

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G10L 21/02 (2006.01)



[12] 发明专利说明书

专利号 ZL 02826919.5

[45] 授权公告日 2007 年 6 月 6 日

[11] 授权公告号 CN 1320522C

[22] 申请日 2002.12.9 [21] 申请号 02826919.5

[30] 优先权

[32] 2002.1.9 [33] EP [31] 02075081.6

[86] 国际申请 PCT/IB2002/005295 2002.12.9

[87] 国际公布 WO2003/058607 英 2003.7.17

[85] 进入国家阶段日期 2004.7.8

[73] 专利权人 皇家飞利浦电子股份有限公司

地址 荷兰艾恩德霍芬

[72] 发明人 R·黑布-昂巴赫 C·P·扬斯

D·A·C·M·罗弗斯

[56] 参考文献

EP 0856833 A 1999.3.24

WO 0207151 A 2002.4.4

ENHANCEMENT OF SPEECH DEGRADED-
BY COHERENT AND INCOHERENT NOISE US-
ING ACROSS - SPECTR - BOUQUIN.
JEANNES LE R ET AL, IEEE TRANSACTIONS
ON SPEECH AND AUDIO PROCESSING, No.
5 1997

审查员 刘红梅

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 程天正 张志醒

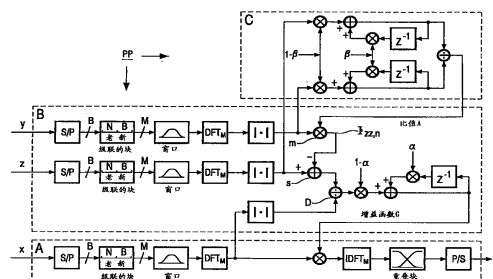
权利要求书 2 页 说明书 6 页 附图 3 页

[54] 发明名称

具有与频谱功率比值相关的处理器的音频增强系统

[57] 摘要

描述了音频增强系统，它包括用于失真的想要的信号和至少一个参考信号的音频信号输入端，以及被耦合到话筒阵列的频谱处理器，它利用该至少一个参考信号作为对于想要的信号的失真的估值来处理失真的想要的信号。频谱处理器被安排来修正所述处理，以使得对于失真的估值取决于 A 乘以该参考信号的频谱功率，其中 A 是在失真的想要的信号的失真的时间平均的频谱功率与参考信号的时间平均的频谱功率之间的比值。被包括在失真估值中的比值 A 对频率的依赖性导致改进的音频增强系统，它更好地适用于干扰信号与想要的信号中的失真之间的关系是事先不知道的情形(诸如在汽车环境下)。



1. 音频增强系统(1), 包括用于一个失真的想要的信号(z, r)和至少一个参考信号(y)的音频信号(z, y, r)输入端, 以及被耦合到音频信号(z, y, r)输入端的频谱处理器(PP), 它藉助于作为对想要的信号(z, r)的失真的估值的该至少一个参考信号(y)来处理失真的想要的信号(z, r), 该系统的特征在于, 频谱处理器(PP)被配置来修正所述处理, 以使得该对于失真的估值的 A 乘以该至少一个参考信号(y)的频谱功率的函数, 其中 A 是想要的信号的失真的时间平均的频谱功率与该至少一个参考信号(y)的时间平均的频谱功率之间的比值。

2. 按照权利要求 1 的音频增强系统(1), 其特征不在于, 对于该失真的估值至少部分地正比于 A 乘以该至少一个参考信号(y)的频谱功率。

3. 按照权利要求 1 的音频增强系统(1), 其特征不在于, 对于该失真的估值至少部分地依赖于失真的想要的信号(z, r)的信号噪声比。

4. 按照权利要求 1-3 之一的音频增强系统(1), 其特征不在于, 各个频谱功率由所涉及的频谱功率的某个正函数来规定, 这些频谱功率属于由频谱幅度、平方的频谱幅度、功率谱密度和 Mel 尺度平滑的谱密度构成的一个组。

5. 按照权利要求 1-3 之一的音频增强系统(1), 其特征不在于, 各个频谱功率由所涉及的频谱功率的某个正函数来规定, 这些频谱功率属于由频谱幅度、平方的频谱幅度、功率谱密度和 Mel 尺度平滑的谱密度构成的一个组, 其中比值 A 是根据在想要的信号不存在的期间所获取的数据计算的。

6. 按照权利要求 5 的音频增强系统(1), 其特征不在于, 语音增强系统(1)包括被耦合到频谱处理器(PP)的语音活动性检测器(DET)。

7. 按照权利要求 1-3 之一的音频增强系统(1), 其特征不在于, 音频增强系统(1)包括被耦合到频谱处理器(PP)的自适应话筒滤波器装置(3)。

8. 按照权利要求 1-3 之一的音频增强系统, 其特征不在于, 各个频谱功率由所涉及的频谱功率的某个正函数来规定, 这些频谱功率属于由频谱幅度、平方的频谱幅度、功率谱密度和 Mel 尺度平滑的谱密度

构成的一个组，其中比值 A 是根据在想要的信号不存在的期间所获取的数据计算的，并且其中该音频增强系统 (1) 包括一个被耦合到该频谱处理器 (PP) 的语音活动性检测器 (DET)。

9. 按照权利要求 1-3 之一的音频增强系统 (1)，其特征在于，音频增强系统 (1) 包括一个或多个扬声器 (6) 以及被耦合在该至少一个扬声器 (6) 与频谱处理器 (PP) 之间的回声抵销滤波器装置 (7)。

10. 按照权利要求 1-3 之一的音频增强系统，该系统被用于属于由免提通信设备、移动电话和话音控制系统构成的一个组的通信系统。

11. 用于改进失真的想要的信号 (z, r) 的方法，该信号被按频谱处理，由此至少一个参考信号 (y) 用作为对于想要的信号的失真的估值，该方法的特征在于，频谱处理是这样执行的，以使得对于失真的估值与 A 乘以该至少一个参考信号 (y) 的频谱功率的函数相关，其中 A 是想要的信号的失真的时间平均的频谱功率与该至少一个参考信号 (y) 的时间平均的频谱功率之间的比值。

具有与频谱功率比值相关的处理器的音频增强系统

本发明涉及音频增强系统，该系统包括用于失真的想要的信号和至少一个参考信号的音频信号输入端，以及一个被耦合到音频信号输入端的频谱处理器，用于借助于该至少一个参考信号来处理该失真的想要的信号，该至少一个参考信号用作为对于想要的信号的失真的估值。

本发明还涉及用于改进失真的想要的信号的方法，该信号经频谱处理，由此至少一个参考信号用作为对于该想要的信号的失真的估值。

被体现为用于抑制干扰分量（诸如失真的噪声）的装置的这样的音频增强系统可从 W0 97/45995 获知。已知的系统包括多个被耦合到音频信号输入端的话筒。这些话筒包括用于失真的想要的信号的一个主话筒，和用于接收干扰信号的一个或多个参考话筒。系统还包括被体现为耦合到话筒的信号处理装置的频谱处理器。在这个信号处理装置中，把干扰信号从失真的信号中按频谱减去，以便在它的输出端呈现包括已减小了干扰噪声分量的输出信号。

已知的音频增强系统的缺点在于，在干扰信号与想要的信号中的失真之间的关系是事先不知道的情形下，诸如例如在汽车环境下，它的干扰抵销能力是不够的。

所以，本发明的目的是提供具有扩展应用领域的、改进的音频增强系统和相关的方法。

为此，按照本发明的音频增强系统的特征在于，频谱处理器被配置来修正所述处理，以使得对于失真的估算是 A 乘以该至少一个参考信号的频谱功率的函数，其中 A 是想要的信号的失真的时间平均的频谱功率与该至少一个参考信号的时间平均的频谱功率之间的比值。

类似地，按照本发明的方法的特征在于，执行频谱处理，以使得对于失真的估值取决于 A 乘以该至少一个参考信号的频谱功率的函数，其中 A 是失真的想要的信号的失真的时间平均的频谱功率与该至少一个参考信号的时间平均的频谱功率之间的比值。

本发明人发现，如上规定的比值在该至少一个参考信号与现有技

术装置中未计及的失真的想要的信号中的失真的估值之间的关系中引入一个有利的频率函数。由于功能相关性，音频增强系统较好地适用于例如在工厂或车辆（诸如汽车），飞机等等中可靠的应用，因为比值项 A 能够更精确地描述对于失真的估值，而不需要有关在干扰信号与想要的信号中的失真之间的关系的先验知识。这改进了失真的抵销，特别是在其中一个或多个参考信号包括失真，诸如噪声，回声，竞争的语音，想要的语音的交混回响等等的情形下。有利地，在其中某些参考信号是可提供的任何情形下，可以计算与频率相关的失真估值。

另一个优点在于，不需要对于各个失真分量（诸如基底噪声或回声拖尾）清楚地估值，而可以容易地实现这些分量的组合技术，如果需要的话。在不存在良好的估值技术的失真的情形下，诸如话筒波束成形（microphone beam forming）应用中，这是特别有利的。此外，在按照本发明的音频增强系统中在很大程度上不再需要对直观推断的熟知的过相减因子（over subtraction factor）的协调。

按照本发明的音频增强系统的实施例的特征在于，对于失真的估值至少部分地正比于 A 乘以该至少一个参考信号的频谱功率。

该比例可以由过相减因子表示，该过相减因子可以小于、等于或大于 1。通过该过相减因子，可以影响失真抑制的量。这样，可以在失真抑制量与处理器的输出信号的感知品质之间作出折衷。

按照本发明的音频增强系统的另一个精心设计的实施例的特征在于，对于失真的估值至少部分地与失真的想要的信号的信号噪声比相关。

在这个实施例中以及在上面提到的实施例中，其中出现相关的部分可能涉及例如现有的频谱的低频或高频部分。

按照本发明的音频增强系统的另一个实施例的特征在于，各个频谱功率由所述的频谱功率的某个正函数规定，诸如频谱幅度，平方的频谱幅度，功率谱密度或经 Mel 尺度平滑的谱密度。

通常，想要的信号的失真的估值可以由某个正函数表示，例如，按照信号功率或信号能量，而它们又由以上的频谱单元之一规定。

按照本发明的音频增强系统的又一个实施例的特征在于，比值 A 是根据在不存在想要的信号期间获取的数据计算的。

在不存在想要的信号期间（该信号通常是语音信号），失真的想要的语音信号代表在失真的想要的语音信号中的失真。所以，比值 A 可以在不存在想要的语音时作为失真的想要的语音信号的时间平均的频谱功率与该至少一个参考信号的时间平均的功率之间的比值而被测量。通常， A 的数值至少将在想要的语音信号重新出现后的某一时间期间被使用。

按照本发明的音频增强系统的另一个示例性简单的实施例的特征在于，语音增强系统包括被耦合到频谱处理器的语音活动性检测器。

按照本发明的音频增强系统的另一个实施例的特征在于，音频增强系统包括被耦合到频谱处理器的自适应话筒滤波器装置。

这些自适应话筒滤波器装置可以与音频增强系统相组合，以便提供足够的频谱处理从而抵销失真。

按照本发明的音频增强系统的又一个实施例的特征在于，音频增强系统包括一个或多个扬声器以及被耦合在该至少一个扬声器与频谱处理器之间的回声抵销滤波器装置。

有利地，这个实施例除了可能的话筒信号处理以外还组合了回声抵销、扬声器信号处理和失真抵销。

现在在参照附图的同时将进一步阐述按照本发明的音频增强系统和方法以及它们的附加优点，其中藉助于相同的参考数字来表示相似的元件。

在图上：

图 1 显示按照本发明的音频增强系统的基本图；

图 2a 和 2b 分别显示带有和不带有话筒自适应滤波器装置的、图 1 的音频增强系统的实施例；

图 3 显示具有话筒波束成形器的、按照本发明的音频增强系统的再一个实施例；

图 4 显示具有回声抵销器、按照本发明的音频增强系统的又一个实施例；以及

图 5 显示图 1 的音频增强系统的详细的实施例。

图 1 显示由后处理器 PP 实现的音频增强系统 1 的基本图，图上显示频域信号 z , y , r , 和 q 。这些频域信号在处理器 PP（在图 5 上示意地表示为 A 和 B）藉助于离散富立叶变换（FT）（例如短时间 DFT，简称

为 STFT) 被按块地按频谱计算。这个 STFT 是时间和频率的函数, 它由宗量 kB 和 lw_0 表示。 k 表示离散时间帧指数, B 表示帧移位, l 表示(离散)频率指数, 以及 w_0 表示基本频率间隔。输入信号 z 表示失真的想要的信号。它包括总的具有语音的形式的想要的信号和失真, 诸如噪声, 回声, 竞争语音, 想要的语音的交混回响等的总和。信号 y 表示参考信号, 通过它得出在失真的想要的信号 z 中失真的估值。信号 z 和 y 可以起源于一个或多个话筒 2, 如图 2a, 2b, 3 和 4 所示。在多话筒音频增强系统 1 中, 有两个或多个分开的话筒 2, 以便从一个或多个话筒得出参考信号。

音频增强系统 1 在图 2a 所示的情形下可包括自适应话筒滤波器装置 3, 而图 2b 显示其中系统 1 缺乏自适应滤波器装置的情形。这两种情形在图 1 上藉助于原理性的开关 S 而被组合, 该开关 S 可以断开或接通。如果开关 S 接通, 则信号 y 被从 z 中减去, 展现信号 r , 该减法在相减单元 4 中进行, 如果存在滤波器装置 3 的话。如果开关 S 被断开, 则该情形反映图 2b 的实施例。信号 z 和 y 以及可能 r 被馈送到频谱后处理器 PP, 以便利用参考信号 y 根据频谱来处理失真的想要的信号 z 或 r 。来自后处理器 PP 的信号 q 是输出信号, 它实际上是无失真的。它的运行将在后面解释。

现在参照图 3, 该图显示具有几个话筒 2 的音频增强系统 1。这里, 自适应滤波器装置由通用边瓣抵销器 (GSC) 3 来实现, 它被耦合到话筒 2 和后处理器 PP。在 GSC 中, 使用了由各个转移函数 $f_1(w)$, $f_2(w)$ 和 $f_3(w)$ 表示的滤波器与和波束成形器 5-1, 以便从各个话筒阵列信号 u_1 , u_2 , 和 u_3 的线性组合得出失真的想要的信号 z 。参考信号 y 是一个通过阻塞矩阵 (blocking matrix) $B(w)$ 把各个阵列信号投影到与想要的信号正交的子空间而从这些信号得出的。理想地, 矩阵 $B(w)$ 的输出信号 x_1 和 x_2 不包含想要的语音而只包含失真。由 $w_1(w)$ 和 $w_2(w)$ 表示的多信道自适应滤波器 5-2 被利用来得出参考信号 y , 在相加后, 该信号 y 然后从信号 z 中被减去, 正如前面解释的。

图 4 显示音频增强系统 1 的实施例, 这里它除了具有自适应回声抵销器滤波器装置 7 以外还具有一个话筒 2, 以及在这种情形下一个扬声器 6。按本身已知的方式, 自适应滤波器 7 在它的输出端生成回声复制信号, 它在通过滤波器 7 对远端信号自适应滤波而得到的参考信号 y

中被反映。当然，一个或多个话筒和或扬声器可包括在音频增强系统 1 的可能的实施例中。音频增强系统 1 可被包括在一个系统中，特别是通信系统中，例如免提通信设备（诸如移动电话）或话音控制系统。

现在参照图 5 解释频谱后处理器 PP 的运行。原理上，处理器 PP 起到对于由如上所述的离散富立叶变换（DFT）生成的后续频率仓（frequency bin）的可控增益函数的作用。这个增益函数被施加到失真的想要的语音 r ，而信号 r 的相位保持不变。这些信号的每一个都要受到以下处理步骤的处理。在串行-并行（S/P）转换后，在大小为 B 的块中进行块处理。每个新的块 B 被附着到以前的块，导致级联的块。这些块重叠起来，并被称为具有大小为 M 的帧，然后，在例如取 FFT 系数的幅度或平方的幅度以后它们要被开窗口，并被具有大小为 M 的 DFT 进行变换。可以使用任何其他可能的频谱功率的正函数。

为了音频增强的良好的性能，增益函数的类型和存在于输入信号（这里由 r 表示）中的失真的估值是重要的。取决于所涉及的最佳化准则，可以采用各种增益函数。例子包括频谱相减，Wiener 滤波或例如基于所涉及的信号的频谱幅度、平方的频谱幅度、功率谱密度、或 Mel 尺度平滑的谱密度的最小均方误差（MMSE）估值或对数 MMSE 估值。这些技术可以与以上说明的应用相组合，以用于具有一个或多个话筒和或扬声器的音频增强系统。

在 Wiener 型滤波器的情形下，增益函数具有以下形式：

$$G(kB, l w_0) = \{P_{zz}(kB, l w_0) - P_{zz,n}(kB, l w_0)\} / P_n(kB, l w_0) \quad (1)$$

其中 $P_{zz}(kB, l w_0)$ 和 $P_{rr}(kB, l w_0)$ 分别是对于信号 z 和 r 的功率分布的度量。如果例如短时间功率谱密度（PSD）被取为对于频谱功率分布的度量，则下式成立：

$$P_{zz}(kB, l w_0) = |z(kB, l w_0)|^2$$

在公式（1）中， $P_{zz,n}(kB, l w_0)$ 是在信号 z 中失真的 PSD，它通常是未知的，所以必须被估值。建议从下式得出估值 p ：

$$P_{zz,n}(kB, l w_0) = A(kB, l w_0) * P_{yy}(kB, l w_0) \quad (2)$$

其中比值项：

$$A(kB, l w_0) = P_{zz}(kB, l w_0) / P_{yy}(kB, l w_0). \quad (3)$$

这里， $P_{zz}(kB, lw_0)$ 是失真的想要的信号 z 的失真的时间平均频谱功率（在不存在想要的信号（如语音）期间测量的），以及 $P_{yy}(kB, lw_0)$ 是参考信号 y 的时间平均频谱功率。作为对于频谱功率的正的测量值，例如可以取所涉及的信号的频谱幅度或大小、平方的频谱幅度、功率谱密度、或 Mel 尺度平滑的谱密度。

接着，对于 Wiener 型滤波器的公式 (1) 的增益函数 $G(kB, lw_0)$ 在图 5 的块 B 的其余部分被实施，而在块 C 中比值项 A 按照公式 (3) 被实施。在比值项 A 的分子和分母中的频谱是通过在块 C 中实施的一阶递归中用平滑常数 β 来平滑功率谱而得到的。递归实施方案包括乘法器 $\times$ ，加法器 $+$ ，延时线 z^{-1} ，和除法器 $/$ 。它们按如图所示被耦合来得出 y 和 z 信号的平滑的 PSD 版本。例如， y 信号谱遵从平滑法则：

$$P_{yy}(kB, lw_0) = \beta P_{yy}((k-1)B, lw_0) + (1-\beta)P_{yy}(kB, lw_0)$$

其中如果在帧 kB 中不存在想要的语音，则平滑常数 β 假设为 0 与 1 之间的数值，而在其他情况，则 $\beta=1$ 。相同的法则应用于 z 频谱。典型地，对于 16ms 的帧移位， $\beta=0.9$ 。任何耦合到处理器 PP 的语音检测器 DET 可用来控制 β 的数值。除法器输出端展现比值 A，如图所示。

在块 B 的其余部分的乘法器 M 中，比值项 A 与 Y 的频谱相乘，以实施公式 (2)，此后最终得到的估值 $p_{zz,n}$ 在减法器 S 中被从信号 z 的频谱中减去，此后该结果在除法器 D 中被除以信号 r 的频谱，以展现在一阶平滑运行中被平滑后的增益函数。这一运算和对信号 y 和 z 的平滑是相似的。对于 16ms 的帧的移位，典型的平滑值 $\alpha=0.6$ 。平滑运行帮助减小音乐音调。在与信号的频谱相乘以后，执行逆 DFT，然后块被重新构建，以及进行从并行到串行的变换，得出想要的输出信号 $q(kB, lw_0)$ 。

虽然以上内容是参照实质上的优选实施例和最好的可能的模式描述的，但应当看到，这些实施例决不被看作为所涉及的系统和方法的限制例子，因为属于附属权利要求的范围内的各种修正，特性和特性的组合现在是在本领域技术人员所能达到的范围内的。

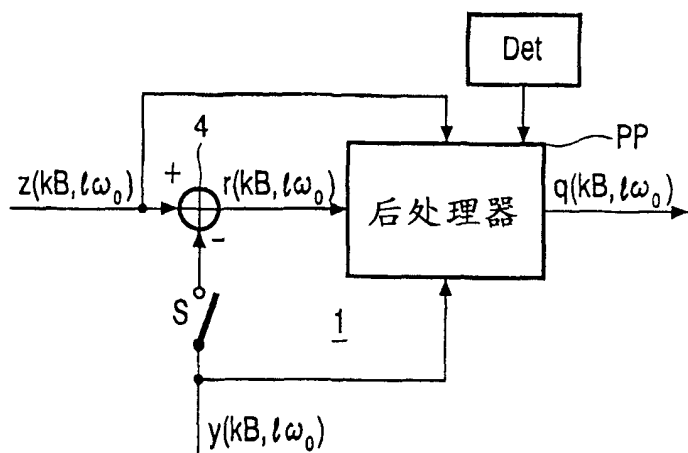


图 1

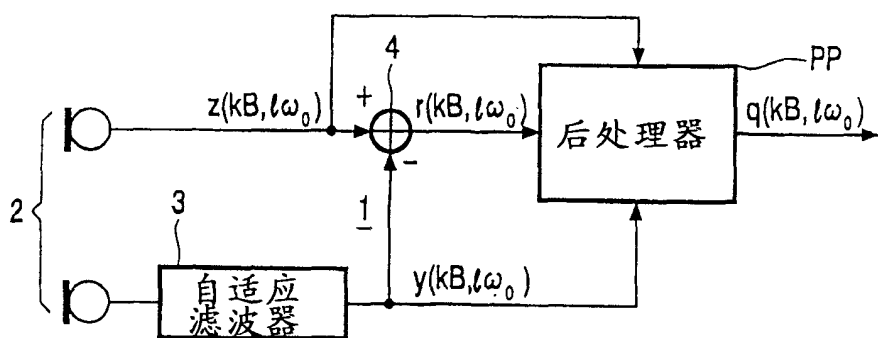


图 2a

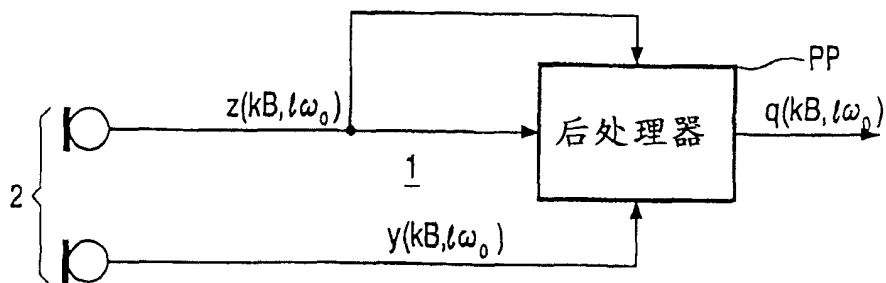


图 2b

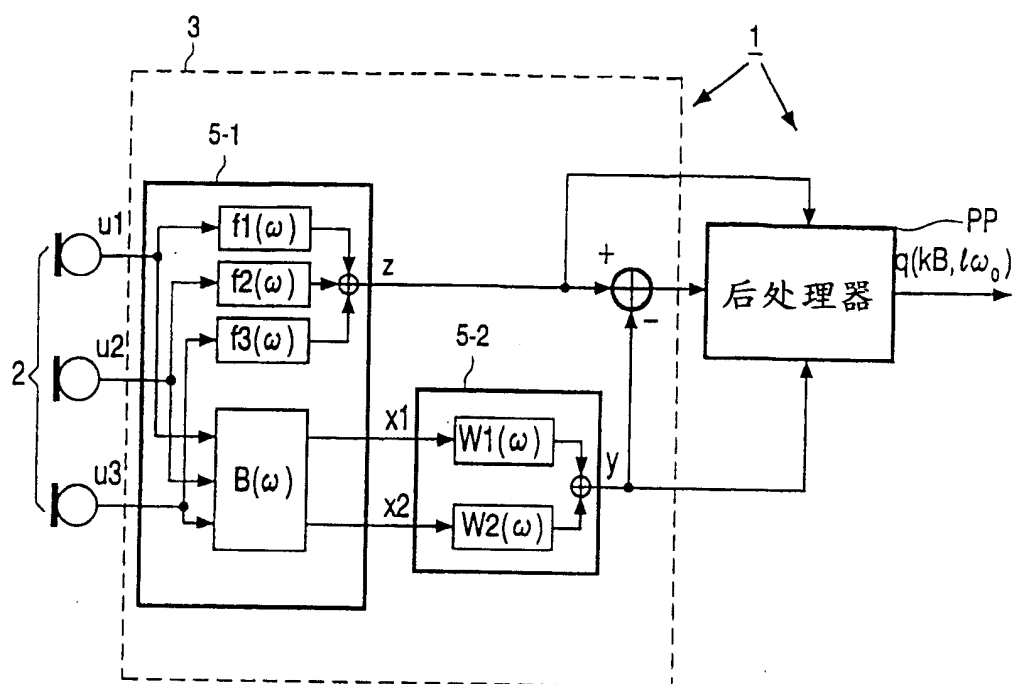


图 3

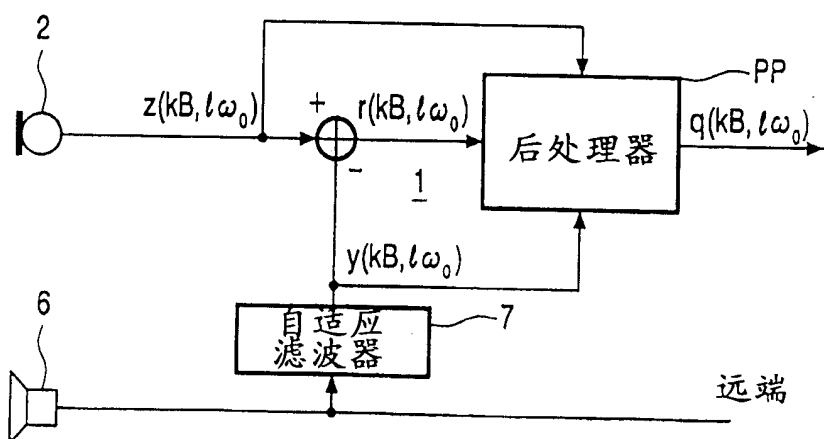


图 4

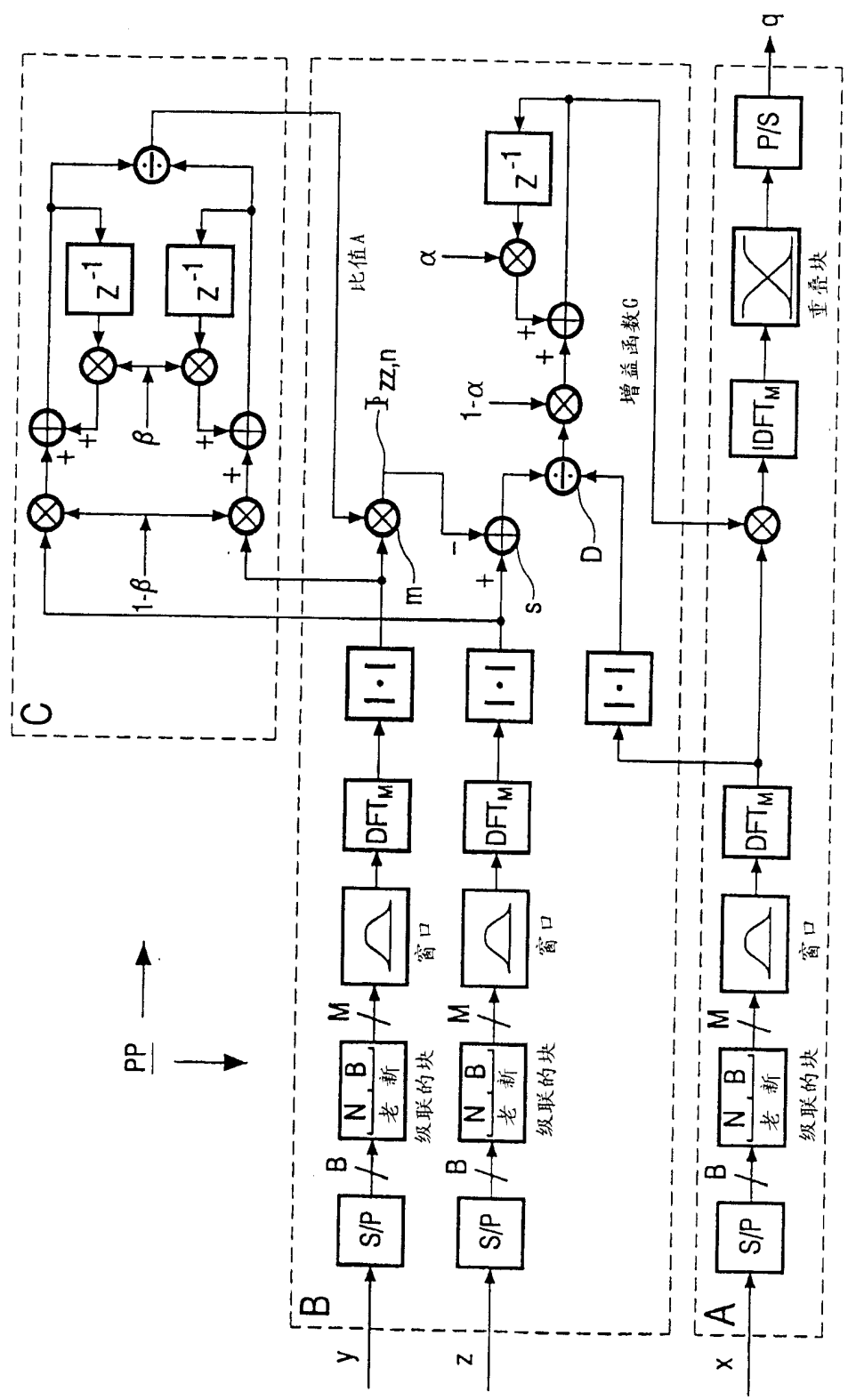


图 5