[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 02801102.3

[45] 授权公告日 2006 年 2 月 1 日

[11] 授权公告号 CN 1240051C

[22] 申请日 2002.3.25 [21] 申请号 02801102.3
[30] 优先权
[32] 2001. 4. 9 [33] EP [31] 01201304.1
[86] 国际申请 PCT/IB2002/001050 2002.3.25
[87] 国际公布 WO2002/082427 英 2002.10.17
[85] 进入国家阶段日期 2002.12.6
[71] 专利权人 皇家菲利浦电子有限公司
地址 荷兰艾恩德霍芬
[72] 发明人 E·F·吉吉
审查员 刘亚斌
[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 陈 霁

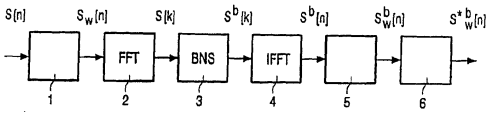
权利要求书 2 页 说明书 6 页 附图 3 页

[54] 发明名称

语音增强设备

[57] 摘要

用于减少背景噪声的语音增强系统包含将各音频信号的时域样值帧变换到频域的时间到频率变换单元(2)、在频域中执行噪声减少的背景噪声减少装置(3)，以及将噪声减少的信号变换回时域的频率到时间变换单元(4)。 在背景噪声减少装置(3)中对于每个频率分量，根据从时间到频率变换单元(2)的测量输入幅度并根据先前计算的背景幅度来计算预测背景幅度，由此对于该每个频率分量，根据该预测背景幅度并根据该测量输入幅度来计算信号噪声比，以及根据该信号噪声比来计算用于该测量输入幅度的滤波器幅度。 语音增强设备可应用于语音编码系统中，特别是 P^2CM 编码系统中。



1. 用于减少背景噪声的语音增强设备，包含：

将音频信号的时域样值帧变换到频域的时间到频率变换单元（2），在频域中执行噪声减少的背景噪声减少装置（3），以及

将噪声减少的音频信号从频域变换到时域的频率到时间变换单元（4），

其中该背景噪声减少装置（3）包含：背景电平更新模块（8），它根据来自时间到频率变换单元（2）的测量输入幅度 S[k]并且根据先前计算的背景幅度 $B_{-1}[k]$来计算当前音频信号帧中每个频率分量的预测背景幅度 B[k]；信噪比模块（9），它根据预测背景幅度 B[k]并根据该测量输入幅度 S[k]来计算该每个频率分量的信噪比 SNR[k]；以及滤波器更新模块（10），它根据信噪比 SNR[k]来为该每个频率分量计算对该测量输入幅度 S[k]的滤波器幅度 F[k]，其特征在于背景电平更新模块（8）包含：存储单元（20）来获得先前计算的背景幅度 $B_{-1}[k]$，处理装置（12－16）和比较器装置（17）以按照下一关系式更新前面预测的背景幅度：

$$B[k] = \max\{ \min\{ B'[k], B''[k]\}, B_{min} \},$$

这里 B_{min}为允许的最小背景电平，而

$B'[k]=B_{-1}[k] \cdot U[k]$和 $B''[k]= (B'[k] \cdot D[k])+(|S[k]| \cdot C \cdot (1-D[k]))$，

其中 U[k]和 D[k]是依赖于频率的缩放因子，而 C 是常数。

2. 按照权利要求 1 的语音增强设备，其特征在于，其中 U[k] = a + k/b，其中 a 和 b 为常数。

3. 按照权利要求 1 的语音增强设备，其特征在于，D[k] = c－k/d，其中 c 和 d 为常数。

4. 按照权利要求 1 的语音增强设备，其特征在于信噪比模块（9）包含用于根据预测背景幅度 B[k]并根据测量输入幅度 S[k]来按照下一关系式计算信噪比 SNR[k]的装置：

$SNR[k] = |S[k]|/B[k]$。

5. 按照权利要求 1 的语音增强设备，其特征在于滤波器更新模块（10）包含第一装置来计算一内部滤波器数值 F'[k]和第二装置来由该数值得到该测量输入幅度的滤波器幅度，该第一装置包含存储单元（31）来获得先前计算的内部滤波器幅度 $F'_{-1}[k]$并包含处理装置（21－23，25－27）来更新先前计算的内部滤波器幅度，这里

$F''[k] = F'_{-1}[k] \cdot E$

$\delta[k] = (1 - F''[k]) \cdot SNR[k]$ 以及

$F'[k] = F''[k]$， 若 $\delta[k] < 1$ ，或者

$F'[k] = F''[k] + G \cdot \delta[k]$,对于其他，

其中 E 和 G 为常数，

第二装置包含用于按照下一关系式对滤波器幅度进行缩放和限幅的比较器装置（28）：

$F[k] = \max\{\min\{H \cdot F'[k], 1\}, F_{min}\}$，

这里 H 是常数、F_{min}是最小滤波器数值而 F'[k]是内部滤波器数值。

6. 用于一语音编码系统的语音编码器，它配备有按照前述权利要求中任何一个的语音增强设备。

7. 语音编码系统，它配备了具有按照权利要求 1-5 中任何一个的语音增强设备的语音编码器。

8. 按照权利要求 7 的语音编码系统，其中该语音编码器是包含预处理器和 ADPCM 编码器的 P^2CM 编码器，该预处理器包括谱振幅扭曲装置，该语音增强设备具有集成于预处理器的谱振幅扭曲装置中的背景噪声减少装置（3）。

语音增强设备

本发明涉及用于减少背景噪声的语音增强设备，该设备包含将各帧的音频信号时域样值变换到频域的时间到频率变换单元、在频域中执行噪声减少的背景噪声减少装置和将噪声减少的音频信号从频域变换到时域的频率到时间变换单元。

这样的语音增强设备可应用于语音编码系统中，该系统比如可用于像数字电话回复机器和话音邮件应用那样的存储应用、可用于比如“车内”导航系统中的话音响应系统和用于如互联网电话的通信应用。

为了增强有噪声语音记录的质量，必须知道噪声的电平。对于单个麦克风记录，只能获得有噪声的语音。噪声电平必须仅仅从这一个信号估计出来。测量噪声的一种方式是使用没有语音活动的记录区域，并且将语音活动期间样值帧的频谱和非语音活动期获得的比较而用后者来更新前者。比如参见 US-A-6,070,137。这种方法的问题在于必须使用语音活动检测器。但是即便是在信号噪声比值相对较高时，也很难建立一个能很好工作的鲁棒性的语音检测器。另一个问题在于非语音活动区域可能非常短或甚至就不出现。当噪声是非平稳时，在语音活动期间其特征会改变，这就使这种方法甚至更加困难了。

已知的还可使用一个统计模型，该模型测量信号中每个谱份量的方差但不采用语音或非语音的二进制选择；参见：Ephraim，Malah 在 1984 年 12 月的 IEEE Trans. On ASSP 期刊第 32 卷第 6 期上发表的论文“使用 MMSE 短时谱振幅估计器的语音增强”（“Speech Enhancement Using MMSE Short-Time Spectral Amplitude Estimator”）。这种方法的问题在于，当背景噪声是非平稳时，估计必须基于最相邻的时间帧。在语音出现的长度内某些区域的语音频谱总高于实际噪声电平。对于这些频谱区域这就导致噪声电平的错误估计。

本发明的目的是预测单个麦克风语音记录中的背景噪声电平，但不使用语音活动检测器并且可显著减少噪声电平的错误估计。

因此，按照本发明，如开始段落描述的语音增强设备，其特征在于背景噪声减少装置包含：背景电平更新模块，它根据来自时间到频率变换单元的测量输入幅度 $S[k]$ 并且根据先前计算的背景幅度 $B_{-1}[k]$ 来计算

当前音频信号帧中每个频率分量的预测背景幅度 B[k]；信噪比模块，它根据该预测背景幅度 B[k]并根据该测量输入幅度 S[k]来计算该每个频率分量的信噪比 SNR[k]；以及滤波器更新模块，它根据信噪比 SNR[k]来为该每个频率分量计算对该测量输入幅度 S[k]的滤波器幅度 F[k]。

本发明还涉及配备了按照本发明的语音增强设备的语音编码系统，并涉及用于这种语音编码系统特别是 P^2CM 音频编码系统的语音编码器。特别是该 P^2CM 音频编码系统的编码器配备了带有上述语音增强系统的自适应差分脉冲编码调制（ADPCM）编码器和预处理器单元。

参考此后描述的附图和实施方案可清楚和说明本发明的这些和其他方面。在附图中：

图 1 示意了带有按照本发明的独立的背景噪声减法器（BNS）的语音增强设备的基本框图；

图 2 示意了 BNS 中的成帧和加窗；

图 3 是 BNS 中频域自适应滤波的框图；

图 4 是 BNS 中背景电平更新的框图；

图 5 是 BNS 中滤波器更新的框图；以及

图 6 示意了被背景噪声污染的发声的语音片断和测量背景电平，以及频域滤波结果。

举一个例子，在语音增强设备中，其音频输入信号被分成如 10 毫秒的帧。按如 8 kHz 的抽样频率，一帧包含 80 个样值。每个样值用如 16 个比特来表示。

BNS 基本上是一个频域自适应滤波器。在实际滤波前，语音增强设备的输入帧必须变换到频域。在滤波后，频域信息变换回时域。必须要特别注意防止帧边界出现不连续，因为 BNS 的滤波器特征会随时间而变化。

图 1 示意了带有 BNS 的语音增强设备的框图。语音增强设备包含输入窗形成单元 1、FFT 单元 2、背景噪声减法器（BNS）3、反 FFT（IFFT）单元 4、输出窗形成单元 5 以及重叠和相加单元 6。在这个例子中输入窗形成单元 1 的 80 样值输入帧移进两倍于帧长，即 160 样值的缓冲器，构成输入窗 s[n]。该输入窗用正弦窗 w[n]来加权。在本例中用 256 点 FFT2 来计算谱 S[k]。BNS 模块 3 对该谱应用频域滤波。得到的 $S^6[k]$ 用

IFFT4 变换回时域。这就得到时域表示 $s^b[n]$。在单元 5 中时域输出用输入处使用的相同正弦窗来加权。用正弦窗两次加权的净结果是用汉宁（Hanning）窗加权。单元 5 的输出用 $s^b_w[n]$表示。汉宁窗是下一处理模块 6，即重叠和相加，优选的窗类型。重叠和相加用于在两个连续的输出帧之间得到平滑变换。对于帧“i”，重叠和相加单元 6 的输出表示为：

$$s^{*b}_{w,i}[n] = s^b_{w,i}[n] + s^b_{w,i-1}[n+80] \quad \text{有 } 0 \leqslant n \leqslant 80$$

图 2 示意了采用的成帧和加窗。语音增强设备的输出是总的延迟为一帧，即本例中为 10 毫秒，的输出信号的处理后版本。

图 3 示意了频域中自适应滤波的框图，包含幅度模块 7、背景电平更新模块 8、信噪比模块 9、滤波器更新模块 10 和处理装置 11。其中对频谱 S[k]的每个频率分量 k 应用下列操作。首先，在幅度模块 7 中用下列关系式计算绝对幅度|S[k]|

$$|S[k]| = [\ (R\{S[k]\})^2 + (I\{S[k]\})^2\ ]^{1/2}$$

这里 R{S[k]}和 I{S[k]}分别是频谱的实部和虚部，本例中 $0 \leqslant k < 129$。然后，背景电平更新模块用输入幅度|S[k]|来计算当前帧的预测背景幅度 B[k]。

信噪比（SNR）用下式计算：

$$SNR[k] = |S[k]|/B[k]$$

并且滤波器更新模块 10 用其计算滤波器幅度 F[k]。

最后，用下列公式进行滤波：

$$R^b\{S^b[k]\} = R\{S[k]\}\cdot F[k] \text{ 和 } I^b\{S^b[k]\} = I\{S[k]\}\cdot F[k]$$

假设背景噪声的总相位贡献在频谱的实部和虚部均匀分布以致于频域振幅的本地减少也减少了添加的相位信息。然而，只改变背景信号的振幅谱而不改变相位分布是否就足够还值得讨论。如果背景只包含周期信号，就很容易测量其振幅和相位分量，并向合成信号添加相同的周期性和振幅，但有 180°的相位旋转。因为在分析期间内的有噪信号的相位分布不是恒定的并且因为只测量信噪比，所以能做的一切就是对每个频率区域用分别的因子抑制输入信号的能量。这通常不仅会抑制背景能量还会抑制语音信号的能量。然而，对感知语音信号很重要的成分通常具有比其他区域要大的信噪比，以致于实际当中本方法已足够高效了。

图 4 更详细地示意了背景电平更新模块 8。模块 8 包含处理装置 12

－16、含比较器 18 和 19 的比较器装置 17 以及存储器单元 20。

背景电平按下列步骤更新:

－首先,经由存储器单元 20 和处理装置 14,前面的背景电平值 $B_{-1}[k]$ 增加因子 U[k]得到 B'[k]。

－然后结果与 B''[k]值相比较，后者是增加的背景电平 B'[k]与经由处理装置 12、13、15 和 16 得到的当前绝对输入电平|S[k]|的按比例合并。通过比较器 18，选择两者中较小的一个作为背景电平 B'''[k]的候选。

－最后，通过比较器 19，用允许的最小背景电平 Bmin 来限制背景电平 B'''[k]，从而得到新的背景电平。这也是背景电平更新模块 8 的输出。

因此，计算的背景幅度可用下一关系式表示:

$$B[k] = \max\{ \min\{ B'[k], B''[k]\}, Bmin \}$$

这里 Bmin 为允许的最小背景电平，而

$B'[k] = B_{-1}[k] \cdot U[k]$ 和 $B''[k] = (B'[k] \cdot D[k]) + (|S[k]| \cdot C \cdot (1-D[k]))$

其中 U[k]和 D[k]是依赖于频率的缩放因子，而 C 是常数。

在本实施方案中输入比例因子 C 设置为 4。Bmin 设置为 64。缩放函数 U[k]和 D[k]对每帧都不变并且只依赖于频率下标 k。这些函数定义为:

$$U[k] = a + k/b \quad 和 D[k] = c - k/d$$

这里 a 可设置为 1.002、b 设置为 16384、c 设置为 0.97 而 d 设置为 1024.

图 5 更详细地示意了滤波器更新模块 10。模块 10 包含处理装置 21－27、包含比较器 29 和 30 的比较器装置 28 以及存储器单元 31。

模块 10 包含两级: 一级用于内部滤波器值 F'[k]的适配而一级用于输出滤波器值的缩放和限幅。内部滤波器值 F'[k]的适配是按照下列关系式将前一帧的向下缩放的内部滤波器值增加依赖于输入和滤波器电平的步长值:

$$F''[k] = F'_{-1}[k] \cdot E$$

$$\delta[k] = (1 - F''[k]) \cdot SNR[k] \quad 以及$$

$F'[k] = F''[k]$， 若 $\delta[k] < 1$ ，或者 $F'[k] = F''[k] + G \cdot \delta[k]$ 对于其他

这里 E 可设置为 0.9375 而 G 可设置为 0.0416。

用下式对输出滤波器值进行缩放和限幅:

$$F[k] = \max\{\min\{H \cdot F'[k], 1\}, F_{min}\}$$

这里 H 可设置为 1.5 而 F_{min} 可设置为 0.2。

对输出滤波器进行额外缩放和限幅的原因是想使滤波器具有比背景的能量明显要高的谱区域带通特征。

图 6 对于受到背景噪声污染的一帧发声的语音片断，示意了背景电平和滤波器更新模块的输出。

如上所述的带有独立的背景噪声减法器（BNS）的语音增强设备可应用于语音编码系统特别是 P^2CM 编码系统的编码器中。该 P^2CM 编码系统的编码器包含预处理器和 ADPCM 编码器。在编码前预处理器修改音频输入信号的信号谱，特别是通过应用振幅扭曲，比如像 R. Lefebre 和 C.Laflamme 在 1997 年 ICASSP 第 1 卷第 335 到 338 页上发表的论文“用于音频编码中噪声谱整形的谱振幅扭曲(SAW)”(“Spectral Amplitude Warping(SAW) for Noise Spectrum Shaping in Audio Coding”)描述的那样。因为这种振幅扭曲是在频域中进行的，所以背景噪声减少可集成到预处理器中。在时间到频率变换后相继实现背景噪声减少和幅度扭曲，这之后可进行频率到时间变换。在这种情况下，语音增强设备的输入信号由预处理器的输入信号构成。在预处理器中，以产生的信号中可获得噪声减少的方式来变化此输入信号，这样对噪声减少了的信号进行扭曲。根据该输入信号而获得的预处理器输出构成输入帧的延迟版本并将其提供给 ADPCM 编码器。本例中为 10 毫秒的这一延迟基本上是源于 BNS 的内部处理。ADPCM 编码器的其他输入信号由编译码器模式信号构成，该编译码器模式信号决定 ADPCM 编码器的比特流输出中码字的比特分配。ADPCM 编码器对于预处理的信号帧中的每个样值产生一个码字。然后将该码字分组成本例中为 80 个码的帧。根据选择的编译码器模式，得到的比特流可具有比如 11.2、12.8、16、21.6、24 或 32kbit/s 的比特率。

上述的实施方案由算法来实现，该算法的形式可以是能在 P^2CM 音频编码器中的信号处理装置上运行的计算机程序。在迄今为止示意了执行特定可编程功能的单元的部分附图中，这些单元必须视为计算机程序的子部分。

这里描述的本发明并不受限于所描述的实施方案。可能会对其进行修改。特别是可注意到，给出的 a、b、c、d、E、G 和 H 的数值只是举例；也可能给出其他数值。

5

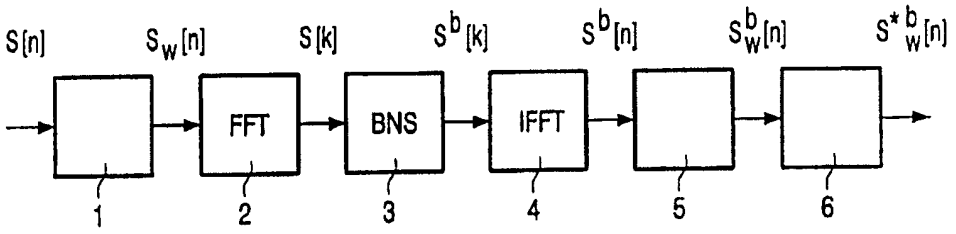

S[n]
$S_w[n]$
S[k]
$S^b[k]$
$S^b[n]$
$S^b_w[n]$
$S^{*b}_w[n]$
FFT
BNS
IFFT
1
2
3
4
5
6

图 1

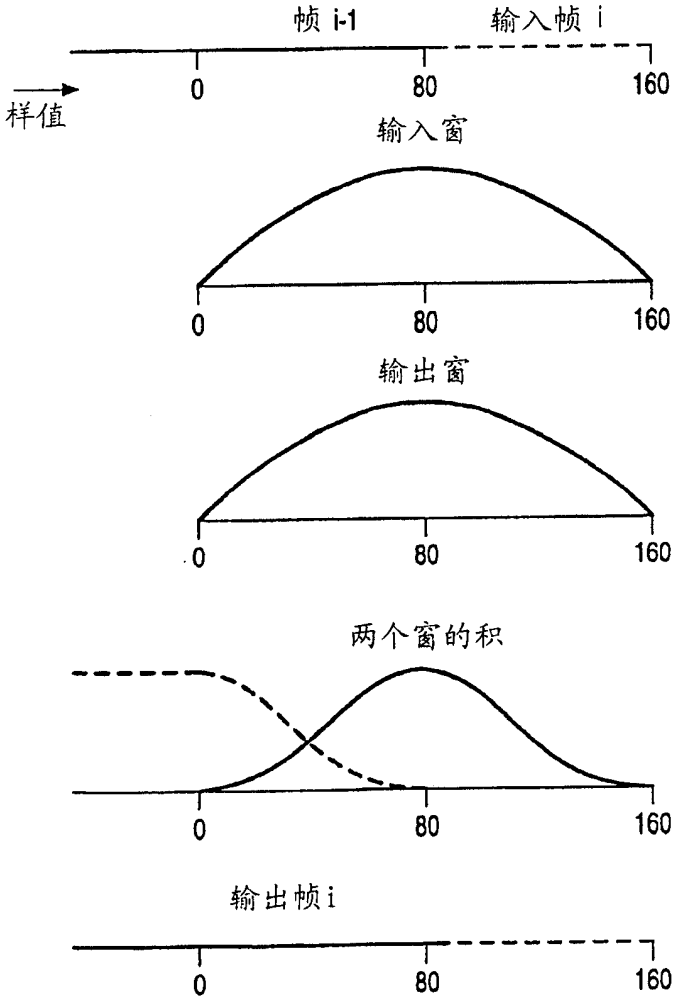

帧 i-1
输入帧 i
样值
0
80
160
输入窗
输出窗
两个窗的积
输出帧i

图 2

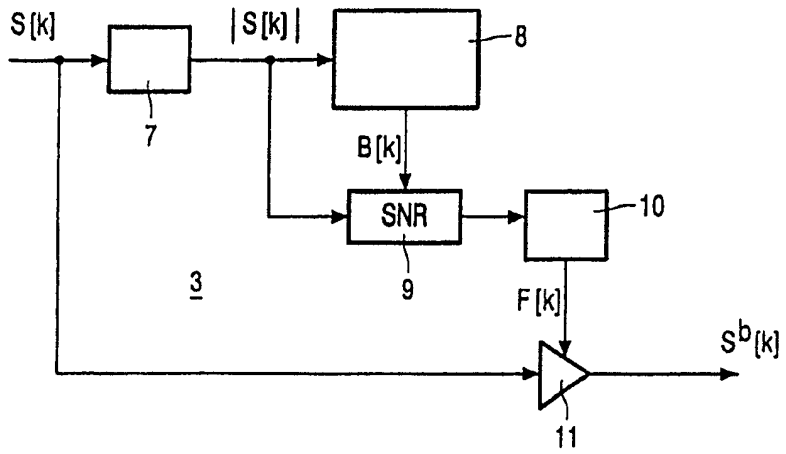

S[k]
|S[k]|
8
7
B[k]
SNR
10
9
3
F[k]
S^b[k]
11

图 3

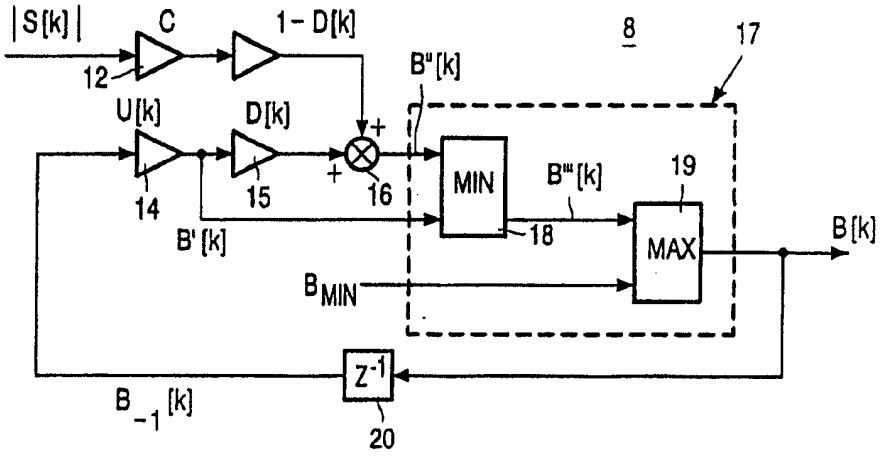

|S[k]|
C
1-D[k]
8
17
12
B''[k]
U[k]
D[k]
+
+
+
MIN
B'''[k]
19
14
15
16
18
B'[k]
MAX
B[k]
B_MIN
Z^-1
B_-1[k]
20

图 4

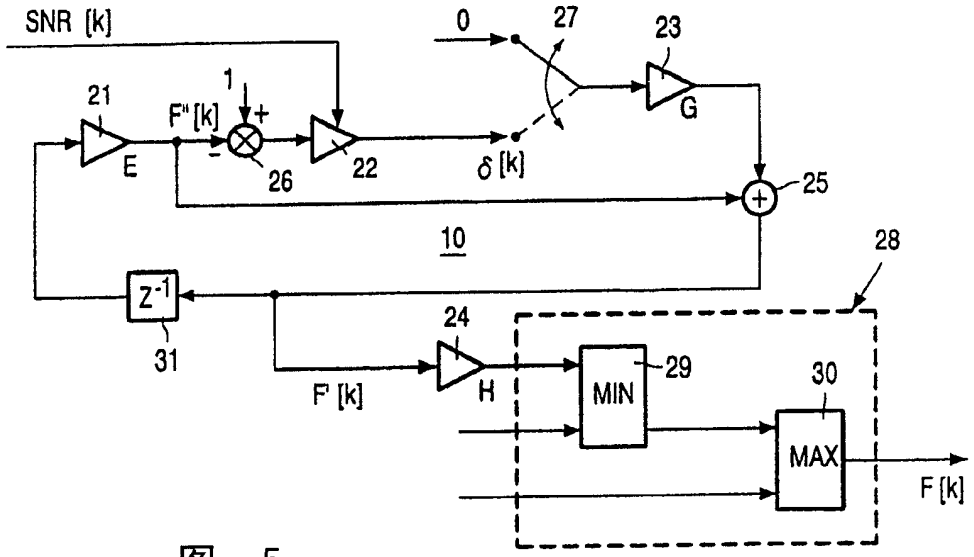

SNR [k]
0
27
23
21
1
F" [k]
+
G
E
26
22
δ [k]
25
10
28
Z-1
31
24
F' [k]
H
MIN
29
30
MAX
F [k]

图 5

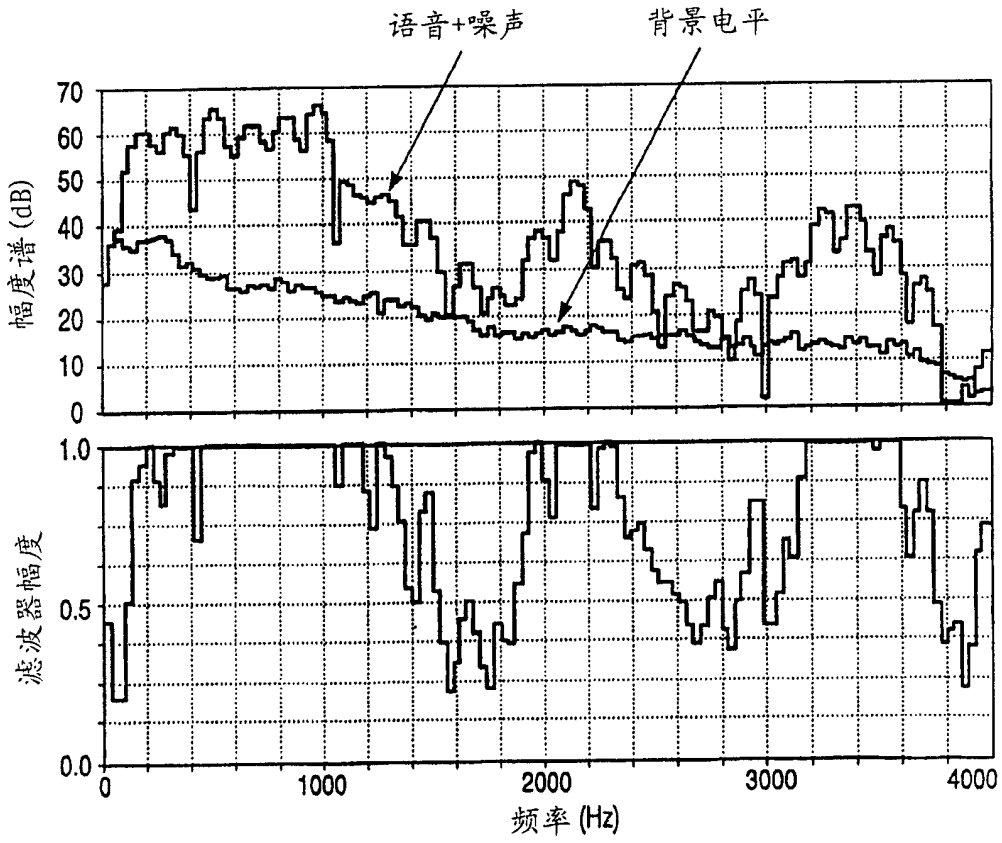

语音+噪声
背景电平
幅度谱 (dB)
70
60
50
40
30
20
10
0
滤波器幅度
1.0
0.5
0.0
0
1000
2000
3000
4000
频率 (Hz)

图 6