



[12] 发明专利说明书

[21] ZL 专利号 00815729.4

[45] 授权公告日 2004 年 8 月 11 日

[11] 授权公告号 CN 1161752C

[22] 申请日 2000.11.14 [21] 申请号 00815729.4

[30] 优先权

[32] 1999.11.15 [33] FI [31] 19992453

[86] 国际申请 PCT/FI2000/000996 2000.11.14

[87] 国际公布 WO2001/037254 英 2001.5.25

[85] 进入国家阶段日期 2002.5.15

[71] 专利权人 诺基亚有限公司

地址 芬兰埃斯波

[72] 发明人 B·阿亚德

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

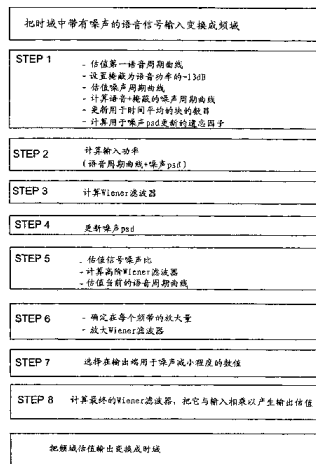
代理人 程天正 张志醒

权利要求书 2 页 说明书 16 页 附图 5 页

[54] 发明名称 噪声抑制器

[57] 摘要

一种抑制包含语音和噪声的信号中的噪声、以便提供噪声被抑制的语音信号的方法。对于噪声进行估值和对于语音连同某些噪声进行估值。被包括在语音连同某些噪声的估值中的噪声电平是可变的,以便把想要的噪声量包括在噪声被抑制的信号中。



I S S N 1 0 0 8 - 4 2 7 4

1. 一种抑制在含有噪声的信号中的噪声以便提供噪声被抑制的信号的方法，其中对噪声进行估值，和对语音连同某些噪声一起进行估值。
- 5 2. 按照权利要求1的方法，其中被包括在语音连同某些噪声一起的估值中的噪声电平是可变的，从而把想要的噪声量包括在噪声被抑制的信号中。
3. 按照权利要求2的方法，其中噪声电平提供可接受的前后关系信息量。
- 10 4. 按照任何前述的权利要求的方法，其中噪声电平低于语音的掩蔽极限，并且因而对于收听者是听不见的。
5. 按照权利要求1到3的任一项的方法，其中噪声电平接近于语音的掩蔽极限，并且因而某些噪声前后关系信息被保留在信号中。
6. 一种产生用于噪声抑制的增益系数的方法，其中该增益系数的
- 15 第一估值是自适应地获得的，这个第一估值被使用来产生噪声估值，后者然后被使用来产生增益函数的第二估值。
7. 按照权利要求6的方法，其中估值的噪声是功率谱密度。
8. 按照权利要求6或7的方法，其中第一估值被使用来更新估值的噪声。
- 20 9. 用于抑制在含有噪声的信号中的噪声以便提供噪声被抑制的信号噪声抑制器，该噪声抑制器包括对噪声进行估值的装置和对语音连同某些噪声一起进行估值的装置。
10. 按照权利要求9的噪声抑制器，其中被包括在语音连同某些噪声的估值中的噪声电平是可变的，从而把想要的噪声量包括在噪声
- 25 被抑制的信号中。
11. 按照权利要求10的噪声抑制器，其中噪声电平提供可接受的前后关系信息量。
12. 按照权利要求9到11的任一项的噪声抑制器，其中噪声电平低于语音的掩蔽极限，并且因而对于收听者是听不见的。
- 30 13. 按照权利要求9到11的任一项的噪声抑制器，其中噪声电平接近于语音的掩蔽极限，并且因而某些噪声前后关系信息被保留在信号中。

14. 噪声抑制器，包括用来产生用于噪声抑制的增益系数的装置，其中增益系数的第一估值是自适应地获得的，这个第一估值被使用来产生噪声估值，后者然后被使用来产生增益函数的第二估值。

5 15. 按照权利要求 14 的噪声抑制器，其中估值的噪声是功率谱密度。

16. 按照权利要求 14 或 15 的噪声抑制器，其中第一估值被使用来更新估值的噪声。

17. 包括按照权利要求 9 到 16 的任一项的噪声抑制器的通信终端。

10 18. 包括按照权利要求 9 到 16 的任一项的噪声抑制器的通信网。

噪声抑制器

5 本发明涉及噪声抑制，以及具体地（但不唯一地）涉及对移动终端获取的语音信号中的噪声抑制。

当通信终端被使用来记录或发送包含语言的语音信号时，不可避免的是，通信终端中的话筒将从讲话人所位于的周围环境中拾取环境噪声或背景噪声。背景噪声降低收听者听见和理解语音的能力，以及在某些情形下，如果噪声电平足够高，它将阻止收听者听见除了背景噪声以外的任何其它声音。另外，这样的背景噪声对于通信终端中或相关的通信网络中的数字信号处理系统（诸如语音编码或语音识别）的性能可能具有负面影响。典型地，在通信终端中引入噪声抑制系统来限制背景噪声的影响。

15 多年来，噪声抑制是熟知的。已经提出了许多不同的方法来达到三个主要目的：

- (i) 抑制噪声而同时保留良好的语音质量；
- (ii) 与处理的噪声性质无关地快速收敛到最优解决方案；以及
- (iii) 改进对于非常低的语音-噪声比(SNR)时语音的可理解度。

一种基于线性最小平均平方误差(MMSE)准则的噪声抑制方法将在下面1进行描述。该方法作用在包含语音信号 $s(t)$ 和噪声信号 $n(t)$ 的带有噪声的语音信号 $x(t)$ 上， $x(t)=s(t)+n(t)$ 。带有噪声的语音信号 $x(t)$ 是在时域中的。通过使用窗函数，把它变换成具有接连的帧号 k 的一系列帧。然后，通过使用快速富立叶变换(FFT)把每个帧变换到频域，以便产生一系列带有噪声的语音帧，其中在频域中带有噪声的语音信号 $X(f, k)$ 包含语音信号 $S(f, k)$ 和噪声信号 $N(f, k)$ ，这样， $X(f, k)=S(f, k)+N(f, k)$ 。在频域中的帧包括多个频率仓(frequency bin) f 。在频域中，MMSE方法包括使得以下的误差函数最小化：

$$\varepsilon^2(f, k) = E \left\{ (S(f, k) - \hat{S}(f, k)) \cdot (S(f, k) - \hat{S}(f, k))^* \right\} \quad (1)$$

其中 $E\{\cdot\}$ 是预期算子, $(*)$ 表示复数共轭, 以及 $\hat{S}(f, k)$ 代表输入语音信号的线性估值。误差 $\varepsilon^2(f, k)$, 由公式 1 定义, 代表在被包含在带有噪声的语音信号内的真正的语音分量与该语音分量的估值 $\hat{S}(f, k)$ (即, 无噪声的语音分量) 之间差值的平方。因此, 使得 $\varepsilon^2(f, k)$ 最小化, 就等价于得到语音分量的最好的可能的估值。 $\hat{S}(f, k)$ 被给出为:

$$\hat{S}(f, k) = G(f, k) \cdot X(f, k) \quad (2)$$

其中 $G(f, k)$ 是增益系数。对于使得每个帧的 $\varepsilon^2(f, k)$ 最小化的相应的解, 可以采取计算增益系数 $G(f, k)$ 的形式, 把 $G(f, k)$ 与该帧的相关的输入频率仓相乘, 以便产生估值的无噪声语音分量 $\hat{S}(f, k)$ 。这个增益系数 (被称为频域 Wiener (维纳) 滤波器) 由以下的比值给出:

$$G(f, k) = \frac{E\{S(f, k) \cdot X^*(f, k)\}}{E\{X(f, k) \cdot X^*(f, k)\}} \quad (3)$$

Wiener 滤波器 $G(f, k)$ 是对于每个帧的每个频率仓 f 生成的。

噪声被抑制的帧然后在方块 14 中被变换回时域, 然后被组合在一起, 以便给出噪声被抑制的语音信号 $\hat{s}(t)$ 。理想地, $\hat{s}(t) = s(t)$ 。

当导出 Wiener 滤波器时, MMSE 方法等价于正交性原理。这个原理规定: 对于每个频率, 输入信号 $X(f, k)$ 正交于误差 $S(f, k) - \hat{S}(f, k)$ 。这意味着:

$$E\{(S(f, k) - \hat{S}(f, k)) \cdot X^*(f, k)\} = 0 \quad (4)$$

因为估值处理是线性的, 通过估值带有噪声的信号 (它包含信号分量和噪声分量) 的信号分量, 噪声的估值 $\hat{N}(f, k)$ 也被有效地得到。而且, 以下的正交性关系也将是正确的:

$$E\{(N(f, k) - \hat{N}(f, k)) \cdot X^*(f, k)\} = 0 \quad (5)$$

其中 $\hat{N}(f, k)$ 表示噪声估值。由此也得出，对于每个频率，有以下的等式：

$$S(f, k) - \hat{S}(f, k) = \hat{N}(f, k) - N(f, k) \quad (6)$$

5 也就是，与噪声分量的估值 $\hat{N}(f, k)$ 有关的误差是和与估值的无噪声语音分量 $\hat{S}(f, k)$ 有关的误差相同的。

在本文件的其余部分，采用以下的表示法： $P_{uv}(f, k)$ 是在 $U(f, k)$ 和 $V(f, k)$ 之间的交叉功率谱密度 ($P_{uv}(f, k) = E\{U(f, k) \cdot V^*(f, k)\}$)。
 $P_{uu}(f, k)$ 是 $U(f, k)$ 的功率谱密度 (psd), ($P_{uu}(f, k) = E\{U(f, k) \cdot U^*(f, k)\}$)。 10

由于上述的正交性原理的结果，有可能得出对于交叉 psd $P_{..}(f, k)$ 的表示式，它是对于计算由公式 3 描述的 Wiener 滤波器所需要的：

$$P_{sx}(f, k) = E\{(X(f, k) - \hat{N}(f, k)) \cdot X^*(f, k)\} \quad (7)$$

15 而且，交叉 psd $P_{rx}(f, k)$ 被给出为：

$$P_{rx}(f, k) = E\{(X(f, k) - \hat{S}(f, k)) \cdot X^*(f, k)\} \quad (8)$$

在记住明显的等式 $P_{rx}(f, k) = P_{sx}(f, k) + P_{rx}(f, k)$ 后，公式 3, 6, 7 和 8 引出和显示自适应计算的概念，因为公式 3 中的 Wiener 滤波器
 20 ($P_{sx}(f, k) / P_{rx}(f, k)$) 取决于估值的信号 $\hat{S}(f, k)$ (6, 7) 和 (8)。

当得到最小值时，描述公式 2 中的误差的表示式取以下的形式：

$$\varepsilon_{\min}^2(f, k) = \frac{P_{ss}(f, k) \cdot P_{rx}(f, k) - |P_{sx}(f, k)|^2}{P_{xx}(f, k)} \quad (9)$$

显然，最小误差 (即 $\varepsilon_{\min}^2(f, k)$) 只有在想要的信号 $S(f, k)$ 完全
 25 与输入信号 $X(f, k)$ 相干时才等于零 (即， $P_{rx}(f, k)$ 趋近于零)。这是希望的。否则，当应用 Wiener 滤波器时有一个误差。这个误差的上限是 $P_{ss}(f, k)$ 。这是不希望的。换句话说，无误差结果只有在输入信号

$X(f, k)$ 中实际上无噪声时才得到。对于任何有限的噪声电平，得到有限的误差。由此得出，当在 $X(f, k)$ 中没有语音信号 $S(f, k)$ 时出现最坏情形的误差。

5 按照本发明的第一方面，提供一种在包含噪声的信号中抑制噪声以便提供噪声被抑制的信号的方法，其中对噪声作出估值，以及对语音连同某些噪声一起作出估值。

优选地，所述信号包括语音。

优选地，在语音连同某些噪声一起的估值中包括的噪声电平是可变的，以便把想要的噪声量包括在噪声被抑制的信号中。

10 噪声电平提供可接受的前后关系信息 (context information)。

优选地，噪声电平低于语音的掩蔽极限，因此对于收听者是听不见的。替换地，噪声电平接近于语音的掩蔽极限，因此某些噪声前后关系信息被保留在信号中。

15 优选地，如果信号噪声比足够高，以使得噪声电平已提供可接受的前后关系信息电平或已低于掩蔽极限，则本方法并不抑制噪声。

优选地，估值的噪声是功率谱密度。

20 按照本发明的第二方面，提供一种生成用于噪声抑制的增益系数的方法，其中增益系数的第一估值自适应地产生，以及这个第一估值被使用来生成噪声估值，然后使用该噪声估值来生成增益函数的第二估值。

25 在这个方面，本发明提供了重要的优点。它有效地消除在按照本发明实施的噪声抑制器中对于语音活动性检测器 (VAD) 的需要。VAD 基本上是一个能量检测器。它接收带有噪声的信号，把已滤波的信号的能量与预定的门限值进行比较，以及当门限值被超过时它表示在接收信号中存在语音。在许多语音编码/译码系统中，特别是在移动通信领域，VAD 的运行可以改变处理语音信号中的背景噪声的方式。具体地，在没有检测到语音的时间间隔期间内，可以切断传输，以及在接收端生成所谓“悦耳的噪声”。因此，这种间断发送和语音活动性检测方案的使用使得噪声抑制的使用复杂化，导致不想要的效果。因此，
30 消除对于语音活动检测器的需要和创建自动地适应于噪声条件改变的噪声抑制方案是高度希望的。因为本发明引入了一种其中获得语音和背景噪声的估值的噪声抑制方法，所以它实际上不需要判决：输入信

号究竟包含语音和噪声还是只包含噪声。结果，VAD 功能成为冗余的。
优选地，第一估值被使用来更新估值的噪声。

按照本发明的另一个方面，提供了按照本发明的第一方面运行的噪声抑制器、按照本发明的第二方面运行的噪声抑制器、按照本发明的第一和第二方面运行的噪声抑制器、包括按照本发明的第一和/或第二方面的噪声抑制器的无线终端、以及包括按照本发明的第一和/或第二方面的噪声抑制器的通信网。

优选地，通信终端是移动的。替换地，本发明可以在网络或固定的通信终端中使用。

10 按照本发明的另一个方面，提供了计算 Wiener 滤波器的方法，其中对语音和背景噪声作出估值，以及噪声远低于语音，以使得它整个地或部分地被掩盖成低于用户可听见的电平或可被觉察到。

优选地，该方法被用于频域中的噪声抑制。本方法可包括计算要被使用于噪声减小系统的 Wiener 滤波器的分子和分母。本说明书描述的噪声抑制系统特别适合于应用在包含单个传感器（例如话筒）的系统中。

优选地，滤波器是 Wiener 滤波器。优选地，它建立在包括语音和噪声的组的周期曲线图的估值的基础上。优选地，本方法涉及连续更新噪声 psd。

20 现在将参照附图仅仅作为例子来描述本发明的实施例，其中：

图 1 显示按照本发明的移动终端；

图 2 显示按照本发明的噪声抑制器；

图 3 显示人的听觉系统的依赖于频率和声级的掩蔽效果；

图 4 显示按照本发明的算法的方框图；以及

25 图 5 显示按照本发明的算法的功能性方框图。

在下面，符号 P 总的代表功率。在带撇的场合，即 P' ，它代表周期曲线；在不带撇的场合，即 P ，它代表功率谱密度 (psd)。按照它们总的可接受的意义，术语“周期曲线”被使用来表示在短的时间间隔上计算的平均值，以及术语功率谱密度被使用来代表较长的时段的平均。

30 现在参照图 1 描述包括按照本发明的教导的噪声抑制器 20 的移动终端 10 的实施例。图 1 相应于按照现有技术的移动终端的安排，虽然

这样的现有技术终端包括传统的现有技术噪声抑制器。移动终端和与其通信的无线通信系统按照全球移动通信系统 (GSM) 标准运行。

移动终端 10 包括发送 (语音编码) 支路 12 和接收 (语音译码) 支路 14。在发送 (语音编码) 支路 12, 语音信号由话筒 16 拾取和由
5 模拟-数字 (A/D) 变换器 18 采样, 以及在噪声抑制器 20 中被进行噪声抑制, 以便产生增强的信号。这需要估值背景噪声的频谱, 以使得在采样的信号中的背景噪声可被抑制。典型的噪声抑制器在频域中运行。时域信号首先被变换成频域, 这可以通过使用快速富立叶变换 (FFT) 有效地实现。在频域中, 把话音活动性与背景噪声区分开, 当
10 没有话音活动性时, 背景噪声的频谱被估值。然后, 根据当前输入信号的频谱和背景噪声估值来计算噪声抑制增益系数。最后, 通过使用逆 FFT (IFFT) 把信号变换回时域。

增强的 (噪声被抑制的) 信号被语音编码器 22 编码, 以便提取一组语音参量, 后者然后在信道编码器 24 中被编码, 其中把冗余性添加到编码的语音信号中, 以便提供某种程度的错误防护。最终得到的信号
15 然后被上变频成射频 (RF) 信号, 以及由发射/接收单元 26 发送。发射/接收单元 26 包括被连接到天线的双工滤波器 (未示出), 以使得能够进行发送和接收。

在已公布的文件 WO 97/22116 中描述了适合于在图 1 的移动终端
20 中使用的噪声抑制器。

为了延长电池寿命, 在移动通信系统中典型地应用不同种类的、依赖于输入信号的低功率运行模式。这些安排通常被称为间断发送 (DTX)。DTX 的基本思想是在非语音时间间隔内断开语音编码/译码过程。典型地, 产生某种悦耳的噪声信号, 它仿造在发射端处的背景
25 噪声, 以作为对于实际的背景噪声的替代。

语音编码器 22 被连接到发送 (TX) DTX 处理器 28。TX DTX 处理器 28 从话音活动性检测器 (VAD) 30 接收一个输入, 它表示在作为噪声抑制器块 20 的输出而被提供的噪声抑制信号中是否有话音分量。如果在信号中检测到语音, 则它的发送继续进行。如果没有检测到语音,
30 则噪声抑制的信号的发送被停止, 直至检测到语音为止。

在移动终端的接收 (语音译码) 支路 14, RF 信号被发射/接收单元 26 接收, 以及从 RF 被下变频到基带信号。基带信号被信道译码器

32 进行信道译码。如果信道译码器在已信道译码的信号中检测到语音，则信号被语音译码器 34 进行语音译码。

移动终端也包括坏帧处理单元 38 以便来处理坏的(即弄乱的)帧。

5 由语音译码器产生的信号，不管是译码的语音、悦耳的噪声还是重复的和衰减的帧，都将由数字模拟变换器 40 从数字形式变换成模拟形式，然后通过扬声器或耳机 42 播放给收听者。

图 2 上显示噪声抑制器 20 的进一步的细节。它包括快速富立叶变换、增益系数或 Wiener 滤波器计算块、以及逆快速富立叶变换。噪声抑制是通过把帧与增益系数/Wiener 滤波器相乘而在频域上执行的。

10 现在描述噪声抑制器 20 的运行。按照本发明，不是试图去估值在带有噪声的语音信号中的“真正的”语音分量 $S(f, k)$ ，而是使用 Wiener 滤波器按照关系式 $S(f, k) + \xi \cdot N(f, k)$ 来估值语音与一定量的噪声的组合。这样建立的、修正的 Wiener 滤波器取以下形式：

$$G(f, k) = \frac{P_{(S+\xi \cdot N)}(f, k)}{P_{XX}(f, k)} \quad (10)$$

$$= \frac{P_{SX}(f, k) + \xi \cdot P_{NX}(f, k)}{P_{SX}(f, k) + P_{NX}(f, k)}$$

15

假设语音和噪声分量是不相关的(也就是，在语音与噪声分量之间的交叉的 psd 必须等于零， $P_{SN}(f, k) = 0$)，公式 10 可被重新表示为以下形式：

$$G(f, k) = \frac{P_{SS}(f, k) + \xi \cdot P_{NN}(f, k)}{P_{SS}(f, k) + P_{NN}(f, k)} \quad (11)$$

20

下面说明因子 ξ 的作用。

如前所述，估值语音与一定量的噪声的组合的主要优点在于，应当具有较小的、与估值有关的误差。这个好处结合下面给出的公式 12 可变得更明显，这个公式规定在这种情形下得到的最小误差：

$$\varepsilon_{\min}^2(f, k) = (1 - \xi)^2 \cdot \frac{P_{SS}(f, k) \cdot P_{NN}(f, k)}{P_{SS}(f, k) + P_{NN}(f, k)} \quad (12)$$

现在可以明白，当 $P_{NN}(f, k)$ 趋向于零时，公式 12 趋向于零，这样，误差趋向于零，正如现有技术的情形那样。与现有技术一样，这是所希望的。然而，因为公式 12 包括 $(1 - \xi)^2$ 的因子，它比起现有技术的情形更快速地达到零。另一方面，当 $P_{NN}(f, k)$ 增加时， ε_{\min}^2 趋向于 $(1 - \xi)^2 \cdot P_{SS}(f, k)$ 。与现有技术一样，这是不希望的。然而，按照本发明的方法给出的误差总是小于先前所述的现有技术方法给出的误差。得出这个优点因为相乘因子 $(1 - \xi)^2$ 总是用来减小误差量。而且，通过设置 ξ 为适当的数值，因子 $(1 - \xi)^2$ 可被最小化，在这种情形下，误差被进一步最小化。

在本发明中，已经认识到： ξ 的数值可被确定，以达到以下的结果：

1. 给出乘积 $\xi \cdot P_{NN}(f, k)$ 的数值，它被 $P_{SS}(f, k)$ “掩蔽的”。虽然计算出了组合的语音和噪声的估值，但收听者将只听见语音，因为乘积 $\xi \cdot P_{NN}(f, k)$ 低于他的听觉的可听见电平。这样，利用人的听觉系统的性质，允许连同被掩蔽的噪声周期曲线的最大值一起计算语音的周期曲线。可以把正好用来达到这个结果的 ξ 称为 ξ_1 。

“掩蔽”效果是人的听觉系统的一个性质，这有效地设置一个取决于频率和取决于声级的听觉下限或门限值。因此，低于掩蔽门限值的任何噪声或语音分量都不被收听者觉察（听见）。通常可接受的是，不管频率如何，掩蔽门限值比当前的输入电平低约 13dB。图 3 上显示了这一点。按照本发明，为了估值纯语音信号（即，当尝试消除所有的背景噪声时），把纯语音信号连同刚好低于掩蔽门限值的那部分噪声合起来进行估值是足够的。

2. 允许自由选择输出端处噪声减小的电平。这可被使用来为远端收听者恢复信号的近端前后关系。可以把正好用来达到这个结果的 ξ 称为 ξ_2 。这意味着，可以这样选择 ξ ，以使得确保适当的噪声抑制，但也许可在接收端处在信号中保持一定的噪声分量，从而使背景噪声呈现为自然地代表在发射端的环境中存在的背景噪声。换句话说，有

可能选择 ξ 的数值，以使得带有噪声的语音信号中的噪声分量由于掩蔽效果而不会完全被消除。

在实际情形中，语音信号是非稳态的，所以需要短期的估值。因此，不是使用 psd 函数，如公式 11 所示，而是由周期曲线来代替某些项。噪声也可能是非稳态的，但通常认为是稳态的，所以仍旧可以使用长期估值。因此，想要的 Wiener 滤波器的形式为：

$$G(f, k) = \frac{P'_{ss}(f, k) + \xi \cdot P'_{nn}(f, k)}{P'_{ss}(f, k) + P_{nn}(f, k)} \quad (13)$$

应当指出，在公式 13 的分母中也有可能使用背景噪声功率谱密度项 P_{nn}。也应当看到，当在以上的公式 13 中使用 $\xi = \xi_1$ 时，项 $P'_{ss}(f, k) + \xi_1 \cdot P'_{nn}(f, k)$ 代表语音周期曲线和掩蔽的噪声周期曲线的组合，以及当使用 $\xi = \xi_2$ 时，项 $P'_{ss}(f, k) + \xi_2 \cdot P'_{nn}(f, k)$ 代表语音周期曲线和许可的噪声周期曲线的组合。分母 $P'_{ss}(f, k) + P_{nn}(f, k)$ 分别包括语音周期曲线和噪声 pnd。

对当前帧 k 的 Wiener 滤波器的计算是如下地根据先前的帧 k-1 来进行的。噪声 psd $P_{nn}(f, k-1)$ 、语音周期曲线 $P'_{ss}(f, k-1)$ 、和用于先前的帧的时间平均的帧的数目 $T(f, k-1)$ 都是已知的。对于当前的帧 k，输入语音和噪声周期曲线的组合 $|X(f, k)|^2$ 也是已知的。如果利用平方根或对数量度，则可以不使用 $P_{nn}(f, k-1)$ ，而使用 $R_{nn}(f, k-1)$ 或 $L_{nn}(f, k-1)$ ，正如在本说明中后面描述的。

下面使用一种 8 步骤算法来计算 Wiener 滤波器。8 步骤被显示于图 4，以及被描述如下：

步骤 1: 估算语音和噪声周期曲线的组合 $\bar{P}'_{ss}(f, k)$

这个周期曲线被如下地计算：

$$\bar{P}'_{ss}(f, k) = \alpha \cdot P'_{ss}(f, k-1) + (1-\alpha) \cdot |X(f, k)|^2 \quad (14)$$

应当指出， $\bar{P}'_{ss}(f, k)$ 是根据先前的语音周期曲线 $P'_{ss}(f, k-1)$ 以及当前的带有噪声的语音信号量 $|X(f, k)|^2$ ，它是由因子 α 确定的。 α 的数值被选择为从带有噪声的语音信号 $|X(f, k)|^2$ 的当前的语音分量

$|S(f, k)|^2$ 中提供最大可能的贡献, 但它被限于确保因子 $(1-\alpha) \cdot |N(f, k)|^2$ (该因子代表将被包括在内的当前的噪声信号量) 被和值 $\alpha \cdot P'_{ss}(f, k-1) + (1-\alpha) \cdot |S(f, k)|^2$ 掩蔽, 该和值代表当前的语音周期曲线的估值。所以, 应当看到, 对于每个帧 k 的每个频率仓 f 必须重新计算遗忘因子 α 。也应当指出, 在公式 14 中提到的因子 $(1-\alpha)$ 类似于 ξ_1 。

实际上, 步骤 1 是通过使用频谱相减方法首先估值当前的语音周期曲线而实施的, 该频谱相减方法是在“Suppression of Acoustic Noise in Speech Using Spectral Subtraction (使用频谱相减抑制语音中的声音噪声)”, IEEE Trans. On Acoustics Speech and Signal Processing, vol.27, No.2, pp.113-120, April 1979 中描述的。然后, 掩蔽电平被设置为低于被估值的语音周期曲线值约 13dB 的数值。噪声周期曲线与语音周期曲线相同的方式被估值。 α 的数值然后通过使用掩蔽、噪声周期曲线和输入周期曲线而被计算。

步骤 2: 估值语音和噪声的组合 psd $\bar{P}_{xx}(f, k)$

这个 psd 代表输入的总功率, 以及被如下地估值:

$$\bar{P}_{xx}(f, k) = \alpha \cdot \left[P'_{ss}(f, k-1) + \frac{\lambda}{\alpha} P_{nn}(f, k-1) \right] + (1-\alpha) \cdot |X(f, k)|^2 \quad (15)$$

这个 psd 组合了短期平均值 (对于语音的周期曲线) 和长期平均值 (对于噪声的 psd)。

步骤 3: 对 Wiener 滤波器的估值

公式 11 的 Wiener 滤波器可被重写为以下的形式:

$$G_1(f, k) = \frac{\bar{P}'_{ss}(f, k)}{\bar{P}_{xx}(f, k)} \quad (16)$$

因此可以根据公式 14 和 15 的结果进行计算。因为 $\hat{S}_1(f, k) = G_1(f, k) \cdot X(f, k)$, 应当看到, 估值的语音 $\hat{S}(f)$ 包含语音和掩蔽的噪声部分。增益 $G_1(f, k)$ 的最小值被设置为 $(1-\alpha)$ 。

步骤 4: 对噪声 psd $P_{nn}(f, k)$ 的更新

为了更新噪声 psd, 使用在公式 8 中给出的理论结果, 在必要时,

用乘积 $(1-G_1(f, k)) \cdot |X(f, k)|^2$ 替换乘积 $(X(f, k) - \hat{S}(f, k)) \cdot X^*(f, k)$ 。可以使用以下的三种方法：

- 5 (i) 功率 psd 估值；
 (ii) 平方根 psd 估值；以及
 (iii) 对数 psd 估值。

在下面描述的所有的的方法中， λ 代表在 0 与 1 之间的遗忘因子。

- (i) 功率 psd 估值

这个方法使用正交性原理，以及它是基于 Welch 方法的，Welch 方法被描述在 “The Use of Fast Fourier Transform for the
 10 Estimation of Power Spectra: A Method Based on Time Averaging
 Over Short, Modified Periodograms (用于功率谱估值的快速富立
 叶变换的使用：基于修正的短周期曲线的时间平均的方法)”，IEEE
 Trans. On Audio and Electroacoustics, vol. Au-15, No. 2,
 pp. 70-72, June 1967 中。它使用按照下式的、被称为“指数时间平均”
 15 的技术：

$$P_{NN}(f, k) = \lambda \cdot P_{NN}(f, k-1) + (1-\lambda) \cdot (1-G_1(f, k)) \cdot |X(f, k)|^2 \quad (17)$$

其中 $G_1(f, k)$ 是按照公式 16 计算的 Wiener 滤波器。

- (ii) 平方根 psd 估值

20 这个方法使用 Welch 方法的修正形式，以及它是基于幅度平均的：

$$\begin{cases} R_{NN}(f, k) = \lambda \cdot R_{NN}(f, k-1) + (1-\lambda) \cdot \sqrt{(1-G_1(f, k))} \cdot |X(f, k)| \\ P_{NN}(f, k) = R_{NN}(f, k) \cdot R_{NN}(f, k) \end{cases} \quad (18)$$

$R_{NN}(f, k)$ 代表平均噪声幅度。

- (iii) 对数 psd 估值

25 这个方法使用对数域中的时间平均：

$$\begin{cases} L_{NN}(f, k) = \lambda \cdot L_{NN}(f, k-1) + (1-\lambda) \cdot \text{Log}[(1-G_1(f, k)) \cdot |X(f, k)|^2] \\ P_{NN}(f, k) = \exp[L_{NN}(f, k) + \gamma] \end{cases} \quad (19)$$

$L_{nn}(f, k)$ 是指对数功率域中的平均值。 γ 是欧拉常数, 以及具有 0.5772156649 的数值。

在上述的三个方法的每个方法中, 遗忘因子在更新噪声时起重要的作用, 以及被加以定义以便在噪声幅度快速地变化时提供良好的 psd 估值。这是通过把 λ 与当前的输入周期曲线 $|X(f, k)|^2$ 与先前的帧中的噪声 psd $P_{nn}(f, k-1)$ 之间的差值相联系而完成的。 λ 取决于数值 $T(f, k)$, 该数值规定对于时间平均所使用的帧的数目, 以及被如下地确定:

$$\begin{cases} \text{if } |X(f, k)|^2 > 10 \cdot P_{nn}(f, k-1) & T(f, k) = 5 \\ \text{elseif } |X(f, k)|^2 < 0.1 \cdot P_{nn}(f, k-1) & T(f, k) = 5 \\ \text{else} & T(f, k) = \text{Min}[T(f, k-1) + 1, 20] \end{cases} \quad (20)$$

10

以及 λ 可以如下地根据 $T(f, k)$ 被得出:

$$\lambda = \frac{T(f, k)}{T(f, k) + 1} \quad (21)$$

应当指出, 对于每个帧 k 和每个频率仓 f , 必须重新计算遗忘因子。显然, 由于 λ 是步骤 2 中所需要的, 它必须被计算以使得它对于该步骤是可供使用的。也应当看到, 因为噪声 psd 连续地被更新, 这就消除了噪声抑制器 20 中对话音活动性检测器的需求。

步骤 5: 对当前的语音周期曲线 $P'_{ss}(f, k)$ 的估值

当前的语音周期曲线 $P'_{ss}(f, k)$ 在算法中起重要的作用。对于当前的帧来估值这个 $P'_{ss}(f, k)$, 以使得它可被使用于下一个帧, 也就是在公式 14 和 15 中被使用。如下所述, $P'_{ss}(f, k)$ 只应当包含语音, 而不应当包含任何噪声。

实际上, 在步骤 3 得出语音幅度估值 $\hat{S}(f, k)$ 后, 这个步骤需要 $P'_{ss}(f, k)$ 的估值, 它代表当前的语音周期曲线。

广泛地可被接受的是, $P'_{ss}(f, k)$ 可以简单地用平方估值的语音幅度代替, 也就是: $P'_{ss}(f, k)$ 等于 $|S(f, k)|^2$ 的估值 $|\hat{S}(f, k)|^2$ 。不幸地, 良好的估值 $\hat{S}(f, k)$ 实际上并不意味着: 对于 $|S(f, k)|^2$ 的良好估值可以通过只取平方而得出。因此, 按照本发明的方法寻求通过应用 MMSE

准则得出 $|S(f, k)|^2$ 的更精确的估值 $P'_{ss}(f, k)$ 。

通过考查组合的语音和噪声周期曲线，可以看到：

$$Y(f, k) = |X(f, k)|^2 = |S(f, k)|^2 + |N(f, k)|^2 + S^*(f, k) \cdot N(f, k) + S(f, k) \cdot N^*(f, k).$$

- 5 因此， $|S(f, k)|^2$ 的良好的估值可以通过使得以下的误差（MMSE 准则）最小化而得出：

$$\chi^2(f, k) = E \left\{ \left| |S(f, k)|^2 - H(f, k) \cdot Y(f, k) \right|^2 \right\} \quad (22)$$

其中 $H(f, k) \cdot |X(f, k)|^2$ 代表语音周期曲线 $|S(f, k)|^2$ 的估值。

- 10 方程 22 的直接解需要高阶方程的解，但这个解可以通过假设语音和噪声是高斯过程并与零均值是非相关的而被简化，从而给出相应的高阶 Wiener 滤波器 $H(f, k)$ 的近似值。在这个方法中使用的近似在下面公式 23 中呈现。（应当看到，在这一个阶段可以使用不同的近似而不背离本发明的原理的基本特性）。

15

$$H(f, k) = \frac{3 \cdot SNR(f, k) \cdot SNR(f, k) + SNR(f, k)}{3 \cdot SNR(f, k) \cdot SNR(f, k) + 6 \cdot SNR(f, k) + 3} \quad (23)$$

这里， $SNR(f, k)$ 是指信号噪声比，以及被如下地计算：

$$SNR(f, k) = \frac{G_1(f, k)}{1 - G_1(f, k)} \quad (24)$$

- 20 公式 24 是关联着 Wiener 滤波器和信号噪声比的已知函数的倒数。（ $Wiener = SNR / (SNR + 1)$ ）

因此，语音周期曲线被如下地计算：

$$P'_{ss}(f, k) = H(f, k) \cdot |X(f, k)|^2 \quad (25)$$

- 25 步骤 6：放大函数

在高的 SNR 条件下, 当带有噪声的输入信号的语音分量比起噪声分量很大时, 估值的 Wiener 滤波器 $G_1(f, k)$ 趋向于 1. 而且, 当语音对噪声的比值很高时, $G_1(f, k)$ 可被相当精确地估值. 因此, 可以具有良好的确信度: 在步骤 3 确定的 Wiener 滤波器能提供最佳滤波, 以及
5 给出一个包含带有残余的(掩蔽的)噪声量的高度精确的语音估值 $\hat{S}_1(f)$ 的输出. 在这种情形下, 当滤波器的增益接近于 1 时, 提供小量的放大以使得增益仍旧接近于 1 是有利的. 然而, 附加的放大也应当被加以限制, 以便确保 Wiener 滤波器增益在任何情形下不超过 1.

另一方面, 在其中带有噪声的输入信号中的语音分量小于噪声分量
10 量的情形下, 相反的结果是正确的. Wiener 滤波器增益是小的, 并且很可能不会像在高 SNR 的条件下那样被精确地确定. 在这种情形下, 放大 Wiener 滤波器输出值并不是有利的, 因而估值的 Wiener 滤波器应当保持为在步骤 3 中原先估值的形式.

考虑到在不同的 SNR 条件下存在的这两个相反的需要, 步骤 3 中
15 确定的 Wiener 滤波器按照下式被更新:

$$G_2(f, k) = G_1(f, k)^{\text{Min}\{K_b(f), 1 - G_1(f, k)\}} \quad (26)$$

从而, 产生出在估值该最后输出时要被使用的 Wiener 滤波器 $G_2(f, k)$. $G_2(f, k)$ 是 $G_1(f, k)$ 的函数.

20 公式 26 利用这一事实: 诸如 $y = x^{1-x}$ ($x > 0$) 的函数在 x 小于 1 时提供放大. 所以, 它满足在良好的 SNR 条件下提供更大的放大和在低的 SNR 的条件下提供低的放大的需要.

25 变量 $K_b(f)$ 可以取 0 与 1 之间的数值, 并且该变量是被包括在公式 26 中的指数中, 以便使得能够使用对于不同的频带 f 的不同的 (例如预定的) 放大值, 如果需要的话.

步骤 7: 选择噪声减小程度

在这个步骤中, 选择想要的噪声减小程度. 对于公式 11 中给定的 Wiener 滤波器, 相应的理想的时间输出具有 $\hat{s}(t) = s(t) + \xi \cdot n(t)$ 的形式. 回忆带有噪声的输入信号具有 $x(t) = s(t) + n(t)$ 的形式, 由滤波器
30 提供的噪声减小理论上约为 $20 \cdot \log[\xi]$ dB. 通过考虑输入信号的噪声电平与输出信号 (即在噪声抑制后获得的信号) 的噪声电平的比值,

这个结果可被证明是正确的。这个比值就是 $\xi \cdot n(t)/n(t)$ ，当以分贝形式的功率比值来表示时，它等于 $20 \cdot \log[\xi]$ 。因此，因子 $0 < \xi < 1$ 对应于由滤波器引入的噪声减小。

选择想要的噪声减小程度和确定对于达到该噪声减小所必须的 ξ 的数值（例如，对于 -12dB 噪声减小， $\xi = 0.25$ ）后，确定因子 η 以使得：

$$G_1(f, k) + \eta \cdot (1 - G_1(f, k)) \Leftrightarrow \frac{P_s(f, k) + \xi \cdot P_n(f, k)}{P_s(f, k) + P_n(f, k)} \quad (27)$$

公式 27 给出一种把被最佳化成能够提供只包含掩蔽的噪声的输出的 Wiener 滤波器与能提供包含一定量的许可噪声的输出的 Wiener 滤波器相关联的方法。按照步骤 1-3，Wiener 滤波器 $G_1(f, k)$ 被构建为可提供带有噪声的语音信号的语音分量加上可被该语音分量有效地掩蔽的噪声量的估值。因此，在那些在输出中一定量的噪声是被许可（想要）的条件下，Wiener 滤波器必须随之被修正。在公式 27 中， $G_1(f, k)$ 代表在步骤 3 Wiener 滤波器被最佳化，以便提供包含被语音掩蔽的噪声的输出。项 $\frac{P_s(f, k) + \xi \cdot P_n(f, k)}{P_s(f, k) + P_n(f, k)}$ 代表提供噪声减小量 ξ 的 Wiener 滤波

器，它产生包含语音和想要的/许可的噪声量的输出信号。项 $\eta \cdot (1 - G_1(f, k))$ 因此代表未掩蔽的噪声量，并且在实际上是 $\frac{P_s(f, k) + \xi \cdot P_n(f, k)}{P_s(f, k) + P_n(f, k)}$

与 $G_1(f, k)$ 之间的差值。考虑到 $G_1(f, k)$ 包含的噪声值约等于 $(1 - \alpha)$ 乘以在原先的带有噪声的语音信号中存在的噪声的事实，在 α 、 η 与 ξ 之间存在以下的关系式：

$$1 - \alpha + \eta \cdot \alpha \Leftrightarrow \xi \quad (28)$$

25

步骤 8：对最终估值的 Wiener 滤波器的估值

通过使用公式 16、26 和 28，施加到输入端上的最终的 Wiener 滤波器 $G(f, k)$ 被给出为：

$$\begin{cases} \text{if } \alpha > (1-\xi) & \eta = \frac{\alpha + \xi - 1}{\alpha} \\ \text{else} & \eta = 0 \end{cases} \quad (29)$$

$$G(f, k) = G_e(f, k) + \eta \cdot (1 - G_e(f, k))$$

虽然在 η 取决于 α ，以及对于每个帧 k 的每个频率仓 f ，它具有不同的数值，但总的噪声减小程度被保持为约 $20 \cdot \log[\xi]$ dB 的恒定值。

5 替换地，步骤 1 到 8 可以通过使用包括信号噪声比的式子的式子来实施。在以上给出的步骤 1-8 的详细的实施方案中，讨论是建立在噪声 psd 函数、语音周期曲线和输入功率（周期曲线 + psd）的基础之上的。然而，通过把公式 11 和/或公式 13 除以噪声 psd，可以得出另一个表示式。这另一个表示式需要对（信号+掩蔽的噪声）对噪声的
10 比值估值，而不是对语音周期曲线估值。

图 5 上显示了实施本发明的算法 50。算法 50 被显示为可以划分成一组自适应处理过程的步骤 52 和一组非自适应处理过程的步骤 54。自适应处理过程使用对 Wiener 滤波器的计算来重新计算 Wiener 滤波器。因此，对 Wiener 滤波器的计算的步骤对于自适应处理过程和非自
15 适应处理过程是共同的。

这种 Wiener 滤波器计算也适用于在包括一个传感器和一个扬声器的组合的回声和噪声控制系统中使得残余回波最小化。

虽然显示和描述了本发明的优选实施例，但应当看到，这样的实施例只是通过实例描述的。例如，虽然本发明被描述为使用在位于移
20 动终端的上行链路路径上的噪声抑制器中，它提供噪声抑制的信号给语音编码器，但本发明同样可使用于处在移动终端的下行链路路径上的噪声抑制器中，以便代替或附加到上行链路路径上的噪声抑制器上。在这种情形下，它可对由语音译码器提供的信号起作用。而且，虽然本发明被描述为用在移动终端中，但替换地，它可以存在有在通
25 信网内的噪声抑制器中，无论是在与语音编码器或语音译码器相关的情况下进行使用。

对于本领域技术人员将出现各种变例、改变和替换，而不背离本发明的范围。因此，以下的权利要求打算覆盖属于本发明的精神和范围内的所有的这样的等同物或变例。

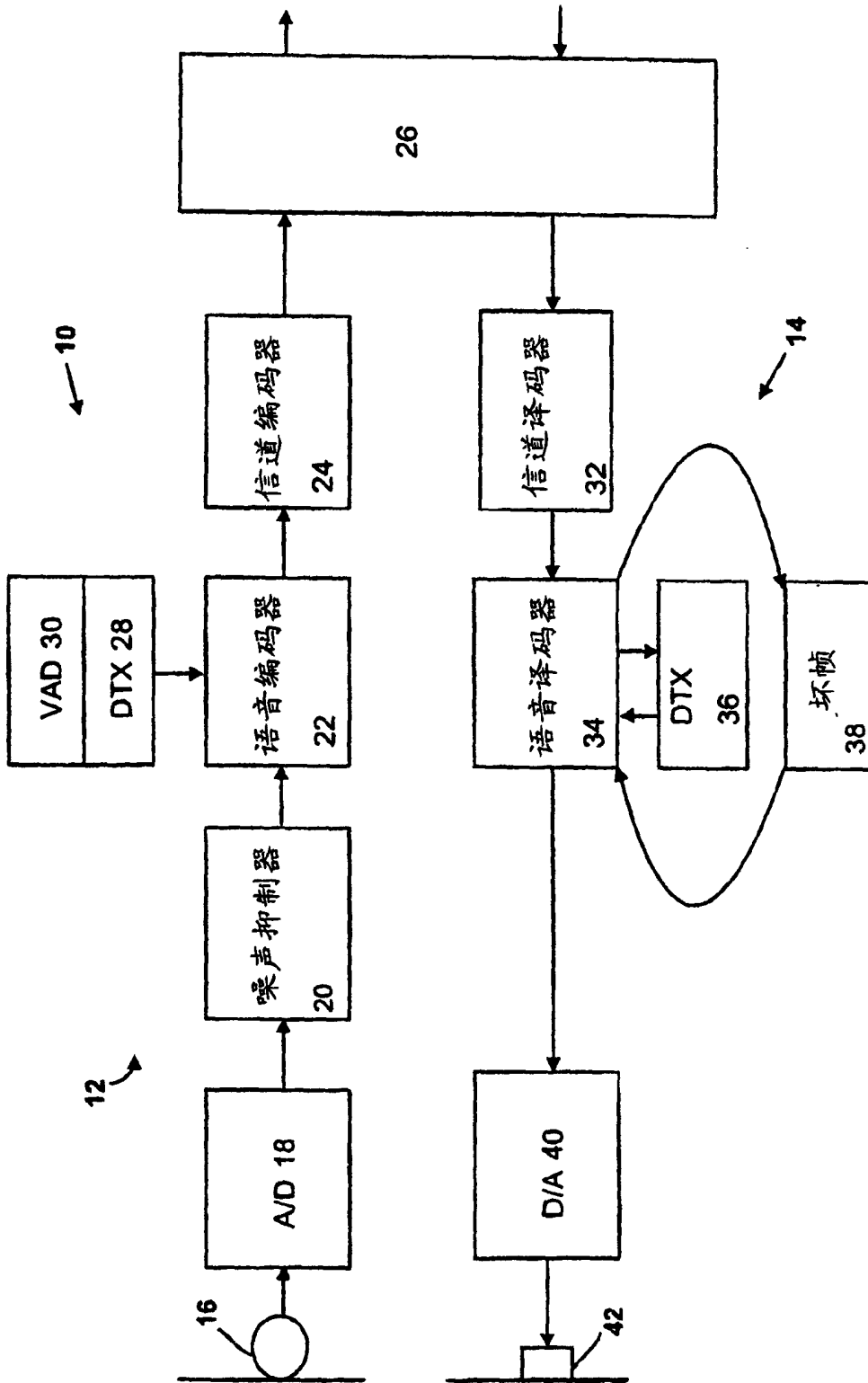
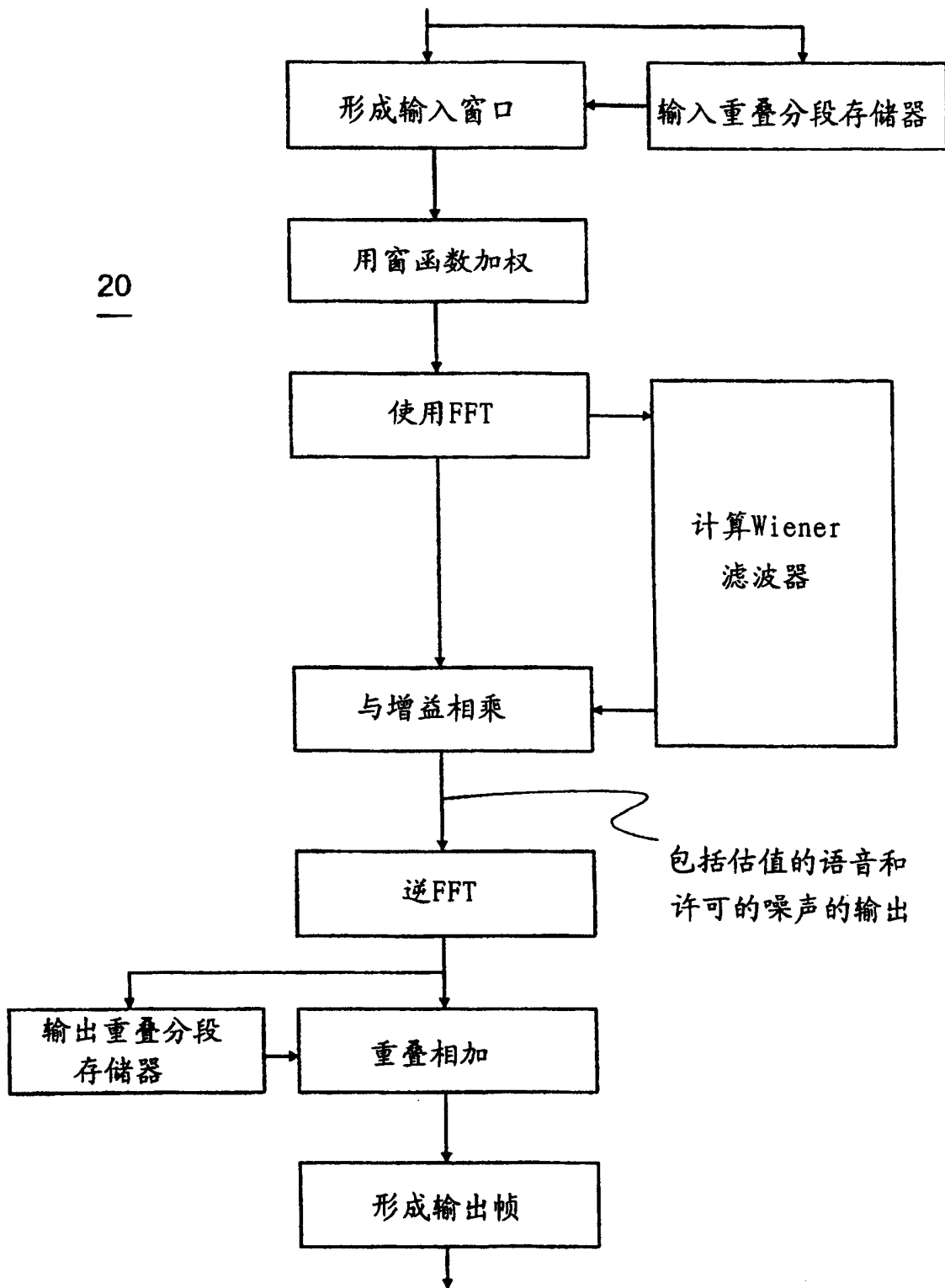


图 1



包括估值的语音和
许可的噪声的输出

图 2

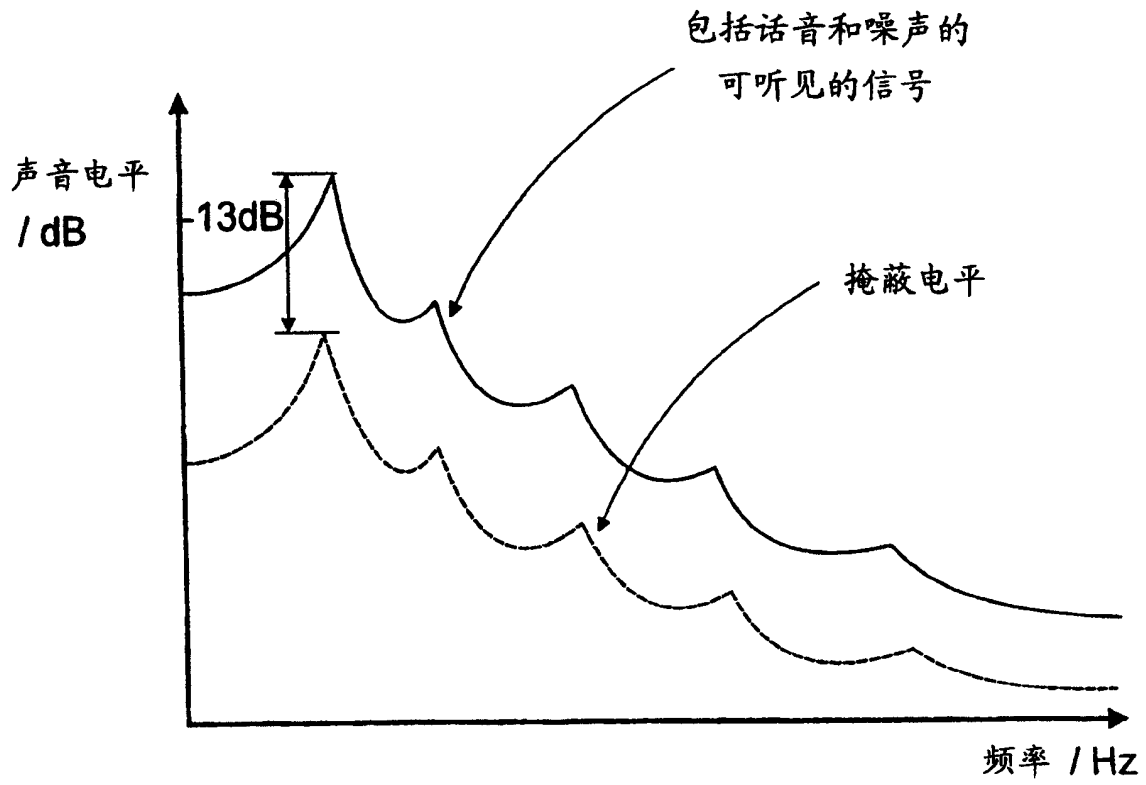


图 3

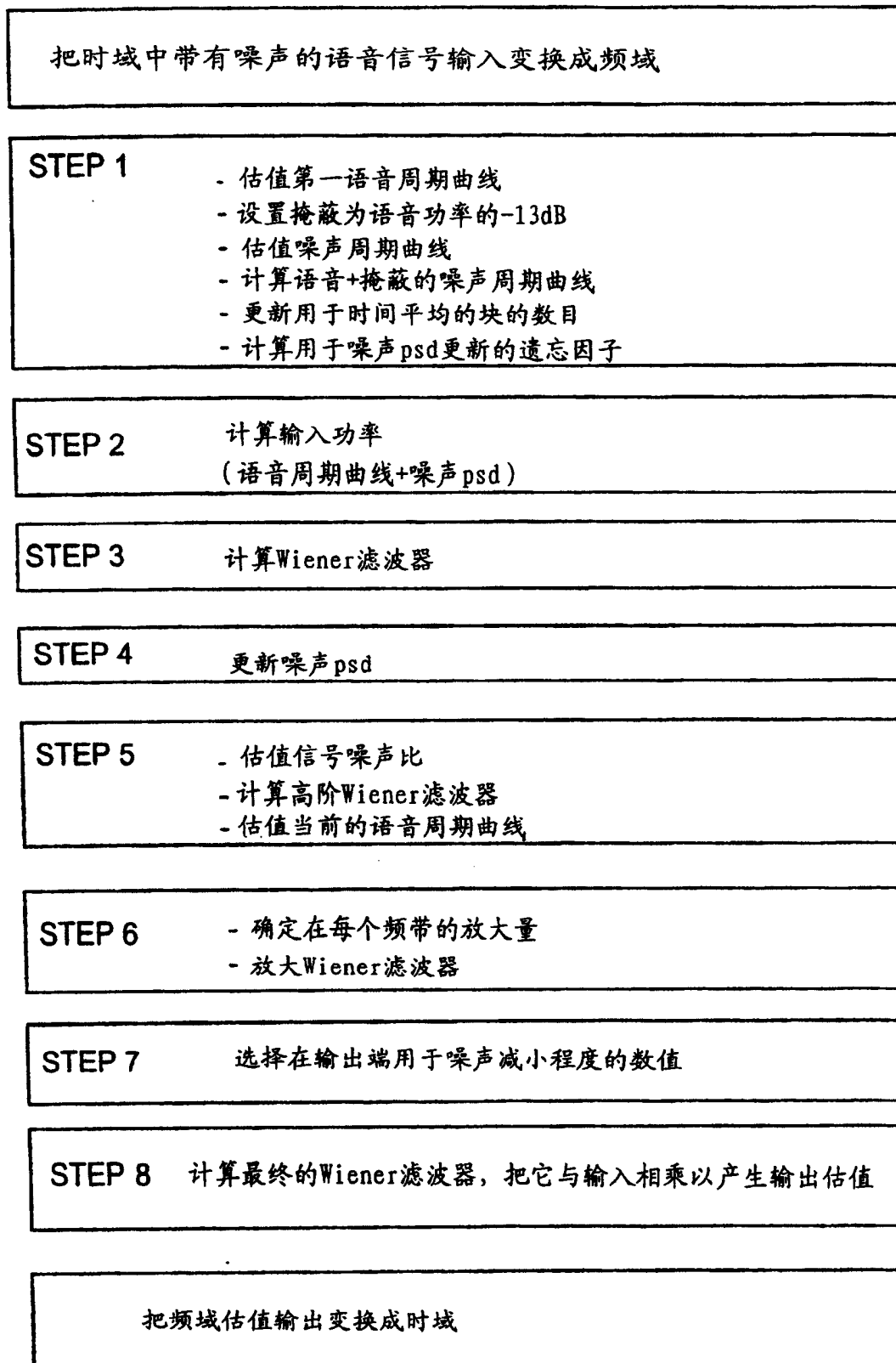


图 4

