



# 权 利 要 求 书

---

1. 一种用于噪声抑制的降低输入语音信号中噪声的方法, 包括:  
将输入语音信号转换为一个频谱;

根据一个第一值和一个第二值确定滤波器特征, 所述的第一值是以该频谱的一个电平与该频谱中包含的噪声的一个估计电平之比为基础获得的, 所述第二值是从该频谱的基于帧的信号电平与该估计噪声电平之比的最大值和从所述估计噪声电平中求出的; 和  
响应该滤波器特征通过滤波降低所述输入语音信号中的噪声。

2. 根据权利要求1的降低输入语音信号中噪声的方法, 其中所述第一值是利用从包含该输入信号的预定电平和该噪声频谱的估计电平的表中获得的值求出的。

3. 根据权利要求1的降低输入语音信号中噪声的方法, 其中所述第二值是响应该信号电平与该估计噪声电平的比率的最大值和该基于帧的噪声电平获得的值, 并是一个通过根据滤波器特征滤波调整最大噪声减少量的值, 以便该最大噪声降低量在1dB范围内将基本线性地变化。

4. 根据权利要求1的降低输入语音信号中噪声的方法, 其中所述估计噪声电平是根据基于帧的输入信号的幅度的均方根值和该均方根值的最大值获得的值, 该信号电平与估计噪声电平之比率的 $\frac{1}{2}$ 最大值是一个根据该均方根值的最大值和该估计值计算获得的值, 其中该均方根值的最大值是基于帧的输入信号的幅度的均方根值,

根据紧接前一帧的该均方根值的最大值获得的值, 和一个预定值中的一个最大值。

5. 一种用于噪声抑制的降低输入语音信号中噪声的装置, 包括:  
将输入语音信号转换成频谱的装置;

用于根据一个第一值和一个第二值确定滤波器特征的装置, 所述的第一值是以该频谱的一个电平与该频谱中包含的噪声的一个估计电平之比率为基础获得的, 所述第二值是从该频谱的基于帧的信号电平与该估计噪声电平之比的最大值和从所述估计噪声电平中求出的; 和

响应所述滤波器特征通过滤波降低所述输入语音信号中的噪声的装置。

# 说明书

---

## 用于降低噪声的方法

本发明涉及除去包含在语音信号中噪声的方法,用于抑制或降低其中的噪声。

在便携电话机或语音识别的技术领域中,人们感到对包含在所收集语音信号中的诸如背景噪声或环境噪声的噪声进行抑制以增强其语音成分是必要的。

作为一种增强语音或降低噪声的技术, R. J. McAulay和 M. L. Maplass 提出了一种把条件概率函数用于衰减因子调整的技术。其发表在1980年4月IEEE 传输,声学,语音信号处理杂志 (Vol. 28, pp. 137 to 145) 上, 题目为"使用软判决噪声抑制滤波器的语音增强"。

在上述的噪声抑制技术中,由于不适当的抑制滤波器或根据不适当的固定信噪比(SNR)进行的操作产生非自发的声调或失真的语音是经常发生的。在实现最佳性能的实际操作中,不得不把信噪比作为噪声抑制装置的参数之一来进行调整是用户所不希望有的。此外,对于普通的语音信号增强技术,要有效地消除噪声而在易受短时间内SNR显著变化影响的语音信号中不产生失真是困难的。

这些语音增强或噪声降低技术使用了一种通过将输入功率或电平与一预定门限值比较鉴别噪声域的技术。然而,如果该门限值的时间常数由于该技术防止门限值跟踪该语音而被增加,则一个变

化的噪声电平,尤其是一个增长的噪声电平不能被合适地跟踪,这样偶然会导致误鉴别。

为了克服这一缺陷,本发明人在日本专利申请平—6—99869(1994)中已提出了一种在语音信号中降低噪声的噪声降低方法。

对于该用于语音信号的噪声降低方法,通过自适应地控制一个最大似然滤波器实现噪声抑制,该最大似然滤波器被构造用于根据由输入的语音信号导出的SNR和该语音当前概率计算语音成分。该方法使用了一个与输入语音频谱对应的信号,该输入语音频谱小于在其中计算该当前概率的估计噪声频谱。

对于该用于语音信号的噪声降低方法,由于该最大似然滤波器取决于该输入语音信号的SNR被调整到一个最佳抑制滤波器,所以可以实现输入语音信号的有效噪声降低。

然而,由于计算语音当前概率需要复杂和大量的处理运算,因此需要对这种处理运算进行简化。

因而,本发明的目的是提供一种用于输入语音信号的噪声降低方法,通过此方法可以简化用于输入语音信号的噪声抑制的处理运算。

按照本发明的一个方案,提供一种用于噪声抑制的降低输入语音信号中噪声的方法,该方法包括将输入语音信号转换成一个频谱,根据一个第一值和一个第二值确定滤波器特征,所述的第一值是以该频谱的一个电平与该频谱中包含的噪声的一个估计电平之比为基础获得的,所述第二值是从该频谱的基于帧的信号电平与该

估计噪声电平之比的最大值和从该估计噪声电平中找出的, 和响应该滤波器特征通过滤波降低输入语音信号中的噪声。

按照本发明的另一个方案, 提供一种用于噪声抑制的降低输入语音信号中噪声的装置, 该装置包括用于将输入语音信号转换成频谱的装置; 用于根据一个第一值和一个第二值确定滤波器特征的装置, 所述的第一值是以该频谱的一个电平与该频谱中包含的噪声的一个估计电平之比为基础获得的, 所述第二值是从该频谱的基于帧的信号电平与该估计噪声电平之比的最大值和从该估计噪声电平中找出的; 和响应该滤波器特征通过滤波降低输入语音信号中的噪声的装置。

按照本发明, 对于降低语音信号中噪声的方法和装置, 所述第一值是根据通过从该输入语音信号转换获得的输入信号频谱与在该输入信号频谱中包含的估计噪声频谱之比计算获得的值, 它设定了滤波器特征的一个初始值, 以确定在用于噪声降低的滤波中的噪声降低量。所述第二值是根据该输入频谱的信号电平与该估计噪声电平之比的值(即最大SNR)和该估计噪声电平计算获得的, 它是一个用于可变化地控制该滤波器特征的值。按照由所述第一和第二值可变控制的滤波器特征通过滤波可以从该输入语音信号中消除掉一个与该最大SNR相应量的噪声。

由于可以使用一个具有输入信号频谱的预定电平和加在该输入信号频谱之中的噪声频谱的估计电平的表来查找该第一值, 因而可有助于减少处理量。

而且,响应该最大SNR和基于帧的噪声电平获得该第二值,所述滤波器特征可被调整,以便响于应该最大SNR,该滤波的最大噪声减少量在1dB范围内将基本线性地变化。

对于上述本发明的降低噪声方法,所述第一和第二值被用于控制用于滤波从输入语音信号中消除噪声的滤波器特征,由此,按照该输入语音信号中的最大SNR通过滤波从该输入语音信号中消除噪声。特别地,在语音信号中由于以高SNR滤波导致的失真可被削弱,并且实现该滤波器特征的处理运算量也可被减少。

此外,根据本发明,用于控制滤波器特征的第一值可使用一个具有输入信号频谱的电平和加在该输入信号频谱之中的估计噪声频谱的电平的表计算得到,以减少实现该滤波器特征的处理量。

而且,根据本发明,响应该最大SNR和基于帧的噪声电平获得的第二值可被用于控制滤波器特征,以减少实现该滤波器特征的处理量。通过滤波器特征实现的最大噪声降低量可相应于该输入语音信号的N比率改变。

图1示出了本发明的用于降低语音信号的噪声方法的第一实施例,它被应用于一降低噪声装置中。

图2示出了图1的实施例中能量 $E[k]$ 和衰减能量 $E_{\text{decay}}[k]$ 的一个特定例子。

图3示出了图1的实施例中RMS值 $\text{RMS}[k]$ ,估计噪声电平值 $\text{MinRMS}[k]$ 和最大RMS值 $\text{MaxRMS}[k]$ 的特定例子。

图4示出了图1的实施例中该相对能量 $B_{rel}[k]$ ,以dB为单位的最大SNR  $MaxSNR[k]$  和一个用于噪声鉴别的门限值  $dB_{thres_{rel}}[k]$ 的特定例子。

图5是图1所示实施例中一个表示作为相应于最大SNR  $MaxSNR[k]$ 定义的函数 NR-电平 $[k]$ 的曲线。

图6示出了图1所示实施例中NR  $[w, k]$ 和以dB 为单位的最大噪声减少量之间的相对关系。

图7示出了图1所示实施例中 $Y[w, k]/N[w, k]$ 的比率和以dB 为单位的相应于NR  $[w, k]$ 的 $H_n[w, k]$ 之间的相对关系。

图8示出了本发明的用于语音信号的噪声降低方法的第二实施例,它被应用于一噪声减低装置中。

图9是一个表示根据由图1和图8的噪声降低装置实现的噪声抑制获得的语音信号的各分段部分的失真相应于该分段的SNR的曲线图。

参考附图,将详细描述按照本发明的用于降低语音信号中噪声的方法和装置。

图1 示出了根据本发明用于降低语音信号中噪声的噪声降低装置的实施例。

该噪声降低装置中作为主要元件包括:一个快速富立叶变换单元3,用于转换输入语音信号成为频域信号或频谱;一个 $H_n$ 值计算单元7,用于在通过滤波从输入语音信号中消除噪声部分期间控制滤

波器特征; 和一个频谱校正单元10, 用于响应由Hn值计算单元7产生的滤波器特征通过滤波降低输入语音信号中的噪声。

输入到所述噪声降低装置的语音信号输入端13的输入语音信号 $y[t]$ 被提供给成帧单元1。由该成帧单元1输出的一个已成帧信号 $y$ -帧 $j, k$ 被提供到开窗单元2, 噪声估计单元5内的均方根 (RMS) 计算单元21, 和滤波单元8。

开窗单元2的输出提供到快速富立叶变换单元3, 它的输出被同时提供给频谱校正单元10和一个频带分割单元4。该频带分割单元4的输出被施加到频谱校正单元10, 噪声估计单元5 内的噪声频谱估计单元26, 以及Hn值计算单元7。频谱校正单元10 的输出经反向快速富立叶变换单元11和一个重叠相加单元12被提供一个语音信号输出端14。

均方根 (RMS) 计算单元21的输出被提供到相对能量计算单元22, 最大RMS计算单元23, 估计噪声电平计算单元24, 以及噪声频谱估计单元26。最大RMS计算单元23的输出被提供给估计噪声电平计算单元24和最大SNR计算单元25。相对能量计算单元22的输出被提供到噪声频谱估计单元26。估计噪声电平计算单元24的输出被提供给滤波单元8, 最大SNR计算单元25, 噪声频谱估计单元26, 以及一个NR值计算单元6。最大SNR计算单元25的输出被提供到NR值计算单元6和噪声频谱估计单元26, 该噪声频谱估计单元26的输出被施加到Hn值计算单元7。

NR值计算单元6的输出被再次提供给NR值计算单元6自身, 而且还被施加到Hn值计算单元7。

该Hn值计算单元7的输出经滤波单元8和频带转换单元9 提供到频谱校正单元10。

对所述噪声降低装置的上述第一实施例的操作进行说明。

在语音信号输入端13施加一个包含语音成分和噪声成分的输入语音信号y[t]。该输入语音信号y[t] (它是一个按例如取样频率FS取样的数字信号) 被提供给成帧单元1, 在其中它被分割成多个帧, 每一帧都具有FL个取样的帧长。然后在上述帧的基础上处理被这样分割的输入语音信号y[t]。作为沿时间轴的帧的位移量的帧间隔是FI个取样, 以便第(k+1)帧从第k帧的FI个取样之后开始。举例来说明取样频率和取样数量, 如果取样频率是8KHz, 则80个取样的帧间隔FI相应为10ms, 而160个取样的帧长相应为20ms。

在由快速富立叶变换单元3进行正交变换计算之前, 开窗单元2把来自成帧单元1的每个已成帧信号y-帧j, k与窗函数W<sub>input</sub>相乘。在基于帧信号处理操作的终端级执行的随后的反向FFI时(如将在后面进行描述的), 使输出信号与窗函数W<sub>output</sub>相乘。所述窗函数W<sub>input</sub>和W<sub>output</sub>可由下列等式(1)和(2)分别示例:

$$W_{input}[j] = \left( \frac{1}{2} - \frac{1}{2} \cos\left(\frac{2\pi j}{FL}\right) \right)^{\frac{1}{4}}, \quad 0 \leq j \leq FL \quad \dots (1)$$

$$W_{output}[j] = \left( \frac{1}{2} - \frac{1}{2} \cos\left(\frac{2\pi j}{FL}\right) \right)^{\frac{3}{4}}, \quad 0 \leq j \leq FL \quad \dots (2)$$

然后,快速富立叶变换单元3执行256点快速富立叶变换,以产生频谱幅度值,然后,该频谱幅度值然后被频带分割单元4分割成例如18个频带。这些频带的频率区域作为例子被显示在表1中:

表 1

频 带	频 率 区 域
0	0到125Hz
1	125到250Hz
2	250到275Hz
3	375到563Hz
4	563到750Hz
5	750到938Hz
6	938到1125Hz
7	1125到1313Hz
8	1313到1563Hz
9	1563到1813Hz
10	1813到2063Hz
11	2063到2313Hz
12	2313到2563Hz
13	2563到2813Hz
14	2813到3063Hz
15	3063到3375Hz
16	3375到3688Hz
17	3688到4000Hz

由频谱分割产生的频带的幅值变为输入信号频谱的帧值 $Y[w, k]$ , 它们被输出给前述的各相应部分。

上述频率区域是以下面的事实为基础的,即频率越高,人的听力器官的感知分辨率就变得越差。对于各频带的幅度,使用相应频率区域中最大FFT幅度。

在噪声估计单元5中,已成帧信号  $y\text{-frame}_{j,k}$  的噪声被从语音中分离出来,被假定为噪声的帧被检测出,而最大 SNR 被提供到 NR 值计算单元6。噪声域估计或噪声帧检测由例如三种检测运算实现。现在对噪声域估计的一个说明例子进行解释。

RMS计算单元21计算信号每帧的RMS值并输出该被计算的RMS值。第k帧的RMS值,或  $RMS[k]$ , 由下列等式(3)计算:

$$RMS[k] = \sqrt{\frac{1}{FL} \sum_{j=0}^{FL-1} (y\text{-frame}_{j,k})^2} \quad \dots (3)$$

在相对能量计算单元22中,计算与来自在前帧的衰减能量,即  $dB_{rel}[k]$  相对应的第K帧的相对能量,并且该结果值被输出。该以 dB 为单位的相对能量即  $dB_{rel}[k]$  由下式求出:

$$dB_{rel}[k] = 10 \log_{10} \left( \frac{E_{decay}[k]}{E[k]} \right) \quad \dots (4)$$

而能量值 $E[k]$ 和衰减能量值 $E_{decay}[k]$ 则由下面的等式(5)和(6)求出:

$$E[k] = \sum_{l=1}^{FL} (y-frame_{j,k})^2 \quad \dots(5)$$

$$E_{decay}[k] = \max\left(E[k], \left(\exp \frac{-FT}{0.65 * FS}\right) * E_{decay}[k - 1]\right) \quad \dots(6)$$

等式(5)可根据等式(3)表示为  $FL * (RMS[k])^2$ 。当然,在等式(3)的计算期间由RMS计算单元21获得的等式(5)的值可被直接提供到相对能量计算单元算单元22。在等式(6)中,衰减时间设定为0.65秒。

图2示出了能量值 $E[k]$ 和衰减能量 $E_{decay}[k]$ 的说明示例。

最大RMS计算单元23求出并输出用于估计信号电平与噪声电平之比(即最大SN比率)的最大值所需要的一个最大RMS值。该最大RMS值  $MaxRMS[k]$ 可由等式(7)求出:

$$\text{MaxRMS}[k] = \max(4000, \text{RMS}[k], \theta * \text{MaxRMS}[k-1] + (1-\theta) * \text{RMS}[k])$$

...(7)

其中  $\theta$  是一个衰减常量. 对于  $\theta$ , 取其使最大RMS值以3.2 秒衰减到1/e的  $\theta$  值, 即  $\theta = 0.993769$ .

估计噪声电平计算单元24 求出并输出一适于估计背景噪声电平的最大RMS值. 该估计噪声电平值  $\text{minRMS}[k]$  是在当前时间点之前的五个局部最小值中的最小值. 这五个值满足等式(8):

$$\begin{aligned} &(\text{RMS}[k] < 0.6 * \text{MaxRMS}[k] \text{ 和} \\ &\text{RMS}[k] < 4000 \text{ 和} \\ &\text{RMS}[k] < \text{RMS}[k+1] \text{ 和} \\ &\text{RMS}[k] < \text{RMS}[k-1] \text{ 和} \\ &\text{RMS}[k] < \text{RMS}[k-2]) \text{ 或} \\ &(\text{RMS}[k] < \text{MinRMS}) \end{aligned}$$

...(8)

该估计噪声电平值  $\text{minRMS}[k]$  被如此设定, 以便升高语音的背景噪声反馈. 这种高噪声电平的升高率是指数的, 而固定升高率被用作低噪声电平, 以实现更显著的升高.

图3示出了RMS值RMS [k], 估计噪声电平值minRMS [k], 和最大RMS值MaxRMS [k]的示例。

最大SNR计算单元25使用最大RMS 值和估计噪声电平值通过下面的等式(9)估计和计算最大SN比率MaxSNR [k]:

$$\text{MaxSNR}[k] = 20 \log_{10} \left( \frac{\text{MaxRMS}[k]}{\text{MinRMS}[k]} \right) - 1 \quad \dots (9)$$

根据该最大SNR值MaxSNR, 计算表示相对能量噪声电平的范围在0到1的一个规一化参数NR\_电平. 对于NR\_电平, 使用下列函数:

$$\text{NR-level}[k] = \begin{cases} \left( \frac{1}{2} + \frac{1}{2} \cos \left( \pi \frac{\text{MaxSNR}[k] - 30}{20} \right) \right) \times (1 - 0.002 (\text{MaxSNR}[k] - 30)^2) & 30 < \text{MaxSNR}[k] \leq 50 \\ 0.0 & \text{MaxSNR}[k] > 50 \\ 1.0 & \text{MaxSNR}[k] : \text{othe} \end{cases} \quad \dots (10)$$

现在来说明对噪声频谱估计单元26的操作。在相对能量计算单元22, 估计噪声电平计算单元24和最大SNR计算单元25 中求出的各个值被用于从背景噪声中识别语音。如果下列条件成立:

$$\begin{aligned}
 & ((\text{RMS}[k] < \text{NoiseRMS}_{\text{thres}}[k]) \text{ 或} \\
 & (\text{dB}_{\text{rel}}[k] > \text{dB}_{\text{thres}}[k])) \text{ 和} \\
 & (\text{RMS}[k] < \text{RMS}[k-1]+200) \dots(11)
 \end{aligned}$$

其中

$$\begin{aligned}
 \text{NoiseRMS}_{\text{thres}}[k] &= 1.05+0.45*\text{NR\_level}[k]*\text{MinRMS}[k] \\
 \text{dB}_{\text{thres rel}}[k] &= \max(\text{MaxSNR}[k]-4.0, 0.9*\text{MaxSNR}[k])
 \end{aligned}$$

则第K帧的信号被划归为背景噪声。对如此划归的背景噪声的幅度进行计算并输出作为该噪声频谱的平均时间估计值 $N[w, k]$ 。

图4示出了图11中所示的以dB为单位的相对能量,即 $\text{dBrel}[k]$ ,最大 $\text{SNR}[k]$ ,和作为用于噪声识别的门限值之一的 $\text{dBthresrel}[k]$ 的说明性示例。

图6示出了作为等式(10)中的 $\text{MaxSNR}[k]$ 的函数的 $\text{NR\_level}[k]$ 。

如果第K帧的信号被划归为背景噪声或噪声,则噪声频谱的平均时间估计值 $N[w, k]$ 通过下列等式(12)由当前帧信号的输入信号频谱的幅度 $Y[w, k]$ 修改:

$$\begin{aligned}
 N[w, k] &= \alpha*\max(N[w, k-1], Y[w, k]) \\
 &+ (1 - \alpha)*\min(N[w, k-1], Y[w, k]) \dots(12)
 \end{aligned}$$

$$\alpha = \exp\left(\frac{-FI}{0.5*FS}\right)$$

其中W表示频带分离中的频带数。

如果第K帧的信号被划归为语音, N[W, K-1]的值被直接用于 N[w, k]。

NR值计算单元6计算NR[w, k], 它是一个用于防止滤波器响应突变的值, 并输出该产生的值NR[w, k]。该NR[w, k]是一个范围从0到1的值并由等式(13)确定:

$$NR[w, k] = \begin{cases} adj[w, k] & NR[w, k-1] - \delta_{NR} < adj[w, k] < NR[w, k-1] + \delta_{NR} \\ NR[w, k-1] - \delta_{NR} & NR[w, k-1] - \delta_{NR} \geq adj[w, k] \\ NR[w, k-1] + \delta_{NR} & NR[w, k-1] + \delta_{NR} \leq adj[w, k] \end{cases} \quad \dots(13)$$

在等式(13)中, adj[w, k]是用于考虑如下所述的效果的参数, 并由等式(14)确定:

$$\delta_{NR} = 0.004 \text{ and} \\ adj[w, k] = \min(adj1[k], adj2[k]) - adj3[w, k] \quad \dots(14)$$

在等式(14)中, adj1[k]是一个具有由下面所述的滤波以高SNR实现的抑制噪声效果的值, 并由下面的等式(15)定义:

$$adj1[k] = \begin{cases} 1 & MaxSNR[k] < 29 \\ 1 - \frac{MaxSNR[k] - 29}{40} & 29 \leq MaxSNR[k] < 43 \\ 0 & MaxSNR[k] : otherwise \end{cases} \dots (15)$$

在等式(14)中, adj2[k]是一个具有抑制效果的值, 由上述滤波操作实现的噪声抑制率相应于极低的噪声电平或极高的噪声电平, 并且该adj2[k]由下面的等式(16)确定:

$$adj2[k] = \begin{cases} 0 & MinRMS[k] < 20 \\ \frac{MinRMS[k] - 20}{40} & 20 \leq MinRMS[k] < 60 \\ 1 & 60 \leq MinRMS[k] < 1000 \\ 1 - \frac{MinRMS[k] - 1000}{1000} & 1000 \leq MinRMS[k] < 1800 \\ 0.2 & 1800 \leq MinRMS[k] \end{cases} \dots (16)$$

在等式(14)中, adj3[k]是一个在2375Hz和4000Hz 之间具有抑制最大噪声降低量从18dB到15dB的效果的值, 并由下面的等式(17)确定:

$$adj3[w, k] = \begin{cases} 0 & W < 2375\text{Hz} \\ \frac{0.059415(w-2375)}{4000-2375} & W: \text{其它情况} \end{cases} \dots (17)$$

而且, 可以看到, 以dB为单位以上的NR[w, k]的各值和最大噪声降低量之间的相对关系在dB范围内基本上是线性的, 如图6所示。

Hn值计算单元7根据分裂成多个频带的输入信号频谱的幅度Y[w, k]产生噪声谱N[w, k]的平均时间估计值和值NR[w, k], 值Hn[w, k]确定被构造用于从输入语音信号中消除噪声部分的滤波器特征。该值Hn[w, k]根据下面的等式(18)计算:

$$H_n[w, k] = 1 - (2 * NR[w, k] - NR^2[w, k]) * (1 - H[w] [S/N=\gamma]) \dots (18)$$

当SNR被固定在值r时, 上述等式(18)中的值H[W] [S/N=r] 等效于一噪声抑制滤波器的最佳特征, 并且由下面的等式(19)求出:

$$H[w] [S/N=\gamma] = \frac{1}{2} \left( 1 + \sqrt{1 - \frac{1}{X^2[w, k]}} \right) * P(H1 | Y_w) [S/N=\gamma] + G_{min} * P(H0 | Y_w) [S/N=\gamma] \dots (19)$$

而且, 该值可事先求出, 并根据Y[w, k]/N[w, k]的值列成一个表。此外, 等式(19)中的X[w, k]等效于Y[w, k]/N[w, k], 而Gmin是一个指示H[W] [S/N=r]的最小增益的参数。另一方面, P(Hi | Yw) [S/N=r]和P(H0 | Yw) [S/N=r]是定义幅度Y[w, k]的状态的参数, P(H1 | Yw) [S/N=r]是一个定义其中语音分量和噪声分量被一起混合在Y[w, k]中的状态的参数, 而P(H0 | Yw) [S/N=r]是一个定义仅有噪声分量被包含在Y[w, k]中的参数。这些值根据下面的等式(20)计算:

$$\begin{aligned}
 P(H1 | Y_w) [S/N=r] &= 1 - P(H0 | Y_w) [S/N=r] \\
 &= \frac{P(H1) * (\exp(-\gamma^2)) * I_0(2 * \gamma * x[w, k])}{P(H1) * (\exp(-\gamma^2)) * I_0(2 * \gamma * x[w, k]) + P(H0) * (\exp(-x^2))}
 \end{aligned}$$

§§(…(20)

其中  $P(H1) = P(H0) = 0.5$

从等式(20)可以看到,  $P(H1 | Y_w) [S/N=r]$  和  $P(H0 | Y_w) [S/N=r]$  是  $X[w, k]$  的函数, 而  $I_0(2 * r * X[w, k])$  是一个贝塞尔函数, 并且相应于  $r$  和  $X[w, k]$  的值求出。  $P(H1)$  和  $P(H0)$  都被固定在 0.5。 通过如上所述对参数进行的简化, 处理量可被减小到大约通常方法的五分之一。

由  $H_n$  值计算单元 7 产生的  $H_n[w, k]$  值和  $X[w, k]$  值 (即比率  $Y[w, k] / N[w, k]$ ) 之间的相对关系是这样的, 对于高的比率  $Y[w, k] / N[w, k]$  的值, 即对于语音分量高于噪声分量的情况, 值  $H_n[w, k]$  被增加, 即抑制被减弱, 相反, 对于较低的比率  $Y[w, k] / N[w, k]$  的值, 即对于语音分量低于噪声分量的情况, 值  $H_n[w, k]$  被减小, 即抑制被增强。 在上面的等式中, 实线代表  $r=2.7$ ,  $G_{min} = -18\text{dB}$  和  $NR[w, k]=1$  的情况。 还可以看到, 说明上述关系的曲线在取决于  $NR[w, k]$  值的区域  $L$  内改变, 并且该  $NR[w, k]$  值的各自曲线以与  $NR[w, k]=1$  相同的趋势变化。

滤波单元 8 执行滤波, 沿频率轴和时间轴对  $H_n[w, k]$  进行平滑, 以便产生一被平滑信号  $H_{t\_smooth}[w, k]$  作为输出信号。 在沿频率轴方向的滤波具有减少信号  $H_n[w, k]$  的有效脉冲响应长度的效果。 这避免

了由于滤波器的实现导致的循环卷积在频域相乘而产生的混淆。在沿时间轴方向的滤波具有限制滤波器特征中变化率的效果,以抑制突发噪声的产生。

首先对沿频率轴方向的滤波进行说明。在每个频带的 $H_n[w, k]$ 上执行平均值滤波。本方法由下面的等式(21)和(22)表示:

$$\text{step 1: } H_1[w, k] = \max(\text{median}(H_n[w-i, k], H_n[w, k], H_n[w+1, k]), H_n[w, k]) \quad \dots (21)$$

$$\text{step 2: } H_2[w, k] = \min(\text{median}(H_1[w-i, k], H_1[w, k], H_1[w+1, k]), H_1[w, k]) \quad \dots (22)$$

在等式(21)和(22)中,如果 $(w-1)$ 或 $(w+1)$ 不存在,则 $H_1[w, k] = H_n[w, k]$ 和 $H_2[w, k] = H_1[w, k]$ 。

在步骤1中, $H_1[w, k]$ 是没有唯一或单独的零(0)频带的 $H_n[w, k]$ ,相反,在步骤2中, $H_2[w, k]$ 是没有唯一,单独或突出频带的 $H_1[w, k]$ 。以这种方式, $H_n[w, k]$ 被转换成 $H_2[w, k]$ 。

下面对沿时间轴方向的滤波进行说明。对于沿时间轴方向的滤波,对这样的情况进行考虑,即输入信号包含三个分量,即语音,背景噪声和表示语音的渐高部分的瞬态的瞬时状态。语音信号 $H_{\text{speech}}[w, k]$ 沿时间轴被平滑,如等式(23)所示:

$$H_{\text{speech}}[w, k] = 0.7 * H_2[w, k] + 0.3 * H_2[w, k-1] \quad \dots (23)$$

背景噪声沿时间轴被平滑, 如等式 (24) 所示:

$$H_{\text{noise}}[w, k] = 0.7 * \text{Min\_H} + 0.3 * \text{Max\_H} \quad \dots (24)$$

在上述的等式 (24) 中, Min\_H 和 Max\_H 可分别通过  $\text{Min\_H} = \min(H_2[w, k], H_2[w, k-1])$  和  $\text{Max\_H} = \max(H_2[w, k], H_2[w, k-1])$  求出。

在沿时间轴方向处在瞬态的信号不被平滑。

利用上述的被平滑信号, 通过等式 (25) 产生被平滑的输出信号

$H_{t\_smooth}$ :

$$H_{t\_smooth}[w, k] = (1 - \alpha_{tr}) (\alpha_{sp} * H_{\text{speech}}[w, k] + (1 - \alpha_{sp}) * H_{\text{noise}}[w, k]) + \alpha_{tr} * H_2[w, k] \quad \dots (25)$$

在上面的等式 (25) 中,  $\alpha_{sp}$  和  $\alpha_{tr}$  可以分别由等式 (26) 和等式 (27) 求出:

$$\alpha_{sp} = \begin{cases} 1.0 & SNR_{inst} > 4.0 \\ \frac{1}{3} (SNR_{inst} - 1) & 1.0 < SNR_{inst} < 4.0 \\ 0 & SNR_{inst} : otherwise \end{cases}$$

其中

... (26)

$$SNR_{inst} = \frac{RMS[k]}{MinRMS[k-1]}$$

$$\alpha_{sp} = \begin{cases} 1.0 & \delta_{rms} > 3.5 \\ \frac{2}{3} (\delta_{rms} - 2) & 1.0 < \delta_{rms} < 3.5 \\ 0 & \delta_{rms} : otherwise \end{cases} \dots (27)$$

其中

$$\delta_{rms} = \frac{RMS_{local}[k]}{RMS_{local}[k-1]}$$

然后, 在频带转换单元9中, 来自滤波单元8的18个频带的平滑信号  $H_{t.smooth}[w, k]$  通过内插例如一个要被输出的128-频带信号  $H_{128}[w, k]$  被扩展。

$$RMS_{local}[k] = \sqrt{\frac{1}{FI} \sum_{j=\frac{FI}{2}}^{FL-\frac{FI}{2}} (y_{frame_{j,k}})^2}$$

这种转换例如通过两级执行, 而从18扩展到64个频带和从64扩展到128个频带分别通过零阶保持和低通滤波器型内插来完成。

频谱校正单元10然后将利用FFT单元3获得的已成帧信号 $y$ -帧 $j, k$ 的快速富立叶变换得到的FFT的实部和虚部与上述的信号 $H_{128}[w, k]$ 相乘, 以执行频谱校正, 即噪声成分衰减, 合成的信号被输出。其结果是频谱幅度被校正, 而不改变相位。

然后, 反向FFT单元11对频谱校正单元10的输出信号执行反向FFT, 以便输出合成的被IFFT的信号。

重叠相加单元12重叠和相加基于帧的被IFFT 的信号的帧边缘部分。该合成的输出语音信号在语音信号输出端14被输出。

图8示出了按照本发明执行用于语音信号的噪声降低方法的另一个实施例。由图1所示的噪声降低装置公用的元件或部件用相同的数字表示, 为了简化删除了有关操作的描述。

该噪声降低装置拥有用于将输入语音信号变换成频域信号的快速富立叶变换单元3, 用于对从输入语音信号中消除噪声成分的滤波运算的滤波器特征进行控制的 $H_n$ 值计算单元7, 和用于根据由 $H_n$ 值计算单元7产生的滤波器特征通过滤波降低输入语音信号中的噪声的频谱校正单元10。

在拥有 $H_n$ 值计算单元7的噪声抑制滤波器特征发生单元35中, 频带分离部分4将自FFT单元3输出的频谱的幅度分离成例如18个频带, 并将基于频带的幅度 $Y[w, k]$ 输出给用于计算RMS, 估计噪声电平

和最大SNR的计算单元31, 以及输出给噪声频谱估计单元26和初始滤波器响应计算单元33。

计算单元31根据从成帧单元1输出的 $y$ -帧 $j, k$ 和由频带分离部分4输出的 $Y[w, k]$ 计算基于帧的RMS值 $RMS[k]$ , 估计噪声电平值 $MinRMS[k]$ 和最大RMS值 $Max[k]$ , 并将这些值传送到噪声频谱估计单元26和 $adj1, adj2,$ 和 $adj3$ 计算单元32。

初始滤波器响应计算单元33将从噪声频谱估计单元26输出的平均时间噪声值 $N[w, k]$ 和从频带分离部分4输出的 $Y[w, k]$ 提供给滤波器抑制曲线表单元34, 用于求出与存储在滤波器抑制曲线表单元34中的 $Y[w, k]$ 和 $N[w, k]$ 相对应的 $H[w, k]$ 的值, 以将如此求得的价值传送到 $H_n$ 值计算单元7。在滤波器抑制曲线表单元34中存储有一个 $H[w, k]$ 值的表。

由图1和8所示的噪声降低装置获得的输出语音信号被提供给一个信号处理电路, 诸如用于便携电话机的各种编码电路, 或提供给一个语音识别装置。可替换地, 可以对便携电话机的解码器输出信号执行噪声抑制。

图9和10分别示出了用本发明的噪声抑制方法实施噪声抑制获得的语音信号中的失真(以黑色所示) 和用通常的噪声抑制方法实施噪声抑制获得的语音信号中的失真(以白色所示)。在图9的图示中, 相对于每20ms取样的各分段的失真绘制了这些分段的SNR值。在图10的图示中, 相对于整个输入语音信号的失真绘制了各分段的

SNR值。在图9和10中,纵坐标代表随自源点的高度逐渐变小的失真,而横坐标代表向右逐渐变高的各分段的SNR。

从这些图中可以看出,与用通常的噪声抑制方法实施噪声抑制获得的语音信号相比,用本发明的噪声抑制方法实施噪声抑制获得的语音信号经受失真到一个很小的程度,尤其是在超过20的一个高SNR值的情况下。

说明书附图

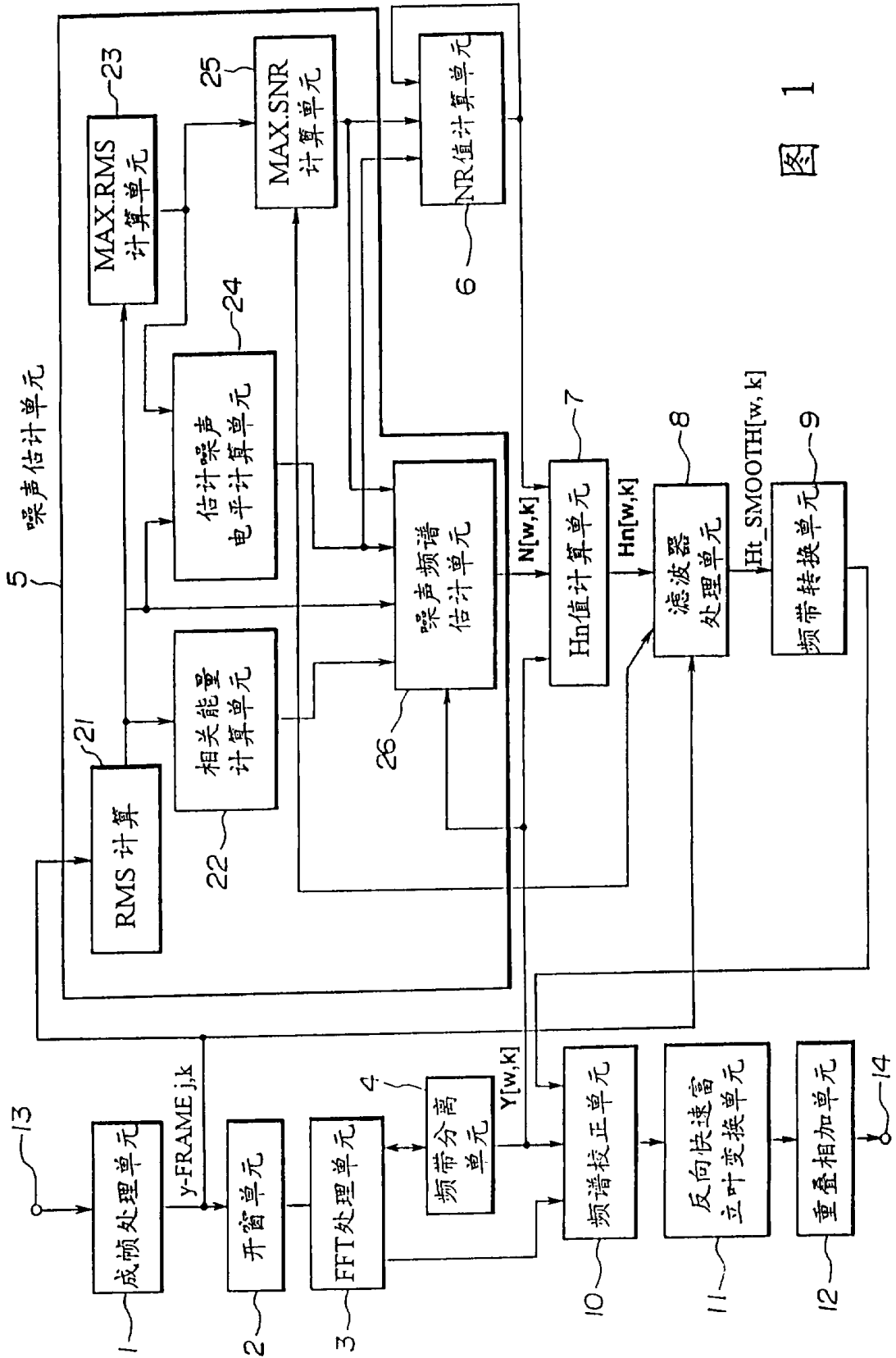


图 1

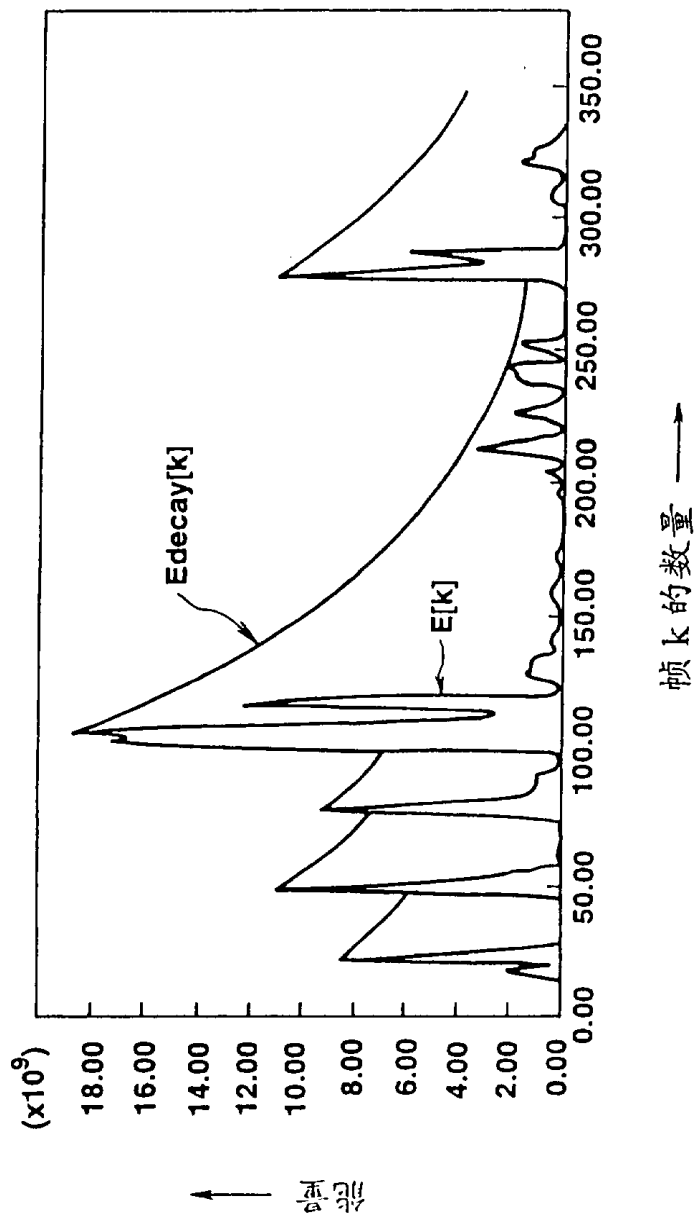
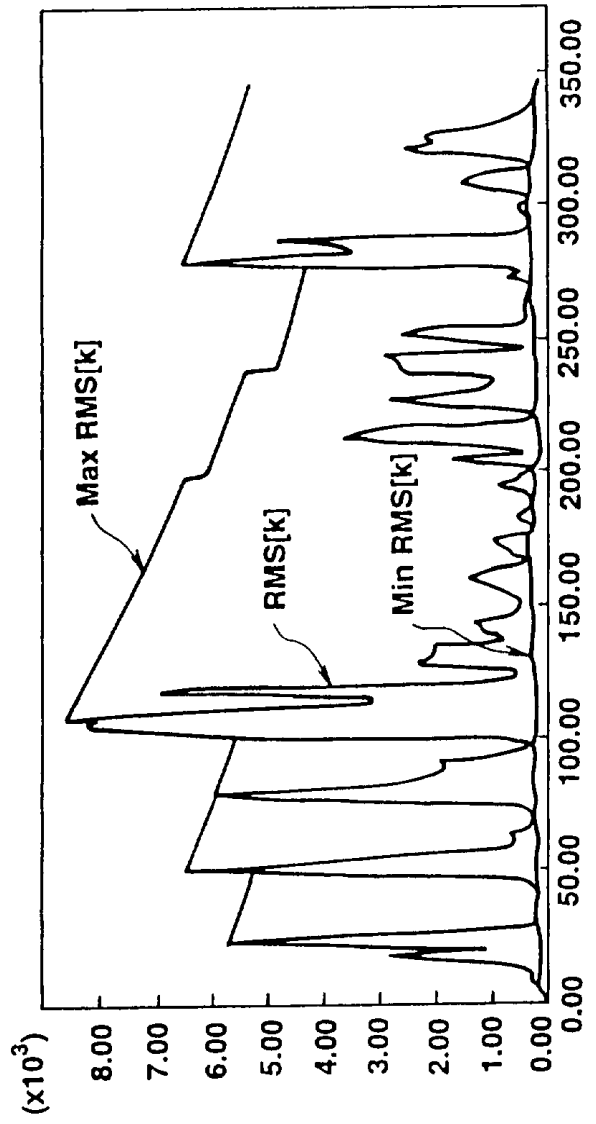


图 2



RMS 电平  $\uparrow$

帧 k 的数量  $\rightarrow$

图 3

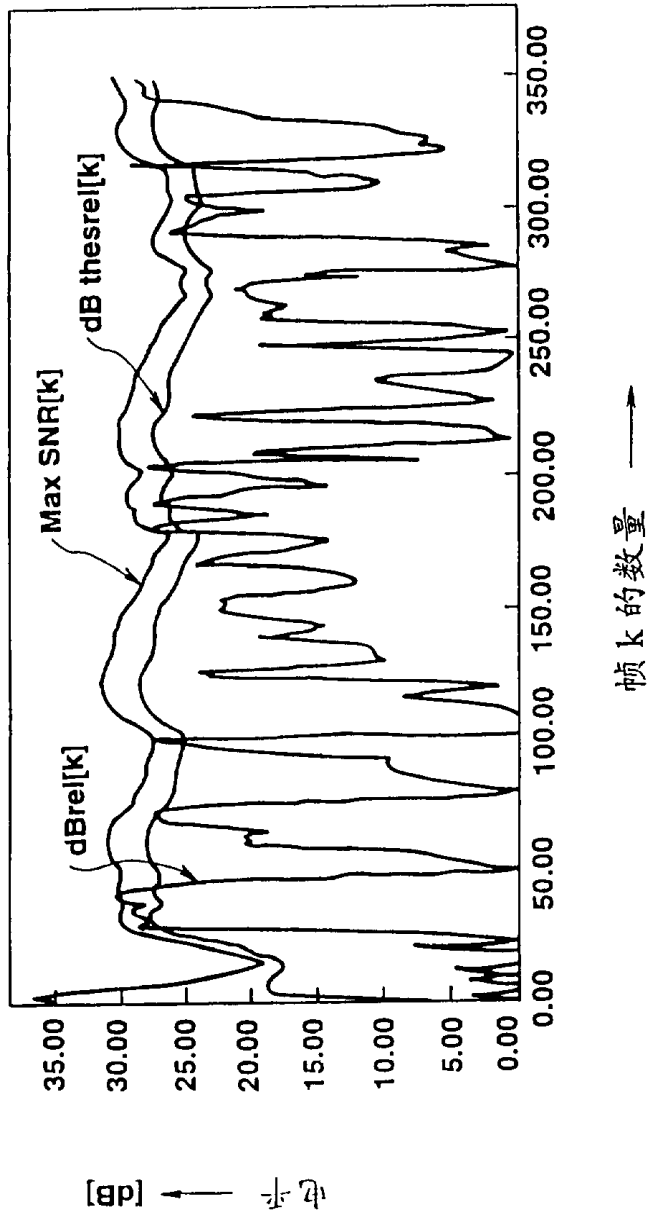


图 4

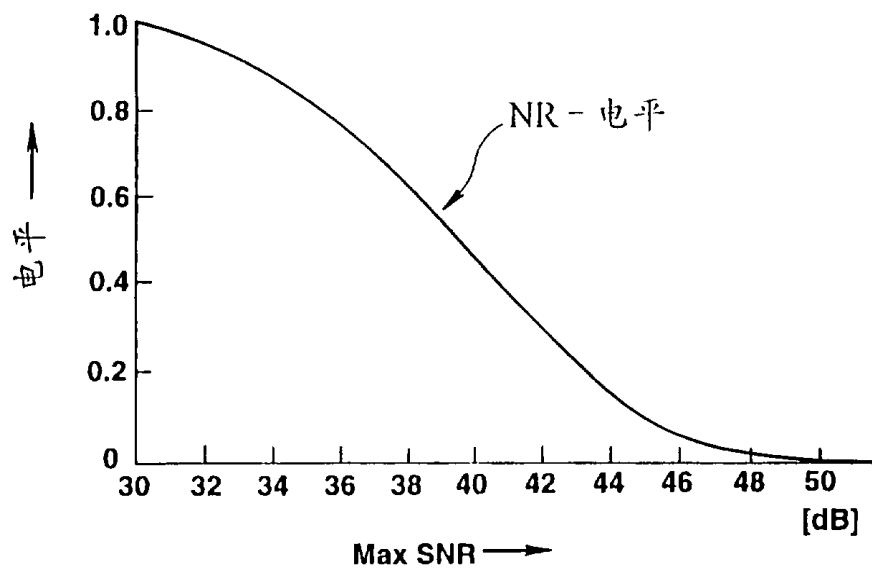


图 5

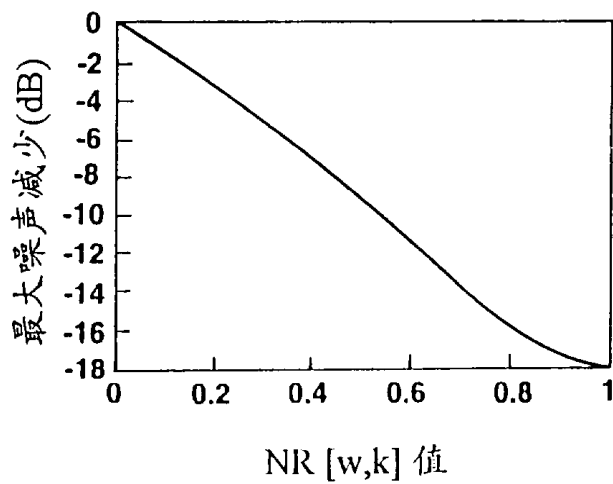


图 6

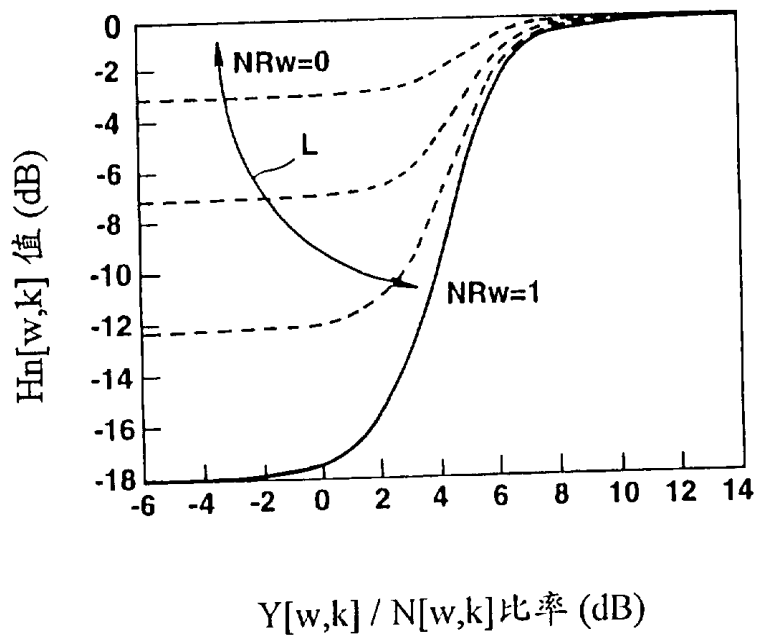


图 7

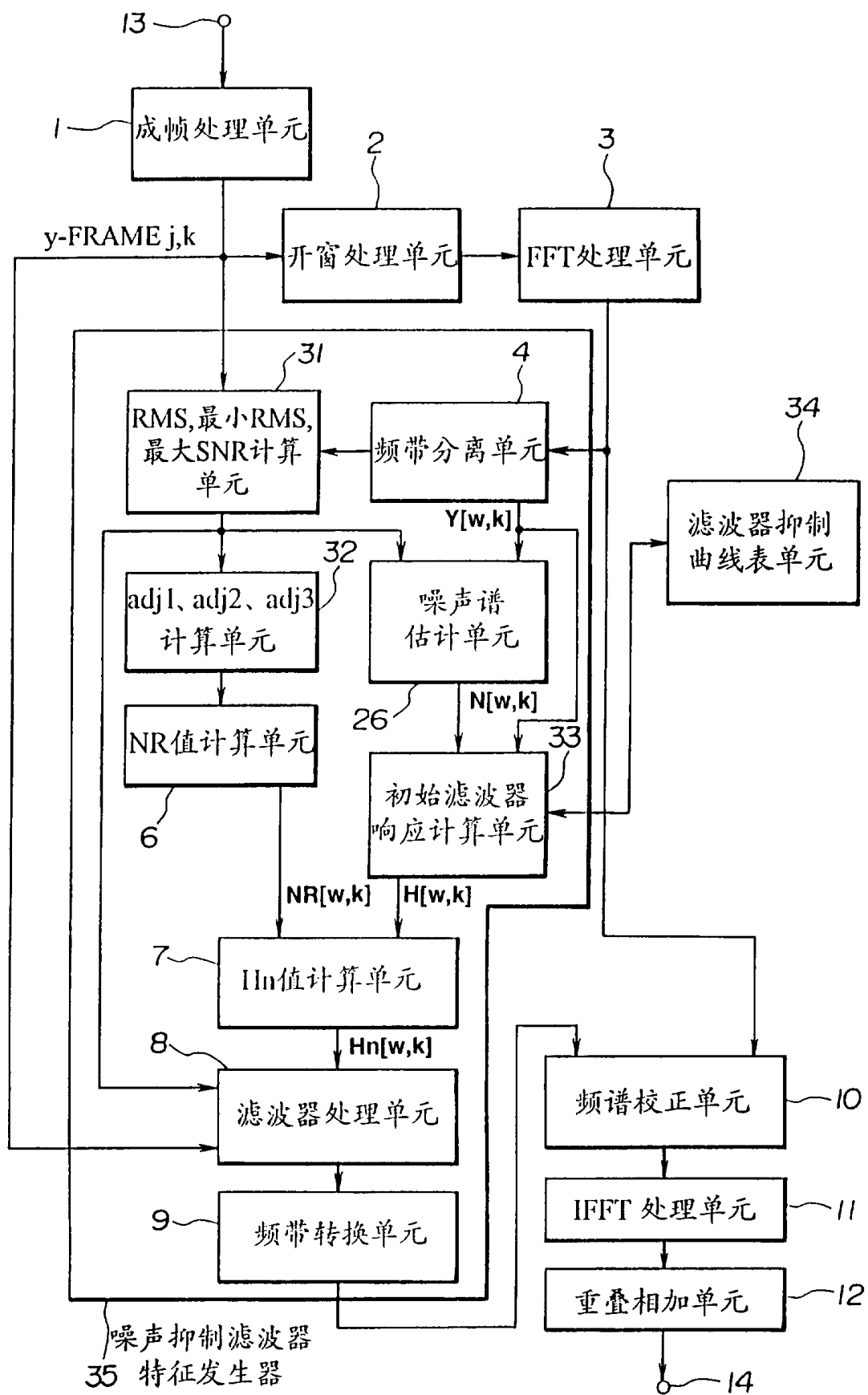


图 8  
7

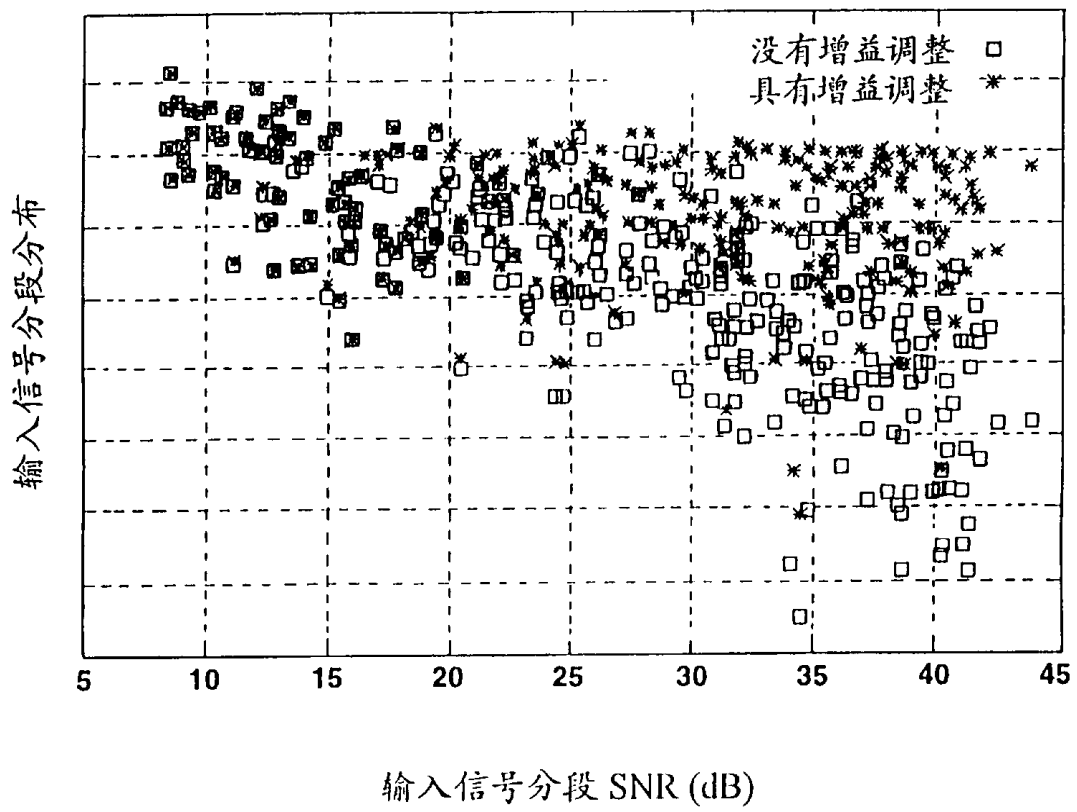


图 9

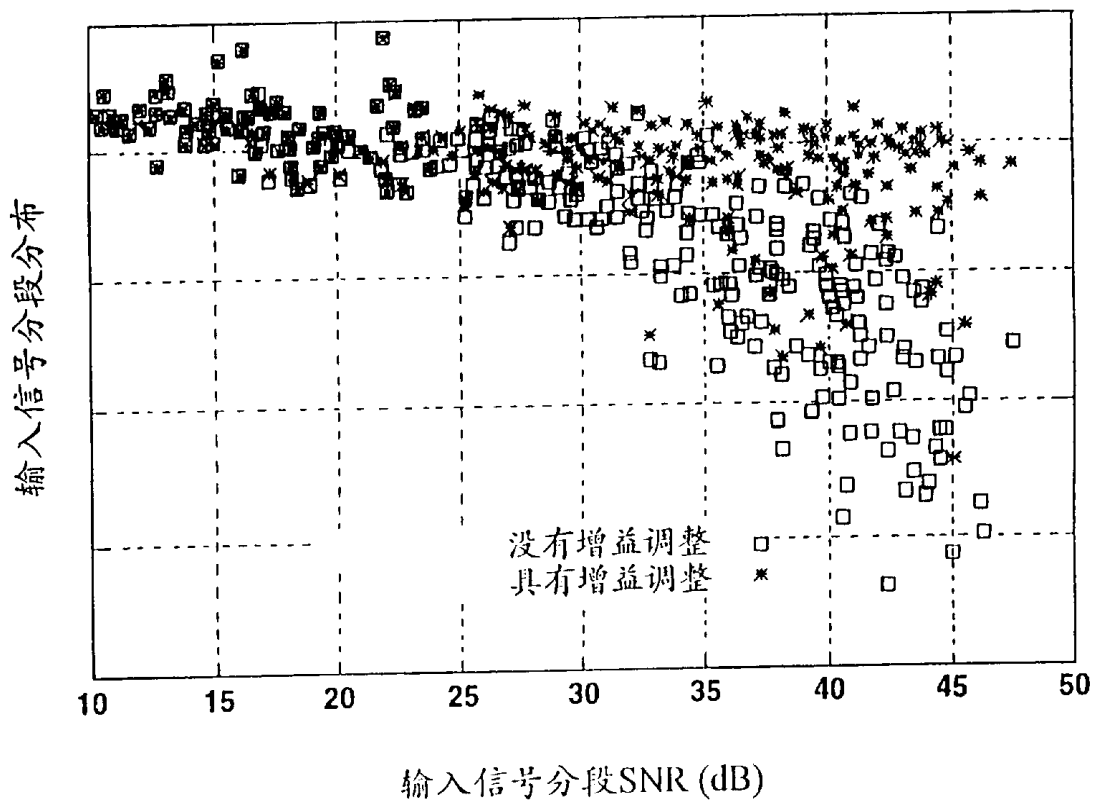


图 10