[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
G10L 21/02

[12] 发明专利说明书

[21] ZL 专利号 99809229.0

[45] 授权公告日 2004 年 4 月 14 日

[11] 授权公告号 CN 1145931C

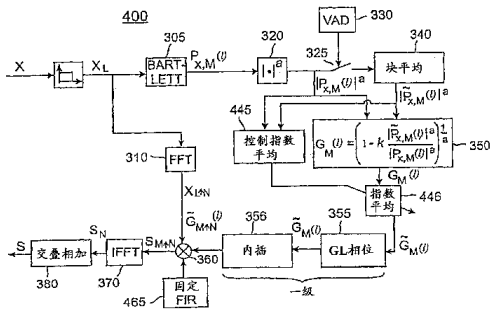
[22] 申请日 1999.5.27 [21] 申请号 99809229.0
[30] 优先权
[32] 1998. 5.27 [33] US [31] 09/084,387
[86] 国际申请 PCT/SE1999/000899 1999.5.27
[87] 国际公布 WO99/62054 英 1999.12.2
[85] 进入国家阶段日期 2001.1.31
[71] 专利权人 艾利森电话股份有限公司
地址 瑞典斯德哥尔摩
[72] 发明人 H·古斯塔夫松 I·克莱松
S·诺尔德霍尔姆
审查员 杨艳兰
[74] 专利代理机构 中国专利代理(香港)有限公司
代理人 邹光新 李亚非

权利要求书 4 页 说明书 23 页 附图 15 页

[54] 发明名称 降低语音信号中的噪声的方法和采用该方法的系统和电话机

[57] 摘要

本发明提出的在降噪系统中提供语音增强的方法和设备包括频谱减除算法，利用了线性卷积、因果滤波和/或根据频谱对频谱减除增益函数执行指数平均。 按照典型实施例，所开发的一些低阶谱估计与传统的频谱减除系统中的谱估计相比具有较低的频率分辨率和较小的方差。 这种低阶频谱用来形成一个具有所需的小方差的增益函数，从而降低了在频谱减除输出信号中的乐音。 有益的是，这种增益函数可以利用取决于频谱的指数平均进一步就一些块予以平滑。 此外，由于增益函数阶数低，就允许在内插期间添加一个相位，使得频谱减除增益滤波器成为因果的，从而防止了块间间断。



ISSN1008-4274

知识产权出版社出版

1. 一种降噪系统，所述降噪系统包括一个配置成对一个有噪声的语音输入信号（X）进行滤波从而提供一个噪声降低了的语音输出信号（S）的频谱减除处理器（300，400），其中所述噪声降低了的语音输出信号的一个样点块（S_{MIN}）根据所述输入信号的一个相应样点块（X_{LIN}）计算，所述降噪系统的特征在于所述处理器包括:

将增益函数加到所述输入信号一个相应样点块（X_{LIN}）的装置，其中增益函数是作为增益函数的样点块加上去的，而增益函数则根据所述输入信号的谱密度估计和所述输入信号的噪声分量的谱密度估计计算；以及

一个将相位添加到增益函数上的装置,使得所述频谱减除处理器提供因果滤波，

其中所述计算得的语音输出信号样点块（S_{MIN}）的阶大于所述语音输入信号的相应样点块（X_{LIN}）的阶与所述增益函数的相应样点块的阶之和。

2. 权利要求1的降噪系统，其中所述计算得的语音输出信号样点块（S_{MIN}）是根据所述语音输入信号的相应样点块（X_{LIN}）与所述增益函数的相应样点块的卷积计算的。

3. 权利要求1的降噪系统，其中语音输出语音信号样点块（S_{MIN}）有N个样点,语音输入信号样点的块（X_{LIN}）有L个输入样点，其中L小于N。

4. 权利要求1的降噪系统，其中一个语音输出信号样点块（S_{MIN}）有N个输出样点，而增益函数有M个样点,其中M小于N。

5. 权利要求1的降噪系统，其中一个输出信号样点块（S_{MIN}）有N个输出语音信号，而输入信号样点块（X_{LIN}）有M个增益函数,其中L与M之和小于N。

6. 权利要求5的降噪系统，其中所述有L个输入信号样点的块通过补零提供一个有N个输入信号样点的块（X_{LIN}），所述有N个输出信号样点的块（S_{MIN}）是根据这个有N个输入信号样点的块（X_{LIN}）计算的。

7. 权利要求5的降噪系统，其中所述有M个增益函数样点的块通过内插（356）提供一个有N个增益函数样点的块，所述有N个输出信号样点的块（S_{MIN}）是根据这个有N个增益函数样点的块计算的。

8. 权利要求5的降噪系统，其中所述有M个增益函数样点的块是根据所述L个输入信号样点通过谱估计计算的。

9. 权利要求8的降噪系统，其中所述谱估计是利用Bartlett方法（305）实现的

10. 权利要求8的降噪系统，其中所述谱估计是利用Welsh方法实现的。

11. 权利要求1的降噪系统，其中相继的输出信号块（S_{MIN}）是利用一个交叠相加方法（380）适配在一起的。

12. 权利要求1的降噪系统，其中所述增益函数具有线性相位。

13. 权利要求1的降噪系统，其中所述增益函数具有最小相位。

14. 一种对一个有噪声的语音输入信号（X）进行处理从而提供一个噪声降低了的语音输出信号（S）的方法，所述方法包括利用频谱减除根据所述有噪声的语音语音输入信号（X）和根据一个利用谱密度估计计算得的增益函数计算所述噪声降低了的语音输出信号（S）的步骤，其中所述噪声降低了的输出信号的一个样点块（S_{MIN}）根据所述语音输入信号的一个相应样点块（X_{LIN}）和所述增益函数的一个相应样点块计算，所述方法的特征是包括下列步骤：

计算所述语音输入信号的谱密度估计和所述语音输入信号的噪声分量的谱密度估计；以及

将一个相位（355）添加给所述增益函数从而使得所述利用频谱减除的步骤提供因果滤波，

其中所述计算得的语音输出信号的样点块（S_{MIN}）的阶大于所述语音输入信号的相应样点块（X_{LIN}）的阶与所述增益函数的相应样点块的阶之和。

15. 权利要求14的方法，所述方法包括将所述计算得的语音输出信号样点块（S_{MIN}）计算为所述语音输入信号的相应样点块（X_{LIN}）与所述增益函数的相应样点块的卷积的步骤。

16. 权利要求14的方法，所述方法包括用一个有L个语音输入信号样点的块计算一个有N个语音输出信号样点的块（S_{MIN}）的步骤，其中L小于N。

17. 权利要求14的方法，所述方法包括用一个有M个增益函数样点的块计算一个有N个输出信号样点的块（S_{MIN}）的步骤，其中M小于N。

18. 权利要求14的方法，所述方法包括用一个有L个输入信号样点的块和一个有M个增益函数样点的块计算一个有N个输出信号样点的块（S_{MIN}）的步骤，其中L与M之和小于N。

19. 权利要求18的方法，所述方法包括对所述有L个输入信号样点的块补零从而提供一个用来计算所述有N个输出信号样点的块（S_{MIN}）的有N个输入信号样点的块（X_{LIN}）的步骤。

20. 权利要求18的方法，所述方法包括对所述有M个增益函数样点的块进行内插（356）从而提供一个用来计算所述有N个输出信号样点的块（S_{MIN}）的有N个增益函数样点的块的步骤。

21. 权利要求18的方法，所述方法包括根据所述L个输入信号样点利用谱估计计算所述有M个增益函数样点的块的步骤。

22. 权利要求21的方法，其中所述利用谱估计的步骤是利用Bartlett算法（305）实现的。

23. 权利要求21的方法，其中所述利用谱估计的步骤是利用Welsh算法实现的。

24. 权利要求14的方法，所述方法包括利用一个交叠相加方法（380）将相继的输出信号块（S_{MIN}）适配在一起的步骤。

25. 权利要求14的方法，其中所述增益函数具有线性相位。

26. 权利要求14的方法，其中所述增益函数具有最小相位。

27. 一种移动电话机，所述移动电话机包括一个配置成对一个有噪声的近端语音信号（X）进行滤波从而提供一个噪声降低了的近端语音信号（S）的频谱减除处理器（300，400），其中所述噪声降低了的近端语音信号的一个样点块（S_{MIN}）根据所述有噪声的近端语音信号的一个相应样点块（X_{LIN}）和一个增益函数的一个相应样点块计算，所述移动电话机的特征是：所述降噪系统的特征在于所述处理器包括：

将增益函数加到所述输入信号一个相应样点块（X_{LIN}）的装置，其中增益函数是作为增益函数的样点块加上去的，而增益函数则根据所述输入信号的谱密度估计和所述输入信号的噪声分量的谱密度估计计算；以及

一个将相位添加到增益函数上的装置，使得所述频谱减除处理器提供因果滤波，

其中所述计算得的噪声降低了的语音信号样点块（S_{MIN}）的阶大于

所述有噪声的近端语音信号的相应样点块（X_{LIN}）的阶与所述增益函数~~(350)~~的相应样点块的阶之和。

28. 权利要求27的移动电话机，其中所述增益函数~~(350)~~的一个样点块是根据所述有噪声的近端语音信号的一个样点块利用谱估计计算的。

29. 权利要求28的移动电话机，其中所述谱估计是利用Bartlett算法（305）和Welch算法之一实现的。

30. 权利要求27的移动电话机，其中所述增益函数~~(350)~~具有线性相位和最小相位之一。

降低语音信号中的噪声的方法和采用该方法的系统和电话机

技术领域

本发明与通信系统有关，具体地说，与降低通信信号中破坏性背景噪声分量的影响有关。

技术背景

现在，使用移动电话机及其他通信设备中的免提装置日益普遍。一个与免提解决方案关联的众所周知的问题，特别是在汽车应用中，是破坏性的背景噪声会被免提机送话器拾取，发送给远端用户。也就是说，由于免提机送话器与近端用户之间的距离可能比较大，免提机送话器拾取的不仅是近端用户的语音，而且还有刚好在近端处出现的所有噪声。例如，在汽车电话应用中，近端送话器通常拾取周围的交通、道路和乘客厢内的噪声。所得到的有噪声的近端语音对于远端用户可能是恼人的，甚至是无法忍受的。因此所希望的是，尽可能降低背景噪声，最好是在近端信号处理链中的前部（例如，在接收的近端送话器信号送至近端语音编码器前）。

这样，许多免提系统包括一个设计成能在近端信号处理链的输入端除去背景噪声的降噪处理器。图1为诸如免提系统100那样的总体方框图。在图1中，降噪处理器110安置在免提机送话器120的输出端和近端信号处理通道（未示出）的输入端。在操作中，降噪处理器110接收一个来自送话器120的有噪声的语音信号x，对这个有噪声的语音信号x进行处理，得到一个较干净的噪声降低了的语音信号S_{NR}，通过近端信号处理链，最终传送给远端用户。

一种实现图1的降噪处理器110的众所周知的方法在这个技术领域内称为频谱减除（spectral subtraction）。例如，见S. F. Boll的“利用频谱减除抑制语音中的噪声”（“Suppression of Acoustic Noise in Speech using Spectral Subtraction”，IEEE Trans. Acoust. Speech and Sig. Proc.，27:113 - 120，1979），该文在这里列作参考。通常，频谱减除利用对噪声频谱和有噪声的语音频谱的估计形成一个基于信噪比（SNR）的增益函数，将输入频谱与这个增益函数相乘来抑制SNR低的那些频率。虽然频谱减除确实明显地减小了噪声，但它有一些

众所周知的缺点。例如，频谱减除的输出信号通常含有在这个技术领域中称为乐音（musical tone）的不自然音。此外，在处理后的信号块之间的间断往往导致使远端用户感到语音质量降低了。

近年来已经开发了许多对这种基本频谱减除方法的改进方法。例如：N. Virage的“利用听觉系统的遮蔽特性的语音增强”（“Speech Enhancement Based on Masking Properties of the Auditory System”，IEEE ICASSP. Proc. 796 - 799 vol.1，1995）；D. Tsoukalas，M. Paraskevas和J. Mourjopoulos的“利用心理声学准则的语音增强”（“Speech Enhancement using Psychoacoustic Criteria” IEEE ICASSP. Proc. 359 - 362 vol.2，1993）；F. Xie和D. Van Compernolle的“通过频谱幅度估计进行语音增强的统一方法”（“Speech Enhancement by Spectral Magnitude Estimation-A Unifying Approach” IEEE Speech Communication，89-104 vol.19，1996）；R. Martin的“根据极小统计量的频谱减除”（“Spectral Subtraction Based on Minimum Statistics” UESIPCO，Proc.，1182-1185 vol. 2，1994）；以及S.M. McOlash，R.J. Niederjohn和 J.A. Heinen的“增强受有色非稳定噪声污染的语音的频谱减除方法”（“A Spectral Subtraction Method for Enhancement of Speech Corrupted by Nonwhite，Nonstationary Noise”，IEEE IECON. Proc.，872-877 vol. 2，1995）。

Rabiner 等人的“系统识别的短时谱分析方法的实现”（“On the Implementation of a Short-Time Spectral Analysis for System Identification”，IEEE Transactions on ASSP，vol. 28，1980，pages 69-78）揭示了一种谱估计方法的实现。该文揭示了防止在使用快速付里叶变换（FFT）时出现折叠的交叠相加（overlap-and-add）方法。该文没有涉及将一个相位施加到增益函数上来提供因果滤波。

虽然这些方法确实提供了不同程度的语音增强，但如果可以开发出另一些技术针对上述频谱减除中与乐音和块间间断有关的问题，那仍然是有益的。因此，有必要改进通过频谱减除执行降噪的方法和设备。

发明内容

本发明提供的用频谱减除执行降噪的改进方法和设备满足了上述

及其他需要。按照典型实施例，频谱减除是利用线性卷积、因果滤波和/或根据频谱对频谱减除增益函数进行指数平均来实现的。有益的是，按照本发明构成的系统与现有技术的系统相比明显地改善了语音质量而并不过分复杂。

按照本发明，所开发的一些低阶谱估计与传统的频谱减除系统中的谱估计相比具有较低的频率分辨率和较小的方差。按照本发明，这种频谱用来形成一个具有所需的小方差的增益函数，从而降低了在频谱减除输出信号中的乐音。按照典型实施例，增益函数再根据输入谱利用指数平均就一些块予以平滑。低分辨率的增益函数内插成全块长增益函数，但是仍然与一个具有低阶长度的滤波器相应。有益的是，由于增益函数阶数低，就允许在内插期间添加一个相位。这个增益函数相位，按照典型实施例可以是线性相位或最小相位，使增益滤波器成为因果的，防止了块间间断。在典型实施例中，因果滤波器乘以输入信号频谱，再利用交叠相加技术将这些块适配在一起。此外，帧长安排成尽可能小，以便使引入的延迟减到最小而不会使频谱估计有过分的变化。

在一个典型实施例中，降噪系统包括一个配置成对一个有噪输入信号进行滤波以提供一个噪声降低了的输出信号的频谱减除处理器。这个频谱减除处理器的增益函数根据输入信号的谱密度估计和输入信号的噪声分量的谱密度估计计算。此外，噪声降低了的输出信号的一个样点块根据输入信号的一个相应样点块和增益函数的一个相应样点块计算， 计算得的输出信号样点块的阶大于输入信号的相应样点块的阶与所述增益函数的相应样点块的阶之和。

在典型实施例中，计算得的输出信号样点块是根据输入信号的相应样点块与增益函数样的相应点块的正确卷积计算的。例如，一个有N个输出信号样点的块是根据一个有L个输入信号样点的块和一个有M个增益函数样点的块计算的，其中L与M之和小于N。这个有M个增益函数样点的块可以例如根据L个输入信号样点利用谱估计计算。按照典型实施例，谱估计是利用Bartlett方法或Welch方法实现的。相继的输出信号块利用一个交叠相加方法适配在一起，而且将一个相位添加给增益函数，使得频谱减除处理器提供因果滤波。有益的是，增益函数可以有线性相位，也可以有最小相位。

本发明的方法包括下列步骤：计算一个输入信号的谱密度估计和这个输入信号的噪声分量的谱密度估计；以及根据有噪声的输入信号和根据一个利用谱密度估计计算得的增益函数，利用频谱减除计算噪声降低了的输出信号。按照这种方法，噪声降低了的输出信号的样点块是根据输入信号的一个相应样点块和增益函数的一个相应样点块计算的，计算得的输出信号样点块的阶大于输入信号的相应样点块的阶与增益函数的相应样点块的阶之和。

本发明提供了一种降噪系统，所述降噪系统包括一个配置成对一个有噪声的语音输入信号（X）进行滤波从而提供一个噪声降低了的语音输出信号（S）的频谱减除处理器（300，400），其中所述噪声降低了的语音输出信号的一个样点块（S_{MIN}）根据所述输入信号的一个相应样点块（X_{LIN}）计算，所述降噪系统的特征在于所述处理器包括：

将增益函数加到所述输入信号一个相应样点块（X_{LIN}）的装置，其中增益函数是作为增益函数的样点块加上去的，而增益函数则根据所述输入信号的谱密度估计和所述输入信号的噪声分量的谱密度估计计算；以及

一个将相位添加到增益函数上的装置，使得所述频谱减除处理器提供因果滤波，

其中所述计算得的语音输出信号样点块（S_{MIN}）的阶大于所述语音输入信号的相应样点块（X_{LIN}）的阶与所述增益函数的相应样点块的阶之和。

本发明提供了一种对一个有噪声的语音输入信号（X）进行处理从而提供一个噪声降低了的语音输出信号（S）的方法，所述方法包括利用频谱减除根据所述有噪声的语音语音输入信号（X）和根据一个利用谱密度估计计算得的增益函数计算所述噪声降低了的语音输出信号（S）的步骤，其中所述噪声降低了的输出信号的一个样点块（S_{MIN}）根据所述语音输入信号的一个相应样点块（X_{LIN}）和所述增益函数的一个相应样点块计算，所述方法的特征是包括下列步骤：

计算所述语音输入信号的谱密度估计和所述语音输入信号的噪声分量的谱密度估计；以及

将一个相位（355）添加给所述增益函数从而使得所述利用频谱减除的步骤提供因果滤波，

其中所述计算得的语音输出信号的样点块（S_{MIN}）的阶大于所述语音输入信号的相应样点块（X_{LIN}）的阶与所述增益函数的相应样点块的阶之和。

本发明还提供了一种移动电话机，所述移动电话机包括一个配置成对一个有噪声的近端语音信号（X）进行滤波从而提供一个噪声降低了的近端语音信号（S）的频谱减除处理器（300，400），其中所述噪声降低了的近端语音信号的一个样点块（S_{MIN}）根据所述有噪声的近端语音信号的一个相应样点块（X_{LIN}）和一个增益函数的一个相应样点块计算，所述移动电话机的特征是：所述降噪系统的特征在于所述处理器包括：

将增益函数加到所述输入信号一个相应样点块（X_{LIN}）的装置，其中增益函数是作为增益函数的样点块加上去的，而增益函数则根据所述输入信号的谱密度估计和所述输入信号的噪声分量的谱密度估计计算；以及

一个将相位添加到增益函数上的装置，使得所述频谱减除处理器提供因果滤波，

其中所述计算得的噪声降低了的语音信号样点块（S_{MIN}）的阶大于所述有噪声的近端语音信号的相应样点块（X_{LIN}）的阶与所述增益函数（350）的相应样点块的阶之和。

下面将结合附图所示的说明性例子详细说明本发明的上述及其他特征和优点。熟悉该技术的人员可以理解，所说明的实施例只是用来例示的，可以设想有大量的与在这里说明的等效的实施方式。

附图说明

图1为一个其中可以实现本发明的精神的降噪系统的方框图。

图2示出了一个传统的频谱减除降噪处理器。

图3-4示出了按照本发明设计的典型频谱减除降噪处理器。

图5示出了利用本发明提出的频谱减除技术得出的典型频谱图。

图6-7示出了利用本发明提出的频谱减除技术得出的典型增益函数。

图8-28示出了按照本发明提出的典型频谱减除技术的模拟结果。

具体实施方式

为了理解本发明的各个特征和优点，首先来看一下传统的频谱减

除技术是有益的。通常，频谱减除是建立在通信应用中的噪声信号和语音信号是随机的、非相关的，加在一起形成有噪声的语音信号的假设的基础上。例如，如果s（n）、w（n）和x（n）为分别表示语音、噪声和有噪声的语音的统计短时平稳过程，于是有

$$x(n) = s(n)+w(n) \qquad (1)$$

$$R_x(f) = R_s(f)+R_w(f) \qquad (2)$$

其中R（f）表示随机过程的功率谱密度。

噪声功率谱密度R_W（f）可以在语音暂停（即在 x（n） = w（n）时）期间估计出来。为了估计语音的功率谱密度，可以如下那样形成一个估计量

$$\hat{R}_s(f) = \hat{R}_x(f)-\hat{R}_w(f) \qquad (3)$$

估计功率谱密度的传统方式是利用周期图（periodogram）。例如，如果X_N（f_u）是x（n）的N点傅里叶变换，而W_N（f_u）是w（n）的相应傅里叶变换，于是:

$$\hat{R}_x(f_u) = P_{x,N}(f_u) = \frac{1}{N}|X_N(f_u)|^2,\ f_u = \frac{u}{N},\ u=0,\ \ldots,\ N-1 \qquad (4)$$

$$\hat{R}_w(f_u) = P_{w,N}(f_u) = \frac{1}{N}|W_N(f_u)|^2,\ f_u = \frac{u}{N},\ u=0,\ \ldots,\ N-1 \qquad (5)$$

式（3）、（4）和（5）可以合并成:

$$|S_N(f_u)|^2 = |X_N(f_u)|^2-|W_N(f_u)|^2 \qquad (6)$$

或者，更普遍的形式为：

$$|S_N(f_u)|^a = |X_N(f_u)|^a - |W_N(f_u)|^a \qquad (7)$$

其中功率谱密度替换为谱密度的一般形式。

由于人类的耳朵对语音的相位误差并不敏感，因此有噪声的语音的相位 $\phi_X(f)$ 可以用洁净语音的相位 $\phi_S(f)$ 来近似：

$$\phi_s(f_u) \approx \phi_x(f_u) \qquad (8)$$

因此，估计洁净语音的傅里叶变换的一般表示式为：

$$S_N(f_u) = \left(|X_N(f_u)|^a - k \cdot |W_N(f_u)|^a\right)^{\frac{1}{a}} \cdot e^{j\phi_x(f_u)} \qquad (9)$$

其中参数 k 是为控制噪声减除量引入的。为了简化表示法，引入向量形式：

$$X_N = \begin{pmatrix} X_N(f_0) \\ X_N(f_1) \\ \vdots \\ X_N(f_{N-1}) \end{pmatrix} \qquad (10)$$

向量间的运算是逐元计算的。为了清晰起见，向量的逐元相乘在这里用⊙表示。因此，式（9）可以用一个增益函数 G_N 按向量记法写成：

$$S_N = G_N \odot |X_N| \odot e^{j\phi_x} = G_N \odot X_N \qquad (11)$$

其中增益函数为：

$$G_N = \left(\frac{|X_N|^a - k \cdot |W_N|^a}{|X_N|^a} \right)^{\frac{1}{a}} = \left(1 - k \cdot \frac{|W_N|^a}{|X_N|^a} \right)^{\frac{1}{a}} \qquad (12)$$

式（12）表示了传统的频谱减除算法，例示于图2。在图2中，传统的频谱减除降噪处理器200包括快速傅里叶变换处理机210、幅度平方处理器220、语音活动检测器230、按块平均器240、按块增益计算处理器250、乘法器260和快速傅里叶逆变换处理器270。

如同所示，一个有噪声的语音输入信号加到快速傅里叶变换处理器210的输入端，快速傅里叶变换处理器210的输出端与幅度平方处理器220的输入端和乘法器260的第一输入端连接。幅度平方处理器220的输出端与开关225的第一触点和增益计算处理器250的第一输入端连接。语音活动检测器230的输出端与开关225的投掷控制输入端连接，而开关225的第二触点与按块平均器240的输入端连接。按块平均器240的输出端与增益计算处理器250的第二输入端连接，而增益计算处理器250的输出端与乘法器260的第二输入端连接。乘法器260的输出端与快速傅里叶逆变换处理器270的输入端连接，而快速傅里叶逆变换处理器270的输出端提供传统的频谱减除系统200的输出。

在操作中，传统的频谱减除系统200利用上述的传统频谱减除算法处理进入的有噪声的语音信号，提供噪声降低了的较为洁净的语音信号。在实践中，图2中的各个组成部分可以利用任何已知的数字信号处理技术实现，包括通用计算机、集成电路和/或专用集成电路（ASIC）。

注意，在传统的频谱减除算法中，有两个参数，a 和 k，控制噪声减除量和语音质量。将第一参数a设置为2，可提供功率谱减除，而将第一参数a设置为1，可提供幅度频谱减除。另外，将第一参数a设置为0.5，可使降噪增大，而只有不大的语音失真。这是由于在从有噪声的语音中减去噪声前频谱受到了压缩。

第二参数k可以调整成达到所需的降噪效果。例如，如果选择一个较大的k，语音失真就会增大。在实践中，参数k通常按照第一参数的

选择情况设置。减小a通常会导致也需减小参数k，以便使语音失真小。在功率谱减除的情况下，通常采用过减除（即k > 1）。

传统频谱减除的增益函数（见式（12））从一个全块估计得出，具有零相位。结果，相应的脉冲响应g_N（U）是非因果的，具有长度N（等于块长度）。因此，增益函数G_N（1）与输入信号 X_N相乘（见式（11））会导致与一个非因果滤波器进行周期性的循环卷积。如上面所说明的那样，周期性的循环卷积可以导致不合要求的时域混叠，而滤波器的非因果性质可以导致块间间断，因此语音质量较差。有益的是，本发明提供了与因果增益的滤波器进行正确卷积的方法和设备，从而消除了上述时域混叠和块间间断的问题。

就时域混叠问题来说，注意到在时域的卷积相应于在频域的相乘。也就是说:

$$x(u) * y(u) \leftrightarrow X(f) \cdot Y(f), \quad u = -\infty, \ldots, \infty \qquad (13)$$

在变换是通过一个N点的快速傅里叶变换（FPT）获得时，相乘的结果就不是正确的卷积。相反，结果是一个周期为N的循环卷积:

$$x_N \textcircled{N} y_N \qquad (14)$$

其中，符号⊗表示循环卷积。

为了在采用快速傅里叶变换时获得正确的卷积，脉冲响应 x_N 和y_N的积累阶数必须小于或等于一个小于块长度N-1的阶数。

因此，按照本发明，由周期性的循环卷积引起的时域混叠问题可以利用总阶数小于或等于N-1的一个增益函数G_N（1）和一个输入信号块X_N来解决。

按照传统的频谱减除，输入信号的谱X_N具有全块长度N的性质。然而，按照本发明，用一个长度为L （L<N）的输入信号块 x_L 来构成一个阶为L的谱。长度L称为帧长，因此 x_L 是一个帧。由于与长度为N的增益函数相乘的谱也应该具有长度N，因此通过填零将帧 x_L 充填到全块长度N，结果形成 X_{LIN}。

为了构成一个长度为N的增益函数，按照本发明提出的增益函数可以根据一个长度为M的增益函数$G_M(l)$内插形成$G_{MIN}(l)$，其中$M < N$。为了得出遵从本发明的低阶增益函数$G_{MIN}(l)$，任何已知的或还有待于开发的频谱估计技术都可以用来代替上述简单的傅里叶变换周期图。一些已知的频谱估计技术使所得到的增益函数具有较小的方差。例如，见J.G. Proakis和D.G. Manolakis的“数字信号处理：原理、算法和应用”(“Digital Signal Processing; Principles, Algorithms, and Applications”，Macmillan，Second Ed，1992)。

按照众所周知的Bartlett方法，例如，将长度为N 的这块分成K个长度为M 的子块。于是，每个子块的周期图可以计算出来，将这些结果平均后就可以得到一个对于总块的长为M 的周期图，为：

$$\begin{aligned} P_{x,M}(f_u) &= \frac{1}{K}\sum_{k=0}^{K-1} P_{x,M,k}(f_u),\quad f_u=\frac{u}{M},\ u=0,\ \ldots,\ M-1 \\ &= \frac{1}{K}\sum_{k=0}^{K-1} \left|\mathcal{F}(x(k\cdot M+u))\right|^2 \end{aligned} \tag{15}$$

有益的是，在这些子块非相关时，与全块长周期图相比方差降低了一个因子K 。频率分辨率也降低了相同的因子。

或者，也可以采用Welch方法。Welch方法类似于Bartlett方法，只是每个子块加了一个Hanning窗，而这些子块允许相互交叠，结果形成较多的子块。与Bartlett方法相比，Welch方法得出的方差更小一些。Bartlett和Welch方法是两种谱估计技术，而其他已知的谱估计技术也可以采用。

不论用的是怎样精确的谱估计技术，还有可能利用平均技术减小噪声周期图估计的方差，这正是所希望的。例如，在噪声是长期平稳的假设下，能对从上述Bartlett和Welch方法得到的这些周期图进行平均。一种技术是采用如下的指数平均：

$$\bar{P}_{x,M}(l) = \alpha\cdot\bar{P}_{x,M}(l-1) + (1-\alpha)\cdot P_{x,M}(l) \tag{16}$$

在式（16）中，函数$P_{x,M}(l)$是利用Bartlett或Welch方法计算的，函数$\overline{P}_{x,M}(l)$是当前块的指数平均，而函数$\overline{P}_{x,M}(l-1)$是前一块的指数平均。参数α控制指数记忆的长度，通常不应该超过噪声可以被认为是平稳的长度。α接近1导致较长的指数记忆，而周期图方差减小得也更可观一些。

长度M 称为子块长度，得到的低阶增益函数具有长度为M 的脉冲响应。因此，用于组成增益函数的噪声周期图估计$\overline{P}_{x_L,M}(l)$ 和有噪声的语音周期图估计$P_{x_L,M}(l)$长度也为M ：

$$G_M(l) = \left(1 - k \cdot \frac{\overline{P}^{a}_{x_L,M}(l)}{P^{a}_{x_L,M}(l)}\right)^{\frac{1}{a}} \qquad (17)$$

按照本发明，这是通过用例如Bartlett方法从输入帧X_L 得到短周期图估计再加以平均实现的。Bartlett方法（或其他适当的估计方法）减小了所估计的周期图的方差，也降低了频率分辨率。分辨率从L个频率分辨区间降低到M个分辨区间意味着周期图估计$P_{x_L,M}(l)$长度也为M。此外，噪声周期图估计$\overline{P}_{x_L,M}(l)$的方差还可以用上述的指数平均减小。

为了满足总阶数小于或等于N-1的要求，使被加到子块长度M的帧长L小于N。结果，可以形成所需的输出块，为：

$$S_N = G_{M1N}(l) \odot X_{L1N} \qquad (18)$$

有益的是，本发明的低阶滤波器还提供了对付由于传统的频谱减除算法中增益滤波器的非因果特性引起的问题（即块间间断和语音质量降低）的可能。具体地说，按照本发明，增益函数可以添加一个相位，从而提供一个因果滤波器。按照典型实施例，这个相位可以由幅度函数构成，按需要可以是线性相位或者最小相位。

为了按照本发明构成一个线性相位滤波器，首先注意到FFT的块长是否为M，然后注意到在时域内的循环移位在频域内是与一个相位函数

相乘:

$$g(n-l)_M \leftrightarrow G_M(f_u)\cdot e^{-j2\pi ul/M},\ f_u = \frac{u}{M},\ u = 0, \ldots, M-1 \qquad (19)$$

在本情况下， 1 等于M/2 +1，因为在脉冲响应中的第一个位置应该有零延迟（即是一个因果滤波器）。因此:

$$g(n-(M/2+1))_M \leftrightarrow G_M(f_u)\cdot e^{-j\pi u(1+\frac{2}{M})} \qquad (20)$$

从而可以得到线性相位滤波器$\overline{G}_M(f_u)$，为:

$$\bar{G}_M(f_u) = G_M(f_u)\cdot e^{-j\pi u(1+\frac{2}{M})} \qquad (21)$$

按照本发明，增益函数还内插到长度为N，这是例如用平滑内插法完成的。添加给增益函数的相位因此改变，从而有:

$$\bar{G}_{M|N}(f_u) = G_{M|N}(f_u)\cdot e^{-j\pi u(1+\frac{2}{M})\cdot\frac{M}{N}} \qquad (22)$$

有益的是，这种线性相位滤波器的设计还可以在时域进行。在这种情况下，增益函数$G_M(f_u)$利用IFFT变换到时域，完成循环移位。将经移位的脉冲响应补零到长度为N，然后用N点FFT变换回来。这就可以按要求得到一个经内插的因果线性相位滤波器$\overline{G}_{M|N}(f_u)$。

可以根据增益函数利用Hilbert变换关系构成本发明的因果最小相位滤波器。例如见A.V. Oppenheim 和R.W. Schafer的“离散时间信号处理”（“Discrete - Time Signal Processing”，Perntic - Hall，Inter. Ed.，1989）。Hilbert变换关系意味着一种在一个复值函数的

实部与虚部之间的唯一关系。有益的是，在使用复信号的对数时，这还可以用于幅度与相位之间的关系，有：

$$\begin{aligned}\ln\left(|G_M(f_u)|\cdot e^{j\cdot\arg(G_M(f_u))}\right) &= \ln(|G_M(f_u)|)+\ln\left(e^{j\cdot\arg(G_M(f_u))}\right)\\ &= \ln(|G_M(f_u)|)+j\cdot\arg(G_M(f_u))\end{aligned} \tag{23}$$

在给出的这种情况下，相位是零，从而得到一个实函数。函数$\ln(|G_M(f_u)|)$利用M 点的IFFT变换到时域，形成$g_M(n)$。这个时域函数整理后为：

$$\bar{g}_M(n)=\begin{cases}2\cdot g_M(n), & n=1,\ 2,\ \ldots,\ M/2-1\\ g_M(n), & n=0,\ M/2\\ 0, & n=M/2+1,\ \ldots,\ M-1\end{cases} \tag{24}$$

函数$\bar{g}_M(n)$利用M 点的FFT变换回频域，得到$\ln(|\bar{G}_M(f_u)|\cdot e^{j\arg(\bar{G}_M(f_u))})$。根据这个形成函数。然后，将这个因果最小相位滤波器$\bar{G}_M(f_u)$内插到长度为N。内插用与在以上对线性相位说明的同样的方法进行。得到的经内插的滤波器$G_{M|N}(f_u)$是因果的，大致具有最小相位。

上述按照本发明设计的频谱减除方案示于图3。在图3中，提供线性卷积和因果滤波的频谱减除降噪处理器300示为包括Bartlett处理器305、幅度平方处理器320和语音活动检测器330。按块平均处理器340、低阶增益计算处理器350、增益相位处理器355、内插处理器356、乘法器360、快速傅里叶逆变换处理器370和交叠相加处理器380。

如图所示，有噪声的语音输入信号加到Bartlett处理器305的输入端和快速傅里叶变换处理器310的输入端上。Bartlett处理器305的输出端与幅度平方处理器320的输入端连接，而快速傅里叶变换处理器310的输出端与乘法器360的第一输入端连接。幅度平方处理器320的输出端与开关325的第一触点和低阶增益计算处理器350的第一输入端连接。语音活动检测器330的控制输出端与开关325的投掷控制输入端连接，而开关325的第二触点与按块平均器340的输入端连接。

按块平均器340的输出端与低阶增益计算处理器350的第二输入端连接，而低阶增益计算处理器350的输出端与增益相位处理器355的输入端连接。增益相位处理器355的输出端与内插处理器356的输入端连接，而内插处理器356的输出端与乘法器360的第二输入端连接。乘法器360的输出端与快速傅里叶逆变换处理器370的输入端连接，而快速傅里叶逆变换处理器370的输出端与交叠相加处理器380的输入端连接。交叠相加处理器380的输出端为典型降噪处理器300提供一个噪声降低了的洁净语音输出。

在操作中，按照本发明设计的频谱减除降噪处理器300利用以上说明的线性卷积、因果滤波算法处理进入的有噪声的语音信号，得出噪声降低了的洁净语音信号。在实践中，图3中的各个组成部分可以利用任何已知的数字信号处理技术实现，包括通用计算机、集成电路和/或专用集成电路（ASIC）。

有益的是，可以用按照本发明设计的受控指数增益函数平均方案进一步减小本发明的增益函数$G_M(l)$的方差。按照典型实施例，这个平均是根据当前块的谱$P_{x,M}(l)$与平均噪声谱$\overline{P}_{x,M}(l)$之间的偏差进行的。例如，在有一个小的偏差时，可以对增益函数$G_M(l)$进行长时间的平均，相当于平稳背景噪声情况。相反，在有一个大的偏差时，可以对增益函数$G_M(l)$进行短时平均或者不平均，相当于有语音或者背景噪声变化很大的情况。

为了处理从一个语音期到一个背景噪声期过渡转换，对增益函数的平均不是与偏差的减小成正比，这样做就引入了一个听得到的阴影语音（因为适合一个语音谱的增益函数会保持一段较长的时间）。代之以允许平均渐渐增大，以便提供时间使增益函数适应这个平稳输入。

按照典型实施例，谱间偏差的度量定义为

$$\beta(l) = \frac{\sum_u |P_{x,M,u}(l) - \bar{P}_{x,M,u}(l)|}{\sum_u \bar{P}_{x,M,u}(l)} \qquad (25)$$

其中，$\beta(l)$受以下限制

$$\beta(l) = \begin{cases} 1, & \beta(l) > 1 \\ \beta(l), & \beta_{min} \le \beta(l) \le 1, \ 0 \le \beta_{min} \ll 1 \\ \beta_{min}, & \beta(l) < \beta_{min} \end{cases} \qquad (26)$$

其中，$\beta(l)=1$导致不对增益函数进行指数平均，而$\beta(l)=\beta_{min}$提供最大程度的指数平均。

参数$\overline{\beta}(l)$为谱间偏差的指数平均，为

$$\overline{\beta}(l) = \gamma \cdot \overline{\beta}(l-1) + (1-\gamma) \cdot \beta(l) \qquad (27)$$

式（27）中的参数γ用来保证在出现从有大谱间偏差期到有小偏差期的转换时使增益函数适应新的情况。如上所述，这样做是为了防止阴影语音。按照典型实施例，这种适应是在由于$\beta(l)$减小而开始增大对增益函数的指数平均前完成的。因此：

$$\gamma = \begin{cases} 0, & \bar{\beta}(l-1) < \beta(l) \\ \gamma_c, & \bar{\beta}(l-1) \geq \beta(l),\ 0 < \gamma_c < 1 \end{cases} \qquad (28)$$

在偏差$\beta(l)$增大时，参数$\beta(l)$直接跟着增大，但是在偏差减小时，就对$\beta(l)$进行指数平均，形成经平均的参数$\beta(l)$。增益函数的指数平均为：

$$\bar{G}_M(l) = (1-\bar{\beta}(l)) \cdot \bar{G}_M(l-1) + \bar{\beta}(l) \cdot G_M(l) \qquad (29)$$

以上这些表示式对于不同的输入信号情况可以解释如下。在噪声期间，方差减小。只要噪声频谱对于每个频率都有一个稳定的平均值，可以予以平均，减小方差。噪声电平改变导致平均噪声谱$\bar{P}_{x,M}(l)$与当前块的谱$P_{x,M}(l)$之间的偏差。因此，这种受控指数平均方法减小增益函数平均，直到噪声电平稳定到一个新的电平。这样就允许对噪声电平改变进行处理，从而在平稳噪声期间给出一个减小的方差，而对噪声改变可以迅速响应。高能量的语音往往具有一些随时间变化的频谱峰。在对不同的块的频谱峰进行平均时，由于它们的频谱估计含有这些峰的平均，所以看来像一个较宽的频谱，这将使语音质量降低。因此，在高能量的语音期期间使指数平均最小。由于平均噪声谱$\bar{P}_{x,M}(l)$与当前高能量语音谱$P_{x,M}(l)$之间的偏差大，因此不对增益函数执行指数平均。在低能量的语音期期间，根据当前低能量语音谱与平均噪声谱之间的偏差，采用短记忆的指数平均。因此，对于高能量语音，方差缩减比在背景噪声期期间的小而比高能量语音期的大。

上述按照本发明设计的频谱减除方案示于图4。在图4中，提供线性卷积、因果滤波和受控指数平均的频谱减除降噪处理器400示为包括图3所示系统300的Bartlett处理器305、幅度平方处理器320、语音活动检测器330、按块平均器340、低阶增益计算处理器350、增益相位处理器355、内插处理器356、乘法器360、快速傅里叶逆变换处理器370和交叠相加处理器380，还有平均控制处理器445、指数平均处理器446

和可选用的固定FIR后置滤波器465。

如图所示，有噪声的语音输入信号加到Bartlett处理器305的输入端和快速傅里叶变换处理器310的输入端上。Bartlett处理器305的输出端与幅度平方处理器320的输入端连接，而快速傅里叶变换处理器310的输出端与乘法器360的第一输入端连接。幅度平方处理器320的输出与开关325的第一触点、低阶增益计算处理器350的第一输入端和平均控制处理器445的第一输入端连接。

语音活动检测器330的控制输出端与开关325的投掷控制输入端连接，而开关325的第二触点与按块平均器340的输入端连接。按块平均器340的输出端与低阶增益计算处理器350的第二输入端和平均控制器445的第二输入端连接。低阶增益计算处理器350的输出端与指数平均处理器446的信号输入端连接，而平均控制器445的输出端与指数平均处理器446的控制输入端连接。

指数平均处理器446的输出端与增益相位处理器355的输入端连接，而增益相位处理器355的输出端与内插处理器356的输入端连接。内插处理器356的输出端与乘法器360的第二输入端连接，而可选用的固定FIR后置滤波器465的输出端与乘法器360的第三输入端连接。乘法器360的输出端与快速傅里叶逆变换处理器370的输入端连接，而快速傅里叶逆变换处理器370的输出端与交叠相加处理器380的输入端连接。交叠相加处理器380的输出端为这个典型系统400提供一个洁净的语音信号。

在操作中，按照本发明设计的频谱减除降噪处理器400利用以上说明的线性卷积、因果滤波和受控指数平均算法处理进入的有噪声的语音信号，得出经改善的噪声降低了的语音信号。就象图3中的实施例，图4的各个组成部分可以利用任何已知的数字信号处理技术实现，包括通用计算机、集成电路和/或专用集成电路（ASIC）。

注意，由于按照典型实施例将帧长L与子块长M之和选择成短于N-1，因此可以如图4所示额外增添长度为J＜N-1-L-M的固定FIR滤波器465。后置滤波器465是如同所示通过将这个滤波器的经内插的脉冲响应乘以信号频谱施加的。内插到长度N是通过对滤波器补零再使用N点FFT来完成的。后置滤波器465可用来滤出电话带宽或一个恒定的音频频段成分。或者，也可以将后置滤波器465的功能直接包括在增

益函数内。

上述算法的这些参数在实践中根据实现这种算法的具体应用设定。作为举例，以下以免提GSM汽车移动电话机为背景说明参数选择。

首先，根据GSM规范，帧长L 设置为160个样点，得出一些20ms的帧。在其他系统中可以采用对L 的其他选择。然而，应指出的是，应指出的是，增加帧长L 相应于增加延迟。将子块长度M （例如，Bartlett处理器的周期图长度）设置得小，可以得到较大的方差缩减M。由于用FFT计算周期图，长度M 可以方便地设定为2的幂。于是频率分辨率确定为：

$$B = \frac{F_s}{M} \qquad (30)$$

GSM系统的采样率为8000 Hz。因此，长度M = 16、M = 32 和 M =64分别给出频率分辨率为500 Hz、250 Hz 和 125 Hz，如图5中所示。在图5中，图（a）示出了一个洁净语音信号的简单周期图，而图（b）、（c）和（d）分别示出了利用具有32、16和8个频带的Bartlett方法对一个洁净语音信号计算出的周期图。对于语音和噪声信号来说，频率分辨率为250 Hz是合理的，因此 M = 32。这就得出长度L+M = 160 + 32 = 192，如上所述应该小于N - 1。因此，将N选择成例如比192大的2的幂（例如，N = 256）。在这种情况下，可以按需要选用长度J < 63的FIR后置滤波器。

如上所述，噪声减除量受a和k参数控制。选择参数a = 0.5 （即平方根频谱减除）可以提供强的降噪效果，而有不大的语音失真。这种情况示于图6 （其中，语音加噪声估计为1，而且k为1）。由图6可见，a = 0.5与较高的a值相比提供较好的降噪效果。为了清晰起见，图6只给出一个频率分辨区间，它是对于这个频率分辨区间的SNR，这在下面还要引用。

按照典型实施例，在采用a = 0.5时参数k可以安排成比较小。在7中，例示了对于a= 0.5 （同样，语音加噪声估计为1）的情况不同的k值的增益函数。在向较低的SNR移动时增益函数应该连续地降低，这是在k 1时的情况。模拟表明，k = 0.7可以在维持高的降噪效果的同时

提供小的语音失真。

如上所述，噪声频谱估计是经指数平均的，而参数α控制指数记忆的长度。由于增益函数经过平均，因此不很需要对噪声频谱估计进行平均。模拟表明，$0.6<\alpha<0.9$提供了所需的方差缩减，得出大约为2到10帧的时间常数τ_{frame}，为：

$$\tau_{frame} \approx -\frac{1}{\ln \alpha} \quad (31)$$

噪声估计的指数平均例如选择为：$\alpha = 0.8$。

参数$\beta_{\min}$确定增益函数指数平均的最大时间常数。规定在几秒的时间常数$\tau_{\beta_{\min}}$用来确定$\beta_{\min}$，有：

$$\beta_{min} = 1 - e^{-\frac{L}{F_s \cdot \tau_{\beta_{min}}}} \quad (32)$$

对于一个平稳的噪声信号来说时间常数为2分钟是合理的，这相当于$\beta_{\min} = 0$。也就是说，对$\beta(l)$不需要有下限（在式（32）中），因为$\beta(l) \geq 0$（按照式（25））。

参数γ_c控制在有从语音到一个平稳的输入信号的转换时允许受控指数平均的记忆增加能多快（即，允许$\overline{\beta}(l)$参数减小能多快，见式（27）和（28））。在用长时间记忆对增益函数进行平均时，会产生阴影语音，因为增益函数还记得这语音谱。

例如，设想一种极端的情况，有噪声的语音谱估计$P_M(l)$与噪声谱估计$\overline{P}_M(l)$之间的偏差从一个极端值变到另一个极端值。在第一种情况下，这个偏差大，使得在上一段长时间内对于所有频率有$G_M(l) = 1$。因此，$\beta(l) = \overline{\beta}(l) = 1$。其次，对这些谱估计进行操作，使得$P_M(l) = \overline{P}_M(l)$，以便模拟$\beta(l) = 0$和$G_M(l) = (1-k)^{1/2}$的极端情况。$\overline{\beta}(l)$参数将根据参数$\gamma_c$减小到零。因此，这些参数值为：

$$\begin{aligned}&\bar{\beta}(-1)=1,\ \bar{G}_M(-1)=1,\\&\beta(-1)=1,\ G_M(-1)=1,\\&\beta(l)=0,\quad G_M(l)=0.09,\ l=0,\ 1,\ 2,\ \ldots\end{aligned}\tag{33}$$

将给定的参数代入式（27）和（29），得出：

$$\bar{\beta}(l)=\gamma_c^{(l+1)}\tag{34}$$

$$\bar{G}_M(l)=(1-\bar{\beta}(l))\cdot\bar{G}_M(l-1)+0.09\cdot\bar{\beta}(l)\tag{35}$$

其中，l为在能量减小后的块块数目。如果增益函数选择成在2帧后达到时间常数电平e^{-1}，就有$\gamma_c=0.506$。这种极端情况对于不同的γ_c值示于图8的图（a）和（b）。能量减少较慢的较为实际的模拟示于图8的图（c）和（d）。e^{-1}电平线表示一个时间常数的电平（即，在越过这个电平时，经过了一个常数的时间）。利用录制的输入信号进行实际模拟的结果示于图9，可见$\gamma_c=0.8$对于防止阴影语音来说是一个良好的选择。

下面，提供利用以上建议的参数选择得到的结果。有益的是，模拟结果表明与其他频谱减除方法相比在语音质量和残余背景噪声质量上都有改进，而且仍能提供强的降噪效果。增益函数的指数平均主要是用来提高残余噪声的质量。正确的卷积与因果滤波相结合提高了整体音质，使得有一个短的延迟成为可能。

在这些模拟中，对有噪声的语音信号使用了众所周知的GSM语音活动检测器（例如见European Digital Cellular Telecommunications Systems（Phase 2）；Voice Activity Detection（VAD）（GSM 06.32），European Telecommunications Standards Institute，1994）。在这些模拟中所用的信号是将在一个汽车内分别录制的语音和噪声的录音合成的。语音录音是在一个安静的汽车内利用免提设备和模拟

电话带宽滤波器执行的。噪声序列利用相同的设备在一个运动的汽车内录制。

对执行的降噪与接收的语音质量作了比较。参数选择成以上值与大的降噪效果相比音质良好更为突出。在采用更为积极的选择时，可以得到更好一些的降噪效果。图10和11分别分别示出了输入的语音和噪声，这两个输入用1:1关系加在一起。所得到的有噪声的输入语音信号示于图12。噪声降低了的输出信号例示于图13。这些结果还可以就能量意义上给出，使得计算降噪效果更为容易，揭示出有些语音期是否没有增强。图14、15和16分别示出洁净语音、有噪声的语音和在降噪后得到的输出语音。如同所示，降噪效果达到了13 dB左右。在输入是用语音和汽车噪声以2:1关系加在一起形成时，输入SNR的增大如图17和19所示。所得到的信号示于图18和20，从中可以估计降噪效果接近18 dB。

还进行了其他一些模拟，清楚地表明增益函数要有适当的脉冲响应长度以及因果性质的重要性。以下给出的这些序列都是根据长度为30秒的有噪声的语音给出的。这些序列作为IFFT输出的绝对平均值$|s_N|$给出（参见图4）。IFFT给出256点长的数据块，取每个数据值的绝对值再加以平均。因此，可以清楚地看出增益函数的不同选择（即，非因果滤波器、较短的和较长的脉冲响应、最小相位或线性相位）的影响。

图21示出了由脉冲响应具有较短长度M的增益函数得出的平均值$|s_N|$，由于增益函数具有零相位，因此是非因果的。这可以从在经平均的块的结尾处的M = 32个样点有着高电平看出。

图22示出了由脉冲响应具有全长N的增益函数得出的平均值$|s_N|$，由于增益函数具有零相位，因此是非因果的。这可以从在经平均的块的结尾处的这些样点有着高电平看出。这种情况在相位和长度上相应于传统的频谱减除的增益函数。全长的增益函数是通过对噪声和有噪声的语音的周期图进行内插代替增益函数得到的。

图23示出了由脉冲响应具有较短长度M的最小相位增益函数得出的平均值$|s_N|$。加到增益函数上的最小相位使它成为因果的。这个因果关系可以从在经平均的块的结尾处的这些样点有着低电平看出。最小相位滤波器给出M = 32个样点的最大延迟，这在图23中由从样点160到192

的斜坡可以看出。延迟在增益函数是因果的约束下是最小的。

图24示出了由脉冲响应具有全长N的增益函数得出的平均值$|s_N|$，限制成有最小相位。限制为最小相位给出N = 256个样点的最大延迟，这个块可以持有96个样点的最大线性延迟，因为这个帧是在256个样点的全块的开始处的160个样点。这可以在图24中由从样点160到255的不为零的斜坡看出，由于延迟可以比96点长，因此产生一个循环延迟，而在最小相位的情况下很难检测到覆盖这个帧部分的延迟样点。

图25示出了由脉冲响应具有较短长度M的线性相位增益函数得出的平均值$|s_N|$。加到增益函数的线性相位使它成为因果的。这可以从在经平均的块的结尾处的这些样点有着低电平看出。有着线性相位增益函数的延迟为 M/2 = 16个样点，这可以由从样点0到15和160到175的斜坡看到。

图26示出了由脉冲响应具有全长N的增益函数得出的平均值$|s_N|$，限制成具有线性相位。限制成线性相位给出 N/2 = 128个样点的最大延迟。这个块可以持有96个样点的最大线性延迟，因为帧是在256个样点的全块开始处的160个样点。这些延迟得比96个样点更长的样点导致可以看出的循环延迟。

在块中的与交叠相应的一些小样点值的好处是块间干扰较小，因为这种交叠不会引起间断。在是传统的频谱减除情况的采用全长的脉冲响应时，线性相位或最小相位引入的延迟超过块的长度。得到的循环延迟引起延迟样点的折叠，从而输出样点可能次序错误。这表示在采用线性相位或最小相位增益函数时，应该选择较短的脉冲响应长度。采用线性或最小相位使增益函数成为因果的。

在输出信号的音质是最重要的因素时，应该采用线性相位滤波器。在延迟重要时，应该采用非因果的零相位滤波器，虽然与采用线性相位滤波器相比语音质量要差一些。一个良好的折中方案是最小相位滤波器，它具有短的延迟和良好的语音质量，但是与采用线性相位滤波器相比复杂许多。始终应该用与具有短长度M 的脉冲响应相应的增益函数来提高音质。

增益函数的指数平均在信号是平稳时提供较小的方差。主要的优点是降低乐音和残余噪声。有和没有指数平均的增益函数示于图27和28。如同所示，在采用指数平均时，在噪声期以及低能量语音期

期间信号的改变较小。增益函数改变较小导致输出信号中显著的不自然声音较少。

总而言之，本发明提供了采用线性卷积、因果滤波和/或增益函数的受控指数平均进行频谱减除的改进方法和设备。这些典型的方法提供了改进的降噪效果，而且用不一定要是2的幂的帧长进行工作。这在这种降噪方法与其他语音增强方法以及语音编码器结合时可能是一个重要的特性。

这些典型的方法以两个有效的方式减小了在这种情况下是一个复值函数的增益函数的变化。第一，用以方差缩减换取频率分辨率的频谱估计方法（例如Bartlett或Welch方法）减小当前块谱估计的方差。第二，提供了一种增益函数的指数平均，这种指数平均取决于估计的噪声谱与当前输入信号谱估计之间的偏差。增益函数在平稳的输入信号期间改变小可以给出一个声音残余噪声较小的输出。增益函数分辨率较低还有利于执行正确的卷积，得到改善的音质。通过使增益函数还具有因果性质进一步提高了音质。有益的是，质量改善在输出块中就可以看出。音质的改善是由于输出块的交叠部分具有减小得很多的样点值，从而在这些块用交叠相加方法适配在一起时干扰较小。采用以上说明的典型参数选择，输出噪声可以降低13-18 dB。

熟悉该技术的人员可以理解，本发明并不局限于在这里为了例示起见进行说明的这些具体的典型实施例，可以设想还有许多可供选择的实施方案。例如，虽然本发明以免提通信应用为背景进行说明，但是熟悉该技术的人员可以理解本发明的精神同样可应用于需要消除一个特定信号分量的信号处理应用。因此，本发明的专利保护范围由所附权利要求书限定，而不是上述说明，而所有与权利要求的含意一致的那些等效实施方式都应列入本发明的专利保护范围。

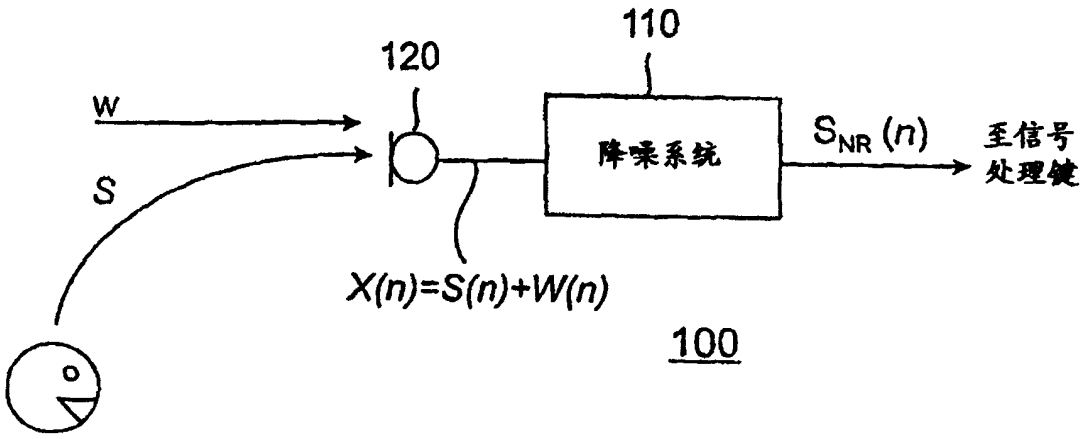

110
120
W
S
降噪系统
$S_{NR}(n)$
至信号处理键
X(n)=S(n)+W(n)
100

图 1

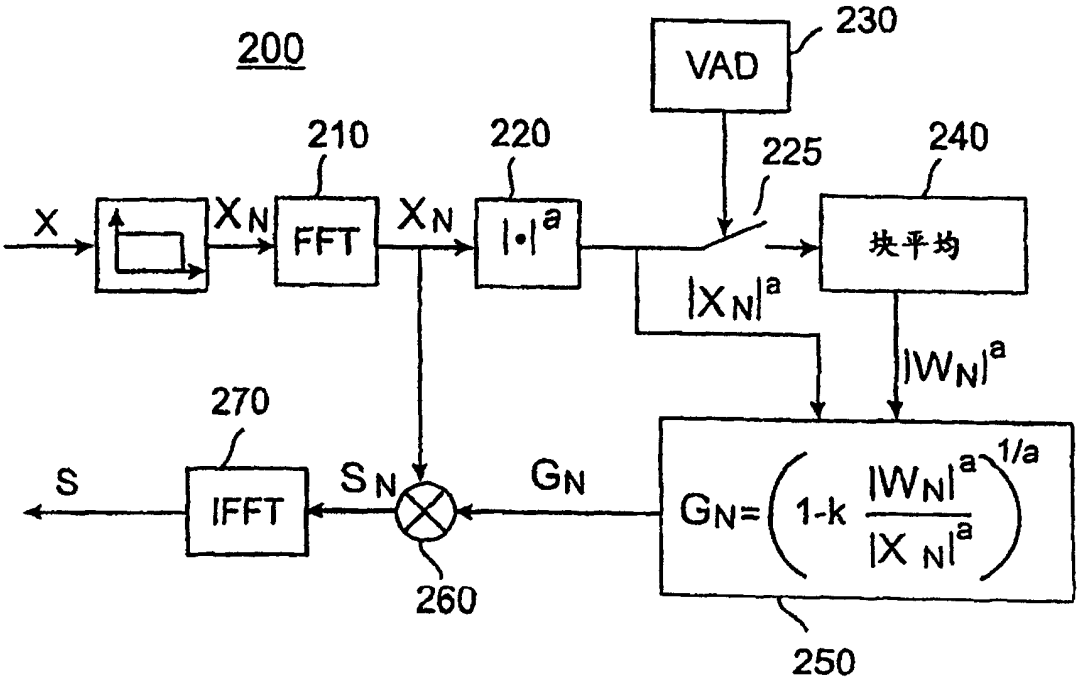

200
VAD
230
210
220
225
240
X
X_N
FFT
X_N
$|\cdot|^a$
块平均
$|X_N|^a$
$|W_N|^a$
270
S
IFFT
S_N
G_N
$G_N=\left(1-k\frac{|W_N|^a}{|X_N|^a}\right)^{1/a}$
260
250

图 2

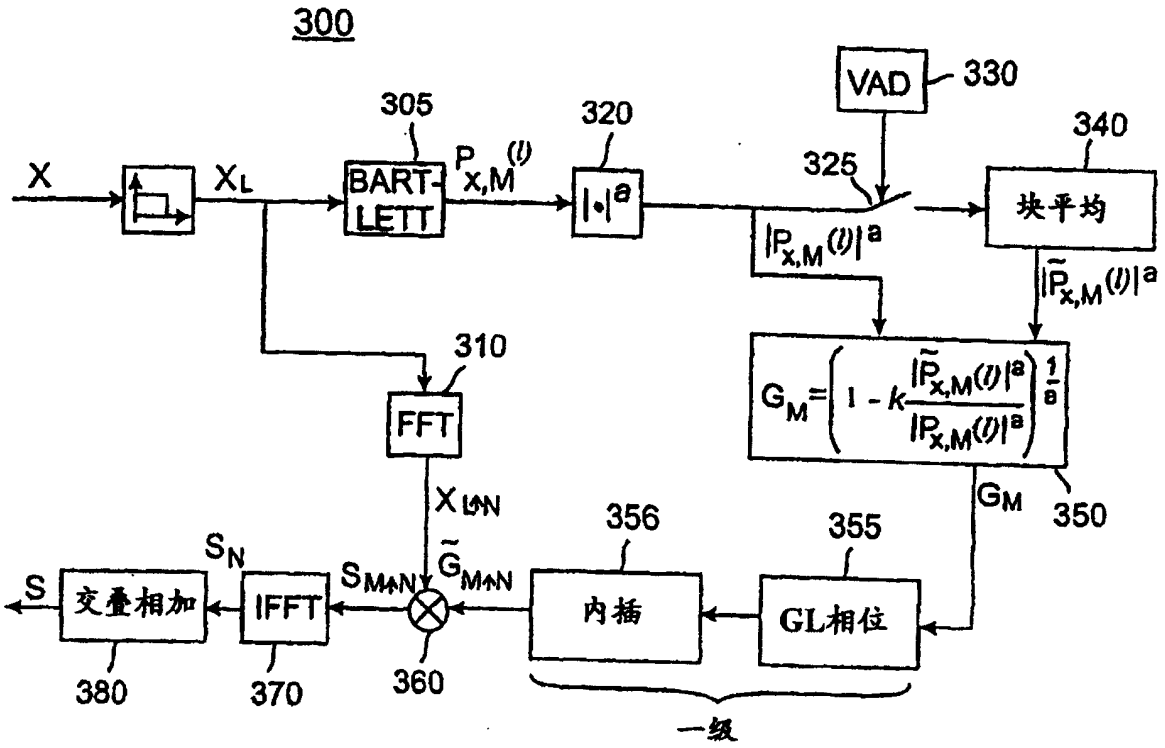

300
VAD
330
305
320
340
X
XL
BART-LETT
$P_{x,M}^{(l)}$
|·|a
325
块平均
$|P_{x,M}(l)|^a$
$|\tilde{P}_{x,M}(l)|^a$
310
$G_M=\left(1-k\frac{|\tilde{P}_{x,M}(l)|^a}{|P_{x,M}(l)|^a}\right)^{\frac{1}{a}}$
FFT
$X_{L\Delta N}$
GM
350
356
355
SN
S
交叠相加
IFFT
$S_{M\Delta N}$
$\tilde{G}_{M\Delta N}$
内插
GL相位
380
370
360
一级

图 3

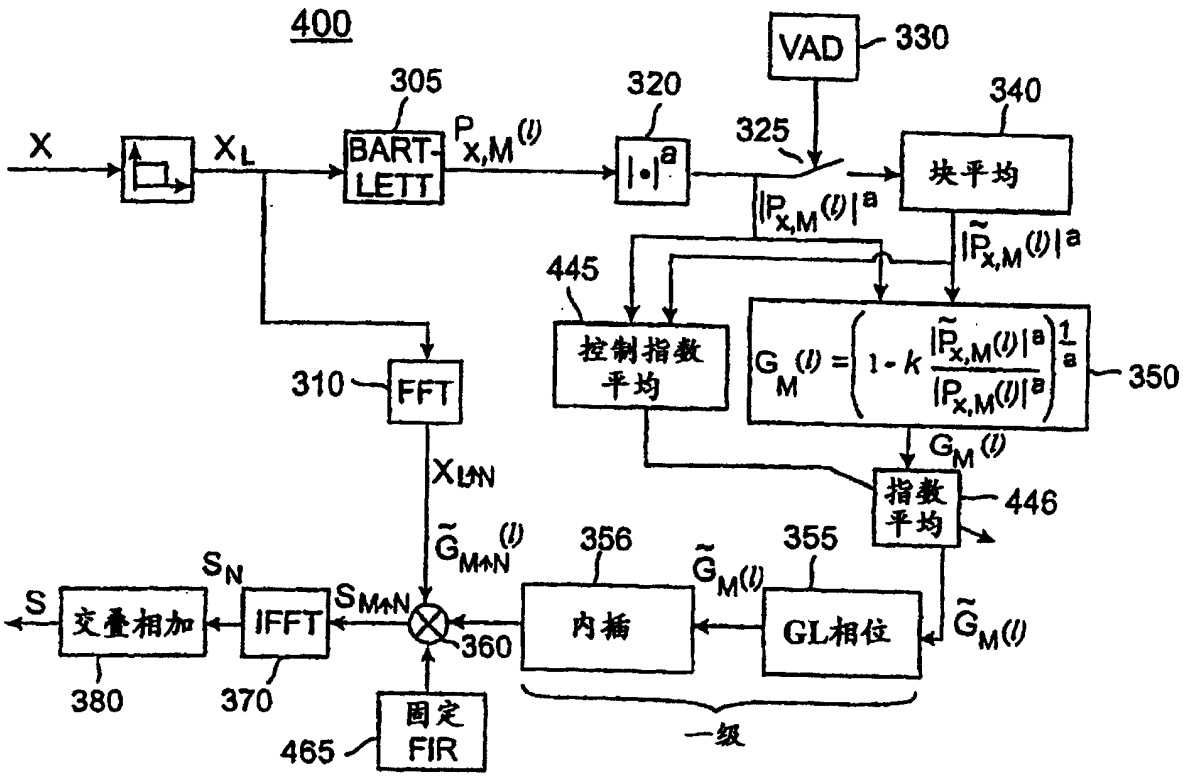

400
VAD
330
305
320
340
X
XL
BART-LETT
$P_{x,M}^{(l)}$
|·|a
325
块平均
$|P_{x,M}(l)|^a$
$|\tilde{P}_{x,M}(l)|^a$
445
控制指数平均
$G_M^{(l)}=\left(1-k\frac{|\tilde{P}_{x,M}(l)|^a}{|P_{x,M}(l)|^a}\right)^{\frac{1}{a}}$
350
310
FFT
$G_M(l)$
$X_{L\Delta N}$
指数平均
446
$\tilde{G}_{M\Delta N}^{(l)}$
356
355
$\tilde{G}_M(l)$
SN
S
交叠相加
IFFT
$S_{M\Delta N}$
内插
GL相位
$\tilde{G}_M(l)$
360
380
370
固定FIR
465
一级

图 4

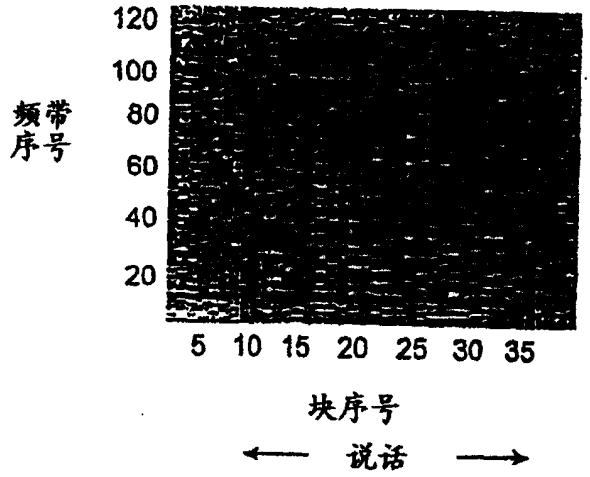


图 5a

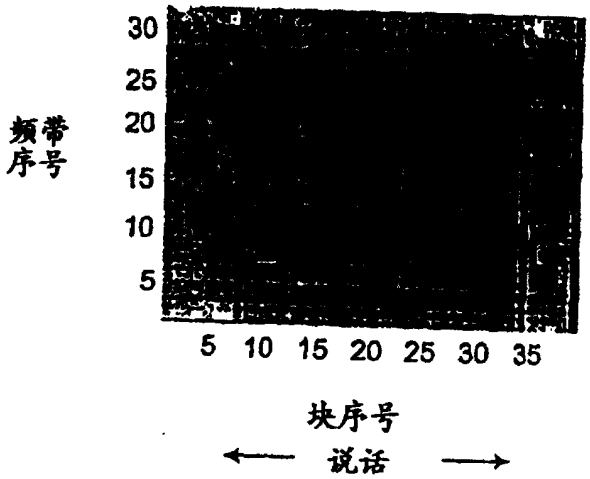


图 5b

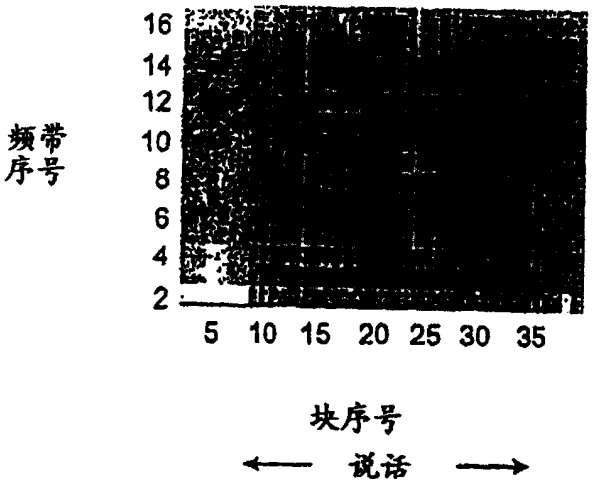


图 5c

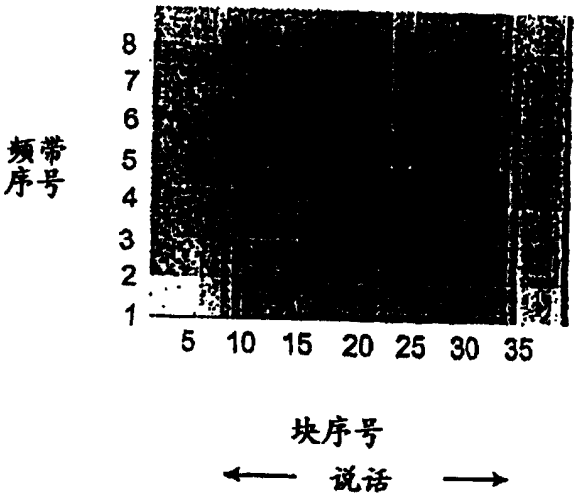


图 5d

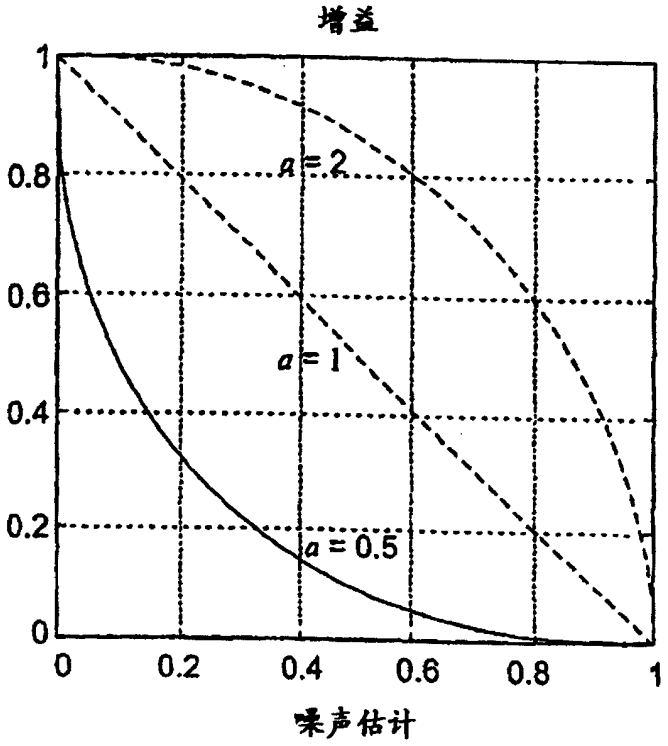

增益
1
0.8
0.6
0.4
0.2
0
a = 2
a = 1
a = 0.5
0
0.2
0.4
0.6
0.8
1
噪声估计

图 6

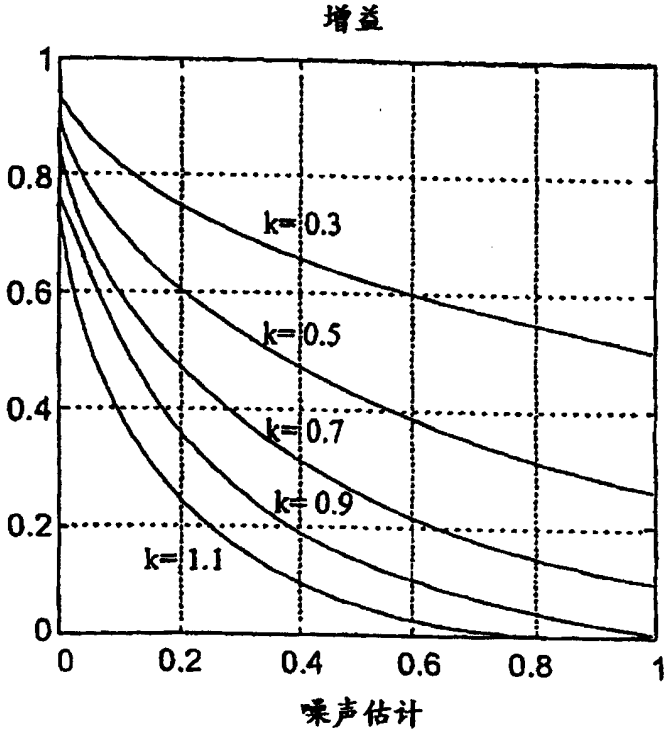

增益
1
0.8
0.6
0.4
0.2
0
k= 0.3
k= 0.5
k= 0.7
k= 0.9
k= 1.1
0
0.2
0.4
0.6
0.8
1
噪声估计

图 7

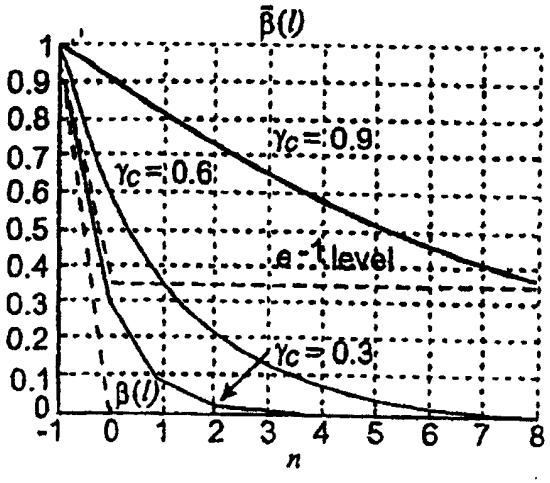


图 8a

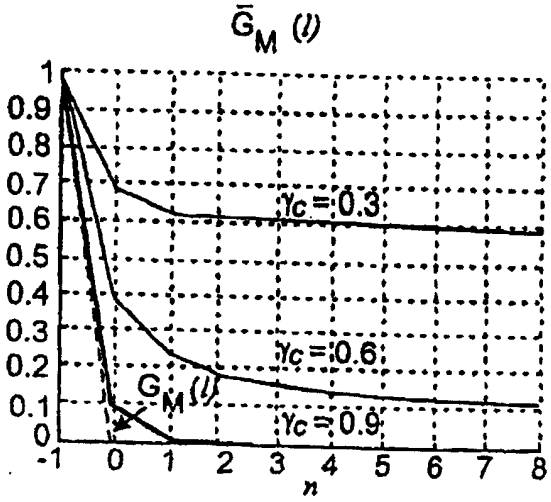


图 8b

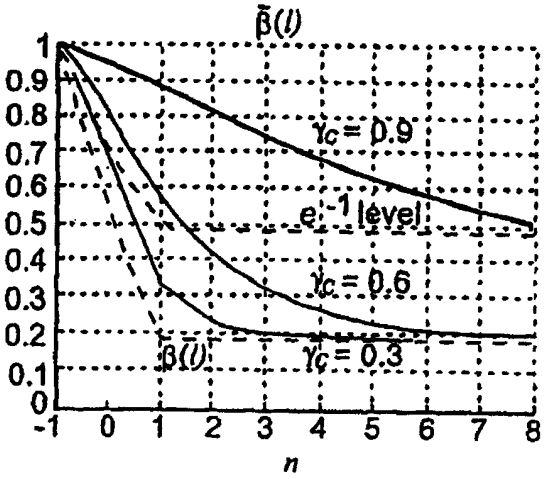


图 8c

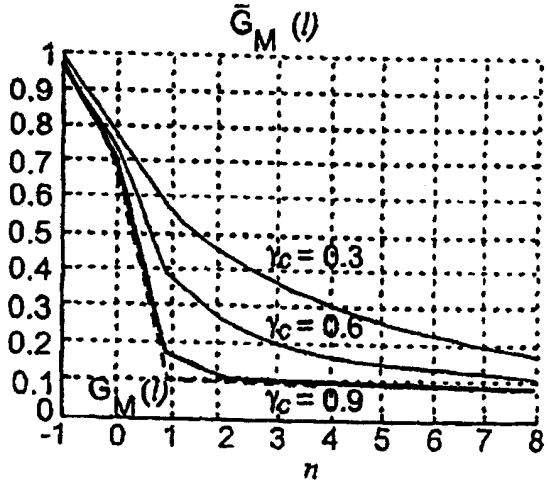


图 8d

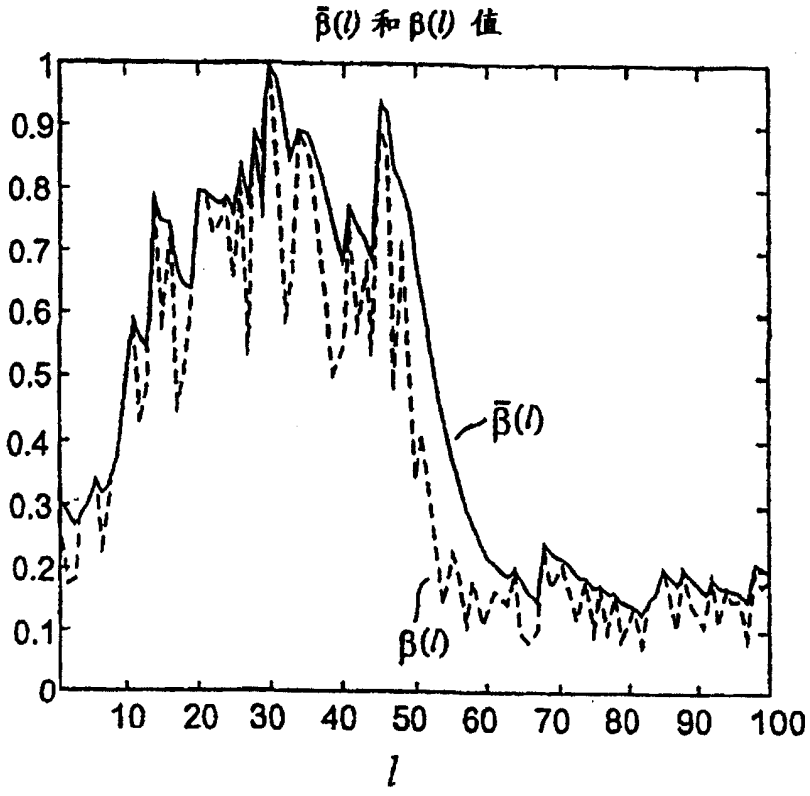

β̄(l) 和 β(l) 值
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10 20 30 40 50 60 70 80 90 100
l
β̄(l)
β(l)

图 9

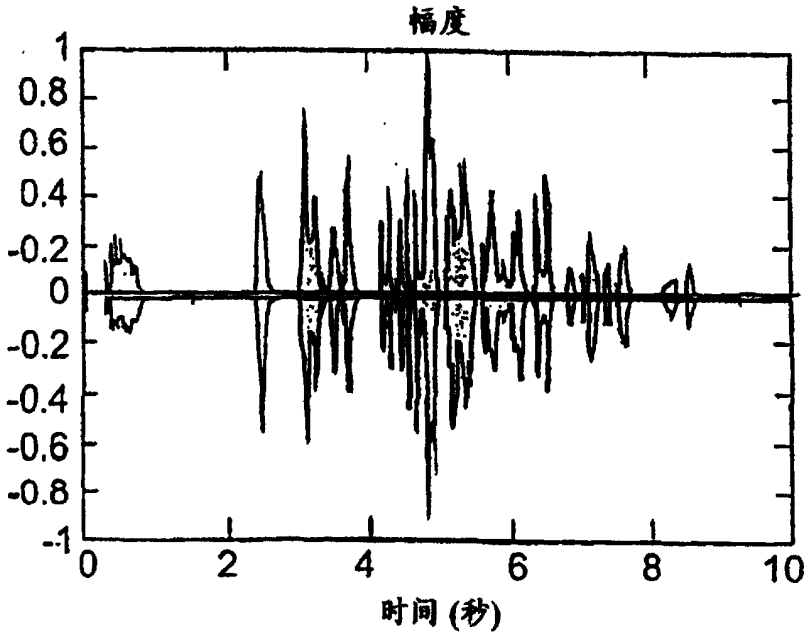

幅度
1
0.8
0.6
0.4
-0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0 2 4 6 8 10
时间 (秒)

图 10

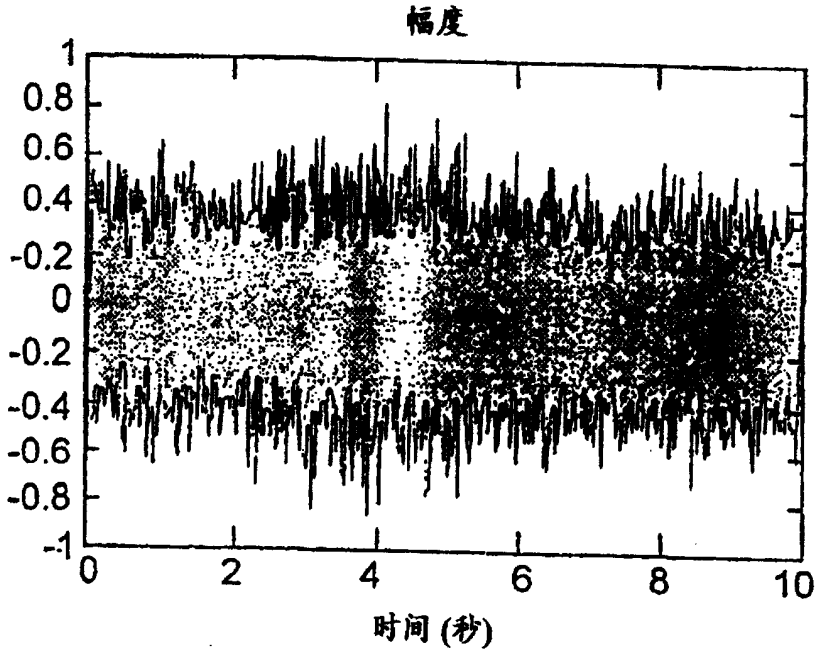

幅度
1
0.8
0.6
0.4
-0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
2
4
6
8
10
时间 (秒)

图 11

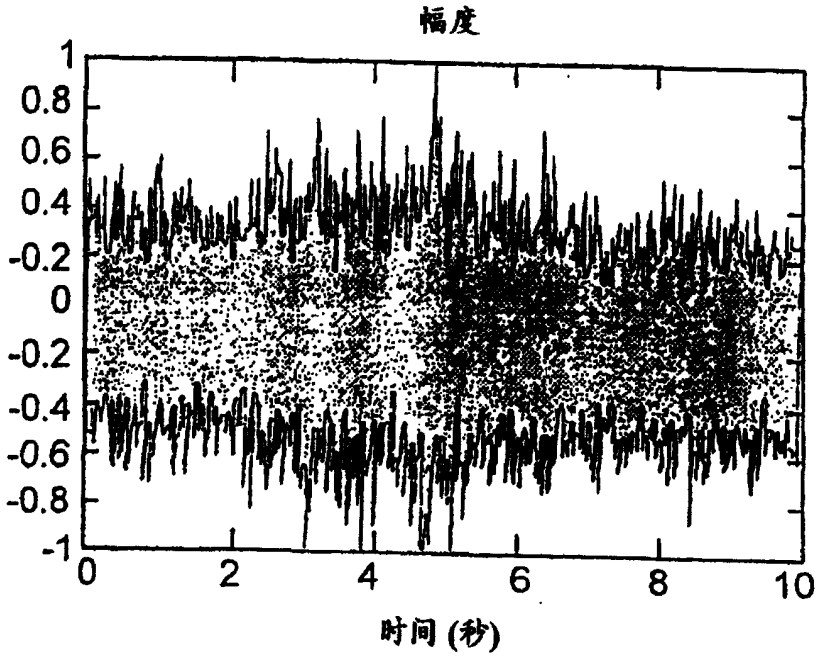

幅度
1
0.8
0.6
0.4
-0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
2
4
6
8
10
时间 (秒)

图 12

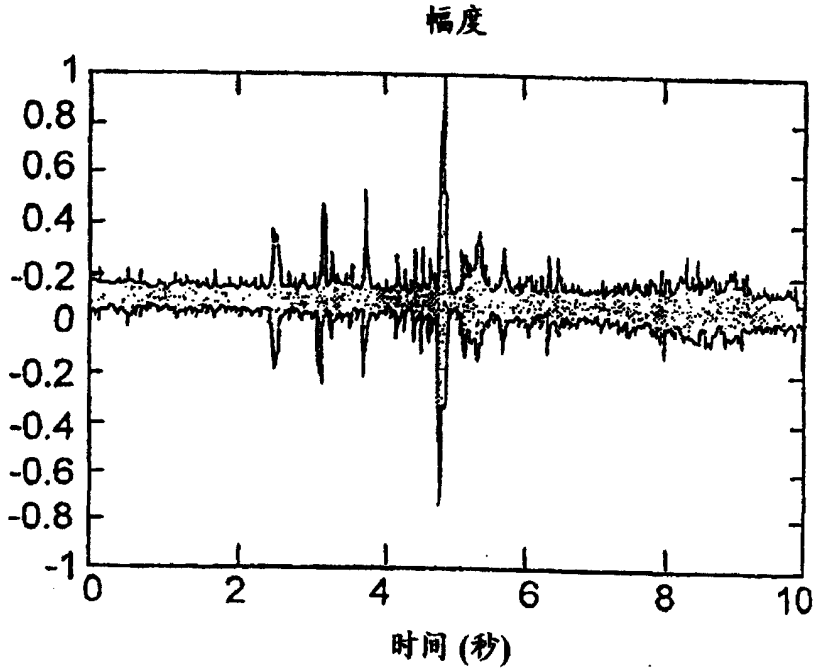

幅度
1
0.8
0.6
0.4
-0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
2
4
6
8
10
时间 (秒)

图 13

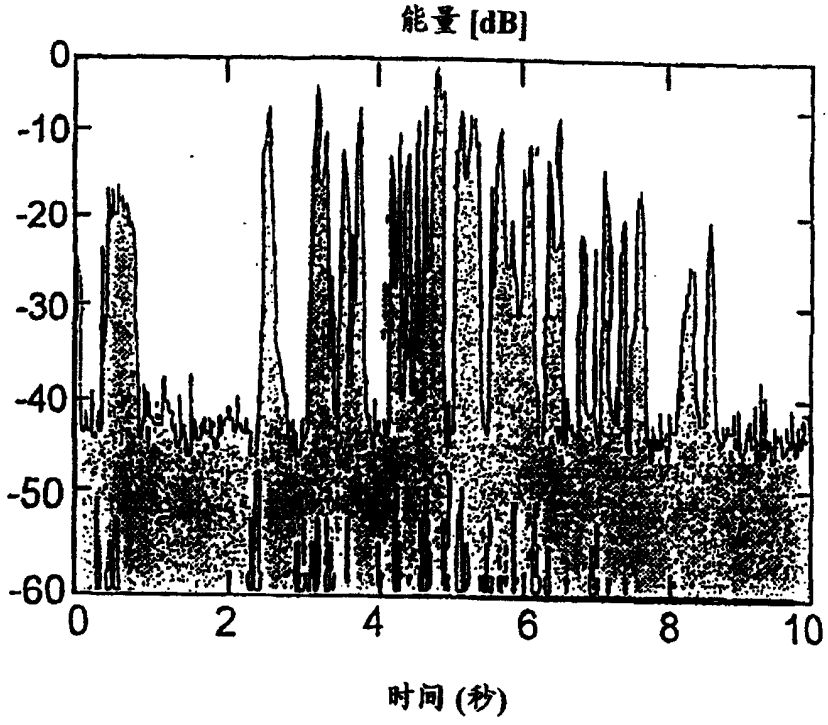

能量 [dB]
0
-10
-20
-30
-40
-50
-60
0
2
4
6
8
10
时间 (秒)

图 14

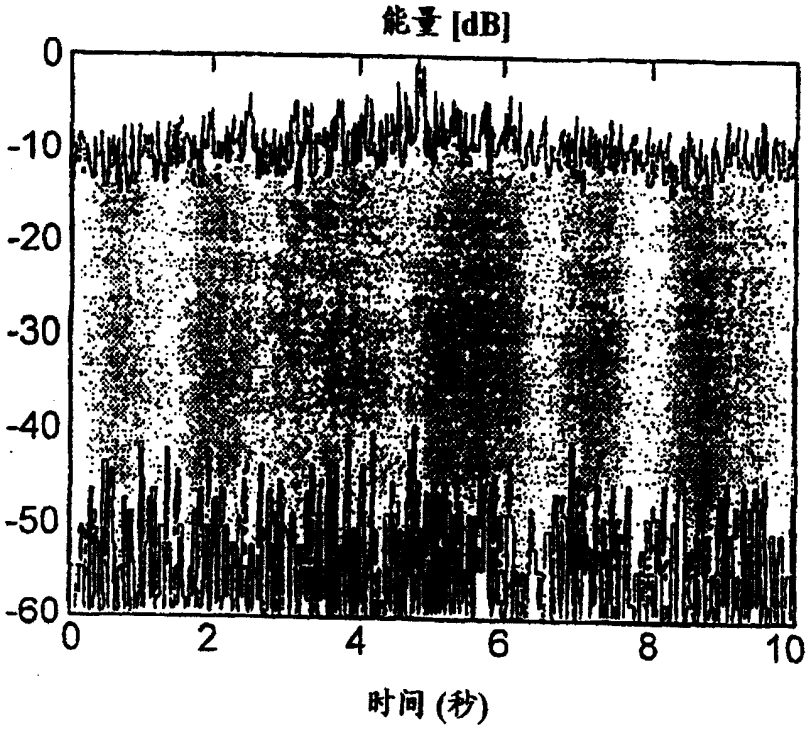

能量 [dB]
0
-10
-20
-30
-40
-50
-60
0
2
4
6
8
10
时间 (秒)

图 15

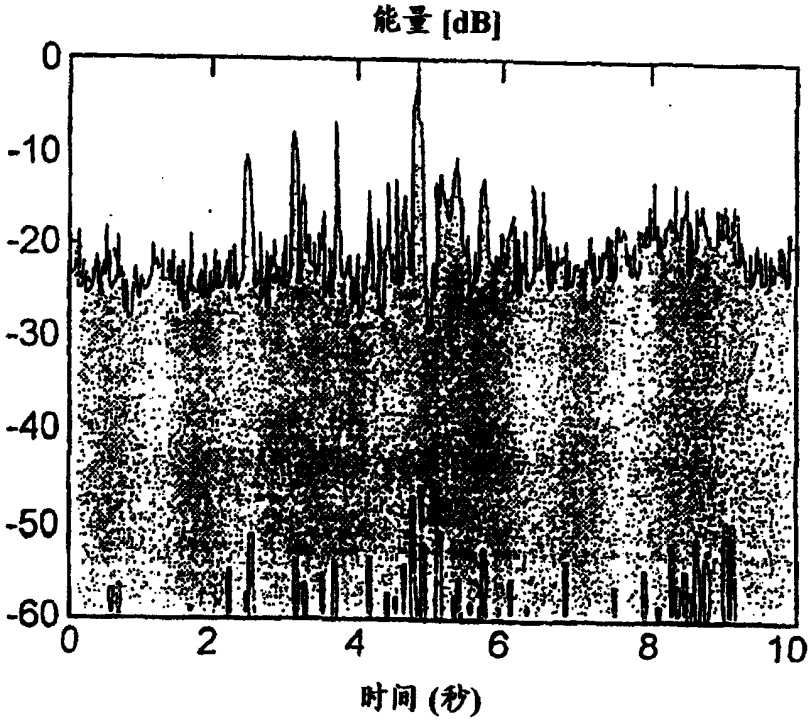

能量 [dB]
0
-10
-20
-30
-40
-50
-60
0
2
4
6
8
10
时间 (秒)

图 16

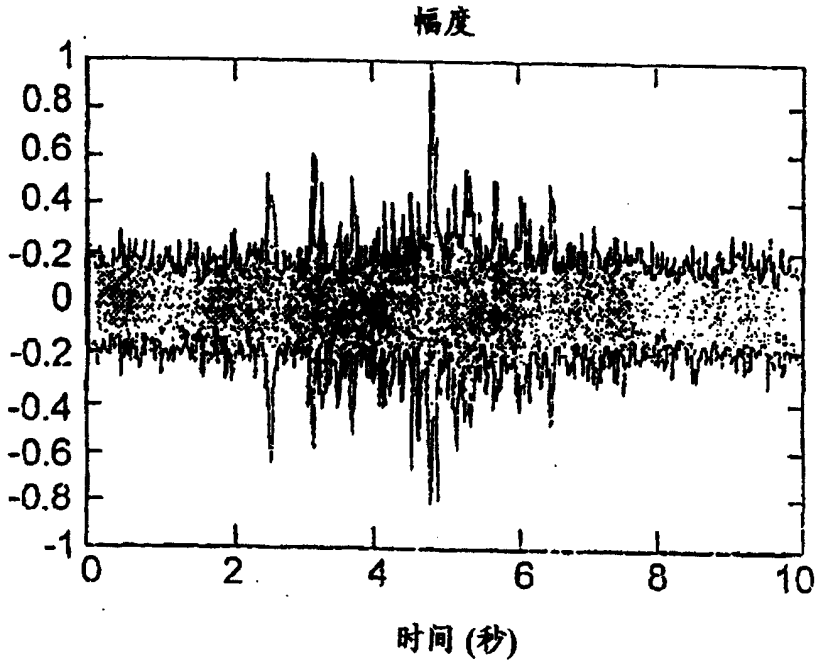

幅度
1
0.8
0.6
0.4
-0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
2
4
6
8
10
时间 (秒)

图 17

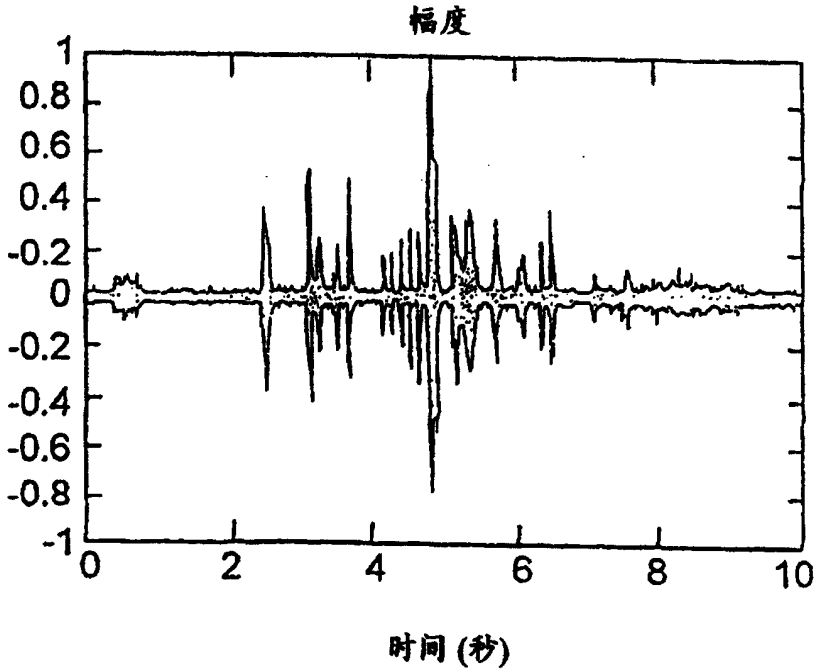

幅度
1
0.8
0.6
0.4
-0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
2
4
6
8
10
时间 (秒)

图 18

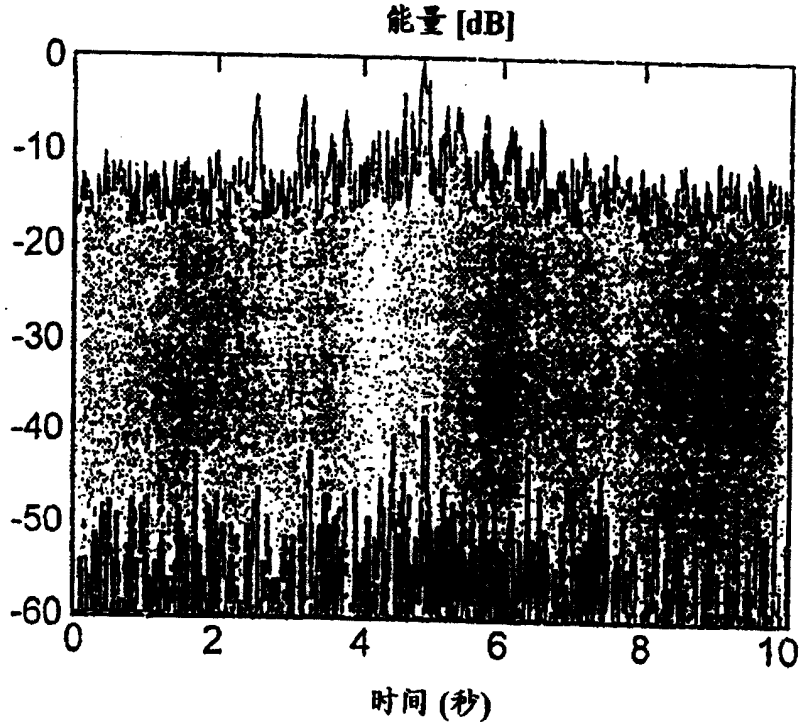

能量 [dB]
0
-10
-20
-30
-40
-50
-60
0
2
4
6
8
10
时间 (秒)

图 19

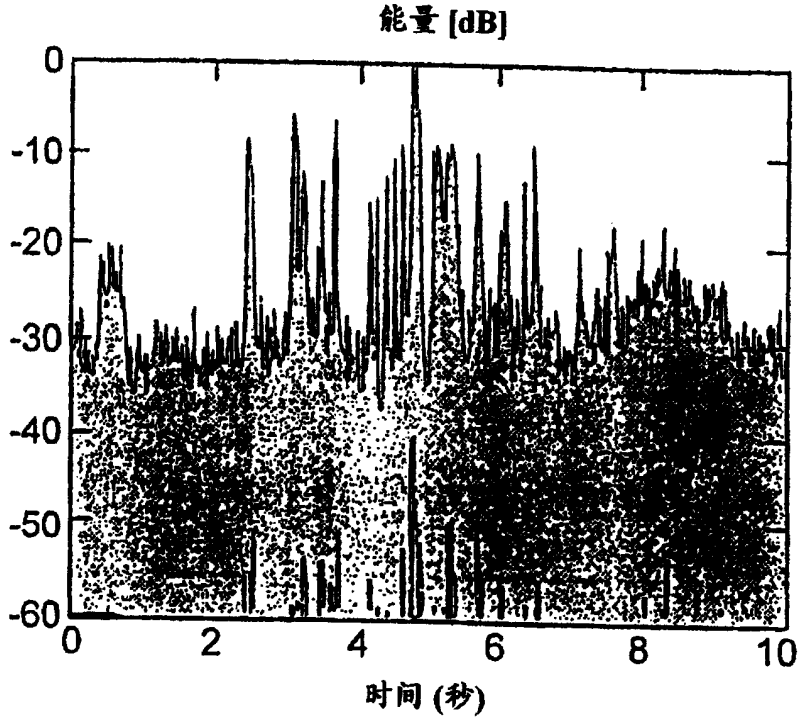

能量 [dB]
0
-10
-20
-30
-40
-50
-60
0
2
4
6
8
10
时间 (秒)

图 20

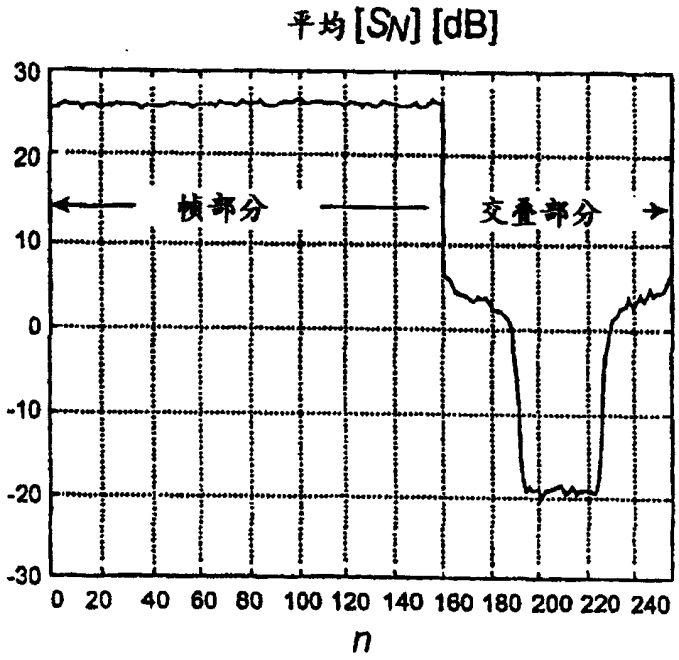

平均 [SN] [dB]
30
20
10
0
-10
-20
-30
帧部分
交叠部分
0 20 40 60 80 100 120 140 160 180 200 220 240
n

图 21

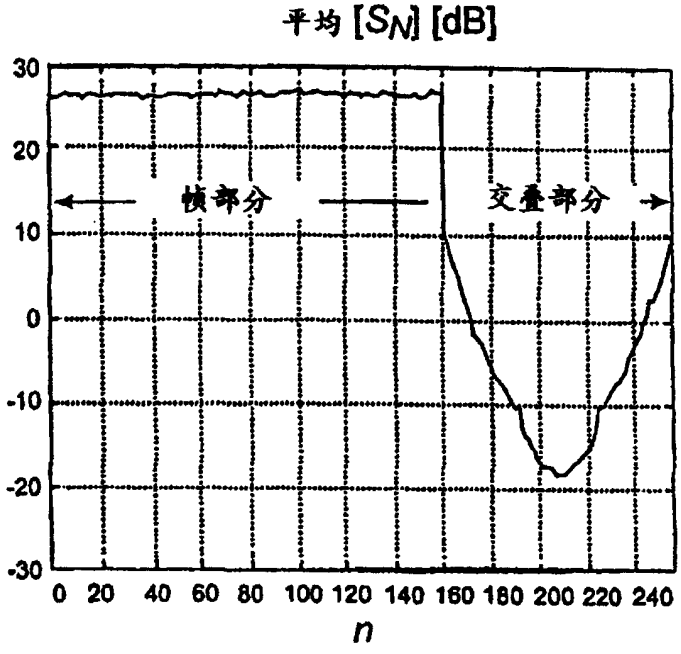

平均 [SN] [dB]
30
20
10
0
-10
-20
-30
帧部分
交叠部分
0 20 40 60 80 100 120 140 160 180 200 220 240
n

图 22

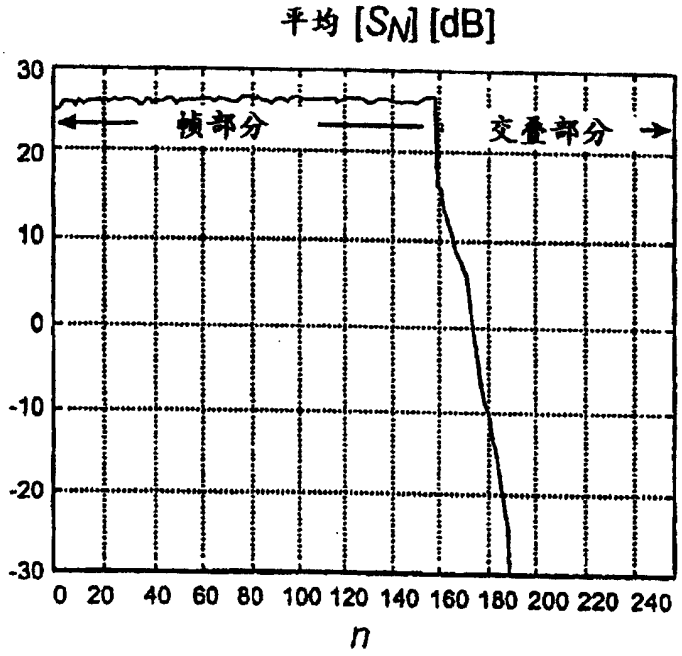

平均 [SN] [dB]
帧部分
交叠部分
30
20
10
0
-10
-20
-30
0 20 40 60 80 100 120 140 160 180 200 220 240
n

图 23

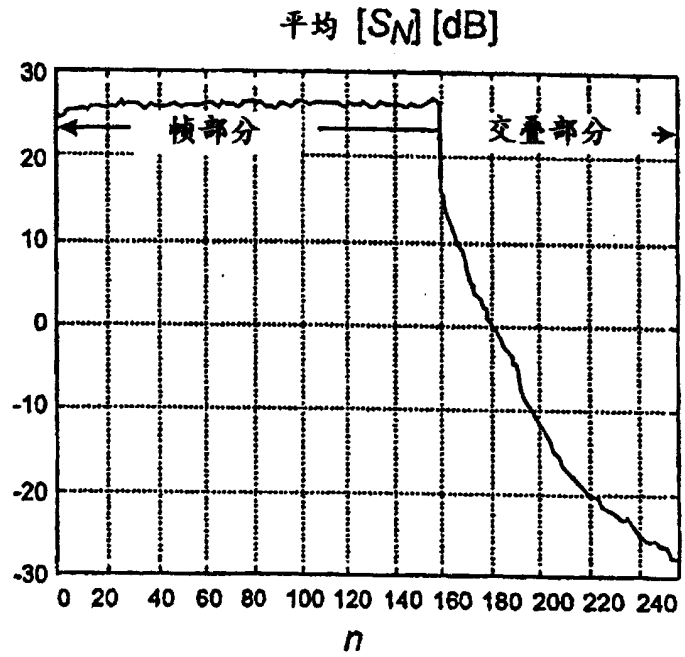

平均 [SN] [dB]
帧部分
交叠部分
30
20
10
0
-10
-20
-30
0 20 40 60 80 100 120 140 160 180 200 220 240
n

图 24

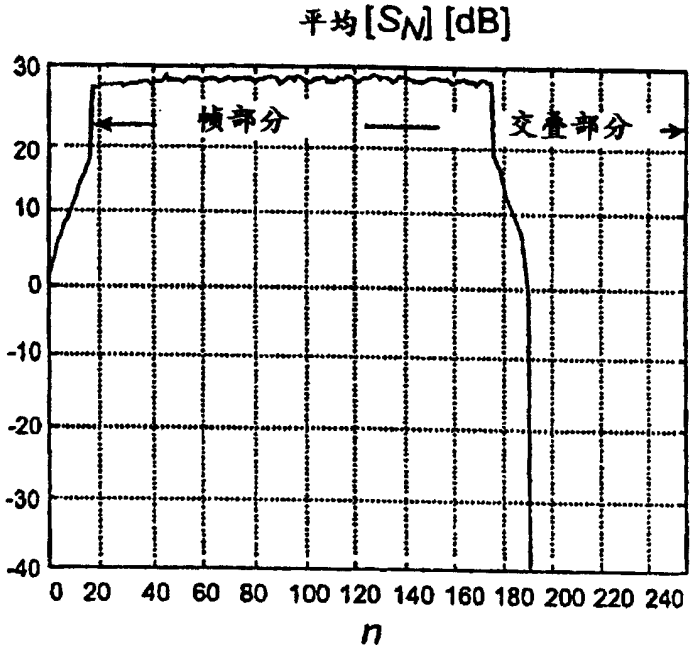

平均[S_N] [dB]
帧部分
交叠部分
30
20
10
0
-10
-20
-30
-40
0 20 40 60 80 100 120 140 160 180 200 220 240
n

图 25

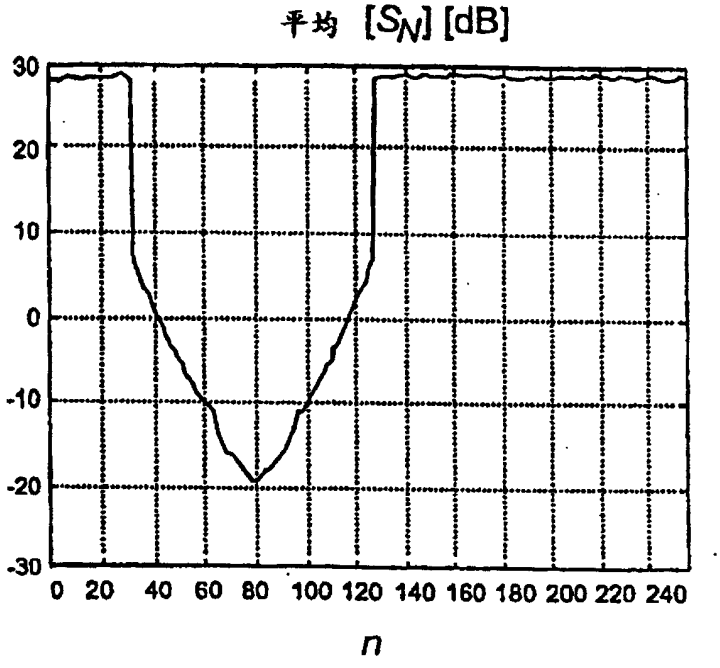

平均 [S_N] [dB]
30
20
10
0
-10
-20
-30
0 20 40 60 80 100 120 140 160 180 200 220 240
n

图 26

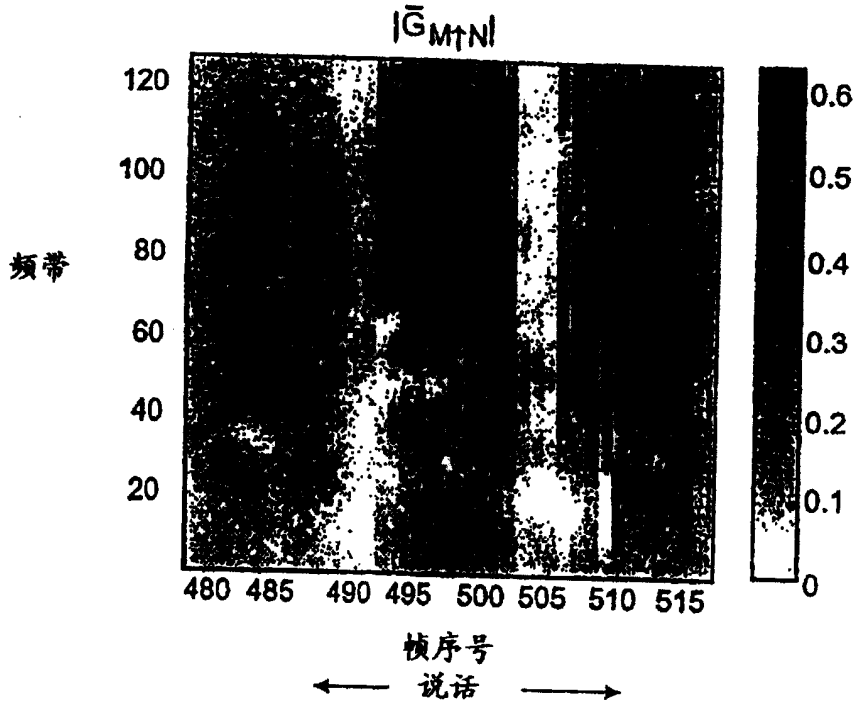

|G̃M↑N|
120
100
80
60
40
20
频带
480 485 490 495 500 505 510 515
帧序号
说话
0.6
0.5
0.4
0.3
0.2
0.1
0

图 27

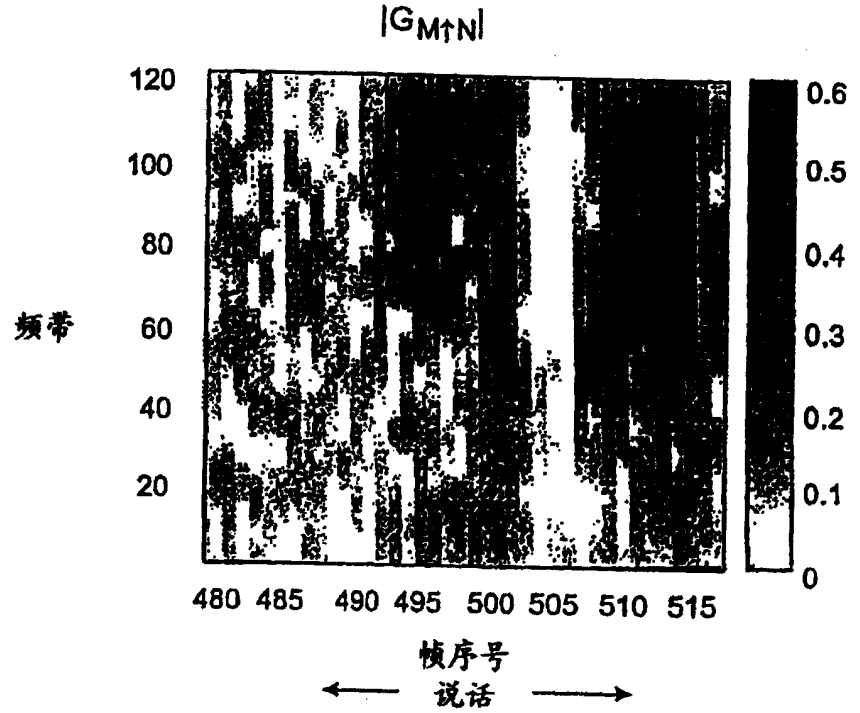

|GM↑N|
120
100
80
60
40
20
频带
480 485 490 495 500 505 510 515
帧序号
说话
0.6
0.5
0.4
0.3
0.2
0.1
0

图 28