

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G10L 11/00 (2006.01)

H03G 3/32 (2006.01)



[12] 发明专利申请公开说明书

[21] 申请号 200480014684.5

[43] 公开日 2006 年 6 月 28 日

[11] 公开号 CN 1795490A

[22] 申请日 2004.5.27

[21] 申请号 200480014684.5

[30] 优先权

[32] 2003.5.28 [33] US [31] 60/474,077

[86] 国际申请 PCT/US2004/016964 2004.5.27

[87] 国际公布 WO2004/111994 英 2004.12.23

[85] 进入国家阶段日期 2005.11.28

[71] 申请人 杜比实验室特许公司

地址 美国加利福尼亚州

[72] 发明人 阿伦·J·西非尔特

迈克尔·J·施弥兹尔斯

布莱特·G·克罗克特

[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

代理人 李德山

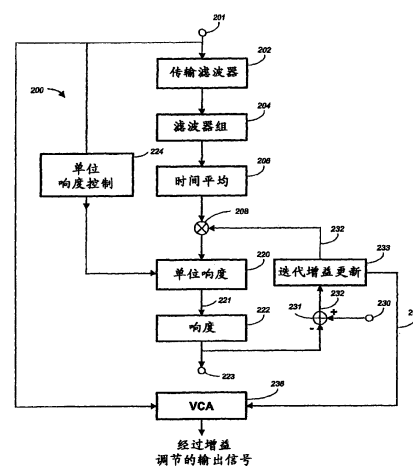
权利要求书 5 页 说明书 25 页 附图 10 页

[54] 发明名称

用于计算和调节音频信号的感觉响度的方法、
设备和计算机程序

[57] 摘要

在计算音频信号感觉响度中应用从一组两个或更多这些函数选择的两个或更多单位响度模型函数之一或其组合。例如，通过测量音频信号是窄带或宽带的程度可以选择函数。替换地或者利用从一组功能的这种选择，计算增益值 $G[t]$ ，当该增益应用于所述音频信号时，其产生基本上和参考响度相同的感度响度。所述增益计算应用包括感觉响度计算的迭代处理循环。



1、一种处理音频信号的方法，包括：

响应于所述音频信号产生激励信号，以及

5 响应于激励信号和音频信号特性的度量计算音频信号的感觉响度，其中所述计算从一组两个或更多单位响度模型函数中选择两个或更多所述单位响度模型函数之一或其组合，通过输入音频信号特性的度量控制这种选择。

2、根据权利要求 1 所述的方法，其中音频信号特性的度量是输入信号是窄带或宽带的程度的度量。

3、根据权利要求 2 所述的方法，还包括通过计算输入信号的频谱平坦度计算该输入信号是窄带或宽带的程度。

4、根据权利要求 1 所述的方法，其中所述计算从两个单位响度模型函数选择或组合它们，通过窄带输入信号产生的特性度量选择第一响度模型函数，通过宽带输入信号产生的特性度量选择第二响度模型函数，以及通过部分窄带、部分宽带输入信号产生的特性度量选择第一和第二响度模型函数的组合。

5、根据权利要求 4 所述的方法，其中所述第一和第二响度模型函数在无声的门限之上根据幂律随增加激励单调地增加，所述第一响度模型函数的增加快于所述第二响度模型函数。

6、根据权利要求 1 所述的方法，其中所述计算从一组两个或更多单位响度模型中选择在激励信号的每一各自频带中的两个或更多所述单位响度模型之一或其组合。

7、根据权利要求 1 所述的方法，其中所述计算从一组两个或更多单位响度模型中选择在激励信号的一组各自频带中的两个或更多所述单位响度模型之一或其组合。

8、根据权利要求 7 所述的方法，其中该组各自频带是所述激励信号的所有频带。

9、根据权利要求 1 所述的方法，其中从所述激励信号推导音频

信号特性的度量。

10、根据权利要求 1 所述的方法，其中所述计算包括计算所述激励信号的每一各自频带中的单位响度。

11、根据权利要求 10 所述的方法，其中所述计算还包括选择频带的单位响度以提供感觉响度，或者组合一组频带的单位响度以提供感觉响度。

12、一种用于处理音频信号的方法，包括：

响应于所述音频信号产生激励信号，以及

至少响应于该激励信号，计算增益值 $G[t]$ ，如果将该增益值应用于音频信号，将会产生基本上与参考响度相同的感觉响度，所述计算包括迭代处理循环，该循环包括至少一个非线性处理。

13、根据权利要求 12 所述的方法，其中所述迭代处理循环包括计算感觉响度。

14、根据权利要求 12 所述的方法，其中所述计算还响应于音频信号特性的度量。

15、根据权利要求 14 所述的方法，其中所述至少一个非线性处理包括单位响度计算，该计算从一组两个或更多单位响度模型函数选择两个或更多所述单位响度模型函数之一或其组合，通过输入音频信号的特性度量控制所述选择。

16、根据权利要求 12 所述的方法，其中所述激励信号是时间平滑的和/或所述方法还包括时间平滑该增益值 $G[t]$ 。

17、根据权利要求 16 所述的方法，其中所述激励信号是线性时间平滑的。

18、根据权利要求 16 所述的方法，其中所述方法还包括平滑增益值 $G[t]$ ，所述平滑应用直方图技术。

19、根据权利要求 12 所述的方法，其中所述迭代处理循环包括时间平滑。

20、根据权利要求 12 到 19 中任一项所述的方法，其中所述迭代处理循环包括：

响应于迭代增益值 G_i 的函数调节激励信号的幅度, 以便所述激励信号的调节幅度随 G_i 的值增加而增加以及随 G_i 的值减小而减小,

响应于经过幅度调节的激励信号计算感觉响度,

比较该计算的音频信号的感觉响度和参考感觉响度以产生差值,

5 以及

响应于该差值调节增益值 G_i , 以便减小该计算的感觉响度和参考感觉响度之间的差值。

21、根据权利要求 20 所述的方法, 其中所述迭代处理循环根据最小算法, 重复地调节所述激励信号的幅度, 计算感觉响度, 比较该
10 计算的感觉响度和参考感觉响度, 以及调节增益值 G_i 成最终值 $G[t]$ 。

22、根据权利要求 21 所述的方法, 其中所述最小算法是根据最小的梯度下降方法。

23、根据权利要求 12 到 22 中任一项所述的方法, 还包括:

利用增益 $G[t]$ 控制输入音频信号的幅度, 以便所产生的输入音频
15 信号的感觉响度基本上与所述参考响度相同。

24、根据权利要求 12 到 23 中任一项所述的方法, 其中通过用户设置所述参考响度。

25、根据权利要求 12 到 23 中任一项所述的方法, 其中所述参考响度是通过根据权利要求 13 所述的计算的感觉响度。

26、根据权利要求 1 或 12 所述的方法, 其中响应于所述音频信号产生激励信号包括:
20

通过模拟外耳和人耳中部特性的函数线性滤波所述音频信号, 以产生线性滤波的音频信号, 以及

将该线性滤波的音频信号划分成频带, 该频带模拟和内耳的基膜一起产生的激励特性曲线以产生所述激励信号。
25

27、根据权利要求 12 所述的方法, 其中所述至少一个非线性处理包括计算所述激励信号的每一频带中的单位响度。

28、根据权利要求 27 所述的方法, 其中所述计算激励信号的每一频带中的单位响度从一组两个或更多单位响度模型函数中选择两个

或更多所述单位响度模型函数之一或其组合，通过所述输入音频信号特性的度量控制所述选择。

29、根据权利要求 20 所述的方法，其中响应于幅度调节的激励信号计算感觉响度包括计算所述激励信号的各自频带中的单位响度。

5 30、根据权利要求 29 所述的方法，其中所述计算激励信号的每一频带中的单位响度从一组两个或更多单位响度模型函数中选择两个或更多所述单位响度模型函数之一或其组合，通过所述输入音频信号特性的度量控制所述选择。

31、根据权利要求 30 所述的方法，其中响应于幅度调节的激励
10 信号计算感觉响度还包括组合每一频带的单位响度成感觉响度的度量。

32、根据权利要求 13、20、21 和 23 中任何一项所述的方法，其中从所计算的感觉响度的度量推导出所述参考感觉响度。

33、根据权利要求 32 所述的方法，其中所述参考感觉响度是所
15 计算的感觉响度的标度形式。

34、根据权利要求 32 所述的方法，其中当所计算的感觉响度低于门限时，所述参考感觉响度大于该计算的感觉响度，以及当所计算的感觉响度高于门限时，所述参考感觉响度小于该计算的感觉响度。

35、一种用于处理多个音频信号的方法，包括：

20 多个处理，每一处理接收各自的一个音频信号，其中每一处理包括：

响应于各自的音频信号产生激励信号，

至少响应于该激励信号计算增益值 $G[t]$ ，其中如果将该增益值应用于音频信号，将会产生基本上与参考响度相同的感度响度，所述计
25 算包括迭代处理循环，该循环包括至少一个非线性处理，以及

利用增益值 $G[t]$ 控制各自音频信号的幅度，以便所产生的各自音频信号的感觉响度基本上与参考响度相同，以及

应用该相同的参考响度给多个处理中的每一处理。

36、一种设备，适于执行根据权利要求 1 到 35 中任一项所述的

方法。

37、在计算机可读媒体上存储的计算机程序，用于使计算机执行权利要求 1 到 35 中任一项所述的方法。

用于计算和调节音频信号的 感觉响度的方法、设备和计算机程序

5

技术领域

本发明涉及音频信号的响度测量，以及响应于这种测量而控制音频信号响度的设备、方法、和计算机程序。

10

背景技术

响度是听觉的主观感觉属性，通过它能够在从无声扩展到大声的标度上排序声音。由于响度是听众所感觉的感觉，因此它不适合于直接的物理测量，所以难以进行量化。另外，由于响度的感觉分量，具有“正常”听力的不同听众可能对同一种声音产生不同的感觉。减少个体感觉所引入的变化和达到音频材料响度的一般测量的唯一方法是集合一群听众并导出响度图形或进行统计排序。这对于逐日标准的响度测量显然是一种不切实际的方法。

已经进行了许多尝试来开发一种令人满意的测量响度的客观的方法。在1933年Fletcher和Munson确定了人类听力在低频和高频比在中间（或声音）频率较不敏感。他们还发现在声音级增加时，灵敏度中的相对变化减小。早先的响度测量仪由麦克风、放大器、测量仪和设计用于在低、中和高声音级粗略模仿听力的频率响应的滤波器组合而组成。

即使这些设备提供了单一的、恒定级独立音调的响度测量，但是更复杂声音的测量不是非常地匹配响度的主观印象。这种类型的声音级测量仪已经标准化，但是其仅仅用于特定的任务，诸如工业噪声的监控和控制。

早在20世纪50年代，其中Zwicker和Stevens在开发响度感觉处理的更为逼真的模型中扩展了Fletcher和Munson的工作。Stevens

在 1956 年美国声学协会杂志上公布了一种方法“**Calculation of the Loudness of Complex Noise**”，以及 Zwicker 在 1958 年的声学文章中公布了他的“**Psychological and Methodical Basis of Loudness**”。在 1959 年，Zwicker 公布了用于响度计算的图像处理以及不久之后又公布了若干的类似文章。Stevens 和 Zwicker 方法（分别）标准化为 ISO 532 部分 A 和 B。两种方法都包含标准的心理声学现象诸如临界频带、频率屏蔽和单位响度。这些方法基于将复合声音划分成落入频率的“临界频带”内的分量，以允许有可能一些信号分量屏蔽其他分量，以及在每一临界频带中增加单位响度以达到声音的完全响度。

正如澳大利亚广播管理局 (ABA) “**Investigation into Loudness of Advertisements**” (2002 年 7 月) 所证明的，近来的研究已经显示出许多广告（和一些节目）被感觉与其他节目相比太大声了，所以对于听众来说非常令人烦恼。ABA 的调查研究仅仅是最近的尝试来解决在实际上所有广播材料和国家多年存在的问题。这些结果显示能够减小或消除由于在节目材料上不相容的响度而引起的听众烦恼，如果可靠的话，能够进行节目响度的相容测量，并用来减小令人烦恼的响度变化。

Bark 标度是在临界频带原理中使用的测量单位。所述临界频带标度是基于这样的事实即人类听力分析广谱成对应于较小临界子频带的部分。以这种方式给下一临界频带增加一个临界频带即较低临界频带的上限是下一较高临界频带的下限，这导致了临界频带速率的标度。如果通过这种方式将临界频带相加起来，那么特定频率对应于每一交叉点。第一临界频带跨越的范围从 0 到 100Hz，第二频带的范围从 100Hz 到 200Hz，第三频带的范围从 200Hz 到 300Hz 等等直到 500Hz，其中每一临界频带的频率范围增加。从 0 到 16kHz 的可听频率范围能够被划分成 24 个邻接的临界频带，其通过增加频率增加了带宽。所述临界频带编号为 0 到 24 并具有定义 Bark 标度的单位“Bark”。临界频带速率和频率之间的关系对于理解人耳的许多特性是重要的。例如，参见 E.Zwicker 和 H.Fastl 所著的 **Psychoacoustics - Facts and Models**,

1990 年柏林 Springer-Verlag 公布。

等效矩形带宽 (ERB) 标度是测量人听力的频率的方式, 它与 Bark 标度相类似。通过 Moore、Glasberg 和 Baer 开发, 它是 Zwicker 响度工作的改进。参见 Moore、Glasberg 和 Baer (B.C.J.Moore, 5 B.Glasberg, T.Baer 的, “A model for the Prediction of Thresholds, Loudness, and Partial Loudness”, 音频工程协会杂志, 45 卷, NO.4, 1997 年 4 月, 第 224 - 240 页)。由于在这样低的频率测量低于 500Hz 的临界频带是困难的, 因此人听觉系统的效率和灵敏度迅速地减小。听觉滤波器带宽的改进测量已经产生了 ERB 速率标度。这些测量使用 10 V 字形噪声屏蔽器测量听觉滤波器带宽。一般而言, 对于 ERB 标度, 听觉滤波器带宽 (以 ERB 为单位表示) 小于 Bark 标度上的带宽。对于较低频率则这种差别变得更大。

低于人听力系统的频率选择性能够通过将声音强度细分为落入临界频带内的部分来近似。这种近似产生了临界频带强度的概念。如果代替理想的临界频带滤波器的无限陡峭斜率, 考虑在人听力系统中 15 产生的实际斜率; 然后这种处理产生了称为激励的强度中间值。通常, 这些值不用作线性值而是用作类似于声压级的对数值。临界频带和激励级是在许多模型中起中间值作用的相应值。(参见 Psychoacoustics - Facts and Models, Supra)。

20 响度级可以以“方”为单位进行测量。一个方定义为在 1dB 声压级 (SPL) 播放的 1kHz 纯正弦波的感觉响度, 其对应于 2×10^{-5} 帕的均方根压强。N 方是在 N dB SPL 播放的 1kHz 音调的感觉响度。在比较除了 1kHz 以外频率的音调响度和 1kHz 音调中使用这种定义, 能够为给定级的方确定等响度轮廓。图 7 示出了用于 20Hz 和 12.5kHz 之间的频率, 以及用于 4.2 方 (被认为是听力的门限) 和 120 方之间方级的 25 等响度级轮廓 (ISO226: 1987 (E), “Acoustics - Normal Equal Loudness Level Contours”)。

响度级还可以以单位“宋”进行测量。如在图 7 中所指示, 在方单位和宋单元之间存在一一映射。1 宋定义为 40dB (SPL) 1kHz 纯正弦

波的响度并且等效于 40 方。宋的单位是这样的：宋单位的两倍增加对应于感觉响度的翻倍。例如，4 宋被感觉为 2 宋两倍的响度。因此，以宋表示响度级更能提供信息。

5 由于宋是音频信号响度的度量，单位响度简单地说是每单位频率的响度。因此，当使用 bark 频率标度时，单位响度具有每 bark 的宋单位，并且同理当使用 ERB 频率标度时，所述单位是每 ERB 宋。

在本文的以下部分，这里使用的术语诸如“滤波器”和“滤波器组”将包括基本上任何形式的递归和非递归滤波诸如 IIR 滤波器和变换，以及“滤波的”信息是应用这些滤波器的结果。以下描述的实施例应用
10 通过 IIR 滤波器和变换实现的滤波器组。

发明内容

根据本发明的一个方面，一种用于处理音频信号的方法包括：响应于所述音频信号产生激励信号，以及响应于激励信号和音频信号特性的测量计算音频信号的感觉响度，其中所述计算从一组两个或更多
15 单位响度模型函数中选择一个或两个或更多单位响度模型函数的组合，并通过输入音频信号特性的测量控制这种选择。

根据本发明的再一个方面，一种用于处理音频信号的方法包括：响应于所述音频信号产生激励信号，以及至少响应于该激励信号，计算增益值 $G[t]$ ，如果将该增益值应用于音频信号，将会产生基本上与
20 参考响度相同的感觉响度，所述计算包括迭代处理循环，该处理包括至少一非线性处理。

根据本发明的又一个方面，一种用于处理多个音频信号的方法包括多个处理，每一处理接收各自的音频信号，其中每一处理响应于各自的音频信号产生激励信号，至少响应于该激励信号计算增益值 $G[t]$ ，
25 如果将该增益值应用于音频信号，将会产生基本上与参考响度相同的感觉响度，所述计算包括迭代处理循环，该处理包括至少一非线性处理，以及利用增益值 $G[t]$ 控制各自音频信号的幅度，以便所产生的各自音频信号的感觉响度基本上与参考响度相同，并应用该相同的参考

响度给多个处理的每一处理。

在应用本发明这些方面的实施例中，一种用于信号处理的方法或设备接收输入音频信号。该信号通过模拟外和中耳特性的滤波器或滤波器功能，以及将滤波的信号划分成频带的滤波器组或滤波器组功能进行线性滤波，其中所述频带模拟和内耳的基膜一起产生的激励特性曲线。对于每一频带，使用一个或更多单位响度函数或模型计算单位响度，其选择是通过从输入音频信号提取的特性或特征来控制的。用于每一频带的单位响度组合成表示宽带输入音频信号的响度测量。可以为输入信号的一些有限的时间范围计算响度测量的单一值，或者在输入音频信号的时段或块上重复计算该响度测量。

在应用本发明这些方面的另一个实施例中，用于信号处理的方法或设备接收输入音频信号。该信号通过模拟外和中耳特性的滤波器或滤波器功能，以及将滤波的信号划分成频带的滤波器组或滤波器组功能进行线性滤波，其中所述频带模拟和内耳的基膜一起产生的激励特性曲线。对于每一频带，使用一个或更多单位响度函数或模型计算单位响度，其选择是通过从输入音频信号提取的特性或特征来控制的。用于每一频带的单位响度组合成表示宽带输入音频信号的响度测量。比较该响度测量与参考响度值，并将其差别用于标度和增益调节先前输入到单位响度计算的划分频带的信号。重复单位响度计算、响度计算和参考比较，直到响度和参考响度值基本上相同为止。因此，应用于划分频带信号的增益表示当增益应用于输入音频信号时，产生输入音频信号的基本上等于参考响度的感觉响度的增益。可以为输入信号的一些有限的时间范围计算响度测量的单一值，或者在输入音频信号的时段或块上重复计算该响度测量。优选增益的递归应用，这是因为感觉响度的非线性特性以及响度测量处理的结构。

通过参考以下公开的内容和附图可以更好地理解本发明的各个方面及其优选实施例，其中在若干附图中相同的附图标记表示相同的单元。示例各个设备和处理的附图示出了有助于理解本发明的主要单元。为了简明起见，附图省略了许多其他特征，这些特征可能在实际

实施中是重要的并且对于本领域的普通技术人员来说是熟知的，但是这对于理解本发明的构思不重要。用于实践本发明的信号处理可以通过大量的方式来实现，其中包括微处理器执行的程序、数字信号处理器、逻辑阵列和其他形式的计算电路。

5

附图说明

图 1 是本发明一个方面的实施例的示意功能方框图。

图 2 是本发明再一个方面的实施例的示意功能方框图。

图 3 是本发明又一个方面的实施例的示意功能方框图。

10 图 4 是在本发明实施例中适于作为传输滤波器的线性滤波器 $P(z)$ 理想的特性响应，其中垂直轴是单位为分贝 (dB) 的衰减以及水平轴是单位为赫兹 (Hz) 的对数底为 10 的频率。

图 5 示出了 ERB 频率标度 (垂直轴) 和单位为赫兹的频率 (水平轴) 之间的关系。

15 图 6 示出了一组近似 ERB 标度上临界频带的理想听觉滤波器特性响应。水平标度是单位为赫兹的频率以及垂直轴是单位为分贝的级。

图 7 示出了 ISO226 的等响度轮廓。水平标度是单位为赫兹的频率 (对数底为 10 的标度) 以及垂直标度是单位为分贝的声压级。

20 图 8 示出了通过传输滤波器 $P(z)$ 规格化的 ISO266 的等响度轮廓。水平轴是单位为赫兹的频率 (对数底为 10 的标度) 和垂直轴是单位为分贝的声压级。

25 图 9 (实线) 示出了用于均匀激励噪声和 1kHz 音调的响度曲线，其中实线是根据本发明的实施例绘制的，参数是选择匹配根据 Zwicker 的实验数据的 (方块和圆圈)。垂直标度是单位为宋的响度 (对数底为 10) 以及水平轴是单位为分贝的声压级。

图 10 是本发明再一个方面的实施例的示意功能方框图。

图 11 是本发明又一个方面的实施例的示意功能方框图。

图 12 是本发明又一个方面的实施例的示意功能方框图。

图 13 是本发明又一个方面的实施例的示意功能方框图。

具体实施方式

如以下更详细所描述的，图 1 示出了本发明第一方面的实施例，包括分析和推导输入音频信号特性的单位响度控制器或控制器功能（“单位响度控制”）124。该音频特性用于控制在单位响度转换器或转换器功能（“单位响度”）120 中的参数。通过使用信号特性调节单位响度参数，本发明的客观响度测量技术可以更近似地匹配通过使用多个听众统计测量响度所产生的主观响度结果。使用信号特性控制响度参数还可以减少不正确测量的发生，这种不正确的测量产生被认为是使听众烦恼的信号响度。

如以下更详细地描述，图 2 中示出了本发明第二方面的实施例，增加增益设备或功能（“迭代增益更新”）233 其目的是为了迭代调节从输入音频信号导出的时间平均的激励信号的增益，直到在图 2 中 223 的相关响度匹配在图 2 中 230 的期望的参考响度。由于感觉响度的客观测量包含固有的非线性处理，迭代循环可以有利地应用于确定适当的增益，以匹配输入音频信号的响度和期望的响度级。但是，围绕整个响度测量系统的迭代增益循环，以至于应用增益调节到用于每一响度迭代的原始输入音频信号，由于产生长期响度的精确测量所需的时间积分，迭代增益循环将会昂贵地实现。一般而言，在这种安排中，所述时间积分需要在迭代中对增益的每一变化进行重新计算。但是，正如以下说明的，在图 2 以及还有图 3，和图 10-12 的实施例中所示的本发明的这些方面中，可以在线性处理通道中执行时间积分，该处理在形成迭代增益循环部分的非线性处理之前和/或之后。线性处理通道不需要形成迭代循环的一部分。因此，例如在图 2 的实施例中，从输入 201 到单位响度转换器或转换器功能（“单元响度”）220 的响度测量通道可以包括在时间平均功能（“时间平均”）206 中的时间积分，并且它是线性的。因此，增益迭代仅仅需要应用于一组减少的响度测量设备或函数，并且不需要包括任何时间积分。在图 2 的实施例中，传输滤波器或传输滤波器功能（“传输滤波器”）202，滤波器组或滤波

器组功能(“滤波器组”)204, 时间平均器或时间平均功能(“时间平均”)206 以及单位响度控制器或单位响度控制功能(“单位响度控制”)224 不是迭代循环的一部分, 这允许在有效和精确的实时系统中实现迭代增益控制。

5 再次参考图 1, 其中示出了根据本发明第一方面的响度测量仪或响度测量处理 100 实施例的功能方框图。要确定响度测量的音频信号施加于响度测量仪或响度测量处理 100 的输入端 101。该输入施加于两条通道 - 第一(主)通道, 计算多个频带的每一频带中的单位响度, 该频带模拟和内耳的基膜一起产生的激励特性曲线, 以及具有单位响
10 度控制器的第二(侧)通道, 用于选择在主通道中应用的单位响度函数或模型。

在优选的实施例中, 在数字域中执行音频处理。相应地, 通过离散时间序列 $x[n]$ 表示音频输入信号, 该序列已经以某采样频率 f_s 从音频源进行采样。假定已适当地标度序列 $x[n]$, 以便 $x[n]$ 以分贝表示的
15 rms 功率:

$$RMS_{dB} = 10 \log_{10} \left(\frac{1}{L} \sum_{n=0}^L x^2[n] \right)$$

等于以 dB 表示的声压级, 在此分贝听众可听得见所述音频。另外, 为了简化说明假定该音频信号是单声道的。但是, 本实施例可以通过以下描述的方式适于多声道音频。

20 传输滤波器 102

在主通道中, 音频输入信号施加于传输滤波器或传输滤波器功能(“传输滤波器”)102, 其输出是音频信号的滤波形式。传输滤波器 102 通过应用线性滤波器 $P(z)$ 模拟通过外耳和耳朵中间部分传输音频的效果。如图 4 中所示, $P(z)$ 的一个适当的幅度频率响应在低于
25 1kHz 时为 1, 以及高于 1kHz 时该响应跟随如在 ISO226 标准中规定的听力门限的倒数, 该门限被规格化成在 1kHz 处等于 1。通过应用传输滤波器, 响度测量处理所处理的音频更接近地近似人听力所感觉的音频, 借此改进了客观响度测量。因此, 传输滤波器 102 的输出是时域输入音频信号 $x[n]$ 的依赖于频率标度的形式。

滤波器组 104

滤波的音频信号施加于滤波器组或滤波器组功能(“滤波器组”) 104 (图 1)。滤波器组 104 被设计用于模拟和内耳的基膜一起产生的激励特性曲线。滤波器组 104 可以包括一组线性滤波器, 其带宽和间隔在等效矩形带宽 (ERB) 频率标度上连续, 如 Moore、Glasberg 和 Baer (B.C.J.Moore, B.Glasberg, T.Baer 的“A model for the Prediction of Thresholds, Loudness, and Partial Loudness”, 如前述) 所定义的。

尽管 ERB 频率标度更接近匹配人的感觉并显示出在产生匹配于主观响度结果的客观响度测量方面的改进性能, 但是也可以应用降低性能的 Bark 频率标度。

对于单位为赫兹的中心频率 f , 单位是赫兹的一个 ERB 频带的宽度可以近似为:

$$ERB(f) = 24.7(4.37f/1000 + 1) \quad (1)$$

从这种关系, 定义变形的频率标度, 以便在任何点和变形的标度一起, 变形标度单位中的相应 ERB 等于 1。用于从单位为赫兹的线性频率转换到该 ERB 频率标度的函数通过对方程 1 的倒数积分而获得:

$$HzToERB(f) = \int \frac{1}{24.7(4.37f/1000 + 1)} df = 21.4 \log_{10}(4.37f/1000 + 1) \quad (2a)$$

通过解方程 2a 求解 f 表示从 ERB 标度转换回线性频率标度是有用的:

$$ERBToHz(e) = f = \frac{1000}{4.37} 10^{(e/21.4-1)} \quad (2b)$$

这里 e 是以 ERB 标度的单位。图 5 示出了 ERB 标度和单位为赫兹的频率之间的关系。

用于滤波器组 104 的听觉滤波器的响应可以使用标准的 IIR 滤波器表征和实现。更具体而言, 在单位为赫兹的中心频率 f_c 处并在滤波器组 104 中实现的各个听觉滤波器可通过第十二阶 IIR 传递函数定义:

$$H_{f_c}(z) = G \frac{(1-z^{-1})(1-2r_B \cos(2\pi f_B/f_c)z^{-1} + r_B^2 z^{-2})}{(1-2r_A \cos(2\pi f_A/f_c)z^{-1} + r_A^2 z^{-2})^6} \quad (3)$$

其中

$$f_A = \sqrt{f_c^2 + B_w^2} \quad (4a)$$

$$r_A = e^{-2\pi B_w / f_A} \quad (4b)$$

$$B_w = \min\{1.55\text{ERB}(f_c), 0.5f_c\} \quad (4c)$$

$$f_B = \min\{\text{ERBscale}^{-1}(\text{ERBscale}(f_c) + 5.25), f_c/2\} \quad (4d)$$

$$r_B = 0.985 \quad (4e)$$

f_s 是单位为赫兹的采样频率，以及 G 是保证每个滤波器在其频率响应峰值处具有单位增益的规格化系数；进行选择以便：

$$\max_{\omega} \{ |H_{f_c}(e^{j\omega})| \} = 1 \quad (4f)$$

- 5 滤波器组 104 可以包括 M 个这种听觉滤波器，称为频带，沿 ERB 标度在中心频率 $f_c[1] \dots f_c[M]$ 均匀间隔。更具体而言，

$$f_c[1] = f_{\min} \quad (5a)$$

$$f_c[m] = f_c[m-1] + \text{ERBToHz}(\text{HzToERB}(f_c[m-1]) + \Delta) \quad m = 2 \dots M \quad (5b)$$

$$f_c[M] < f_{\max} \quad (5c)$$

- 这里 Δ 是期望的滤波器组 104 的 ERB 间隔，其中 $f_{\min}$ 和 $f_{\max}$ 分别是期望的最小和最大中心频率。可以选择 $\Delta = 1$ ，并考虑人耳敏感的频率范围，可以设置 $f_{\min} = 50\text{Hz}$ 和 $f_{\max} = 20000\text{Hz}$ 。利用这些参数，例如，应用方程 6a - c 产生 $M = 40$ 个听觉滤波器。图 6 示出了在 ERB 标度上近似临界频带的这 M 个听觉滤波器的幅度。

- 可替换地，使用有限长度离散傅里叶变换，通常称之为短时离散傅里叶变换 (STDFT)，可以充分地近似所述滤波操作，这是因为实现以音频信号的采样速率运行滤波器，称之为全速率实现，这被认为比精确的响度测量提供更高的时间分辨率。通过使用 STDFT 代替全速率实现，可以实现提高效率 and 减小计算的复杂度。

输入音频信号 $x[n]$ 的 STDFT 定义为：

$$X[k, t] = \sum_{n=0}^{N-1} w[n] x[n + tT] e^{-j \frac{2\pi k n}{N}} \quad (6)$$

- 20 这里 k 是频率索引， t 是时间块索引， N 是 DFT 大小， T 是跳跃大小，以及 $w[n]$ 是长度 N 的窗，其被规格化使得

$$\sum_{n=0}^{N-1} w^2[n] = 1 \quad (7)$$

值得注意的是方程 6 中的变量 t 是表示 STDFT 的时间块的离散索引，相对于以秒为单位的时间测量。 t 中的每一增加表示沿信号 $x[n]$ 的 T 个采样的跳跃。接下来参考索引 t 假设这种定义。尽管可以取决于实现的细节而使用不同的参数设置和窗形状，对于 $f_s = 44100\text{Hz}$ ，选择 $N=4096$ ， $T=2048$ ，并让 $w[n]$ 为产生最佳结果的汉明窗。使用快速傅里叶变换（FFT）上述的 STDFT 可以更加有效。

为了计算输入音频信号的响度，需要测量滤波器组 104 中每一滤波器的音频信号能量。通过将频域中的滤波器响应与输入信号的功率谱相乘，可以近似滤波器组 104 中每一滤波器的瞬时能量输出：

$$E[m, t] = \frac{1}{N} \sum_{k=k_1}^{k_2} |H_{f_m}(e^{j\frac{2\pi}{N}kt})|^2 |P(e^{j\frac{2\pi}{N}kt})|^2 |X[k, t]|^2 \quad (8)$$

这里 m 是频带数量， t 是方块号，以及 p 是传输滤波器。应该注意到除了方程 3 中规定以外的听觉滤波器的幅度响应形式可以在方程 8 中使用以获得类似的结果。例如，Moore 和 Glasberg 建议了一种通过指数函数描述的滤波器形状，该函数的操作类似于方程 3。另外，通过稍微地降低性能，可以将每一滤波器近似为通过一个 ERB 带宽的“砖墙”频带，以及作为另一种近似，传输滤波器 P 可以从求和中拉出。在这种情况下，方程 8 简化成

$$E[m, t] = \frac{1}{N} |P(e^{j2\pi f_m[m]/f_s})|^2 \sum_{k=k_1}^{k_2} |X[k, t]|^2 \quad (9a)$$

$$k_1 = \text{round}(\text{ERBToHz}(\text{HzToERB}(f_s[m]) - 1/2)N / f_s) \quad (9b)$$

$$k_2 = \text{round}(\text{ERBToHz}(\text{HzToERB}(f_s[m]) + 1/2)N / f_s) \quad (9c)$$

因此，滤波器组 104 的激励输出是在每时间周期 t 的各个 ERB 频带 m 中的能量 E 的频域表示。

多声道

对于当输入音频信号是要在多个扬声器上可听得见的多声道格式的情况下，对于每一声道一个扬声器，可以首先如上所述计算用于每一单独声道的激励。为了接下来计算组合的所有声道的感觉响度，可以将单独的激励一起相加成单一激励，以近似到达听众耳朵的激励。然后，对该单一的、求和的激励执行所有接下来的处理。

时间平均 106

心理声学的研究和主观响度测试表明当比较各种音频信号之间的响度时，听众执行某种类型的短期或“瞬时”信号响度的时间积分，以达到用于比较中使用的长期感觉响度的值。当构造一种响度感觉的模型时，已经表明在将激励非线性变换成单位响度之后执行这种时间积分。但是，本发明人已经确定在变换激励成单位响度之前，对该激励使用线性平滑可以充分地模型化这种时间积分。根据本发明的一个方面，通过在计算单位响度之前执行这种平滑，当计算所需应用于信号的增益时，实现显著的优点，以便以预定的方式调节其测量的响度。如以下进一步说明的，可以通过使用迭代循环计算增益，该循环不仅排除激励计算还优选排除这种时间积分。通过这种方式，迭代循环通过计算可以产生增益，该计算仅仅取决于为其计算增益的当前时间帧，正如该计算对立于取决于时间积分的整个时段的计算。该结果节约了处理时间和存储器。使用迭代循环计算增益的实施例包括以下结合图 2、3 和 10-12 描述的内容。

返回到图 1 的描述，可以各种方式来实现激励的线性平滑。例如，可以使用应用以下方程的时间平均设备或功能（“时间平均”）106 递归地执行平滑：

$$\tilde{E}[m,t] = \tilde{E}[m,t-1] + \frac{1}{\tilde{\sigma}[m,t]} (E[m,t] - \tilde{E}[m,t-1]) \quad (10a)$$

$$\tilde{\sigma}[m,t] = \lambda_m \tilde{\sigma}[m,t-1] + 1 \quad (10b)$$

这里初始条件是 $\tilde{E}[m,-1] = 0$ 和 $\tilde{\sigma}[m,-1] = 0$ 。平滑滤波器的唯一特征是通过改变平滑参数 λ_m ，平滑的能量 $\tilde{E}[m,t]$ 可以从 $E[m,t]$ 的真实时间平均变化成 $E[m,t]$ 的衰落记忆平均。如果 $\lambda_m = 1$ ，那么根据 (10b) 可以看出 $\tilde{\sigma}[m,t] = t$ ，那么对于从 0 到 t 的时间块， $\tilde{E}[m,t]$ 等于 $E[m,t]$ 的真实时间平均。如果 $0 \leq \lambda_m < 1$ ，那么当 $t \rightarrow \infty$ 时， $\tilde{\sigma}[m,t] \rightarrow 1/(1 - \lambda_m)$ ，以及 $\tilde{E}[m,t]$ 简化为应用一个极点平滑器到 $E[m,t]$ 的结果。对于期望描述有限长度音频分段的长期响度的单数字应用来说，对于所有 m 可以设置 $\lambda_m = 1$ 。对于愿意实时跟踪连续音频流的时间变化长期响度的实时应用，对于所有 m ，可以设置 $0 \leq \lambda_m < 1$ 并且设置 λ_m 成相同的值。

在计算 $E[m, t]$ 的时间平均中, 可以期望省略被认为“太静”的瞬时分段并且不会促进感觉响度。为了实现这, 可以在方程 10 中与平滑器并行地运行第二门限平滑器。如果 $E[m, t]$ 小于 $\bar{E}[m, t]$, 则该第二平滑器保持它的当前值:

$$\bar{E}[m, t] = \begin{cases} \bar{E}[m, t-1] + \frac{1}{\sigma[m, t]} (E[m, t] - \bar{E}[m, t-1]), & \sum_{m=1}^M E[m, t] > 10^{\frac{tdB}{10}} \sum_{m=1}^M \bar{E}[m, t] \\ \bar{E}[m, t-1], & \text{其他} \end{cases} \quad (11a)$$

$$\sigma[m, t] = \begin{cases} \lambda_m \sigma[m, t-1] + 1, & \sum_{m=1}^M E[m, t] > 10^{\frac{tdB}{10}} \sum_{m=1}^M \bar{E}[m, t] \\ \sigma[m, t-1], & \text{其他} \end{cases} \quad (11b)$$

5

这里 tdB 是以分贝为单位规定的相对门限。尽管对于本发明不是关键, 但是已经发现 tdB 的值 = -24 产生好的结果。如果不存在并行运行的第二平滑器, 那么 $\bar{E}[m, t] = \bar{E}[m, t]$ 。

单位响度 120

10

它对于频带时间平均的激励保留能量 $\bar{E}[m, t]$ 被转换成感觉单位, 在这种情况下单位为宋的单一响度测量。在单位响度转换器或转换功能 (“单位响度”) 120 中, 激励的每一频带被转换成单位响度的值, 其中每 ERB 以宋测量该值。在响度组合器或响度组合功能 (“响度”) 122 中, 可以积分单位响度的值或者在频带上进行求和以产生总的感觉响度。

15

单位响度控制 124/单位响度 120

多模型

20

在一个方面, 本发明在方块 120 中使用多个模型用于转换激励成频带单位响度。经由单位响度控制 124 在侧通道中从输入音频信号导出的控制信息选择模型或控制模型促进单位响度的程度。在方块 124 中, 从所述音频中提取有利于从可用的模型中选择一个或更多单位响度模型的特定特征或特性。根据该提取的特征或特性产生指示应该使用哪一模型或模型组合的控制信号。在可能期望使用不止一个模型的情况下, 所述控制信息还可以指示应该如何组合这些模型。

25

例如, 每频带单位响度 $N'[m, t]$ 可以表示为用于每一模型的每频带单位响度 $N'_q[m, t]$ 的线性组合:

$$N'[m, t] = \sum_{q=1}^Q \alpha_q[m, t] N'_q[m, t] \quad (12)$$

这里 Q 指示模型的总数量以及控制信息 $\alpha_q[m, t]$ 表示每一模型的加权或贡献。加权的求和可以或不可以等于 1，这取决于要使用的模型。

尽管本发明并不限制于此，但是已经发现两个模型给出了精确的结果。当所述音频信号表征为窄带时，一个模型执行最佳，以及当该音频信号表征为宽带时，另一个模型执行最佳。

初始，在计算单位响度中，可以将 $\bar{E}[m, t]$ 的每一频带中的激励级转换成在 ISO266 (图 7) 的等响度轮廓所规定的 1kHz 处相同的激励级，通过传输滤波器 $P(z)$ (图 8) 规格化该等响度轮廓：

$$\bar{E}_{1kHz}[m, t] = L_{1kHz}(\bar{E}[m, t], f_c[m]) \quad (13)$$

这里 $L_{1kHz}(E, f)$ 是产生在 1kHz 处级的函数，其等于频率 f 处的级 E 。实际上， $L_{1kHz}(E, f)$ 被实现为通过传输滤波器规格化的等响度轮廓查询表的内插。变换到 1kHz 处的相同级简化了以下单位响度计算。

接下来，每一频带中的单位响度可以如下计算：

$$N'[m, t] = \alpha[m, t] N'_{NB}[m, t] + (1 - \alpha[m, t]) N'_{WB}[m, t] \quad (14)$$

其中 $N'_{NB}[m, t]$ 和 $N'_{WB}[m, t]$ 分别是基于窄带和宽带信号模型的单位响度值。该值 $\alpha[m, t]$ 是根据音频信号计算的位于 0 和 1 之间的内插系数，以下将描述细节。

使用以下指数函数从频带激励可以估计窄带和宽带单位响度值 $N'_{NB}[m, t]$ 和 $N'_{WB}[m, t]$ ：

$$N'_{NB}[m, t] = \begin{cases} G_{NB} \left(\left(\frac{\bar{E}_{1kHz}[m, t]}{TQ_{1kHz}} \right)^{s_{NB}} - 1 \right), & \bar{E}_{1kHz}[m, t] > 10^{\frac{TQ_{1kHz}}{10}} \\ 0, & \text{其他} \end{cases} \quad (15a)$$

$$N'_{WB}[m, t] = \begin{cases} G_{WB} \left(\left(\frac{\bar{E}_{1kHz}[m, t]}{TQ_{1kHz}} \right)^{s_{WB}} - 1 \right), & \bar{E}_{1kHz}[m, t] > 10^{\frac{TQ_{1kHz}}{10}} \\ 0, & \text{其他} \end{cases} \quad (15b)$$

这里 TQ_{1kHz} 是在用于 1kHz 音调的无声门限的激励级。根据等响度轮廓 (图 7 和 8)， TQ_{1kHz} 等于 4.2dB。注意到当该激励等于无声的门限时，这些单位响度函数都等于零。对于大于无声门限的激励，这些函

数根据 Stevens 的强度敏感定律以幂律单调增加。选择窄带函数的指数大于宽带函数的指数，使得窄带函数比宽带函数更迅速地增加。以下将讨论对用于窄带和宽带情况的指数 β 和增益 G 的具体选择。

响度 122

- 5 响度 122 使用单位响度 120 的频带单位响度以产生音频信号的单一响度测量，即在端 123 的输出感觉单位的响度值。该响度测量可以具有任意的单位，只要用于不同音频信号的响度值的比较指示哪一更大声和哪一更温柔。

可以将以宋为单位表示的总响度计算为用于所有频带的单位响度的之和；

$$S[t] = \Delta \sum_{m=1}^M N[m, t] \quad (16)$$

这里 Δ 是方程 6b 中规定的 ERB 间隔。选择方程 15a 中的参数 G_{NB} 和 β_{NB} ，以便当 $\alpha[m, t] = 1$ 时，以宋为单位的 S 的曲线与用于 1kHz 音调的 SPL 的关系曲线基本上匹配 Zwicker 所提供的相应实验数据（图 9 中的圆圈）（Zwicker, H.Fastl, “Psychoacoustics - Facts and Models,” 如前述）。选择方程 15b 中的参数 G_{WB} 和 β_{WB} ，以便当 $\alpha[m, t] = 0$ 时，以宋为单位的 N 的曲线与用于均匀激励噪声（在每一 ERB 具有相等功率的噪声）的 SPL 的关系曲线基本上匹配来自 Zwicker 的相应结果（图 9 中的方块）。适于 Zwicker 数据的最小二乘法产生：

$$G_{NB} = 0.0404 \quad (17a)$$

$$\beta_{NB} = 0.279 \quad (17b)$$

$$G_{WB} = 0.058 \quad (17c)$$

$$20 \quad \beta_{WB} = 0.212 \quad (17d)$$

图 9（实线）示出了用于均匀激励噪声和 1kHz 音调的响度曲线。
单位响度控制 124

如上所述，在实际的实施（方程 15a 和 15b）中使用单位响度的两种模型，一种模型用于窄带信号以及另一种模型用于宽带信号。在侧通道中的单位响度控制 124 计算度量 $\alpha[m, t]$ ，它指示在每一频带中
25 中输入信号是窄带或宽带的程度。一般而言，当信号是接近于频带中

心频率 $f_c[m]$ 的窄带时, $\alpha[m, t]$ 应该等于 1, 以及当该信号是接近于频带中心频率 $f_c[m]$ 的宽带时, $\alpha[m, t]$ 应该等于 0。所述控制应该在改变这些特征混合的两个极限之间连续地变化。作为一种简化, 控制 $\alpha[m, t]$ 可以选择为频带上的常数, 在这种情况下 $\alpha[m, t]$ 基本上被称为 $\alpha[t]$, 而省略了频带索引 m 。然后, 控制 $\alpha[t]$ 表示信号在整个频带上如何窄带的测量。尽管以下将描述用于产生这种控制的适合方法, 但是特定的方法并不关键, 而是可以应用其他适合的方法。

控制 $\alpha[t]$ 可以在滤波器组 104 的输出端根据激励 $E[m, t]$ 计算, 而不通过信号 $x[n]$ 的一些其他处理。 $E[m, t]$ 可以提供足够的参考, 根据该参考测量 $x[n]$ 的“窄带”和“宽带”, 结果, 利用几乎很少增加的计算就可以产生 $\alpha[t]$ 。

“频谱平坦度”是根据其可以计算 $\alpha[t]$ 的 $E[m, t]$ 的特征。频谱平坦度是几何平均与算术平均的比值, 如 Jayant 和 Noll 所定义 (N.S.Jayant, P.Noll, Digital Coding Of Waveforms, Prentice Hall, 新泽西州, 1984 年), 其中在频率 (在 $E[m, t]$ 情况中的索引 m) 上采用该平均。当 $E[m, t]$ 在 m 上为常数时, 几何平均等于算术平均, 并且频谱平坦度等于 1。这对应于宽带信号的情形。如果 $E[m, t]$ 在 m 上显著地变化, 那么几何平均显著小于算术平均, 以及频谱平坦度近似零。这对应于窄带信号的情形。通过计算 1 减去频谱平坦度, 可以产生“窄带”的测量, 其中零对应于宽带以及 1 对应于窄带。具体而言, 可以计算 1 减去 $E[m, t]$ 的修改的频谱平坦度:

$$NB[t] = 1 - \frac{\left(\prod_{m=M_l[t]}^{M_u[t]} \frac{E[m, t]}{|P[m]|^2} \right)^{\frac{1}{M_u[t]-M_l[t]+1}}}{\frac{1}{M_u[t]-M_l[t]+1} \sum_{m=M_l[t]}^{M_u[t]} \frac{E[m, t]}{|P[m]|^2}} \quad (18)$$

其中 $P[m]$ 等于在频率 $\omega = 2\pi f_c[m]/f_s$ 所采样的传输滤波器 $P(z)$ 的频率响应。通过传输滤波器规格化 $E[m, t]$ 可以提供更好的结果, 因为应用该传输滤波器在 $E[m, t]$ 中引入了“凸起 (bump)”, 它有助于“窄带”测量。另外, 在 $E[m, t]$ 的频带子集上计算频谱平坦度可以产生更好的结果。方程 18 中求和的下限和上限, $M_l[t]$ 和 $M_u[t]$, 定义可以小

于所有 M 个频带的范围的区域。期望 $M_l[t]$ 和 $M_u[t]$ 包括 $E[m, t]$ 的部分, 该部分包含它的大部分能量, 以及 $M_l[t]$ 和 $M_u[t]$ 所定义的范围仅仅是在 ERB 标度上的 24 单位宽。更具体而言 (并且回想 $f_c[m]$ 是单位赫兹的频带 m 的中心频率), 期望:

$$5 \quad \text{HzToERB}(f_c[M_u[t]]) - \text{HzToERB}(f_c[M_l[t]]) \cong 24 \quad (19a)$$

并且需要

$$\text{HzToERB}(f_c[M_u[t]]) \geq CT[t] \geq \text{HzToERB}(f_c[M_l[t]]) \quad (19b)$$

$$\text{HzToERB}(f_c[M_l[t]]) \geq \text{HzToERB}(f_c[l]) \quad (19c)$$

$$\text{HzToERB}(f_c[M_u[t]]) \leq \text{HzToERB}(f_c[M]) \quad (19d)$$

其中 $CT[t]$ 是在 ERB 标度上测量的 $E[m, t]$ 的频谱重心:

$$CT[t] = \frac{\sum_{m=1}^M \text{HzToERB}(f_c[m]) E[m, t]}{\sum_{m=1}^M E[m, t]} \quad (19e)$$

10 理想情况下, 当在 ERB 标度上测量时, $M_l[t]$ 和 $M_u[t]$ 求和的极限定位于 $CT[t]$ 周围的中心, 但是当 $CT[t]$ 接近它范围的下限或上限时, 并不总是有可能出现这种情况。

接下来, $NB[t]$ 可以以一种类似于方程 11a 的方式在时间上进行平滑:

$$15 \quad \overline{NB}[t] = \begin{cases} \overline{NB}[t-1] + \frac{1}{\sigma[t]} (NB[t] - \overline{NB}[t-1]) & \sum_{m=1}^M E[m, t] > 10^{\frac{NB}{10}} \sum_{m=1}^M \tilde{E}[m, t] \\ \overline{NB}[t-1], & \text{其他} \end{cases} \quad (20)$$

这里 $\sigma[t]$ 在整个 m 上等于方程 11b 中定义的 $\sigma[m, t]$ 的最大值。最后, 根据 $\overline{NB}[t]$ 计算 $\alpha[t]$ 如下:

$$\alpha[t] = \begin{cases} 0, & \Phi\{\overline{NB}[t]\} < 0 \\ \Phi\{\overline{NB}[t]\}, & 0 \leq \Phi\{\overline{NB}[t]\} \leq 1 \\ 1, & \Phi\{\overline{NB}[t]\} \geq 1 \end{cases} \quad (21a)$$

其中

$$20 \quad \Phi\{x\} = 12.2568x^3 - 22.8320x^2 + 14.5869x - 2.9594 \quad (21b)$$

尽管 $\Phi\{x\}$ 的精确形式并不关键, 但是可以通过相对大量音频资料的主观测量的响度优化 $\alpha[t]$ 得到方程 21b 中的多项式。

图 2 示出了根据本发明第二方面的响度测量仪或响度测量处理

200 实施例的功能方框图。图 2 的设备或函数 202、204、206、220、222、223 和 224 对应于图 1 的各个设备或函数 102、104、106、120、122、123 和 124。

根据本发明的第一方面，图 1 示出了实施例，响度测量仪或计算产生感觉单位的响度值。为了调节输入信号的响度，有用的测量是增益 $G[t]$ ，当将其与输入信号 $x[n]$ 相乘时（例如，在以下描述的图 3 的实施例中），使其响度等于参考响度级 S_{ref} 。可以任意规定或通过根据本发明第一方面操作的另一设备或处理根据一些“已知的”参考音频信号测量参考响度 S_{ref} 。让 $\Psi\{x[n], t\}$ 表示对信号 $x[n]$ 执行的所有计算以产生响度 $S[t]$ ，想要得到 $G[t]$ 以便

$$S_{ref} = S[t] = \Psi\{G[t]x[n], t\} \quad (23)$$

由于在 $\Psi\{\cdot\}$ 中实施的处理部分是非线性的，存在用于 $G[t]$ 的非封闭形式解决方案，因此可以使用其代替迭代技术找到近似的解决方案。在处理中的每一迭代 i ，让 G_i 表示 $G[t]$ 的当前估计。对于每一迭代，更新 G_i ，以便与参考响度的绝对误差减小：

$$|S_{ref} - \Psi\{G_i x[n], t\}| < |S_{ref} - \Psi\{G_{i-1} x[n], t\}| \quad (24)$$

存在许多用于更新 G_i 的适当技术，以便获得误差中的上述减小。一种这样的方法是梯度下降（参见 Dimitri P. Bertsekas 的 Nonlinear Programming, Athena Scientific, 贝尔蒙特, 1995 年 MA），其中通过与前一代的误差成比例的量更新 G_i ：

$$G_i = G_{i-1} + \mu(S_{ref} - \Psi\{G_{i-1} x[n], t\}), \quad (25)$$

这里 μ 是迭代的步进大小。上述迭代继续直到上述绝对误差低于某门限，迭代的次数到达某预定的最大限制，或者已经过了规定的时间。在该点上设置 $G[t]$ 等于 G_i 。

再次参考方程 6-8，注意到通过对信号 STDFT 幅度的平方进行线性操作获得信号 $x[n]$ 的激励，即 $|X[k, t]|^2$ 。接着，由增益修改的信号 $Gx[n]$ 产生的激励等于 $x[n]$ 的激励乘以 G^2 。而且，所需用于估计长期感觉响度的时间积分可以通过激励的线性平均而执行，所以，对应于 $Gx[n]$ 的时间平均激励对等于 $x[n]$ 的时间平均激励乘以 G^2 。结果，在

上述的迭代处理中，不需要在整个输入信号历史上对于 $\Psi\{G_i x[n], t\}$ 的每一重新估计重新计算所述时间平均。相反，在该迭代中，根据 $x[n]$ 可以仅仅计算一次 $\bar{E}[m, t]$ 的时间平均激励，通过直接应用更新增益的平方给 $\bar{E}[m, t]$ 可以计算响度的更新值。具体而言，让 $\Psi_E\{\bar{E}[m, t]\}$ 表示对时间平均激励 $\bar{E}[m, t]$ 执行的产生 $S[t]$ 的所有处理，对于通常的相乘增益 G 保持以下关系；

$$\Psi_E\{G^2 \bar{E}[m, t]\} = \Psi\{G x[n], t\} \quad (26)$$

使用这种关系，通过利用 $\Psi_E\{G_i^2 \bar{E}[m, t]\}$ 代替 $\Psi\{G_i x[n], t\}$ 可以简化迭代处理。如果在非线性变换到单位响度之后，已经执行了所需用于估计长期感觉响度的时间积分，则这种简化将会不可能。

图 2 中描述了计算 $G[t]$ 的迭代处理。在减法组合器或组合函数 231 中可以从端 230 的参考响度 S_{ref} 减去在端 223 的输出响度 $S[t]$ 。所产生的误差信号 232 馈送到在迭代中产生下一增益 G_i 的迭代增益更新器或更新功能（“迭代增益更新”）233。该增益的平方， G_i^2 ，然后在输出 234 反馈到相乘组合器 208，在那里将 G_i^2 乘以来自方块 206 的时间平均激励信号。在迭代中，然后根据时间平均激励的该增益修改形式通过方块 220 和 222 计算 $S[t]$ 的下一值。上述循环迭代直到满足终止条件，即端 235 的增益 $G[t]$ 被设置为等于 G_i 的当前值。例如对于每一 FFT 帧 t ，或者在该分段的整个长度上平均所述激励之后，仅仅一次在音频分段的结尾，通过上述迭代处理可以计算最终值 $G[t]$ 。

如果希望结合该迭代处理计算非增益修改信号响度，能够在用于每一时间周期 t 的每一迭代处理的开始初始化增益 G_i 为 1。该方法中，在循环中计算的 $S[t]$ 的第一值表示原始信号响度并且能够将其如此记录。但是，如果不希望记录该值，可以利用任何值初始化 G_i 。在相继时间帧上计算 $G[t]$ 以及不希望记录原始信号响度的情况下，可能希望从前一时间周期初始化 G_i 等于 $G[t]$ 的值。该方法中，如果从前一时间周期信号没有显著地变化，有可能 $G[t]$ 值将仍然保持基本上相同。所以，将只需要少许迭代以收敛于正确的值。

一旦完成迭代， $G[t]$ 表示在 201 通过一些外部设备应用于输入音

频信号的增益，以便修改信号的响度匹配所述参考响度。图3示出了一种适合的安排，其中来自迭代增益更新233的增益 $G[t]$ 应用于信号电平控制设备或函数诸如压控放大器(VCA)236的控制输入，以便提供经过增益调节的输出信号。图3中的VCA234可以通过控制增益调节器的操作员响应于线路235上的增益 $G[t]$ 的传感指示进行替换。传感指示可以例如由测量仪提供。增益 $G[t]$ 可以经过时间平滑(未示出)。

对于某些信号，可以希望方程10和11中所描述平滑的替换用于计算长期感觉响度。听众倾向于将信号的长期响度于该信号的最高声部分关联起来。结果，方程10和11中表示的平滑可能低估信号的感觉响度，该信号包含由较高声资料的较短分段所中断的长周期的相对静默。通常在具有周围场景嘈杂的较长周期所围绕对话的短分段的电影声轨中找到这种信号。即使利用方程11表示的门限，这些信号的安静部分也可能极大地有助于时间平均激励 $\bar{E}[m, t]$ 。

为了处理这种问题，在本发明的另一个方面可以应用计算长期响度的统计技术。首先，方程10和11中的平滑时间常数非常小，并设置 τ dB成负无穷大，从而 $\bar{E}[m, t]$ 表示所述“瞬时”激励。在这种情况下，可以选择平滑参数 λ_m 在频带 m 上变化成更精确的模型，其方式是瞬时响度的感觉在频率上变化。但是，在实际中，选择 λ_m 为 m 上的常数仍然产生可接受的结果。如在方程16中所规定的，前面描述算法的剩余部分操作不变，这将产生瞬时响度信号 $S[t]$ 。在某范围 $t_1 \leq t \leq t_2$ 上，然后对于该范围中时间值的 p 百分比，定义长期响度 $S_p[t_1, t_2]$ 为大于 $S[t]$ 的值，以及对于该范围中时间值的 $100 - p$ 百分比，定义长期响度 $S_p[t_1, t_2]$ 为小于 $S[t]$ 的值。实验已经表明设置 p 等于大致90%匹配主观感觉的长期响度。利用这种设置，只有10%的 $S[t]$ 的值需要显著地影响长期响度。该值的其他90%相对安静而不会降低长期响度测量。

通过升序排序值 $S[t]$ ， $t_1 \leq t \leq t_2$ ，成列表 $S_{\text{sort}}\{i\}$ ， $0 \leq i \leq t_2 - t_1$ 能够计算值 $S_p[t_1, t_2]$ ，这里 i 表示排序列表的第 i 个单元。然后通过 p 百分比的单元到列表中给出长期响度：

$$S_p[t_1, t_2] = S_{\text{sort}}\{\text{round}(p(t_2 - t_1)/100)\} \quad (27)$$

上述计算相对简单。但是，如果希望计算增益 $G_p[t_1, t_2]$ ，其中当其乘以 $x[n]$ 时产生了等于某参考响度 S_{ref} 的 $S_p[t_1, t_2]$ 时，该计算变得更加复杂。正如前所述，需要迭代方法，但是现在长期响度测量 $S_p[t_1, t_2]$ 取决于值 $S[t]$ 的整个范围， $t_1 \leq t \leq t_2$ ，在所述迭代中必须利用 G_i 的每一更新来更新其中的每个值。为了计算这些更新，必须在整个范围 $t_1 \leq t \leq t_2$ 上存储信号 $E[m, t]$ 。另外，由于 $S[t]$ 对 G_i 的依赖是非线性的，因此 $S[t]$ 的相对顺序， $t_1 \leq t \leq t_2$ ，可以利用每一迭代改变，并且所以，还必须重新计算 $S_{\text{sort}}\{i\}$ 。当对于迭代中的特定增益考虑短时间信号分段时，显然需要重新排序，其中该分段的频谱正好低于听力的门限。当增加所述增益时，分段频谱的重要部分可以变得可听得见，这可能使得分段的总响度大于先前可听信号的其他窄带分段。当范围 $t_1 \leq t \leq t_2$ 变大或者如果希望连续计算增益 $G_p[t_1, t_2]$ 为滑动时间窗的函数时，这种迭代处理的计算和存储器成本可能变为禁止。

通过实现 $S[t]$ 是 G_i 的单调增加函数从而达到了计算和存储器的显著节约。换言之，在每一时间瞬间增加 G_i 总是增加短期响度。利用这种知识，想要的匹配增益 $G_p[t_1, t_2]$ 能够如下有效地计算。首先，使用所述的迭代根据 $E[m, t]$ ，对于范围 $t_1 \leq t \leq t_2$ 中的所有值 t 计算前一定义的匹配增益 $G[t]$ 。注意到对于每一值 t ，通过在单值 $E[m, t]$ 上迭代来计算 $G[t]$ 。接下来，通过升序排序值 $G[t]$ ， $t_1 \leq t \leq t_2$ ，到列表 $G_{\text{sort}}\{i\}$ ， $0 \leq i \leq t_2 - t_1$ ，计算长期匹配增益 $G_p[t_1, t_2]$ ，然后设置

$$G_p[t_1, t_2] = G_{\text{sort}}\{\text{round}((100 - p)(t_2 - t_1)/100)\}. \quad (28)$$

现在论证 $G_p[t_1, t_2]$ 等于增益，当其乘以 $x[n]$ 时产生等于期望的参考响度 S_{ref} 的 $S_p[t_1, t_2]$ 。从方程 28 注意到，对于范围 $t_1 \leq t \leq t_2$ 中的时间值 $100 - p$ 百分比， $G[t] < G_p[t_1, t_2]$ ，对于其他 p 百分比则 $G[t] > G_p[t_1, t_2]$ 。对于 $G[t]$ 的这些值诸如 $G[t] < G_p[t_1, t_2]$ ，注意到如果将 $G_p[t_1, t_2]$ 应用于 $E[m, t]$ 而不是 $G[t]$ 的相应值，那么所产生的 $S[t]$ 值将会大于期望的参考响度。由于 $S[t]$ 是增益的单调增加函数，因此这是事实。同理，如果将 $G_p[t_1, t_2]$ 应用于对应于 $G[t]$ 的 $E[m, t]$ 值诸如 $G[t] > G_p[t_1, t_2]$ ，所产

生的 $S[t]$ 值将会小于期望的参考响度。所以，应用 $G_p[t_1, t_2]$ 到范围 $t_1 \leq t \leq t_2$ 中的所有值 $E[m, t]$ ，产生的 $S[t]$ 大于期望的时间的参考 $100 - p$ 百分比并小于时间的参考 p 百分比。换言之， $S_p[t_1, t_2]$ 等于期望的参考。

5 这种替换的计算匹配增益的方法消除了的范围 $t_1 \leq t \leq t_2$ 上存储 $E[m, t]$ 和 $S[t]$ 的需要。只需要存储 $G[t]$ 。另外，对于计算的每一值 $G_p[t_1, t_2]$ ，只需要执行一次在范围 $t_1 \leq t \leq t_2$ 上排序 $G[t]$ ，这与前一方法相反，其中每一迭代都需要重新排序 $S[t]$ 。在某长度 T 滑动窗（即， $t_1 = t - T$ ， $t_2 = t$ ）上连续计算 $G_p[t_1, t_2]$ 的情况下，对于每一新的时间瞬间，通过
10 简单地从排序列表中删除和增加单个值能够有效地保持列表 $G_{\text{sort}}\{i\}$ 。当范围 $t_1 \leq t \leq t_2$ 变得非常大时（例如，整个歌曲或电影的长度），仍可能禁止所需用于存储 $G[t]$ 的存储器。在这种情况下，根据 $G[t]$ 的离散直方图可以近似 $G_p[t_1, t_2]$ 。实际上，根据 $G[t]$ 以分贝为单位创建该直方图。该直方图可以计算为：

15 $H[i] = t_1 \leq t \leq t_2$ 范围内的样本数，使得

$$\Delta_{\text{dB}} i + dB_{\text{min}} \leq 20 \log_{10} G[t] < \Delta_{\text{dB}} (i+1) + dB_{\text{min}} \quad (29)$$

这里 Δ_{dB} 是直方图分辨率以及 dB_{min} 是直方图最小值。然后匹配增益可以近似为：

$$G_p[t_1, t_2] \cong \Delta_{\text{dB}} I_p + dB_{\text{min}} \quad (30a)$$

20 其中

$$100 \frac{\sum_{i=0}^I H[i]}{\sum_{i=0}^I H[i]} \cong p \quad (30b)$$

以及 I 是最大直方图索引。使用该离散的直方图，只需要存储 I 值，以及利用 $G[t]$ 的每一新值容易地更新 $G_p[t_1, t_2]$ 。

可以构思根据 $G[t]$ 近似 $G_p[t_1, t_2]$ 的其他方法，并且本发明旨在
25 包括这些技术。本发明该部分的关键方面在于对匹配增益 $G[t]$ 执行某种类型的平滑以产生长期匹配增益 $G_p[t_1, t_2]$ ，而不是处理瞬时响度 $S[t]$ 来产生长期响度 $S_p[t_1, t_2]$ ，从该长期响度然后通过迭代处理估计 $G_p[t_1, t_2]$ 。

图 10 和 11 显示了分别类似于图 2 和 3 的系统，但是匹配增益 $G[t]$ 的平滑（设备或功能 237）用于产生平滑的增益信号 $G_p[t_1, t_2]$ （信号 238）。

在输入 230 的参考响度（图 2、3、10、11）可以是“固定的”或“可变的”，参考响度的源可以在实施本发明这些方面的安排的内部或外部。例如，所述参考响度可由用户设置，在这种情况下该参考响度的源是外部的并且它可以保持“固定”一段时间周期直到用户重新设置它。可替换地，参考响度可以是根据本发明从响度测量处理或设备推导出的另一音频源的响度测量，诸如图 1 的实例中示出的结构。

音频产生设备的正常音量控制可代替为根据本发明这些方面的处理或设备诸如图 3 或图 11 的实例。在这种情况下，用户操作的音量按钮或滑动器等等将会控制图 3 或图 11 的 230 处的参考响度，因此，音频产生设备将具有与音量控制的用户调节相匹配的响度。

图 12 中示出了可变参考的实例，其中参考响度 S_{ref} 被代替为例如通过可变参考响度设备或功能（“可变参考响度”）239 根据响度信号 $S[t]$ 计算的可变参考 $S_{ref}[t]$ 。在这种安排中，在对于每一时间周期 t 的每一迭代的开始，在 208 将任何增益已经应用于所述激励之前，根据未修改的响度 $S[t]$ 可以计算可变参考 $S_{ref}[t]$ 。通过可变响度参考功能 239 的 $S_{ref}[t]$ 和 $S[t]$ 的相关性可以采取各种形式来实现各种效果。例如，所述函数可以简化标度 $S[t]$ 来产生为原始响度的某固定比率的参考。可替换地，当 $S[t]$ 低于某一门限时该函数可能产生大于 $S[t]$ 的参考，以及当 $S[t]$ 高于某一门限时产生小于 $S[t]$ 的参考，因此减少了音频感觉响度的动态范围。不论这种函数的形式如何，都执行前面所述的迭代以计算 $G[t]$ ，以便

$$\Psi_E\{G^2[n]\bar{E}[m,n]\}=S_{ref}[n] \quad (31)$$

然后，可以如上所述或者通过一些其他适当的技术来平滑匹配增益 $G[t]$ 以获得期望的感觉效果。最后，可以引入音频信号 201 和 VCA 方块 236 之间的延迟 240 补偿计算平滑增益中的任何等待时间。这种延迟还可以在图 3 和 11 的结构中提供。

图 3 结构的增益控制信号 $G[t]$ 和图 11 结构的平滑增益控制信号 $G_p[t_1, t_2]$ 在大量应用中是有用的, 例如包括, 感觉响度在不同声道上变化的广播电视或卫星电台。在这些环境中, 本发明的设备或方法可以比较来自每一声道的音频信号与参考响度级 (或参考信号的响度)。
5 操作员或自动化设备可以使用这种增益调节每一声道的响度。因此所有声道将会具有基本上相同的感觉响度。图 13 示出了这种结构的一个实例, 其中来自多个电视或音频声道的音频, 1 到 N, 被施加于根据如图 3 或 11 所示本发明方面的处理或设备 250、252 的各种输入 201。为每一处理或设备 250、252 应用相同的参考响度级, 以在每一输出
10 236 产生经过响度调节的第 1 声道到第 N 声道音频。

测量和增益调节技术还可以应用于监控输入音频资料, 执行识别主要包含人语音信号的音频内容处理, 以及计算增益, 以便该语音信号基本上匹配前面定义的参考级的实时测量设备。在 2002 年 8 月 30 日申请的序列号为 10/233073 的美国专利申请以及 2004 年 3 月 4 日公布
15 的美国专利申请公开 US2004/0044525A1 中, 阐述了用于识别音频资料中语音的适当技术。所述申请在此全部包含引作参考。由于大声音频内容带来的听众烦恼倾向于集中于节目资料的语音部分, 因此测量和增益调节方法可以极大地减小在电视、电影和音乐资料中通常使用的音频中的烦恼级差。

20 实现

本发明可以以硬件或软件, 或二者的组合 (例如, 可编程逻辑阵列) 实现。除非特别规定, 所包括的作为本发明一部分的算法本质上并不与任何特定的计算机或其他设备有关。特别是, 通过根据本发明的教导写入程序可以使用各种通用机器, 或者可以更方便地构造更专
25 门的设备 (例如, 集成电路) 来执行所需的方法步骤。因此, 可以在一个或更多可编程计算机系统上执行的一个或更多计算机程序中实现本发明, 每一系统包括至少一个处理器、至少一个数据存储系统 (包括易失性和非易失性存储器和/或存储单元)、至少一个输入设备或端口、以及至少一个输出设备或端口。应用程序代码给输入数据来执行

这里所述的功能并产生输出信息。该输出信息以已知的方式应用于一个或更多输出设备。

每一这种程序可以用任何想要的计算机语言（包括机器、汇编或高级程序、逻辑、或面向对象的程序设计语言）与计算机系统通信。

5 在任何情况下，所述语言可以是编辑或解释的语言。

每一这种计算机程序优选地存储或下载到通用或专用可编程计算机可读的存储媒体或设备（例如，固体存储器或媒体，或磁或光媒体）上，以当计算机系统读取存储媒体或设备执行这里描述的程序时，配置和操作所述计算机。本发明的系统还可以认为将其实现为配置有计算机程序的计算机可读存储媒体，其中如此配置的存储媒体使
10 计算机系统以特定和预定的方式操作来执行这里描述的功能。

已经描述了本发明的大量实施例。无论如何，将会明白可以作出各种修改而不背离本发明的精神和范围。例如，可以独立地排序上述的一些步骤，因此能够以不同于所述的顺序执行这些步骤。相应地，
15 其他实施例都位于以下权利要求书的范围内。

图1

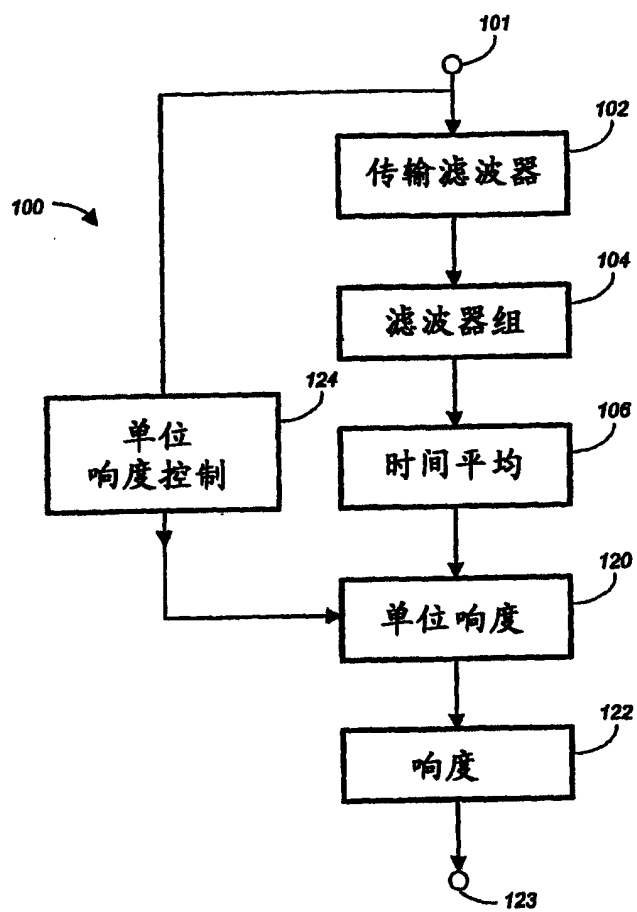


图2

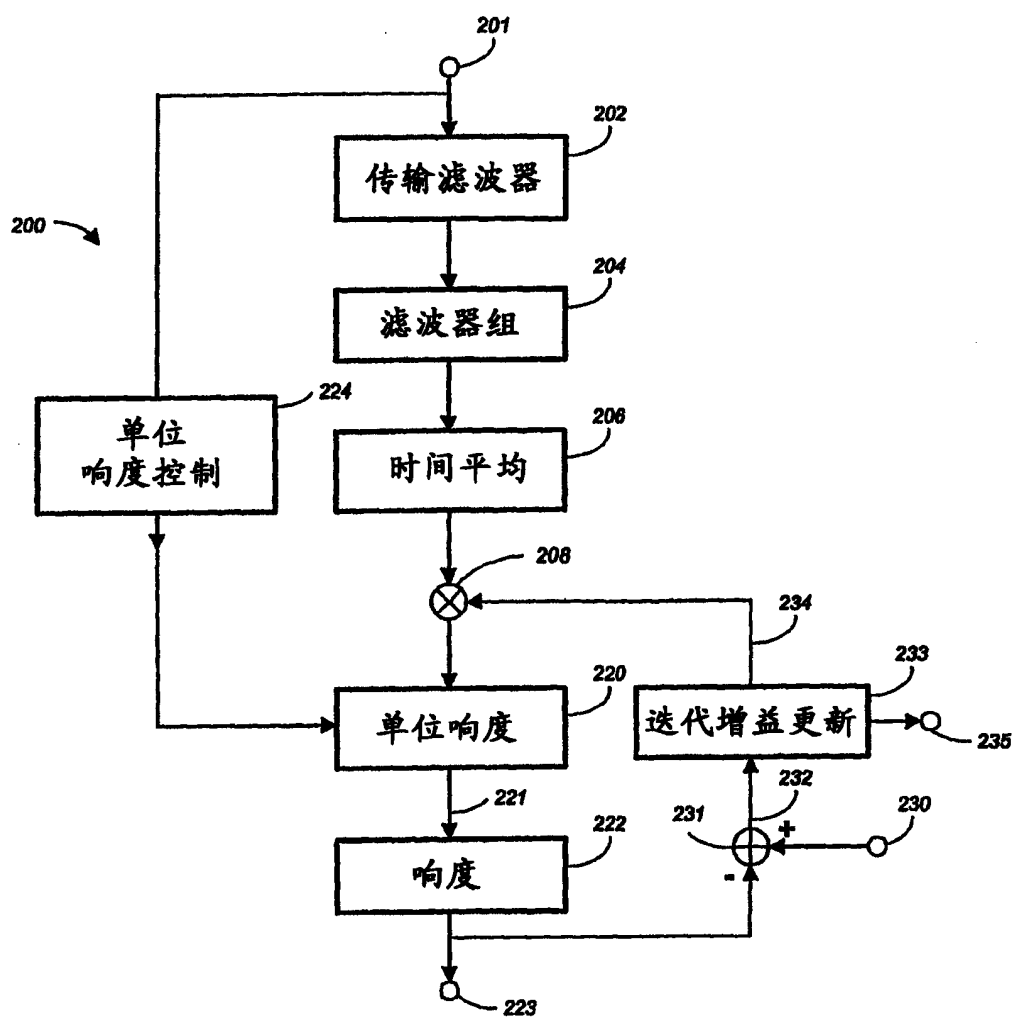


图3

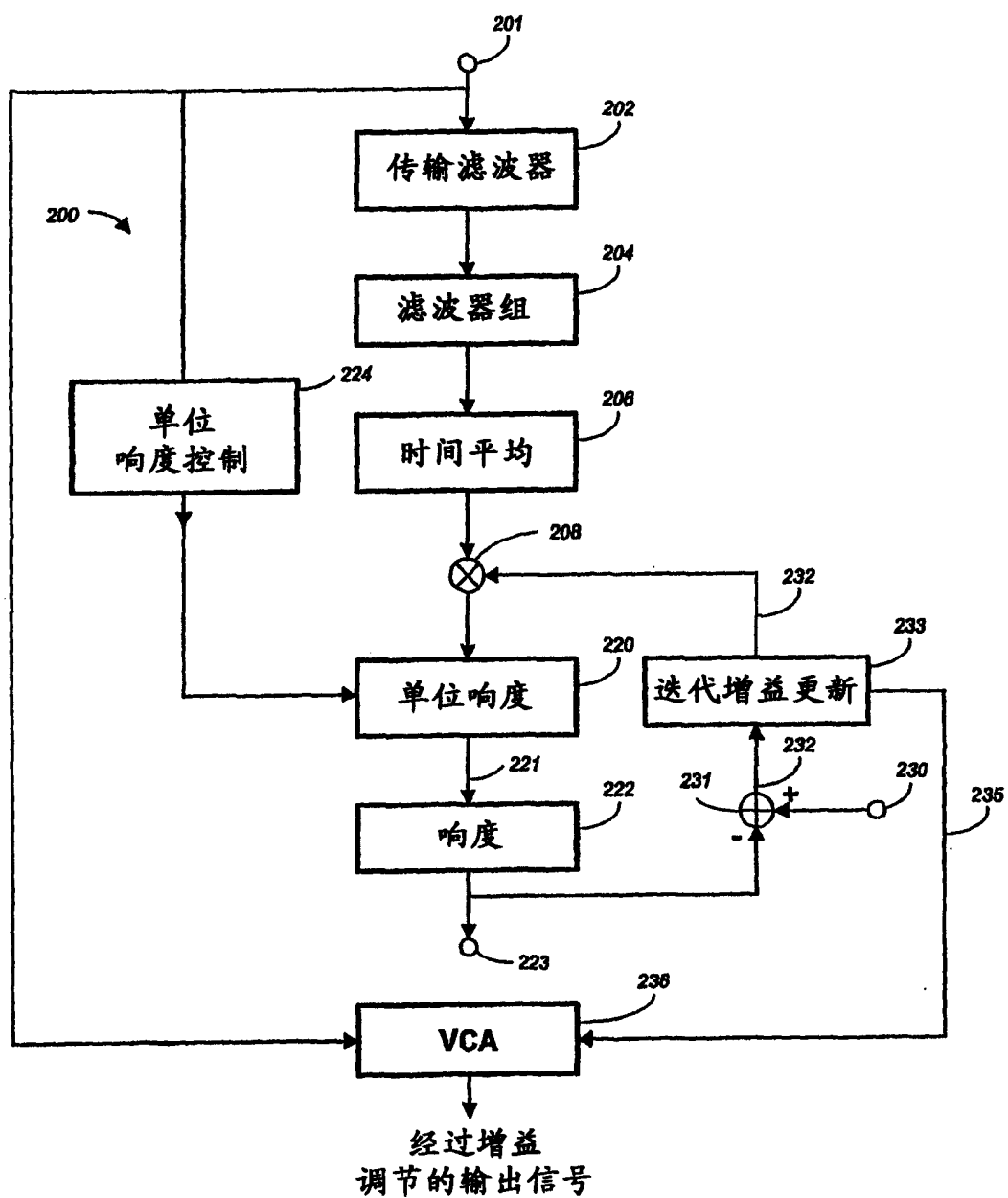


图 4

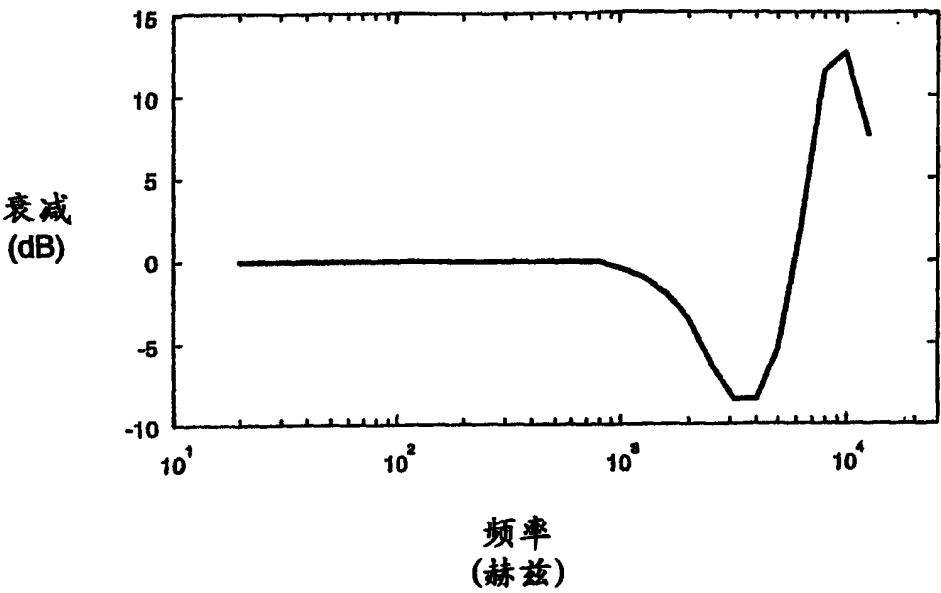


图 5

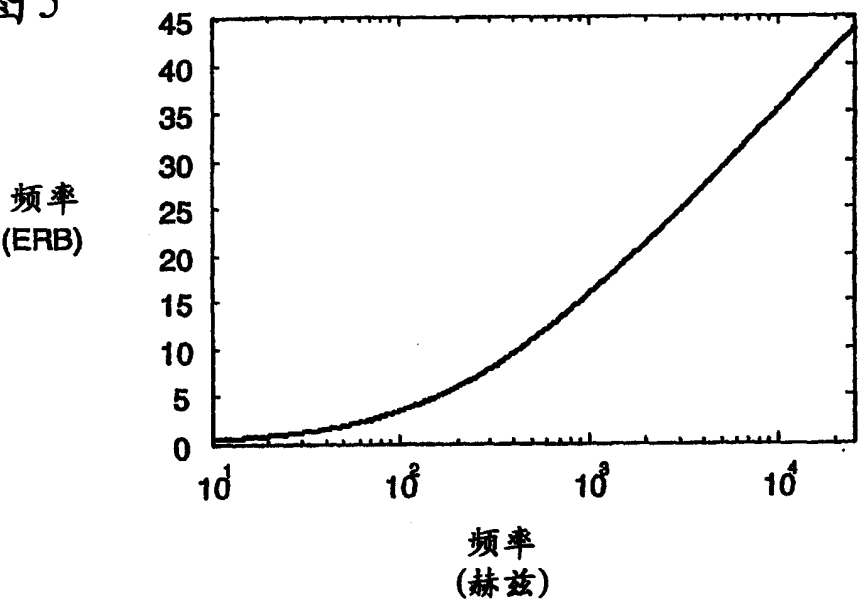


图 6

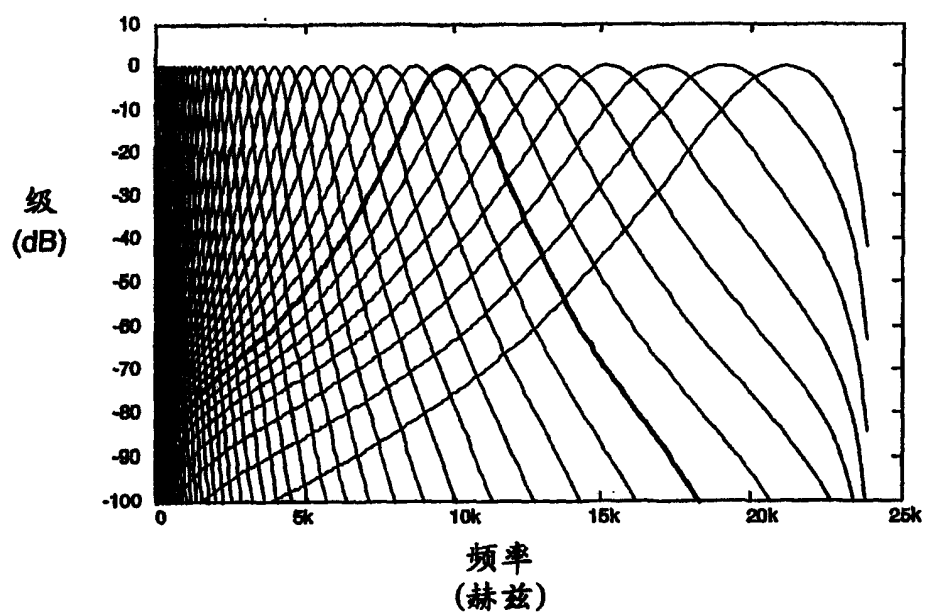


图 7

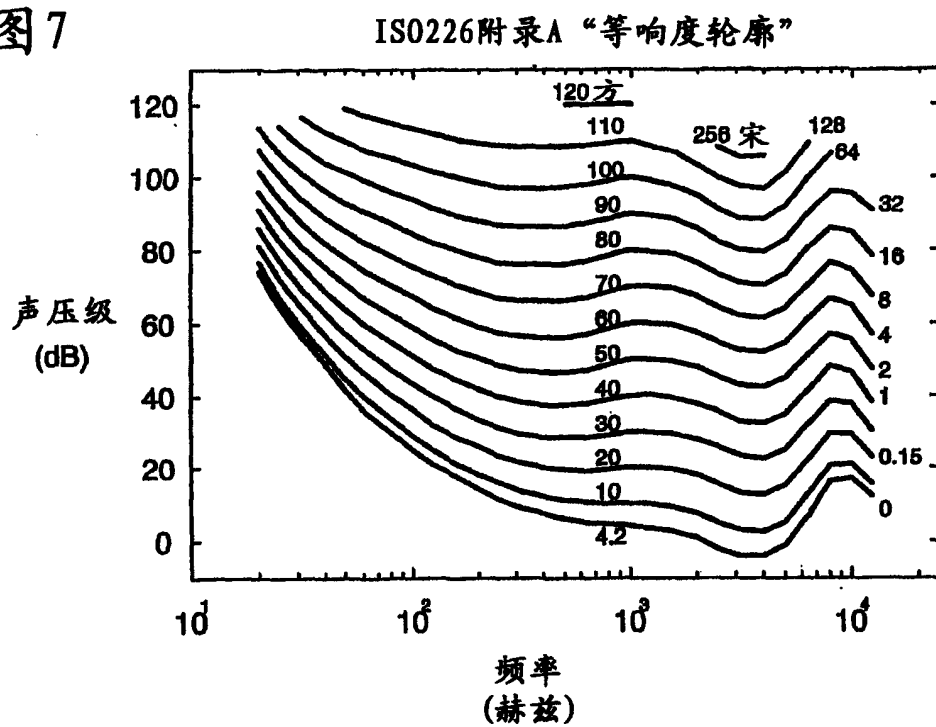


图 8

ISO226附录A"等响度轮廓"
(通过传输滤波器规格化的)

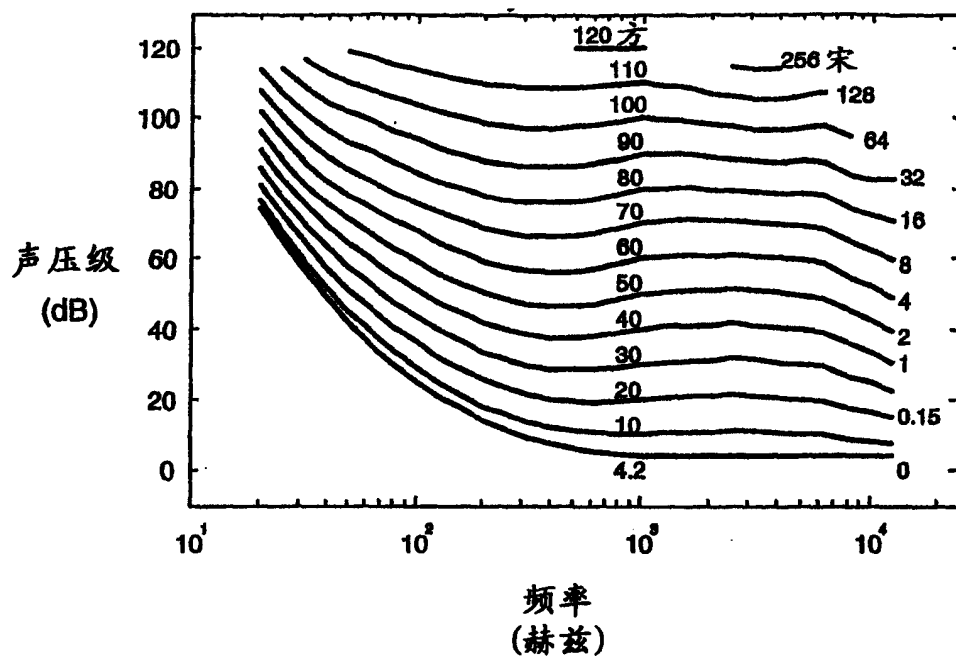


图 9

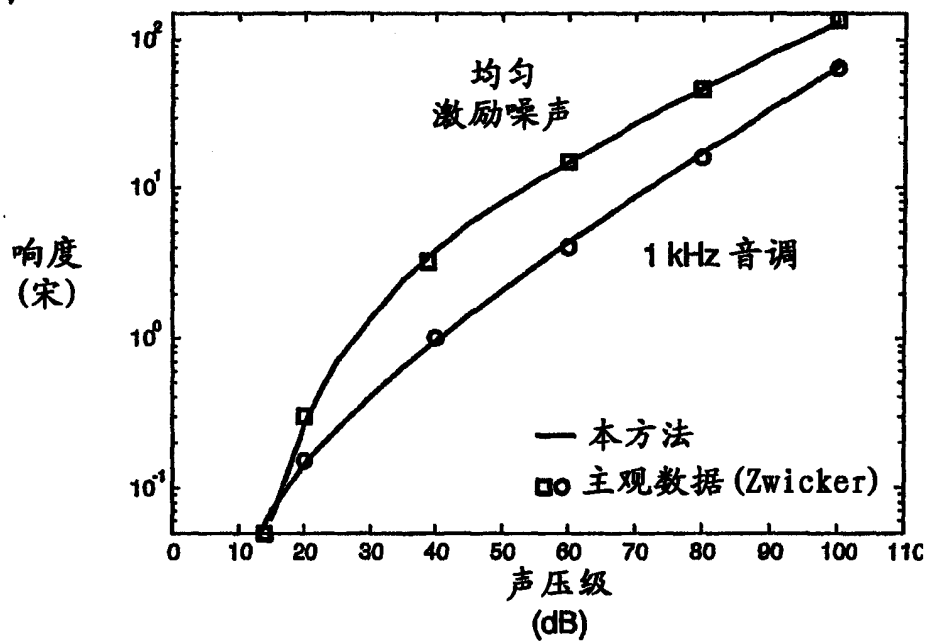


图 10

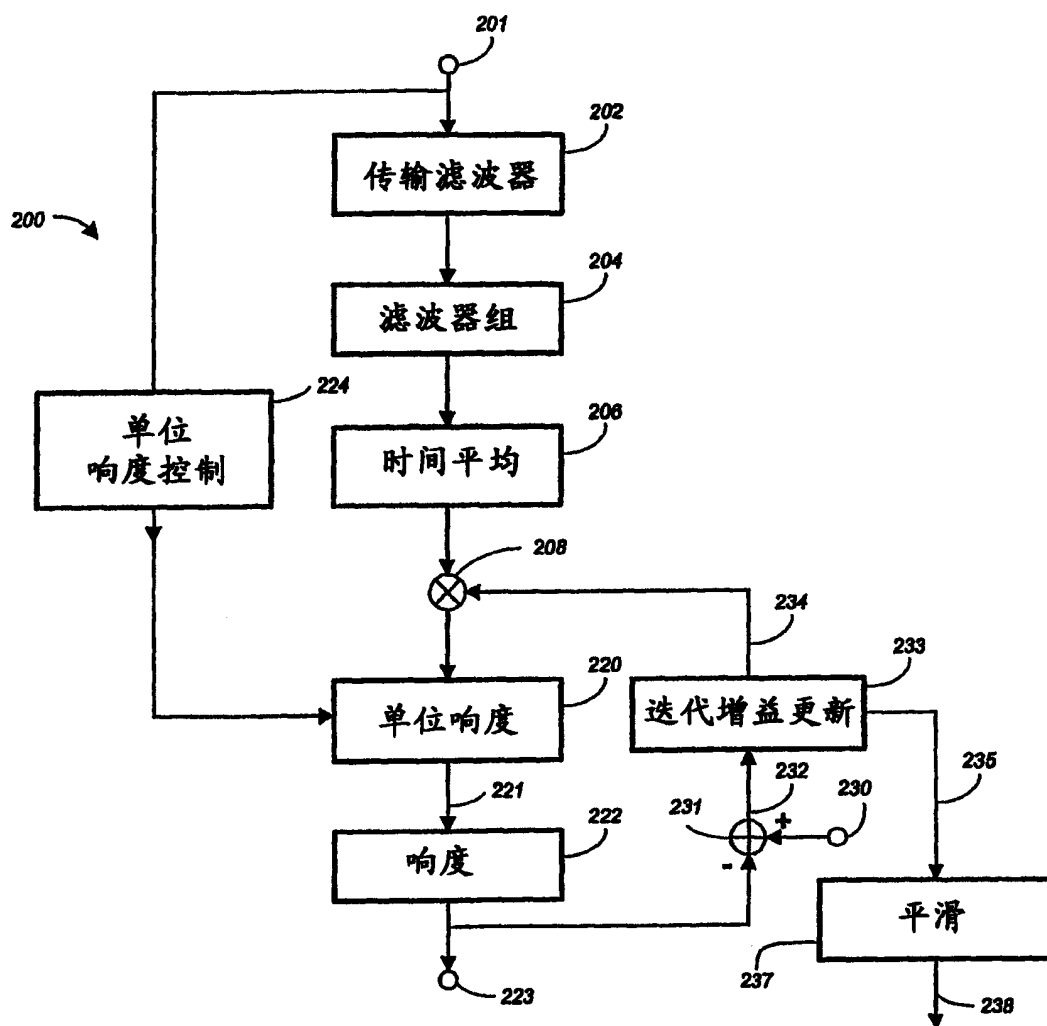


图 11

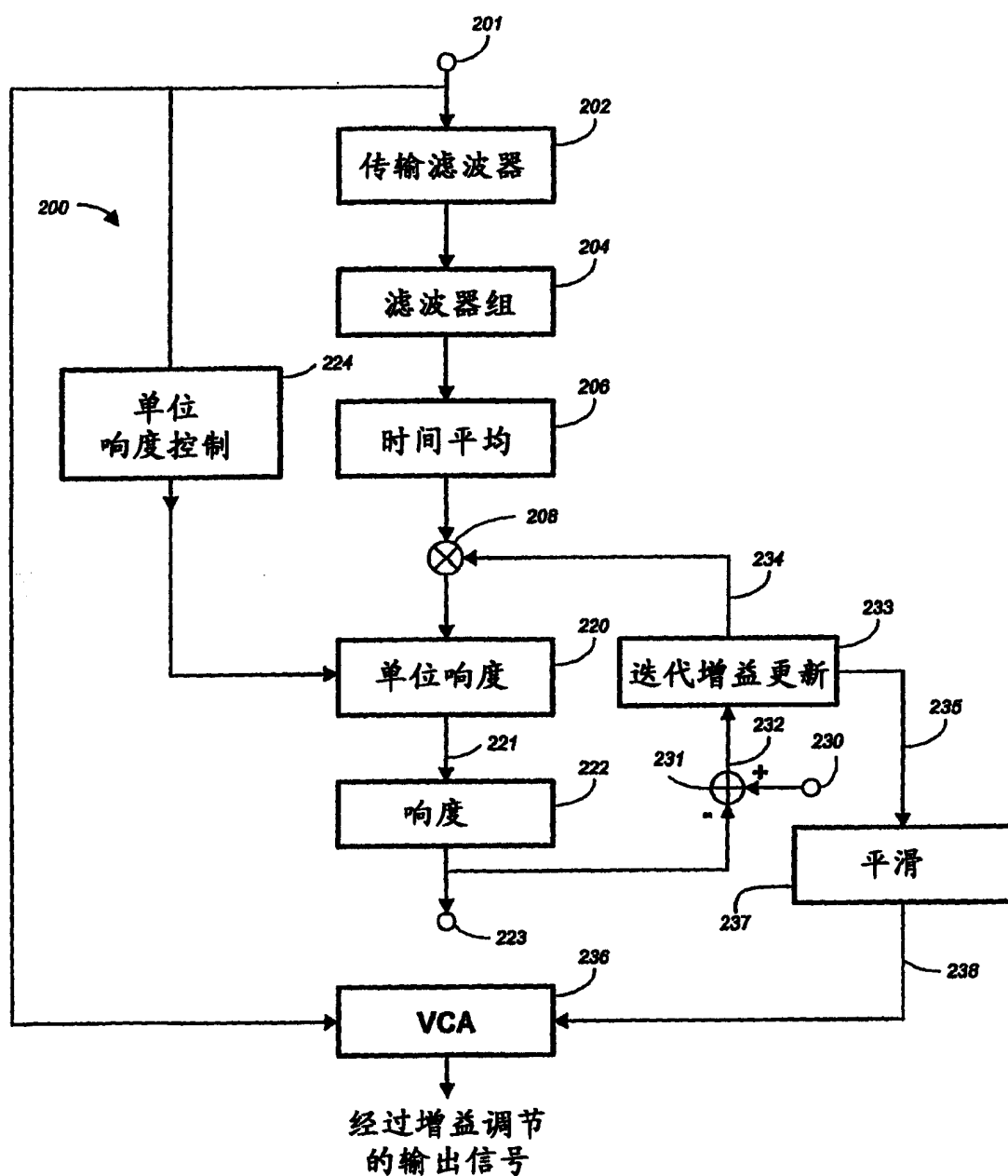


图12

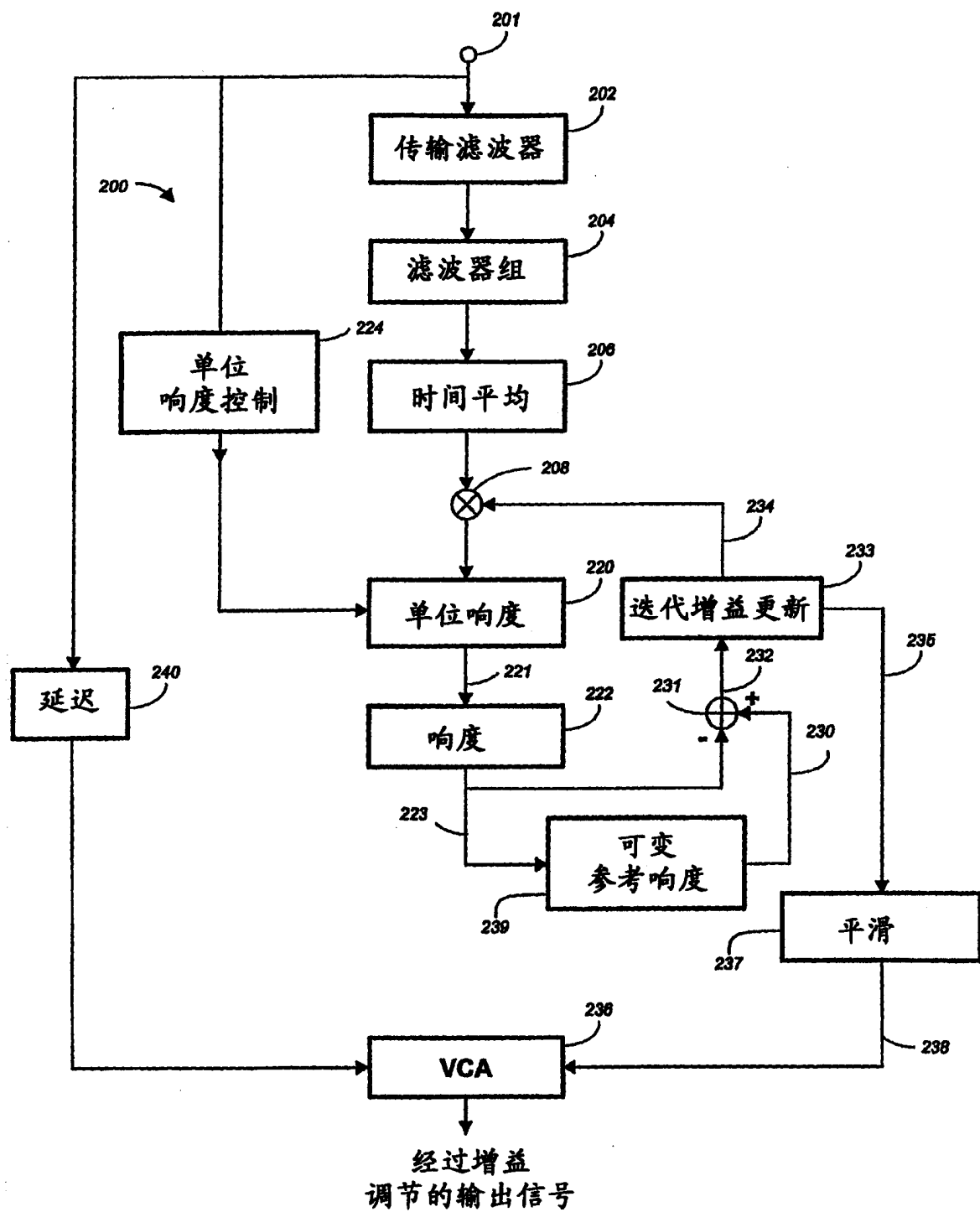


图13

