31个临界频带,采样率为16kHz。

[0057] 在第二操作中,为每个频率索引( $i=0,1,2,\dots$ )估计相对掩蔽阈值(THD2) 513。可以通过考虑不同频率的掩蔽效应来设置THD2。对于每个表示频率索引*i*的系数,优选地不单独确定THD2,而是针对每组频率索引确定THD2。可以按照上述任一种方法将所述系数分组在一起。在每个组中,将当前组中具有最大功率的系数设置为“掩蔽者”512。THD2可以设置为所述掩蔽者的功率减去某个预定量 $\alpha$ ,其中, $\alpha$ 可以根据心理声学原理设置。例如,所述预定量 $\alpha$ 的合适值可以是13dB。

[0058] 然后,将每个系数索引的最终掩蔽阈值选择为THD1和THD2的最小值514,即  $THD(i) = \min\{THD1(i), THD2(i)\}$ 。

[0059] 第三操作使用在505中确定的所述二进制掩码。对于相应二进制掩码元素为 $M(i)=1$ 的系数,所述增益可以设置为1,即不对所述频谱分量进行更改。对于相应二进制掩码元素为 $M(i)=0$ 的系数,通过比较在504中为所述频谱分量确定的功率与在514中确定的所述阈值THD来确定适当的增益。框515示出了所述比较,其可以表达为如下等式:

$$[0060] \quad g(i) = \begin{cases} 1, & \text{if } P(i) < THD(i) \\ \frac{THD(i)}{P(i)}, & \text{if } P(i) > THD(i) \end{cases}$$

[0061] 其中, $g(i)$ 是所述自适应增益。

[0062] 基本上,如果 $P(i) < THD(i)$ ,则说明所述频谱分量太弱而无法听到,并且不需要增益控制。如果 $P(i) > THD(i)$ ,则说明所述频谱分量足够强可以被听到,因此通过在频域中对其系数应用适当的增益来调整其功率。通过计算新系数 $X'(i)$ 在508中应用所述自适应增益控制:

$$[0063] \quad X'(i) = g(i) * X(i)$$

[0064] 所述新系数 $X'(i)$ ( $i=0,1,2,\dots$ )形成509中构造所述输出帧510的傅里叶逆变换的基础。在此示例中,所述调整的和未调整的系数在频域中组合,然后一起转换到时域。所述组合同样可以在转换到时域之后发生,使得所述调整的和未调整的系数先分别转换到时域,然后再组合在一起以形成单个输出帧。

[0065] 在其它实现方式中,可以不同地确定所述阈值 $THD(i)$ 。例如,图5b示出了一种实现

方式,在所述实现方式中所述最终阈值由所述绝对阈值设置。在此示例中,在514中 $\text{THD}(i) = \text{THD1}(i)$ ,并且不需要计算所述相对掩蔽阈值 $\text{THD2}$ 。图5c示出了另一种实现方式。在此示例中,所述最终阈值由所述相对阈值设置。因此,在514中 $\text{THD}(i) = \text{THD2}(i)$ ,并且不需要计算所述绝对掩蔽阈值 $\text{THD1}$ 。在此实现方式中,希望增加掩蔽者和被掩蔽者之间的功率差以获得更好的掩蔽效果,因此优选地增加所述掩蔽者的功率所减去的所述预定量。例如, $\alpha$ 可以从13dB增加到16dB。

[0066] 上述系数增益控制改变所述频域频谱并掩蔽所述噪声分量,使得它们变得不易感知。图7示出了所述效果。原始信号中的掩蔽者701(增益控制前)属于需要分量。被掩蔽者702属于不需要分量。增益控制后,掩蔽者703的功率不变,而被掩蔽者704的功率减小了(如虚线所示)。因此调整了所述两个分量的相对功率,其效果是所述需要分量更容易与所述不需要分量区分开。这在所述输入信号是带噪语音信号的实现方式中尤其有用,因为其结果是具有改进语音清晰度的增强信号。所述增强信号可以用作任何多麦克风语音增强技术中的目标信号。

[0067] 图8a和图8b示出了基于多麦克风阵列的系统的两个示例。在这两个示例中,所述多麦克风阵列具有两个麦克风。这仅用于示例目的。应当理解的是,本文描述的技术可以有益地在具有任何数量的麦克风的系统中实现,包括基于单通道增强的系统或具有含三个或更多麦克风的阵列的系统。

[0068] 在图8a和图8b中,两个系统均用于从两个麦克风信号接收实时语音帧: $x_1(i)$ 和 $x_2(i)$ (801、802、806、807)。两个系统均具有感知分析器(perception analyser,简称PA)块803、808,所述感知分析器块803、808用于实现根据上述一个或多个实施例的增强技术。每个PA块输出目标信号 $t(i)$  811、812。每个系统还包括最佳滤波块805、810。每个最佳滤波块用于服从所述目标信号 $t(i)$ 的约束实现用于滤波 $x_1(i)$ 和 $x_2(i)$ 的算法。可以使用任何合适的算法。例如,广义自适应滤波算法可以基于以下项定义的最佳滤波问题:

$$[0069] \quad \min (\|\tilde{A} \cdot X - T\|^2)$$

[0070] 其中, $X$ 是麦克风信号 $x_1(i)$ 和 $x_2(i)$ 的矩阵表达式, $T$ 是目标信号 $t(i)$ 的矩阵表达式。所述最佳滤波器参数由矩阵表达式 $\tilde{A}$ 定义。然后,所述一组最佳滤波器参数 $\tilde{A}$ 可以用于滤波所述麦克风信号以产生单个增强信号。

[0071] ASR场景的主要目的在于增加输入到所述ASR块的所述音频信号的清晰度。对原始麦克风信号进行最佳滤波。优选地,不执行额外的降噪以避免移除关键语音信息。对于语音通信场景,应保持主观质量和清晰度之间的良好平衡。本申请应考虑降噪。因此,所述麦克风信号可以在进行最佳滤波之前进行降噪。图8a和图8b示出了这些替代方案。

[0072] 在图8a中,所述麦克风阵列(microphone array,简称MA)处理块804的主要功能是噪声估计(noise estimation,简称NE)。其用于生成每个系数的所述参考功率密度。所述目标信号 $t(i)$ 在所述PA块803中生成。所述最佳滤波块805通过对由目标信号 $t(i)$ 约束的所述麦克风信号进行滤波来完成所述语音增强处理。

[0073] 在图8b中,MA块809的主要功能同样是噪声估计(noise estimation,简称NE)。其也用于实现适当的降噪(noise reduction,简称NR)方法。这产生初步的增强信号 $y_1(i)$ 和 $y_2(i)$ (813、814)。所述目标信号 $t(i)$ 在PA块808中生成。所述最佳滤波块810通过对由目标信号 $t(i)$ 约束的所述麦克风信号进行滤波来完成所述语音增强处理。

[0074] 在图8a和图8b所示的双麦克风布置中,任一麦克风信号均可以用作所述PA块803、808的直接输入。然而,也可以使用任何其它信号。例如, $y_1(i)$ 和 $y_2(i)$ 之一可以用作所述PA块803、808的所述输入。另一种选择是故意选择某个信号以实现特定效果。例如,可以选择特定信号以便实现波束成形,其目的在于通过考虑麦克风信号的不同通道之间的空间相关性来实现进一步增强。

[0075] 应当理解的是,此解释和所附权利要求指的是,所述设备通过执行某些步骤或程序或通过实施特定技术来执行某些操作,而这不会妨碍所述设备执行其它步骤或程序或者实施其它技术(作为同一流程的一部分)。换句话说,在所述设备被描述为“通过”某些指定方法执行某些操作时,“通过”一词的意思是所述设备执行“包括”指定方法而不是“只包含”这些方法的流程。

[0076] 申请方在此单独公开本文描述的每一个体特征及两个或两个以上此类特征的任意组合。以本领域技术人员的普通知识,能够基于本说明书将此类特征或组合作为整体实现,而不考虑此类特征或特征的组合是否能解决本文所公开的任何问题;且不对权利要求书的范围造成。本申请表明本发明的各方面可由任何这类单独特征或特征的组合构成。鉴于前文描述可在本发明的范围内进行各种修改对本领域技术人员来说是显而易见的。

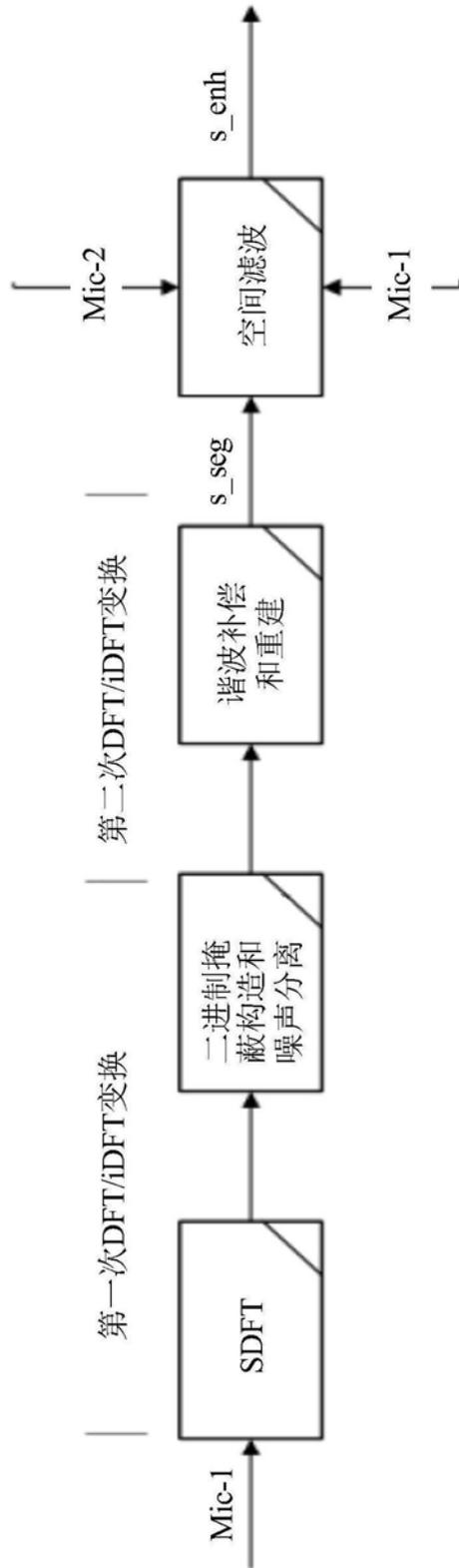


图1

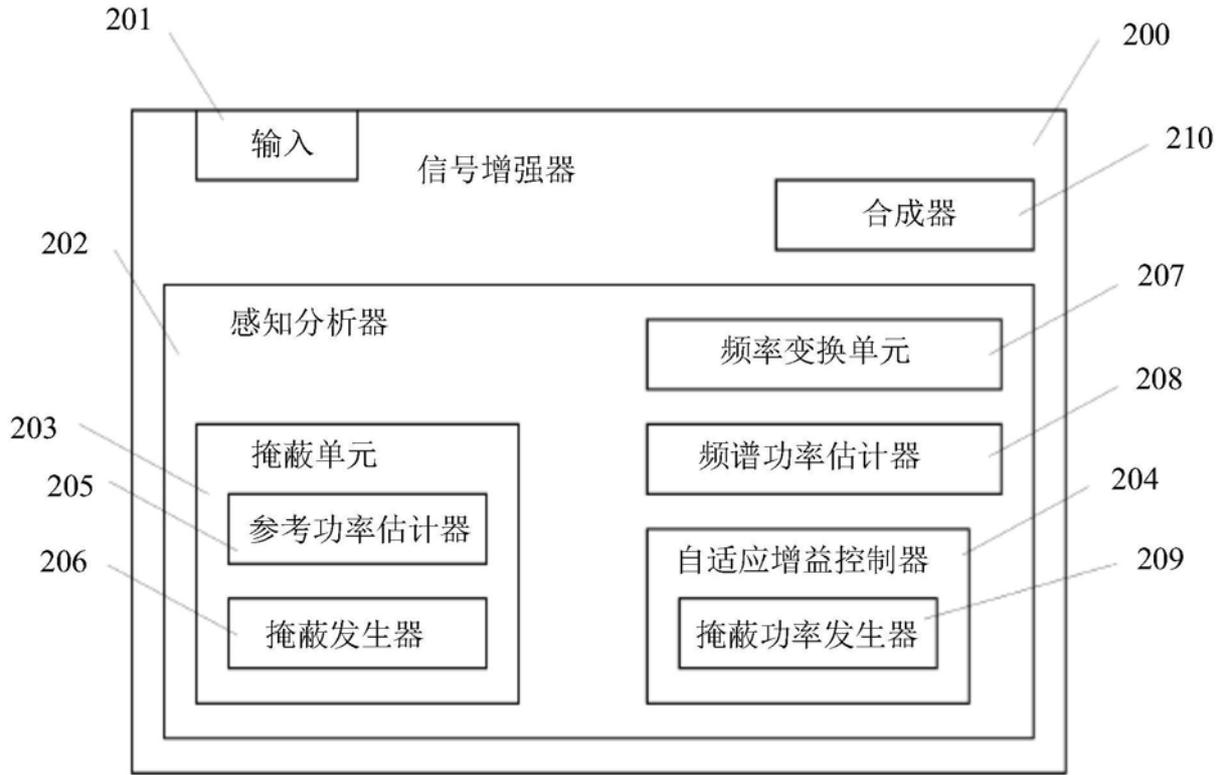


图2

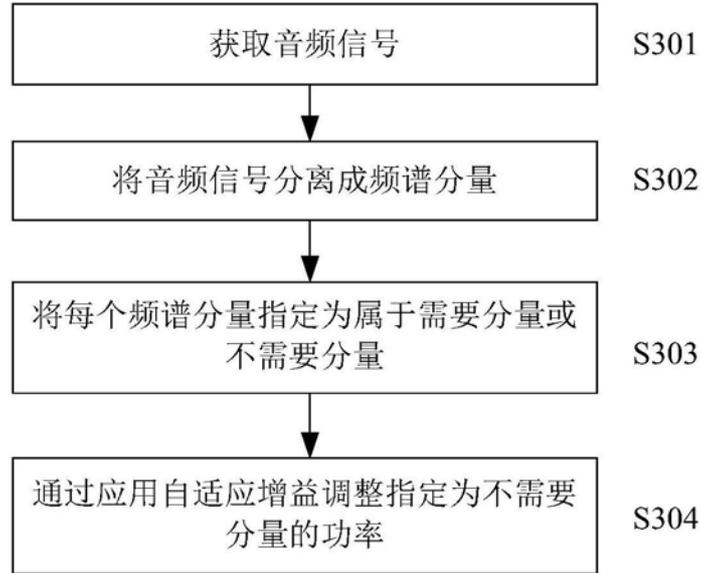


图3

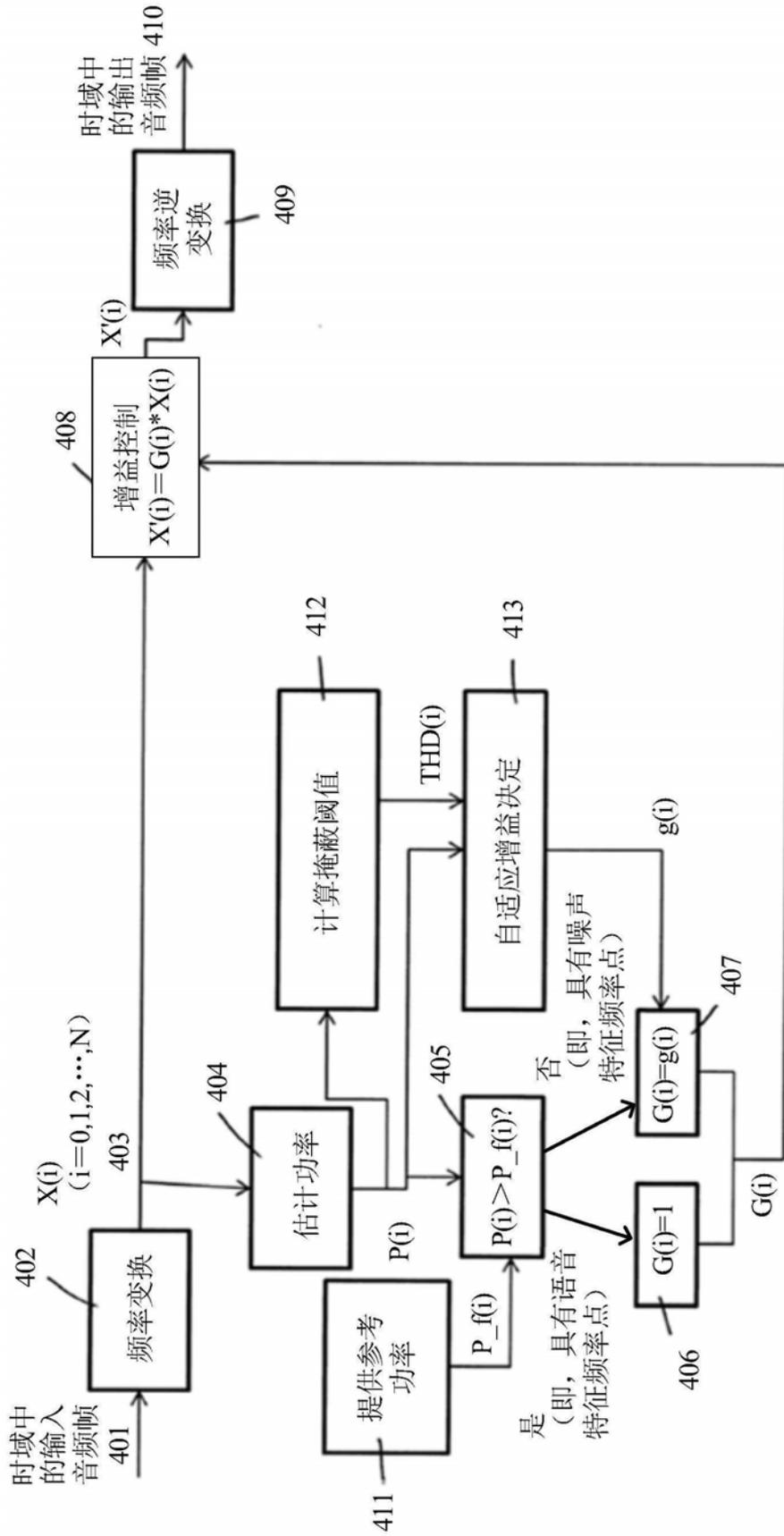


图4

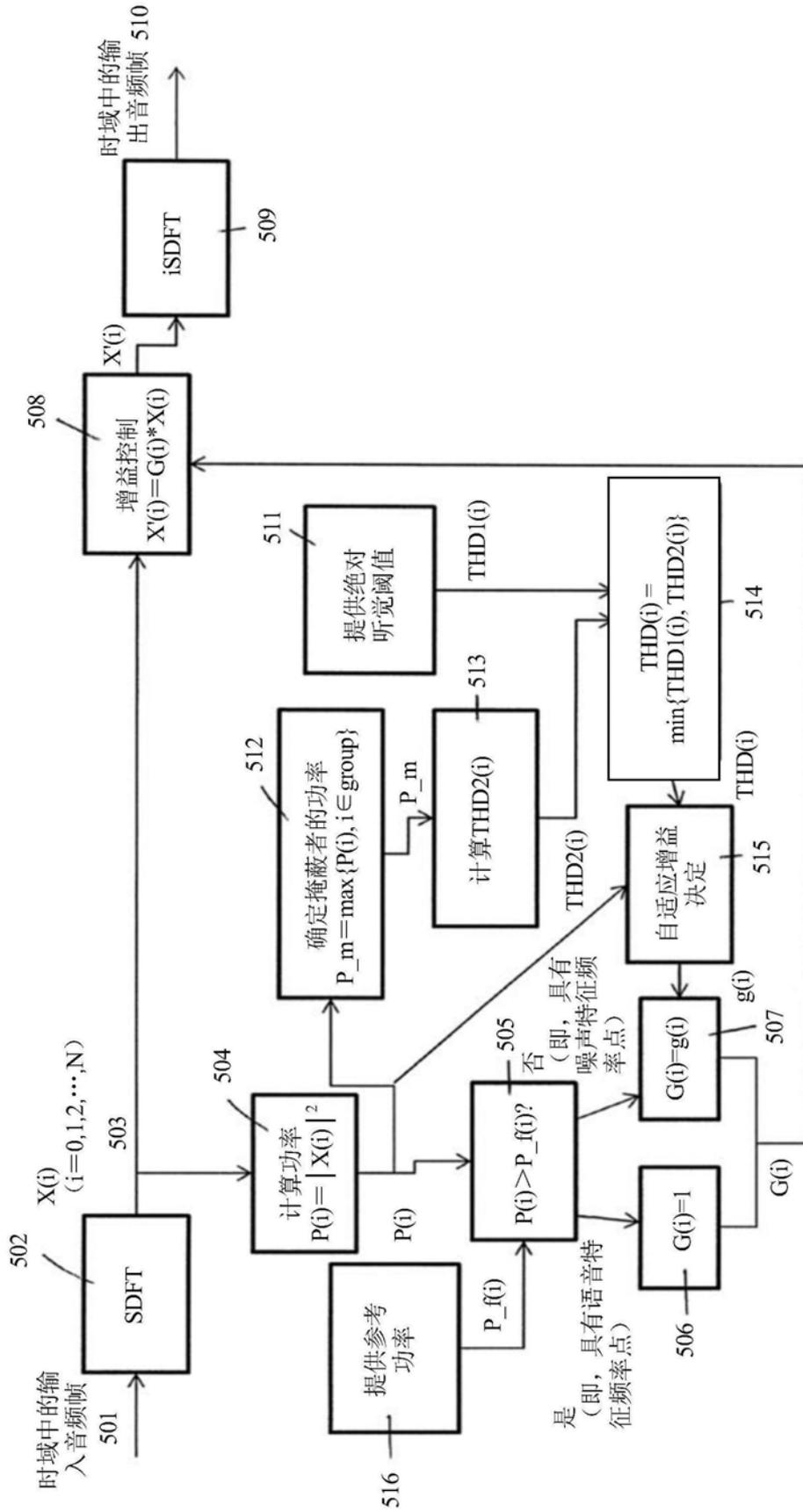


图5a

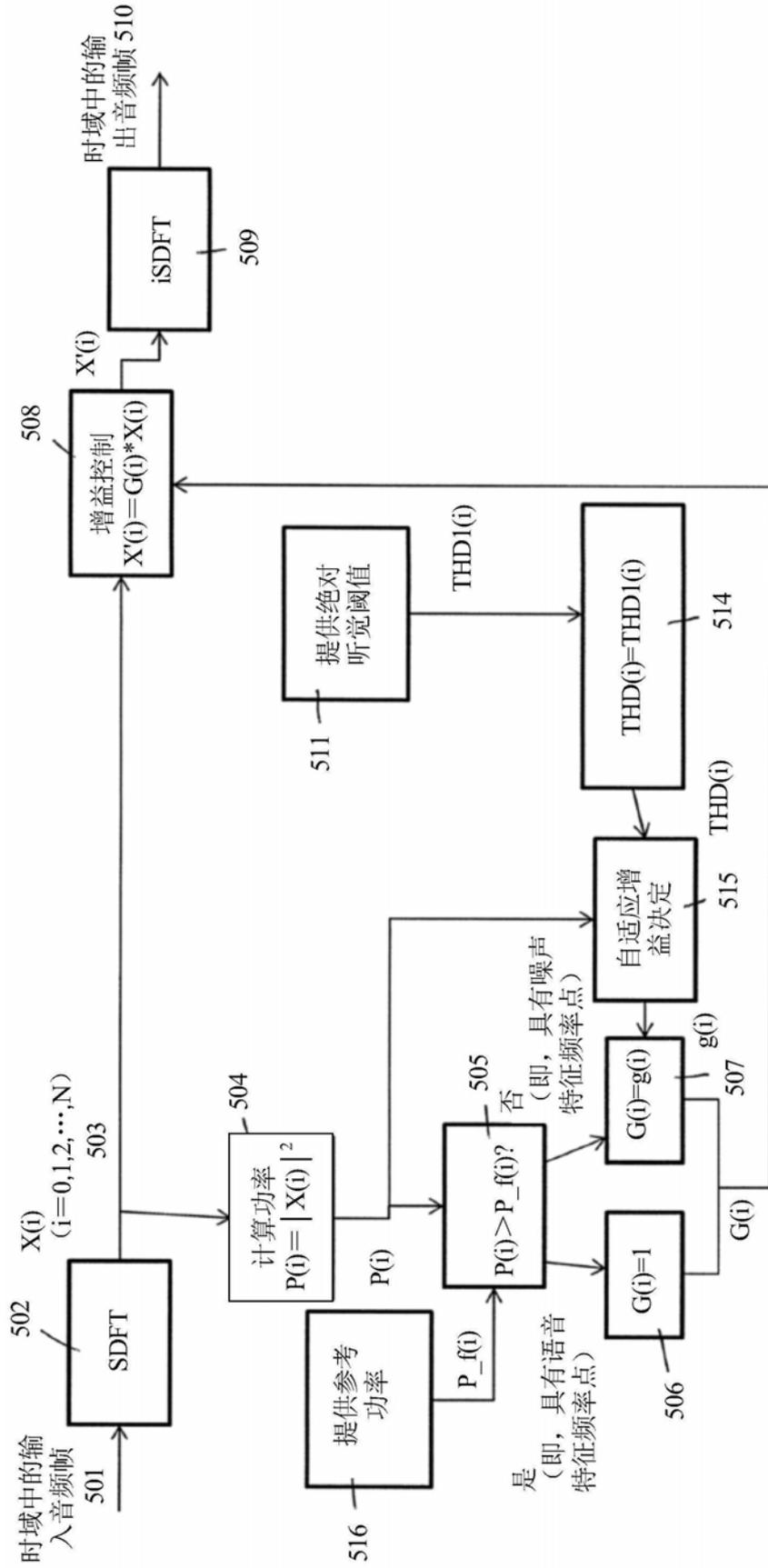


图5b

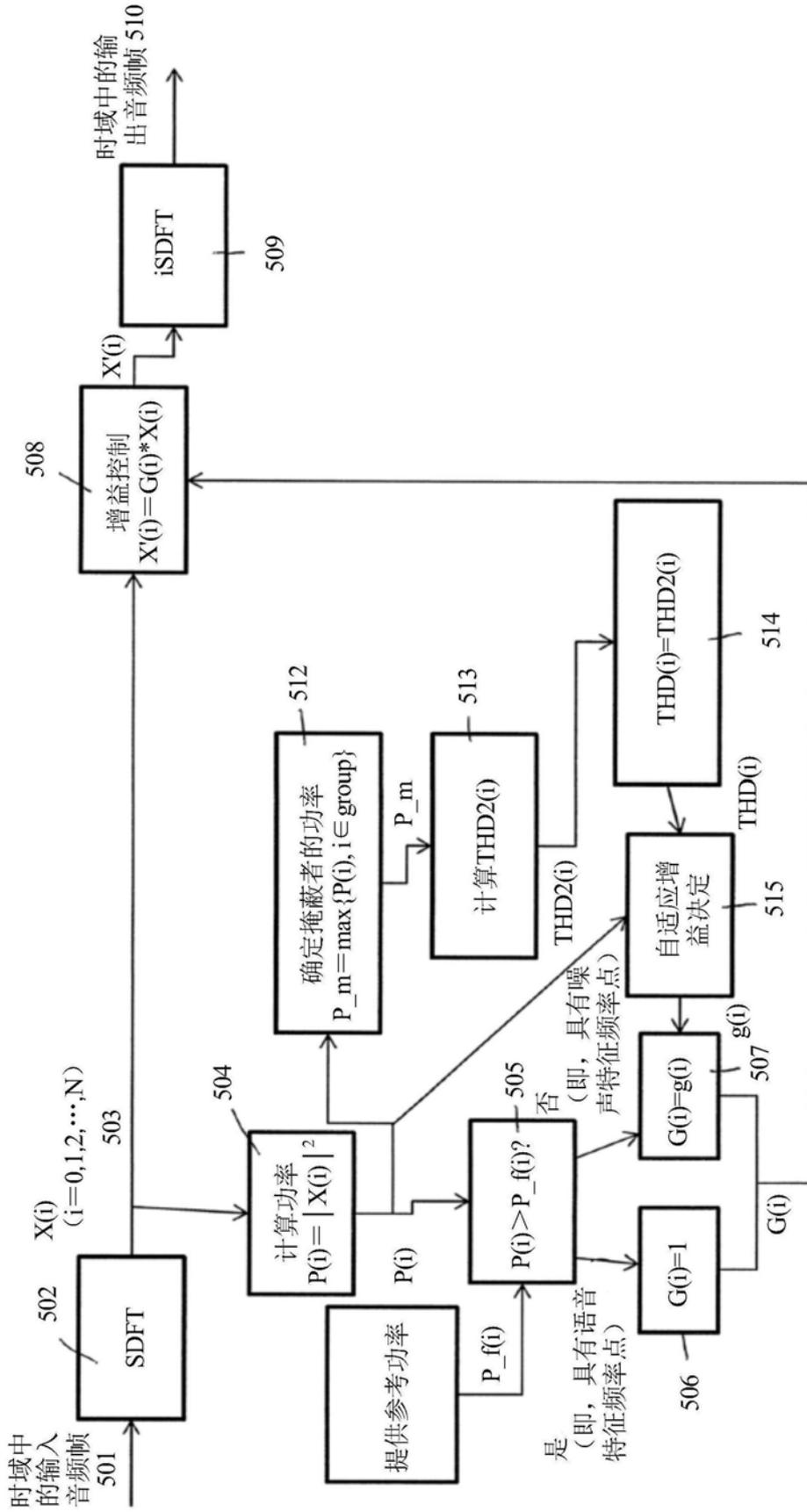


图5c

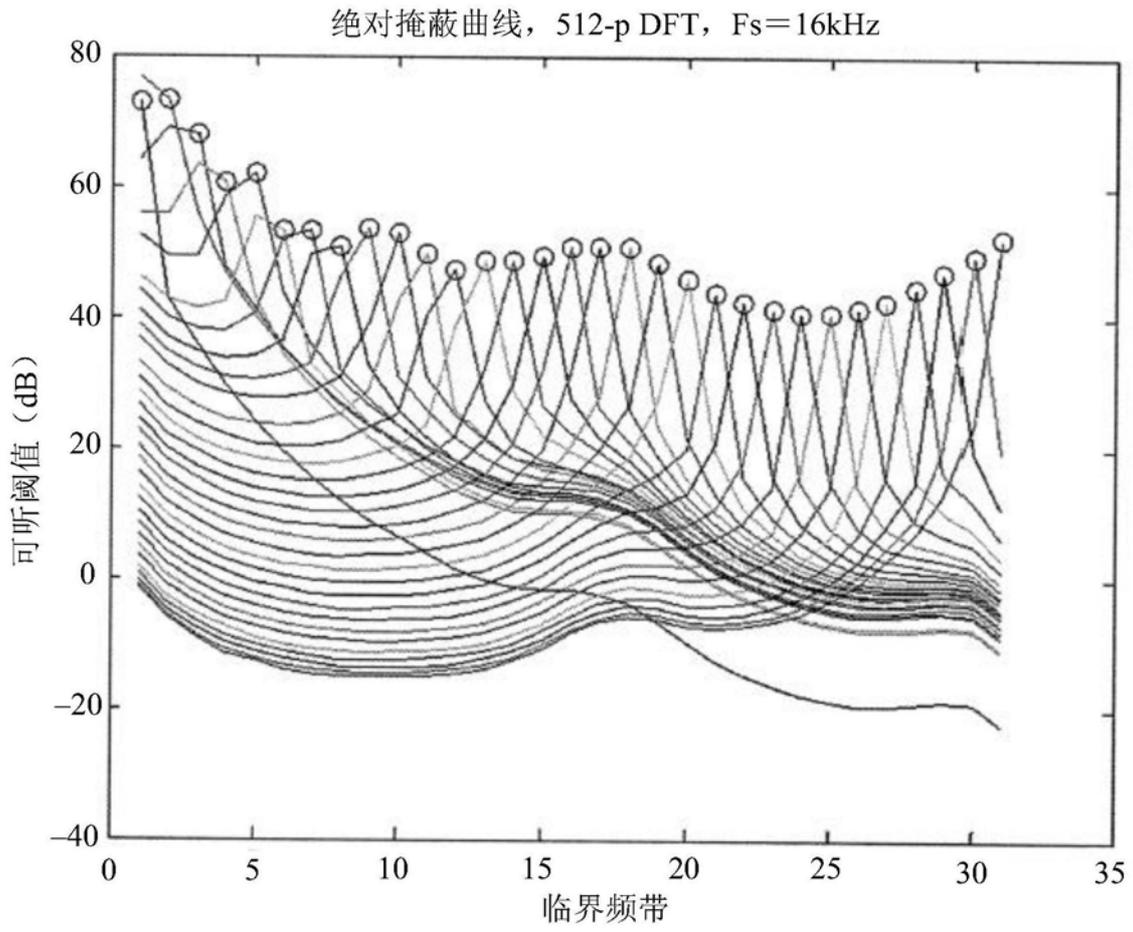


图6

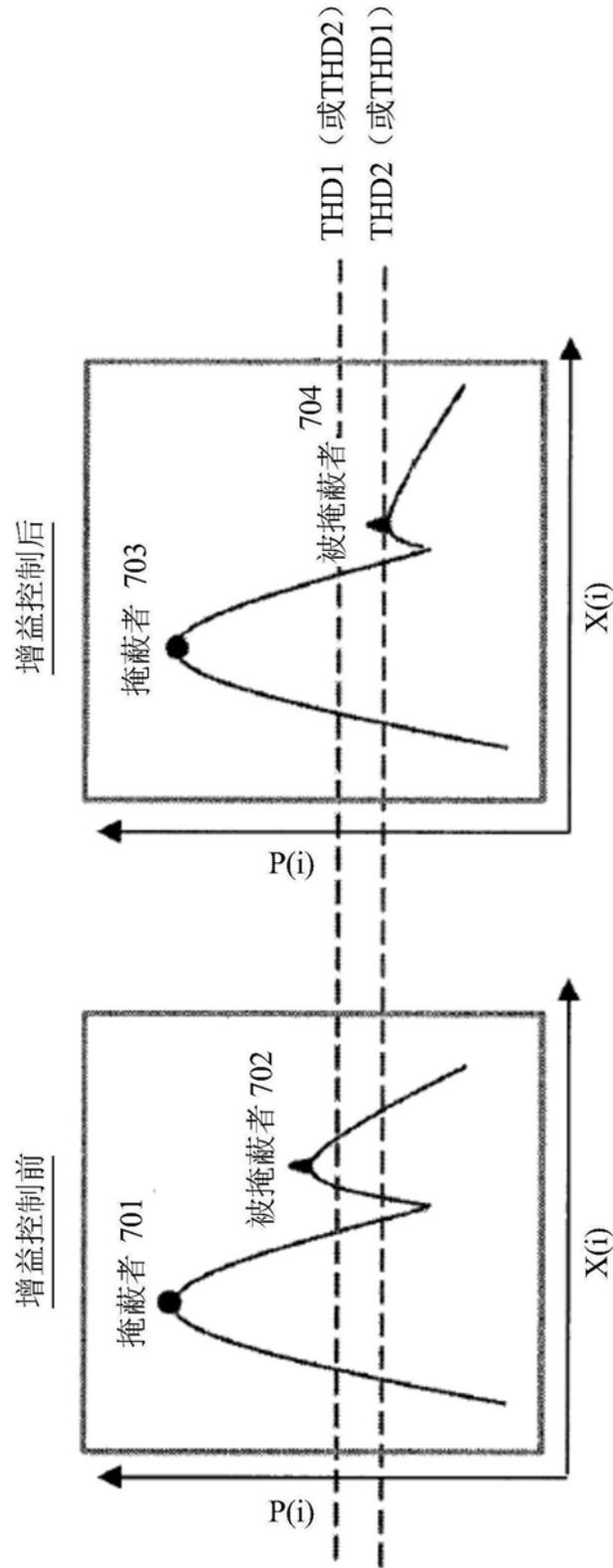


图7

